Title

Understanding Structure and Chemistry in Enzyme Active Sites Through the Lens of NMR Crystallography

Author

Holmes, Jacob

Publication Date

2023

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Understanding Structure and Chemistry in Enzyme Active Sites Through the Lens of NMR Crystallography

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

Doctor of Philosophy

in

Chemistry

by

Jacob Brian Holmes

March 2023

Dissertation Committee:
Dr. Leonard J. Mueller, Chairperson
Dr. Gregory J. O. Beran
Dr. Christopher J. Bardeen

The Dissertation of Jacob Brian Holmes is approved:

____

____

____
Committee Chairperson

University of California, Riverside

Acknowledgments

There are many people I want to thank for more reasons than I can express here. I thank my family who supported me in this endeavor and kept me looking forward. I want to thank the Heisinger’s that gave me a friendship I could never forget. I thank all the lab mates I had, there was much time filled with fun and laughter. I give a special thanks to Rittik who is an exceptional scientist and the source of much knowledge. He can always be turned to when questions need answers.

Most of all I want to thank Len. He inspired me to be the best scientist I could be and he always encourage me to pursue my scientific interests.

The work in this thesis comes entirely or in part from the following publications

[1] **Holmes, J. B.,** Liu, V., Caulkins, B. G., Hilario, E., Ghosh, R. K., Drago, V. N., Young, R. P., Romero, J. A., Gill, A. D., Bogie, P. M., Paulino, J., Wang, X., Riviere, G., Bosken, Y. K., Struppe, J., Hassan, A., Guidoulianov, J., Perrone, B., Mentink-Vigier, F., Chang, C. A., Long, R. J., Hooley, H. J.,Mueser, T. C., Dunn M. F., Mueller, L. J. **(2022)**. Imaging active site chemistry and protonation states NMR crystallography of the tryptophan synthase α-aminoacrylate intermediate. Proceedings of the National Academy of Sciences of the United States of America, 119(2), e2109235119. https://doi.org/10.1073/pnas.2109235119

[2] Klein, A., Rovó, P., Sakhrani, V. V., Wang, Y., **Holmes, J. B**., Liu, V., Skowronek, P., Kukuk, L., Vasa, S. K., Güntert, P., Mueller, L. J., & Linser, R. **(2022)**. Atomic-resolution chemical characterization of (2x)72-kDa tryptophan synthase via four- and five-dimensional 1H-detected solid-state NMR. Proceedings of the National Academy of Sciences of the United States of America, 119(4), e2114690119. https://doi.org/10.1073/pnas.2114690119

[3] Ghosh, R. K., Hilario, E., Liu, V., Wang, Y., Niks, D., **Holmes, J. B**., Sakhrani, V. V., Mueller, L. J., & Dunn, M. F. **(2021)**. Mutation of βGln114 to Ala Alters the Stabilities of Allosteric States in Tryptophan Synthase Catalysis. Biochemistry, 60(42), 3173–3186. https://doi.org/10.1021/acs.biochem.1c00383

Dedication

To my sister, Jenny. A constant source of joy.

ABSTRACT OF THE DISSERTATION

Understanding Structure and Chemistry in Enzyme Active Sites Through the Lens of NMR Crystallography

by

Jacob Brian Holmes

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, March 2023
Dr. Leonard J. Mueller, Chairperson

NMR-assisted crystallography is the integrated use of solid-state NMR, X-ray crystallography, and first-principles calculations to define the protonation states in an enzyme active site to give a chemically detailed three-dimensional structure. Determining hydrogen positions is a notable feature of the technique, complementing neutron diffraction but here NMR-assisted crystallography applies to catalytically active samples. This integrated approach was applied to the intermediates of Tryptophan synthase and Toho-1-β-lactamase. To determine an experimental structure, an unbiased set of protonation states are modeled and their predicted chemical shifts are compared to experimental shifts using the reduced-χ^2 statistics. The confidence in the experimental structure can be quantified using Bayesian probability. From the structure the positional uncertainties are determined for the shift-rich region as well as the surrounding surface. NMR crystallography gives a remarkably clear picture of the chemistry of the β-subunit active site in tryptophan synthase. This level of detail reveals why BZI, an indole analog, does not proceed with the bond formation step despite being the better nucleophile: BZI is held in the wrong orientation by hydrogen bonds to the charged βLys87 and βGlu109 residues. This chemically-detailed view also reveals water positioned for nucleophilic attack on C^{β} of the substrate.

Based on its position and alignment, we posit that it sits in the initial binding pocket for the β-hydroxyl leaving group.

Table of Contents

List of Figures

Schemes

List of Tables

Broad Overview of this Thesis: Applying NMR to PLP-dependent Enzymes

NMR is a useful technique because it provides chemical information of a molecule or system which is sensitive to the local environment. This can be applied to biomolecules for investigating allostery and reaction mechanisms and provides direct information about each residue in these large systems. The enzyme system primarily studied in this work is tryptophan synthase (TS), a pyridoxial-5' phosphate (PLP) dependent enzyme, which is involved in the last two step of the biosynthesis of L-tryptophan. We are targeting the β-site reaction which involves the PLP cofactor as a model for or PLP-dependent system. Specifically, we choose PLP because it is involved in a diverse array of reactions including transamination, lyases, hydrolases, isomerases, and oxidoreductase. Interestingly, each PLP-dependent enzyme is reaction specific and this specificity is well controlled despite using the same cofactor. The Dunathan hypothesis proposes this control is regulated by stereoelectronic affects which places the group to be eliminated perpendicular to the PLP ring but it has also been observed that protonation states are not uniform across systems. From X-ray and NMR, the pyridine nitrogen of aspartate aminotransferase is protonated with a interaction to an aspartate while in TS the pyridine nitrogen is deprotonated and interacts strongly with serine. It is not fully understood how these difference in protonation states affect reaction specificity. There are five potential protonation sites common to all PLP enzymes, the pyridine nitrogen, the phenolic oxygen, the Schiff base nitrogen, a lysine residue, which binds to the cofactor in the internal aldimine, and the phosphate group. Our group has already shown that NMR can be used to determine the protonation states of these sites as well as other potential protonation sites in the active site. The general approach is to introduce isotopic labels into the active site to delineate the chemical shifts of cofactor and substrate nuclei from surrounding environments. The isotropic chemical shifts are enough information to determine the protonation but to understand the reaction mechanism the chemical structure from NMR needs to be mapped onto the structure from X-ray crystallography. To do this we are using an integrated method between NMR, X-ray crystallography, and first-principles calculations referred to as NMR-crystallography.

This approach has been applied to the 2-aminophenol and indoline carbanion intermediate of tryptophan synthase. The major findings in that work were determining a fast proton exchange between the phenolic oxygen and the Schiff base nitrogen and showing this intermediate is a true carbanion intermediate which was thought to be a quinonoid like other conical PLP-dependent enzymes. In this thesis we will expand on the technique of NMR-crystallography applied to the active site of enzymes. We focus on two intermediates of tryptophan synthase, the internal aldimine and aminoacrylate. For the internal aldimine we are using a combination of isotropic labels from isotopically enriched PLP and chemical shifts from unambiguous assignments. We are dubbing the approach to use shifts from the surrounding residues as an “outside-in” approach to NMR-crystallography. For the aminoacrylate intermediate we seek to answer why benzimidazole does not proceed with the bond formation step despite it being a better nucleophile than the natural substrate indole. We also extend our structural analysis by determining anisotropic displacement parameters to atoms in the active site including hydrogen which have been placed through NMR-crystallography.

Chapter 1 – Introduction and Basics of Nuclear Magnetic Resonance

Spectroscopy The basics of Nuclear Magnetic Resonance (NMR) Spectroscopy have been thoroughly covered in numerous works and their applications (1-5). It is a powerful analytical tool both in solution and solids for structure elucidation, understating dynamics, and chemical information such as hybridization and protonation state. Solution-state NMR and solid-state NMR (ssNMR) have their strengths some of which are covered in this chapter. Commonly, NMR uses information from the isotropic chemical shifts to determine the local environment of an atom. The isotropic chemical shifts can be sensitive enough to distinguish protonation state, connectivity, and hybridization. Important is how this data is gathered and how to interpret the result to gain the necessary information. This Chapter is a brief summary of how information in NMR, and in particular ssNMR, is collected and interpreted.

The signal in NMR arises when looking at the magnetization generated by the nuclear spin of an atom. We can use energy levels to simplify our understanding of how signals arise. In the absence of a strong magnetic field energy levels for the “up” and “down” state, E_α and E_β, are degenerate. In the presence of a strong external magnetic field, the energy levels lose their degeneracy and the transition between these states is where our signal is derived from. To understand how these energy levels are defined, a Hamiltonian for a single-spin one-half will be defined. This Hamiltonian ($\widehat{\mathrm{H}}_{\mathrm{single\ spin}}$) can be written as where γ is the gyromagnetic ratio and B_0 is the external magnetic field.

$$\hat{\mathrm{H}}_{\mathrm{single\ spin}} = -\frac{\gamma \mathrm{B}_0}{2\pi}\hat{\mathrm{I}}_\mathrm{z}$$

The external field B_o is applied along the z-direction and we also include the nuclear spin angular momentum operator $\hat{\mathrm{I}}_\mathrm{z}$. In this case the operator is spin one-half and has two eigenfunctions, $\psi_{+1/2}$ and $\psi_{-1/2}$ described with their eigenvalues as

$$\hat{\mathrm{I}}_\mathrm{z}\psi_{+\frac{1}{2}} = +\frac{1}{2}\frac{\mathrm{h}}{2\,\pi}\psi_{+\frac{1}{2}} \ \text{ and } \hat{\mathrm{I}}_\mathrm{z}\psi_{-\frac{1}{2}} = -\frac{1}{2}\frac{\mathrm{h}}{2\,\pi}\psi_{-\frac{1}{2}}$$

This eigenfunction of $\hat{\mathrm{I}}_\mathrm{z}$ is also an eigenfunction of $\hat{\mathrm{H}}_{\mathrm{single\ spin}}$ and the eigenvalues of this Hamiltonian correspond the E_α and E_β with E_α traditionally defined using $\psi_{-1/2}$. With energy levels defined, we can now look at the transition E_α going to E_β. This allowed transition can be used to define the Larmor frequency which we will measure in NMR experiments. The Larmor frequency in hertz (Hz) is defined as.

$$\upsilon_0 = -\gamma \mathrm{B}_0/2\pi$$

And so our Hamiltonian of a single spin can be rewritten as

$$\hat{\mathrm{H}}_{\mathrm{single\ spin}} = \upsilon_0\hat{\mathrm{I}}_\mathrm{z}$$

Using this description help determine at what frequency a signal appears but does not describe how the signal is collected in a simple experiment. For that, we must consider how the bulk magnetization of nuclear spins behaves under an applied external field.

