UNIVERSITY OF CALIFORNIA
RIVERSIDE

Structure and Dynamics of Tryptophan Synthase Intermediates via NMR-Crystallography Computational Chemistry

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

Doctor of Philosophy in

Chemistry

by

Yangyang Wang

March 2021

Dissertation Committee:

Dr. Leonard J. Mueller, Chairperson
Dr. Chia-En Chang
Dr. Matthew Conley

The Dissertation of Yangyang Wang is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

When I start reflecting on the five years of the graduate school, I would like to express my deepest appreciation to my advisor, Len, for accepting me as your graduate student. You have shown me what is called the outstanding scientist. You are so knowledgeable, insightful, and full of wisdom and humor. Without your unwavering mentorship and support, I cannot be what I am today.

I'd also like to extend my gratitude to my committee, Chia-en and Matthew, for providing constructive advice. Many thanks go to Mike as well, the projects have benefited from your invaluable insights and guidance.

I cannot leave the lab without mentioning Viktoriia. You are such a patient and responsible person. You have not only trained me from scratch, but also offered invaluable contribution. This research would have been impossible without your tireless assistance.

Thanks should also go to all lab mates. Mengyao, thanks for having you as my lab mate and friend. I still miss the time we sat elbow to elbow in the lab. I also had great pleasure of working with Jacob, your profound work cannot be underestimated. Besides, I very much appreciate the help from Bethany and Robert. You extended a great amount of assistance even after you had left the lab. Varun, Kevin, Manpreet, Rittik, Kate, Corbin, thank you for making joy, laughter, and happiness circulating around the lab.

Special thanks to the remarkable friends I met during these five years. To Tianyi and Fan, you were the people who rushed to my mind whenever I encountered the trouble. Thank you for having lent me the hand to overcome the hardships. To Yinan, thank you for having been putting yourself into my shoes to offer relentless advice and help; you are such a reliable friend whom I always trust. To Chen for unparalleled support and guidance. I will miss the open-minded and inspirational conversations with you. I'm deeply indebted to my family. To my parents, thank you for giving me everything. To my grandparents, uncles, and aunts for your unconditional love.

ABSTRACT OF THE DISSERTATION

Structure and Dynamics of Tryptophan Synthase Intermediates Via NMR-Crystallography Computational Chemistry

by

Yangyang Wang

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, March 2021
Dr. Leonard J. Mueller, Chairperson

This thesis presents progresses in the adoption of NMR crystallography – the synergistic combination of X-ray crystallography, NMR spectroscopy, and Ab initio computational approaches – to study structures and dynamics of enzymes. The objective is to develop exceptionally definite and chemically-detailed structures of the chemical active sites by focusing on the intermediates along the pyridoxal-5'-phosphate catalyzed reaction pathway of tryptophan synthase.

NMR can supply direct chemical shifts to certain selected atoms in enzyme-substrate complexes; however, it is extremely sensitive to the local environment, making large complexes, such as enzyme complexes, unable to be interpreted. When the NMR shifts combine with X-ray crystallography, they can provide the framework to build computational models of active sites. The results of Ab initio calculations can reveal the unprecedented level of structural details. This is addressed in particular by the ketoenamine and enolimine tautomerization, in which a proton transfer points to the importance of protonation/deprotonation at ionizable sites on the coenzyme, substrates, and side residues to activate key steps in the catalytic process.

Solid-state NMR suffers from low sensitivity due to low polarization and slow polarization recovery times. With aid of dynamic nuclear polarization at the low temperatures, solid-state NMR can not

only enhance sensitivity, but also capture some kinetic intermediates. These intermediates can be tautomers if solid-state NMR with dynamic nuclear polarization can selectively stabilize a tautomer. The use of NMR crystallography, which builds models for tautomers, provides holistic view for the protonation states and dynamics of tautomers.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1. Introduction

1.1 Spin

Spin and magnetism are two physical but invisible phenomena in matter because electrons and nuclei, which possess spin and magnetism, compose matter. Nonetheless, spin and magnetism are not as apparent as other visible phenomena such as mass. They play an indispensable role in our daily life, with applications in computers, credit cards, magnetic resonance imaging machines, and business equipment. Spin and magnetism are the topics of this thesis.

In physics, angular momentum is produced by the rotation of an object; for example, the pivoting of an electron around the nucleus generates angular momentum, which is normally called electronic orbital angular momentum and usually represented by a quantum number, the integer ℓ. Different ℓ values represent different quantum states: $\ell = 0$ is known as s-orbital, $\ell = 1$ is known as p-orbital, and so on. Aside from the electronic orbital angular momentum, an atom also has two other sources of angular momentum: electron spin and nuclear spin. Spin angular momentum is an intrinsic characteristic not related to any motion in space and is fundamentally different from orbital angular momentum; however, it is seen as a rotation around itself and has the corresponding angular momentum attributes.[1,2] It is also denoted by the vector *I*, and the direction of *I* is called the spin polarization axis. All elementary particles such as protons, neutrons, and electrons carry characteristic spins.

The magnitude of the spin angular momentum is given by[3]

$$|I| = \hbar\sqrt{I(I+1)} \tag{1.1}$$

where *I* is the nuclear spin quantum number and $\hbar = \frac{h}{2\pi}$ (h is Planck's constant).

In a molecule, there are various possible sources of angular momentum: the spinning of electrons around nuclei, the motion of the nuclear framework around the center of mass of the molecule, and in some cases, the rotation of internal molecular groups, the electron spins, and the nuclear spins. However, the nuclear

framework's motion can very often be treated classically: the molecule is treated as an ordinary object rotating in space. Because of the Pauli principle[4] and the quantum rules for chemical bonding, in the lowest energy state of a chemically stable molecule, the electron orbital angular moment and the electron spin angular moment always cancel out, thus only singling out the molecular rotation and nuclear spins as sources of angular momentum in the molecular ground state.

1.2 Magnetism

Magnetism is a class of physical phenomena associated with magnetic fields, as electricity is related to electronic fields. At every point of space, each object is associated with two vectors, *E* and *B*. The field *E* is called the electric field and interacts with electric charges, and the field *B* is called the magnetic field and interacts with magnetic moments. The behaviors of the fields *E* and *B* in time and space are described by Maxwell's equations, a set of coupled partial differential equations that describe how electric and magnetic fields are generated by charges, currents, and changes of the fields. Magnetism and electricity are proved to be different manifestations of the phenomenon of electromagnetism.

Macroscopically, all objects can interact with magnetic fields, and this ability can be quantified by the magnetic energy, which is reversely proportional to the magnetic moment and the *B* field,

$$E_{mag} = -\mu \cdot B. \tag{1.2}$$

Even some substances, such as a bar magnet or a compass needle, have a permanent magnetic moment. Most substances, on the other hand, induce magnetism, indicating that the magnetic moment appears only when an external magnetic field is applied:

$$\mu_{\text{induced}} = \mu_0^{-1} V_\chi B, \tag{1.3}$$

where μ_0 denotes the constant of vacuum permeability and V_χ is the volume of the object. The dimensionless number χ denotes the magnetic susceptibility of the material. If a substance has a positive value of χ, it is called paramagnetic. Most materials have a negative value of χ and are thus called diamagnetic.

From the microscopical perspective, magnetism can have any of the following three origins: (i) circulated electric currents produced by the movement of electrons around the nucleus, (ii) electrons' magnetic moments, and (iii) the nucleus' magnetic moments. Effect (i) may be understood from elementary physics: when an electric current is made to flow in a loop, a magnetic field is generated. (Figure 1.1)

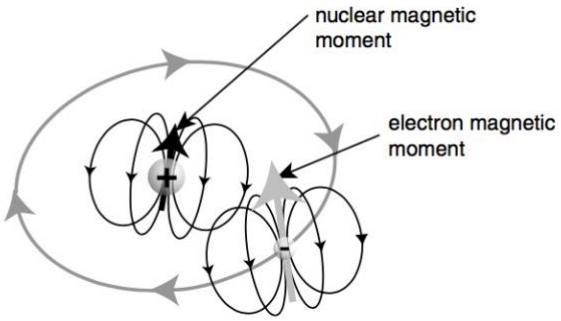


Figure 1.1 Electron and proton spin magnetism in a hydrogen atom, adapted from Levitt5. Effects (ii) and (iii) are intrinsic and the same as intrinsic spin. Magnetism is linearly proportional to spin,6

$$\mu = \gamma I = \gamma\hbar[i(i+1)]^{1/2}, \qquad (1.4)$$

where I is spin angular momentum, μ is magnetic momentum, the proportionality constant γ is called the isotope-specific gyromagnetic ratio, and i is a half-integer value representing the quantization of rotational energy levels in a given system.

In the case of effect (i), each electron produces a small magnetic field by orbiting the nucleus of an atom; however, the orbiting direction is random, making the angular momentum evenly distributed in all directions. Therefore, if there is no external perturbation, atoms with paired electrons are not magnetic. When an atom has unpaired electron spin, an entirely different phenomenon, namely, unpaired electron spin line up in parallel with each other in a domain, is very much likely to occur. This is called ferromagnetism or anti-ferromagnetism depending on how the magnetic moments align with the neighboring moments. Magnetism of electrons or effect (i) is beyond the scope of this thesis and will not be discussed here.

For effect (ii), the relatively weak interaction of electron magnetic moments allows electron paramagnetic resonance experiments to be conducted. The majority of substances have no electron magnetic moment in the electronic ground state.

Effect (iii) is the topic of this thesis. The magnetic moments of the atomic nucleus have weak diamagnetism ($\chi < 0$), which arise from the electron orbital currents induced by the applied magnetic field, enabling nuclear magnetic resonance (NMR) experiments to be conducted. The following chapters will focus on NMR.

1.3 Dynamics of Nuclear Spin

1.3.1 Spin Hamiltonian Hypothesis

In Dirac notation[7], the quantum state of the entire sample is fully described by a wave function $|\Psi_{\text{full}}\rangle$ that contains information on the positions, velocities, and spin states of all the electrons and nuclei. This wavefunction obeys the time-dependent Schrodinger equation,[8,9]

$$\frac{d}{dt}|\Psi_{\text{full}}(\text{t})\rangle = -i\mathcal{H}_{\text{full}}|\Psi_{\text{full}}(\text{t})\rangle, \tag{1.5}$$

where Hamiltonian $\mathcal{H}_{\text{full}}$ contains all interactions in the system. By contrast, in the NMR of the diamagnetic substance, the spin states of all the electrons will be blurred out, leaving out only the nuclear spin states,

$$\frac{d}{dt}\left|\Psi_{\text{spin}}(\text{t})\right\rangle \cong -i\mathcal{H}_{\text{spin}}\left|\Psi_{\text{spin}}(\text{t})\right\rangle, \tag{1.6}$$

where $\left|\Psi_{\text{spin}}(\text{t})\right\rangle$ represents the spin state of the nuclei, and $\mathcal{H}_{\text{spin}}$ is the nuclear spin Hamiltonian. The nuclear spin Hamiltonian contains only terms that depend on the directions of the nuclear spin polarization, which assumes that the magnetic and electrical influences of the rapidly moving electrons are blurred out so that only their average is seen.

This massive simplification is called the spin Hamiltonian hypothesis. It rests on a time-scale separation of nuclear and electronic motions. Electronic motions are so rapid that the nuclear spin only senses the time average of the fields they generate. Furthermore, nuclear spin energies are assumed to be too small to affect the motions of the electrons within the molecules or the molecules' motions themselves. Practically, the nuclear spin Hamiltonian is a secure concept for almost all systems at ordinary temperatures.

An atomic nucleus is seen as a small lumpy magnet with a magnetic moment and a non-uniform distribution of positive electric charges to interact with its environment. Having a magnetic moment enables it to interact with magnetic fields. Having an electric charge likewise makes interaction with electric fields possible. The nucleus interacts with its surrounding fields in two ways: moving and rotating.

The NMR leverages the rotation of nuclei. If the nucleus rotates, the nuclear magnetic moment and nuclear electric charges will rotate accordingly, differing the orientation with respect to the surrounding fields and changing the energy of the nucleus. The nuclear spin Hamiltonian uses two terms to describe nuclear energy change with respect to the orientation: (i) an electric spin Hamiltonian, which describes the way the nuclear electric energy changes as the nucleus rotates, and (ii) a magnetic spin Hamiltonian, which depicts the way the nuclear magnetic energy changes as the nucleus rotates. The spin Hamiltonian operator for nucleus I_j may thus be written as

$$\mathcal{H}_j = \mathcal{H}_j^{\mathrm{ele}} + \mathcal{H}_j^{\mathrm{mag}}. \tag{1.7}$$

Spin-1/2 nuclei are the focus of this thesis, in which the distribution of positive electric charges is even, and there are no electric energy terms that rely on the orientation or internal structure of the nucleus, that is, $\mathcal{H}_j^{\mathrm{ele}} = 0$; therefore, the nuclear spin energy $\mathcal{H}_j$ for the nucleus I_j is only involved in the magnetic spin Hamiltonian $\mathcal{H}_j^{\mathrm{mag}}$. Thereby, Eq.(1.7) can be simplified by Eq.(1.2) as

$$\mathcal{H}_j = \mathcal{H}_j^{\mathrm{mag}} = -\mu \cdot B = \gamma_i\left(B_x \hat{I}_{jx} e_x + B_y \hat{I}_{jy} e_y + B_z \hat{I}_{jz} e_z\right) \tag{1.8}$$

The electric and magnetic fields that can influence the spin may stem from external equipment and the sample itself. For spin-1/2, the external spin interactions are between the spin and external magnetic fields; internal spin interactions are purely the spin's interaction with the magnetic fields produced within

the molecule. In the case of spin >1/2, electric quadrupolar interaction is also involved. What characterizes NMR is external interactions that are intentionally produced to be considerably larger than internal interactions. The reasons will be discussed in the following sections.

1.3.2 External Spin Hamiltonian

In the NMR spectrometer, there are two primary external magnetic fields: (i) B_0, a quite strong, homogeneous, and static magnetic field generated by a superconducting solenoid and (ii) B_1, a radiofrequency(RF) oscillating field provided by the radiofrequency coil on the probe. The external spin Hamiltonian $\mathcal{H}_{\text{ext}}$ is time-dependent and written as

$$\mathcal{H}_{\text{ext}}(\text{t}) = \mathcal{H}_{\text{static}}(\text{t}) + \mathcal{H}_{\text{RF}}(\text{t}), \tag{1.9}$$

where

$$\begin{aligned} \mathcal{H}_{\text{static}}(\text{t}) &= \sum_j \mathcal{H}_{\text{j}}^{\text{static}} \\ \mathcal{H}_{\text{RF}}(\text{t}) &= \textstyle\sum_{\text{j}} \mathcal{H}_{\text{j}}^{\text{RF}}. \end{aligned} \tag{1.10}$$

Sums are taken over all spins in the sample. Here, $\mathcal{H}_j^{\text{static}}$ is the interaction of each spin I_j with the longitudinal static field B_0, and $\mathcal{H}_j^{\text{RF}}$ is the interaction of each spin with the radiofrequency field B_1. By standard, a laboratory reference frame is usually chosen in which the static field is along the z-plane, and each spin Hamiltonian for the interaction with the static longitudinal field is given by

$$\mathcal{H}_j^{\text{static}} = -\gamma_j B_0 I_{jz} \tag{1.11}$$

This is called the nuclear Zeeman interaction.

The radiofrequency coil generates a field $B_{\text{RF}}(\text{t})$ along the tilted axis, as shown in Figure (1.2).

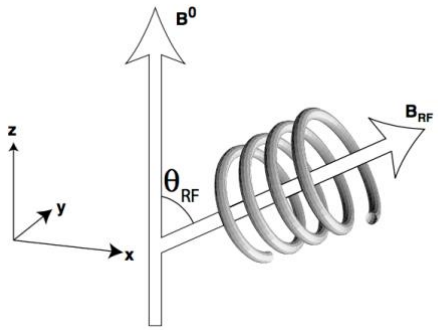


Figure 1.2 The laboratory frame, the radiofrequency coil, and the applied fields B0 and BRF, adapted from Levitt5

This field oscillates at the spectrometer reference frequency ω_{ref} during the radiofrequency; otherwise, it is equal to zero. If the pulse is perfectly rectangular, then the radiofrequency field is represented as

$$B_{\mathrm{RF}}(\mathrm{t}) = \begin{cases} B_{\mathrm{RF}}(\mathrm{e_z cos\theta_{RF}} + \mathrm{e_x sin\theta_{RF}})\cos\left(\omega_{\mathrm{ref}}\mathrm{t} + \phi_{\mathrm{p}}\right) & \text{during an radiofrequency pluse} \\ 0 & \text{otherwise} \end{cases}, \quad (1.12)$$

where ϕ_{p} is the angle between the z-axis and the xy-plane.

1.3.3 Internal Spin Hamiltonian

In addition to the external magnetic and electric fields, the nuclei are also affected by the internal magnetic and electric fields originating from themselves. These interactions are denoted by the internal spin Hamiltonian $\mathcal{H}_{\mathrm{int}}$ and contain the following terms:

1. Chemical shift. This term is generated with the involvement of the external magnetic field but is directly created by the electrons themselves. The chemical shift Hamiltonian $\mathcal{H}_j^{\mathrm{CS}}$ derived from Eq.(1.2) for spin I_j is written as

$$\mathcal{H}_j^{\mathrm{CS}} = -\mathcal{u}_j \cdot B_j^{\mathrm{induced}}, \quad (1.13)$$

where $\mathcal{u}_j$ is the magnetic moment for spin I_j, and $B_{\mathrm{j}}^{\mathrm{induced}}$ is the indirectly induced magnetic field from the external magnetic field B_0, which can be expressed as

$$B_j^{\text{induced}} = \delta^j \cdot B^0. \tag{1.14}$$

δ^j *is* usually a non-diagonal symmetrical 3×3 matrix representing the influence that includes the direction and the magnitude from B^0 to B_j^{induced}. The direction of the magnetic field may not coincide with a principal axis direction, thus producing a chemical shift anisotropy tensor:

$$\delta^j = R^j(\Theta) \cdot \begin{pmatrix} \delta_{XX}^j & 0 & 0 \\ 0 & \delta_{YY}^j & 0 \\ 0 & 0 & \delta_{zz}^j \end{pmatrix} \cdot R^j(\Theta)^{-1} \tag{1.15}$$

where $R^j(\Theta)$ is a 3×3 rotation matrix that converts the non-diagonal symmetrical matrix δ^j to a diagonal matrix called the chemical shift anisotropy tensor.

2. Quadrupolar couplings. These interactions only appear in the presence of spin > 1/2 nuclei. It is not the topic of this thesis.

- Direct dipole–dipole couplings. It is also called dipolar coupling. This interaction results from the classical through-space spin–spin coupling. These couplings build additional local fields at the nucleus. The dipolar coupling may be either intramolecular or intermolecular. Its magnitude depends primarily on the internuclear distance between the coupled spins, yet its influence on the NMR spectrum also depends on the orientation of this internuclear vector with respect to B^0. The Hamiltonian for a two-spin system, *I* and *S*, the dipolar coupling Hamiltonian, can be expressed as

$$\mathcal{H}_{IS}^{\text{DD,full}} = 2\gamma_I\gamma_S\hbar I\widehat{D}_{IS}S, \tag{1.16}$$

where $\widehat{D}_{IS}$ is the anisotropic dipolar coupling tensor between *I* and *S*.

For a homonuclear dipole coupled system, where I_i and I_j are of the same isotopic type but with different chemical shifts, the secular part of the dipolar Hamiltonian can be written as

$$\mathcal{H}_{I_iI_j}^{\text{DD,homo}} = \frac{\gamma_I\gamma_S}{r_{IS}^3}\frac{1}{2}(3cos^2\theta - 1)\left\{\underbrace{2I_iI_j}_{\text{static}} - \frac{1}{2}\underbrace{\left(I_+^iI_-^j + I_-^iI_+^j\right)}_{\text{flip-flop}}\right\}. \tag{1.17}$$

Note that in the above expression, θ corresponds to the angle of the internuclear vector connecting the two spins with respect to the magnetic field B_0. In Figure (1.3), e_{ik} represents the internuclear vector connecting the two spins.

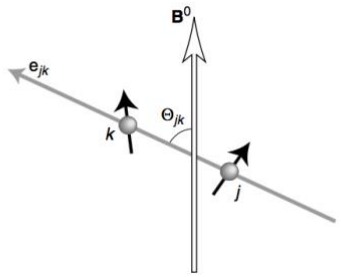


Figure 1.3 The angle θik is used in the calculation of the secular dipole-dipole coupling. Adapted from Levitt5

Eq.(1.18) has two terms in the square brackets. The static spin term attributes to the interaction between vertical component of each nuclei's spin. The vertical direction normally aligns with the z-axis, B_0. This dipolar contribution increases or decrease from the magnitude of the external field felt at the nucleus, hence impacting its local precession frequency. The flip-flop spin term represents the interaction between the transverse magnetization components of spins. when the two angular frequencies are sufficiently close, orientation of the nuclear spin will change. It can only easily occur when the system is homonuclear. As each nucleus establishes its own rotating field within B_0, the energy exchange occurring via flip-flop is effective only in the case of conserved, such as when it commutes with $\mathcal{H}_Z$.

The secular component of the interaction Hamiltonian for a heteronuclear dipole-coupled spin system takes the following form:

$$\mathcal{H}_{IS}^{\mathrm{DD,heter}} = \frac{\gamma_I \gamma_S}{r_{IS}^3}(3cos^2\theta - 1)I_z S_z \qquad (1.18)$$

This is exactly similar to the homonuclear dipolar Hamiltonian; however, the flip-flop term does not commute with I_Z. The difference in Larmor frequencies of two unique spins, *I* and *S*, is usually much more significant than that of the dipolar interaction $\mathcal{H}_{IS}^{\mathrm{DD,heter}}$. Therefore, the feasibility of the occurrence of an energy-conserving exchange is negligible.

- J-couplings. These represent indirect spin–spin interactions through the involvement of electrons. The full form of the intramolecular J-coupling interaction between spins I_j and I_k in a molecule is

$$\mathcal{H}_{jk}^{J} = 2\pi I_j \cdot J_{jk} I_k, \tag{1.19}$$

where J_{jk} is the J-coupling 3×3 real tensor whose form is similar to δ^j in Eq.(1.15) and is written as

$$\mathcal{H}_{jk}^{J} = 2\pi\begin{pmatrix}\hat{I}_{jx} & \hat{I}_{jy} & \hat{I}_{jz}\end{pmatrix} \cdot \begin{pmatrix} J_{xx}^{jk} & J_{xy}^{jk} & J_{xz}^{jk} \\ J_{yx}^{jk} & J_{yy}^{jk} & J_{yz}^{jk} \\ J_{zx}^{jk} & J_{zy}^{jk} & J_{zz}^{jk} \end{pmatrix} \cdot \begin{pmatrix} \hat{I}_{kx} \\ \hat{I}_{ky} \\ \hat{I}_{kz} \end{pmatrix}. \tag{1.20}$$

The mathematical forms of nuclear spin interactions are quite intricate. Fortunately, using simplified form of the internal Hamiltonian is usually possible. This is because the strong external magnetic field makes secular approximation possible, and the rapid molecular motion in liquids, gases, and some solids leads to motional averaging.

In most solids, atomic motion is highly restricted, and very little averaging of internal spin interactions exists. Both intramolecular and intermolecular spin interactions survive, and the internal spin Hamiltonian terms depend on the position of the sample with respect to the magnetic field.

1.4 Quantum Statistical Mechanics and Spin Density Operator

According to the quantum theory of angular momentum, angular momentum operators are Hermitian, and the eigenstate of $\hat{I}_z$ is $|\ell, m\rangle$. It obeys

$$\hat{I}_z|\ell, m\rangle = m|\ell, m\rangle, \tag{1.21}$$

where ℓ and m are quantum numbers and follow

$$\ell = 0,1,2\cdots,$$

$$m = -\ell, \ell+1, \ell+2\cdots, +\ell$$

Spin-1/2 is most common in organic chemicals, such as ^{13}C, ^{15}N, ^{1}H, and ^{31}P. One-spin systems, in which isolated nuclei are assumed not to interact with one another, are a good premise for the ensemble system of spin -1/2. A single spin-1/2 system has two eigenstates of the angular momentum along the z-axis, named Zeeman eigenstates $|\alpha\rangle$ and $|\beta\rangle$. They obey the following:[10]

$$\begin{aligned} \hat{I}_Z|\alpha\rangle &= +\frac{1}{2}|\alpha\rangle \\ \hat{I}_Z|\beta\rangle &= -\frac{1}{2}|\beta\rangle \end{aligned} \tag{1.22}$$

If the magnetic field is along the z-axis, and its magnitude is B_0, the spin Hamiltonian of the nucleus j is proportional to $\hat{I}_{jz}$:

$$\mathcal{H}_j^0 = -\gamma_j B_0 \hat{I}_{jz} = \omega^0 \hat{I}_{jz}, \tag{1.23}$$

where the (chemically shifted) Larmor frequency is given by $\omega^0 = -\gamma_j B_0$. The states $|\alpha\rangle$ and $|\beta\rangle$, which are eigenstates of the spin Hamiltonian, if j is ignored, obey the eigenequations of

$$\begin{aligned} \mathcal{H}_j^0|\alpha\rangle &= \omega^0 \hat{I}_{jz}|\alpha\rangle = +\frac{1}{2}\omega^0|\alpha\rangle \\ \mathcal{H}_j^0|\beta\rangle &= \omega^0 \hat{I}_{jz}|\beta\rangle = -\frac{1}{2}\omega^0|\beta\rangle. \end{aligned} \tag{1.24}$$

The eigenvalues ± ω^0 are two state energies. The energy level splitting of the spin in the magnetic field is known as Zeeman splitting and is shown in Figure (1.4).

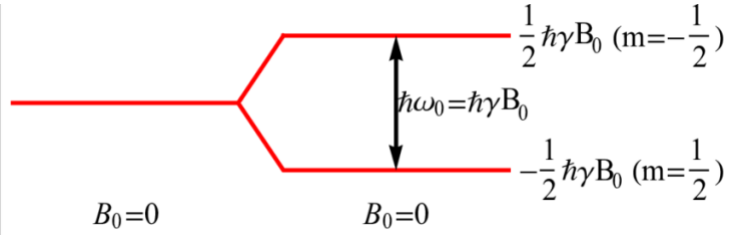


Figure 1.4 Diagram of Zeeman splitting.

However, in a macroscopic system having a collection of N identical spin $\frac{1}{2}$, where N is so vast that it is comparable to the Avogadro number, $N \cong 6 \times 10^{23}$, each proton has a different polarization state at any given point. If each spin has two potential spin states, $|\alpha\rangle$ or $|\beta\rangle$, the state of all the spins will span the vast 2^N dimensions, making it impractical to calculate. In this case, there is an elegant alternative, which is a statistical approach that does not refer to individual spin states and only requires information about the probabilities of each spin being $|\alpha\rangle$ and $|\beta\rangle$; this is because an ensemble of a vast collection of independent spins shows the average response of spins instead of individual observations, making quantum statistical mechanics an elegant and powerful approach. The classical statistical ensemble states that the relative statistic weight of the systems in which the point lies in an element of the hypervolume dv around point r of their probabilistic space equals to

$$\mathcal{P}(r)dv.$$

Assume a physical variable $\mathcal{F}(r)$ whose possible value is determined by the probability distribution of the position of states; the ensemble average is defined as

$$\bar{\mathcal{F}} = \int F(r)\,\mathcal{P}(r)dv. \tag{1.25}$$

In the quantum system, the representative states are represented in Hilbert space as Dirac notation with discrete and finite dimensions. The distribution of probability for a point located in the probability space $d\tau$ around each point τ is defined as

$$\mathcal{P}(\psi)d\tau.$$

The expectation of an observable $\mathcal{A}$ is equal to

$$\langle\psi|\mathcal{A}|\psi\rangle. \tag{1.26}$$

For a statistical distribution of the states, the average is

$$\langle\mathcal{A}\rangle = \int \mathcal{P}(\psi)\langle\psi|\mathcal{A}|\psi\rangle d\tau. \tag{1.27}$$

Consider that a single spin $|\psi\rangle$ in a general superposition state can be projected as the sum of the basis state $|\varphi_i\rangle$,

$$|\psi\rangle = \sum_i^n a_i|\varphi_i\rangle, \tag{1.28}$$

where the expectation value of an observable $\mathcal{A}$ by closure theorem and circular permutation can be further expressed as

$$\begin{aligned}\langle\mathcal{A}\rangle &= \sum_i\sum_j \int \mathcal{P}(\psi)\langle\psi|\varphi_i\rangle\langle\varphi_i|\mathcal{A}|\varphi_j\rangle\langle\varphi_j|\psi\rangle d\tau \\ &= \sum_i\sum_j \langle\varphi_i|\mathcal{A}|\varphi_j\rangle \int \mathcal{P}(\psi)\langle\varphi_j|\psi\rangle\langle\psi|\varphi_i\rangle d\tau\end{aligned} \tag{1.29}$$

This results in the definition of the density operator σ,

$$\sigma = \int \mathcal{P}(\psi)|\psi\rangle\langle\psi|d\tau, \tag{1.30}$$

where $|\psi\rangle\langle\psi|$ is also called the projection operator.

Eq.(1.29) is rewritten by substituting Eq.(1.30) and taking advantage of the definition of the trace as

$$\begin{aligned}\langle\mathcal{A}\rangle &= \sum_i\sum_j \langle\varphi_i|\mathcal{A}|\varphi_j\rangle \int \mathcal{P}(\psi)\langle\varphi_j|\psi\rangle\langle\psi|\varphi_i\rangle d\tau \\ &= \sum_i\sum_j \langle\varphi_i|\mathcal{A}|\varphi_j\rangle \langle\varphi_i|\sigma|\varphi_j\rangle = Tr\{\sigma\mathcal{A}\} = Tr\{\mathcal{A}\sigma\}\end{aligned} \tag{1.31}$$

Eq.(1.31) implies that density spin operators can deduce any macroscopic observation, with a density operator representing the state of the entire spin ensemble rather than a vast number of independent microscopic spins. Quantum statistical mechanics stems from the density operator concept, which underpins NMR.

In a given basis, a matrix element of σ is

$$\langle i|\sigma|j\rangle = \int \mathcal{P}(\psi)\langle i|\psi\rangle\langle\psi|j\rangle d\tau. \tag{1.32}$$

Expanding $|\psi\rangle$ as the superposition of eigenstates using Eq.(1.28), we get

$$\langle i|\psi\rangle\langle\psi|j\rangle = a_i a_j. \tag{1.33}$$

Substituting Eq.(1.33) to Eq.(1.32), we get

$$\langle i|\sigma|j\rangle = \int \mathcal{P}(\psi) a_i a_j \, d\tau = \overline{a_i a_j}, \tag{1.34}$$

where the bar means an ensemble average.

The density operator is the Hermitian operator, which has no relationship with the physical density of the sample. In the system with spin-1/2, in which two eigen spin states are denoted as $|\alpha\rangle$ and $|\beta\rangle$, two coherences are created and denoted as σ_+ and σ_-:

$$\begin{aligned} \sigma_+ &= \langle\alpha|\sigma|\beta\rangle = \overline{a_\alpha a_\beta} \\ \sigma_- &= \langle\beta|\sigma|\alpha\rangle = \overline{a_\beta a_\alpha}. \end{aligned} \tag{1.35}$$

The coherence orders −1 and +1 are present via the difference in z-angular momentum of the connected states in a high magnetic field; the coherences σ_+ and σ_- indicate transverse spin magnetization, which occurs normally in the xy-plane if an external field is aligned with the z-axis.

The diagonal matrix elements are

$$\langle i|\sigma|i\rangle = \overline{a_i a_i}, \tag{1.36}$$

where $\overline{a_i a_i}$ is a positive real number, meaning probability. In a system with spin-1/2, $\sigma_{\alpha\alpha}$ and $\sigma_{\beta\beta}$ are the populations of states $|\alpha\rangle$ and $|\beta\rangle$, respectively.[4] As spin polarizations are derived from spin populations, and there is a link between spin polarization and magnetization, the bulk magnetization M can be represented by the spin density operator. The longitudinal component M_z is related to the population difference between the states of $|\alpha\rangle$ and $|\beta\rangle$:

$$M_Z = 2B^{-1}\left(\overline{\sigma_{\alpha\alpha}} - \overline{\sigma_{\beta\beta}}\right). \tag{1.37}$$

The transverse components M_x and M_y can be represented by (−1)-quantum coherence:

$$M_x = 4B^{-1}\mathrm{Re}\{\sigma_-\}$$
$$M_y = 4B^{-1}\mathrm{Im}\{\sigma_-\}. \tag{1.38}$$

Recall the Schrodinger equation and its related conjugation expressed as

$$\frac{d}{dt}|\psi\rangle = -i\mathcal{H}|\psi\rangle \tag{1.39}$$

and

$$\frac{d}{dt}\langle\psi| = -i\mathcal{H}\langle\psi|. \tag{1.40}$$

The time derivative of the projection operator $|\psi\rangle\langle\psi|$ can be derived as

$$\begin{aligned}\frac{d}{dt}|\psi\rangle\langle\psi| &= \left(\frac{d}{dt}|\psi\rangle\right)\langle\psi| + |\psi\rangle\left(\frac{d}{dt}\langle\psi|\right)\\ &= -i\mathcal{H}|\psi\rangle\langle\psi| + i|\psi\rangle\langle\psi|\mathcal{H}\\ &= -i[\mathcal{H}, |\psi\rangle\langle\psi|],\end{aligned} \tag{1.41}$$

that is,

$$\frac{d}{dt}\sigma = -\mathrm{i}[\mathcal{H}, \sigma]. \tag{1.42}$$

Eq(1.42) is called the Liouville-von Neumann equation,[11,12] and its solution can be separately discussed when $\mathcal{H}$ is time-dependent and time-independent.

1.4.1 $\mathcal{H}$ Is Time-independent

When $\mathcal{H}$ is time-independent, and the basis and eigenstate of $\mathcal{H}$ are assumed,

$$\mathcal{H}|i\rangle = a_i|i\rangle; \tag{1.43}$$

By substituting Eq.(1.42) and Eq.(1.43), we can obtain the time derivative of the diagonal matrix elements of density operator σ,

$$\frac{d}{dt}\langle i|\sigma|i\rangle = \left\langle i\left|\frac{d}{dt}\sigma\right|i\right\rangle = \langle i|-\mathrm{i}[\mathcal{H},\sigma]|i\rangle = -i\{\langle i|\mathcal{H}|i\rangle\langle i|\sigma|i\rangle - \langle i|\sigma|i\rangle\langle i|\mathcal{H}|i\rangle\} = 0. \tag{1.44}$$

This means that the diagonal matrix elements of the density matrix are constant as a function of time,

$$\langle i|\sigma(t)|i\rangle = \langle i|\sigma(0)|i\rangle. \tag{1.45}$$

Similarly, by substituting Eq.(1.42) and Eq.(1.43), we can obtain the time derivative of the off-diagonal matrix elements of the density operator σ,

$$\frac{d}{dt}\langle i|\sigma|j\rangle = -i\{\langle i|\mathcal{H}|i\rangle\langle i|\sigma|j\rangle - \langle i|\sigma|j\rangle\langle j|\mathcal{H}|j\rangle\} = -i\big(a_i - a_j\big)\langle i|\sigma|j\rangle. \tag{1.46}$$

Because $a_i \neq a_j$ based on the eigenvalues that are different, we get

$$\langle i|\sigma(t)|j\rangle = \langle i|\sigma(0)|j\rangle exp\big\{-i\big(a_i - a_j\big)t\big\}. \tag{1.47}$$

Eq.(1.46) shows that the off-diagonal elements of the density matrix oscillate without decaying as a function of time.

We combine the results of Eq.(1.45) and Eq.(1.47) into a general operator form:

$$\sigma(t) = \exp(-i\mathcal{H}t)\sigma(0)\exp(i\mathcal{H}t). \tag{1.48}$$

Eq.(1.48) represents the evolution of the density matrix when $\mathcal{H}$ is time-independent.

For the physical observable $\mathcal{A}$, the substitution of Eq.(1.42) obtains the time evolution:

$$\frac{d}{dt}\langle\mathcal{A}\rangle = \mathrm{Tr}\left\{\mathcal{A}\frac{d\sigma}{dt}\right\} = \mathrm{Tr}\{\mathcal{A}\times -\mathrm{i}[\mathcal{H},\sigma]\}. \tag{1.49}$$

When $\mathcal{H}$ is time-independent, we get

$$\langle\mathcal{A}\rangle = \mathrm{Tr}\{\mathrm{e}^{-\mathrm{i}\mathcal{H}\mathrm{t}}\mathcal{A}e^{-i\mathcal{H}t}\sigma(0)\} \tag{1.51}$$

1.4.2 $\mathcal{H}$ Is Time-dependent

When $\mathcal{H}$ is time-dependent, the solution is quite complicated and not definitely solvable. In NMR, the overall Hamiltonian of a spin is

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1, \tag{1.52}$$

where

$\mathcal{H}_0$ in the first term of Eq.(1.53) is a static field applied along the z-axis, and $\mathcal{H}_1$, which is much smaller than $\mathcal{H}_0$, is applied perpendicular to the transversal plane and rotates around it at frequency ω. $\mathcal{H}_1$ is time-dependent and often called the radiofrequency pulse.