Signal Acquisition

For NMR we must consider the bulk magnetization of our nuclear spins because not all spins are oriented as “up” or “down” relative to the external field. When viewing individual spins, their alignment seems random due to thermal motion but when the magnetization is summed over, the bulk magnetization is aligned along the applied external field. At equilibrium, the magnetization is pointed along the z-axis and no signal will be observed yet. These signals come from magnetization observed along the xy- plane. If we start with magnetization along the x-axis the magnetization will process around the direction of the external field. In free precession, this is along the z-axis and we see an oscillating signal along the x- and y-axis. Magnetization along x (M_x) and y (M_y) can be described as

$$\mathrm{M_x = M_0 Sin(\beta) Cos(2\pi\upsilon_0 t)}$$

and

$$\mathrm{M_y = M_0 Sin(\beta) Sin(2\pi\upsilon_0 t)}$$

Where M_0 is the magnitude of the equilibrium magnetization, β is the angle between the magnetization vector and the z-axis. The Larmor frequency is also included as it determines the rate of procession as a function of time (t). The magnetization also has a relaxation component which will be expanded upon later in this chapter but for simplicity, it will not be considered yet. The next step is to perturb the equilibrium magnetization using a pulse to bring the magnetization into the xy-plane for detection. To apply a pulse, radiofrequency near the Larmor frequency is applied to generate a magnetic field along either the x- or y-axis causing the equilibrium magnetization to rotate into the xy-plane and then a receiver detects the signal. This signal collected is a sum of potentially several different nuclei with different procession frequencies and so the overall signal (S(t)) can be represented by the following.

$$S(t) = \sum_{j}^{k} S_j(t)$$

Since our signal comes from both M_x and M_y, the signal of a single nucleus can be written as

$$S_j(t) = (M_{j,x} + i\, M_{j,y}) \exp\left(-\frac{t}{T_{j,2}}\right)$$

Where T_2 is the relaxation term which describes the rate for the signal to decay or return to the equilibrium position. The recorded signal is called the free induction decay (FID), shown in figure 1.1. It is not typical to analyze the data from FID but instead we Fourier transform the FID to give a spectrum composed of peaks with a Lorentzian line shape. The position of these peaks are at the procession frequency and the widths are determined by T_2 and other relaxation terms.

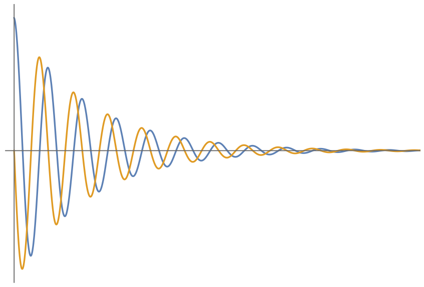


Figure 1.1: Shown is a simulated FID. The blue line is real part and the orange is the imaginary portion.

Coupling

To this point, we have only considered the signal of a single spin, but this is often not the case. When two NMR active nuclei are interacting, they have an effect on each other known as coupling. For this work, we focus on two types of coupling: scalar coupling (or j-coupling) and dipolar coupling. Both interactions occur between two nuclei in one of two ways: the homonuclear, when the spins are the same isotope such as ^{1}H-^{1}H interactions, and the heteronuclear case when spins are different isotopes such as ^{1}H-^{13}C. To describe the coupling interaction, we use the simplified spin Hamiltonian and expand on the one-spin Hamiltonian.

$$\widehat{H}_{\text{two spin}} = \upsilon_1 \hat{I}_{1,z} + \upsilon_2 \hat{I}_{2,z} + 2\pi J_{12} \hat{I}_{1,z} \cdot \hat{I}_{2,z} + d_{12}(3\hat{I}_{1,z}\hat{I}_{2.z} - \hat{I}_1 \cdot \hat{I}_2)$$

Where J_{12} is the scalar coupling term usually on the order of Hz and d_{12}, part of the dipolar term is

$$d_{12} = \frac{\mu_0 \gamma_1 \gamma_2 h}{8\pi^2 r_{12}^3}(3\text{Cos}^2\theta_{12} - 1)$$

Where μ_0 is vacuum permeability, $4\pi \cdot 10^{-7}$ H m^{-1}, γ_1, and γ_2 are the respective gyromagnetic ratios, r_{12} is the distance between the two spins, and θ_{12} is the angle between the vector of spin 1 and spin 2 and the external field. The dipolar coupling is usually on the order of kHz. A wealth of information can be gained by measuring couplings or these couplings can be exploited to reduce spectral complexity(6-8).

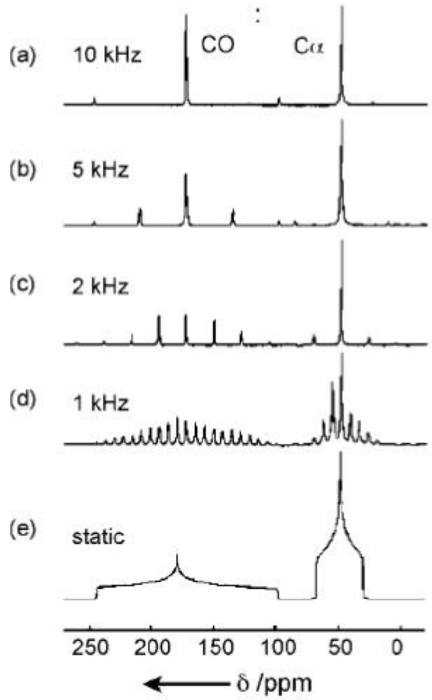


Figure 1.2: Spectra of $^{13}C_2$-glycine at various MAS rates. Reprinted from ref 4

Solid State NMR

In solution NMR, the isotropic chemical shift is observed but this is an average of the chemical shift tensor components. The molecular motion in solution is fast enough to average the different orientations during signal acquisition. With ssNMR it is typical to have a powder for organic molecules, macromolecule precipitates, or microcrystalline protein samples where the molecules have restricted molecular motion and now the chemical shift tensor is not averaged over. Instead, we get a powder pattern as shown in figure 1.2. The powder spectrum is useful because it clearly shows the chemical shift anisotropy which contains a plethora of information lost by motional averaging, however, with multiple overlapping spins it becomes difficult to interpret or deconvolve the spectrum. We can use magic angle spinning (MAS) to bring the signal intensity to the isotropic peak. This process averages both the dipolar coupling and the secular part of the chemical shift anisotropy. If the MAS rate is less than the width CSA the spectra will contain “spinning sidebands” which resemble to shape of the powder pattern. At sufficiently high MAS rates, the sideband will no longer be present, and intensity is brought to the isotropic peak.

Cross Polarization

At slow or moderate spinning speeds (<20 kHz) it is typical to detect the carbon or nitrogen nuclei in biosolids despite the relatively low sensitivity compared to hydrogen. While hydrogens have a higher sensitivity, they have broader signals, due to the strong ^{1}H-^{1}H dipolar couplings, increasing spectral overlap. At slower MAS rates coupled with ^{1}H decoupling, line widths of carbon are greatly narrowed making the ^{13}C nucleus ideal for detection. To increase sensitivity, Cross Polarization(9, 10) is used to enhance the insensitive nuclei, carbon or nitrogen, by taking advantage of higher sensitivity protons and transferring polarization. The sensitivity increase is related to the ratio of the gyromagnetic ratio of the two nuclei. This experiment works by first giving a 90-degree pulse on ^{1}H nuclei and then spin-locking with the X nuclei, ^{13}C in this example. In CPMAS(11) the spin-lock is typically set as

$$\omega_{\mathrm{H}} = \omega_{\mathrm{C}} + \mathrm{n}\omega_{\mathrm{r}}$$

Where ωH and ωC are the nutation rates of the magnetization from an rf pulse, ωr is the MAS spinning rate, and n is some nonzero integer value. One additional advantage to using cross-polarization is the reduced spin-lattice relaxation time. The relaxation time is related to the gyromagnetic ratio where a lower value corresponds with a longer relaxation time. Since this experiment relies on the polarization of hydrogen, less time is required between experiments.

Multi-Dimensional Experiments

As the number of atoms in a molecule increases the spectral complexity increases and it becomes difficult to assign peaks to atoms based solely on their chemical shift or because of overlapping peaks. Protein Co backbone shifts generally have a resonance between 170-180 ppm, shown in figure 1.3, it is obvious that determining individual peaks and then assigning them to the corresponding atoms is impossible. Instead, we can make correlations between two or more nuclei using multi-dimensional experiments. There is a multitude of ways to make correlations and

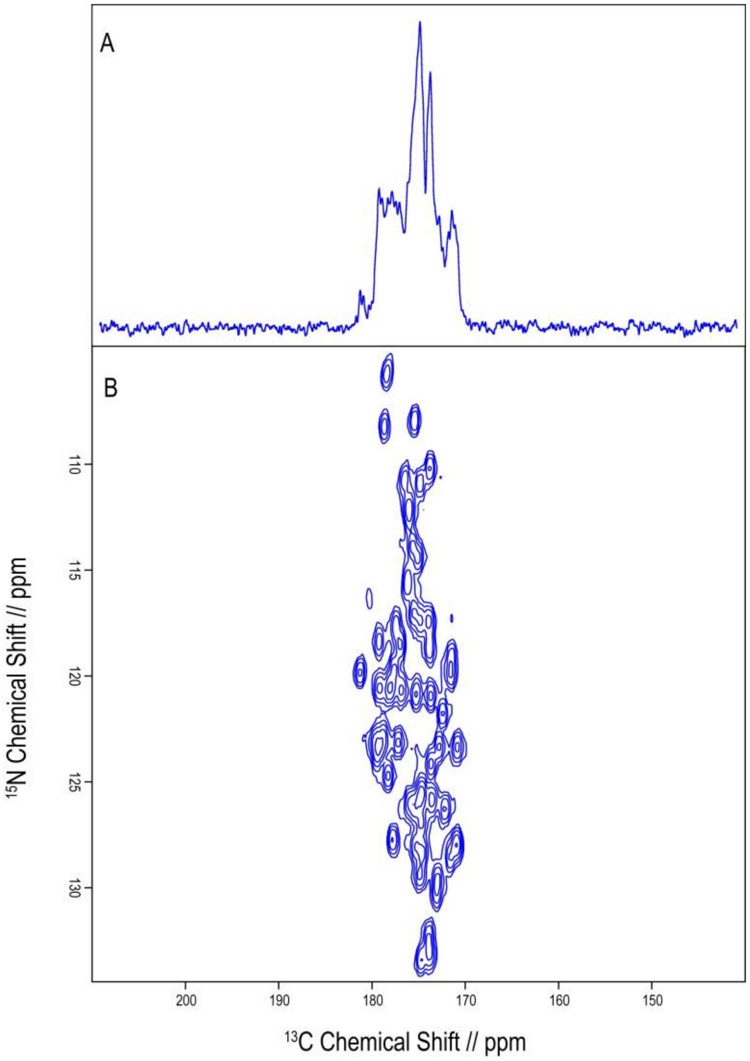


Figure 1.3: A double cross polarization experiment of (H)NCo in a one-dimensional experiment (A) and a two-dimensional experiment (B). Sample is U-^{13}C,^{15}N-GB1

the theory behind this technique is discussed in length by Ernst (2) but for this work, I will only focus on the application in biosolids. A common experiment in protein assignmentsise the double cross-cross polarization (DCP) experiments, (H)NCo or (H)NCa. These experiments start with a CP transfer from hydrogens to the amide backbone nitrogen then another CP transfer takes place to either Co or Ca. The resulting spectrum leads to distinct peaks for a small protein. In large proteins, the spectrum will still be too cluttered to be useful and a three-dimensional or n-dimensional experiment can be used to unambiguously determine assignment. To assign a protein backbone, several two- and three-dimensional experiments are required(12).