The trick in replacing the time-dependent $\mathcal{H}$ with time-independent $\widetilde{\mathcal{H}}$ is the rotating frame. Recall that any operator Q has an associated unitary operator U(t),

$$\tilde{Q} = \mathrm{U}(t) Q U^*(t), \tag{1.55}$$

where $U^*(t)$ satisfies

We conveniently choose

$$\begin{aligned} \mathrm{U}(t) &= e^{iAt}\mathrm{U}(0), \\ U^*(t) &= e^{-iAt}U^*(0). \end{aligned} \tag{1.56}$$

$$\mathrm{U}(0) = U^*(0) = \hat{1}$$

so that

$$\begin{aligned} \frac{d}{dt}U(t) &= iAU(t), \\ \frac{d}{dt}U^*(t) &= -iU^*(t)A. \end{aligned} \tag{1.57}$$

We return to the evolution of the density matrix–Liouville-von Neumann equation (1.42):

$$\mathcal{H}_0 = -\gamma B_0 I_z = \omega_0 I_z \tag{1.53}$$

$$\mathcal{H}_1 = -\gamma B_1 \left(I_x \cos \omega t + I_y \sin \omega t\right) = \omega_1 \left(I_x \cos \omega t + I_y \sin \omega t\right). \tag{1.54}$$

$$\frac{d}{dt}\sigma = -i[\mathcal{H}, \sigma]$$

When replacing Q with σ in Eq.(1.55), we get

$$\tilde{\sigma} = \mathrm{U}(t) \sigma U^*(t). \tag{1.58}$$

Then, applying Eq.(1.42) to $\tilde{\sigma}$, which is denoted in Eq.(1.58), we get

$$\frac{d}{dt}\tilde{\sigma} = \frac{d}{dt}\mathrm{U}\sigma U^* = \left(\frac{d}{dt}\mathrm{U}\right)\sigma U^* + \mathrm{U}(\frac{d}{dt}\sigma)U^* + \mathrm{U}\sigma\left(\frac{d}{dt}U^*\right). \tag{1.59}$$

According to Eq.(1.57), we rewrite Eq.(1.59) as

$$\frac{d}{dt}\tilde{\sigma} = iA\mathrm{U}\sigma U^* - i\mathrm{U}[\mathcal{H},\sigma]U^* - i\mathrm{U}\sigma U^* A. \tag{1.60}$$

Note that

$$\mathrm{U}\sigma U^* = \tilde{\sigma} \tag{1.61}$$

$$U\sigma U^* = \widetilde{\mathcal{H}} \tag{1.62}$$

$$\mathrm{U}[\mathcal{H},\sigma]U^* = [\mathrm{U}\mathcal{H}U^*, \mathrm{U}\sigma U^*] = \left[\widetilde{\mathcal{H}},\tilde{\sigma}\right]. \tag{1.63}$$

Substituting Eq.(1.61) and Eq.(1.62) to Eq.(1.60), we get

$$\frac{d}{dt}\tilde{\sigma} = iA\tilde{\sigma} - \left[\widetilde{\mathcal{H}},\tilde{\sigma}\right] - i\tilde{\sigma}A = -\mathrm{i}\left[\left(\widetilde{\mathcal{H}} - A\right),\tilde{\sigma}\right] \tag{1.64}$$

This transformation defines a change of representation in which the evolution of new density matrix $\tilde{\sigma}$ is subject to an effective Hamiltonian. This new representation is called Heisenberg representation:

$$\widetilde{\mathcal{H}}_{\mathrm{eff}} = \widetilde{\mathcal{H}} - A. \tag{1.65}$$

Then, we turn to the original problem—the overall Hamiltonian with the presence of the radiofrequency field is expressed as Eq.(1.52). Operator *A* in Eq.(1.56) is chosen as

$$A = \omega I_z. \tag{1.66}$$

Eq.(1.56) is written as

$$\begin{aligned} U &= e^{iAt} = e^{i\omega I_z t} \\ U^* &= e^{-iAt} = e^{-i\omega I_z t}. \end{aligned} \tag{1.67}$$

In the substitution of U and U^* to observables, observables have a literally unitary transformation. This is how the rotating frame works. It is equivalent to rotating the reference axes around the Z-axis at the frequency -ω. We rewrite Eq.(1.54) as having a unitary transformation in Eq.(1.66):

$$\mathcal{H}_1 = \omega_1(I_x \cos\omega t + I_y \sin\omega t) = \omega_1(e^{-i\omega I_z t} I_x e^{i\omega I_z t}) = \omega_1 U^*(t) I_x U(t). \tag{1.68}$$

The effect Hamiltonian in the rotating frame from Eq.(1.65) by substituting Eq.(1.62) and Eq.(1.66) can be described as

$$\widetilde{\mathcal{H}}_{\text{eff}} = \widetilde{\mathcal{H}} - A = U\mathcal{H}\, U^* - \omega I_z. \tag{1.69}$$

Substituting Eq.(1.52), Eq.(1.53), and Eq.(1.68) to $\mathcal{H}$, we represent Eq.(1.69) as

$$\begin{aligned} \widetilde{\mathcal{H}}_{\text{eff}} &= U(\omega_0 \mathrm{I_z} + \omega_1\, U^*(t) I_x U(t)) U^* - \omega I_z \\ &= (\omega_0 - \omega) I_z + \omega_1 U U^* I_x\, U U^* = (\omega_0 - \omega) I_z + \omega_1 I_x. \end{aligned} \tag{1.70}$$

In Eq.(1.70), the time *t* disappears, making the effective Hamiltonian time-independent. It is the form of a Zeeman interaction in an effective field $\widetilde{\mathcal{H}}_{\text{eff}}$ of components:

$$\widetilde{\mathcal{H}}_0 = -(\omega_0 - \omega) I_z \tag{1.71}$$

$$\widetilde{\mathcal{H}}_1 = \omega_1 I_x. \tag{1.72}$$

In summary, the evolution of the nuclear magnetic moment in the rotating frame is a precession around $\widetilde{\mathcal{H}}_{eff}$ with a frequency ω_{eff} of the components $\omega_0 - \omega$ along the Z-axis and ω_1 along the X-axis.

In particular, when $\omega_0 = \omega$, the effective filed in the rotating frame is reduced to $\mathcal{H}_1$ (i.e., the vertical magnetic moment disappears, leaving out the pure transversal magnetic moment). This is called resonance.

1.5 Quantum Description of the Standard NMR Process

The standard NMR is composed of a starting point of the thermal equilibrium system, and then it goes through a single radiofrequency pulse and a free induction decay. Finally, it returns to the thermal equilibrium system.

1.5.1 The Density Matrix at Thermal Equilibrium

A spin system is a thermal equilibrium with a thermostat at temperature T, implying that the system reaches a steady state in which there is no macroscopic change in the net flow of the thermal energy between the system and its surroundings. The probability of finding the system in a state of energy E_i obeys the Boltzmann distribution:

$$p_i = \zeta \exp\left(-\frac{E_i}{kT}\right), \tag{1.73}$$

where T is the absolute temperature, and K is the Boltzmann constant. ζ is the normalization constant to adjust

$$\sum_i p_i = 1$$

In the perpendicular direction to the magnetic field, the spins are equal and uniform to point to, making the transverse spin magnetization zero via general statistical arguments; thus, there is no coherence. The density matrix satisfying quantum statistical mechanics is

$$\sigma_{eq} = Z\begin{pmatrix} e^{-\hbar\omega_0/2kt} & 0 \\ 0 & e^{\hbar\omega_0/2kt} \end{pmatrix}, \tag{1.74}$$

where the partition function is $Z = \frac{1}{(e^{-\hbar\omega_0/2kt} + e^{\hbar\omega_0/2kt})}$. Because $\frac{\hbar\omega_0}{2kt} \ll 1$, Eq.(1.74) can be simplified to:

$$\sigma_{eq} = Z\begin{pmatrix} 1-\hbar\omega_0/2kt & 0 \\ 0 & 1+\hbar\omega_0/2kt \end{pmatrix} = \hat{1} - \frac{\hbar\omega_0}{2kt}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \hat{1} - \beta I_z, \tag{1.75}$$

where $\hat{1}$ is the identity matrix, and the βI_z, term is the secular part of the Hamiltonian σ_{eq}.

Therefore, the spin density operator in the rotating frame is derived as follows:

$$\widetilde{\sigma_0} = \mathrm{U}\sigma_0 U^* = e^{i\omega I_z t}\left(\hat{1} - \beta I_z\right)e^{-i\omega I_z t} = \sigma_0. \tag{1.76}$$

1.5.2 A Single Radiofrequency Pulse

Consider the case in which the pulse is applied on resonance, and the $(\frac{\pi}{2})_x$ pulse rotates the spin vector from the z-axis to the y-axis. The spin density operator is

$$\begin{aligned}\widetilde{\sigma_1} &= e^{i\left(\frac{\pi}{2}\right)I_x}\widetilde{\sigma_0}e^{-i\left(\frac{\pi}{2}\right)I_x} = e^{i\left(\frac{\pi}{2}\right)I_x}\hat{1}e^{-i\left(\frac{\pi}{2}\right)I_x} - \beta e^{i\left(\frac{\pi}{2}\right)I_x}\hat{I}_z e^{-i\left(\frac{\pi}{2}\right)I_x} \\ &= \hat{1} - \beta\hat{I}_y.\end{aligned} \tag{1.77}$$

This result shows that the effect of a $(\frac{\pi}{2})$ pulse aligned along the *x*-axis on the z- component of bulk magnetization is its conversion into the y-component of the magnetization. This is equivalent to rotating the bulk magnetization vector M from the z-axis into the y-axis.

When rewritten as the matrix representation, the radiofrequency pulse process is described as

$$\widetilde{\sigma_0} = \begin{pmatrix} \frac{1}{2}+\beta & 0 \\ 0 & \frac{1}{2}-\beta \end{pmatrix} \xrightarrow{(\pi/2)_x} \widetilde{\sigma_1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{i}\beta \\ \frac{1}{i}\beta & \frac{1}{2} \end{pmatrix}.$$

Therefore, the pulse eliminates the z-axis polarization in the thermal equilibrium, making the population of states $|\alpha\rangle$ and $|\beta\rangle$ both $1/2$. The pulse also creates coherences. There is no coherence in $\widetilde{\sigma_0}$, but it appears in $\widetilde{\sigma_1}$.

After the radiofrequency pulse, if relaxation is not taken into consideration, the process can be expressed as

$$\begin{aligned}\widetilde{\sigma_2} &= e^{i(\Omega^0\tau)I_z}\widetilde{\sigma_1}e^{-i(\Omega^0\tau)I_z} = e^{i(\Omega^0\tau)I_z}\left(\frac{1}{2}\hat{1} - \beta\hat{I}_y\right)e^{-i(\Omega^0\tau)I_z} \\ &= \frac{1}{2}\hat{1} - (\hat{I}_y\cos\Omega^0\tau + \hat{I}_x\sin\Omega^0\tau).\end{aligned} \tag{1.78}$$

Therefore, the population in the $|\alpha\rangle$ and $|\beta\rangle$ states are retained, that is,

$$\tilde{\sigma}_2^{\alpha} = \tilde{\sigma}_1^{\alpha}$$

$$\tilde{\sigma}_2^{\beta} = \tilde{\sigma}_1^{\beta}.$$

For (-1)-quantum coherence,

$$\tilde{\sigma}_2^{-1} = \exp\{+i\Omega^0\tau\}\tilde{\sigma}_1^{-1}$$

1.5.3 Free-Induction Decay

In practice, the phenomenon of relaxation works in the longitudinal and transversal directions. (1) Fluctuating the molecular surroundings causes the thermal equilibrium state to be gradually re-established after radiofrequency pulses convert the population difference into coherences, and (2) the coherences gradually decay to zero.

The coherences' decay is expressed by a mathematically exponential term:

$$\tilde{\sigma}_2^{-1} = \tilde{\sigma}_1^{-1}\exp\{(i\Omega^0 - \lambda)\tau\}, \tag{1.79}$$

where τ is the time interval between time points $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$, and λ, expressed by the reverse of the transverse relaxation time constant T_2, is called the damping rate constant.

The populations evolving after the pulse are given by

$$\begin{aligned} \sigma_3^{\alpha} &= \frac{1}{2} + \beta(1 - 2e^{-\tau/T_1}) \\ \sigma_3^{\beta} &= \frac{1}{2} - \beta\left(1 - 2e^{-\tau/T_1}\right), \end{aligned} \tag{1.80}$$

where the time constant T_1 is the longitudinal or spin-lattice relaxation time constant.

1.6 NMR Signals and NMR Spectrua

After relaxation, the NMR signal is generated and collected by utilizing transverse nuclear magnetization to induce an electric current in the coil surrounding the sample.[11] The current is amplified and transferred to the digit signal, conducting a Fourier transform to obtain the NMR spectrum.

The complex NMR signal, assumed to be denoted as $\tilde{\sigma}_3^{-1}$, is produced from the (−1)-quantum coherence in the rotating frame,

$$\mathrm{s(t)}\sim 2\mathrm{i}\tilde{\sigma}_3^{-1}(\mathrm{t})\exp\{-\mathrm{i}\phi_{\mathrm{rec}}\} = 2\mathrm{i}\tilde{\sigma}_3^{-1}(0)\exp\{-\mathrm{i}\phi_{\mathrm{rec}} + (\mathrm{i}\Omega^0 - \lambda)\mathrm{t}\}, \tag{1.81}$$

where ϕ_{rec} is the receiver phase shift, and $\lambda = \frac{1}{T_2}$.

We then extract the term without time t from Eq.(1.81); it is called signal amplitude and expressed as

$$\mathrm{a} = 2\mathrm{i}\tilde{\sigma}_3^{-1}(0)\exp\{-\mathrm{i}\phi_{\mathrm{rec}}\}.$$

Thus, the amplitude of the NMR signal is the product of the initial value of the (−1)-quantum coherence and a phase factor. The Fourier transformation[13] decomposes it into its constituent frequencies from the time domain to the frequency domain, obtaining the NMR spectrum.

1.7 Advanced Techniques

1.7.1 Magic-Angle Spinning (MAS)

Magic-angle spinning (MAS) is a widely used averaging technique to eliminate the anisotropic interactions of internal Hamiltonians. It makes a sample rapidly spin around an axis at an inclined angle to B_0. This angle is so unusual that anisotropic interaction, which depends on the angle, partially or wholly disappears. The internal interactions we have discussed in Section (1.3.3), including $\mathcal{H}_{\mathrm{q}}$, $\mathcal{H}_{\mathrm{DD}}$, $\mathcal{H}_{\mathrm{CS}}$, and $\mathcal{H}_{\mathrm{j}}$, all have similar angular dependencies. We take heteronuclear dipolar coupling as an example; if we only consider the z-component of the angular momentum of spins *I* and *S*, the dipolar coupling is described as Eq.(1.18). In this equation, θ is the angle of the internuclear vector:

$$\mathcal{H}_{IS}^{\mathrm{DD,heter}} = \frac{\gamma_I\gamma_S}{r_{IS}^3}(3\cos^2\theta - 1)I_zS_z$$

Mathematically, this equation can be zero if the value of expression $(3cos^2\theta - 1)$ equals zero. It is 54.74°, which makes $(3cos^2\theta - 1)$ zero. Thus, 54.74° for θ is called the magic angle to make the dipolar interaction average zero, as the solution-state NMR processes a rapidly tumbling moving, enabling the narrowing of the linewidth on account of this interaction's suppression. In addition, rotating the sample at the magic angle is usually associated with faster spinning; this removes anisotropic interactions as solution NMR and makes the spectrum average out to a narrow isotropic peak with a manifold of spinning sidebands appearing at integer intervals of the spinning speed.[14–16]

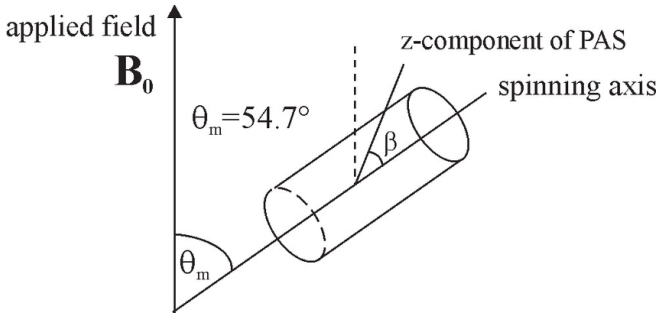


Figure 1.5 Scheme of magic angle spinning.

1.7.2 Cross-Polarization(CP)

In solids, the technique of cross-polarization(CP) involves transferring magnetization between two different spins, *I* and *S*, via heteronuclear dipole–dipole coupling. The reason for transferring magnetization is that one spin, which is our interest, is usually of low abundance, limiting the NMR application. Therefore, this magnetization transfer happens typically from a spin with high natural abundance, such as ^{1}H to another spin with rare abundance, such as ^{13}C. In this way, the sensitivity of nuclei with a low magnetogyric ratio is boosted, thus improving the signal-to-noise ratio.

As cross-polarization leverages *I–S* dipolar coupling, in solid, it still suffers from severe issues of broadening the spectra coming from anisotropic interactions. MAS is usually experimentally coupled with the CP process. The rapid rotation of MAS averages out anisotropic interactions; at the same time, CP enhances the sensitivity to acquire relatively accurate quantitative information for spin-1/2 systems.

Consider a system in which dipole–dipole interactions dominate; the total spin Hamiltonian can be represented as

$$\mathcal{H}_{\mathrm{tot}} = \mathcal{H}_{\mathrm{I}} + \mathcal{H}_{\mathrm{S}} + \mathcal{H}_{\mathrm{DD}}^{\mathrm{II}} + \mathcal{H}_{\mathrm{DD}}^{\mathrm{SS}} + \mathcal{H}_{\mathrm{DD}}^{\mathrm{IS}}, \tag{1.82}$$

where *I* spins are abundant, and *S* spins are rare. $\mathcal{H}_I$ and $\mathcal{H}_s$ denote the Hamiltonian of the chemical shift corresponding to spins *I* and *S*, respectively, $\mathcal{H}_{\mathrm{DD}}^{\mathrm{SS}}$ represents these two spins' homonuclear dipole–dipole coupling, and $\mathcal{H}_{\mathrm{DD}}^{\mathrm{IS}}$ is the heteronuclear dipolar coupling.

The CP experiment usually initially applies a 90-degree pulse to *I* spins at frequency $\omega_{oI} = \gamma_I B_0$, where γ_I is the magnetogyric ratio of spins *I*. Next, as the process discussed in standard NMR, transverse magnetization, such as I_x and I_y, is generated. Then, we utilize a spin-locking technique, which includes a pulse with radiofrequency strength B_1 and phase-shifting by 90° relative to the first pulse, to lock up *I* spin magnetization intentionally.

$$\omega_I = \gamma_I B_{1I} = \gamma_s B_{1s} = \omega_s. \tag{1.83}$$

At the same time, mixing happens with spin-locking. It has two phases: (1) A pulse of radiofrequency with strength B_{1s} is applied to the *S* spins with magnetogyric ratio γ_I. (2) Polarization transfer proceeds if the Hartmann-Hahn (H.H.) matching condition is satisfied [17],

The time during polarization transfer is described as mixing time τ_{CP}. At the end of mixing time τ_{CP}, *S* spins build up transverse magnetization. Finally, the two spin systems are decoupled by applying nonstop radiofrequency irradiation to the *I* spins, and then the *S* spins alone go thorough free-induction decay(FID) acquisition.

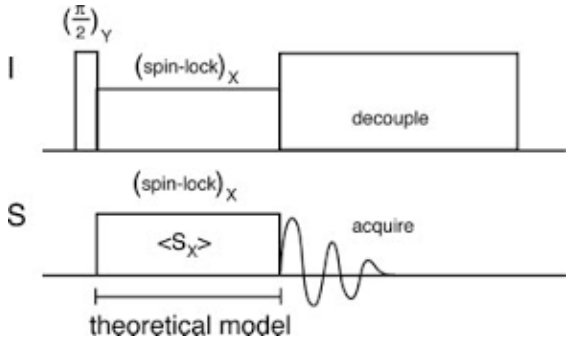


Figure 1.6 Schematic diagram of the basic experiment employing cross polarization. Adapted from Rovnyak. 18

From Eq.(1.8), the Hartmann-Hahn matching condition shows that efficient cross-polarization transfer occurs on the condition of equality of the angular frequency, $\omega = \gamma B$, of the two spin systems. γ, which is the intrinsic characteristic of the spins of two spin systems, is usually of a vast discrepancy; therefore, adjusting two radiofrequencies, B_{1I}, and B_{1S}, is necessary to meet the requirement of the

Hartmann-Hahn matching condition. With this, polarization shift can occur, and the benefits of the CP technique can be realized.

For instance, the majority of common nuclei, such as ^{1}H and ^{13}C, have a gyromagnetic ratio, γ_H ~267.5 and γ_C ~67.2828; if CP takes place between them, there is an enhancement to ^{13}C (*S* spins) by γ_H/γ_C~4. With the faster spin-lattice relaxation time, T_1, a spectrum's S/N beyond 1,000 is easily realized in the absence of augmenting the experimental time. Furthermore, as ^{3}C nuclei are rare, the $\mathcal{H}_{\mathrm{DD}}^{\mathrm{SS}}$ interaction can be legitimately ignored. If a radiofrequency pulse is maintained on the H^1 nuclei during FID acquisition, the $\mathcal{H}^{IS}$interaction is also completely removed, thereby leaving out information about the system's chemical shift and the smaller second-order effects.

1.8 References

(1) Uhlenbeck, G. E.; Goudsmit, S. Ersetzung der Hypothese vom unmechanischen Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons. *Naturwissenschaften* **1925**, *13* (47), 953–954. https://doi.org/10.1007/BF01558878.

(2) Uhlenbeck, G. E.; Goudsmit, S. Spinning Electrons and the Structure of Spectra. *Nature* **1926**, *117* (2938), 264–265. https://doi.org/10.1038/117264a0.

(3) Protein NMR Spectroscopy - 2nd Edition https://www.elsevier.com/books/protein-nmr-spectroscopy/cavanagh/978-0-12-164491-8 (accessed Apr 9, 2020).

(4) Pauli, W. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Z. Für Phys.* **1925**, *31* (1), 765–783. https://doi.org/10.1007/BF02980631.

(5) Levitt, M. H. *Spin Dynamics Basics of Nuclear Magnetic Resonance*.

(6) Forsén, S.; Lindman, B. Ion Binding in Biological Systems as Studied by NMR Spectroscopy. *Methods Biochem. Anal.* **1981**, *27*, 289–486. https://doi.org/10.1002/9780470110478.ch5.

(7) Dirac, P. A. M. A New Notation for Quantum Mechanics. *Math. Proc. Camb. Philos. Soc.* **1939**, *35* (3), 416–418. https://doi.org/10.1017/S0305004100021162.

(8) Phys. Rev. 28, 1049 (1926) - An Undulatory Theory of the Mechanics of Atoms and Molecules https://journals.aps.org/pr/abstract/10.1103/PhysRev.28.1049 (accessed Apr 9, 2020).

(9) Discussion of Probability Relations between Separated Systems | Mathematical Proceedings of the Cambridge Philosophical Society | Cambridge Core https://www.cambridge.org/core/journals/mathematical-proceedings-of-the-cambridge-philosophical-society/article/discussion-of-probability-relations-between-separated-systems/C1C71E1AA5BA56EBE6588AAACB9A222D (accessed Apr 9, 2020).

(10) Keeler, J. *Understanding NMR Spectroscopy*.

(11) Breuer, H.-P.; Breuer, P. I. H.-P.; Petruccione, F.; Petruccione, S. of P. and A. P. F. *The Theory of Open Quantum Systems*; Oxford University Press, 2002.

(12) Schwabl, F. *Statistical Mechanics*; Springer Science & Business Media, 2002.

(13) LINDON, J.; JC, L.; AG, F. DIGITISATION AND DATA PROCESSING IN FOURIER TRANSFORM NMR. *Digit. DATA Process. FOURIER TRANSFORM NMR* **1980**.

(14) Lowe, I. J. Free Induction Decays of Rotating Solids. *Phys. Rev. Lett.* **1959**, *2* (7), 285–287. https://doi.org/10.1103/PhysRevLett.2.285.

(15) Mehring, M.; Waugh, J. S. Magic-Angle NMR Experiments in Solids. *Phys. Rev. B* **1972**, *5* (9), 3459–3471. https://doi.org/10.1103/PhysRevB.5.3459.

(16) Andrew, E. R.; Richards, R. E.; Packer, K. J. Magic Angle Spinning in Solid State n.m.r. Spectroscopy. *Philos. Trans. R. Soc. Lond. Ser. Math. Phys. Sci.* **1981**, *299* (1452), 505–520. https://doi.org/10.1098/rsta.1981.0032.

(17) Hartmann, S. R.; Hahn, E. L. Nuclear Double Resonance in the Rotating Frame. *Phys. Rev.* **1962**, *128* (5), 2042–2053. https://doi.org/10.1103/PhysRev.128.2042.

(18) Rovnyak, D. Tutorial on Analytic Theory for Cross-Polarization in Solid tate NMR. *Concepts Magn. Reson. Part A* **2008**, *32A* (4), 254–276. https://doi.org/10.1002/cmr.a.20115.

Chapter 2. Theoretical Calculations of NMR

2.1 The Origin of Quantum Mechanics

At the end of the 19th century and beginning of the 20th century, classical physics, such as Newton's mechanism, thermodynamics and statistical mechanics, met its Waterloo when it got into trouble by explaining certain experimental observations, such as black body radiation[1,2], photoelectronic effect[3,4] and atomic spectra[5–7]. To remedy the chaos, the matrix mechanism (inspired by the Bohr model) and wave mechanics (derived from de Broglie's matter-wave) were developed nearly simultaneously. Modern quantum mechanics was born when Schrödinger demonstrated the equivalence[8] between the matrix formalism and his wave formulation[9].

In quantum mechanics, things behave differently from those we have experienced because they are on the microcosmic scale. Feynman explains:

'Historically, the electron, for example, was thought to behave like a particle, and then it was found that in many respects it behaved like a wave. So it really behaves like neither. Now we have given up. We say: "It is like neither."'[10]

Since neither a wave nor a particle accurately describes it, a novel way to represent the world underpinning quantum mechanics was introduced: the probability of a particle occurring at a given time and in a given place is proportional to the square of the magnitude of the particle's wave-function Ψ [11].

$$\left(\frac{-\mathrm{h}^2}{8\pi^2\mathrm{m}}\,\nabla + \mathcal{V}\right)\Psi_{\mathrm{tot}}(\mathrm{X},\mathrm{t}) = \frac{\mathrm{ih}}{2\pi}\frac{\partial\Psi_{\mathrm{tot}}(\mathrm{X},\mathrm{t})}{\partial\mathrm{t}}. \qquad (2.1)$$

Here Ψ is the wave function, h is Plank's constant, m is the particle's mass, $\mathcal{V}$ is the potential field and ∇ is the Laplacian: $\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$.

A particle's distribution of likelihood at time *t* can be written as the product of the wave function coupled with its complex conjugate. The form of the Hamiltonian operator is derived from classical mechanics, where kinetic and potential energies add up to the Hamiltonian function- $\mathcal{H}=\mathcal{T}+\mathcal{V}$ for a single particle in a non-relativistic circumstance.

In chemistry, the state in which the wave function is stationary (referred to as the stationary state) is atomic and molecular orbitals; thus, the time-dependent Schrödinger equation, Eq.(2.1), can be simplified to the time-independent Schrödinger equation in which $\Psi_{\mathrm{tot}}(\mathrm{X},\mathrm{t})$ can be decomposed as $\Psi_{\mathrm{tot}}(\mathrm{X})\cdot\zeta(\mathrm{t})$ and the variable $\zeta(\mathrm{t})$ can be dropped out, producing

$$\left(\frac{-\mathrm{h}^2}{8\pi^2\mathrm{m}}\nabla+\mathcal{V}\right)\Psi_{\mathrm{tot}}(\mathrm{X})=E\Psi_{\mathrm{tot}}(\mathrm{X}), \tag{2.2}$$

where *E* represents the ground–state energy of the system.

2.2 Molecular Hamiltonian

The molecular system is a multi-particle system wherein each electron interacts with multiple nuclei and the other electrons. In this case, spatial position X is separated into r for the electronic position and R for the nucleus position; therefore, a complete description of the Schrödinger equation is expressed as

$$\mathcal{H}_{\mathrm{tot}}\Psi(\mathrm{R},\mathrm{r})_{\mathrm{tot}}=\mathrm{E}\Psi(\mathrm{R},\mathrm{r})_{\mathrm{tot}}, \tag{2.3}$$

where

$$\begin{aligned}\mathcal{H}_{\mathrm{tot}} &= \mathcal{T}_{\mathrm{e}}+\mathcal{T}_{\mathrm{N}}+\mathcal{V}_{\mathrm{eN}}+\mathcal{V}_{\mathrm{ee}}+\mathcal{V}_{\mathrm{NN}}\\ &= \sum_{\mathrm{i}}\frac{\mathrm{h}}{2\mathrm{m}}\nabla+\sum_{\mathrm{I}}\nabla-\sum_{\mathrm{i}}\sum_{\mathrm{I}}\frac{\mathrm{Z_I e^2}}{\mathrm{r_{iI}}}+\sum_{\mathrm{i<j}}\frac{\mathrm{e^2}}{\mathrm{r_{ij}}}+\sum_{\mathrm{I<J}}\frac{\mathrm{Z_I Z_J e^2}}{\mathrm{R_{IJ}}}.\end{aligned} \tag{2.4}$$

The five terms in Eq.(2.4), in order, are the kinetic energy electrons, the kinetic energy of the nucleus, the overall interaction energy between each electron and each nucleus, the overall interaction energy between any electron and the other electrons, and the total interaction energy between any nucleus and the other nuclei. Ψ is the molecular wave function, and E is the ground-state energy of the molecular system. r_{iI} denotes the distance between the particle i and the particle j, and Z represents the atom's unique atomic number.

2.3 The Born-Oppenheimer Approximation

The full Hamiltonian Eq.(2.3) that neglects spin interactions, can be written as

$$\mathcal{H}_{\mathrm{tot}}\Psi_{\mathrm{tot}}(\mathrm{R},\mathrm{r}) = (\mathcal{T}_{\mathrm{N}} + \mathcal{H}_{\mathrm{e}})\Psi_{\mathrm{tot}}(\mathrm{R},\mathrm{r}), \tag{2.5}$$

where $\mathcal{T}_{\mathrm{N}}$ is the kinetic energy term for the nuclei, as in

$$\mathcal{T}_{\mathrm{N}} = \sum_{I<J} \frac{e^2 Z_I Z_J}{R_{IJ}}; \tag{2.6}$$

and $\mathcal{H}_e$ contains all interaction related to the electrons:

$$\begin{aligned} \mathcal{H}_{\mathrm{e}} &= \mathcal{T}_{\mathrm{e}} + \mathcal{V}_{\mathrm{eN}} + \mathcal{V}_{\mathrm{ee}} \\ &= \sum_{\mathrm{i}} \frac{\mathrm{h}}{2\mathrm{m}} \nabla \quad - \sum_{\mathrm{i}} \sum_{\mathrm{I}} \frac{\mathrm{Z_I e^2}}{\mathrm{r_{iI}}} + \sum_{\mathrm{i<j}} \frac{\mathrm{e^2}}{\mathrm{r_{ij}}}. \end{aligned} \tag{2.7}$$

Eq.(2.5) gives the double-particle system (i.e., one electron and one nucleus) a mathematically accurate solution. Nevertheless, for higher-order systems, certain approximations have to be applied. Because the mass of the nucleus is much larger than the mass of the electron but the velocity of the nucleus is much slower than that of the electron, the velocity of the nucleus can be neglected; thus, the nucleus is assumed fixed relative to the electrons, and the fixed nucleus provides the nucleus field to moving electrons. This approximation is the Born-Oppenheimer approximation[13].

After separating the nucleus and electrons as

$$\Psi_{\mathrm{tot}}(\mathrm{R},\mathrm{r}) = \Psi_{\mathrm{e}}(\mathrm{R},\mathrm{r})\Psi_{\mathrm{N}}(\mathrm{R}), \tag{2.8}$$

By substituting Eq.(2.8) into Eq.(2.5), the Hamiltonian of electrons $\mathcal{H}_{\mathrm{e}}$, after dropping $\Psi_{\mathrm{N}}(\mathrm{R})$, can be written as

$$\mathcal{H}_{\mathrm{e}}(\mathrm{R}) = E(\mathrm{R})\Psi_{\mathrm{e}}(\mathrm{R},\mathrm{r}). \tag{2.9}$$

In Eq(2.8), $\Psi_{\mathrm{N}}(\mathrm{R})$ is the wave function of the nucleus, which is the function of the position of the nucleus alone, and $\Psi_{\mathrm{e}}(\mathrm{R},\mathrm{r})$ is the wave function of electrons, which is the function of the position of the nucleus and electrons.

The solution to Eq.(2.9) for electronic wave function $\Psi_{\mathrm{e}}(\mathrm{R},\mathrm{r})$ generates an effective potential capacity $\mathrm{E}_{\mathrm{eff}}(\mathrm{R})$,which is an expression of variable nuclear coordinates and describes the potential energy surface of a chemical system.

2.4 Independent Electrons and the Hartree Function

The interaction between the electrons still causes a dilemma after the Born-Oppenheimer approximation to the Schrödinger equation because at this point the molecule usually has too many electrons. Another blunt and coarse approximation assumes that every electron is independent, which means no electron has an impact on any other electron and is also free from the effect of the other electrons. $\Psi_{\mathrm{e}}(\mathrm{R},\mathrm{r})$ can be rewritten as a product of monoelectronic functions called spin-orbitals:

$$\Psi_{\mathrm{e}}(\mathrm{r}) = \Psi_{\mathrm{e}}(\mathrm{r}_1 \cdots \mathrm{r}_{\mathrm{n}}) = \varphi_1(\mathrm{r}_1)\cdots\varphi_{\mathrm{n}}(\mathrm{r}_{\mathrm{n}}). \tag{2.10}$$

Should $\mathcal{V}_{\mathrm{ee}}$ be excluded from $\mathcal{H}_{\mathrm{e}}$, the multi-electrons problems would degrade to a single-electron problem, with the result being

$$\sum_{\mathrm{i}} \hbar_{\mathrm{i}}\Psi_{\mathrm{e}}(\mathrm{r}) = E\Psi_{\mathrm{e}}(\mathrm{r}), \tag{2.11}$$

where $\hbar_i$ is the separate Hamiltonian for the electron *i*.

Obviously, $\Psi_e(r)$ fails to be a solution for Eq.(2.9) due to the presence of $\mathcal{V}_{ee}$, but it can be an approximation. The expected energy calculated from $\Psi_e(r)$, in which the basis wave function is orthonormal and spin is ignored, is

$$\langle \Psi_e | \mathcal{H}_e | \Psi_e \rangle = \sum_i \langle \varphi_i | \hbar_i | \varphi_i \rangle + \sum_{i,j} \left\langle \varphi_i \varphi_j \middle| \mathcal{V}_{ij} \middle| \varphi_i \varphi_j \right\rangle. \tag{2.12}$$

Given the orthonormality constraint: $\varphi_i(r_i) * \varphi_j(r_j) \equiv \delta_{ij}$, the solution applies the variational principle[14], which states that the minimal expected energy is estimated to be the exact ground-state when trying all wave functions,

$$\tilde{E} = \langle \Psi_e | \mathcal{H}_e | \Psi_e \rangle \rightarrow \tilde{E} \geq \langle \Psi_e | \mathcal{H}_e | \Psi_e \rangle = E_0. \tag{2.13}$$

The minimum E_0 is searched for by adding Lagrangian multipliers[15] to Eq.(2.12),

$$\delta \left[E - \sum_i \varepsilon_i \left(\varphi_i(r_i) * \varphi_j(r_j) - 1 \right) \right] = 0. \tag{2.14}$$

By differentiating Eq.(2.13) with respect to φ_i while disregarding $\delta\varphi_i$ terms beyond the second-order and forcing the minimum condition, we acquire

$$\langle \delta\varphi_i | \hbar_i | \varphi_i \rangle + \sum_{i \neq j} \left\langle \delta\varphi_i \varphi_j \middle| \mathcal{V}_{ij} \middle| \varphi_i \varphi_j \right\rangle - \varepsilon_i \langle \delta\varphi_i | \varphi_i \rangle = \left\langle \delta\varphi_i \middle| \hbar_i + \sum_{i \neq j} \left\langle \varphi_j \middle| \mathcal{V}_{ij} \middle| \varphi_j \right\rangle - \varepsilon_i \middle| \varphi_i \right\rangle = 0.$$

φ_i is arbitrary, so $\hbar_i + \sum_{i \neq j} \left\langle \varphi_j \middle| \mathcal{V}_{ij} \middle| \varphi_j \right\rangle - \varepsilon_i = 0$ always holds and can be rewritten as

$$\left\{ \hbar_i + \sum_{i \neq j} \int dr_j \frac{\left| \varphi_j \right|^2}{\left| r_i - r_j \right|} \right\} \varphi_i = \varepsilon_i \varphi_j, \tag{2.15}$$

Where

$$\hbar_i = \nabla + \mathcal{V}_{eN}.$$

Eq.(2.15) is called the Hartree function. It shows that independent electron approximation does not mean completely ignoring the interactions between electrons but rather focuses on how an electron moves in the potential $\mathcal{V}_{\mathrm{eN}}$ and the average field from the Coulomb interactions with other electrons**.**

However, the chicken or egg paradox looms with the realization that the average field of the Hamiltonian is built from the wave function, but the wave function comes from solving the Schrödinger function using Eq.(2.15). The solution is the self-consistent field method(SCF), which works in an iterative way beginning with the tentative orbitals used to set up the $\hbar_i$, and then Eq.(2.15) is solved to get new orbitals. This cycle goes on until the orbitals converge.

2.5 Slater Determinants and Hartree-Fock Self-consistent Fields

Even though the orbitals embracing the Pauli exclusion principle can be represented as the product of the independent orbital of each electron, they are incapable of satisfying anti-symmetry, which serves the intrinsic constraint that swapping coordinates of any two particles leads to a change in the sign of the wave function,

$$\varphi_1(\mathrm{r}_1)\ldots\varphi_{\mathrm{k}}(\mathrm{r}_{\mathrm{k}})\varphi_{\mathrm{l}}(\mathrm{r}_{\mathrm{l}})\ldots\varphi_{\mathrm{n}}(\mathrm{r}_{\mathrm{n}}) = -\varphi_1(\mathrm{r}_1)\ldots\varphi_{\mathrm{k}}(\mathrm{r}_{\mathrm{l}})\varphi_{\mathrm{l}}(\mathrm{r}_{\mathrm{k}})\ldots\varphi_{\mathrm{n}}(\mathrm{r}_{\mathrm{n}}).$$

The most straightforward way to ensure anti-symmetry is the Slater determinant, which builds upon the combination of the orbitals. Here, only the closed-shell system, in which electrons fully occupy every orbital with different spins, is considered. The wave function, considering a spin orbital of $\sigma(\omega)$, of the n electrons for the system can be expressed as

where q contains the spatial position $\varphi(\mathrm{r})$ and the spin $\sigma(\omega)$,

$$\phi(q) = \frac{1}{\sqrt{N!}}\begin{bmatrix} \varphi_1(q_1) & \varphi_2(q_1) & \cdots & \varphi_n(q_1) \\ \varphi_1(q_2) & \varphi_2(q_2) & \cdots & \varphi_n(q_2) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_1(q_N) & \varphi_2(q_N) & \ldots & \varphi_n(q_N) \end{bmatrix}, \tag{2.16}$$

$$\phi(\mathrm{q}) = \phi(\mathrm{r},\omega) = \varphi(\mathrm{r})\sigma(\omega). \tag{2.17}$$

For a spin-1/2, ω has two values:$+\frac{1}{2}$ and $-\frac{1}{2}$.