References

1. J. Cavanagh, J. Cavanagh, Protein NMR spectroscopy: principles and practice (Academic, San Diego, CA ; London, 1996), pp. xxiii, 587 p. : figs, tables.

2. R. R. Ernst, G. Bodenhausen, A. Wokaun, Principles of nuclear magnetic resonance in one and two dimensions, International series of monographs on chemistry 14 (Clarendon Press, Oxford, ed. Pbk. , correction of errors . 1989), pp. xxiv, 610 pages : illustrations.

3. J. Keeler, VleBooks, Understanding NMR spectroscopy (John Wiley and Sons, Chichester, U.K, ed. 2nd edition, 2010).

4. M. H. Levitt, Spin dynamics: basics of nuclear magnetic resonance (Wiley-Blackwell, ed. 2nd, 2013).

5. E. O. Stejskal, J. D. Memory, High resolution NMR in the solid state : fundamentals of CP/MAS (Oxford University Press, Inc, New York, 1994), pp. xii, 189 pages : illustrations.

6. P. R. Costa, J. D. Gross, M. Hong, R. G. Griffin, Solid-state NMR measurement of Psi in peptides: a NCCN 2Q-heteronuclear local field experiment. Chem Phys Lett **280**, 95-103 (1997).

7. Y. Tian et al., J-Based 3D sidechain correlation in solid-state proteins. Phys Chem Chem Phys **11**, 7078-7086 (2009).

8. Y. Pan, T. Gullion, J. Schaefer, Determination of C-N Internuclear Distances by Rotational-Echo Double-Resonance Nmr of Solids. J Magn Reson **90**, 330-340 (1990).

9. S. R. Hartmann, E. L. Hahn, Nuclear Double Resonance in the Rotating Frame. Physical Review **128**, 2042-2053 (1962).

10. A. Pines, M. G. Gibby, J. B. S. Waugh, Proton-enhanced nuclear induction spectroscopy 13C chemical shielding anisotropy in some organic solids. Chem Phys Lett **15**, 373-376 (1972).

11. J. Schaefer, E. O. Stejskal, C. F. Beard, Carbon-13 Nuclear Magnetic Resonance Analysis of Metabolism in Soybean Labeled by CO(2). Plant Physiol **55**, 1048-1053 (1975).

12. J. Hoffmann et al., Protein resonance assignment by BSH-CP-based 3D

solid-state NMR experiments: A practical guide. Magn Reson Chem **58**, 445-465 (2020).

Chapter 2 – Integrating Methods with NMR in Biosolids

This chapter focuses on some recent work that integrates solid-state Nuclear Magnetic Resonance (ssNMR) with structure tools of biology, cryoEM and x-ray crystallography, and computational methods, first principles and molecular dynamics. The applications of these integrated approaches in biosolids can range from whole structure refinement to active sites or focus on specific residues(1-9).

NMR provides a wealth of information (10) about the local environment as well as some distant restraints up to 10 Å but as a protein gets larger, the spectral complexity increases and the number of restraints required also increases. This makes structure determination for large enzymes a difficult process but can be simplified with structures from either X-ray or CryoEM, even with low-resolution structures. Focusing on active sites, NMR alone can determine protonation states of a ligands or substrate but lacks information about the 3D structure which is critical in understanding enzyme mechanism. The powe of NMR is ability to apply the learned information to other methods and improve the chemical understanding of a system. For proteins the use of X-ray or CryoEM provides a scaffolding to for initial models.

The advantage of X-ray and CryoEM is the ability to provide a framework for atom positions. Even in low-resolution structures where exact atom positions are uncertain, secondary structure motifs like alpha-helices are distinguishable. It becomes possible to model some restrains from NMR into the positions and improve the structural information. As technologies from these methods improve and higher resolution is achieved one critical component from a mechanistic standpoint is the position of hydrogen atoms. The hydrogen bond network can be inferred form these structures but the initial protonation states in acid-base catalyst is need to map out intermediates especially when the protonation state differs from the typical states. One example is the protonation state of Lys87 in Tryptophan synthase. For the aminoacrylate intermediate, the ε-amino group in neutral contrary to expected positive charge for lysine at pH 7.8 (3). Neutron structures alleviated some of the burden for determine protonation states but these often required large single crystal of perdeuterated protein and multiweek acquisitions (11). The protonation states aid in building initial models, but these structures lack the dynamic information which can identify if any chemical exchange is occurring and what the exchange partners are. Already there is some advantage to combining NMR with structural methods to extend our understanding of a biological system. What is needed now is to use a way to generate and test models, for that we turn to computational methods.

There are two computational methods I will focus on, first principles calculations and molecular dynamics (MD). MD uses molecular mechanics to generate and ensemble of structures which access conformations in sides chains not obvious in crystal structures. One example from tryptophan synthase is Gln114 where it was show that the side chain was able to rotate so the sidechain amide was able to interact with the cofactor despite there being little evidence in crystal structures, even those with relatively high resolution (12). Through kinetic and NMR studies it was reveal that when this residue was mutated to alanine, the steady-state rate of L-trp formation was reduced approximately 6.5 fold (13). MD has also been used in structure refinement using chemical shifts as factor for the generated ensembles. The power in MD is the large number of chemically reasonable models it can produce in a short amount of time which point out potential targets for future studies. There are more advantages to molecular dynamics but I will limit the discussion here.

Many groups use integrated methods to understand the structure and chemistry of their biologic system so to highlight some uses of integrated approaches I summarize work from a few groups including work from our lab.

NMR Crystallography

Tryptophan synthase is 144kDa $\alpha_2\beta_2$ bienzyme complex that catalyzes the last two steps in L-tryptophan formation. In the α-subunit, indole-D-glycerol 3'-phosphate is cleaved to for indole and glyceraldehyde 3-phophate. The newly formed indole moves from the α-subunit to beta -subunit through a tunnel connecting the two active sites. The beta-subunit starts with betaLys87 bound to the cofactor pyridoxial-5'-phosphate (PLP) through a Schiff base linkage, the internal aldimine intermediate (E(Ain)). L-serine enters the beta-subunit active site and displaces betaLys87 forming a new Schiff base link between L-ser and PLP. The subsequent steps are extraction C^{α} proton from L-ser (E(Aex)) and the elimination of the beta-hydroxyl from L-ser to give the alpha-aminoacrylate intermediate (E(A-A)). The indole from the alpha-subunit is now positioned to make the nucleophilic attack at C^{β} of the aminoacrylate forming L-trp. L-trp is released from the beta-subunit and the internal aldimine is restored. Much of the work discussed comes from a larger work by Mueller and Linser (14) where the unambiguous assignment of tryptophan synthase was accomplished using 4D and 5D methods but focus will be on the NMR crystallography applied to the beta-subunit active site.

In brief, NMR crystallography is the integrated approach of NMR, X-ray crystallography and first principles(3-5). X-ray provides a scaffolding of heavy atoms which is used as the building block for models in first principles calculations. NMR is used to measure chemical shifts in the active site. The calculated shifts from first principles are compared to experimental data to determine the best fit and equally importantly, those models which can be excluded. The method has shown to be exceptional for studying enzyme active site with positional uncertainties of 0.11-0.16 Å.

This study focuses on the internal aldimine of tryptophan synthase, and intermediate common to all PLP-dependent enzymes. Chemical shifts on the cofactor, pyridine nitrogen (N1), C2, C2', and C3, and side chain shifts for betaLys87, SBN/N^{ζ}, C^{ε}, C^{δ}, and C^{γ} are used to test models of the active site. The initial finding is a chemical shift for the Schiff base nitrogen (SBN) is 227.3 ppm at 30 ^{0}C which suggestion an proton exchange between the nitrogen and some partner. To determine a structure which describes the experimental shifts models

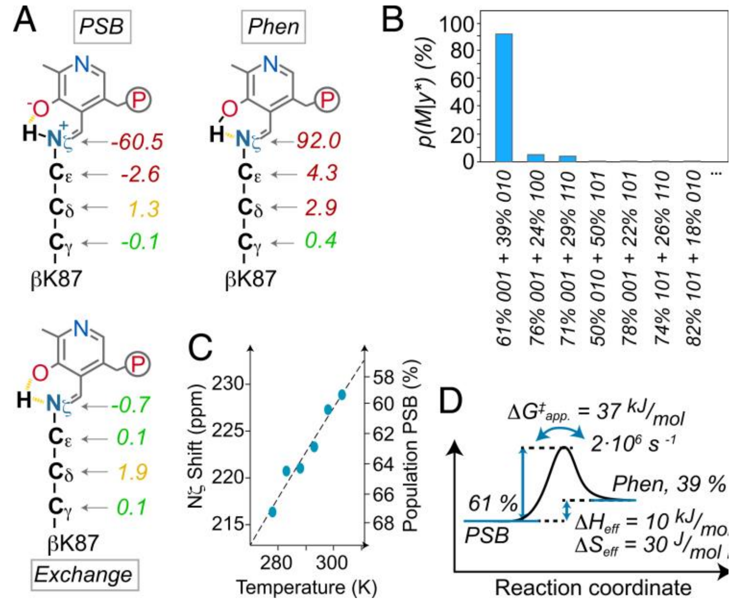


Figure 2.1: Tautomers of the internal aldimine and the probability of each exchange pair. Figure reproduced from ref 14

of the active site used to explore various protonation states. This is done by taking a residues within 7 Å of the PLP cofactor using the heavy atom coordinates from the X-ray structure 4HT3. These models are geometry optimized using DFT level of theory with the outer residues held in their crystallographic location. Next NMR calculations are performed to compare models to the experimental shift by using the reduced-χ^2 statistic. The candidate structures were generated by systematically varying the protonation states of the phenolic oxygen, Schiff base nitrogen, and the pyridine nitrogen. What is found is that none of the candidate structures have acceptable agreement between the predicted chemical shifts and the experimental shifts with the lowest reduced-χ^2 of 28.7 for the protonated Schiff base. Fast exchange models are considered which is consistent with the temperature dependence observed. Only one exchange pair was found to satisfy the 95% confidence limit, a proton exchange between the phenolic oxygen and the Schiff base with respective populations of 39% and 61%. Furthermore a Bayesian analysis show the best fit model is most probable with an 85%

proability. This finding is consistent with the hypothesis that the Schiff base is protonated in the internal aldimine to activate C4' of the cofactor so that L-serine can make a nucleophilic attack and begin the catalytic cycle.

Using first principles also has another subtle advantage which is the prediction of theoretical shift. The model compound studies offer a good estimation of the chemical shifts of a system with 100% population but the calculations offer system specific theoretical values. These values were using in the analysis of the equilibrium constant for the tautomeric exchange and in estimation of the activation energy for the proton exchange between the phenolic oxygen and the Schiff base. By measuring the linewidth as a function of temperature in the following equation the free energy of activation was estimated to 37 kJ mol^{-1} and the enthalpy and entropy were estimated to be 10 kJ mol^{-1} and 30 j mol^{-1} K^{-1} respectively using the relationship between tautomer population and temperature.

One of the interesting finds was the strong temperature dependence of the Schiff base given a simple two-state model for the proton exchange. This significant entropy term and large free energy of activation were not expected for this exchange pair, but an offered explanation was coupled dynamics to large scale motion of TS. It is known there are two conformations, open and closed, which are present in different intermediates of TS. Some $R_{1\rho}$ experiments shown elevated chemical exchange rates for residues at the opening of the beat-subunit supporting that some large-scale motion is occurring.