Basically, the anti-symmetry works because the swapping of any two electrons also swaps two associated columns in the Slater determinant, and the switching of two columns flips the sign of the determinant. Also, if two or more electrons are identical, then $\phi(q) = 0$, which represents the nonexistent state, echoing the Pauli exclusion principle.

Then, Eq.(2.12) is modified with the Slater matrix; all the particles fail to differentiate, so all of them are represented by q_1 and q_2:

$$\langle\phi(q)|\mathcal{H}|\phi(q)\rangle = \sum_i \int dq_1 \phi_i^*(q_1)\, h_i \phi_i(q_1) + \frac{1}{2}\sum_{i,j} dq_1 dq_2 \frac{|\phi_i(q_1)|^2 \left|\phi_j^*(q_2)\right|^2}{|r_1 - r_2|} - \frac{1}{2}\sum_{i,j} dq_1 dq_2 \frac{\phi_i^*(q_1)\phi_i(q_2)\phi_j^*(q_2)\phi_j(q_1)}{|r_1 - r_2|}. \tag{2.18}$$

Still, by utilizing the variational principle and substituting Eq.(2.17), as in

$$\delta\left[E - \sum_i \varepsilon_{ij}\left(\left\langle\phi_i\middle|\phi_j\right\rangle - \delta_{ij}\right)\right] = 0,$$

we get

$$[-\nabla^2 + \mathcal{V}(r_1)]\phi_i(q_1) + \sum_j \int dq_2 \frac{\left|\phi_j(q_2)\right|^2}{|r_i - r_j|}\phi_i(q_1) - \sum_j \int dq_2 \frac{\phi_j^*(q_2)\phi_i(q_2)}{|r_i - r_j|}\phi_j(q_1) = \varepsilon_{ij}\,\phi_j(q_1). \tag{2.19}$$

Separating $\phi(q)$ to $\varphi(r)$ and $\sigma(\omega)$ following Eq.(2.17), the second term in Eq.(2.19) will vanish because of the orthogonality of $\sigma(\omega)$, whose values are $+\frac{1}{2}$ and $-\frac{1}{2}$ and so they cancel each other out in the integral, and the third term leaves the part where q_2 and q_1 are parallel spins. Replacing $\phi(q)$ with $\varphi(r)$ and putting $\|$ denotes parallel spinning in Eq.(2.19), and thus we obtain

(2.20)

$$[-\nabla^2 + \mathcal{V}(\mathrm{r}_1)]\varphi_\mathrm{i}(\mathrm{r}_\mathrm{i}) + \sum_{\mathrm{j}(\neq\mathrm{i})} \int \mathrm{dr}_2 \frac{\left|\varphi_2(\mathrm{r}_2)\right|^2}{\left|\mathrm{r}_\mathrm{i} - \mathrm{r}_\mathrm{j}\right|} \varphi_\mathrm{i}(\mathrm{r}_2) - \sum_{\mathrm{j}(\neq\mathrm{i}),\Vert} \int \mathrm{dr}_2 \frac{\varphi_\mathrm{j}^*(\mathrm{r}_2)\varphi_\mathrm{i}(\mathrm{r}_2)}{\left|\mathrm{r}_\mathrm{i} - \mathrm{r}_\mathrm{j}\right|} \varphi_\mathrm{j}(\mathrm{r}_1)$$
$$= E_i\, \varphi_\mathrm{i}(\mathrm{r}_1).$$

This is the Hartree-Fock (HF) equation for a single electron. Compared with the Hartree equation (2.15), the HF equation has one more term called the exchange interaction.

We can rearrange the second term of Eq.(2.20) as

$$\sum_{\mathrm{j}(\neq\mathrm{i})} \int \mathrm{dr}_2 \frac{\left|\varphi_\mathrm{j}(\mathrm{r}_2)\right|^2}{\left|\mathrm{r}_\mathrm{i} - \mathrm{r}_\mathrm{j}\right|} \varphi_\mathrm{i}(\mathrm{r}_1) = \sum_{\mathrm{j}} \int \mathrm{dr}_2 \frac{\left|\varphi_\mathrm{j}(\mathrm{r}_2)\right|^2}{\left|\mathrm{r}_\mathrm{i} - \mathrm{r}_\mathrm{j}\right|} \varphi_\mathrm{i}(\mathrm{r}_1) - \int \mathrm{dr}_2 \frac{\left|\varphi_\mathrm{i}(\mathrm{r}_2)\right|^2}{\left|\mathrm{r}_\mathrm{i} - \mathrm{r}_\mathrm{j}\right|} \varphi_\mathrm{i}(\mathrm{r}_1). \tag{2.21}$$

The first term in Eq.(2.21) is the interaction of $\varphi_\mathrm{i}(\mathrm{r}_1)$ with all electrons including itself, and the second term is the inexistent interaction of the electron with itself:

where

$$\int \mathrm{dr}_2 \frac{\left|\varphi_\mathrm{i}(\mathrm{r}_2)\right|^2}{\left|\mathrm{r}_\mathrm{i} - \mathrm{r}_\mathrm{j}\right|} \varphi_\mathrm{i}(\mathrm{r}_1), \tag{2.23}$$

$$\sigma_\mathrm{i}^\mathrm{H}(\mathrm{r}_2) = -\left|\varphi_\mathrm{i}(\mathrm{r}_2)\right|^2 \tag{2.24}$$

$$\int \mathrm{dr}_2\, \sigma_\mathrm{i}^\mathrm{H}(\mathrm{r}_2) = -1. \tag{2.25}$$

The last term in Eq.(2.20) can be rewritten as

$$-\sum_{\mathrm{j}(\neq\mathrm{i}),\Vert} \int \mathrm{dr}_2 \frac{\varphi_\mathrm{j}^*(\mathrm{r}_2)\varphi_\mathrm{i}(\mathrm{r}_2)}{\left|\mathrm{r}_\mathrm{i} - \mathrm{r}_\mathrm{j}\right|} \varphi_\mathrm{i}(\mathrm{r}_1)$$
$$= -\sum_{\mathrm{j},\Vert} \int \mathrm{dr}_2 \frac{\varphi_\mathrm{j}^*(\mathrm{r}_2)\varphi_\mathrm{i}(\mathrm{r}_2)}{\left|\mathrm{r}_\mathrm{i} - \mathrm{r}_\mathrm{j}\right|} \varphi_\mathrm{j}(\mathrm{r}_2) + \int \mathrm{dr}_2 \frac{\left|\varphi_\mathrm{i}(\mathrm{r}_2)\right|^2}{\left|\mathrm{r}_\mathrm{i} - \mathrm{r}_\mathrm{j}\right|} \varphi_\mathrm{i}(\mathrm{r}_1). \tag{2.26}$$

The absolute value of the second term in Eq.(2.26) is the same as one of the second terms in Eq.(2.21) but in the opposite sign so that these two terms cancel out each other when summing them up.

We rearrange the first term of Eq.(2.26) as

$$-\sum_{j,\|}\int \mathrm{d}r_2 \frac{\varphi_j^*(r_2)\varphi_i(r_2)}{|r_i - r_j|}\varphi_j(r_2) = \int \mathrm{d}r_2 \frac{\sigma_i^{HF}(r_1,r_2)}{|r_i - r_j|}\varphi_i(r_1), \tag{2.27}$$

where

$$\sigma_i^{HF}(r_1,r_2) = -\sum_{j,\|}\frac{\varphi_j^*(r_2)\varphi_i(r_2)\varphi_i^*(r_1)\varphi_j(r_1)}{\varphi_i^*(r_1)\varphi_j(r_2)} \tag{2.28}$$

$$\int \mathrm{d}r_2\, \sigma_i^{HF}(r_1,r_2) = -1 \tag{2.29}$$

σ_i^{HF} can be interpreted as the exchange electron density; its integral of space is also -1. σ_i^{HF} differs from σ_i^{H} in that $\varphi_i(r_1)$is distributed in the same space with the other n-1 electrons and that σ_i^{HF} involves the location of electron r, proving the moving electron has a relation with the spinning due to the Pauli exclusion principle.

By substituting $\sigma(r_2) = \sum_i \sigma_i^{H}(r_2)$ and $\sigma^{HF}(r_1,r_2) = \sum_i \sigma_i^{H}(r_2)$ into Eq.(2.20), we get

$$\left\{-\nabla^2 + \mathcal{V}(r_1) - \int \mathrm{d}r_2 \frac{\sigma(r_2) - \sigma^{HF}(r_1,r_2)}{|r_1 - r_2|}\right\}\varphi_i(r_1) = E_i\,\varphi_i(r_1). \tag{2.30}$$

The third term in Eq.(2.20), which is the function of r_1 after the integral, combines with the second term, forming the effective potential $\mathcal{V}_{\mathrm{eff}}$, and, thus, Eq.(2.30) can be written as

$$[-\nabla^2 + \mathcal{V}_{\mathrm{eff}}(r)]\varphi_i(r) = E_i\,\varphi_i(r), \tag{2.31}$$

where

$$\mathcal{V}_{\mathrm{eff}}(r) = \mathcal{V}(r_1) - \int \mathrm{d}r_2 \frac{\sigma(r_2) - \sigma^{HF}(r_1,r_2)}{|r_1 - r_2|}.$$

At this point, the multi-electron Schrödinger equation is simplified as the single-electron effective potential equation by the Hartree–Fock approximation, which includes the exchange interaction between two electrons but lacks the exclusion interaction between two electrons with antiparallel spins. The

exclusion interaction means the density of electrons is not $\sigma(r_2)$ but smaller than $\sigma(r_2)$ if r_1 is occupied by one electron, that is, electronic correlation.

To get around the issue of calculating coefficients that minimize the resultant wave functions energy in SCF calculations, Roothann and Hall[16,17] claimed one-electron wave functions φ_j can be a linear combination of a non-orthonormal basis set with a Gaussian type or Slater type as

in

$$\varphi_j = \sum_m c_{mj}\phi_m, \tag{2.32}$$

where ϕ_m denotes an either Gaussian type or Slater type basis set.

At this point, the HF-SCF matrix is built up, and the above equations describing the molecular orbital expansion coefficient based on the variational principle have been derived.

The Hartree–Fork theory fully materializes, standing as the pillar for quantum chemistry in the following steps:

$$\left(\frac{-h^2}{8\pi^2 m}\nabla + \mathcal{V}\right)\Psi_{tot}(X,t) = \frac{ih}{2\pi}\frac{\partial\Psi_{tot}(X,t)}{\partial t}$$

↓ Time-Independent

$$\left(\frac{-h^2}{8\pi^2 m}\nabla + \mathcal{V}\right)\Psi_{tot}(X) = E\Psi_{tot}(X)$$

↓ Born-Oppenheimer Approximation

$$\mathcal{H}_e(R) = E(R)\Psi_e(R,r)$$

↓ Independent Electrons

$$h_i\varphi_i = \varepsilon_i\varphi_i$$

2.6 The Exchange-Correlation Energy Problem and Post-Hartree-Fock

The Hartree–Fock theory has achieved massive success based on the assumption of the independent electron; however, the assumption of the independent electron also poses limitations to the Hartree–Fock theory because it is insufficient to describe the correlation between electrons that results from the anti-symmetrization of the wave function within the molecular system. Three main approaches to combat the limitations are elucidated in the sections below.

2.6.1 Configuration Interaction (CI)

In contrast to the Hartree–Fock approach, in which the wave function is described as one single determinant or called configuration, configuration interaction (CI) utilizes a variational wave function composed of a linear combination of configuration state functions. Each configuration state function is built from spin orbitals, and the full configuration interaction method spans all possible substituted configurations; it can be written as

$$\Psi = \sum_{I} C_I \Psi_I^{SO} = C_0 \Psi_0^{SO} + C_1 \Psi_1^{SO} + \cdots, \tag{2.33}$$

where Ψ is usually the electronic ground state of the system, Ψ_I^{SO} is the configuration state and C_I is associated with the weight.

The first configuration state function in the above expansion is normally the Hartree–Fock determinant. The rest can be replaced with some spin orbitals using virtual orbitals from the Hartree–Fock determinant. For example, the single excitation determinant is built by swapping only one spin orbital, the double excitation determinant by two, and so on.

Although the full CI is the complete non-relativistic treatment of the molecular system, listing all configuration is only feasible for the tiny system. In practice, the CI expansion is usually truncated at a certain point in the substitution[18–20]. For example, in configuration interaction with single excitations, CI expansion is truncated after expanding only a single excitation. Other typical calculations include coupled-cluster single and double excitation (CCSD) as well as Brueckner double (BD) excitation.

2.6.2 Møller-Plesset Perturbation Theory

Møller-Plesset perturbation theory is the second-most famous post-Hartree-Fock(post-HF) approach[21]. In this technique, the total Hamiltonian considers an unperturbed Hamiltonian operator $\mathcal{H}_0$., to which a small perturbation $\mathcal{V}$. denoted correlation potential is added:

$$\mathcal{H} = \mathcal{H}_0 + \lambda\mathcal{V}, \tag{2.34}$$

where λ is an arbitrary real parameter that controls the size of the perturbation.

Since the perturbation is usually small, the perturbed wave function and energy can be expanded as a power series of ψ and *E* respectively in λ :

$$\Psi = \lim_{m\to\infty} \sum_{i=0}^{m} \lambda^i \psi^{(i)}$$

$$E = \lim_{m\to\infty} \sum_{i=0}^{m} \lambda^i E^{(i)}.$$

By substituting those perturbed wave functions and energies back into Eq.(2.34),we get

$$(\mathcal{H}_0 + \lambda\mathcal{V}) \sum_{i=0}^{m} \lambda^i \psi^{(i)} = \sum_{i=0}^{m} \lambda^i E^{(i)} \sum_{i=0}^{m} \lambda^i \psi^{(i)}. \tag{2.35}$$

The kth-order perturbation equation is singled out by equating the factors of λ in the left-hand and right-hand side of the equation. Eq.(2.35) can also be truncated at a different level because of the trade-off between computational cost and accuracy. If it is truncated at the second order, it is abbreviated as MP2. MPs are not as computationally costly as the CI, but are still genuinely constrained in their practical use.

2.6.3 Semi-Empirical Methods

Semi-empirical methods rely on Hartree–Fock formalism yet make many approximations, whereas the full Schrödinger equation is truncated to the simplified version and obtains some parameters from empirical data to address errors generated by simplification. A variety of numbers and types constitute

many semi-empirical methods, which can be classified into two main categories. The first one makes corrections limited to π-electrons only, such as in the Hückel method that treats the Hamiltonian of the conjugated system as a one-electron Hamiltonian and decides corrections experimentally. The other category is restricted to the *s* and *p* atomic orbitals and contains a larger number of methods, such as MNDO, AM1, and PM3.

While having the shortcoming of requiring a good estimation of empirical parameters, semi-empirical methods are comparatively cheap, and thus are ideal for large systems with hundreds of atoms.

2.7 Density-Functional Theory

Even though the Hartree–Fock method remains the pillar SCF method in computational chemistry in which the exact spin orbitals are approximated by adding a finite set of functions, the accuracy of calculation stems from a considerable amount of basis sets and plenty of iterative cycles, demanding a vast of amount of computation resources.

Density-functional theory (DFT) was created to solve two problems in Hartree–Fock approaches: 1) correlation and 2) inefficiency overcome without loss of accuracy.

2.7.1 Thomas-Fermi-Dirac Approximation

The Thomas–Fermi–Dirac(TFD) model was developed by Thomas[22], Fermi [23], and Dirac [24] in 1927–1930. Rather than using the wave function as in the Hartree–Folk method, this method uses electron

$$\mathrm{E_{TFD}}[\sigma(\mathrm{r})] = \mathrm{A_1}\int \sigma(\mathrm{r})^{5/3}\,\mathrm{dr} + \int \sigma(\mathrm{r})\mathcal{V}_{\mathrm{ext}}(\mathrm{r})\mathrm{dr} + +\frac{1}{2}\int\int \frac{\sigma(\mathrm{r})\sigma(\mathrm{r}')}{|\mathrm{r}-\mathrm{r}'|}\mathrm{dr}\mathrm{dr}' + \mathrm{A_2}\int \sigma(\mathrm{r})^{4/3}\,\mathrm{dr}, \tag{2.36}$$

density $\sigma(\mathrm{r})$, which is the basic variable of the function, to represent the total energy of a system:

where the first term is the kinetic energy of electrons assumed to interact with each other in a homogeneous electron gas, and the last term is a local exchange term. A_1 and A_2 are constants used to make the approximation. The second and third terms are nucleus–electron Coulomb interaction and classical electron–electron Coulomb repulsion, respectively. The last term denotes exchange and correlation, which includes all the neglected parts in the previous Hartree–Folk method. The novelty of this TF model is that it represents the kinetic and exchange energies of the many-electron system by their uniform electron gas energy densities.

Applying the technique of Lagrange multipliers to Eq.(2.36), the solution of the ground-state density and energy can be found in the stationary condition:

$$\delta\left\{E_{\mathrm{TFD}}[\sigma(\mathrm{r})]-\mu\left(\int \sigma(\mathrm{r})\mathrm{dr}-\mathrm{N}\right)\right\}=0, \tag{2.37}$$

where μ is the Lagrange multiplier and physically represents the chemical potential. After deriving Eq.(2.37) from the TFD equation (2.36), we get

$$\frac{5}{3}\mathrm{A}_1\sigma(\mathrm{r})^{5/3}+\mathcal{V}_{\mathrm{ext}}(\mathrm{r})+\int\frac{\sigma(\mathrm{r}')}{|\mathrm{r}-\mathrm{r}'|}\mathrm{dr}'+\frac{4}{3}\mathrm{A}_2\sigma(\mathrm{r})^{1/3}-\mu=0. \tag{2.38}$$

From here, the ground-state density can be acquired.

The TFD approach is such a crude approximation that it suffers from numerous issues. For instance, assuming electrons do not interact with each other fails to consider the bonding relationships between atoms; thus, the TFD approach cannot be applied to molecules and solids[25]. On the other hand, the TFD approach breaks new ground by using electron density as the primary variable, which is a core part of DFT.

2.7.2 Hohenberg-Kohn (HK) Theorems

Even though the TFD approximation, which is expressed as the function of electronic density $\sigma(\mathrm{r})$, is magnificently simple, it fails in the qualitative aspect in that TFD energies $E_{\mathrm{TFD}}[\sigma(\mathrm{r})]$ have errors of around 10%, which is too large for computational purposes.

In 1964, Hohenberg, Kohn and Sham[26] developed the theory that finally legitimized the application of electron density. Their work is composed of two theorems acting as the foundation of DFT. The first theorem states that a unique electron density defines the ground-state energy from the Schrödinger equation, implying that the relationship between the ground-state wave function and the ground-state electron density is a one-to-one mapping. The second defines a decisive property of the functional: Among all solutions of the Schrödinger equation, the one minimizing the overall functional energy is the actual electron density.

The full energy under the Hohenberg-Kohn (HK) theorem can be defined as

$$E[\sigma] = \mathcal{T}[\sigma] + \mathcal{U}[\sigma] + \mathcal{V}_{\text{ext}}[\sigma]. \tag{2.39}$$

$E[\sigma]$, a functional of $\sigma(r)$, denotes the ground-state energy of the system, $\mathcal{T}$ represents the kinetic energy of the electrons and $\mathcal{V}_{\text{ext}}$ is the interaction between the external potential, which could be the potential from the nuclei or an external field such as an electric or magnetic field. $\mathcal{U}[\sigma]$ represents the classic electrostatic energy (also called the Hartree) of electrons. The kinetic energy and electrostatic energy are determined by the density function because the density function uniquely produces kinetic and electrostatic operators. Kinetic and electrostatic energy introduce simplification and approximation to the DFT.

2.7.3 Kohn-Sham Ansatz

Despite the fact that the HK theorems introduced a novel paradigm into quantum mechanics, the functional $E[\sigma(r)]$ still conceptually exists but is impossible to write down. This problem was solved by the Kohn–Sham (KS) ansatz.[27]

The KS ansatz replaces the original functional, which has the electronic mutual interaction, with the new and separate functional, which gets rid of electronic mutual interaction, and to put all the difference of the new functional off the original and exact functional into the functional $E_{XC}[\sigma(r)]$. That is to say, the true kinetical energy $\mathcal{T}[\sigma]$ and true electrostatic energy $\mathcal{U}[\sigma]$, both of which fail the calculation, are replaced with fictitious and non-interacting $\mathcal{T}_{\mathrm{S}}[\sigma(\mathrm{r})]$ and $\mathcal{U}_{\mathrm{S}}[\sigma(\mathrm{r})]$:

$$E_{XC}[\sigma(r)] = \mathcal{T}[\sigma(r)] - \mathcal{T}_S[\sigma(r)] + \mathcal{U}[\sigma(r)] - \mathcal{U}_s[\sigma(r)]. \tag{2.40}$$

By substituting Eq.(2.40) to Eq.(2.39), we obtain:

$$E[\sigma(r)] = \mathcal{T}_S[\sigma(r)] + \mathcal{U}_s[\sigma(r)] + \mathcal{V}_{\mathrm{ext}}[\sigma(r)] + E_{XC}[\sigma(r)]. \tag{2.41}$$

$[\sigma(r)]$ contains both known distinctions as well as unknown distinctions, such as the non-classical contribution to the electron-electron interactions, making the ground state energy $E[\sigma]$ exact.

After the original interacting system with real potential is mapped onto an auxiliary multi-particle system where each electron moves independently, the idea solving the Hartree–Fock equation is discussed in section (2.5).

The density of the auxiliary system is calculated using

$$\sigma(\mathrm{r}) = \sum_{\mathrm{i}=1}^{\mathrm{N}} |\psi_{\mathrm{i}}(\mathrm{r})|^2, \tag{2.42}$$

in which $\sigma(r)$ is subject to the conservation condition

The non-interacting independent-particle kinetic energy $\mathcal{T}_S[\sigma(r)]$ is given by

$$\int \sigma(r)\, dr = N. \tag{2.43}$$

$$\mathcal{T}_{\mathrm{S}}[\sigma(\mathrm{r})] = -\sum_{\mathrm{i}=1}^{\mathrm{N}} \int \psi_{\mathrm{i}}^{*}(\mathrm{r}) \nabla \psi_{\mathrm{i}}(\mathrm{r}) \mathrm{dr}, \tag{2.44}$$

and $\mathcal{U}_s[\sigma(r)]$ is derived from the classic electrostatic energy of the electrons:

The variational principle used in the Hartree–Fock equation is also applied here by minimizing the energy

$$\mathcal{U}_s[\sigma(\mathrm{r})] = -\frac{1}{2}\int\int \frac{\sigma(\mathrm{r})\sigma(\mathrm{r}')}{|\mathrm{r}-\mathrm{r}'|}\mathrm{d}\mathrm{r}\mathrm{d}\mathrm{r}'. \tag{2.45}$$

functional $E[\sigma(r)]$ subject to the constraint of Eq.(2.41):

$$\delta\left\{E\,[\sigma(r)] + \mathcal{E}(\int \sigma(r)\,dr - N)\right\} = 0. \tag{2.46}$$

where $\mathcal{E}$ is the Lagrange multiplier.

By substituting Eq.(2.44) and Eq.(2.45) with Eq.(2.41) and deriving using $\delta\psi_i^*$, we get

$$\begin{aligned}\delta\mathrm{E}\,[\sigma(\mathrm{r})] = -\sum_{\mathrm{i}=1}^{\mathrm{N}}\int \delta\psi_\mathrm{i}^*(\mathrm{r})\nabla\psi_\mathrm{i}(\mathrm{r}) \\ + \int\int \frac{\delta\psi_\mathrm{i}^*(\mathrm{r})\psi_\mathrm{i}(\mathrm{r})\sigma(\mathrm{r}')}{|\mathrm{r}-\mathrm{r}'|}\mathrm{d}\mathrm{r}\mathrm{d}\mathrm{r}' + \sum_{\mathrm{i}=1}^{\mathrm{N}}\int \mathcal{V}(\mathrm{r})\delta\psi_\mathrm{i}^*(\mathrm{r})\psi_\mathrm{i}(\mathrm{r})dr + \delta\mathrm{E}_{\mathrm{XC}}[\sigma(\mathrm{r})].\end{aligned} \tag{2.47}$$

We can rearrange $\delta\mathrm{E}_{\mathrm{XC}}[\sigma(\mathrm{r})]$ as

$$\delta\mathrm{E}_{\mathrm{XC}}[\sigma(\mathrm{r})] = \frac{\delta\mathrm{E}_{\mathrm{XC}}[\sigma(\mathrm{r})]\delta\sigma(\mathrm{r})}{\delta\sigma(\mathrm{r})} = \sum_{\mathrm{i}=1}^{\mathrm{N}}\int \frac{\delta\mathrm{E}_{\mathrm{XC}}[\sigma(\mathrm{r})]\delta\psi_\mathrm{i}^*(\mathrm{r})\psi_\mathrm{i}(\mathrm{r})}{\delta\sigma(\mathrm{r})}. \tag{2.48}$$

We can then substitute Eq.(2.47) and Eq.(2.48) with Eq.(2.46):

$$\begin{aligned}\sum_{\mathrm{i}=1}^{\mathrm{N}}\int \mathrm{d}\mathrm{r}\,\delta\psi_\mathrm{i}^*(\mathrm{r})\left\{-\nabla\psi_\mathrm{i}(\mathrm{r}) + \int \frac{\psi_\mathrm{i}(\mathrm{r})\sigma(\mathrm{r}')}{|\mathrm{r}-\mathrm{r}'|}\mathrm{d}\mathrm{r}' + \mathcal{V}(\mathrm{r})\psi_\mathrm{i}(\mathrm{r}) + \frac{\delta\mathrm{E}_{\mathrm{XC}}[\sigma(\mathrm{r})]\psi_\mathrm{i}(\mathrm{r})}{\delta\sigma(\mathrm{r})} - \varepsilon_\mathrm{i}\psi_\mathrm{i}(\mathrm{r})\right\} \\ = 0.\end{aligned} \tag{2.49}$$

The term in the curly brackets must equal to zero, so we obtain

$$\left\{\nabla + \int \frac{\psi_\mathrm{i}(\mathrm{r})\sigma(\mathrm{r}')}{|\mathrm{r}-\mathrm{r}'|}\mathrm{d}\mathrm{r}' + \mathcal{V}(\mathrm{r}) + \frac{\delta E_{XC}[\sigma(r)]}{\delta\sigma(r)}\right\}\psi_i(r) = \varepsilon_i\psi_i(r). \tag{2.50}$$

This is the equation for the one-electron wave function, and it can be rewritten as

$$\left(-\frac{1}{2}\nabla^2 + \mathcal{V}_{KS}(r)\right)\psi_i(r) = \varepsilon_i\psi_i(r), \tag{2.51}$$

where

$$\mathcal{V}_{KS}(r) = \int \frac{\psi_i(\mathrm{r})\sigma(\mathrm{r}')}{|\mathrm{r}-\mathrm{r}'|}\mathrm{dr}' + \mathcal{V}(\mathrm{r}) + \frac{\delta E_{XC}[\sigma(r)]}{\delta\sigma(r)}. \tag{2.52}$$

Eq. (2.51), (2.42), and (2.43) are the well-known KS equations and can be solved by self-consistent method because $\mathcal{V}_{KS}(r)$ depends on the density through the XC potential. In order to calculate the density, the *N* equations in Eq.(2.51) must be solved using KS theory as opposed to the one equation in the TF approach.

The XC potential $\mathcal{V}_{XC}(r)$can be expresses as

$$\mathcal{V}_{XC}(r) = \frac{\delta E_{XC}[\sigma(r)]}{\delta\sigma(r)}, \tag{2.53}$$

where there is one electron in each of the N orbitals $\psi_i(\mathrm{r})$ with the lowest eigenvalues ε_i.

Kohn–Sham equations, such as Eq.(2.51), are often solved by expanding the single-electron wave function into a finite basis set, as in the Hartree–Fock method:

$$\psi_i = \sum_{\alpha}^{2M} c_{\alpha i}\,\phi_\alpha, \tag{2.54}$$

Eq.(2.54) is now transformed into finding the LCAO expansion coefficients $\{c_{\alpha i}\}$.

However, a benefit of the KS method over the Hartree–Fock approach is that as *N*, the count of electron s, increases, only the number of Eq.(2.51), the independent single-electron equation, increases accordingly, so the complexity of the system does not increases dramatically

2.7.4 Exchange-correlation Functional Approximations

After the production of Eq.(2.52), approximating the $E_{XC}[\sigma(r)]$ was a core part of the development of the DFT, in which many expressions have been and are still being proposed. Perdew

proposed a hierarchically classified Jacob's Ladder to describe approaches approximating $E_{XC}[\sigma(r)]$ (Fig. 2.1), from the oldest and simplest to more recent and complex[28].

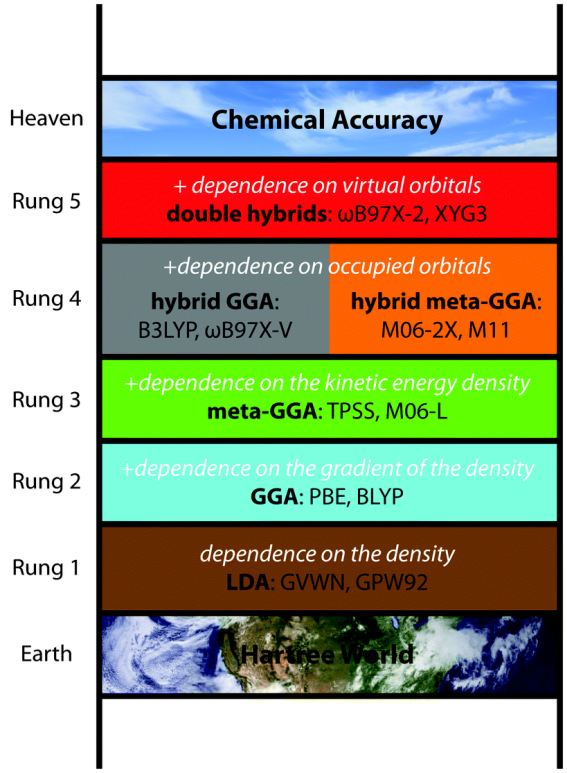


Figure 2.1 Perdew's "Jacob's Ladder." Adapted from Mardirossian. [29]

The ladder spans multiple rungs, with the lowest rung representing the HF approximation where correlation is not considered, to the highest rung of the holy grail, "Chemical Accuracy", which ideally depicts exact quantum properties and incorporates true electron correlations.

The Jacob's ladder rungs are as follows:

- 1st, LDA: Local Density Approximation (LDA) comes from the notion that the simplest expression for $E_{XC}[\sigma(r)]$ is based on the intuition that exchange-correlation energy depends solely upon the value of the electronic density at each point in space, as in

$$E_{XC}[\sigma(r)] = \int dr\sigma(r)\,\epsilon_{xc}[\sigma(r)],$$

where $\epsilon_{xc}[\sigma(r)]$ is the exchange-correlation energy for one electron.

Although the idea of the uniform electron gas is overwhelmingly successful in approximating homogeneous, non-defective systems with a uniform distribution of charge density, but fizzles at inhomogeneous systems.

- 2nd, GGA: Generalized Gradient Approximation (GGA) acts as a more precise approximation for $E_{XC}[\sigma(r)]$ than LDA. It is described as

$$E_{XC}[\sigma(r)] = \int dr\sigma(r)\,\epsilon_{xc}[\sigma(r), \nabla\sigma(r)].$$

It is the gradient, which is the first derivative with respect to r, of the electron density that makes GGA ideal at estimating the nonuniform distribution of charge density.Furthermore, GGA usually has an enhancement factor F_x^{en} to enhance the local density to represent inhomogeneous systems:

$$E_{XC}[\sigma(r)] = \int dr\sigma(r)\,\epsilon_{xc}[\sigma(r)]F_x^{en}\big[s\big(\nabla\sigma(r)\big)\big].$$

Here,$s\big(\nabla\sigma(r)\big)$is the reduced density gradient.

- 3rd, meta-GGA: Meta-GGA goes further, taking the Laplacian of density $\nabla^2\sigma(r)$ into consideration. Recently, more typical mGGAs include the Laplacian of occupied orbitals (τ), as in

$$E_{XC}[\sigma(r)] = \int dr\sigma(r)\,\epsilon_{xc}[\sigma(r), \nabla\sigma(r), \tau(r)],$$

where τ, called kinetic energy density, is:

$$\tau(r) = \sum_i^{occ} \frac{1}{2}|\nabla\psi_i(r)|^2.$$

- 4th, hyper-GGA: these functionals were devised to have the partial or full HF exact exchange beyond the DFT exchange-correlation terms based on the HF/KS method which is described in

Section 2.6. A lot of common hybrid functionals are considered semi-empirical approximations to this 4th rung.

- 5th rung: the last rung entails functionals that contribute to the exact exchange and a partial exact correlation. This is accomplished by considering unoccupied (virtual) orbitals.

In reality, the widely used post-HF-like techniques normally put HF and DFT together, approximating $E_{XC}[\sigma(r)]$ in DFT and incorporating a portion of the exact exchange from the Hartree–Folk method with the rest of the exchange-correlation energy from other sources, such as first-principle or empirical, creating a hybrid semi-empirical way to obtain such correlation. Not only do Post-HF-like techniques bring about accurate approximation, they are also computationally costly. One of the most popular versions is B3LYP, which stands for "Becke, 3-parameter, Lee-Yang–Parr", which uses both GGA and LDA approximation[30,31].

2.8 NMR Calculation

2.8.1 Chemical Shifts

Starting from the basic principles of electromagnetism, nuclear magnetic shielding depicts all electron–nuclear moment interaction with an applied magnetic field, with the electron spin excluded. It can be found as a different part between the applied magnetic field and the actual field at the site of the nucleus.

The energy of the interaction between the nuclear moment and the shielding field is neglectable compared to the electron orbital energies; therefore, interaction between the nuclear moment is often considered as a perturbation.

When a magnetic field is present, the momentum of an electron transforms into

$$p = (-i\hbar)\nabla - (e/c)A, \tag{2.55}$$

where ∇ is the gradient operator, and A is the magnetic vector potential. A is related to the magnetic field B by:

$$B = \nabla \times A. \tag{2.56}$$

For diamagnetic molecules where all electrons are paired, the magnetic interactions in the Hamiltonian are expressed as

$$H - V = \frac{1}{2m_e}[(-i\hbar)\nabla - (e/c)A]^2. \tag{2.57}$$

When a nuclear moment is present, A is expressed as the sum of two vector potentials, one associated with the external field, A_0, and another one referring to the field generated by the nuclear moment, A_{n}:

$$A = A_0 + A_n. \tag{2.58}$$

Eq.(2.57) can be expressed in terms of these two vector potentials by introducing Eq.(2.58). For simplicity, a new quantity is introduced:

$$H - V = \frac{\Gamma^2}{2m_e} - \frac{e}{2m_e c}[\Gamma \cdot A_n + A_n \cdot \Gamma] + \frac{e^2}{2m_e c^2}A_n^2, \tag{2.59}$$

where

$$\Gamma = (-i\hbar)\nabla - (e/c)A_0. \tag{2.60}$$

The first term on the right-hand side of Eq.(2.59), $(\Gamma^2/2m_e)$, contains the kinetic energy of the electron,

$$K_e = \left[\left(\frac{-i\hbar}{2m_e}\right)\nabla\right]^2; \tag{2.61}$$

the diamagnetic energy,

$$D_e = \frac{e^2}{2m_e c^2}{A_0}^2; \tag{2.62}$$

and the orbital Zeeman energy of the electron,

$$Z_e = (i\hbar)\frac{e}{2m_e c}(\nabla \cdot A_0 + A_0 \cdot \nabla). \tag{2.63}$$

Thus, the first-order perturbation Hamiltonian can be described as

Because the second term is independent of the applied field and quadratic in the nuclear moment, the first-

$$\mathcal{H}^1 = \left(\frac{e^2}{2m_e c^2}\right)(A_0 \cdot A_n + A_n A_0) + \left(\frac{e^2}{2m_e c^2}\right)A_n^2. \tag{2.64}$$

order perturbation the energy can be written as:

$$E^1 = \left(\frac{e^2}{2m_e c^2}\right)\langle 0|A_0 \cdot A_n + A_n \cdot A_0|0\rangle, \tag{2.65}$$

where $|0\rangle$ denotes the Hartree–Fock ground state wave function.

The magnetic vector potential attributed to the field of the nuclear magnetic dipole moment μ_n is expressed as

$$A_n(r) = \frac{(\mu_n \times r_n)}{r_n^3}. \tag{2.66}$$

Where r_n has the nucleus as its origin. the magnetic vector potential at the presence of the applied external magnetic field can be defined at any position r_0:

$$A_0 = \frac{1}{2}(B_0 \times r_0). \tag{2.67}$$

By substituting Eqs (2.66) and (2.67) into (2.65), we obtain

$$E^1 = \left(\frac{e^2}{2m_e c^2}\right)\mu_n \cdot \left\langle 0\left|\frac{(r_0 \times r_n)}{r_n^3}\right|0\right\rangle \cdot B_0. \tag{2.68}$$

$$\sigma_{ij}^d = \left(\frac{e^2}{2m_e c^2}\right)\left\langle 0\left|\frac{j \times r_0 r_n \times i}{r_n^3}\right|0\right\rangle, \tag{2.69}$$

This summarizes the result of a first-order perturbation approach.

In liquid, gas-phase and solution NMR, free tumbling motion of the molecules results in one value of the shielding property. It is defined as the isotropic shielding value, which is the average of the principal components,

$$\sigma_{iso} = \frac{(\sigma_{11} + \sigma_{22} + \sigma_{33})}{3}.$$

Lastly, based on the definition of the shielding in Eq.(2.69), the absolute shielding is not observable in NMR experiments. Therefore, some compounds have been chosen as convenient references or standards, such as tetramethylsilane (TMS) as H^1 and C^{13}. However, chemical shift, δ, which is the normally reported quantity, is related to the shielding property as follows:

$$\delta = \frac{(\nu - \nu_{ref})}{\nu_{ref}} = \frac{\sigma_{ref} - \sigma}{1 - \sigma_{ref}}.$$

where *v* is the measured resonance frequency directly.

2.8.2 Ab initio Calculations of the Shielding Property

The shielding property is built on the ground and excited electronic states and their respective energies. In this method, even though both diamagnetic and paramagnetic terms depend on this choice of origin, their sum is supposed to be independent of the choice of origin. When the choice of gauge origin has a shift, additional terms will be present in both the diamagnetic and paramagnetic parts. These terms in the diamagnetic and paramagnetic parts is the equal value but with opposite signs.

Without an applied external magnetic field, the ground and excited electronic states can be calculated by a self-consistent field (SCF) calculation. These electronic states in the SCF calculation can be presented by the linear combinations of atomic orbitals (LCAO), which is denoted as molecular orbitals. Mathematically, atomic orbitals usually presented by the basis set, Gaussian functions, due to it is easy to differentiate and integrate. Though these states can be used to calculate the shielding via the perturbation method, computational cost is proportional to N^4 or $(N\log N)^2$, where *N* presents the number of Gaussian functions, therefore restraining the number of Gaussian functions to be used owe to the limited computational resources. Hence, both ground and excited electronic states end up with being calculated as approximations, diamagnetic and paramagnetic components have different sensitivities to these approximations, errors in both components are hard to cancel each other out.

The first use of SCF wave functions in the computation of shielding was introduced by Stevens et al. Compared with Eq.(2.68) and Eq.(2.69), shielding can be calculated as the second derivative of the molecular energy with respect to an external magnetic field B_0 and the nuclear magnetic moment μ:

$$\sigma_{\alpha\beta} = \frac{\partial^2 E}{\partial \mu_\alpha \partial_{B_\beta}}. \tag{2.70}$$

Calculating E can refer to the Hartree–Fock method discussed in section (2.6) or the DFT approaches from section (2.70).

2.8.3 Gauge Including Atomic Orbital (GIAO) Method

Leveraging theoretical method to calculate the diamagnetic and paramagnetic contributions separately is dependent on the choice of gauge origin; even the total shielding should be independent of the choice of origin. There are additional paramagnetic and diamagnetic terms necessary if gauge origin changes. However, these additional terms are exactly equal value but with opposite signs. However, in practice, errors are present due to the fact that the desired cancellation are hardly occur perfectly. One way of getting around such error, which is proposed by Ditchfield[32], is choosing a gauge origin and using field-dependent atomic orbitals as basis sets. In this set, a multiplicative complex factor determined by the gauge of the vector is added to the original atomic basis function:

$$\Upsilon_j = e^{[-(i/c)A_v \cdot r]}\psi_j,$$

Here, A_v is the value of the vector potential A at the nuclear position R_v of the atomic orbital φ_v: $A_v = \frac{1}{2}B_0 \times R_v$.

This approach is called gauge-including atomic orbitals (GIAO)[33] where each atomic orbital has its own local gauge origin placed at its center so that it can eliminate the task of arbitrarily assigning a gauge origin: The gauge origins are all predetermined. The coupled Hartree–Fock (CHF) equations with GIAOs calculates all the electron integrals. But the previous self-consistent field (SCF) calculation is not working since these integrals exclude the extra multiplicative complicated factor. Hence, although the GIAOs was able to achieve the desired gauge independence, the gauge factors also result in extra time and memory-consuming recalculations of integrals.

2.9 References

1. Einstein, A. Zur Elektrodynamik bewegter Körper. *Ann. Phys.* **322**, 891–921 (1905).

2. Condon, E. U. 60 years of quantum physics. *Phys. Today* **15**, 37 (2009).

3. Zhang, Q.-R. Intensity dependence of the photoelectric effect induced by a circularly polarized laser beam. *Phys. Lett. A* **216**, 125–128 (1996).

4. Bubb, F. W. Direction of Ejection of Photo-Electrons by Polarized X-rays. *Phys. Rev.* **23**, 137–143 (1924).

5. Wagner, H. J. Early Spectroscopy and the Balmer Lines of Hydrogen. *J. Chem. Educ.* **82**, 380 (2005).

6. Ph.D, J. R. R. XXXIV. On the structure of the line-spectra of the chemical elements. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **29**, 331–337 (1890).

7. Ritz, W. On a New Law of Series Spectra. *Astrophys. J.* **28**, 237 (1908).

8. Schrödinger, E. Quantisierung als Eigenwertproblem. *Ann. Phys.* **385**, 437–490 (1926).

9. Aspect, A. & Villain, J. The birth of wave mechanics (1923–1926). *Comptes Rendus Phys.* **18**, 583–585 (2017).

10. The Feynman Lectures on Physics Vol. III Ch. 1: Quantum Behavior. https://www.feynmanlectures.caltech.edu/III_01.html#Ch1-S7.

11. born-lecture.pdf.

12. Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.* **28**, 1049–1070 (1926).

13. Born, M. & Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **389**, 457–484 (1927).

14. The variation method in quantum chemistry - 1st Edition. https://www.elsevier.com/books/the-variation-method-in-quantum-chemistry/epstein/978-0-12-240550-1.

15. Beavis, B. & Dobbs, I. *Optimisation and Stability Theory for Economic Analysis*. (Cambridge University Press, 1990).

16. Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **23**, 69–89 (1951).

17. Hall, G. G. & Lennard-Jones, J. E. The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials. *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.* **205**, 541–552 (1951).

18. Čížek, J. On the Use of the Cluster Expansion and the Technique of Diagrams in Calculations of Correlation Effects in Atoms and Molecules. in *Advances in Chemical Physics* 35–89 (John Wiley & Sons, Ltd, 2007). doi:10.1002/9780470143599.ch2.

19. Purvis, G. D. & Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **76**, 1910–1918 (1982).

20. Pople, J. A., Head-Gordon, M. & Raghavachari, K. Quadratic configuration interaction. A general technique for determining electron correlation energies. *J. Chem. Phys.* **87**, 5968–5975 (1987).

21. Møller, Chr. & Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **46**, 618–622 (1934).

22. Thomas, L. H. The calculation of atomic fields. *Math. Proc. Camb. Philos. Soc.* **23**, 542–548 (1927).

23. Fermi, E. Statistical method to determine some properties of atoms. 7.

24. Dirac, P. a. M. Note on Exchange Phenomena in the Thomas Atom. *Math. Proc. Camb. Philos. Soc.* **26**, 376–385 (1930).

25. Teller, E. On the Stability of Molecules in the Thomas-Fermi Theory. *Rev. Mod. Phys.* **34**, 627–631 (1962).

26. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **136**, B864–B871 (1964).

27. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **140**, A1133–A1138 (1965).

28. Perdew, J. P. & Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.* **577**, 1–20 (2001).

29. Mardirossian, N. & Head-Gordon, M. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **16**, 9904–9924 (2014).

30. Kim, K. & Jordan, K. D. Comparison of Density Functional and MP2 Calculations on the Water Monomer and Dimer. *J. Phys. Chem.* **98**, 10089–10094 (1994).

31. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields | The Journal of Physical Chemistry. https://pubs.acs.org/doi/abs/10.1021/j100096a001.

32. Ditchfield, R. Self-consistent perturbation theory of diamagnetism. *Mol. Phys.* **27**, 789–807 (1974).

33. London, F. Théorie quantique des courants interatomiques dans les combinaisons aromatiques. *J. Phys. Radium* **8**, 397–409 (1937).

34. Ziegler, Tom. Approximate density functional theory as a practical tool in molecular energetics and dynamics. *Chem. Rev.* **91**, 651–667 (1991).

35. Ziegler, T. The 1994 Alcan Award Lecture Density functional theory as a practical tool in studies of organometallic energetics and kinetics. Beating the heavy metal blues with DFT. *Can. J. Chem.* **73**, 743–761 (1995).

36. Wolinski, K., Hinton, J. F. & Pulay, P. Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J. Am. Chem. Soc.* **112**, 8251–8260 (1990).

Chapter 3. Pyridoxal-5′-Phosphate and Tryptophan Synthase

3.1 Pyridoxal-5′-Phosphate

Pyridoxal-5′-phosphate acting as the bioactive form of vitamin B6 is an magnet in the nitrogen metabolism of all organisms.[1] Currently, PLP-dependent enzymes have been proved that it catalyzes over 140 distinct activities and found it is in all free-living organisms. [2,3] PLP is composed of one heteroaromatic pyridine ring, one aldehyde group, a hydroxyl group, and a phosphate group. (PLP; Figure 3.1)

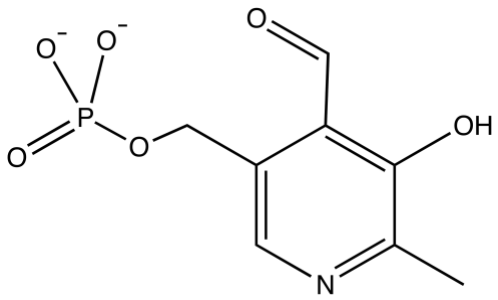


Figure 3.1 Pyridoxal-5'-phosphate.

With the heteroaromatic pyridine ring, it is possible for PLP to make anionic intermediate stable, which is formed in some PLP-dependent enzymes. The pyridine ring acts as an electron sink to stabilized and delocalized the negative electron via resonance in some intermediates of PLP-dependent enzymes. Compared to the analog of pyridine ring-the benzene, which is composed of 6 carbons without the nitrogen in the pyridine ring, unique properties and specific performing reactions of pyridine ring results from the pyridine ring nitrogen. The pronation of the PLP ring nitrogen has been the research topic for PLP-dependent enzymes. The X-ray crystal technique has been used to determine the structures of PLP-dependent enzymes. Different amino acid residues positioned below the pyridine nitrogen showed some PLP-dependent enzymes are indeed protonated but some are not. (In Figure 3.2, the pyridine hydrogen

inside the square brackets represents the uncertainty). The work with these enzymes-but replaced pyridine with benzene, showed reactions of every enzyme was deteriorate in different levels.

The aldehyde group is a critical function group in PLP. An imine is usually formed through the reaction between the aldehyde group and free amino group (Figure 3.2). For example, the reaction of PLP and the conserved lysine residue forms internal aldimine, while the reaction of PLP and substrate amino group forms external aldimine. Furthermore, Internal aldimine and external aldimine are reversible by the attack on another unprotonated primary amino group. (Figure 3.3) This capability enables PLP an ideal catalyzer to go through the cycle of PLP linking, substrate binding, product release and regeneration of enzyme with PLP bond in the active site.

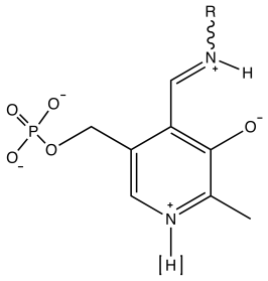
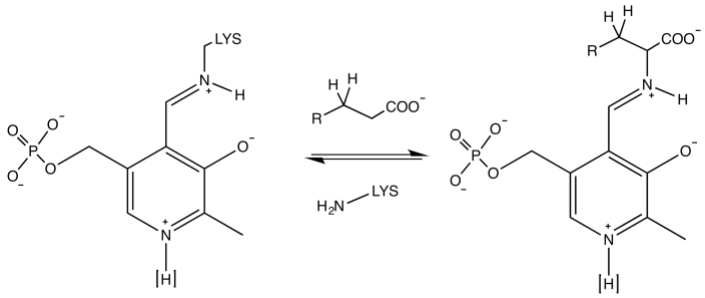


Figure 3.2 PLP-Imine Form.



Internal aldimine

External aldimine

Figure 3.3 Reversion of internal aldimine and external aldimine.

5′-phosphate group has not been explained as much as heteroaromatic pyridine ring and aldehyde group so far. Primary functions of 5′-phosphate group are: (1) In the transaldimination reaction, the 5′-phosphate group is suggested to work with the substrate L-serine hydroxyl group to promote the forming of primary intermediates by making it more stereospecific orientated.[4]

(2)The 5′-phosphate group is also a protonatable group, hence, it is advisable to catalyze acid-base reactions via donating and accepting a proton when some intermediates need a proton to work or speed up.[5]

(3) The 5′-phosphate group can work as the binder to effectively attach to active site of PLP in some exceptional cases as well. [6] Although the function has been rarely discussed, it still represents the necessity of the 5′-phosphate group.

The hydroxyl group normally works as a proton donor or acceptor. Its function will be discussed in depth in the following mechanism parts thoroughly.

With every above decomposed part of PLP combined and coupled with unique enzymatic environment such as active site residues, PLP-dependent enzymes catalyzes a myriad of chemical interactions such as α/β/γ elimination/replacement, [7]racemization,[8,9] transamination,[10] and decarboxylation.[3,11] (Figure 3.4)

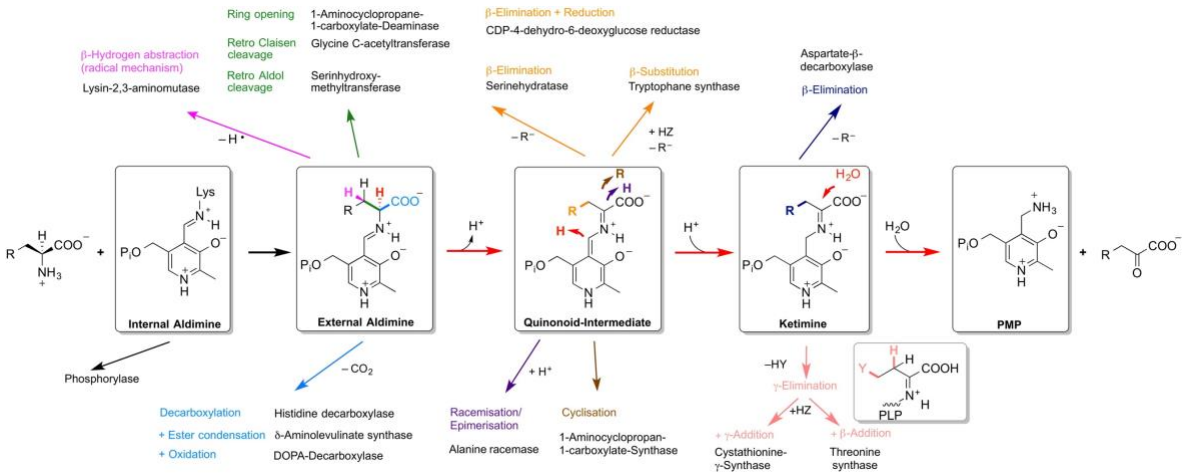


Figure 3.4 PLP-dependent enzymes catalyze a variety of chemical reactions, adapted from Bornscheuer.12

What the α,β-eliminations have in common is the removal of a leaving group at the β position. For example, decarboxylation is done by removing –COO− group which is at the β position of external aldimine intermediate. and an anionic intermediate is produced which can be protonated at Cα position to create an amine. (Figure 3.5.A)

Racemization has nothing to do with leaving group at the β position. It is all concerned with the Cα spot. It starts with the deprotonation of Cα on one side, then ends with reprotonation of Cα on the other side. (Figure 3.5.B)

Differing from racemization's last step which Cα site is reprotonated, transamination accepts a proton at C4′ via the tautomerism of the quinonoid intermediate after the same steps above of racemization. And then in the following step, it is very important for transamination to release a proton at Cα spot and form a ketimine intermediate. (Figure 3.5.C)

Contrary to all above reactions participating with the deprotonation to create the intermediate such as carbanionic or quinonoid intermediate, the retro-aldol condensation breaks covalent bond between Cα and Cβ and creates a carbanionic intermediate.[13] This step marks the retro-aldol condensation a somewhat incomparable reaction.

While PLP catalyzes a flood of reactions, each of which owns distinct attribute and pathway. Some enzymes carry out reactions combined with an assortment of enzymes, e.g., dialkylglycine decarboxylase that performs the transamination, yet it is decarboxylation-dependent.[14,15]

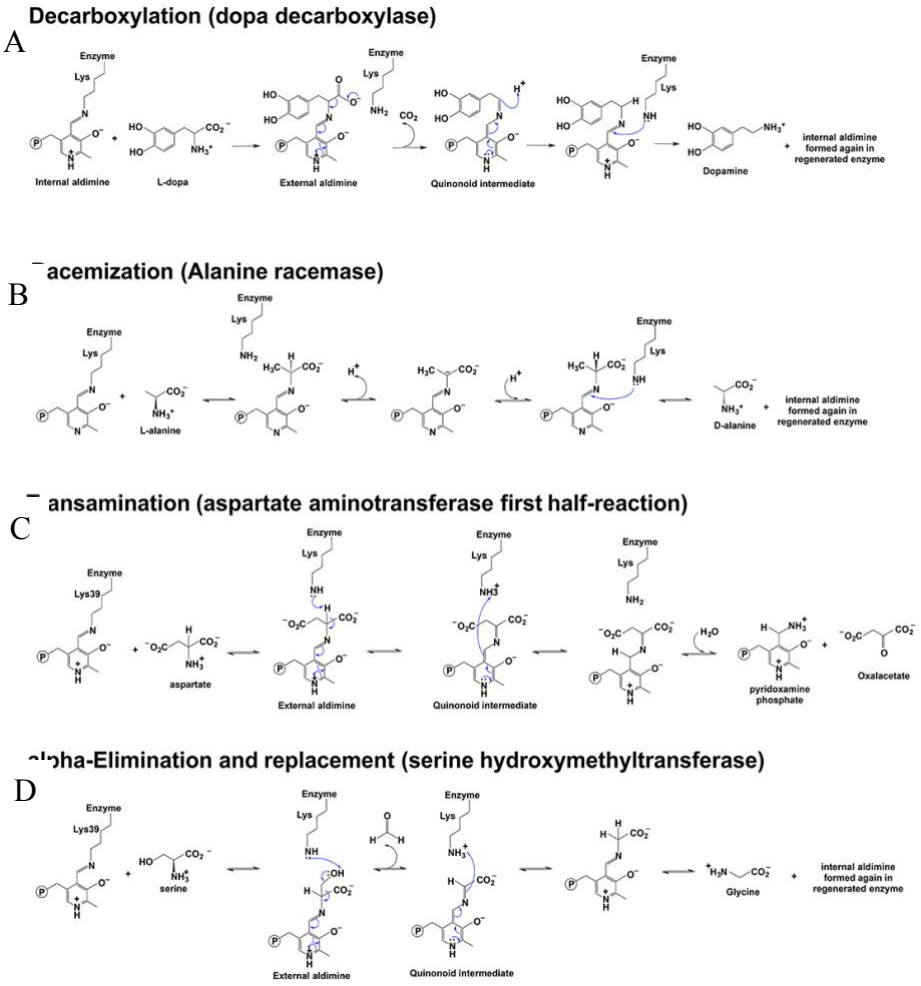


Figure 3.5 Scheme depicting of examples of the reaction mechanisms catalyzed by PLP-dependent enzymes. Adapted from Jianyong. [16]

3.2 Fold Types of PLP-Dependent Enzymes

PLP enzymes were once underrepresented in protein structure databases,[17]but plenty of new structures have tided over it in recent years. At the beginning, it was proposed structures of PLP enzymes correlated with reaction types.[18] Nevertheless, it has been implied every main structural kinds includes representatives of many kinds of reaction types. Grishin et al. divided all PLP-dependent enzymes of which

structures had been solved into 5-fold types.[19] It indicates in Figure1 single representative of Fold Types I-IV, of which mechanisms are described in brief in the thesis. Arranging the PLP enzymes into fold types offers values of mastering lots of conformational limitations of the active sites and the identity of the nearby residues which are participated in catalysis via cofactors or substrate bindings, empowering researchers to distinguish similarities and differences between the various enzyme fold types and their catalyzed representative reactions, and further awaiting unknown PLP enzymes.[19]

Fold type I (most aminotransferases, decarboxylase groups II and III [12] and a few enzymes with α-, β- or γ-elimination activity [20]). These enzymes generally have a Schiff base formed by a nearby Lys residue. The Lys residue is also usually close to C-terminus, forming a loop full of glycine. In the resting state, the loop can bind PLP phosphate group and a hydrophobic β-strand in front of the Lys residue. Besides, the typical character of these enzymes is that the pyridine nitrogen is around 30-amino-acids apart from a conserved aspartate residue, which is proved to have a tendency to interact with the PLP pyridine nitrogen.[21] Though Fold type I enzymes are often composed of two or four subunits and each of them has a PLP molecule, the active site of the enzymes is on the interface between subunits.[22–24] The subunit of Fold type I enzymes includes two kinds of domains that one is large and the other is small: the large one is composed of a seven-stranded β-sheet at the N-terminal, while the small one (the C-terminal of enzymes) includes a 3- or 4-stranded β-sheet secured by helices on one side.[22,24]Aspartate aminotransferase is one of the representatives of Fold type I enzymes.[25–27]

Fold type II enzymes(tryptophan synthase, cysteine synthase, serine dehydratase, threonine dehydratase, O-acetyl serine sulfhydrylase, threonine synthase, etc.[21]) Different from Fold type I enzymes, Fold type II enzymes have the Schiff base formed from Lys residue, which is located closer to the N-terminus, while the loop region for PLP phosphate group attachment is closer to the C-terminus. Compared with conserved aspartate residue in fold type I, a conserved serine residue coordinates Fold type II enzymes. Fold type II enzymes generally work as dimers and tetramers. A N-terminal domain and a C-terminal domain almost shows the same size and function as the region for PLP-binding in fold type II enzymes.[24] Tryptophan synthase is the representative of Fold type II enzymes.

Fold type III enzymes (alanine racemase and eukaryotic ornithine decarboxylase). Same as Fold type II enzymes, Fold type III enzymes carry aldimines created with their Lys residue. The aldimine is located near N-terminus while PLP interacting glycine-rich loop near C-terminus. The distinction is that there is a hydrophobic β-strand in front of the conserved Lys residue while two β-strands in the interface /α/ β units in Fold type III enzymes.[24] Alanine racemase and human ornithine decarboxylase are obligate dimers, because every monomer can supply the two active sites residues. Alanine racemase is made up of two domains in one subunit. One is made up of eight-stranded α/β barrel while the other includes β strands. With regard to alanine racemase, PLP binding site is located in a cleft between the both of domains and one arginine residue interacts with the pyridine ring nitrogen atom by the formation of a hydrogen bond.

The Fold Type IV (D-amino acid aminotransferase family) enzymes, Fold Types I and II are much alike. They usually function as homodimers, and catalytic part is comprised of a small and a large domain. The N-terminal domain is made up of a six-stranded antiparallel β sheet and two helices, while the C-terminal contains a pseudo-β-barrel and some helices.

Fold type V enzymes (PLP-binding domain with a lactate dehydrogenase fold) are mechanistically different in applying the phosphate group of the cofactor for proton transfer in the catalytic reaction. Unlike fold types above, there is not any hydrogen bond reaction in the active site residues of glycogen phosphorylase and the pyridine ring nitrogen atom of PLP.[24] As for glycogen phosphorylase, there are two forms (phosphorylase a and b) existing as tetramer or dimer. Phosphorylase b converts to a by phosphorylase b kinase catalyzing phosphorylation of phosphorylase protomer b while phosphorylase phosphatase mediates the inverse reaction.[28] The research on phosphorylase suggested the two of dimer and tetramer enzymes could combine glycogen, but it showed lower affinity when the tetramer combined the substrate, having not any activity. The binding rate of dimer forming the tetramer is decreased by the glycogen combining, however glycogen combining has no effect on dissociating dimer from tetramer. The dimer and tetramer form of glycogen phosphorylase possibly participate in a regulatory mechanism for the glycogen metabolism.[29] Glycogen phosphorylase has three domains which are N-terminal, glycogen-combining and C-terminal domain. [24]

3.3 Mechanism of Reaction Specificity

An enzyme's specifying reaction types includes the enzyme belongs to a fold type and that fold type contains a variety of reactions. For the majority of these reactions, there are initially two steps: (i) forming of an internal aldimine intermediate which exists in the resting state as a Schiff base by aldehyde group of PLP with lysine residue (Figure 3.2.A). (ii) then, producing of external aldimine (Figure 3.2.B) by the substrate causes a nucleophilic attack at C4` of internal aldimine, replacing lysine residue and forming imine between PLP and substrate amino group. After that, divergence in reaction specificity occurs, leading to different catalyzed mechanistic pathways. These two steps have been explored in the transamination reaction by performing computational approaches. It results in a proposal of three steps sequentially taking place: (i) amino substrate binding to the PLP co-factor attacks at C4 of internal aldimine which is covalent bonded the active site lysine, leading to an unstable tetrahedral intermediate; (ii) a proton switches between the amino substrate and the lysine residue; (iii) lysine residue leaves the tetrahedral intermediate, resulting in the external aldimine(Figure 3.2.B). Under the condition of the PLP, the Cα-proton dissociation from an amino forms the anion,[30] yet the overall reaction in the enzyme is exothermic (−12.0 kcal/mol) [31]. Hence, what makes the formation of anion possible or enhances the reaction efficiency in enzymes whereas it's not accessible or much slower under physiological conditions remains an elusive problem.

The reaction specificity indicates that the nonenzymatic reactions are proposed that they follow the same basic chemical mechanisms, however the enzymes enforce exquisite selectivity of substrate binding and reaction type, which depends on careful control of typical reaction variables including given protein component, structural characteristics of PLP, pH, solvent, temperature, substrate concentration, and even chelation with metal ions. [9,32,33]

In 1966, Dunathan [34] proposed stereoelectronic effect hypothesis to explain the reaction specificity in PLP-dependent enzymes. It stated that, after PLP and substrate forming the external aldimine, a heterolytic bond cleavage happens between Cα and one of its neighboring bonds except Cα-N (Figure 3.6) depending on how much the delocalization energy can be gained.[3,11,35]

It hypothesized that the breakage of the bond with orientation being more perpendicular to the conjugated π-system was to facilitate the formation of Cα. For example, amino acid decarboxylases were predicted to stereospecifically remove the α-carboxylate group whereas elimination/replacement, racemization and transamination abstracted the Cα proton. (Figure 3.4)

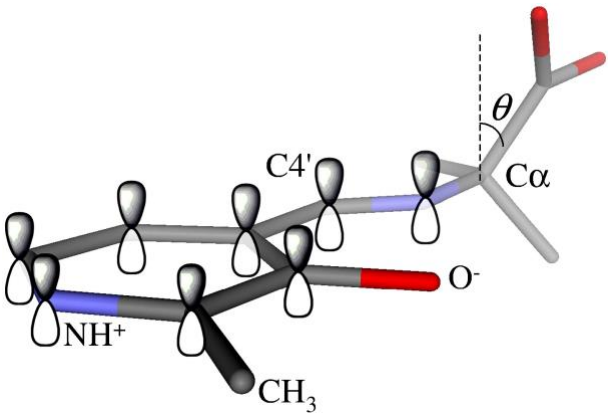


Figure 3.6 Dunathan's stereoelectronic hypothesis.

Even though Dunathan's stereoelectronic hypothesis is able to explain the forming of the external aldimine through limiting the combined substrate to a specific orientation, however, the control of the following steps is less well understood. Breakage of the Cα C bond can lead to multiple reaction pathways: elimination/replacement, racemization, or transamination, therefore, promoting the desired reaction in the subsequent reactions, is complex and requires well-controlled collaboration of the substrate and environment.

The cleavage of the bonds results in an unstable intermediate called the carbanion. The way to stabilize or transfer this carbanion depends on the pyridine ring through a number of resonance forms. the carbanion is stabilized though the quinonoid intermediate if the pyridine nitrogen is protonated (Figure 3.7 A1-A3), while though true carbanion intermediate if pyridine nitrogen is at the absence of protonation (Figure 3.7 B1-B3).The majority of PLP-dependent enzymes, such as AAT, has the quinonoid structure(Figure 3.7.B3) because of its ability to delocalize the negative charge.

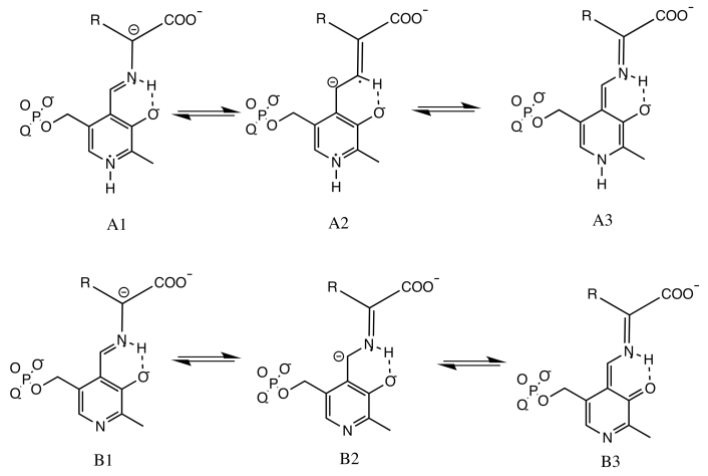


Figure 3.7 Resonance forms of the quinonoid and carbanions intermediate.

Researchers in early found that pyridoxal (Figure 3.8) alone was able to carry out a series of reactions, but at plummet rates.[36–38]Then, a few experiments like UV/vis and IR spectroscopy[37–40] had been conducted, realizing that the protonation states of the pyridoxal were tied to pH-dependent environment at the course of catalysis. Further work in more model compounds confirmed the cationic form of the Schiff base [41], which has three dissociable protons: the ring nitrogen, the Schiff base nitrogen and the phenolic oxygen.

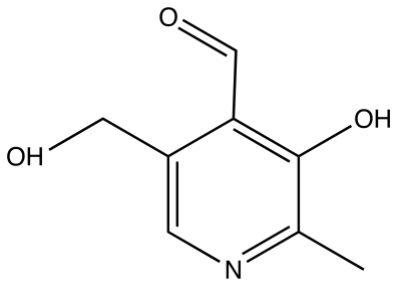


Figure 3.8 Scheme of pyridoxal.

Schiff base nitrogen, one of the ionizable sites, appears to be a main feature in PLP-dependent enzymes. At the presence of the protonation of the pyridine ring nitrogen, it was strikingly more challenging to protonate the phenolic oxygen than Schiff base nitrogen, implying the Schiff base nitrogen

may be the protonated candidate given the pyridine nitrogen is assumed protonated through the course of reaction cycle.[39] Next, the PLP enzyme having protonated Schiff base demonstrates a lot of preference to nucleophilic attack of an incoming amino acid substrate, suggesting the protonated Schiff base played a significant role at the beginning of catalysis.[42] The first PLP-dependent enzyme determined by X-ray crystal structure is aspartate aminotransferase. At 2.8 Å resolution, it tells an aspartate side chain is below the PLP cofactor and interacts with the pyridine nitrogen.[43] Considering aspartate residue being an acid that it is possible to contribute one proton, the pyridine nitrogen is indeed likely protonated in aspartate aminotransferase. Nonetheless, the second crystal structure of a PLP enzyme, tryptophan synthase, indicates that he hydroxyl group of a serine side chain is below the pyridine nitrogen. Hydroxyl group was not a legitimate proton donor to the pyridine nitrogen, calling the question about the protonation of the pyridine nitrogen. As the crystal structure of alanine racemase is published at 1.9 Å resolution,[44] a crystal-clear and positively charged arginine residue is close to the pyridine nitrogen. Certainly, the positive charge excludes the likelihood of protonation of the pyridine nitrogen. All of these X-ray crystal structure observations began to doubt one of the necessities: the protonation of pyridine nitrogen throughout the course of catalysis.

In the other two ionizable sites, the Schiff base nitrogen and the phenolic oxygen, UV/vis are proved a tautomeric equilibrium thanks to the shift of one proton.[45] It shows the protonated Schiff base form absorbs maximally between 410 and 430 nm while the phenolic form absorbs maximally around 330 nm. Later on, X-ray crystal structures of a complex containing a benzoic acid derivative and an internal aldimine analogue without the 5′-phosphate group implied inter- and intramolecular proton transfer coexist.

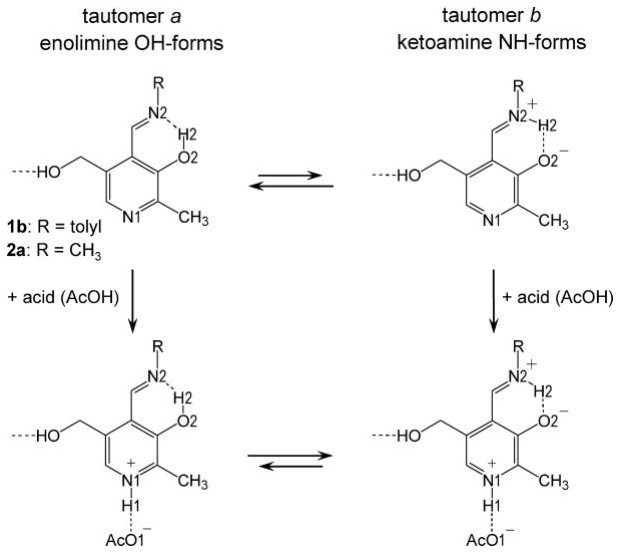


Figure 3.9 Schematic the Keto−Enol Tautomerism of the aldenamine Schiff base (R = Tolyl) and aldimine Schiff base (R = CH3),Adapted from Limbach.46

According to Figure 3.8, an intermolecular proton transfer occurs between the protonated pyridine nitrogen and the acid group of the benzoic acid derivative, and an intramolecular proton transfer occurs between the Schiff base nitrogen along with the phenolic oxygen of the pyridine compound. It suggests inter- and intramolecular proton transfer collaborate to stabilize the complex.[46] Next, the work in the solid-state echoed these findings.[46] Then, modeling work with PLP systems in polar solutions, which allows for the change of acid-base state in the ionizable spots, shows at physiological pH, the phosphate group is dianionic, and the pyridine nitrogen in intramolecular proton transfer lost its proton preceding the intramolecular bond.[47,48] Further more works proved that the polarity of the solvent greatly impacts on determining which sites to be protonated: Nonpolar environments prefers neutral phenolic, while polar environments opts for protonated Schiff base,[49] except that protonated Schiff base form always dominates at whatever environment when pyridine nitrogen protonated.[47] Finally, the study on NMR directly supports the coupling of the inter- and intramolecular proton transfer, meaning protonated Schiff base is the catalytically active tautomer and boosted by protonation of the pyridine. Besides, this study also showed that it is possible to shift protonation to the Schiff base even without protonation of the pyridine nitrogen if

the phenolic oxygen is stabilized by hydrogen bonds with the side chains of active site residues or water molecules in the active site.[50]

It now shows that three main factors account to reaction specificity in PLP enzymes (1) Dunathan hypothesis, a kind of stereoelectronic effect, determines which bond of Cα to be broke.(2) Functional groups of neighboring residues of the active site aim in stabilizing the intermediates and transition states in some sense.(3) The protonation profiles, the protonation states of the cofactor-substrate complex, guarantee the reaction is accompanied with the proper acid-base chemistry. This thesis focuses on the last term.

3.4 Tryptophan Synthase

The PLP-dependent tryptophan synthase (TS) is one of Fold type II enzymes that carry out a-β-elimination and replacement reaction in which the β-hydroxyl group of L-serine is removed and replaced with indole to produce tryptophan. TS discovered by Yanofsky consists of the α-subunit and β-subunit, which are arranged as a linear αββα tetramer, see Figure (3.10).

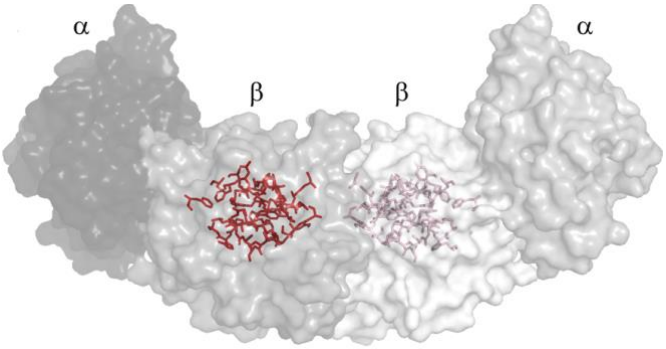


Figure 3.10 Tryptophan synthase full linear αββα tetrameric quaternary structure.

The α-subunit and β-subunit catalyze different reactions. When they are combined, the activity is increased in 30-100 times compared with the individual.[51] The α-subunit converts indoleglycerol phosphate (IGP) to indole and glyceralehyde 3-phosphate (G3P), and the β-subunit transforms the indole and serine to tryptophan along with water.[52] Interestingly, in 1988,the X-ray crystal structure of TS from *S. typhimurium* was revealed at 2.5 Å resolution. It was the first time that substrate channeling(Figure 3.11)

had been definitely shown in an enzyme system and the results caused other researchers to focus on catalytic mechanism.[53].

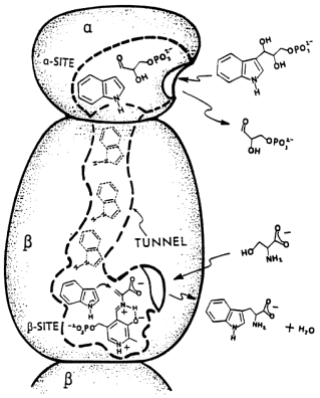


Figure 3.11 Cartoon for the αβ-dimeric unit of tryptophan synthase depicting the binding and dissociation of ligands to the α- and β-sites and the transfer of intermediate indole from the α-site to the β-site.

During the same period, rapid-scanning stopped-flow (RSSF) experiments performed by the Dunn lab generated the UV/vis profiles of the intermediates, lending credence to the first-proposed mechanism for the β-elimination and replacement reaction in 1968.[54–56]

Afterwards, the study on mechanism of tryptophan Synthase was booming. Kinetics experiments led researchers to sort out the order of events in the conformational changes[57–61], which was a switch in closed and open conformations by the allosteric regulations[58,62,63], ensuring the successful analysis. Later, factors like monovalent cation (MVC) [61,62,64] and ligand biding [56,65–67] were soon presented to impact the transition between open and closed conformations as well as the relative stability of formed intermediates. Even though preliminary hypotheses regarding the catalytic cycle assumed a protonated pyridine ring nitrogen as the case in aspartate aminotransferase,[68] solid state NMR measurements and first-principles calculations disproved it and confirmed the unprotonated status of pyridine nitrogen by Mueller's group.[69]

3.4.1 Catalytic Mechanism of Tryptophan Synthase

Tryptophan synthase catalyzes the last two steps in the biosynthesis of the L-tryptophan involving in the elimination of the hydroxyl group of L-serine and its replacement with indole.

In the α-subunit, indole-glycerol-3-phosphate (IGP) is cleft to glyceraldehyde-3-phosphate (G3P) and indole, see Figure (3.12).[70] Then the indole is channeled to an active site of β-subunit via a 25 Å tunnel, conducting the reaction cycle in the β-subunit through a few more intermediate steps.

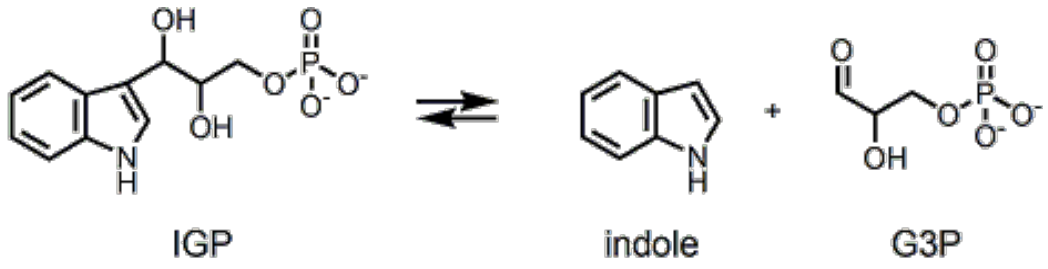


Figure 3.12 The α-site reaction in tryptophan synthase. IGP is cleaved to yield indole and G3P.