The integrated technique of NMR crystallography shows the in-depth analysis possible by combining the NMR, X-ray, and computational chemistry. The proton exchange modeled in this study provide insight to “one possible mechanism for coupling global structure changes with chemical reactivity as part of allosteric regulation”. The level of insight into the reaction mechanism is remarkable especially when considering the application to multiple intermediates. The reaction mechanism become clearer and the bridges the intermediates in a catalytic cycle with significant resolution.

Molecular Dynamics and ssNMR

Next the work by Glaubitz and co-workers (15) where they used NMR and molecular dynamics to investigate the reverse adenylate kinase (rAK) phosphoryl transfer of two ADP to ATP and AMP in the ATP binding cassette (ABC) exporter, MsbA. These ABC transporters are proteins which transport a range of substrates across a membrane. Often these proteins are composed of a substrate-specific transmembrane domain (TMD) and the nucleotide-binding domain (NBD). MsbA transports lipopolysaccharides across the cell membrane typically undergoing ATP-hydrolysis for function. The ATP hydrolysis occurs in the NBD region of the protein and under low ATP conditions it was observed to undergo the rAK reaction. This finding is not exclusive to MsbA but similar activity was shown in other ABC exporter. Given that the conical binding sites for ATP are too far from each other for the rAK reaction to occur, an additional binding site must be present to facilitate the phosphate transfer. The hypothesis is that this other binding site is near the Q-loop based on mutation studies. The reaction in the NBD region is separated into 5 states. State 4 is the step in which the phosphoryl transfer occurs and cannot be directly studied. Analogues of ATP are used to trap the various states of the reaction cycle. For this review, focus will be placed on the AK-binding site and the interplay of NMR and MD to model this site.

To establish ADP binding in the AK-site two regions of the NBD were targeted using unique pair-labeling. This is used to identify a single N-Co connection in the protein backbone to reduced spectral complexity. The two pair identifies are Val426-His427 in the Q-loop and Ala536-His537 in the His-switch region. In states 3 and 5, where ADP of AMP occupy the AK-site, a second peak appears for both C^γ of Val426, shifted -3 ppm, and C^β of Ala 536, shift -4 ppm. The absence of a second peak in states 1 and 2 support the hypothesis of nucleotide binding the AK-site. Next to make direct correlations between residues in the conical sites and the AK-site, the protein was uniformly carbon-13 labeled and ^{15}N-enriched nucleotides were introduced. To measure the direct interaction, ^{15}N-^{13}C TEDOR experiments were employed using DNP MAS NMR. In the conical sites, correlations between the nucleotide and Try351 are observed. This correlation is expected given well characterized π-π stacking interaction between adenosine and the residue. What is notably found is cross peak between Gln424 and ^{15}N-AMP. This correlation provides a direct measurement of nucleotide interaction in the AK-site. What NMR doesn’t provide is a structural model for how AMP is position in this site and from previous ^{31}P-CPMAS NMR data (16) has suggested the possibility of two ADP molecules binding in the AK-site. Here molecular dynamics offers insight into potential binding motifs.

For the phosphoryl transfer to take place, one ADP molecule has to be in the conical binding site, and another has to be position in the AK-site correctly. Because there can be potentially two AMP nucleotides in the AK-binding site there are 3 possible motifs explored by Glaubitz and coworkers. First where only a single AMP is bound in the AK-site and positioned for the phosphoryl transfer. Second is where 2 AMP molecules are in the AK-sites and both oriented to make the transfer. Third is where 2 AMP molecules occupy the AK-site but only one is positioned to make the phosphoryl transfer. The MD were carried out by binding ATP molecules to both conical sites and running simulations to produce stable conformations then AMP was loaded into the AK-site. When was found is that binding of 2 AMP nucleotides was generally favorable then when loaded with a single nucleotide, measure by considering the number of hydrogens formed between the Q-loop and the nucleotide. For both simulations of one and two AMP in the AK-site, the connection between to Q242 observed in the TEDOR experimented was also present. The simulations also reveal the phosphate group is oriented towards H537. From the simulations with two AMP molecules stay in the site, one of the AMP molecules would be in position of the rAK reaction while the second is not in the correct conformation.

Using NMR, direct correlations to residues in the Q-loop and AMP were measured establishing on of the hydrogen bond interactions. The results from NMR, however, remain ambiguous for delineating between the quantity of nucleotides in the target site. This can be modeled using molecular dynamics and many simulations can be performed. The combination of NMR and MD in this study suggest the more likely model is a single nucleotide in the AK-site. This integrated approach provides insight into the AK- binding site for the reverse reaction of MsbA. While the result from Glaubitz do not definitively determine the binding mode of AMP, the models provide excellent insight in the reaction pathway and targets for future studies.

Another example of combining NMR and molecular dynamics comes from McDermott and coworkers (9) where they look at the conformations of KcsA. In the solid state NMR three states have been stabilized for study: the deactivated state, the activated states, and the inactivated state. Two of the states, deactivated and inactivated, have been well characterized and the is structural agreement between X-ray diffraction and ssNMR. The activate state has be postulated to have either a partially open (PO) of a fully open (FO) conformation with the degree of openness defined by the distance between cross bundles with PO being approximately 16 Å and FO being 23 Å. The activated state is the only state which conduct K^+ ions and so it remains of high interest for study and understand the conformation this state has. To do this McDermott and coworkers create a workflow using the assignments from NMR and molecular dynamics to determine if the activated state in NMR is PO or FO.

NMR assignment for the activated state were done with a suite of 3D experiments and the shift were compared against the deactivated state to identify the experimental markers. The activated and deactivated states were simulated using molecular dynamics with separate simulations for the PO and FO state. Then the frames from the simulations were used to predict chemical shifts using SPARTA+ and SHIFTX2. In initial comparisons they do not see a clear difference between the predicted shifts of the PO and FO states but using a statistical approach to determine nuclei which are sensitive to the different conformations they find 49 nuclei which can be used as reports. The determine the most probable structure they compare a difference in experimental activated versus deactivated and the simulated activated versus deactivated but to reduce error they approach it as a double difference as shown.

$$\Delta\Delta \mathrm{CS}_{\mathrm{sim}}^{\mathrm{FO/PO}} = \Delta \mathrm{CS}_{\mathrm{sim}}^{\mathrm{FO/PO}} - \Delta \mathrm{CS}_{\mathrm{exp}}$$

$$= \mathrm{CS}_{\mathrm{sim}}^{\mathrm{FO/PO}} - \mathrm{CS}_{\mathrm{sim}}^{\mathrm{deactivated}} - (\mathrm{CS}_{\mathrm{exp}}^{\mathrm{actiavted}} - \mathrm{CS}_{\mathrm{exp}}^{\mathrm{deactivated}})$$

If the distribution from the double difference approach is generally lower in value and the two distributions are not overlapping for FO and PO then a state can be assigned to the active state studied in NMR. The authors find that the PO state is majority state for KscA. This finding is important in guiding future experiments.

The us of molecular dynamics shows again the power in being able to simulate large numbers of ensembles with different conformations. Here is allowed the authors to distinguish between to confirmations which could not be identified from the chemical shifts alone. Integrating methods with NMR proves a valuable tool to elucidate both structure and chemical understanding.

Conclusion

Combining NMR with other tool of structural biology and computational chemistry furthers our understanding of biological system. Importantly it does two things, elucidates potential models and eliminates others. NMR helps to bridge the gap between mechanism and structure while computational methods allow models to be built and reliably tested. The integration of techniques provides chemical detail that relates to structure and vice versa. The connection between mechanism and structure allows new and interesting chemistry to be understood in biological systems.

Reference

1. D. F. Gauto et al., Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. Nature Communications **10**, 2697 (2019).
2. J. Felix et al., Mechanism of the allosteric activation of the ClpP protease machinery by substrates and active-site inhibitors. Science Advances **5**, eaaw3818 (2019).
3. J. B. Holmes et al., Imaging active site chemistry and protonation states: NMR crystallography of the tryptophan synthase alpha-aminoacrylate intermediate. Proc Natl Acad Sci U S A **119** (2022).
4. B. G. Caulkins et al., NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. J. Am. Chem. Soc. **138**, 15214-15226 (2016).
5. L. J. Mueller, M. F. Dunn, NMR crystallography of enzyme active sites: probing chemically detailed, three-dimensional structure in tryptophan synthase. Accounts of Chemical Research **46**, 2008-2017 (2013).
6. J. R. Perilla, J. A. Hadden-Perilla, A. M. Gronenborn, T. Polenova, Integrative structural biology of HIV-1 capsid protein assemblies: combining experiment and computation. Curr Opin Virol **48**, 57-64 (2021).
7. H. L. Zhang et al., HIV-1 Capsid Function Is Regulated by Dynamics: Quantitative. Atomic-Resolution Insights by Integrating Magic-Angle-Spinning NMR, QM/MM, and MD. J. Am. Chem. Soc. **138**, 14066-14075 (2016).
8. H. Singh et al., Fast Microsecond Dynamics of the Protein-Water Network in the Active Site of Human Carbonic Anhydrase II Studied by Solid-State NMR Spectroscopy. J. Am. Chem. Soc. **141**, 19276-19288 (2019).
9. S. Perez-Conesa, E. G. Keeler, D. Zhang, L. Delemotte, A. E. McDermott, Informing NMR experiments with molecular dynamics simulations to characterize the dominant activated state of the KcsA ion channel. J Chem Phys **154**, 165102 (2021).
10. J. Kraus, S. Sarkar, C. M. Quinn, T. Polenova, Solid-state NMR spectroscopy of microcrystalline proteins. Annu Rep Nmr Spectro **102**, 81-151 (2021).
11. S. Dajnowicz et al., Direct visualization of critical hydrogen atoms in a pyridoxal 5′-phosphate enzyme. Nature communications **8**, 1-9 (2017).
12. Y. M. Huang et al., Protonation states and catalysis: Molecular dynamics

studies of intermediates in tryptophan synthase. Protein Sci **25**, 166-183 (2016).

13. R. K. Ghosh, E. Hilario, C. A. Chang, L. J. Mueller, M. F. Dunn, Allosteric regulation of substrate channeling: Salmonella typhimurium tryptophan synthase. Front Mol Biosci **9**, 923042 (2022).
14. A. Klein et al., Atomic-resolution chemical characterization of (2x)72-kDa tryptophan synthase via four- and five-dimensional (1)H-detected solid-state NMR. Proc Natl Acad Sci U S A **119** (2022).
15. H. Kaur et al., Unexplored Nucleotide Binding Modes for the ABC Exporter MsbA. J. Am. Chem. Soc. **140**, 14112-14125 (2018).
16. H. Kaur et al., Coupled ATPase-adenylate kinase activity in ABC transporters. Nature Communications **7** (2016).

Chapter 3 – The Internal Aldimine Intermediate of Tryptophan Synthase

Introduction

The family of pyridoxal-5′-phosphate (PLP)–dependent enzymes catalyze a wide variety of chemical transformations including transamination, racemization, decarboxylation, elimination, and substitution (1-4). The large number of PLP enzymes and their crucial metabolic functions make them drug targets for the treatment of diseases including tuberculosis, epilepsy, and Parkinson's disease (5, 6). Fig. 1 depicts the crystal structure of Salmonella typhimurium tryptophan synthase (TS) (7). TS itself is both an important drug target in the context of continuously emerging bacterial antibiotics resistance (8) and of great interest in biotechnology (9) as an enantiospecific source of a large variety of unnatural amino acids and their derivatives (10, 11). Wild-type TS catalyzes the final two steps in tryptophan biosynthesis: production of indole from indole-3-glycerol phosphate (IGP) and its subsequent condensation reaction with L-serine to give L-tryptophan (7, 12, 13). As for many other enzymes, X-ray structural data are abundant, but the rational design of therapeutic agents and the understanding and engineering of catalysis, in particular regarding the β-subunit enzymatic reaction, hinge on the availability

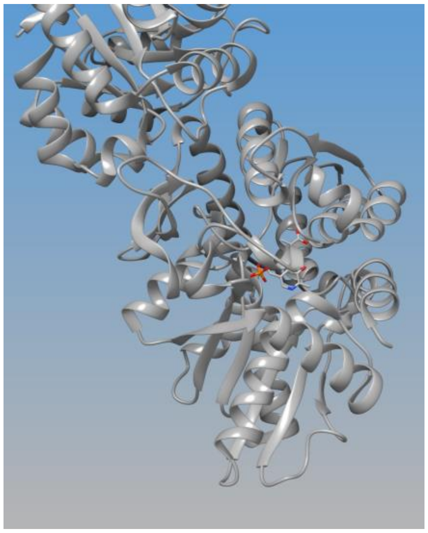


Figure 3.1: The crystal structure of Tryptophan synthase

of detailed knowledge of the chemical and electrostatic properties of the active site. The β-subunit reaction, which acts as a pivot for the overall reaction and selectivity of the catalytic cycle. Nucleophilic attack of the PLP cofactor in the β-subunit active site is thought to involve activation of C4′ by protonation of βLys87 Nζ. However, the thermodynamic and kinetic details of potential tautomeric exchange are currently missing (13-15). Such features, in particular protonation, hybridization, and tautomeric states of the active-site side chains and substrates, cannot be directly determined by protein crystallography or cryoelectron microscopy (cryo-EM) but are accessible from NMR chemical shifts.

Results and Discussion

Protonation of the Schiff-base nitrogen, Nζ, has been proposed to activate the cofactor C4′ carbon for nucleophilic attack by the incoming substrate, serine (13-15). How this activation might be coupled to larger conformational motions responsible for substrate trapping and allosteric signaling remains an intriguing mechanistic question. Here, the intermediate value of 227.3 ppm found for Nζ at 30 °C suggests a dynamic tautomeric exchange between protonated and neutral Schiff-base forms (11). To quantify the exchange and the identity of the chemical structure of the exchanging partners, we turned to NMR-assisted crystallography—the integrated application of ssNMR, X-ray crystallography, and first-principles computational chemistry (13, 16-25).

Our approach follows that of Caulkins et al. (13). Starting with the crystal structure of the TS internal-aldimine form (Protein Data Bank identifier [PDB ID]:

4HT3), a cluster model of the active site was constructed that included all residues within 7 Å of the PLP cofactor. Five models of the active-site chemistry were generated by varying the protonation states of the pyridine ring nitrogen, pyridoxal phenolic oxygen, and Nζ of βLys87 (Scheme 1). Each of these candidate structures was geometry optimized using density functional theory (DFT), with the exterior residues of the cluster fixed at their crystallographic positions. NMR chemical shieldings were calculated using a locally dense basis approach and converted to chemical shifts (26) (Appendix Table 3.1). Finally, the structural models were ranked based on the agreement between their first-principles predicted chemical shifts and the ssNMR

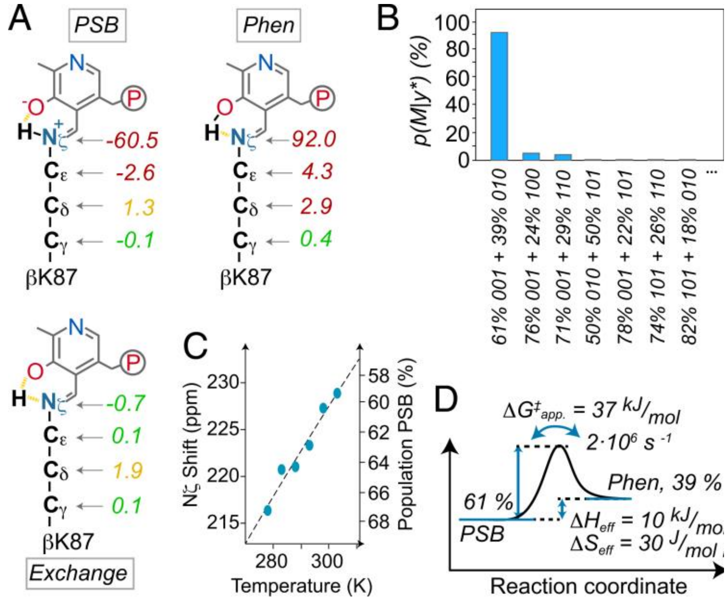


Figure 3.2: Analysis of the tautomeric equilibrium at the βK87 Schiff base in the active site of the β-subunit. (A) Shifts obtained from first-principles calculations for the active site with PSB, protonated phenolic oxygen (Phen), and tautomeric exchange between the two with a 61:39 ratio. (B) Bayesian probability for the best-fitting combinations of different two-site, fast-exchange equilibrium models, comparing experimental and first-principles chemical shifts, with protonation (C) Best fit of temperature-dependent populations to the apparent enthalpy and entropy difference for the entire tautomeric exchange process within the enzyme. The dashed curve shows the dependence consistent with ΔHapp. = 9.6 ± 1 kJ/mol and ΔSapp. = 28.6 ± 3 J/(mol K). (Populations are derived from temperature-dependent 15Nζ chemical shifts measured in a 2D fashion but with universal carbon shifts.) (D) Effective energy profile of the tautomeric exchange of the Schiff base at 30 °C,

assignments for the βLys87 side chain and PLP cofactor using the reduced-χ^2 statistic, the weighted deviation of the model from experimental shifts.

Of the candidate structures, none was found to show acceptable agreement between the predicted and experimental chemical shifts, with the lowest reduced $\chi^2 = 28.4$ (Table 3.1). Based on the temperature dependence and large line width of the Schiff-base nitrogen, fast-exchange equilibrium models were considered next, in which the effective chemical shifts were given as the population-weighted average of the shifts for the individual structures. All models were paired and their populations optimized for best agreement with the experimental chemical shifts. The best-fit fast exchange, with a reduced χ^2 of 1.3, was found to be between the protonated Schiff base (PSB; ketoenamine) and phenolic (Phen; enolimine) forms (Fig. 6A). Table 1 summarizes select experimental and first-principles predicted shifts for this model compared with its parent states. The next-best exchange model had a reduced χ^2 of 4.9, making it statistically unlikely (SI Appendix, Table S7). Bayesian probability analysis (27) confirmed that the best-fit exchange model is the most probable experimental state with 91% confidence (Fig. 6B and SI Appendix, Section 11).

The populations of PSB and phenolic tautomers are strongly temperature dependent which is unexpected for a simple two-state model for the exchanging proton and requires a significant entropy term to accommodate. (See the fit of enthalpy/entropy contributions to the free-energy difference in Fig. 6C and SI Appendix, Section 12.) The temperature dependence of the populations is, however, consistent with larger-scale processes in the surrounding active site being coupled to this exchange. Crystal structures of TS show both open and closed conformations of the β-subunit for various intermediates in the catalytic cycle (SI Appendix, Fig. S13)(3, 7). The open conformation is necessary for the free diffusion of substrate into the active site. It also establishes an aqueous environment proximal to the cofactor that favors the Zwitterionic, PSB form (13). Closed conformations largely exclude water from the active site, favoring the neutral Schiff base, phenolic form. The open and closed states of the active site remain in equilibrium, with a switch between the predominant form for the various intermediates (7). An entire conformational exchange through the crystallographic conformations is unlikely in the absence of a substrate. However, it is noteworthy that already within a single (cryogenic) X-ray structure (PDB ID: 4HT3), conformational plasticity is seen in the entry portal for serine in the β-subunit active site (SI Appendix, Fig. S14), which involves interactions between residues in the carboxy-terminal α-helix and the loop holding the cofactor (SI Appendix, Figs. S14 and S15). Likewise, an isolated tautomeric-exchange process would be associated with a low effective activation barrier and an expected timescale in the picosecond-to-nanosecond regime (28, 29). To assess the effective timescale of the tautomeric exchange in the enzyme and whether contributions from conformational motion in the surrounding could play a role, we conducted line shape analysis of the Schiff-base nitrogen based on the limiting chemical shifts given by the computational modeling as well as $R_{1\rho}$ analysis of the protein backbone via the pseudo-4D, relaxation-edited hCONH experiments. Line shape analysis gives experimental access to the apparent rates of the tautomeric exchange and hence, via the Eyring equation, the effective free energy of activation. Whereas homogeneous nitrogen line widths in the sample in the absence of exchange (including the βK87 backbone amide) generally amount to only around 20 Hz, linewidths for the Schiff-base nitrogen are on the order of 270 Hz. Equally, the Schiff-base proton has a line width of 120 Hz compared to amide H^N widths of, generally, around 50 Hz. Assuming a two-state exchange, the exchange-broadened lines suggest a tautomeric turnover on the microsecond motional timescale, with a forward rate of around 2.4 $*10^6$ s^{-1}, a regime much slower (or a $\Delta G^{\ddagger}$ larger) than expected for an isolated proton-exchange process (28, 29). Even though the ssNMR line width does not purely reflect the exchange contribution as in solution NMR and would have to be scaled down somewhat (to account for residual anisotropic interactions, sample inhomogeneity, and anisotropic bulk magnetic susceptibility) (30), the high apparent effective activation barrier is in line with the Nζ temperature dependence and consistent with a linkage of tautomerism with variations in the surrounding electrostatic environment. Details of conformational exchange dynamics can be provided by relaxation dispersion experiments, which have remained exceedingly demanding for TS so far. However, the simple $R_{1\rho}$ experiments at least reveal moderately elevated rates (5 to 9 s^{-1}, compared to rates of 1 to 2 s^{-1} for inconspicuous residues) at ∼30 °C for entry portal residues like D649 and G352. The increased rates, and likely also the low degree of assignments in the adjacent communication domain residues, are in agreement with the exchange-broadened character of the Schiff-base line shapes. Whereas details of the modulation of the exchange by the environment remain elusive, the data are consistent with a connection between tautomerism and a variable pocket environment, which suggests a modulation of Schiff-base protonation by the surrounding architecture. Possible scenarios in the functional enzyme would be a conformation-dependent fast-exchange equilibrium of the Schiff base directly driven by electrostatics or changes in the water network that differentially stabilize neutral and Zwitterionic forms dependent on pocket conformations and thus indirectly tune the Schiff base properties. These provide one possible mechanism for coupling global structural changes with chemical reactivity as part of the allosteric regulation in TS.