In the β- subunit, the detailed mechanism of site catalysis is shown in Figure (3.13).

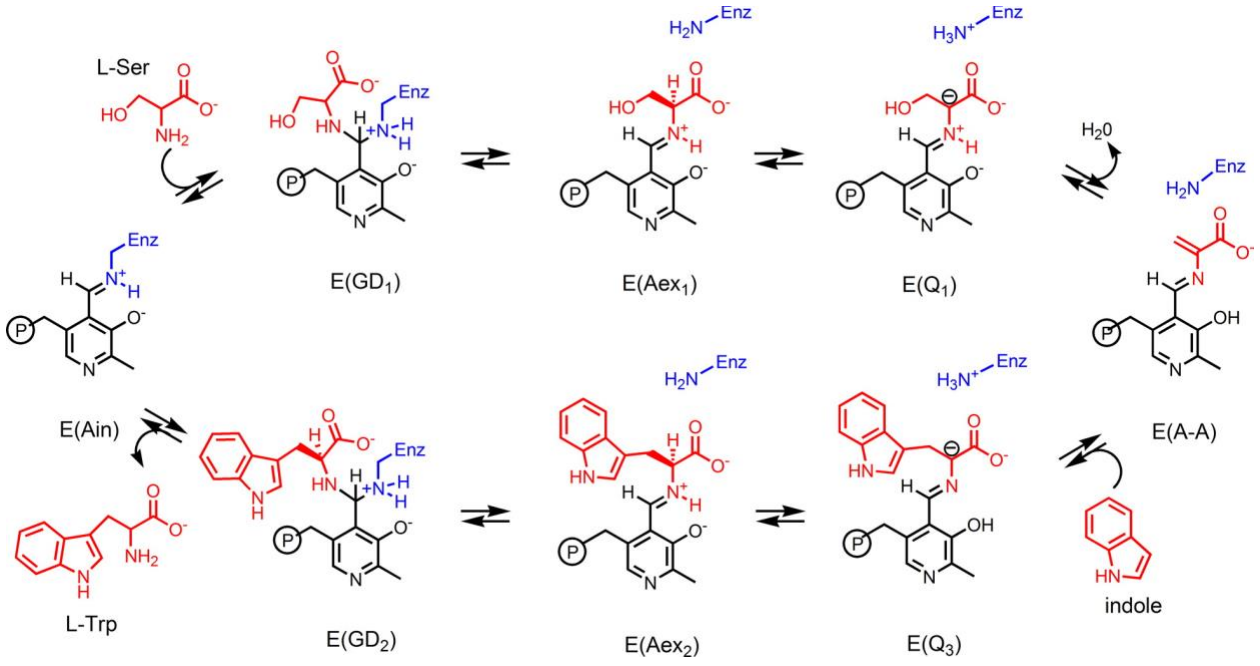


Figure 3.13 The β-site reactions in tryptophan synthase. Figure adapted from Mueller.69

These intermediates corresponding UV/Vis maximum absorbance wavelengths provide where observed. Without substrates, the intermediate internal aldimine is formed by PLP being covalently bound

to the sidechain N^{ε} of β-Lys87. Then, L-serine nucleophilic attacks the C^{α}, through a gem-diamine intermediate E(GD_1), producing the first external aldimine intermediate E(Aex_1).[68]And β-Lys87 pulls the proton away from C^{α} , leading to a first transient quinonoid intermediate, E(Q_1). At almost the same time, β-hydroxyl group of the serine is derived, resulting in the aminoacrylate intermediate, E(A-A). The E(A-A) intermediate goes through nucleophilic attack by the C^3 of the indole which is generated at the α-site and channeled to the β -subunit, creating a $2^{nd}/3^{rd}$ quinonoid E(Q_2) and E(Q_3).[71] After that, C^{α} is reprotoned by βLys87, generating a second external aldimine, E(Aex_2) and a followed second gem-diamine intermediate, E(GD_2). Eventually, a L-tryptophan is released and the E(Ain) is regenerated between PLP and βLys87.[68]

3.4.2 Quasi-Stable Intermediates of Tryptophan Synthase

There are two essential intermediates, internal aldimine E(Ain) and E(Q3) in the β-site analytic cycle of TS identified by the UV/Vis, X-ray, NMR et al.

Internal aldimine is the intermediate generated at the first phase of the catalytic cycle. The cofactor is covalently linked to the enzyme with an imine bond of the ε- nitrogen of a lysine sidechain, forming a Schiff base which is also named as the "internal aldimine,". A hypothesis is stated that the Schiff base should be protonated to speed enzyme-catalyzed reactions in contrast to PLP-catalyzed reactions in solution. [42] Internal aldimine X-ray crystal are firstly published from the structure of the S. typhimurium TS internal aldimine state. Its N-O bond with 2.6 Å distance in the intramolecular proton transfer is completely agreeable to a proton in an N-H-O hydrogen bond.[72] UV/vis optical spectroscopy of TS internal aldimine supports the PSB hypothesis by its 412 nm, which matches the maxima ranging from the 420 to 430 nm, being greatly consistent with the hypothesis of a protonated Schiff base structure conjugated with the PLP π-system.[63]

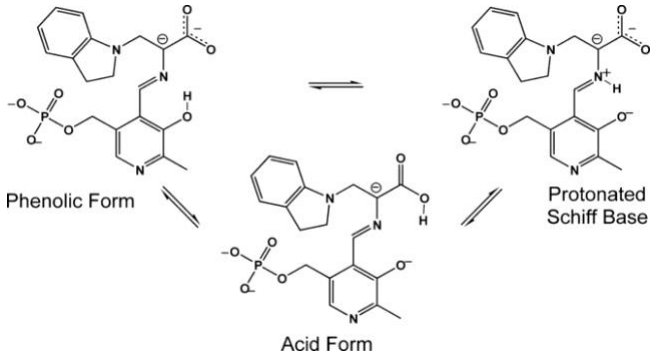


Figure 3.14 Three-site tautomeric equilibrium exchange model is from the original E(Q3) study. (A) Canonical protonated Schiff base quinonoid, (B) acid-form, and (C) phenolic form.

The canonical of E(Q3)(fig) intermediate is initially confirmed by UV/Vis measurements showing a characteristic absorption in the 460–550 nm range and matching absorption characteristics as observed in the natural enzyme.[73,74] E(Q3) is quasi-stable, therefore the lifetime of the intermediate is usually extended to conduct explorations. Addition of N-(4'-trifluoro-methoxybenzenesulfonyl)-2-amino-ethyl phosphate (F9),an α- site substrate analogue that, to E(Q3) can help close the α-subunit and replace monovalent cation with Cs+ to close the β-subunit.[61,68] Besides, the X-ray crystal structures of these quinonoids also clearly supporting the Cα is in expected sp2 hybridization.[68,75] And , first principles calculations have been carried out to predict the most possible structures of the enzyme seen from Figure (3.14).[75] A much longer lived E(Q3)2AP intermediate replaced 2-aminophenol (2AP) with indoline, determining that the assumption of a protonated pyridine ring nitrogen was not correct. A deprotonated pyridine nitrogen on PLP verifies carbanionic intermediate is formed instead of a true quinonoid species. Natural bond orbital(NBO) analysis shows that negative charge is produced at the Cα, while positive charge is formed at C4′, being in accordance with the function as the catalytic tautomer and corresponding to the hypothesis the specificity for β -elimination/replacement versus transamination is controlled in part by the protonation.[76]

3.5 NMR crystallography

NMR crystallography— combining X-ray diffraction, solid-state NMR spectroscopy, and computational chemistry—provides exceptional and holistic insights to three-dimensional and big enzyme systems. Mueller has employed it in the tryptophan synthase enzyme catalysis.[76] Symbolic X-ray crystal structures (1.3 to 2.5 Å resolution) of intermediates that binds the enzyme make a contribution that hydrogen-bond interacts possibly between site residues and the substrate, but fail legitimately to recognize the protonation states. NMR supplies direct chemical shifts to certain chosen atoms in enzyme-substrate complexes, however it requires a framework to explain them. Ab initio calculations and molecular mechanics are capable to build models of enzymatic processes which must be on the ground of the specified framework. Molecular dynamic simulations can provide a good initial guess for molecular' orientations, but they cannot determine the protonation states of cofactor. However, they are combined to supply structure and function of enzyme active sites models which are accordant and testable .[77–79] The previous studies of NMR crystallography has determined the protonation profiles of PLP in carbanionic intermediate states for tryptophan synthase and showed the remarkable results.[75,76]

3.6 References

(1) György, P. Vitamin B 2 and the Pellagra-like Dermatitis in Rats. *Nature* **1934**, *133* (3361), 498–499. https://doi.org/10.1038/133498a0.

(2) Jansonius, J. N. Structure, Evolution and Action of Vitamin B6-Dependent Enzymes. *Curr. Opin. Struct. Biol.* **1998**, *8* (6), 759–769. https://doi.org/10.1016/S0959-440X(98)80096-1.

(3) Toney, M. D. Reaction Specificity in Pyridoxal Phosphate Enzymes. *Arch. Biochem. Biophys.* **2005**, *433* (1), 279–287. https://doi.org/10.1016/j.abb.2004.09.037.

(4) Beattie, A. E.; Clarke, D. J.; Wadsworth, J. M.; Lowther, J.; Sin, H.-L.; Campopiano, D. J. Reconstitution of the Pyridoxal 5'-Phosphate (PLP) Dependent Enzyme Serine Palmitoyltransferase (SPT) with Pyridoxal Reveals a Crucial Role for the Phosphate during Catalysis. *Chem. Commun. Camb. Engl.* **2013**, *49* (63), 7058–7060. https://doi.org/10.1039/c3cc43001d.

(5) Phillips, R. S.; Scott, I.; Paulose, R.; Patel, A.; Barron, T. C. The Phosphate of Pyridoxal-5′-Phosphate Is an Acid/Base Catalyst in the Mechanism of Pseudomonas Fluorescens Kynureninase. *FEBS J.* **2014**, *281* (4), 1100–1109. https://doi.org/10.1111/febs.12671.

(6) Eliot, A. C.; Kirsch, J. F. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations. *Annu. Rev. Biochem.* **2004**, *73* (1), 383–415. https://doi.org/10.1146/annurev.biochem.73.011303.074021.

(7) Dunn, M. F.; Niks, D.; Ngo, H.; Barends, T. R. M.; Schlichting, I. Tryptophan Synthase: The Workings of a Channeling Nanomachine. *Trends Biochem. Sci.* **2008**, *33* (6), 254–264. https://doi.org/10.1016/j.tibs.2008.04.008.

(8) Olivard, J.; Metzler, D. E.; Snell, E. E. Catalytic Racemization of Amino Acids by Pyridoxal and Metal Salts. *J. Biol. Chem.* **1952**, *199* (2), 669–674.

(9) Metzler, D. E.; Snell, E. E. Deamination of Serine. I. Catalytic Deamination of Serine and Cysteine by Pyridoxal and Metal Salts. *J. Biol. Chem.* **1952**, *198* (1), 353–361.

(10) Bender, D. Transaminases. *FEBS Lett.* **1985**, *190* (1), 179–180. https://doi.org/10.1016/0014-5793(85)80462-2.

(11) Metzler, D. E.; Ikawa, M.; Snell, E. E. A General Mechanism for Vitamin B6-Catalyzed Reactions1. *J. Am. Chem. Soc.* **1954**, *76* (3), 648–652. https://doi.org/10.1021/ja01632a004.

(12) Steffen-Munsberg, F.; Vickers, C.; Kohls, H.; Land, H.; Mallin, H.; Nobili, A.; Skalden, L.; van den Bergh, T.; Joosten, H.-J.; Berglund, P.; Höhne, M.; Bornscheuer, U. T. Bioinformatic Analysis of a PLP-Dependent Enzyme Superfamily Suitable for Biocatalytic Applications. *Biotechnol. Adv.* **2015**, *33* (5), 566–604. https://doi.org/10.1016/j.biotechadv.2014.12.012.

(13) Toney, M. D. Controlling Reaction Specificity in Pyridoxal Phosphate Enzymes. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2011**, *1814* (11), 1407–1418. https://doi.org/10.1016/j.bbapap.2011.05.019.

(14) Fogle, E. J.; Toney, M. D. Mutational Analysis of Substrate Interactions with the Active Site of Dialkylglycine Decarboxylase. *Biochemistry* **2010**, *49* (30), 6485–6493. https://doi.org/10.1021/bi100648w.

(15) Taylor, J. L.; Price, J. E.; Toney, M. D. Directed Evolution of the Substrate Specificity of Dialkylglycine Decarboxylase. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2015**, *1854* (2), 146–155. https://doi.org/10.1016/j.bbapap.2014.12.003.

(16) Liang, J.; Han, Q.; Tan, Y.; Ding, H.; Li, J. Current Advances on Structure-Function Relationships of Pyridoxal 5′-Phosphate-Dependent Enzymes. *Front. Mol. Biosci.* **2019**, *6*. https://doi.org/10.3389/fmolb.2019.00004.

(17) John, R. A. Pyridoxal Phosphate-Dependent Enzymes. *Biochim. Biophys. Acta* **1995**, *1248* (2), 81–96. https://doi.org/10.1016/0167-4838(95)00025-p.

(18) Alexander, F. W.; Sandmeier, E.; Mehta, P. K.; Christen, P. Evolutionary Relationships among Pyridoxal-5'-Phosphate-Dependent Enzymes. Regio-Specific Alpha, Beta and Gamma Families. *Eur. J. Biochem.* **1994**, *219* (3), 953–960. https://doi.org/10.1111/j.1432-1033.1994.tb18577.x.

(19) Percudani, R.; Peracchi, A. A Genomic Overview of Pyridoxal-Phosphate-Dependent Enzymes. *EMBO Rep.* **2003**, *4* (9), 850–854. https://doi.org/10.1038/sj.embor.embor914.

(20) Kappes, B.; Tews, I.; Binter, A.; Macheroux, P. PLP-Dependent Enzymes as Potential Drug Targets for Protozoan Diseases. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2011**, *1814* (11), 1567–1576. https://doi.org/10.1016/j.bbapap.2011.07.018.

(21) Grishin, N. V.; Phillips, M. A.; Goldsmith, E. J. Modeling of the Spatial Structure of Eukaryotic Ornithine Decarboxylases. *Protein Sci.* **1995**, *4* (7), 1291–1304. https://doi.org/10.1002/pro.5560040705.

(22) Han, Q.; Ding, H.; Robinson, H.; Christensen, B. M.; Li, J. Crystal Structure and Substrate Specificity of Drosophila 3,4-Dihydroxyphenylalanine Decarboxylase. *PLoS ONE* **2010**, *5* (1). https://doi.org/10.1371/journal.pone.0008826.

(23) Milano, T.; Paiardini, A.; Grgurina, I.; Pascarella, S. Type I Pyridoxal 5′-Phosphate Dependent Enzymatic Domains Embedded within Multimodular Nonribosomal Peptide Synthetase and Polyketide Synthase Assembly Lines. *BMC Struct. Biol.* **2013**, *13* (1), 26. https://doi.org/10.1186/1472-6807-13-26.

(24) Schneider, G.; Käck, H.; Lindqvist, Y. The Manifold of Vitamin B6 Dependent Enzymes. *Structure* **2000**, *8* (1), R1–R6. https://doi.org/10.1016/S0969-2126(00)00085-X.

(25) McPhalen, C. A.; Vincent, M. G.; Jansonius, J. N. X-Ray Structure Refinement and Comparison of Three Forms of Mitochondrial Aspartate Aminotransferase. *J. Mol. Biol.* **1992**, *225* (2), 495–517. https://doi.org/10.1016/0022-2836(92)90935-D.

(26) Danishefsky, A. T.; Onnufer, J. J.; Petsko, G. A.; Ringe, D. Activity and Structure of the Active-Site Mutants R386Y and R386F of Escherichia Coli Aspartate Aminotransferase. *Biochemistry* **1991**, *30* (7), 1980–1985. https://doi.org/10.1021/bi00221a035.

(27) Snell, E. E.; Jenkins, W. T. The Mechanism of the Transamination Reaction. *J. Cell. Comp. Physiol.* **1959**, *54* (S1), 161–177. https://doi.org/10.1002/jcp.1030540413.

(28) Cohen, P.; Duewer, T.; Fischer, E. H. Phosphorylase from Dogfish Skeletal Muscle. Purification and a Comparison of Its Physical Properties to Those of Rabbit Muscle Phosphorylase. *Biochemistry* **1971**, *10* (14), 2683–2694. https://doi.org/10.1021/bi00790a005.

(29) Wang, Z.-X. Kinetic Study on the Dimer–Tetramer Interconvertion of Glycogen Phosphorylase a. *Eur. J. Biochem.* **1999**, *259* (3), 609–617. https://doi.org/10.1046/j.1432-1327.1999.00058.x.

(30) Ho, J.; Easton, C. J.; Coote, M. L. The Distal Effect of Electron-Withdrawing Groups and Hydrogen Bonding on the Stability of Peptide Enolates. *J. Am. Chem. Soc.* **2010**, *132* (15), 5515–5521. https://doi.org/10.1021/ja100996z.

(31) Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. Computational Mechanistic Studies Addressed to the Transimination Reaction Present in All Pyridoxal 5′-Phosphate-Requiring Enzymes. *J. Chem. Theory Comput.* **2011**, *7* (5), 1356–1368. https://doi.org/10.1021/ct1002219.

(32) Metzler, D. E.; Longenecker, J. B.; Snell, E. E. REVERSIBLE CATALYTIC CLEAVAGE OF HYDROXYAMINO ACIDS BY PYRIDOXAL AND METAL SALTS. *J. Am. Chem. Soc.* **1953**, *75* (11), 2786–2787. https://doi.org/10.1021/ja01107a527.

(33) Metzler, D. E.; Longenecker, J. B.; Snell, E. E. The Reversible Catalytic Cleavage of Hydroxyamino Acids by Pyridoxal and Metal Salts1. *J. Am. Chem. Soc.* **1954**, *76* (3), 639–644. https://doi.org/10.1021/ja01632a002.

(34) Dunathan, H. C. Conformation and Reaction Specificity in Pyridoxal Phosphate Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *55* (4), 712–716.

(35) H, H. Pyridoxal Enzymes: Mechanistic Diversity and Uniformity. *J. Biochem. (Tokyo)* **1995**, *118* (3), 463–473. https://doi.org/10.1093/oxfordjournals.jbchem.a124931.

(36) Snell, E. E.; Di Mari, S. J. 7 Schiff Base Intermediates in Enzyme Catalysis. In *The Enzymes*; Boyer, P. D., Ed.; Academic Press, 1970; Vol. 2, pp 335–370. https://doi.org/10.1016/S1874-6047(08)60186-9.

(37) Heinert, D.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. II. [1,2] Infrared Spectra and Hydrogen Bonding. *J. Am. Chem. Soc.* **1959**, *81* (15), 3933–3943. https://doi.org/10.1021/ja01524a031.

(38) Nakamoto, K.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. III. Ultraviolet Absorption Studies and Solution Equilibria of 2- and 4-Hydroxymethyl-3-Hydroxypyridines and Pyridine-2, 3- and 4-Aldehydes [1]. *J. Am. Chem. Soc.* **1959**, *81* (22), 5857–5863. https://doi.org/10.1021/ja01531a005.

(39) Christensen, H. N. Three Schiff Base Types Formed by Amino Acids, Peptides and Proteins with Pyridoxal and Pyridoxal-5-Phosphate1. *J. Am. Chem. Soc.* **1958**, *80* (1), 99–105. https://doi.org/10.1021/ja01534a027.

(40) Heinert, Dietrich.; Martell, A. E. **Pyridoxine and Pyridoxal Analogs. VIII. Synthesis and Infrared Spectra of Metal Chelates**. *J. Am. Chem. Soc.* **1963**, *85* (9), 1334–1337. https://doi.org/10.1021/ja00892a026.

(41) Heinert, Dietrich.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. VII. Acid-Base Equilibria of Schiff Bases. *J. Am. Chem. Soc.* **1963**, *85* (2), 188–193. https://doi.org/10.1021/ja00885a018.

(42) Cordes, E. H.; Jencks, W. P. Semicarbazone Formation from Pyridoxal, Pyridoxal Phosphate, and Their Schiff Bases. *Biochemistry* **1962**, *1* (5), 773–778. https://doi.org/10.1021/bi00911a007.

(43) Ford, G. C.; Eichele, G.; Jansonius, J. N. Three-Dimensional Structure of a Pyridoxal-Phosphate-Dependent Enzyme, Mitochondrial Aspartate Aminotransferase. *Proc. Natl. Acad. Sci.* **1980**, *77* (5), 2559–2563. https://doi.org/10.1073/pnas.77.5.2559.

(44) Shaw, J. P.; Petsko, G. A.; Ringe, D. Determination of the Structure of Alanine Racemase from Bacillus Stearothermophilus at 1.9-A Resolution. *Biochemistry* **1997**, *36* (6), 1329–1342. https://doi.org/10.1021/bi961856c.

(45) Metzler, D. E. Equilibria between Pyridoxal and Amino Acids and Their Imines1. *J. Am. Chem. Soc.* **1957**, *79* (2), 485–490. https://doi.org/10.1021/ja01559a068.

(46) Sharif, S.; Schagen, D.; Toney, M. D.; Limbach, H.-H. Coupling of Functional Hydrogen Bonds in Pyridoxal-5‘-Phosphate−Enzyme Model Systems Observed by Solid-State NMR Spectroscopy. *J. Am. Chem. Soc.* **2007**, *129* (14), 4440–4455. https://doi.org/10.1021/ja066240h.

(47) Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Toney, M. D.; Limbach, H.-H. NMR Studies of the Stability, Protonation States, and Tautomerism of 13C- and 15N-Labeled Aldimines of the Coenzyme Pyridoxal 5′-Phosphate in Water. *Biochemistry* **2010**, *49* (51), 10818–10830. https://doi.org/10.1021/bi101061m.

(48) Sharif, S.; Huot, M. C.; Tolstoy, P. M.; Toney, M. D.; Jonsson, K. H. M.; Limbach, H.-H. 15N Nuclear Magnetic Resonance Studies of Acid-Base Properties of Pyridoxal-5’-Phosphate Aldimines in Aqueous Solution. *J. Phys. Chem. B* **2007**, *111* (15), 3869–3876. https://doi.org/10.1021/jp067334g.

(49) Limbach, H.-H.; Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Shenderovich, I. G.; Denisov, G. S.; Toney, M. D. Critical Hydrogen Bonds and Protonation States of Pyridoxal 5’-Phosphate Revealed by NMR. *Biochim. Biophys. Acta* **2011**, *1814* (11), 1426–1437. https://doi.org/10.1016/j.bbapap.2011.06.004.

(50) 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. NMR Studies of Protonation and Hydrogen Bond States of Internal Aldimines of Pyridoxal 5’-Phosphate Acid-Base in Alanine Racemase, Aspartate Aminotransferase, and Poly-L-Lysine. *J. Am. Chem. Soc.* **2013**, *135* (48), 18160–18175. https://doi.org/10.1021/ja408988z.

(51) Crawford, I. P.; Yanofsky, C. ON THE SEPARATION OF THE TRYPTOPHAN SYNTHETASE OF ESCHERICHIA COLI INTO TWO PROTEIN COMPONENTS*. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44* (12), 1161–1170.

(52) Henning, U.; Helinski, D. R.; Chao, F. C.; Yanofsky, C. The A Protein of the Tryptophan Synthetase of Escherichia Coli. Purification, Crystallization, and Composition Studies. *J. Biol. Chem.* **1962**, *237*, 1523–1530.

(53) Hyde, C. C.; Ahmed, S. A.; Padlan, E. A.; Miles, E. W.; Davies, D. R. Three-Dimensional Structure of the Tryptophan Synthase Alpha 2 Beta 2 Multienzyme Complex from Salmonella Typhimurium. *J. Biol. Chem.* **1988**, *263* (33), 17857–17871.

(54) Drewe, W. F.; Dunn, M. F. Detection and Identification of Intermediates in the Reaction of L-Serine with Escherichia Coli Tryptophan Synthase via Rapid-Scanning Ultraviolet-Visible Spectroscopy. *Biochemistry* **1985**, *24* (15), 3977–3987. https://doi.org/10.1021/bi00336a027.

(55) Drewe, W. F.; Dunn, M. F. Characterization of the Reaction of L-Serine and Indole with Escherichia Coli Tryptophan Synthase via Rapid-Scanning Ultraviolet-Visible Spectroscopy. *Biochemistry* **1986**, *25* (9), 2494–2501. https://doi.org/10.1021/bi00357a032.

(56) Roy, M.; Miles, E. W.; Phillips, R. S.; Dunn, M. F. Detection and Identification of Transient Intermediates in the Reactions of Tryptophan Synthase with Oxindolyl-L-Alanine and 2,3-Dihydro-

L-Tryptophan. Evidence for a Tetrahedral (Gem-Diamine) Intermediate. *Biochemistry* **1988**, *27* (23), 8661–8669. https://doi.org/10.1021/bi00423a023.

(57) Houben, K. F.; Dunn, M. F. Allosteric Effects Acting over a Distance of 20-25 .ANG. in the Escherichia Coli Tryptophan Synthase Bienzyme Complex Increase Ligand Affinity and Cause Redistribution of Covalent Intermediates. *Biochemistry* **1990**, *29* (9), 2421–2429. https://doi.org/10.1021/bi00461a028.

(58) Brzović, P. S.; Sawa, Y.; Hyde, C. C.; Miles, E. W.; Dunn, M. F. Evidence That Mutations in a Loop Region of the Alpha-Subunit Inhibit the Transition from an Open to a Closed Conformation in the Tryptophan Synthase Bienzyme Complex. *J. Biol. Chem.* **1992**, *267* (18), 13028–13038.

(59) Kirschner, K.; Lane, A. N.; Strasser, A. W. Reciprocal Communication between the Lyase and Synthase Active Sites of the Tryptophan Synthase Bienzyme Complex. *Biochemistry* **1991**, *30* (2), 472–478. https://doi.org/10.1021/bi00216a024.

(60) Miles, E. W. The Tryptophan Synthase A2β2 Complex: A Model for Substrate Channeling, Allosteric Communication, and Pyridoxal Phosphate Catalysis. *J. Biol. Chem.* **2013**, *288* (14), 10084–10091. https://doi.org/10.1074/jbc.X113.463331.

(61) Woehl, E.; Dunn, M. F. Mechanisms of Monovalent Cation Action in Enzyme Catalysis: The Tryptophan Synthase α-, β-, and Aβ-Reactions. *Biochemistry* **1999**, *38* (22), 7131–7141. https://doi.org/10.1021/bi982919p.

(62) Leja, C. A.; Woehl, E. U.; Dunn, M. F. Allosteric Linkages between .Beta.-Site Covalent Transformations and .Alpha.-Site Activation and Deactivation in the Tryptophan Synthase Bienzyme Complex. *Biochemistry* **1995**, *34* (19), 6552–6561. https://doi.org/10.1021/bi00019a037.

(63) Peracchi, A.; Bettati, S.; Mozzarelli, A.; Rossi, G. L.; Miles, E. W.; Dunn, M. F. Allosteric Regulation of Tryptophan Synthase: Effects of PH, Temperature, and Alpha-Subunit Ligands on the Equilibrium Distribution of Pyridoxal 5'-Phosphate-L-Serine Intermediates. *Biochemistry* **1996**, *35* (6), 1872–1880. https://doi.org/10.1021/bi951889c.

(64) Weber-Ban, E.; Hur, O.; Bagwell, C.; Banik, U.; Yang, L. H.; Miles, E. W.; Dunn, M. F. Investigation of Allosteric Linkages in the Regulation of Tryptophan Synthase: The Roles of Salt Bridges and Monovalent Cations Probed by Site-Directed Mutation, Optical Spectroscopy, and Kinetics. *Biochemistry* **2001**, *40* (12), 3497–3511. https://doi.org/10.1021/bi002690p.

(65) Houben, K. F.; Kadima, W.; Roy, M.; Dunn, M. F. L-Serine Analogues Form Schiff Base and Quinonoidal Intermediates with Escherichia Coli Tryptophan Synthase. *Biochemistry* **1989**, *28* (10), 4140–4147. https://doi.org/10.1021/bi00436a003.

(66) Hur, O.; Niks, D.; Casino, P.; Dunn, M. F. Proton Transfers in the Beta-Reaction Catalyzed by Tryptophan Synthase. *Biochemistry* **2002**, *41* (31), 9991–10001. https://doi.org/10.1021/bi025568u.

(67) Niks, D.; Hilario, E.; Dierkers, A.; Ngo, H.; Borchardt, D.; Neubauer, T. J.; Fan, L.; Mueller, L. J.; Dunn, M. F. Allostery and Substrate Channeling in the Tryptophan Synthase Bienzyme Complex: Evidence for Two Subunit Conformations and Four Quaternary States. *Biochemistry* **2013**, *52* (37), 6396–6411. https://doi.org/10.1021/bi400795e.

(68) Dunn, M. F. Allosteric Regulation of Substrate Channeling and Catalysis in the Tryptophan Synthase Bienzyme Complex. *Arch. Biochem. Biophys.* **2012**, *519* (2), 154–166. https://doi.org/10.1016/j.abb.2012.01.016.

(69) Caulkins, B. G.; Yang, C.; Hilario, E.; Fan, L.; Dunn, M. F.; Mueller, L. J. Catalytic Roles of BLys87 in Tryptophan Synthase: 15N Solid State NMR Studies. *Biochim. Biophys. Acta* **2015**, *1854* (9), 1194–1199. https://doi.org/10.1016/j.bbapap.2015.02.003.

(70) Kulik, V.; Weyand, M.; Seidel, R.; Niks, D.; Arac, D.; Dunn, M. F.; Schlichting, I. On the Role of AlphaThr183 in the Allosteric Regulation and Catalytic Mechanism of Tryptophan Synthase. *J. Mol. Biol.* **2002**, *324* (4), 677–690. https://doi.org/10.1016/s0022-2836(02)01109-9.

(71) Barends, T. R. M.; Domratcheva, T.; Kulik, V.; Blumenstein, L.; Niks, D.; Dunn, M. F.; Schlichting, I. Structure and Mechanistic Implications of a Tryptophan Synthase Quinonoid Intermediate. *ChemBioChem* **2008**, *9* (7), 1024–1028. https://doi.org/10.1002/cbic.200700703.

(72) Caulkins, B. G.; Bastin, B.; Yang, C.; Neubauer, T. J.; Young, R. P.; Hilario, E.; Huang, Y. M.; Chang, C. A.; Fan, L.; Dunn, M. F.; Marsella, M. J.; Mueller, L. J. Protonation States of the Tryptophan Synthase Internal Aldimine Active Site from Solid-State NMR Spectroscopy: Direct Observation of the Protonated Schiff Base Linkage to Pyridoxal-5′-Phosphate. *J. Am. Chem. Soc.* **2014**, *136* (37), 12824–12827. https://doi.org/10.1021/ja506267d.

(73) Metzler, C. M.; Harris, A. G.; Metzler, D. E. Spectroscopic Studies of Quinonoid Species from Pyridoxal 5'-Phosphate. *Biochemistry* **1988**, *27* (13), 4923–4933. https://doi.org/10.1021/bi00413a050.

(74) Metzler, D. E. Equilibria between Pyridoxal and Amino Acids and Their Imines1. *J. Am. Chem. Soc.* **1957**, *79* (2), 485–490. https://doi.org/10.1021/ja01559a068.

(75) Lai, J.; Niks, D.; Wang, Y.; Domratcheva, T.; Barends, T. R. M.; Schwarz, F.; Olsen, R. A.; Elliott, D. W.; Fatmi, M. Q.; Chang, C. A.; Schlichting, I.; Dunn, M. F.; Mueller, L. J. X-Ray and NMR Crystallography in an Enzyme Active Site: The Indoline Quinonoid Intermediate in Tryptophan Synthase. *J. Am. Chem. Soc.* **2011**, *133* (1), 4–7. https://doi.org/10.1021/ja106555c.

(76) Caulkins, B. G.; Young, R. P.; Kudla, R. A.; Yang, C.; Bittbauer, T. J.; Bastin, B.; Hilario, E.; Fan, L.; Marsella, M. J.; Dunn, M. F.; Mueller, L. J. NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. *J. Am. Chem. Soc.* **2016**, *138* (46), 15214–15226. https://doi.org/10.1021/jacs.6b08937.

(77) Mueller, L. J.; Dunn, M. F. NMR Crystallography of Enzyme Active Sites: Probing Chemically Detailed, Three-Dimensional Structure in Tryptophan Synthase. *Acc. Chem. Res.* **2013**, *46* (9), 2008–2017. https://doi.org/10.1021/ar3003333.

(78) Harris, R. K.; Hodgkinson, P.; Pickard, C. J.; Yates, J. R.; Zorin, V. Chemical Shift Computations on a Crystallographic Basis: Some Reflections and Comments. *Magn. Reson. Chem.* **2007**, *45* (S1), S174–S186. https://doi.org/10.1002/mrc.2132.

(79) Elena, B.; Pintacuda, G.; Mifsud, N.; Emsley, L. Molecular Structure Determination in Powders by NMR Crystallography from Proton Spin Diffusion. *J. Am. Chem. Soc.* **2006**, *128* (29), 9555–9560. https://doi.org/10.1021/ja062353p.

Chapter 4. NMR Crystallography of the Internal Aldimine in Tryptophan Synthase

4.1 Introduction

Pyridoxal-5′-phosphate (PLP), acting as a coenzyme, is ubiquitous in the nitrogen metabolism of all organisms. PLP catalyzes a wide variety of amino acid transformations, including α/β/γ elimination/replacement, racemization, transamination, and decarboxylation.[1,2]

At the start of the PLP-catalyzed reaction cycle, a general internal aldimine intermediate is formed by the aldehyde group with lysine residue (Figure 4.1). In the next step, the substrate initiates a nucleophilic attack at C4`, replacing the lysine residue and forming an external aldimine from the PLP and the substrate amino group.

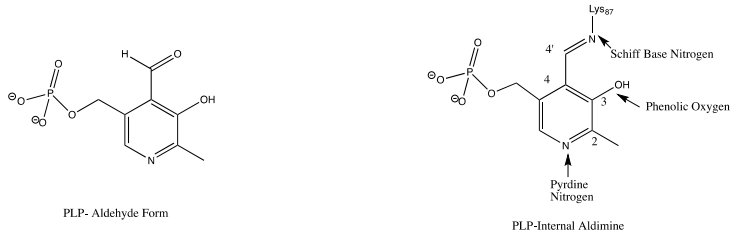


Figure 4.1 Pyridoxal-5'-phosphate.

After the external aldimine is formed, deprotonation usually occurs with the aid of the ε-amino group, yielding a carbanionic intermediate.[3,4] The carbanionic intermediate can undergo several different subsequent reactions, leading to different reaction paths among different fold types of PLP-dependent enzymes.[5,6]

Three major determinant factors can be used to explain the direction of different types of reactions in PLP-dependent enzymes: (1) stereoelectronic control, (2) the role of active site side chains, and (3) the role of the various protonation states of the external aldimine intermediate in PLP-dependent enzymes.[7–9]

Stereoelectronic control has relationship with bond cleave, which usually break the bond that is most perpendicular to the conjugated π-system of the PLP molecule.[5] Electronic modulation through selective protonation asserts that negative charge stabilization in the transition state is a result of the electrophilic strength of the combined Schiff base and pyridine ring π-system, resulting in the local active site environment of the enzyme promoting specific protonation states of the cofactor, suggesting that the proton positions govern reaction diversities. For example, amino acid decarboxylases generally stereo-specifically remove the α-carboxylate group, and elimination/replacement, racemization, and transamination usually abstract the Cα proton according to which bond is more perpendicular to the conjugated π-system in the PLP molecule. After a distinct carbanionic intermediate is conformed, its interaction with the active site residues accounts for the posterior analytical pathways. In aspartate aminotransferase (AAT), which is the fold type I PLP-dependent enzyme and undergoes the transamination reaction, the negative charge at C^{α} in the carbanionic intermediate delocalizes to the whole π-system due to the protonation of pyridine nitrogen. This leads to the protonation at C4′ again and the creation of one ketimine intermediate.[10–12] By contrast, in tryptophan synthase (TS), which belongs to the fold type II PLP-dependent enzyme and promotes β-elimination followed by replacement reactions, the delocalization is not required and the charge distribution favors the protonation at Cα without the protonation of pyridine N. This causes the cleavage of the leaving group hydroxyl from C^{β} in the β-elimination.[2,3,7,13] Thus, active site protonation states, or so-called protonation profiles, of PLP-dependent enzymes are very likely to contribute toward optimizing the desired chemical reactions. A specific protonation profile may come with a specific fold type of PLP-dependent enzyme and pivot to a specific reaction.

The tryptophan synthase (TS) is the fold type II PLP-dependent enzyme in which the hydroxyl group of L-serine is eliminated and replaced with indole to produce tryptophan. At the start of the catalytic cycle of the TS reaction, the protonated Schiff base (PSB) has been suggested to activate catalysis by having a more reactive target for the nucleophilic attack than that of the neutral imine.[14–16] In addition, the protonation profiles, especially of the phenolic oxygen and pyridine nitrogen, are thought to be critical in establishing the specificity of the reaction.

UV/vis optical spectroscopies of TS support the PSB hypothesis. Explanations regarding a PSB nitrogen can initially be derived from peptides and amino acids in solvents with different polar conditions.[13,17] The internal aldimine intermediate in most PLP enzymes absorbs maxima wavelengths below the 420–430 nm range, approximately 410 nm from TS internal aldimine in solution, and single crystals on the TS internal aldimine have provided evidence for the PSB hypothesis.[18] Conversely, since the UV/vis spectrum of TS is not dependent on pH over the range of 5.8–10.4,[18–20] deprotonation of PLP enzyme Schiff bases is feasible through changing pH.[21] Consequently, a shift from a longer wavelength (430 nm) to a shorter wavelength (~360 nm) occurs. [22] It has been proposed that, in this phenomenon, the protonated Schiff base may undergo a transformation because another protonatable site, the internal aldimine pyridine nitrogen, has only negligible effects on the UV/vis spectrum.[23]

The X-ray crystal structures of a TS internal aldimine intermediate derived from *Salmonella typhimurium* in 1.30 Å resolution (PDB code 4HT3) showed the Schiff base nitrogen is in close proximity to the 3' oxygen of the PLP ring.[24–27] (Figure 4.2) The 2.6 Å N–O distance, between the Schiff base nitrogen and the pyridine oxygen, indicated the presence of a proton in an N-H-O hydrogen bond. However, the X-ray crystal structure can not be used to determine the donor and acceptor in this bonding system. In addition, X-ray crystal structures showed the Schiff base nitrogen is stretched slightly out of the PLP plane with a C^3-C^4-$C^{4'}$-N twist angle of 27.2°, and the phenolic oxygen is located near two water molecules, implying the phenolic oxygen may participate in a hydrogen-bonding network, as indicated in Figure 4.1.