Conclusion

We have studied the Internal aldimine of tryptophan synthase, an intermediate common to all PLP-dependent enzymes, using NMR crystallography. A proton exchange was established between the phenolic oxygen and the Schiff base with the dominate species being the protonated Schiff base. The exchange rate and activation barrier do not correspond to a simple proton exchange as expected. This supports a hypothesis that global activity is coupled to the reaction mechanism which can active C4’ of PLP making the carbon more electrophilic.

The detailed analysis of a common intermediate provides insight into the electronics which give PLP-dependent enzyme a diverse array of function.

Methods

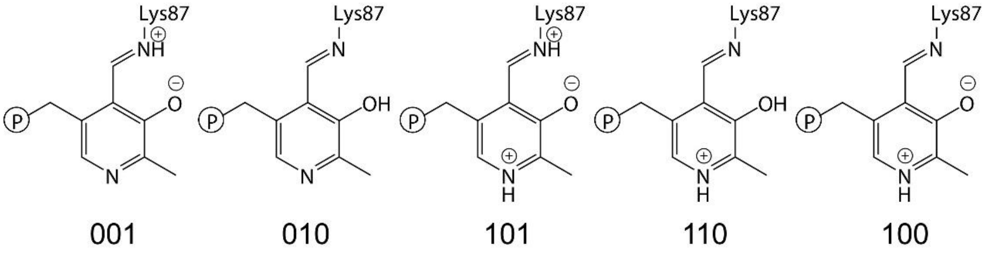
First Principles Cluster Models

The active-site experimental chemical shifts were used as restraints to evaluate computational models of structure and dynamics using “NMR-assisted crystallography,” the synergistic combination of solid-state NMR spectroscopy, X-ray crystallography, and first-principles computational chemistry(24, 25, 31-33). In this approach, models of the active site are built on the framework of an X-ray crystal structure, and these models are quantitatively screened and ranked based on the agreement between their first-principles predicted chemical shifts and the results of solid-state NMR experiments.

Our approach follows that outlined by Caulkins et al. (13). To begin, a cluster model of the enzyme active site was constructed starting with the crystal structure of the resting, internal aldimine form of TS (PDB ID: 4HT3) by selecting all residues that contained atoms within 7 Å of the cofactor. This selection was expanded to include complete residues and was modified as follows: (1) Residues that were not part of continuous backbone segments and with only two atoms within the initial 7 Å cut were deleted; (2) residues with two backbone atoms and no side chain residues within 7 Å were converted to glycine (βL349); (3) residues having two side chain atoms and no backbone atoms within 7 Å were truncated by removing the backbone atoms and replacing Cα with a methyl group (βH313, βD383); (4) N-terminal nitrogen atoms were replaced with a hydrogen atom and C-terminal carbonyls were capped with an $-NH_2$ group (amidated); (5) the Cs^+ cation was replaced with Na^+; and (6) hydrogen atoms were added. The final clusters contained 577-578 atoms depending in the protonation state of the cofactor.

Next, the active-site cluster was used as a framework for building first-principles computational models. Five candidate structures (Scheme S1) were generated by systematically varying the protonation states of the following ionizable sites: the pyridine ring nitrogen; the pyridoxal phenolic oxygen; and the Schiff base nitrogen (Nζ of βLys87). A binary code for the protonation states of the active site was used to label the models, where 1 indicated that the site was protonated and 0 indicated no protonation. The code refers to the following 3 positions in order: (pyridine nitrogen)(phenolic oxygen)(Schiff base nitrogen). Based on its ^{31}P CSA tensor (34), the phosphate was fixed as di-anionic for all calculations. The fully deprotonated (000) and triply protonated (111) models were not considered, nor was the state in which both the phenolic oxygen and Schiff base were protonated (011).

Scheme 3.1: Candidate chemical structures for the internal aldimine state



Models were geometry optimized in Gaussian09 at the DFT B3LYP level of theory using the Grimme D3 empirical dispersion correction and a two-tier, locally dense basis set with 6-31G(d,p) for the cofactor and 6-31G for the remaining atoms. Atoms within 4 Å of the cofactor and all hydrogens were allowed to adjust while remaining atoms were fixed at their crystallographic coordinates.

For each optimized structure, NMR shieldings were calculated using Gauge Independent Atomic Orbital (GIAO) method at the DFT B3LYP level of theory using a three-tier, locally dense basis set assignment with the PLP/Lys87 atoms at 6-311+G(2d,p), atoms within 4 Å of the PLP/Lys87 at 6-311G(d,p), and all remaining atoms at 6-31G. NMR shielding values were converted to chemical shifts using the following linear rescaling

$$^{13}\text{C}: \qquad \delta[\text{DSS(l)}] = \delta[\text{TMS(l)}] + 2 = 175.70 - 0.9685\sigma_{\text{calc}}$$

$$^{15}\text{N}: \qquad \delta[\text{NH}_3\text{(l)}] = \delta[\text{NH}_4\text{Cl(l)}] + 39.27 = 230.45 - 0.9996\sigma_{\text{calc}}$$

Benchmark studies across test sets of solid-state structures for the linear rescalings gave root-mean-square-error (RMSE) for the isotropic shift of 1.5 and 4.3 for ^{13}C and ^{15}N, respectively (26, 35).

Finally, models were ranked based on the agreement between their first-principles predicted chemical shifts and the experimental NMR parameters using the reduced-χ^2 statistic,

$$\chi_r^2(\text{model}) = \frac{1}{N-f}\sum_i \frac{(\delta_i^{\text{model}} - \delta_i^{\text{exp}})^2}{s_i^2}$$

after application of isotope shift correction to the set of synthetic shifts. Here, N is the number of measured NMR shifts, f is the number of adjustable model parameters (0 for static models, 1 for fast-exchange models), δ_i^{exp} is the experimental chemical, δ_i^{model} is the predicted chemical shift from first-principles calculations, and s_i^2 is the nuclide-specific weighting derived by setting s_i to the respective RMSE from benchmark studies (1.5 and 4.3 for ^{13}C and ^{15}N, respectively). The reduced-χ^2 should approach 1 for a large number of measured chemical shifts, provided the weights have been properly assigned. For a finite number of experimental values, the reduced-χ^2 varies due to sampling statistics in a manner consistent with the χ^2 probability distribution.

Experimental chemical shifts for βK87 were obtained via a combination of the 4D hCCNH, the H/N, and a long-range H/C correlation as described in the main text. Experimental shifts for the PLP ring were reported previously at -10 ^{o}C, but were measured again at 30 ^{o}C for consistency (Fig. S12 and Table S6).

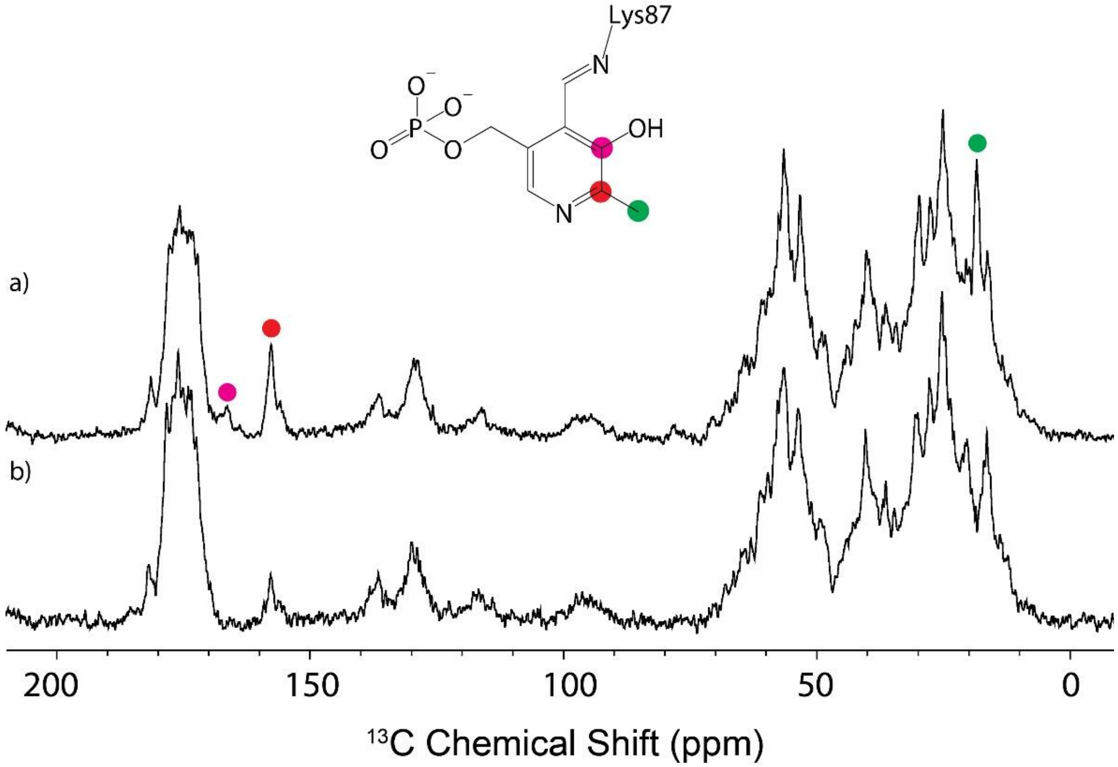


Fig. 3.3: 13C SSNMR CPMAS spectra of microcrystalline TS with the following isotopic labeling: (a) natural abundance isotopomer concentration (b) selectively 15N and 13C enriched on the PLP at N1, C2, C2’, and C3. Spectra were acquired at 9.4 T, 8 kHz MAS, and 30 ^{o}C.

Table 3.1 summarizes the first-principles chemical shifts for the 5 cluster models of the active site with varying protonation and quantifies their agreement with the experimental chemical shifts. None of the single, static models shows acceptable agreement with the experimental shifts, with the Schiff-base nitrogen (βLys87 Nζ) an outlier for all models.

Of the candidate structures, none was found to show acceptable agreement between the predicted and experimental chemical shifts, with the lowest reduced-χ^2 = 28.1 (Table S6). Based on the temperature dependence of the Schiff-base

Table 3.1: Experimental and first-principles predicted chemical shifts (ppm) for cluster models of the internal aldimine with varying protonation.

	PLP N1	PLP C3	PLP C2	PLP C2'	βLys87 Nζ	βK87 Cε	βK87 Cδ	βK87 Cγ	Reduced χ^2
Experiment	294.7	168.3	159.6	20.4	227.3	51.3	32.7	26.9	-
001	304.4	173.9	162.0	23.3	166.8	50.7	34.0	25.8	28.1
010	302.7	158.8	153.3	22.2	319.3	57.6	35.6	26.3	67.8
101	203.6	173.6	158.2	20.2	186.2	52.0	33.5	26.1	69.4
110	199.3	162.5	149.9	18.8	332.2	58.0	35.0	26.6	145.9
100	189.3	167.3	156.0	21.7	376.9	59.5	35.8	26.2	231.6

nitrogen, we next considered fast-exchange equilibrium models in which all models, regardless of protonation state, were paired and their populations optimized for best agreement with the experimental chemical shifts. These models were again ranked, and the results are summarized in Table S7. The best-fit, with a reduced-χ^2 of 1.7, is for the exchange between the protonated Schiff-base (61%) and the protonated phenolic oxygen (39%) forms (shifts summarized in Main Text Table 1). In this dynamic fast-exchange equilibrium, the proton merely changes position between the nitrogen and oxygen atoms in the internal hydrogen bond. The next-best exchange model has an unlikely proton exchange between the Schiff base and pyridine ring nitrogens and with a reduced-χ^2 of 5.1 does not conform to the experimental chemical shifts within the expected error range.