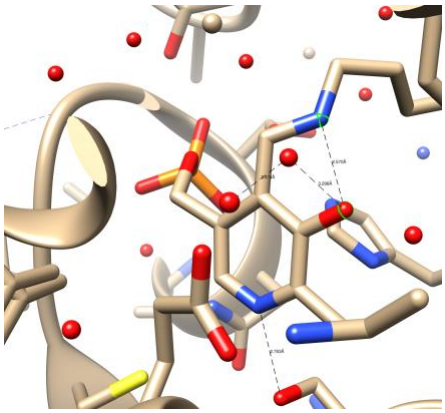


Figure 4.2 X-ray crystal structure of internal aldimine.

The ^{15}N solid-state NMR also directly verified the PSB hypothesis for the internal aldimine. Besides, NMR chemical shift has conformed ^{13}C, ^{15}N and ^{31}P chemical shifts on internal aldimine, restablishing the phosphoryl group, phenolic oxygen, and pyridine ring nitrogen are all deprotonated. The pyridine nitrogen in tryptophan synthase interacts with the hydroxyl of Ser-377. Whether or not pyridine nitrogen is assumed the role of hydrogen bond donor is a question without an additional proton. [28]

Research concerning the pronation of the PLP ring nitrogen has a long history. The first determined crystal structure of a PLP-dependent enzyme was an aspartate aminotransferase (AAT) that performs the transamination reaction. Its pyridine nitrogen interacted with a below positioned aspartate residue, and the acidic property of the aspartate functional group confirmed that the pyridine nitrogen is protonated in AAT (Figure 4.3.A). The second determined crystal structure of a PLP enzyme was tryptophan synthase, which performs a β-elimination and replacement reaction.[29] Its residue, located below the pyridine nitrogen, consisted of serine coupled with the hydroxyl group. The hydroxyl group is less likely to be a proton donor to the pyridine nitrogen.(Figure 4.3.B) When the X-ray crystal structure of alanine racemase(AR) was published, which interacts with a basic arginine side chain and catalyzes the racemization of L-alanine to D-alanine, it clearly showed a charged arginine residue positioned below the pyridine nitrogen, which indeed excludes the protonation of the pyridine nitrogen.(Figure 4.3.C)[46] In general, these observations from X-ray crystal structures raised the question of whether the protonated

pyridine nitrogen is required throughout the course of catalysis. The following work with these three enzymes—but with pyridine ring replaced with deazaPLP, a PLP analog with the pyridine nitrogen substituted with a carbon atom—showed they were incapable of forming a quinonoid intermediate, indicating that the loss of the hydrogen bond between the side chain pyridine nitrogen triggered different reductions in activity. In the AAT, it reduced the activity of the enzyme by greater than 109-fold, while in the AR, activity was reduced ~275-fold, and in O-acetylserine sulfhydrylase (OASS), which interacts with a polar serine hydroxyl sidechain such as the TS and catalyzes the β-elimination of the β-acetyl group (Figure 4.3.B), activity was reduced ~250-fold.[13] These results demonstrated that transamination reactions require a greater electron-withdrawing potential, attributed to the nitrogen in the pyridine, of the PLP cofactor than racemization and β-elimination/replacement reactions.[30] In the transamination reactions performed by fold type I enzymes, such as AAT, acidic residues lead to the protonation of the pyridine nitrogen, hence delocalizing the negative charge at C^{α} over the entire ring system. This delocalization forms the quinonoid intermediate through the resonance (Figure 3.7.A3) and results in the proposed reprotonation at C4′.[10] However, in AR and OASS, results with the deazaPLP implied that this delocalization is not necessary for reactions that do not require reprotonation at C4′.2 In these cases, negative charge accumulates at Cα as the reaction goes forward, thus forming the carbanion intermediate (Figure 3.7.B3).[12] Furthermore, a proposal that OASS exhibits a concerted-E2 type mechanism, which does not include a distinct carbanion/quinonoid intermediate, was made to explain how deazaPLP has little impact on the activities of OASS and AR in β-elimination reactions.[31,32]

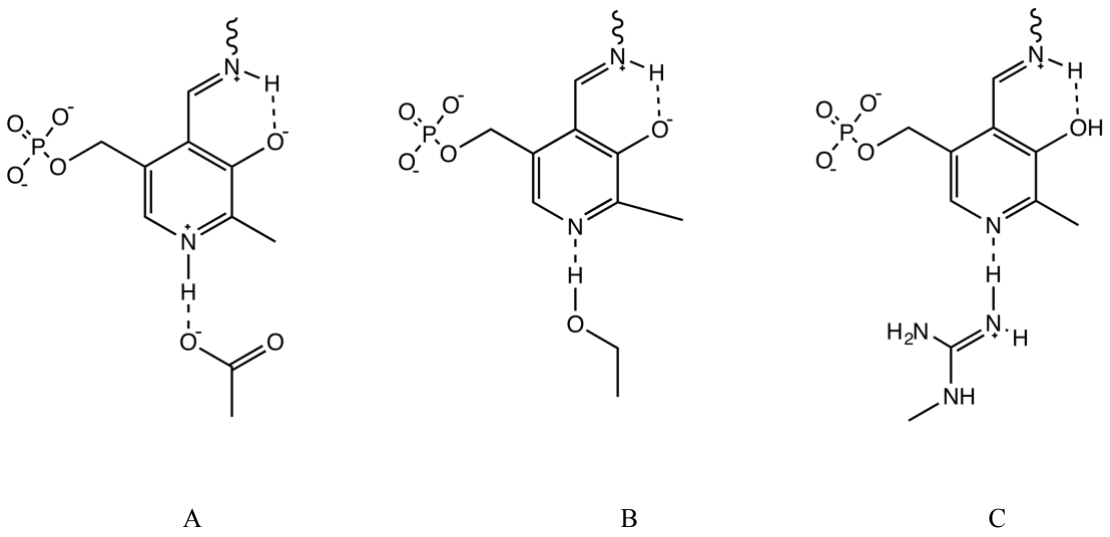


Figure 4.3 Three types of pronation of the PLP ring nitrogen.

The protonation profiles of TS intermediates have been studied using X-ray diffraction, [10,33,34,] computational studies with MD simulations[35,36], and solid-state NMR.[8] Each of these methods has its advantages and limitations. Typical X-ray crystal structures (1.3–2.5 Å resolution) of intermediates that bind the enzyme have shown that hydrogen bonds possibly exist between site residues and substrates, but protonation states cannot be determined from these structures. NMR stimulates direct chemical shifts in specific atoms within enzyme-substrate complexes. However, it requires a framework to explain. Ab initio calculations and molecular mechanics can be used to building models of enzymatic processes that must be based on the specified framework. Molecular dynamic simulations can provide accurate initial estimations for orientations of molecules, but they cannot determine the protonation states of the cofactors. However, they can be combined to describe the structure and function of enzyme active sites in models that are accordant and testable. [37–40] Previous studies on NMR crystallography have involved the characterization of the protonation profiles of PLP in carbanionic intermediate states for tryptophan synthase and shown remarkable results. [3] Here, we used NMR crystallography to explore more internal aldimine of tryptophan synthase by interpreting chemical shifts for selected atoms of the enzyme in terms of the protonation states and elucidating three-dimensional, chemically-detailed structures of the intermediate.

4.2 Experimental Details

4.2.1 CP-MAS SSNMR

The chemical shifts of PLP active sites, including the pyridine nitrogen, the Schiff base(SB) nitrogen and select carbon atoms, in the resting holoenzyme form, the internal aldimine state of tryptophan synthase have been reported by Bethany et al.[28] Table 4.1 lists all the measured chemical shifts of the internal aldimine.

Position	SB N	Pyridine N1	C2	C2`	C3
Measured Value	202.3	294.7	15.8	18.9	168.7

Table 4.1 Reported chemical shifts of internal aldimine from Bethany[28] et al.

Besides, two positions, the Schiff base nitrogen and the pyridine nitrogen(N1), were also performed variable-temperature ^{15}N CP-MAS solid state NMR whose theoretical foundations were elucidated in Chapter 1. Variable-temperature ^{15}N CP-MAS spectra of the microcrystalline internal intermediate Schiff base nitrogen and pyridine nitrogen(N1) were shown in Figure 4.4 and Figure 4.5 respectively. From Figure 4.4, Substantial temperature dependence is observed for the Schiff base nitrogen (yellow dot), while, from Figure 4.5, no evident temperature dependence is observed for the pyridine nitrogen (orange pink dot). The large spectral features around 125 ppm and 30 ppm in these two figures correspond to the spinning sidebands of the labeled backbone.

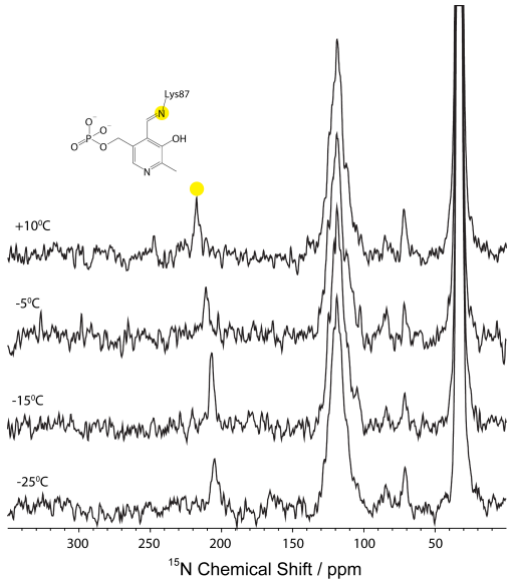


Figure 4.4 Variable-temperature 15N CP-MAS spectra for Schiff base nitrogen.

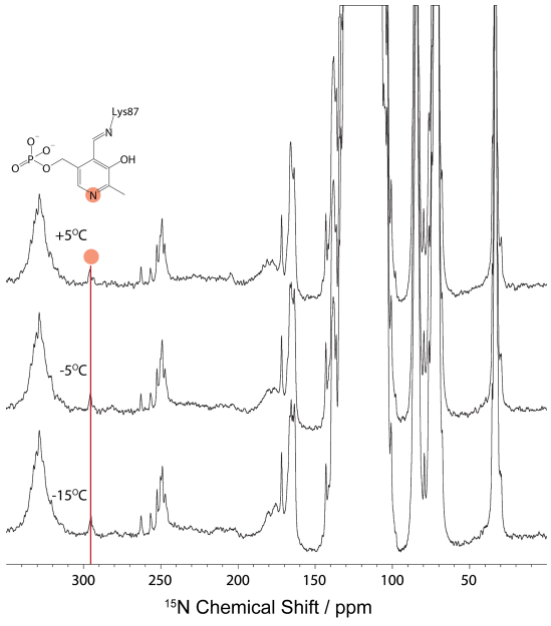


Figure 4.5 Variable-temperature 15N CP-MAS spectra for pyridine nitrogen.

4.2.2 First-Principle Calculations

Using the first-principles methods to calculate NMR chemical shielding has become increasingly reliable. The theoretical foundation is elaborated upon in Chapter 2.

The theoretical simulations of the factors determining the chemical shifts in natural conformations of proteins and peptides were motivated by Spera and Bax [41], who used an empirical correlation between the secondary chemical shifts and the peptide backbone conformation. Later, it was revealed that factors include electronic structures due to variation in the backbone torsion angles, environmental effects due to the electric field inside the protein, and the presence of hydrogen bonds.[42,43] It was the first time that researchers realized only accurate quantum chemical calculations could be used to analyze the relationship between the observed shifts and protein structures became clear. However, the first application of DFT methods to calculate C^{13} absolute shielding in a set of small organic molecules was rather disappointing due to the hard-to-ignore average deviations from the experimental values.[44]

Over time, proteins to be characterized became more and more massive, and the target accuracies of chemical shifts increased, which was quite demanding. Fortunately, quite encouraging results were obtained by combining the approaches of quantum mechanics and molecular mechanics (QM/MM).[45] These were the ONIOM[46] (Our own n-layered integrated molecular orbital and molecular mechanic) approach, the fragment approach (AF), [47] and a combination of these, AF-QM/MM. [48]

Due to the increasingly efficient algorithms and widespread access to supercomputing resources, the pure QM approach was allowed because nuclear magnetic shielding is primarily effected by near interactions. In the pure QM approach**,** the system was truncated to include components usually having experimentally acquired NMR chemical shifts, and adequate surroundings to keep critical interactions,.[49] The challenges of using cluster approaches to calculate NMR parameters in large systems lie in the trade-off between the size of the cluster and the choice of the basis set owing to the conflict between accuracy and computational cost (time), [49] so that NMR calculation in large systems lies in the determination of cluster size.

Hartman[49] et al. built a methodology to achieve the optimum strategy for the selection of the cluster size and the basis set. The experiments were performed in proteins, comparing the agreement between the calculated and experimentally measured chemical shielding for theoretical models with different cluster sizes and basis sets. Their findings led to selection suggestions for biological systems, and those relevant to this work are as follows: (1) the radius of the cluster should be at least ∼7 Å if the key nuclei are assigned as the center; if needed, longer-range molecular mechanical interactions are incorporated as well. For the internal aldimine, the PLP/substrate complex acts as the key interest nuclei. (2) a solitary large basis set can be changed to a multi-layer locally dense basis set, that is, a large basis set, such as triple-zeta, with diffuse functions utilized for crucial interest nuclei, a reduced triple-zeta basis excluded diffuse functions for the adjacent structure, which contains the hydrogen bonding or van der Waals interactions; and a less expensive basis set, such as a double-zeta basis set, is used for the remaining shifts. These two recommendations work together to not only ensure the accuracy of atoms with which the protein interacts strongly, such as hydrogen bond donors/acceptors or charged functional groups, but also to reduce computation cost by utilizing reasonable basis sets in the middle and lower tiers of the enzyme.

This pure QM cluster approach provides several advantages: (1) The ONIOM places the PLP/substrate complex alone at the QM level and leaves all other atoms at the semi-empirical PM3 level, whereas the QM cluster treats every atom at the QM level though the multi-tier method. (2) The QM cluster approach applies geometry optimization to the cofactor within the active site environment, to permit the movement of the PLP/substrate as well as the surrounding layer, which includes intermolecular hydrogen bond partners and charges, while the QM/MM approach (ONIOM) only allows the PLP/substrate atoms to relax their coordinates. (3) Even though the pure QM cluster approach involves a greater computation cost than QM/MM approach (ONIOM), it is still practically feasible by virtue of the multi-tier basis set: it normally takes1–2 days to run the NMR calculations on the cluster with 600–700 atoms on a computer cluster implemented with 32 processors and 48 GB of RAM in parallel.[49]

For the internal aldimine, the first-principles calculation was performed using cluster models of the X-ray crystal structure of the active site for internal aldimine E(Ain) of *S. Typhimurium* downloaded from PDB data bank (PDB ID: 4HT3). The cluster model was created using PyMol software by truncating

the crystallographic structure into the 500–600 atom fragments that include the PLP cofactor and the protein active site within 7 Å distance from the cofactor. A distance of 7 Å provides the optimum balance between cluster size and basis set selection by performing a systematic approach to NMR calculations in proteins. Then, to simplify the modeling work and reduce the required computational resources while maintaining the radius of 7Å, further modifications were made: (i) Residues excluded from the continuous segment of the 7 Å cut and have only one atom within 7 Å were eliminated (β-site residues Val201, Gln205, Ser301, His313). (ii) Similarly, residues with two backbone atoms within 7 Å were replaced with alanine (β-site residues Arg379). (iii) Hydrogen atoms at the C-termini carbonyl carbons of backbone segments were replaced with –NH2 groups to yield amides, while the resulting [NH3] + N-termini of backbone segments were replaced with a hydrogen atom.(iv) Each Cs+ atom was changed to a Na+ atom to carry out the posterior computation. The final cluster (Figure 4.6) secured a total of approximately 570 atoms.

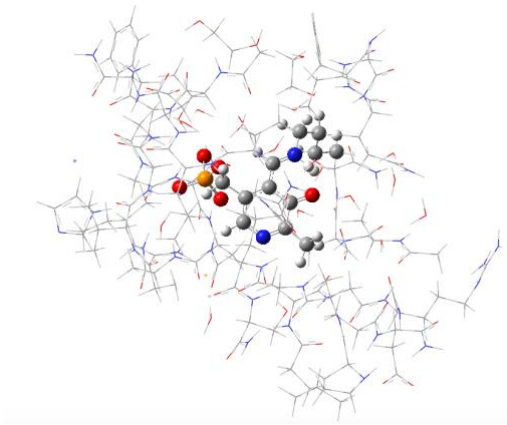
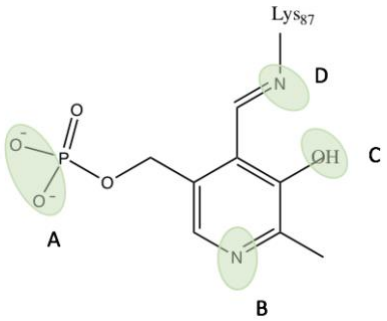


Figure 4.6 The cluster model of the β-subunit active site for first-principles geometry optimization and chemical shift calculations. the cofactor and substrates are represented in ball-and-stick and Protein side chains are shown in wireframe.

Based on the protonation state of the cofactor–substrate complex, different structures were created. A four-digit binary code was assigned to each structure where "0" represented a deprotonated position and "1" represented a protonated position. Potential sites of protonation included 1) the PLP phosphate group, 2) the PLP pyridine nitrogen, 3) the PLP phenolic oxygen, and 4) the Schiff base nitrogen (Figure 4.7).

Although the phosphate group protonation was confirmed in the dianionic state by the ^{31}P CSA tensor measurement, the nomenclature was kept unchanged to be consistent with the nomenclature of the previous study. Thus, the structures that had only one protonated position were "0100" (protonated pyridine nitrogen), "0010" (protonated phenolic oxygen), and "0001" (PSB nitrogen).

The structures with protons in two protonated positions were “0110” (both PLP pyridine nitrogen and protonated phenolic oxygen) and “0101” (both protonated pyridine nitrogen and Schiff base nitrogen). The completely deprotonated ("0000") and protonated ("0111") structures were considered as well. All structures are shown in Figure 4.8.



A-B-CD

Figure 4.7 Potential sites of protonation on and near the cofactor/substrate complex include (A) the PLP phosphate group, (B) the PLP pyridine ring nitrogen, (C) the PLP phenolic oxygen, (D) the Schiff-base nitrogen.

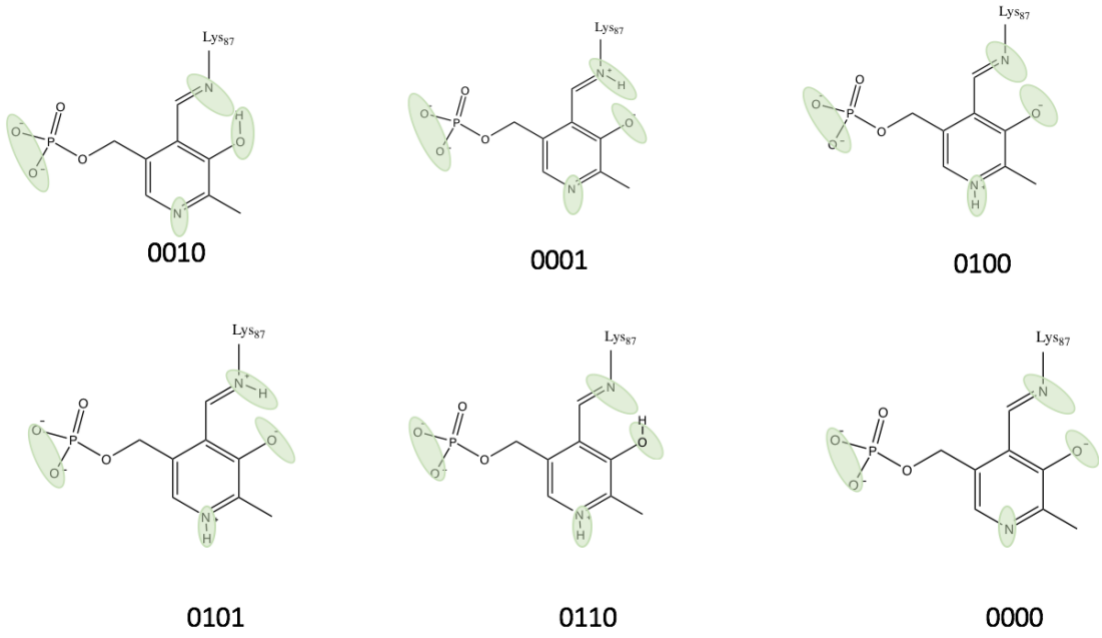


Figure 4.8 Protonation states of 6 candidate structures constructed for NMR shielding calculations.

Each structure was optimized separately using Gaussian 09 software [51]. To minimize calculation time, the initial geometry optimization at a low basis set involved freezing the coordinates of all atoms except the hydrogen atoms and the modified nitrogen atoms at the terminal ends. All calculations were performed via the DFT-hybrid B3LYP functional [51–54] with the 6-31G basis. This optimization allowed a quick pre-optimization of the structures and the setting of the initial orientations for water molecules and amino acids.

All models were then subjected to geometry optimization via a hybrid density-function-theory (DFT) approach: B3LYP (refer to the section). Basis sets were selected using a two-tier method: PLP/substrate complex was paired with locally dense basis set 6-31G(d,p), while the remaining atoms were paired with 6-31G. Atoms within 4 Å, as well as all hydrogen atoms in the cluster, were allowed to move, whereas the rest were fixed at coordinates provided by the crystallography experiments. Between the geometry optimizations, a restriction that adjustable atoms were allowed to move on the order of 0.2 Å, at a maximum, and less than 0.1 Å RMSD from the initial coordinates was imposed.

All optimized structures were observed for the presence of local minimums and possibilities for alternative orientations of hydrogens. The optimization results were compared to the results of the MD simulations, and the next hydrogen atoms were manually rotated into a new orientation.

Following the optimization, the NMR shielding calculation was carried out on the ground of the gauge-independent atomic orbital (GIAO) method at the DFT level, which is implemented in the Gaussian-09 program. The DFT calculation utilized the B3LYP hybrid functional but, compared with the two-layer basis set in the geometry optimization, with a more accurate three-layer basis set[55]: The high layer, including the PLP/substrate complex, was assigned the 6-311+G(2d,p) basis set; the medium layer, comprised of any remaining atom within 4 Å of the center of the high layer, was assigned the 6-311G(d,p) basis set, and all remaining atoms composed of the low layer were given the 6-31G basis set.[56]

The output generated from the gauge-independent atomic orbital (GIAO) calculation offered every atom an absolute magnetic shielding value, yet experimental NMR measurements provided isotropic chemical shifts, δ_{iso}, which is a relative value with the referenced material, such as the adamantane. In order to compare the experimental and calculated results, a map of absolute shielding values to referenced chemical shifts had been produced. It was uniquely determined via the theoretical models and basis sets, which had been performed by Hartman et al. on a variety of test sets of crystal structures. The conversion proved to be a linear rescale between the isotropic chemical shift and the absolute magnetic shielding:

$$\delta_{iso} = A\sigma_{iso} + B,$$

where A and B are two parameters determined from the best fit of all data points[50,7] and are shown on Table 4.2 for ^{13}C, ^{15}N, and ^{17}O.

Atom	A	B(ppm)	δ_{iso} RMSE(ppm)
N^{15}	-0.9996	230.45	4.82
C^{13}	-0.9685	173.70	1.49
O^{17}	-1.0612	265.79	8.25

Table 4.2 Parameters of linear rescale between the isotropic chemical shift and the absolute magnetic shielding

NMR shielding (σ) values of the atoms, which are at the three ionizable sites (βLys87 Nε, pyridine ring nitrogen, and phenolic oxygen), were converted to chemical shifts (δ) using the following linear rescaling relationships with parameters shown above.[56,58]

$$C^{13}: \delta[TMS(l)] = (-0.9685)\sigma_{calc} + 173.70$$

$$N^{15}: \delta[NH3(l)] = (-0.9996)\sigma_{calc} + 230.45$$

In addition to the values of A and B, root-mean-square errors (also attached to Table 4.2) for every type of isotropic shift were acquired after applying best-fit rescaling.[56]

Root-mean-square errors were used in the reduced-χ2 analysis, which is widely used in goodness-of-fit tests, to evaluate every candidate structure's matching between the calculated chemical shifts and experimentally measured counterparts.[58] The formal expression of the reduced χ2 statistic used in the experiment is as follows:

$$\mathrm{x_r^2} = \frac{1}{\mathrm{N-f}} \sum_{\mathrm{i=1}}^{\mathrm{N}} \frac{\left(\delta_\mathrm{i}^{\mathrm{model}} - \delta_\mathrm{i}^{\mathrm{exp}}\right)^2}{\mathrm{s_i^2}},$$

where $\delta_\mathrm{i}^{\mathrm{model}}$ and $\delta_\mathrm{i}^{\mathrm{exp}}$ represent the calculated and the experimental chemical shifts, respectively, at the site *I*; $\mathrm{s_i}$ denotes the deviation of the site *I* expressed as root-mean-square, and s_i with different values corresponds to different varieties of nuclear species, such as, nitrogen, carbon, and oxygen. $N - f$ denotes the degree of freedom derived from statistics, *N* represents the count of all compared $\delta_\mathrm{i}^{\mathrm{model}}$ points, and f is

the number of dependent points in δ_i^{exp} since δ_i^{exp} can be calculated from δ_i^{model}. Here, N is 5, which refers to the number of chemical shifts measured, and f equals to 0 since all δ_i^{exp} were acquired from experiments independently rather than δ_i^{model}.

4.3 Results and Discussion

Figure 4.9 ranked all candidate models with their corresponding reduced χ2 values. Among all candidate structures, none were found to be below the acceptable threshold between the calculated and experimental ^{13}C and ^{15}N chemical shifts (the lowest reduced χ2 =13.08).

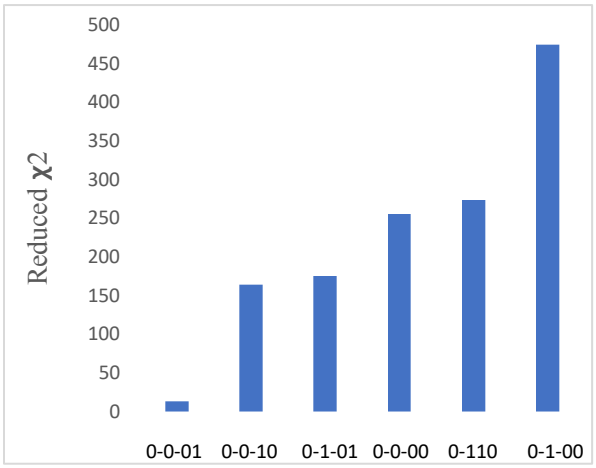


Figure 4.9 Reduced-χ2 comparing the experimental and first-principles isotropic chemical shifts for geometry-optimized active-site models with varying protonation states.

The temperature dependence experiments, which were conducted on the solid-state NMR to measure the chemical shifts at different temperatures, shows that as the temperature increases from 273K to 287K and the measured chemical shifts change, a proton shifts away from the phenolic oxygen and toward the PSB nitrogen, implying a fast tautomeric equilibrium between these two species. This phenomenon is known as the intramolecular hydrogen bond in PLP-dependent enzymes. The two-site model has been proposed to interpret the experimental nitrogen chemical shift with the mutual participation of two

species.[59] Therefore, the two-site exchange model, in which only pairs of structures containing the same number of protons were considered and weighted, produces the lowest reduced χ2 value.

The reduced χ2 for two-site equilibrium model can be written as

$$x_r^2 = \frac{1}{N}\sum_i \frac{\left(\alpha\delta_i^m + \beta\delta_i^n - \delta_i^{exp}\right)^2}{s_i^2}, \quad (4.1)$$

subject to $\alpha + \beta = 1$

$0 \leq \alpha \leq 1$

$0 \leq \beta \leq 1.$

Here the superscript *m* and *n* refer to different candidate structures, so that $\alpha\delta_i^m + \beta\delta_i^n$ means the weight chemical shift at site *i* between candidate structures m and candidate structures n.

The minimum value of the reduced χ2 can be determined by adjusting the α and β since Eq.(4.1) is a mathematically constrained convex equation[60,61]. By applying the CVXOPT quadratic core package[62], the reduced-χ2 values comparing the experimental and computational isotropic chemical shifts for fast-exchange mixture models are list in Table 4.3.

Two models lie within the 95% confidence limit: one that has a reduced χ2 value of 0.798 and represents an 86.13%:13.87% equilibrium of the structure with PSB nitrogen (0-0-01) and structure with protonated pyridine nitrogen (0-1-00), and another with a reduced χ2 value of 1.034 and an equilibrium with structure PSB nitrogen (0-0-01) and the structure with protonated phenolic oxygen (0-0-10) at a ratio of 78.6%:21.38%.

Str1	**Str2**	**Reduced-χ2**	**Ratio of str1 and str2**

0-0-01	0-1-00	0.798	86.13%:13.87%
0-0-01	0-0-10	1.034	78.62%:21.38%
0-0-10	0-1-00	164.029	100.0%: 0.00%
0-1-01	0-1-10	165.062	76.56%: 23.44%

Table 4.3 Reduced-χ2 comparing the experimental and first-principles isotropic chemical shifts for the mixture structures.

Based on variable temperature experiments in solid-state NMR, the substantial chemical shifts in Schiff base nitrogen at different temperatures suggested that this site was likely to be involved in the chemical exchange. On the other hand, the absence of significant temperature dependence on the PLP nitrogen suggests this site was not involved in the chemical exchange. Therefore, the model with 0-0-01 and 0-1-00 can be ruled out, excluding an equilibrium between 0-0-01 and 0-0-10, which reinforces the PSB hypothesis. The experimental and calculated chemical shifts for 0-0-01 and 0-0-10 and their fast-exchange mixture are shown in Table 2. The major species corresponds to the PSB nitrogen (0-0-01), which is ranked as the best single configuration, and the minor species, 0-0-10, represents the transfer of a proton from the PSB nitrogen to the phenolic oxygen, hereafter referred to as the PSB tautomer.

Position	Exp	Calculated 0-0-01	Calculated 0-0-10	Two-site mixture
Lys 87N	202.3	168.5	316.8	200.3
Pyridine N	294.7	302.1	300.3	301.7
C3	168.7	169.8	155.1	166.7
C2	158.0	158.8	150.3	157.0
C2`	18.9	19.3	18.1	19.0

Table 4.4 Experimental and calculated chemical Shifts [in ppm Relative to TMS (13C) and NH3(l) (15N)a] for the Fast-cxchange equilibrium between two sites.

The PSB hypothesis was proposed to boost reactivity toward nucleophiles at the Schiff base carbon compared to the neutral imine. [63–65] The results of X-ray crystallography[2], UV/vis spectroscopy, [66,67] and NMR spectroscopy[28] all suggest a protonated Schiff base form.

Here we revealed that internal aldimine intermediate undergoes fast proton exchange between its Schiff base and phenolic form at room temperature, in agreement with the PSB hypothesis, which proposed the protonated Schiff base activates the nucleophilic attack at C4' of the pyridine.

The canonical mechanism for enzymatic PLP-dependent reactions suggests that protonated pyridine nitrogen (the canonical/true quinonoid form) in the carbanionic intermediate substantially affects posterior reaction specificity for PLP-dependent enzymes: 1) a protonated pyridine nitrogen acts as an electron sink to withdraw the locally located charge and promote electrophilic addition at C4', whereas a deprotonated pyridine is associated with the reaction at Cα;[2,68–71] 2) the PSB hypothesis proposes that proton transfer from the phenolic oxygen to the Schiff-base nitrogen is promoted by the protonation of the PLP nitrogen,[72,73] making C4' the preferred location for the nucleophilic attack during the transamination reaction.[1]

For TS, our calculation consistent with the experimental result shows that pyridine nitrogen is unprotonated, which echoes the next step of carbanionic intermediate, assuming it is maintained during the catalytic cycle, in the β- replacement pathway. The possibility of a hydroxyl group being cleft at Cα because deprotonated pyridine favors the reaction at the substrate Cα, however, is in disagreement with the PSB hypothesis, which proposes the protonation of the PLP nitrogen promotes the protonated Schiff base and activates the nucleophilic attack of the pyridine ring at C4' in the process of forming the external aldimine.

Instead of the protonation of the PLP nitrogen, another mechanism of activation has been proposed: the PLP phenolic oxygen can be stabilized by hydrogen-bonding with surrounding water molecules.[2,20] Regarding the active site of TS, the X-ray crystal structure of the *S. typhimurium* TS internal aldimine [34] indeed shows that the phenolic oxygen participates in a hydrogen-bonding network with two observed water molecules, Wat-684 and Wat-602, as indicated in Figure 4.10.

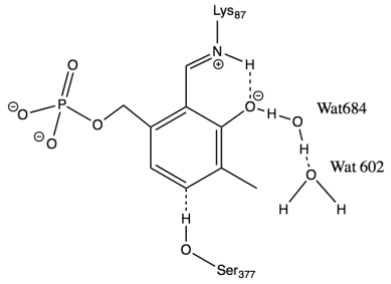


Figure 4.10 Schematic of PLP covalently bound to βLys87 in the β-subunit active site of tryptophan synthase. Distances and water molecules are from the X-ray crystal structure of the ST typhimurium TS internal aldimine state.

Our first-principles calculation is consistent with the X-ray crystal structure and found the pyridine nitrogen in tryptophan synthase forms a hydrogen bond with the positionally lower Ser377. Our calculation confirmed a hydrogen-bonding network with the phenolic oxygen. The distance between the phenolic oxygen of PLP and Wat-684 hydrogen is 1.824 Å, and the distance between the oxygen on Water 684 and the hydrogen on Water 602 is 2.024Å, which is in agreement with the length of the hydrogen bond. This hydrogen-bonding network stabilizes the phenolic oxygen, shifting the proton in the intramolecular OHN H-bond from the phenolic oxygen to the aldimine nitrogen, reinforcing the PSB hypothesis.

4.4 Conclusion

Application of NMR crystallography, the highly combination of solid-state NMR spectroscopy, X-ray crystallography, and computational chemistry, to the tryptophan synthase internal aldimine intermediate offers internal aldimine structure, protonation states, and chemical dynamics an atomic-resolution description that would be impossible to achieve by the individual application of these techniques. With the utilization of solid NMR spectroscopy at different temperatures, only one model through the first principles calculation of NMR parameters confidently established. This model reveals that the internal aldimine is undergoing fast proton exchange between its protonated Schiff base and phenolic tautomeric forms. The deprotonated pyridine ring nitrogen and hydrogen-bonding network to stabilize phenolic oxygen are consistent both with the PSB hypothesis for PLP activation and the role of tryptophan synthase as a β

elimination/replacement catalyst. These findings support the hypothesis that reaction specificity in PLP-dependent enzymes is attributed to the protonation states of ionizable groups on PLP and the reacting substrates. Some PLP-dependent enzymes echo the stabilization of a canonical quinonoid form to maintain this specificity. These results also prove that NMR crystallography play a decisive role in characterizing chemical structure within enzyme active sites, and its ability.

4.5 References

(1) Metzler, D. E.; Ikawa, M.; Snell, E. E. A General Mechanism for Vitamin B6-Catalyzed Reactions1. *J. Am. Chem. Soc.* **1954**, *76* (3), 648–652. https://doi.org/10.1021/ja01632a004.

(2) Toney, M. D. Controlling Reaction Specificity in Pyridoxal Phosphate Enzymes. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2011**, *1814* (11), 1407–1418. https://doi.org/10.1016/j.bbapap.2011.05.019.

(3) Caulkins, B. G.; Young, R. P.; Kudla, R. A.; Yang, C.; Bittbauer, T. J.; Bastin, B.; Hilario, E.; Fan, L.; Marsella, M. J.; Dunn, M. F.; Mueller, L. J. NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. *J. Am. Chem. Soc.* **2016**, *138* (46), 15214–15226. https://doi.org/10.1021/jacs.6b08937.

(4) Frontiers | Current Advances on Structure-Function Relationships of Pyridoxal 5′-Phosphate-Dependent Enzymes | Molecular Biosciences https://www.frontiersin.org/articles/10.3389/fmolb.2019.00004/full (accessed Dec 11, 2019).

(5) Toney, M. D. Reaction Specificity in Pyridoxal Phosphate Enzymes. *Arch. Biochem. Biophys.* **2005**, *433* (1), 279–287. https://doi.org/10.1016/j.abb.2004.09.037.

(6) Sharif, S.; Huot, M. C.; Tolstoy, P. M.; Toney, M. D.; Jonsson, K. H. M.; Limbach, H.-H. 15N Nuclear Magnetic Resonance Studies of Acid−Base Properties of Pyridoxal-5‘-Phosphate Aldimines in Aqueous Solution. *J. Phys. Chem. B* **2007**, *111* (15), 3869–3876. https://doi.org/10.1021/jp067334g.

(7) Richard, J. P.; Amyes, T. L.; Crugeiras, J.; Rios, A. Pyridoxal 5’-Phosphate: Electrophilic Catalyst Extraordinaire. *Curr. Opin. Chem. Biol.* **2009**, *13* (4), 468–476. https://doi.org/10.1016/j.cbpa.2009.06.023.

(8) Lin, Y.-L.; Gao, J.; Rubinstein, A.; Major, D. T. Molecular Dynamics Simulations of the Intramolecular Proton Transfer and Carbanion Stabilization in the Pyridoxal 5’-Phosphate Dependent Enzymes L-Dopa Decarboxylase and Alanine Racemase. *Biochim. Biophys. Acta* **2011**, *1814* (11), 1438–1446. https://doi.org/10.1016/j.bbapap.2011.05.002.

(9) Major, D. T.; Gao, J. A Combined Quantum Mechanical and Molecular Mechanical Study of the Reaction Mechanism and α-Amino Acidity in Alanine Racemase. *J. Am. Chem. Soc.* **2006**, *128* (50), 16345–16357. https://doi.org/10.1021/ja066334r.

(10) Snell, E. E.; Jenkins, W. T. The Mechanism of the Transamination Reaction. *J. Cell. Comp. Physiol.* **1959**, *54* (S1), 161–177. https://doi.org/10.1002/jcp.1030540413.

(11) Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. Computational Mechanistic Studies Addressed to the Transimination Reaction Present in All Pyridoxal 5′-Phosphate-Requiring Enzymes. *J. Chem. Theory Comput.* **2011**, *7* (5), 1356–1368. https://doi.org/10.1021/ct1002219.

(12) Griswold, W. R.; Toney, M. D. Role of the Pyridine Nitrogen in Pyridoxal 5′-Phosphate Catalysis: Activity of Three Classes of PLP Enzymes Reconstituted with Deazapyridoxal 5′-Phosphate. *J. Am. Chem. Soc.* **2011**, *133* (37), 14823–14830. https://doi.org/10.1021/ja2061006.

(13) Griswold, W. R.; Toney, M. D. Role of the Pyridine Nitrogen in Pyridoxal 5′-Phosphate Catalysis: Activity of Three Classes of PLP Enzymes Reconstituted with Deazapyridoxal 5′-Phosphate. *J. Am. Chem. Soc.* **2011**, *133* (37), 14823–14830. https://doi.org/10.1021/ja2061006.

(14) Cordes, E. H.; Jencks, W. P. Semicarbazone Formation from Pyridoxal, Pyridoxal Phosphate, and Their Schiff Bases. *Biochemistry* **1962**, *1* (5), 773–778. https://doi.org/10.1021/bi00911a007.

(15) Heinert, Dietrich.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. VII. Acid-Base Equilibria of Schiff Bases. *J. Am. Chem. Soc.* **1963**, *85* (2), 188–193. https://doi.org/10.1021/ja00885a018.

(16) Metzler, D. E. Equilibria between Pyridoxal and Amino Acids and Their Imines1. *J. Am. Chem. Soc.* **1957**, *79* (2), 485–490. https://doi.org/10.1021/ja01559a068.

(17) Sharif, S.; Powell, D. R.; Schagen, D.; Steiner, T.; Toney, M. D.; Fogle, E.; Limbach, H. H. X-Ray Crystallographic Structures of Enamine and Amine Schiff Bases of Pyridoxal and Its 1:1 Hydrogen-Bonded Complexes with Benzoic Acid Derivatives: Evidence for Coupled Inter- and Intramolecular Proton Transfer. *Acta Crystallogr. B* **2006**, *62* (Pt 3), 480–487. https://doi.org/10.1107/S0108768105040590.

(18) Snell, E. E.; Di Mari, S. J. 7 Schiff Base Intermediates in Enzyme Catalysis. In *The Enzymes*; Boyer, P. D., Ed.; Academic Press, 1970; Vol. 2, pp 335–370. https://doi.org/10.1016/S1874-6047(08)60186-9.

(19) Heinert, D.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. II.1,2 Infrared Spectra and Hydrogen Bonding. *J. Am. Chem. Soc.* **1959**, *81* (15), 3933–3943. https://doi.org/10.1021/ja01524a031.

(20) Chalk, A. J.; Halpern, J.; Harkness, A. C. Medium Effects in the Homogeneous Catalytic Activation of Molecular Hydrogen by Metal Salts. III. Silver and Mercuric Salts1. *J. Am. Chem. Soc.* **1959**, *81* (22), 5854–5857. https://doi.org/10.1021/ja01531a004.

(21) Nakamoto, K.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. IV. Ultraviolet Spectra and Solution Equilibria of 3-Methoxypyridine-2(and 4-)-Aldehydes and of 3-Hydroxypyridine-2 (and 4-)-Aldehydes1,2. *J. Am. Chem. Soc.* **1959**, *81* (22), 5863–5869. https://doi.org/10.1021/ja01531a006.

(22) Peterson, R. F.; Nauman, L. W.; McMeekin, T. L. The Separation and Amino Acid Composition of a Pure Phosphopeptone Prepared from β-Casein by the Action of Trypsin2. *J. Am. Chem. Soc.* **1958**, *80* (1), 95–99. https://doi.org/10.1021/ja01534a026.

(23) Heinert, Dietrich.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. VIII. Synthesis and Infrared Spectra of Metal Chelates. *J. Am. Chem. Soc.* **1963**, *85* (9), 1334–1337. https://doi.org/10.1021/ja00892a026.

(24) Heinert, Dietrich.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. VI. Electronic Absorption Spectra of Schiff Bases. *J. Am. Chem. Soc.* **1963**, *85* (2), 183–188. https://doi.org/10.1021/ja00885a017.

(25) Auld, D. S.; Bruice, T. C. Catalytic Reactions Involving Azomethines. VIII.1 Water and Alanine Catalysis of the Transamination of 3-Hydroxypyridine-4-Aldehyde ba Alanine. *J. Am. Chem. Soc.* **1967**, *89* (9), 2090–2097. https://doi.org/10.1021/ja00985a021.

(26) Auld, D. S.; Bruice, T. C. Catalytic Reactions Involving Azomethines. IX.1 General Base Catalysis of the Transamination of 3-Hydroxypyridine-4-Aldehyde ba Alanine. *J. Am. Chem. Soc.* **1967**, *89* (9), 2098–2106. https://doi.org/10.1021/ja00985a022.

(27) Fischer, E. H.; Kent, A. B.; Snyder, E. R.; Krebs, E. G. THE REACTION OF SODIUM BOROHYDRIDE WITH MUSCLE PHOSPHORYLASE [1]. *J. Am. Chem. Soc.* **1958**, *80* (11), 2906–2907. https://doi.org/10.1021/ja01544a082.

(28) Caulkins, B. G.; Bastin, B.; Yang, C.; Neubauer, T. J.; Young, R. P.; Hilario, E.; Huang, Y. M.; Chang, C. A.; Fan, L.; Dunn, M. F.; Marsella, M. J.; Mueller, L. J. Protonation States of the Tryptophan Synthase Internal Aldimine Active Site from Solid-State NMR Spectroscopy: Direct Observation of the Protonated Schiff Base Linkage to Pyridoxal-5'-Phosphate. *J. Am. Chem. Soc.* **2014**, *136* (37), 12824–12827. https://doi.org/10.1021/ja506267d.

(29) Hyde, C. C.; Ahmed, S. A.; Padlan, E. A.; Miles, E. W.; Davies, D. R. Three-Dimensional Structure of the Tryptophan Synthase Alpha 2 Beta 2 Multienzyme Complex from Salmonella Typhimurium. *J. Biol. Chem.* **1988**, *263* (33), 17857–17871.

(30) Richard, J. P.; Amyes, T. L.; Crugeiras, J.; Rios, A. Pyridoxal 5'-Phosphate: Electrophilic Catalyst Extraordinaire. *Curr. Opin. Chem. Biol.* **2009**, *13* (4), 475–483. https://doi.org/10.1016/j.cbpa.2009.06.023.

(31) Tai, C.-H.; Cook, P. F. Pyridoxal 5'-Phosphate-Dependent α,β-Elimination Reactions: Mechanism of O-Acetylserine Sulfhydrylase. *Acc. Chem. Res.* **2001**, *34* (1), 49–59. https://doi.org/10.1021/ar990169l.

(32) He, Y.; Xiang, S.; Chen, B. A Microporous Hydrogen-Bonded Organic Framework for Highly Selective C2H2/C2H4 Separation at Ambient Temperature. *J. Am. Chem. Soc.* **2011**, *133* (37), 14570–14573. https://doi.org/10.1021/ja2066016.

(33) Burkhard, P.; Jagannatha Rao, G. S.; Hohenester, E.; Schnackerz, K. D.; Cook, P. F.; Jansonius, J. N. Three-Dimensional Structure of O-Acetylserine Sulfhydrylase from Salmonella Typhimurium11Edited by R. Huber. *J. Mol. Biol.* **1998**, *283* (1), 121–133. https://doi.org/10.1006/jmbi.1998.2037.

(34) Niks, D.; Hilario, E.; Dierkers, A.; Ngo, H.; Borchardt, D.; Neubauer, T. J.; Fan, L.; Mueller, L. J.; Dunn, M. F. Allostery and Substrate Channeling in the Tryptophan Synthase Bienzyme Complex: Evidence for Two Subunit Conformations and Four Quaternary States. *Biochemistry* **2013**, *52* (37), 6396–6411. https://doi.org/10.1021/bi400795e.

(35) Protonation states and catalysis: Molecular dynamics studies of intermediates in tryptophan synthase https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4815307/ (accessed Dec 11, 2019).

(36) Investigation of Structural Dynamics of Enzymes and Protonation States of Substrates Using Computational Tools https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5119520/ (accessed Dec 11, 2019).

(37) Harris, R. K.; Hodgkinson, P.; Pickard, C. J.; Yates, J. R.; Zorin, V. Chemical Shift Computations on a Crystallographic Basis: Some Reflections and Comments. *Magn. Reson. Chem.* **2007**, *45* (S1), S174–S186. https://doi.org/10.1002/mrc.2132.

(38) Elena, B.; Pintacuda, G.; Mifsud, N.; Emsley, L. Molecular Structure Determination in Powders by NMR Crystallography from Proton Spin Diffusion. *J. Am. Chem. Soc.* **2006**, *128* (29), 9555–9560. https://doi.org/10.1021/ja062353p.

(39) Mueller, L. J.; Dunn, M. F. NMR Crystallography of Enzyme Active Sites: Probing Chemically Detailed, Three-Dimensional Structure in Tryptophan Synthase. *Acc. Chem. Res.* **2013**, *46* (9), 2008–2017. https://doi.org/10.1021/ar3003333.

(40) Elena, B.; Pintacuda, G.; Mifsud, N.; Emsley, L. Molecular Structure Determination in Powders by NMR Crystallography from Proton Spin Diffusion. *J. Am. Chem. Soc.* **2006**, *128* (29), 9555–9560. https://doi.org/10.1021/ja062353p.

(41) Empirical correlation between protein backbone conformation and C.alpha. and C.beta. 13C nuclear magnetic resonance chemical shifts | Journal of the American Chemical Society https://pubs.acs.org/doi/abs/10.1021/ja00014a071 (accessed Apr 13, 2020).

(42) Oldfield, E. Chemical Shifts in Amino Acids, Peptides, and Proteins: From Quantum Chemistry to Drug Design. *Annu. Rev. Phys. Chem.* **2002**, *53*, 349–378. https://doi.org/10.1146/annurev.physchem.53.082201.124235.

(43) Forsyth, D. A.; Sebag, A. B. Computed 13C NMR Chemical Shifts via Empirically Scaled GIAO Shieldings and Molecular Mechanics Geometries. Conformation anc Configuration from 13C Shifts. *J. Am. Chem. Soc.* **1997**, *119* (40), 9483–9494. https://doi.org/10.1021/ja970112z.

(44) de Dios, A. C.; Pearson, J. G.; Oldfield, E. Secondary and Tertiary Structural Effects on Protein NMR Chemical Shifts: An Ab Initio Approach. *Science* **1993**, *260* (5113), 1491–1496. https://doi.org/10.1126/science.8502992.

(45) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed Engl.* **2009**, *48* (7), 1198–1229. https://doi.org/10.1002/anie.200802019.

(46) ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition | The Journal of Physical Chemistry https://pubs.acs.org/doi/10.1021/jp962071j (accessed Apr 13, 2020).

(47) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112* (1), 632–672. https://doi.org/10.1021/cr200093j.

(48) He, X.; Wang, B.; Merz, K. M. Protein NMR Chemical Shift Calculations Based on the Automated Fragmentation QM/MM Approach. *J. Phys. Chem. B* **2009**, *113* (30), 10380–10388. https://doi.org/10.1021/jp901992p.

(49) Hartman, J. D.; Neubauer, T. J.; Caulkins, B. G.; Mueller, L. J.; Beran, G. J. O. Converging Nuclear Magnetic Shielding Calculations with Respect to Basis and System Size in Protein Systems. *J. Biomol. NMR* **2015**, *62* (3), 327–340. https://doi.org/10.1007/s10858-015-9947-2.

(50) Hartman, J. D.; Neubauer, T. J.; Caulkins, B. G.; Mueller, L. J.; Beran, G. J. O. Converging Nuclear Magnetic Shielding Calculations with Respect to Basis and System Size in Protein Systems. *J. Biomol. NMR* **2015**, *62* (3), 327–340. https://doi.org/10.1007/s10858-015-9947-2.

(51) Gaussian 09 Citation | Gaussian.com https://gaussian.com/g09citation/ (accessed Nov 18, 2019).

(52) Becke, A. D. A New Mixing of Hartree–Fock and Local Density-functional Theories. *J. Chem. Phys.* **1993**, *98* (2), 1372–1377. https://doi.org/10.1063/1.464304.

(53) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785–789. https://doi.org/10.1103/PhysRevB.37.785.

(54) Bonhomme, C.; Gervais, C.; Babonneau, F.; Coelho, C.; Pourpoint, F.; Azaïs, T.; Ashbrook, S. E.; Griffin, J. M.; Yates, J. R.; Mauri, F.; Pickard, C. J. First-Principles Calculation of NMR Parameters Using the Gauge Including Projector Augmented Wave Method: A Chemist's Point of View. *Chem. Rev.* **2012**, *112* (11), 5733–5779. https://doi.org/10.1021/cr300108a.

(55) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98* (45), 11623–11627. https://doi.org/10.1021/j100096a001.

(56) Hartman, J. D.; Kudla, R. A.; Day, G. M.; Mueller, L. J.; Beran, G. J. O. Benchmark Fragment-Based 1H, 13C, 15N and 17O Chemical Shift Predictions in Molecular Crystals. *Phys. Chem. Chem. Phys.* **2016**, *18* (31), 21686–21709. https://doi.org/10.1039/C6CP01831A.

(57) Hartman, J. D.; Monaco, S.; Schatschneider, B.; Beran, G. J. O. Fragment-Based 13C Nuclear Magnetic Resonance Chemical Shift Predictions in Molecular Crystals: An Alternative to Planewave Methods. *J. Chem. Phys.* **2015**, *143* (10), 102809. https://doi.org/10.1063/1.4922649.

(58) Hartman, J. D.; Day, G. M.; Beran, G. J. O. Enhanced NMR Discrimination of Pharmaceutically Relevant Molecular Crystal Forms through Fragment-Based Ab Initio Chemical Shift Predictions. *Cryst. Growth Des.* **2016**, *16* (11), 6479–6493. https://doi.org/10.1021/acs.cgd.6b01157.

(59) 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. NMR Studies of Protonation and Hydrogen Bond States of Internal Aldimines of Pyridoxal 5′-Phosphate Acid–Base in Alanine Racemase, Aspartate Aminotransferase, and Poly- L -Lysine. *J. Am. Chem. Soc.* **2013**, *135* (48), 18160–18175. https://doi.org/10.1021/ja408988z.

(60) Bertsekas, D. P. *Convex Optimization Theory*; Optimization and computation series; Athena Scientific: Belmont, Massachusetts, 2009.

(61) Boyd, S. P.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK ; New York, 2004.

(62) Vandenberghe, L. The CVXOPT Linear and Quadratic Cone Program Solvers. 30.

(63) Cordes, E. H.; Jencks, W. P. Semicarbazone Formation from Pyridoxal, Pyridoxal Phosphate, and Their Schiff Bases. *Biochemistry* **1962**, *1* (5), 773–778. https://doi.org/10.1021/bi00911a007.

(64) Heinert, Dietrich.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. VII. Acid-Base Equilibria of Schiff Bases. *J. Am. Chem. Soc.* **1963**, *85* (2), 188–193. https://doi.org/10.1021/ja00885a018.

(65) Metzler, D. E. Equilibria between Pyridoxal and Amino Acids and Their Imines1. *J. Am. Chem. Soc.* **1957**, *79* (2), 485–490. https://doi.org/10.1021/ja01559a068.

(66) Heinert, Dietrich.; Martell, A. E. **Pyridoxine and Pyridoxal Analogs. VIII. Synthesis and Infrared Spectra of Metal Chelates**. *J. Am. Chem. Soc.* **1963**, *85* (9), 1334–1337. https://doi.org/10.1021/ja00892a026.

(67) Christensen, H. N. Three Schiff Base Types Formed by Amino Acids, Peptides and Proteins with Pyridoxal and Pyridoxal-5-Phosphate1. *J. Am. Chem. Soc.* **1958**, *80* (1), 99–105. https://doi.org/10.1021/ja01534a027.

(68) Crugeiras, J.; Rios, A.; Riveiros, E.; Richard, J. P. Substituent Effects on Electrophilic Catalysis by the Carbonyl Group: Anatomy of the Rate Acceleration for PLP-Catalyzed Deprotonation of Glycine. *J. Am. Chem. Soc.* **2011**, *133* (9), 3173–3183. https://doi.org/10.1021/ja110795m.

(69) Major, D. T.; Gao, J. A Combined Quantum Mechanical and Molecular Mechanical Study of the Reaction Mechanism and α-Amino Acidity in Alanine Racemase. *J. Am. Chem. Soc.* **2006**, *128* (50), 16345–16357. https://doi.org/10.1021/ja066334r.

(70) Major, D. T.; Nam, K.; Gao, J. Transition State Stabilization and α-Amino Carbon Acidity in Alanine Racemase. *J. Am. Chem. Soc.* **2006**, *128* (25), 8114–8115. https://doi.org/10.1021/ja062272t.

(71) Casasnovas, R.; Adrover, M.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Muñoz, F. C–H Activation in Pyridoxal-5′-Phosphate Schiff Bases: The Role of the Imine Nitrogen. A Combined Experimental and Computational Study. *J. Phys. Chem. B* **2012**, *116* (35), 10665–10675. https://doi.org/10.1021/jp303678n.

(72) Sharif, S.; Powell, D. R.; Schagen, D.; Steiner, T.; Toney, M. D.; Fogle, E.; Limbach, H. H. X-Ray Crystallographic Structures of Enamine and Amine Schiff Bases of Pyridoxal and Its 1:1 Hydrogen-Bonded Complexes with Benzoic Acid Derivatives: Evidence for Coupled Inter- and Intramolecular Proton Transfer. *Acta Crystallogr. B* **2006**, *62* (Pt 3), 480–487. https://doi.org/10.1107/S0108768105040590.

(73) Limbach, H.-H.; Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Shenderovich, I. G.; Denisov, G. S.; Toney, M. D. Critical Hydrogen Bonds and Protonation States of Pyridoxal 5′-Phosphate Revealed by NMR. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2011**, *1814* (11), 1426–1437. https://doi.org/10.1016/j.bbapap.2011.06.004.

Chapter 5. Proton Dynamics in the Internal Aldimine State of Tryptophan Synthase via NMR Crystallography and DNP

5.1 Introduction

Pyridoxal-5′-phosphate(PLP) acting as the bioactive form of vitamin B6 is a magnet in the nitrogen metabolism of all organisms.[1] Currently, PLP-dependent enzymes, in which PLP serves as the cofactor, have been identified that it catalyzes over 140 distinct activities and found in all free-living organisms.[2,3] PLP-dependent enzymes contains serval protonatable groups: the heteroaromatic pyridine ring, imine group, the hydroxyl group, and a phosphate group (PLP; Figure 5.1), resulting in different tautomeric forms. One of them comes from the tautomerization between the ketoenamine and enolimine (Figure 5.2) where an intramolecular proton transfer between the imine nitrogen and phenolic oxygen in the active site of the enzymes. Through pyridine ring along with surrounding active site pocket, the dominant tautomer varies,[4] and it is proved to play a large role that affects the reactivity of the PLP Schiff base in different reactions. [5–8]

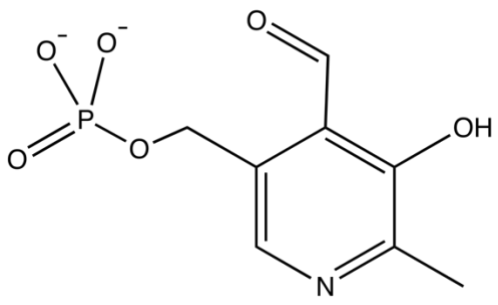


Figure 5.1 Pyridoxal-5'-phosphate.

In some PLP-dependent enzymes such as aspartate aminotransferase (AAT), the protonated Schiff base (PSB) nitrogen, a kind of the ketoenamine, has been demonstrated to boost the enzymatic activity by virtue of forming a reactive intermediate. This intermediate is much more sensitive than the neutral imine

for the nucleophilic attack.[9] On the other hand, electrostatic interactions involving side residues or water may favor the equilibrium shift toward the enolimine side as the case in l-dopa decarboxylase.[10] Furthermore, NMR also shows a extrinsic factor-temperature, has a huge impact on the tautomeric equilibrium of the Schiff base. [11,12]

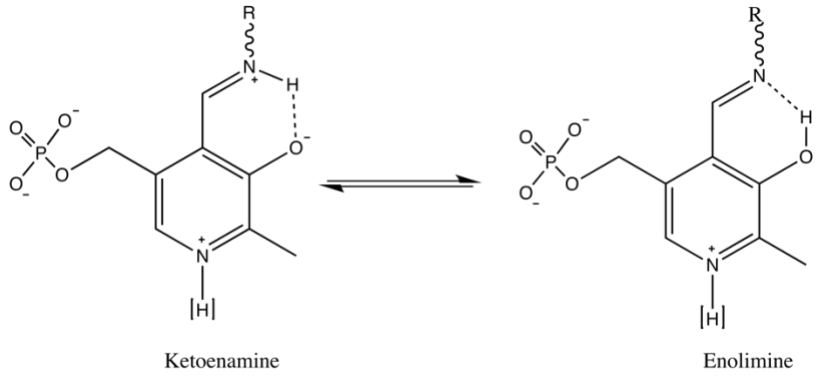


Figure 5.2 Tautomerization between ketoenamine and enolimine.

The tryptophan synthase (TS) is the fold type II PLP-dependent enzyme where β-hydroxyl group of L-serine is eliminated and replaced with indole to produce tryptophan.[13–15] At the start of the catalytic cycle of the TS reaction, as at the initial state of the AAT reaction, PSB is also proposed to be present to assist the reaction going to the next step. Besides, the protonation profiles especially the phenolic oxygen and the pyridine nitrogen in TS in particular are identified to be essential in establishing the specificity of the reaction.[4,16,17] Previously, we reported the NMR chemical shift of the Schiff base nitrogen by conducting solid-state ^{15}N NMR, supporting the PSB hypothesis for the internal aldimine state of TS (Figure 5.3).[18] This was the first direct observation of the PSB via the ^{15}N NMR chemical shifts.

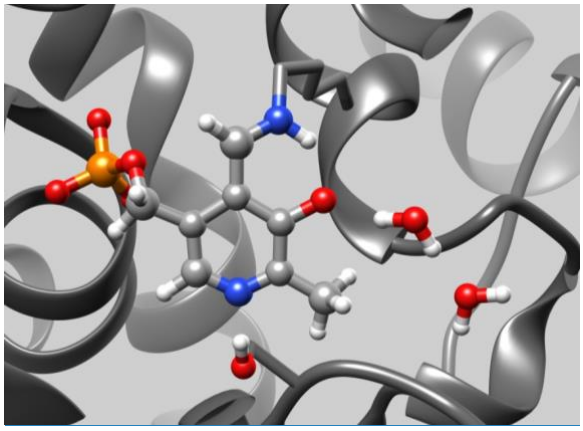


Figure 5.3 Internal aldimine in tryptophan synthase.

Nevertheless, solid-state NMR suffers from low sensitivity due to low polarization and slow polarization recovery times. Thus, in the past two decades, dynamic nuclear polarization (DNP), which transfers spin polarization from electrons to nuclei, has been widely utilized to improve the sensitivity of magic angle spinning (MAS) experiments at moderate fields (B_0 < 9.4 T) and temperatures (T > 90 K). [19–22] Though its signal enhancement makes it ideal for the study of many large and complex systems such as protein systems,[23–25] it also requires very high field conditions, which can be up to B_0 = 21T, to provide better resolution and sensitivity. This is because in one of DNP mechanisms, cross effect, efficiency of DNP is reversely proportional to the external field strength.[26,27] One of the effective solutions of this problem is lowering the sample temperature because electron-to-nuclear polarization transfer is favored by the slower spin relaxation at lower temperatures, and both electron and nuclear spin polarizations at thermal equilibrium are larger at lower temperatures.[20,28] In addition, conducting MAS NMR on colder samples also enable trap some kinetic intermediates and unstable molecules, making the low temperature DNP a promising way to study tautomerization.

Here, we conducted MAS DNP for internal aldimine of tryptophan synthase and showed that the ^{15}N chemical shift of the SBN is perturbed at low-T, indicating that it also is part of a tautomeric exchange. The room T and low T chemical shifts are also confirmed in a first principles analysis.

5.2 Experimental Details

5.2.1 Temperature Dependence of PSB Nitrogen

Figure 5.4 shows the chemical shifts (blue dots) of PSB nitrogen at the different temperatures. At 273K, CP-MAS ^{15}N NMR shows a peak at 202 ppm, while at a low temperature, 105K, DNP ^{15}N NMR shows a peak around 175 ppm, which represents a huge shift from that of the room temperature.

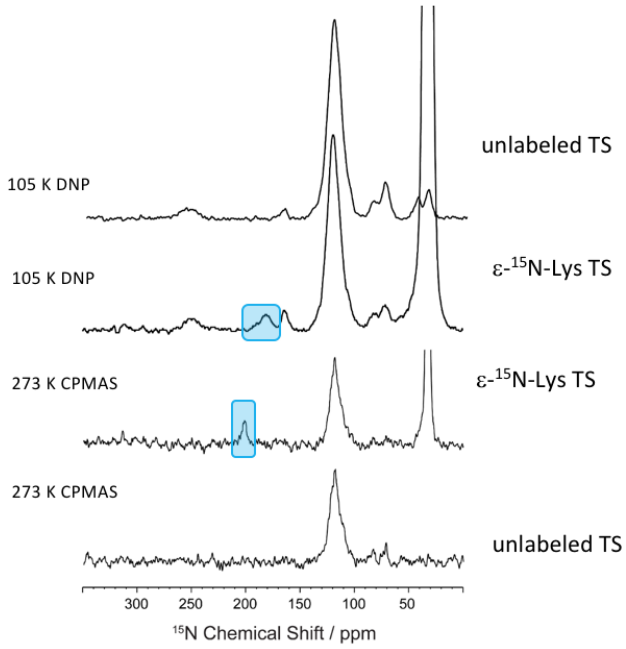


Figure 5.4 15N Lys temperature dependent chemical shifts of the internal aldimine of TS.

5.2.2 First-Principle Calculations

To gain a holistic view of the protonation states and dynamics, we used NMR crystallography, which has determine the protonation profiles of PLP in the carbanionic intermediate state for TS and showed the remarkable results.[29]

Protonation profiles of TS intermediate have been numerously studied involved X-ray diffraction [30–32], computational studies involved MD simulations[33,34], and solid-state NMR[35], each of these methods

has its advantages and limitations. The NMR crystallography — combining X-ray diffractions, solid-state NMR spectroscopies, and the computational chemistry—provides exceptional and holistic insights into three-dimensional and big enzyme systems. Mueller has employed it in the TS enzyme catalysis.[76] Symbolic X-ray crystal structures (1.3 to 2.5 Å resolution) of intermediates that binds have shown the hydrogen-bond interacts between site residues and the substrate, but fail to legitimately recognize the protonation states. NMR supplies direct chemical shifts to certain chosen atoms in enzyme-substrate complexes; however, it requires a framework to explain them. Ab initio calculations and molecular mechanics are capable of building models of enzymatic processes, but it must use specified frameworks. Molecular dynamic simulations can provide an excellent initial guess for molecular' orientations, but they cannot determine the protonation states of the cofactor. However, they are combined to supply the structure and function of enzyme active sites models, which are accordant and testable .[36–39] The previous studies of NMR crystallography have determined the protonation profiles of PLP in carbanionic intermediate states for tryptophan synthase and showed the remarkable results.[40,41]

For the internal aldimine, the first principle calculations was performed using cluster models of the X-ray crystal structure of the active site for internal aldimine of *S. Typhimurium* downloaded from PDB data bank (PDB ID 4HT3).[32] The cluster model was created using PyMol software by truncating the crystallographic structure into the 500–600 atom fragments that include the PLP cofactor and the protein active site within 7 Å distance from the cofactor. A distance of 7 Å provides the optimum balance between cluster size and basis set selection by performing a systematic approach to NMR calculations in proteins. Then, to simplify the modeling work and reduce the required computational resources while maintaining the radius of 7Å, further modifications were made: (i) Residues excluded from the continuous segment of the 7 Å cut and have only one atom within 7 Å were eliminated (β-site residues Val201, Gln205, Ser301, His313). (ii) Similarly, residues with two backbone atoms within 7 Å were replaced with alanine (β-site residues Arg379). (iii) Hydrogen atoms at the C-termini carbonyl carbons of backbone segments were replaced with –NH2 groups to yield amides, while the resulting [NH3] + N-termini of backbone segments

were replaced with a hydrogen atom.(iv) Each Cs+ atom was changed to a Na+ atom to carry out the posterior computation. The final cluster (Figure 5.5) secured a total of approximately 570 atoms.

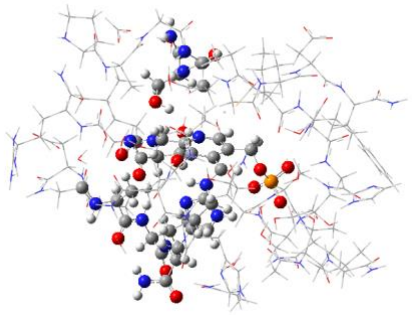


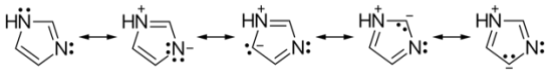
Figure 5.5 Cluster model of the β-subunit active site for first-principles geometry optimization and chemical shift calculations.

Besides, multidimensional NMR techniques have emerged to play a critical role for research of the structures and dynamics of proteins recently. Rasmus's group performed a solid-state NMR 5D experiment in TS, combined with complementary 4D spectra of high resolution, gaining spectral information content for a huge amount of backbone assignments at the absence of long time. The residues within the modeling 7 Å cut, which includes Ala-85, His-86, Lys-87, Glu-350, Ser-351, Ser-377, were selected as modeling active sites as well. The measured backbone chemical shifts for these residues are shown in Table 5.1.

Residue	Amine N	Amine C	C^{α}	C^{β}
Ala-85	117.6	175.3	49.13	21.02
His-86	117.1	173.9	56.0	29.1
Lys-87	118.6	174.7	57.66	31.73
Glu-350	123.5	177.2	59.26	27.22
Ser-351	113.6	175.6	57.76	60.66
Ser-377	116.1	169.9	60.53	60.09

Table 5.1 Reported chemical shifts for backbone atoms of selected residues.

Furthermore, His-86 has a highly polar functional group imidazole. Though a planar 5-membered ring, it is classified as aromatic due to the 6-π electrons plane and exists as a tautomer. Some resonance structures of it are shown below:



Based on the tautomerization of imidazole, Mueser has studied an extended hydrogen bond network in ATT, which is coupled to the pyridinyl nitrogen of the PLP.[43] It showed this network, which contains residues Thr-139, His-189, His-143, Asp-222, and structural waters, has a direct impact to the pyridine nitrogen's protonation state, therefore affecting the rates of catalysis. By comparing the internal aldimine and the external aldimine, the researchers proposed the Grotthüss proton hopping mechanism, in which His-189 is neutral in the internal aldimine but becomes protonated in the external aldimine, is hypothesized to be the spot for a proton shift through the water channel. Given that sigma bonds are defined as having their electron density along the bond axis, rotating the His-86 along the bond between His-86 C^{γ} and His-86 C^{β} doesn't break the bond and makes the flip His-86 possible. Therefore, we also incorporated the chemical shifts of His-86 to model three forms of His-86 (Figure 5.6) as well as three responding flip forms. 6 structures of His-86, including τ, flip τ, π, flip π, cationic and flip cationic, are shown in Figure 5.7.

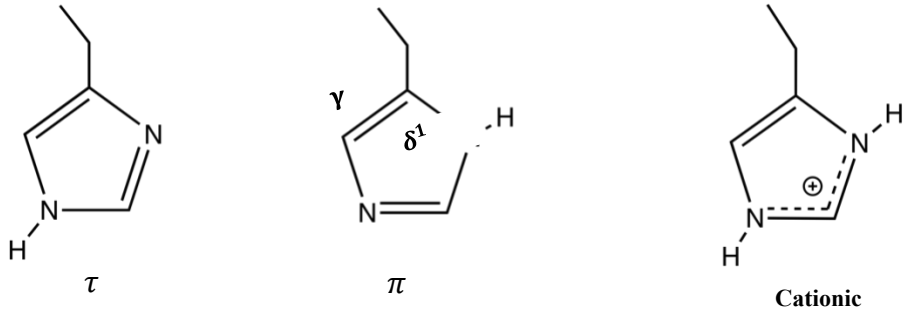
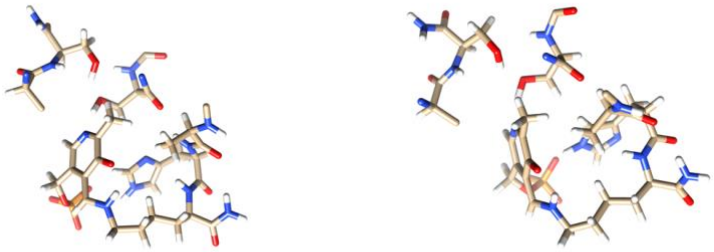
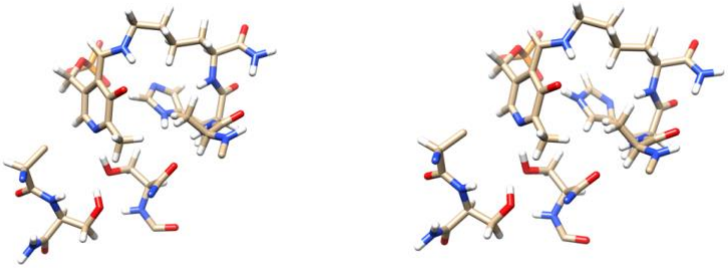


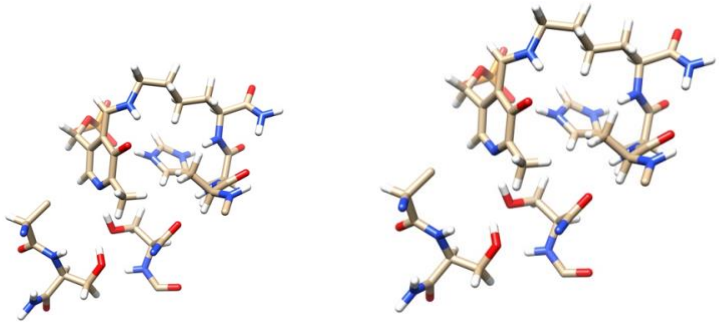
Figure 5.6 Three forms of histidine: (a)τ tautomer, (b)π tautomer, and (c)cationic histidine.



A τ τ-F



B π π-F



C Cat Cat-F

Figure 5.7 Six modeling forms of histidine: (A)τ tautomer and flip τ tautomer, (B)π and flip π , and (C)cationic and flip cationic.

Then, all models were then subjected to geometry optimization via a hybrid density-functional theory (DFT) approach: B3LYP (Refer to the section 2.7 for the theoretical foundation). Basis sets were selected using a two-tier method: PLP/substrate complex was paired with locally dense basis set 6-31G(d,p), while the remaining atoms were paired with 6-31G. Atoms within 4 Å, as well as all hydrogen atoms in the cluster, were allowed to move, whereas the rest were fixed at coordinates provided by the crystallography experiments. Between the geometry optimizations, a restriction that adjustable atoms were allowed to move on the order of 0.2 Å, at a maximum, and less than 0.1 Å RMSD from the initial coordinates was imposed.

All optimized structures were observed for the presence of local minimums and possibilities for alternative orientations of hydrogens. The optimization results were compared to the results of the MD simulations, and the next hydrogen atoms were manually rotated into a new orientation.
Following the optimization, the NMR shielding calculation was carried out on the ground of the gauge-independent atomic orbital (GIAO) method at the DFT level, which is implemented in the Gaussian-09 program. The DFT calculation utilized the B3LYP hybrid functional but, compared with the two-layer basis set in the geometry optimization, with a more accurate three-layer basis set [44]: The high layer, including the PLP/substrate complex and residues whose backbone atoms have chemical shift assignments, was assigned the 6-311+G(2d,p) basis set; the medium layer, comprised of any remaining atom within 4 Å of the center of the high layer, was assigned the 6-311G(d,p) basis set, and all remaining atoms composed of the low layer were given the 6-31G basis set. [45]

The output generated from GIAO calculations offered every atom an absolute magnetic shielding value, yet experimental NMR measurements provide the isotropic chemical shift, δ_{iso} which is a relative value with the referenced material like the adamantane. In order to compare experimental and calculated results, a map from absolute shielding values to referenced chemical shifts had been produced. It was uniquely determined by the theoretical model and basis sets, which had been performed by Hartman et al on a variety of test sets of crystal structures. The conversion is proved to be a linear rescale between isotropic chemical shift and the absolute magnetic shielding:

$$\delta_{iso} = A\sigma_{iso} + B$$

where A and B are two parameters determined from the best fit of all data points[42,46], and are shown on Table 5.2 for ^{13}C, ^{15}N, and ^{17}O

Atom	A	B(ppm)	δ_{iso} RMSE(ppm)
^{15}N	-0.9996	230.45	4.82
^{3}C	-0.9685	173.70	1.49
^{17}O	-1.0612	265.79	8.25

Table 5.2 Parameters values for linear rescale between isotropic chemical shift and the absolute magnetic shielding.

NMR shielding (σ) values of the atoms which are at the three ionizable sites(βLys87 Nε, the pyridine nitrogen and the phenolic oxygen) were converted to chemical shifts (δ) using the following linear rescaling relationships with parameters shown above: [45,47]

$$^{13}C: \delta[TMS(l)] = (-0.9685)\sigma_{calc} + 173.70$$

$$^{15}N: \delta[NH3(l)] = (-0.9996)\sigma_{calc} + 230.45$$

In addition to the values of A and B, root-mean-square errors(also attached to Table 5.2) for every type of isotropic shift is also acquired after applying best-fit rescaling. [45]

Root-mean-square errors were used in reduced χ2 analysis, which is widely used in goodness-of-fit testing, to evaluate every candidate structure's matching between the calculated chemical shifts and experimentally measured ones.[47] The formal expression of reduced χ2 statistic used in the experiment is:

$$x_r^2 = \frac{1}{N - f}\sum_{i=1}^{N}\frac{\left(\delta_i^{model} - \delta_i^{exp}\right)^2}{s_i^2}$$

Where δ_i^{model} and δ_i^{exp} represents the chemical shifts of the calculated and the experimental at site *I* respectively, s_i denotes the deviation at the site *I* expressed as root-mean-square. s_i with the different value corresponds to different varieties of nuclear species, such as, nitrogen, carbon, and oxygen. $N - f$ denotes the degree of freedom derived from statistics, *N* represents the count of all compared δ_i^{model} points, and f is

the number of dependent points in δ_i^{exp} since δ_i^{exp} may be calculated from δ_i^{model}. Here, N is 29, which refers to the number of chemical shifts measured, and f equal to 0 since all δ_i^{exp} were acquired from experiments rather than δ_i^{model}.