Bayesian Probability Analysis

Following the approach of Engel et al., a Bayesian probability analysis was used to quantify the overall confidence in the identification of the experimental structure(27). In this approach, the posterior probability, $p\left(M \middle| y^*\right)$, that M is the experimental structure given the measured set of chemical shifts y^* is given as

$$p\left(M \middle| y^*\right) = \frac{p\left(y^* \middle| M\right) p(M)}{p\left(y^*\right)} = \frac{p\left(y^* \middle| M\right) p(M)}{\sum_{M'} p\left(y^* \middle| M'\right) p(M')}$$

Table 3.2: Reduced-χ^2 and relative probabilities for Bayesian probability analysis for the 7 best-fit fast-exchange models

Model	Red-χ^2	$p\left(y^* \middle| M\right)$ (relative)	$p(M)$ (relative)	$p\left(M \middle| y^*\right)$ (absolute)
61% 001 + 39% 010	1.7	1	1	0.852
76% 001 + 24% 100	5.1	0.093	1	0.0794
71% 001 + 29% 110	5.5	0.077	1	0.0662
50% 101 + 50% 010	20.6	0.0020	1	0.00167
78% 001 + 22% 101	28.2	0.00073	0	0
74% 101 + 26% 110	67.4	0.000041	1	0.0000353
82% 101 + 18% 100	69.9	0.000037	1	0.0000312

where $p(y^*|M)$ is the probability of observing chemical shifts y* for a given structure M, and $p(M)$ is the prior probability of structure M matching the experimental structure. The probabilities $p(y^*|M)$ were determined relative to the probability of the best-fit (BF) model based on the one-sided p-value (z) from the corresponding F-distribution evaluated at the critical value set by the ratio of the reduced-χ^2 for model M and the best fit model (36)

$$\alpha = \frac{\chi_M^2}{\chi_{BF}^2}$$

The p-value was then converted into a relative probability for model M according to

$$\frac{p(y^*|M)}{p(y^*|M) + p(y^*|M^{BF})} = z$$

using a fixed relative probability $p(y^*|M^{BF}) = 1$. Reduced-χ^2 for the 7 best exchange models are given in Table S7, along with their relative probabilities, $p(y^*|M)$.

The prior probability $p(M)$ was used to introduce consistency with the observed temperature dependence of the Schiff base nitrogen. Temperature dependence at this site requires exchange models that include tautomeric exchange at the

Schiff base nitrogen. These models were assigned relative probabilities $p(M)=1$ while all other models were assigned $p(M)=0$

Taking into account both the isotropic chemical shift data on the βLys87 and the PLP cofactor, along with the observed temperature dependence of the Schiff base nitrogen, the proposed exchange model is found to be the most probable experimental structure with 85% confidence.

Section 12: External contributions to Proton Exchange

The free-energy difference between the exchanging PSB and phenolic tautomers was determined by fitting the experimental chemical shift to a fast-exchange average shift for the phenolic and PSB states

$$\langle \nu \rangle = p_{\mathrm{PSB}} \nu_{\mathrm{PSB}} + p_{\mathrm{Phen}} \nu_{\mathrm{Phen}}$$

where the populations were used to define an equilibrium constant, K, and a change in enthalpy and entropy for the exchange

$$K = \frac{p_{\mathrm{Phen}}}{p_{\mathrm{PSB}}} = \exp\left[\frac{-(\Delta H - T\Delta S)}{RT} \right]$$

In this fit, ν_{Phen} and ν_{PSB} were set to their theoretical values of 319.3 ppm and 166.8 ppm, respectively. The best fit to the data gives $\Delta H = 9.6 \pm 1$ kJ/mol and $\Delta S = 28.6 \pm 3$ J/(mol K) (Main Text Fig. 6C).

To estimate the activation energy for the proton exchange, the linewidth of the Schiff-base nitrogen was fit to the two-site exchange model developed by Tian et al. (37) for the residual linewidth (in Hz) in the fast-exchange limit:

$$\Delta \nu_{1/2} = \frac{1}{\pi T_2} + \frac{4\pi K^2 \left(\nu_{\mathrm{PSB}} - \nu_{\mathrm{Phen}}\right)^2}{(1+K)^3 k_f}.$$

In this expression, $1/T_2$ is the spin-spin relaxation rate, ν_{PSB} and ν_{Phen} are the limiting chemical shifts (in Hz) of the Schiff-base nitrogen in the PSB and phenolic tautomeric forms, K is the equilibrium constant for the exchange given above, and k_f is the forward exchange rate (PSB to Phen). A limiting linewidth in the fast-exchange limit of 20 Hz (T_2=16 ms) was assumed. Any contribution due to direct spin-spin interaction and anisotropic bulk magnetic susceptibility, adding to the observed linewidth, are unknown and were ignored for simplicity. The exchange rate can be related to the free energy of activation by Eyring's equation,

$$k_f = \frac{k_B T}{h} e^{-\Delta G^{\neq}/RT},$$

where k_B is Boltzmann's constant, h is Plank's constant, and R is the gas constant, to give

$$\Delta \nu_{1/2} = \frac{1}{\pi T_2} + \frac{4\pi K^2 \left(\nu_{\mathrm{PSB}} - \nu_{\mathrm{Phen}}\right)^2}{(1+K)^3} \frac{h}{k_B T} e^{\Delta G^{\neq}/RT}.$$

To fit the data, ν_{Phen} and ν_{PSB} were again set to their theoretical values of 319.3 ppm and 166.8 ppm (and converted to Hz) and K calculated at each temperature based on the free energy difference determined above. The best fit to the data gives a forward rate of 2.4 x 10^6 /s and a free energy of activation of 36.8 kJ/mol (Fig. S13).

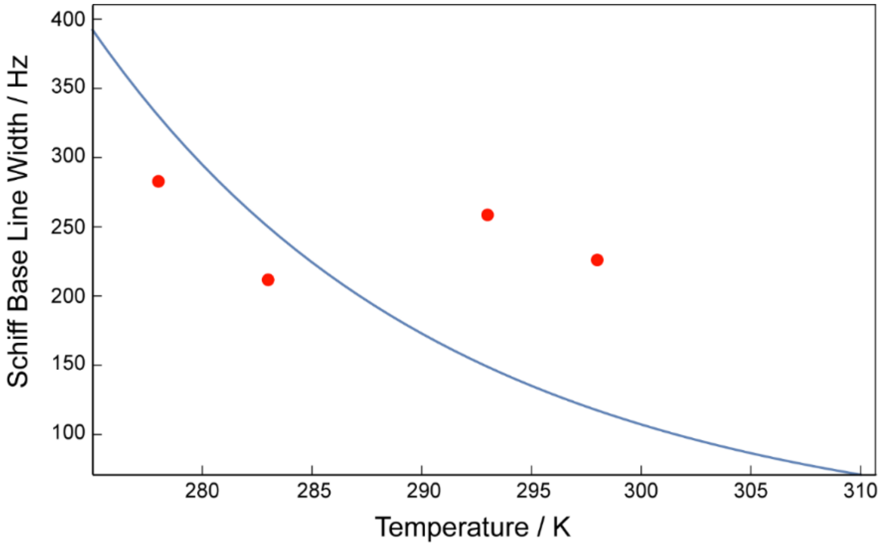


Figure 3.4**.** Temperature dependence of the βK87 Nζ linewidth. Opposed to the 1.3 mm rotor used in the remainder of the study, temperatures down to 5 °C could be achieved using a perdeuterated sample in a 0.7 mm rotor spun at only 55 kHz, half the speed usually employed and needed for fully protonated samples but also used for the 1.3 mm measurements. This procedure comes at the cost of a 50% sensitivity drop due to the smaller volume, which – in conjunction with the exchange-broadened lines – results in relatively poor signal-to-noise ratio for the assessment, prohibiting a more detailed assessment of enthalpic and entropic contributions to the free energy of activation. The general linewidth, however, unambiguously points to an exchange on the 1 µs timescale, which number is hardly affected by assuming a limiting linewidth of 20, 40, or 100 Hz. The data was fit to a two-site exchange model between the PSB and phenolic forms to extract an activation barrier $\Delta G^{\ddagger} = 36.8 \pm 0.5$ kJ mol^{-1} for an assumed limiting linewidth of 20 Hz.

References

1. H. Hayashi, Pyridoxal Enzymes: Mechanistic Diversity and Uniformity. The Journal of Biochemistry **118**, 463-473 (1995).

2. A. C. Eliot, J. F. Kirsch, Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. Annu Rev Biochem **73**, 383-415 (2004).

3. A. Mozzarelli, S. Bettati, Exploring the pyridoxal 5'-phosphate-dependent enzymes. Chem Rec **6**, 275-287 (2006).

4. C. Walsh, Enzymatic Reaction Mechanisms (W. H. Freeman and Company, San Francisco, 1979), pp. 978.

5. A. Sarup, O. M. Larsson, A. Schousboe, GABA transporters and GABA-transaminase as drug targets. Curr Drug Targets CNS Neurol Disord **2**, 269-277 (2003).

6. K. A. Abrahams et al., Inhibiting mycobacterial tryptophan synthase by targeting the inter-subunit interface. Sci Rep **7**, 9430 (2017).

7. D. Niks et al., Allostery and substrate channeling in the tryptophan synthase bienzyme complex: evidence for two subunit conformations and four quaternary States. Biochemistry **52**, 6396-6411 (2013).

8. K. Michalska et al., Allosteric inhibitors of Mycobacterium tuberculosis tryptophan synthase. Protein Sci **29**, 779-788 (2020).

9. J. F. Rocha, A. F. Pina, S. F. Sousa, N. M. F. S. A. Cerqueira, PLP-dependent enzymes as important biocatalysts for the pharmaceutical, chemical and food industries: a structural and mechanistic perspective. Catal Sci Technol **9**, 4864-4876 (2019).

10. D. K. Romney, J. Murciano-Calles, J. E. Wehrmuller, F. H. Arnold, Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. J. Am. Chem. Soc. **139**, 10769-10776 (2017).

11. M. Dick, N. S. Sarai, M. W. Martynowycz, T. Gonen, F. H. Arnold, Tailoring Tryptophan Synthase TrpB for Selective Quaternary Carbon Bond Formation. J. Am. Chem. Soc. **141**, 19817-19822 (2019).

12. B. G. Caulkins et al., Catalytic roles of βLys87 in tryptophan synthase: ^{15}N solid state NMR studies. BBA - Protein Proteom **1854**, 1194-1199 (2015).

13. B. G. Caulkins et al., NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. J. Am. Chem. Soc. **138**, 15214-15226 (2016).

14. B. G. Caulkins et al., 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. **136**, 12824-12827 (2014).

15. E. H. Cordes, W. P. Jencks, Semicarbazone Formation from Pyridoxal, Pyridoxal Phosphate, and Their Schiff Bases. Biochemistry **1**, 773-& (1962).