5.3 Results and Discussion

Theoretical chemical shift calculations of 18 modeling structures, each of which has 29 positions, are shown in Table 5.3; $\chi 2$ of each single structure is also shown in the last row of the table.

Atom	A0001_Tau_F	A0001_Tau	A0001_Cat_F	A0001_Cat	A0001_Pi_F	A0001_Pi	A0010_Tau_F	A0010_Tau	A0010_Cat	A0010_Cat_F	A0010_Pi_F	A0010_Pi	A0100_Tau	A0100_Tau_F	A0100_Pi	A0100_Cat_F	A0100_Cat	A0100_Pi_F
His86 N	125.0554	127.2102	120.0495	118.0335	124.4872	121.5475	125.0537	127.2018	118.0831	120.1023	124.4295	121.6249	126.822	125.1756	120.9442	120.691	118.8435	123.771
Lys 87 N	116.7173	116.2194	119.7843	119.3174	118.7531	118.1405	116.9151	116.5637	119.5954	119.8673	118.9636	118.0749	117.0357	116.5924	117.8691	122.5318	120.3721	118.1632
Glu350 N	122.8353	123.2102	125.6878	124.795	123.0217	122.7873	122.5638	122.9637	124.3282	125.0294	122.8719	122.6355	127.5048	127.5726	127.4407	126.1016	125.9287	130.0237
Ser351 N	110.0683	110.5165	107.0795	107.4636	110.1864	110.0516	109.7678	110.4345	107.4097	107.3891	110.1575	110.0705	109.126	109.0546	109.6536	109.0052	109.2379	108.727
Ser377 N	116.8915	117.4931	118.2852	117.4136	117.6918	117.3829	116.2386	117.2539	117.2265	117.9654	117.5422	117.2475	115.5506	115.8916	117.0152	116.8517	116.7346	119.9053
SN	169.0966	166.817	172.7264	172.8422	166.5813	168.6437	319.6729	319.2959	323.5546	324.858	314.9811	317.4608	376.9149	381.5672	395.4105	347.3415	367.8266	384.3105
PN	304.3966	304.4376	305.9836	304.1917	300.7044	301.8772	301.7142	302.6631	302.4899	304.8291	297.012	299.9826	189.3198	189.1579	187.9929	191.9168	189.5303	186.4571
C3	170.4883	169.8888	169.9148	170.6485	170.4429	170.7385	155.3769	154.8006	156.4654	155.0854	155.3843	155.977	163.2737	163.616	164.8637	165.3592	165.5423	164.407
C2	158.2823	158.0098	157.7796	158.2383	155.9966	158.3693	149.3369	149.2623	149.8468	148.7686	146.9472	149.4535	152.005	152.5962	149.8853	147.7199	147.3026	146.2765
C2'	19.2556	19.3031	19.1991	19.0622	19.2063	19.2085	18.132	18.2491	17.735	17.9249	18.0575	17.9514	17.6947	17.7421	17.8071	16.3705	16.7077	18.0237
Ala85 carbonyl	177.8734	177.7464	179.8921	179.8306	178.3337	178.2993	177.7364	177.624	179.6499	179.7636	178.2217	178.1532	177.3044	177.2513	178.144	179.4751	179.177	177.5063
Ala85 alpha	51.1397	51.0122	51.1017	51.5403	51.5472	51.5559	51.1035	51.0248	51.4925	51.0133	51.5854	51.5479	51.284	51.0752	51.5521	51.106	51.1749	51.6091
Ala85 beta	22.0843	21.3941	22.9845	22.7124	21.7889	21.959	21.8438	21.1071	22.3867	22.8154	21.4571	21.6836	21.629	23.1047	23.3655	23.8127	23.2783	23.2514
His86 carbonyl	172.2419	174.0017	171.4603	170.0658	173.3505	173.0058	172.2057	173.9727	169.8988	171.3125	173.3011	173.03	173.321	171.3324	173.1204	170.4985	168.9351	173.5982
His86 alpha	57.2913	58.4875	55.7108	56.5898	57.2945	58.5436	57.3304	58.544	56.6675	55.8201	57.3437	58.5878	58.9011	57.9038	58.7692	56.6928	56.9466	57.459
His86 beta	30.1375	27.4523	25.7999	25.6693	25.3671	25.1338	30.1226	27.4184	25.7354	25.9638	25.2536	25.1057	27.2279	30.2295	25.0596	26.5854	25.6012	25.1127
His 86 Gamma	135.994	135.1954	128.3115	130.9062	122.9674	123.3422	136.0429	135.1201	130.8672	128.4968	122.8945	123.2555	134.6945	134.6178	122.4764	127.7042	129.9893	122.1338
His86 delta	112.4068	116.0914	117.197	121.2068	127.0738	130.0789	112.2481	116.1141	121.3217	117.3666	127.036	130.2292	116.2247	112.2222	130.2958	114.4848	121.4179	126.4679

Lys87 alpha	57.1526	57.3231	58.144	57.8594	57.8131	57.334	57.2074	57.3251	57.8667	58.1272	57.8021	57.3827	57.6111	57.4099	57.6784	57.6767	57.6878	58.1588
Lys87 beta	35.0094	34.5529	32.8599	33.2169	33.697	34.5812	35.4805	35.1587	33.7742	33.4752	34.2197	35.1575	34.2608	35.0833	35.615	31.6069	33.8835	34.6032
Glu350 carbonyl	173.124	172.5801	173.3742	173.6944	173.0327	173.2575	173.2861	172.6468	173.7421	173.4903	173.0996	173.2343	171.7683	171.7705	171.5462	171.5259	171.5514	171.6307
Glu350 alpha	57.6053	57.5613	57.3051	57.3627	57.6209	57.6016	57.5826	57.59	57.4572	57.2926	57.61	57.5766	58.1023	58.1942	57.702	58.0705	57.8487	58.2133
Glu350 beta	28.6075	28.5646	28.8849	29.4932	28.8622	29.1277	29.1576	29.0644	30.1678	29.598	29.405	29.6388	29.6473	29.6934	29.7472	29.6017	29.8847	29.2789
Ser351 carbonyl	174.0955	174.148	173.5249	173.5562	174.2592	174.1957	173.7696	174.1118	173.3705	173.5232	174.2329	174.1593	175.3192	175.1134	174.8945	175.3019	175.3666	173.1242
Ser351 alpha	57.4168	57.4569	57.4126	57.274	57.214	57.0163	57.6354	57.4169	57.0816	57.3862	57.1745	57.0014	57.6366	57.3226	57.2845	57.7063	57.8939	56.1648
Ser351 beta	60.771	60.9762	61.3955	61.1785	60.8309	60.6094	60.6623	60.9482	61.0458	61.3238	60.8351	60.6162	61.1182	60.9641	60.9166	60.8146	60.6831	61.1608
Ser377 carbonyl	171.9613	171.6287	173.6723	175.5003	171.9759	173.6595	171.9295	171.6847	175.5464	173.7489	172.0872	173.7108	170.1643	170.5272	171.6447	170.9704	172.7095	170.4125
Ser377 alpha	60.4807	60.0807	60.725	60.2507	60.7982	60.7791	60.2591	59.9954	60.082	60.5051	60.6875	60.7092	59.6407	60.2493	60.6967	60.6598	60.2094	57.5132
Ser377 beta	56.2418	56.1	56.4211	56.434	55.9866	56.8696	56.3369	56.2353	56.6664	56.6363	56.0965	57.0134	59.2238	59.5791	60.1134	59.8592	59.7779	61.6524
*******																		
chi	3.544016786	4.015188699	4.688105 83	5.199357708	9.359793138	10.46842662	30.97797458	31.30730838	34.296 1885	35.125 66551	35.162 48694	36.775 17498	79.999 11398	82.947 12118	100.83 00766	63.372 52483	77.049 40861	93.321 59043

Table 5.3 Calculated isotropic chemical shifts, δiso, in parts per million, ppm, ranked by reduced χ2 value for all 18 candidates. structures.

Figure 5.8 ranked all candidate models with their corresponding reduced $\chi 2$ values. Among all candidate structures, none were found to be below the acceptable threshold between the calculated and experimental ^{13}C and ^{15}N chemical shifts (lowest reduced $\chi 2 = 3.5$).

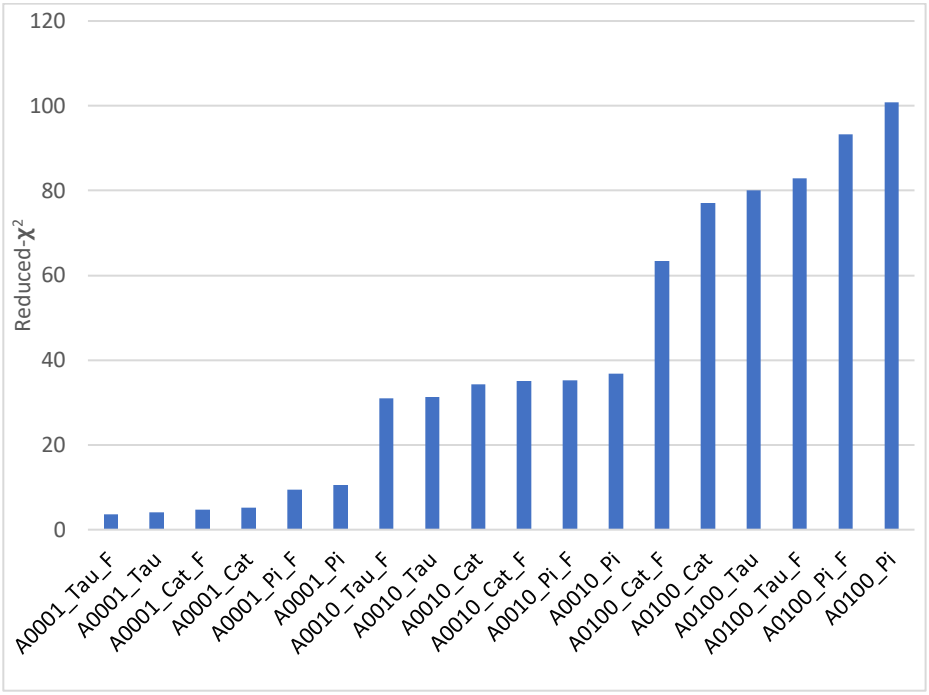


Figure 5.8 Reduced-$\chi 2$ comparing the experimental and first-principles isotropic chemical shifts for 18 geometry-optimized active-site models with varying protonation states.

The temperature dependence experiments, which was conducted on the solid-state NMR to measure the chemical shifts at different temperatures, shows that, as the temperature changing from 273K to 287K, the measured chemical shifts indicates the a proton is hopping away from the phenolic oxygen and toward the PSB nitrogen, implying a fast tautomeric equilibrium between these two species. This phenomenon has been extensively known as the intramolecular hydrogen bond in PLP dependent enzymes. To interpret how this case works, a two-site model has been suggested to interpret the experimental nitrogen chemical shift with the mutual participation of two species. [48] Therefore, the two-site exchange

models, in which only pairs of structures containing the same number of protons and the same histidine form, were considered and weighted to give the lowest reduced χ2 value (Table 5.4).

χ2	**Str1**	**Str2**	**Ratio**
1.25747664	A0001_Tau_F	A0100_Tau_F	('0.8590', '0.1410')
1.40149648	A0001_Tau_F	A0010_Tau_F	('0.7915', '0.2085')
1.4643815	A0001_Tau	A0100_Tau	('0.8498', '0.1502')
1.6579132	A0001_Tau	A0010_Tau	('0.7840', '0.2160')
2.62996517	A0001_Cat_F	A0100_Cat_F	('0.8503', '0.1497')
2.907432	A0001_Cat_F	A0010_Cat_F	('0.8182', '0.1818')
3.20658154	A0001_Cat	A0100_Cat	('0.8657', '0.1343')
3.33018346	A0001_Cat	A0010_Cat	('0.8124', '0.1876')
6.70460714	A0001_Pi_F	A0010_Pi_F	('0.7813', '0.2187')
6.71614366	A0001_Pi_F	A0100_Pi_F	('0.8627', '0.1373')
7.83664156	A0001_Pi	A0010_Pi	('0.7866', '0.2134')
7.85944473	A0001_Pi	A0100_Pi	('0.8699', '0.1301')
28.9794238	A0010_Tau_F	A0100_Tau_F	('1.0000', '0.0000')
29.2875005	A0010_Tau	A0100_Tau	('1.0000', '0.0000')
32.083752	A0010_Cat	A0100_Cat	('1.0000', '0.0000')
32.8599109	A0010_Cat_F	A0100_Cat_F	('0.9996', '0.0004')
32.8940383	A0010_Pi_F	A0100_Pi_F	('1.0000', '0.0000')
34.402786	A0010_Pi	A0100_Pi	('1.0000', '0.0000')

Table 5.4 Reduced-χ2 comparing the experimental and first-principles isotropic chemical shifts for fast-exchange equilibrium models.

Based on variable temperature experiments in solid NMR, a substantial chemical shift for Schiff-base nitrogen at different temperatures suggests that this site is likely to involve in chemical exchange. On

the other hand, the absence of significant temperature dependence on pyridine nitrogen suggests this site is not involved in the chemical exchange. Therefore, the model with 0-0-01 and 0-1-00 can be ruled out, leaving out the equilibriums between 0-0-01 and 0-0-10, which echoes the PSB hypothesis. The two best-fits, both are 0-0-01 and 0-0-10 equilibriums but with τ form and flip τ form of His-86 separately, are shown in Table 5.4. The major species is the PSB nitrogen (0-0-01), which was also shown as the best single modeling form with the least chi-square value, and the minor species, 0-0-10, indicates the possibility of a transfer of a proton from the PSB nitrogen to the phenolic oxygen. Incapability to rule out neither form of τ form nor flip τ form indicates there may exist the combination of these two forms of His-86 in the equilibrium of 0-0-01 and 0-0-10.

The PSB hypothesis was proposed to boost reactivity toward nucleophiles at the Schiff base carbon compared with the neutral [49–51] X-ray crystallography[52], UV/vis spectroscopy [53,54] and NMR spectroscopy[35] suggest a PSB form. Here the consistence between MAS-DNP in lower temperature and first principle calculation revealed that internal aldimine tautomer equilibrium is almost completely dominated by PSB tautomer in the lower temperature(105k) and undergoing fast proton exchange between its Schiff base and phenolic form in the room temperature(273K), agreeing with the PSB hypothesis which proposed the PSB activates the nucleophilic attack at C4` of the pyridine.

The PLP catalyzes a broad variety of amino acid transformations, such as, α/β/γ elimination/replacement, racemization, transamination, and decarboxylation.[55,56]

The first and common step in PLP-catalyzed reactions is the formation of an internal aldimine intermediate in which aldehyde group of PLP forms imine group with lysine residue. In the next step, the substrate makes a nucleophilic attack at C4`, replacing lysine residue and forming imine between PLP and substrate amino group (external aldimine).After the external aldimine is formed, the deprotonation which is commonly carried out by the ε-amino group, resulting in the formation of a carbanionic intermediate. [57,58] The carbanionic intermediate can undertake several different subsequent reactions, leading to different reaction paths among different fold types of PLP-dependent enzymes. [59,60] Between two tautomeric form, factors that PSB is favored than protonated phenolic oxygen, can depend upon the substitutable groups on the Schiff base nitrogen, solute−solvent interactions, the protonation states of the pyridine nitrogen.[61–63] In

AAT, which is the representative of fold-type I PLP-dependent enzymes, X-ray crystallography and ^{15}N solid-state NMR both conformed Schiff base nitrogen is protonated and postulated that the protonated pyridine nitrogen acts as a prerequisite for catalytic activity. On the contrary, Alanine Racemase, one of the representatives of fold type III PLP-dependent, behaves in a very different way. Further work proposed that the PSB acts as the catalytically reactive tautomer without protonation of the pyridine nitrogen, and it is possible to shift protonation to the Schiff base form even if the phenolic oxygen is stabilized by hydrogen bonds with the second aspartic acid or waters. With the techniques we developed here, a variety of PLP-dependent enzymes for other fold types with different routes are able to be studied.

5.4 References

(1) György, P. Vitamin B 2 and the Pellagra-like Dermatitis in Rats. *Nature* **1934**, *133* (3361), 498–499. https://doi.org/10.1038/133498a0.

(2) Jansonius, J. N. Structure, Evolution and Action of Vitamin B6-Dependent Enzymes. *Curr. Opin. Struct. Biol.* **1998**, *8* (6), 759–769. https://doi.org/10.1016/S0959-440X(98)80096-1.

(3) Toney, M. D. Reaction Specificity in Pyridoxal Phosphate Enzymes. *Arch. Biochem. Biophys.* **2005**, *433* (1), 279–287. https://doi.org/10.1016/j.abb.2004.09.037.

(4) Metzler, D. E. Equilibria between Pyridoxal and Amino Acids and Their Imines1. *J. Am. Chem. Soc.* **1957**, *79* (2), 485–490. https://doi.org/10.1021/ja01559a068.

(5) Dutta Banik, S.; Chandra, A. A Hybrid QM/MM Simulation Study of Intramolecular Proton Transfer in the Pyridoxal 5′-Phosphate in the Active Site of Transaminase: Influence of Active Site Interaction on Proton Transfer. *J. Phys. Chem. B* **2014**, *118* (38), 11077–11089. https://doi.org/10.1021/jp506196m.

(6) Karsten, W. E.; Ohshiro, T.; Izumi, Y.; Cook, P. F. Initial Velocity, Spectral, and PH Studies of the Serine-Glyoxylate Aminotransferase from Hyphomicrobiuim Methylovorum. *Arch. Biochem. Biophys.* **2001**, *388* (2), 267–275. https://doi.org/10.1006/abbi.2001.2294.

(7) Internal Proton Transfer in the External Pyridoxal 5′-Phosphate Schiff Base in Dopa Decarboxylase | Biochemistry https://pubs.acs.org/doi/10.1021/bi901790e.

(8) Lin, Y.-L.; Gao, J.; Rubinstein, A.; Major, D. T. Molecular Dynamics Simulations of the Intramolecular Proton Transfer and Carbanion Stabilization in the Pyridoxal 5′-Phosphate Dependent Enzymes l-Dopa Decarboxylase and Alanine Racemase. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2011**, *1814* (11), 1438–1446. https://doi.org/10.1016/j.bbapap.2011.05.002.

(9) Dutta Banik, S.; Chandra, A. A Hybrid QM/MM Simulation Study of Intramolecular Proton Transfer in the Pyridoxal 5′-Phosphate in the Active Site of Transaminase: Influence of Active Site Interaction on Proton Transfer. *J. Phys. Chem. B* **2014**, *118* (38), 11077–11089. https://doi.org/10.1021/jp506196m.

(10) Lin, Y.; Gao, J. Internal Proton Transfer in the External Pyridoxal 5′-Phosphate Schiff Base in Dopa Decarboxylase. *Biochemistry* **2010**, *49* (1), 84–94. https://doi.org/10.1021/bi901790e.

(11) Sharif, S.; Denisov, G. S.; Toney, M. D.; Limbach, H.-H. NMR Studies of Solvent-Assisted Proton Transfer in a Biologically Relevant Schiff Base: Toward a Distinction of Geometric and Equilibrium H-Bond Isotope Effects. *J. Am. Chem. Soc.* **2006**, *128* (10), 3375–3387. https://doi.org/10.1021/ja056251v.

(12) Kolehmainen, E.; Ośmiałowski, B.; M. Krygowski, T.; Kauppinen, R.; Nissinen, M.; Gawinecki, R. Substituent and Temperature Controlled Tautomerism: Multinuclear Magnetic Resonance, X-Ray, and Theoretical Studies on 2-Phenacylquinolines. *J. Chem. Soc. Perkin Trans. 2* **2000**, *0* (6), 1259–1266. https://doi.org/10.1039/A908874A.

(13) Yanofsky, C.; Crawford, I. P. 1 Tryptophan Synthetase. In *The Enzymes*; Boyer, P. D., Ed.; Academic Press, 1972; Vol. 7, pp 1–31. https://doi.org/10.1016/S1874-6047(08)60445-X.

(14) Miles, E. W.; Bauerle, R.; Ahmed, S. A. [49] Tryptophan Synthase from Escherichia Coli and Salmonella Typhimurium. In *Methods in Enzymology*; Metabolism of Aromatic Amino Acids and Amines; Academic Press, 1987; Vol. 142, pp 398–414. https://doi.org/10.1016/S0076-6879(87)42051-X.

(15) Dunn, M. F.; Niks, D.; Ngo, H.; Barends, T. R. M.; Schlichting, I. Tryptophan Synthase: The Workings of a Channeling Nanomachine. *Trends Biochem. Sci.* **2008**, *33* (6), 254–264. https://doi.org/10.1016/j.tibs.2008.04.008.

(16) Heinert, Dietrich.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. VII. Acid-Base Equilibria of Schiff Bases. *J. Am. Chem. Soc.* **1963**, *85* (2), 188–193. https://doi.org/10.1021/ja00885a018.

(17) Semicarbazone Formation from Pyridoxal, Pyridoxal Phosphate, and Their Schiff Bases | Biochemistry https://pubs.acs.org/doi/10.1021/bi00911a007 (accessed Aug 22, 2020).

(18) Caulkins, B. G.; Bastin, B.; Yang, C.; Neubauer, T. J.; Young, R. P.; Hilario, E.; Huang, Y. M.; Chang, C. A.; Fan, L.; Dunn, M. F.; Marsella, M. J.; Mueller, L. J. Protonation States of the Tryptophan Synthase Internal Aldimine Active Site from Solid-State NMR Spectroscopy: Direct Observation of the PSBLinkage to Pyridoxal-5′-Phosphate. *J. Am. Chem. Soc.* **2014**, *136* (37), 12824–12827. https://doi.org/10.1021/ja506267d.

(19) Becerra, L. R.; Gerfen, G. J.; Temkin, R. J.; Singel, D. J.; Griffin, R. G. Dynamic Nuclear Polarization with a Cyclotron Resonance Maser at 5 T. *Phys. Rev. Lett.* **1993**, *71* (21), 3561–3564. https://doi.org/10.1103/PhysRevLett.71.3561.

(20) Hall, D. A.; Maus, D. C.; Gerfen, G. J.; Inati, S. J.; Becerra, L. R.; Dahlquist, F. W.; Griffin, R. G. Polarization-Enhanced NMR Spectroscopy of Biomolecules in Frozen Solution. *Science* **1997**, *276* (5314), 930–932. https://doi.org/10.1126/science.276.5314.930.

(21) Maly, T.; Debelouchina, G. T.; Bajaj, V. S.; Hu, K.-N.; Joo, C.-G.; Mak–Jurkauskas, M. L.; Sirigiri, J. R.; van der Wel, P. C. A.; Herzfeld, J.; Temkin, R. J.; Griffin, R. G. Dynamic Nuclear Polarization at High Magnetic Fields. *J. Chem. Phys.* **2008**, *128* (5), 052211. https://doi.org/10.1063/1.2833582.

(22) Loening, N. M.; Rosay, M.; Weis, V.; Griffin, R. G. Solution-State Dynamic Nuclear Polarization at High Magnetic Field. *J. Am. Chem. Soc.* **2002**, *124* (30), 8808–8809. https://doi.org/10.1021/ja026660g.

(23) Rosay, M.; Lansing, J. C.; Haddad, K. C.; Bachovchin, W. W.; Herzfeld, J.; Temkin, R. J.; Griffin, R. G. High-Frequency Dynamic Nuclear Polarization in MAS Spectra of Membrane and Soluble Proteins. *J. Am. Chem. Soc.* **2003**, *125* (45), 13626–13627. https://doi.org/10.1021/ja036898k.

(24) Goncalves, J. A.; Ahuja, S.; Erfani, S.; Eilers, M.; Smith, S. O. Structure and Function of G Protein-Coupled Receptors Using NMR Spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57* (2), 159–180. https://doi.org/10.1016/j.pnmrs.2010.04.004.

(25) Mak-Jurkauskas, M. L.; Bajaj, V. S.; Hornstein, M. K.; Belenky, M.; Griffin, R. G.; Herzfeld, J. Energy Transformations Early in the Bacteriorhodopsin Photocycle Revealed by DNP-Enhanced Solid-State NMR. *Proc. Natl. Acad. Sci.* **2008**, *105* (3), 883–888. https://doi.org/10.1073/pnas.0706156105.

(26) Maly, T.; Debelouchina, G. T.; Bajaj, V. S.; Hu, K.-N.; Joo, C.-G.; Mak–Jurkauskas, M. L.; Sirigiri, J. R.; van der Wel, P. C. A.; Herzfeld, J.; Temkin, R. J.; Griffin, R. G. Dynamic Nuclear Polarization at High Magnetic Fields. *J. Chem. Phys.* **2008**, *128* (5), 052211. https://doi.org/10.1063/1.2833582.

(27) Barnes, A. B.; Mak-Jurkauskas, M. L.; Matsuki, Y.; Bajaj, V. S.; van der Wel, P. C. A.; DeRocher, R.; Bryant, J.; Sirigiri, J. R.; Temkin, R. J.; Lugtenburg, J.; Herzfeld, J.; Griffin, R. G. Cryogenic Sample Exchange NMR Probe for Magic Angle Spinning Dynamic Nuclear Polarization. *J. Magn. Reson.* **2009**, *198* (2), 261–270. https://doi.org/10.1016/j.jmr.2009.03.003.

(28) Thurber, K. R.; Yau, W.-M.; Tycko, R. Low-Temperature Dynamic Nuclear Polarization at 9.4T with a 30mW Microwave Source. *J. Magn. Reson.* **2010**, *204* (2), 303–313. https://doi.org/10.1016/j.jmr.2010.03.016.

(29) Bg, C.; Rp, Y.; Ra, K.; C, Y.; Tj, B.; B, B.; E, H.; L, F.; Mj, M.; Mf, D.; Lj, M. NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. *J. Am. Chem. Soc.* **2016**, *138* (46), 15214–15226. https://doi.org/10.1021/jacs.6b08937.

(30) Snell, E. E.; Jenkins, W. T. The Mechanism of the Transamination Reaction. *J. Cell. Comp. Physiol.* **1959**, *54* (S1), 161–177. https://doi.org/10.1002/jcp.1030540413.

(31) Burkhard, P.; Jagannatha Rao, G. S.; Hohenester, E.; Schnackerz, K. D.; Cook, P. F.; Jansonius, J. N. Three-Dimensional Structure of O-Acetylserine Sulfhydrylase from Salmonella Typhimurium11Edited by R. Huber. *J. Mol. Biol.* **1998**, *283* (1), 121–133. https://doi.org/10.1006/jmbi.1998.2037.

(32) Niks, D.; Hilario, E.; Dierkers, A.; Ngo, H.; Borchardt, D.; Neubauer, T. J.; Fan, L.; Mueller, L. J.; Dunn, M. F. Allostery and Substrate Channeling in the Tryptophan Synthase Bienzyme Complex: Evidence for Two Subunit Conformations and Four Quaternary States. *Biochemistry* **2013**, *52* (37), 6396–6411. https://doi.org/10.1021/bi400795e.

(33) Protonation states and catalysis: Molecular dynamics studies of intermediates in tryptophan synthase https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4815307/ (accessed Dec 11, 2019).

(34) Investigation of Structural Dynamics of Enzymes and Protonation States of Substrates Using Computational Tools https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5119520/ (accessed Dec 11, 2019).

(35) Caulkins, B. G.; Bastin, B.; Yang, C.; Neubauer, T. J.; Young, R. P.; Hilario, E.; Huang, Y. M.; Chang, C. A.; Fan, L.; Dunn, M. F.; Marsella, M. J.; Mueller, L. J. Protonation States of the Tryptophan Synthase Internal Aldimine Active Site from Solid-State NMR Spectroscopy: Direct Observation of the PSBLinkage to Pyridoxal-5'-Phosphate. *J. Am. Chem. Soc.* **2014**, *136* (37), 12824–12827. https://doi.org/10.1021/ja506267d.

(36) Harris, R. K.; Hodgkinson, P.; Pickard, C. J.; Yates, J. R.; Zorin, V. Chemical Shift Computations on a Crystallographic Basis: Some Reflections and Comments. *Magn. Reson. Chem.* **2007**, *45* (S1), S174–S186. https://doi.org/10.1002/mrc.2132.

(37) Elena, B.; Pintacuda, G.; Mifsud, N.; Emsley, L. Molecular Structure Determination in Powders by NMR Crystallography from Proton Spin Diffusion. *J. Am. Chem. Soc.* **2006**, *128* (29), 9555–9560. https://doi.org/10.1021/ja062353p.

(38) Mueller, L. J.; Dunn, M. F. NMR Crystallography of Enzyme Active Sites: Probing Chemically Detailed, Three-Dimensional Structure in Tryptophan Synthase. *Acc. Chem. Res.* **2013**, *46* (9), 2008–2017. https://doi.org/10.1021/ar3003333.

(39) Elena, B.; Pintacuda, G.; Mifsud, N.; Emsley, L. Molecular Structure Determination in Powders by NMR Crystallography from Proton Spin Diffusion. *J. Am. Chem. Soc.* **2006**, *128* (29), 9555–9560. https://doi.org/10.1021/ja062353p.

(40) Lai, J.; Niks, D.; Wang, Y.; Domratcheva, T.; Barends, T. R. M.; Schwarz, F.; Olsen, R. A.; Elliott, D. W.; Fatmi, M. Q.; Chang, C. A.; Schlichting, I.; Dunn, M. F.; Mueller, L. J. X-Ray and NMR Crystallography in an Enzyme Active Site: The Indoline Quinonoid Intermediate in Tryptophan Synthase. *J. Am. Chem. Soc.* **2011**, *133* (1), 4–7. https://doi.org/10.1021/ja106555c.

(41) Caulkins, B. G.; Young, R. P.; Kudla, R. A.; Yang, C.; Bittbauer, T. J.; Bastin, B.; Hilario, E.; Fan, L.; Marsella, M. J.; Dunn, M. F.; Mueller, L. J. NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. *J. Am. Chem. Soc.* **2016**, *138* (46), 15214–15226. https://doi.org/10.1021/jacs.6b08937.

(42) Hartman, J. D.; Neubauer, T. J.; Caulkins, B. G.; Mueller, L. J.; Beran, G. J. O. Converging Nuclear Magnetic Shielding Calculations with Respect to Basis and System Size in Protein Systems. *J. Biomol. NMR* **2015**, *62* (3), 327–340. https://doi.org/10.1007/s10858-015-9947-2.

(43) Dajnowicz, S.; Parks, J. M.; Hu, X.; Gesler, K.; Kovalevsky, A. Y.; Mueser, T. C. Direct Evidence That an Extended Hydrogen-Bonding Network Influences Activation of Pyridoxal 5'-Phosphate in Aspartate Aminotransferase. *J. Biol. Chem.* **2017**, *292* (14), 5970–5980. https://doi.org/10.1074/jbc.M116.774588.

(44) Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields | The Journal of Physical Chemistry https://pubs.acs.org/doi/10.1021/j100096a001.

(45) Hartman, J. D.; Kudla, R. A.; Day, G. M.; Mueller, L. J.; Beran, G. J. O. Benchmark Fragment-Based 1H, 13C, 15N and 17O Chemical Shift Predictions in Molecular Crystals. *Phys. Chem. Chem. Phys.* **2016**, *18* (31), 21686–21709. https://doi.org/10.1039/C6CP01831A.

(46) Hartman, J. D.; Monaco, S.; Schatschneider, B.; Beran, G. J. O. Fragment-Based 13C Nuclear Magnetic Resonance Chemical Shift Predictions in Molecular Crystals: An Alternative to Planewave Methods. *J. Chem. Phys.* **2015**, *143* (10), 102809. https://doi.org/10.1063/1.4922649.

(47) Hartman, J. D.; Day, G. M.; Beran, G. J. O. Enhanced NMR Discrimination of Pharmaceutically Relevant Molecular Crystal Forms through Fragment-Based Ab Initio Chemical Shift Predictions. *Cryst. Growth Des.* **2016**, *16* (11), 6479–6493. https://doi.org/10.1021/acs.cgd.6b01157.

(48) 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. NMR Studies of Protonation and Hydrogen Bond States of Internal Aldimines of Pyridoxal 5′-Phosphate Acid–Base in Alanine Racemase, Aspartate Aminotransferase, and Poly- L -Lysine. *J. Am. Chem. Soc.* **2013**, *135* (48), 18160–18175. https://doi.org/10.1021/ja408988z.

(49) Cordes, E. H.; Jencks, W. P. Semicarbazone Formation from Pyridoxal, Pyridoxal Phosphate, and Their Schiff Bases. *Biochemistry* **1962**, *1* (5), 773–778. https://doi.org/10.1021/bi00911a007.

(50) Heinert, Dietrich.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. VII. Acid-Base Equilibria of Schiff Bases. *J. Am. Chem. Soc.* **1963**, *85* (2), 188–193. https://doi.org/10.1021/ja00885a018.

(51) Metzler, D. E. Equilibria between Pyridoxal and Amino Acids and Their Imines1. *J. Am. Chem. Soc.* **1957**, *79* (2), 485–490. https://doi.org/10.1021/ja01559a068.

(52) Hyde, C. C.; Ahmed, S. A.; Padlan, E. A.; Miles, E. W.; Davies, D. R. Three-Dimensional Structure of the Tryptophan Synthase Alpha 2 Beta 2 Multienzyme Complex from Salmonella Typhimurium. *J. Biol. Chem.* **1988**, *263* (33), 17857–17871.

(53) Heinert, Dietrich.; Martell, A. E. Pyridoxine and Pyridoxal Analogs. VIII. Synthesis and Infrared Spectra of Metal Chelates. *J. Am. Chem. Soc.* **1963**, *85* (9), 1334–1337. https://doi.org/10.1021/ja00892a026.

(54) Christensen, H. N. Three Schiff Base Types Formed by Amino Acids, Peptides and Proteins with Pyridoxal and Pyridoxal-5-Phosphate1. *J. Am. Chem. Soc.* **1958**, *80* (1), 99–105. https://doi.org/10.1021/ja01534a027.

(55) Metzler, D. E.; Ikawa, M.; Snell, E. E. A General Mechanism for Vitamin B6-Catalyzed Reactions1. *J. Am. Chem. Soc.* **1954**, *76* (3), 648–652. https://doi.org/10.1021/ja01632a004.

(56) Toney, M. D. Controlling Reaction Specificity in Pyridoxal Phosphate Enzymes. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2011**, *1814* (11), 1407–1418. https://doi.org/10.1016/j.bbapap.2011.05.019.

(57) Caulkins, B. G.; Young, R. P.; Kudla, R. A.; Yang, C.; Bittbauer, T. J.; Bastin, B.; Hilario, E.; Fan, L.; Marsella, M. J.; Dunn, M. F.; Mueller, L. J. NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. *J. Am. Chem. Soc.* **2016**, *138* (46), 15214–15226. https://doi.org/10.1021/jacs.6b08937.

(58) Frontiers | Current Advances on Structure-Function Relationships of Pyridoxal 5′-Phosphate-Dependent Enzymes | Molecular Biosciences https://www.frontiersin.org/articles/10.3389/fmolb.2019.00004/full.

(59) Toney, M. D. Reaction Specificity in Pyridoxal Phosphate Enzymes. *Arch. Biochem. Biophys.* **2005**, *433* (1), 279–287. https://doi.org/10.1016/j.abb.2004.09.037.

(60) Sharif, S.; Huot, M. C.; Tolstoy, P. M.; Toney, M. D.; Jonsson, K. H. M.; Limbach, H.-H. 15N Nuclear Magnetic Resonance Studies of Acid−Base Properties of Pyridoxal-5‘-Phosphate Aldimines in Aqueous Solution. *J. Phys. Chem. B* **2007**, *111* (15), 3869–3876. https://doi.org/10.1021/jp067334g.

(61) Limbach, H.-H.; Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Shenderovich, I. G.; Denisov, G. S.; Toney, M. D. Critical Hydrogen Bonds and Protonation States of Pyridoxal 5’-Phosphate Revealed by NMR. *Biochim. Biophys. Acta* **2011**, *1814* (11), 1426–1437. https://doi.org/10.1016/j.bbapap.2011.06.004.

(62) Sharif, S.; Schagen, D.; Toney, M. D.; Limbach, H.-H. Coupling of Functional Hydrogen Bonds in Pyridoxal-5‘-Phosphate−Enzyme Model Systems Observed by Solid-State NMR Spectroscopy. *J. Am. Chem. Soc.* **2007**, *129* (14), 4440–4455. https://doi.org/10.1021/ja066240h.

(63) Sharif, S.; Denisov, G. S.; Toney, M. D.; Limbach, H.-H. NMR Studies of Coupled Low- and High-Barrier Hydrogen Bonds in Pyridoxal-5‘-Phosphate Model Systems in Polar Solution. *J. Am. Chem. Soc.* **2007**, *129* (19), 6313–6327. https://doi.org/10.1021/ja070296+.

Chapter 6. Concluding Remarks

In this dissertation, the application of NMR-Crystallography techniques – the synergistic combination of X-ray crystallography, NMR spectroscopy, and Ab initio computational approach– to enzymatic systems has been deeply explored. It provides tryptophan synthase exceptionally definite and chemically-detailed structures of the chemical active sites by focusing on the Internal aldimine intermediates of tryptophan synthase along the pyridoxal-5'-phosphate catalyzed reaction pathway.

In the solid-state NMR experiment, dynamic nuclear polarization assists to enhance sensitivity, building models for histidine-86 tautomers and proving dynamics of TS, the multidimensional NMR techniques have been performed in TS, gaining spectral information content for a huge amount of backbone assignments at the absence of long time.

Ab initio calculations has been mainly used throughout the research. This is addressed in particular by the ketoenamine and enolimine tautomerization, in which a proton transfer points to the importance of protonation/deprotonation at ionizable sites on the coenzyme, substrates, and side residues to activate key steps in the catalytic process.