16. H. Singh et al., Fast Microsecond Dynamics of the Protein-Water Network in the Active Site of Human Carbonic Anhydrase II Studied by Solid-State NMR Spectroscopy. J. Am. Chem. Soc. **141**, 19276-19288 (2019).

17. A. Bertarello et al., Picometer Resolution Structure of the Coordination Sphere in the Metal-Binding Site in a Metalloprotein by NMR. J. Am. Chem. Soc. **142**, 16757-16765 (2020).

18. L. J. Mueller, M. F. Dunn, NMR crystallography of enzyme active sites: probing chemically detailed, three-dimensional structure in tryptophan synthase. Accounts of Chemical Research **46**, 2008-2017 (2013).

19. J. B. Holmes et al., Imaging active site chemistry and protonation states: NMR crystallography of the tryptophan synthase alpha-aminoacrylate intermediate. Proc Natl Acad Sci U S A **119** (2022).

20. D. Bryce, NMR crystallography: structure and properties of materials from solid-state nuclear magnetic resonance observables. IUCrJ **4**, 350-359 (2017).

21. J. C. Facelli, D. M. Grant, Determination of molecular symmetry in crystalline naphthalene using solid-state NMR. Nature **365**, 325-327 (1993).

22. P. Hodgkinson, NMR crystallography of molecular organics. Progress in Nuclear Magnetic Resonance Spectroscopy **118-119**, 10-53 (2020).

23. H. L. Zhang et al., HIV-1 Capsid Function Is Regulated by Dynamics: Quantitative. Atomic-Resolution Insights by Integrating Magic-Angle-Spinning NMR, QM/MM, and MD. J. Am. Chem. Soc. **138**, 14066-14075 (2016).

24. J. A. Gascon, E. M. Sproviero, V. S. Batista, QM/MM Study of the NMR Spectroscopy of the Retinyl Chromophore in Visual Rhodopsin. J Chem Theory Comput **1**, 674-685 (2005).

25. C. Bonhomme et al., First-principles calculation of NMR parameters using the gauge including projector augmented wave method: a chemist's point of view. Chem Rev **112**, 5733-5779 (2012).

26. J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller, G. J. Beran, Benchmark fragment-based ^{1}H, ^{13}C, ^{15}N and ^{17}O chemical shift predictions in molecular crystals. Physical Chemistry Chemical Physics **18**, 21686-21709 (2016).

27. E. A. Engel et al., A Bayesian approach to NMR crystal structure determination. Physical Chemistry Chemical Physics **21**, 23385-23400 (2019).

28. M. Chan-Huot et al., 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. **135**, 18160-18175 (2013).

29. S. Sharif, G. S. Denisov, M. D. Toney, H. H. Limbach, 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 **128**, 3375-3387 (2006).

30. U. Sternberg et al., (1)H line width dependence on MAS speed in solid state NMR - Comparison of experiment and simulation. J Magn Reson **291**, 32-39 (2018).

31. J. Herzfeld, A. E. Berger, Sideband intensities in NMR spectra of samples spinning at the magic angle. Journal of Chemical Physics **73**, 6021-6030 (1980).

32. J. Lai et al., X-ray and NMR crystallography in an enzyme active site: the indoline quinonoid intermediate in tryptophan synthase. J. Am. Chem. Soc. **133**, 4-7 (2011).

33. M. Baias et al., De novo determination of the crystal structure of a large drug molecule by crystal structure prediction-based powder NMR crystallography. J. Am. Chem. Soc. **135**, 17501-17507 (2013).

34. K. D. Schnackerz, B. Andi, P. F. Cook, ^{31}P NMR spectroscopy senses the microenvironment of the 5'-phosphate group of enzyme-bound pyridoxal 5'-phosphate. BBA - Protein Proteom **1814**, 1447-1458 (2011).

35. J. D. Hartman, T. J. Neubauer, B. G. Caulkins, L. J. Mueller, G. J. Beran, Converging nuclear magnetic shielding calculations with respect to basis and system size in protein systems. J. Biomol. NMR **62**, 327-340 (2015).

36. D. P. Shoemaker, C. W. Garland, Experiments in physical chemistry (ed. 2, 1967).

37. Y. Tian et al., 31P NMR investigation of backbone dynamics in DNA binding sites. J Phys Chem B **113**, 2596-2603 (2009).

Chapter 4 – The Aminoacrylate Intermediate of Tryptophan Synthase

Introduction

Pyridoxal-5′-phosphate (PLP; Fig. 1) participates in numerous enzyme-catalyzed reactions essential for amino acid metabolism, including transamination, decarboxylation, and α/β/γ-elimination and substitution(1-3). The power of PLP as a cofactor comes from its ability to act as an electron sink, allowing for the stabilization of carbanionic intermediates. A more subtle aspect of PLP chemistry demonstrated in β-elimination and replacement reactions is the ability of the cofactor to fine tune the polarity at the β-carbon of amino acids, facilitating the elimination of poor leaving groups and their replacement with weak nucleophiles.

Tryptophan synthase (TS) is the prototypical example of a PLP-dependent

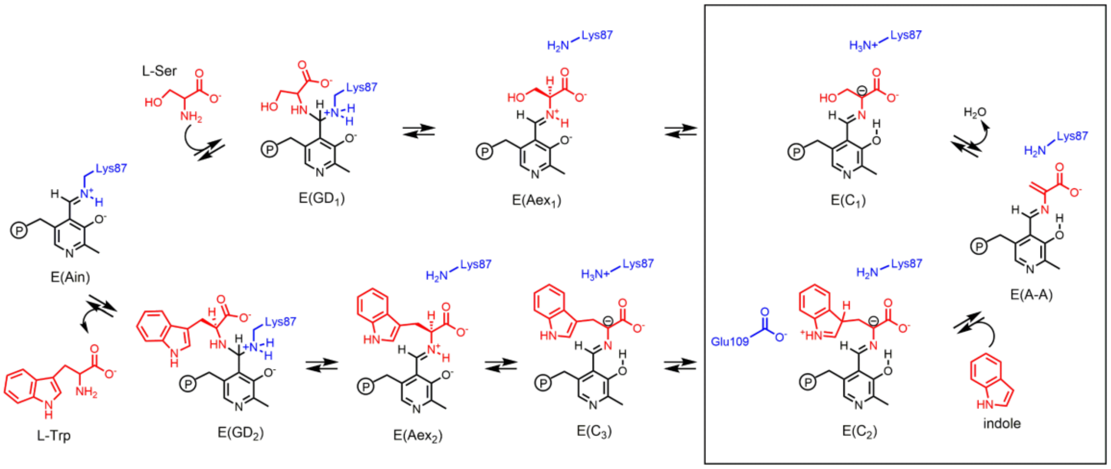


Figure 4.1: TS β-site reaction, highlighting the α-aminoacrylate intermediate (41, 82). The PLP cofactor is drawn in black. In stage I of the reaction, L-Ser reacts with the internal aldimine E(Ain) to give the gem-diamine, E(GD1), and external aldimine, E(Aex1), intermediates. Subsequent proton abstraction via βLys87 leads to the first carbanionic intermediate, E(C1), and elimination of the β-hydroxyl group as water gives the α-aminoacrylate intermediate, E(A-A). In stage II, indole makes a nucleophilic attack at Cβ of E(A-A) to yield a new C–C bond and the L-Trp carbanionic intermediate, E(C2), and, upon deprotonation, E(C3). In the final stages, E(C3) is reprotonated, leading to the eventual release of L-Trp and the regeneration of E(Ain).

enzyme that catalyzes β-elimination and substitution reactions. The Salmonella typhimurium tryptophan synthase (StTS) studied here is a 143 kDa, αββα bienzyme complex(5). TS performs the final two steps in the biosynthesis of L-tryptophan: the α-subunit cleaves indole-3-glycerol 3′-phosphate to glyceraldehyde-3-phosphate and indole, while the β-subunit catalyzes the PLP-dependent β-elimination and replacement of the substrate L-Ser hydroxyl with indole to produce L-Trp (Fig. 1) (6, 7). The fidelity of the proton transfers in the TS catalytic cycle are critical for maintaining the β-elimination and substitution pathway. In addition to the stereoelectronic control of the initial bond-breaking step(8), it has been proposed that reaction specificity in PLP-dependent enzymes is modulated by the protonation states of the PLP cofactor–substrate complex(2, 3, 9), which are in turn directed by chemical interactions with acid–base groups such as βLys87 and βGlu109 in TS. Hence, the catalytic residues interacting with the cofactor establish the appropriate chemical and electrostatic environment to favor a particular protonation state and reaction pathway.

The determination of protonation states remains a significant challenge to the tools of structural biology. Even high-resolution X-ray crystal structures are challenged to place hydrogen atoms. Neutron crystallography can locate hydrogen atoms, offering remarkable insights into enzyme mechanism(10, 11). However, the requirements for neutron crystallography include large, preferably perdeuterated crystals and multiweek-long acquisition times at room temperature. The latter currently precludes analysis of reactive intermediates such as those studied here. Cryogenic electron microscopy (cryo-EM) is also pushing to the boundaries of resolution necessary for hydrogen atom detection (12), an extraordinary achievement yet currently far from routine. In combined application with NMR spectroscopy, however, cryo-EM and diffraction methods continue to work toward complete atomic-resolution descriptions of structure and function(13-25). For delineating the chemistry of the active site, NMR and diffraction become even more powerful when combined with first-principles computational chemistry (26-32).

We are developing NMR-assisted crystallography—the joint application of solid-state NMR (SSNMR), X-ray crystallography, and first-principles computational chemistry—to solve for the chemically rich, three-dimensional structures of enzyme active sites(28-30) . By “chemically rich,” we mean structures in which the location of all atoms, including hydrogens, are specified. NMR crystallography was originally developed within the context of molecular organic and inorganic solids(33-39). Our group (28-30) and others(26, 31, 32) have been working to extend this approach to structural biology, in which it can provide consistent and testable models of enzyme structure and function. Our approach is threefold. First, X-ray crystallography is used to provide a structural framework for the active site. Second, chemically detailed models are built on this framework using computational chemistry, and various active site chemistries are explored. Third, these models are quantitatively distinguished by comparing their predicted NMR chemical shifts to the results of SSNMR experiments. Provided that a sufficient number of chemical shifts are measured, NMR-assisted crystallography can identify a unique, consistent structure or, equally important, determine that none of the candidates is consistent with the experimental observations. For the case of the “quinonoid” intermediate in tryptophan synthase, this approach demonstrated that the intermediate is better described as a carbanionic species with a deprotonated pyridine ring nitrogen—a structure that is fundamental to understanding reaction specificity in TS (30).

Here, NMR-assisted crystallography is used to characterize the TS α-aminoacrylate intermediate [E(A-A); Fig. 3.1], a species that marks a divergent step in PLP catalysis, as only enzymes that perform β-elimination reactions generate this intermediate. For E(A-A), NMR-assisted crystallography is able to identify the protonation states of ionizable sites on the cofactor, substrates, and catalytic side chains as well as the location and orientation of active site waters. From this, a detailed, three-dimensional picture of structure and reactivity emerges, highlighting βLys87 as the acid–base catalytic residue and delineating the reaction coordinate for the elimination of the substrate β-hydroxyl. Subsequent characterization of the Michaelis complex formed with the indole isostere benzimidazole (BZI) shows BZI bound in the active site and poised for, but unable to initiate, the subsequent bond formation step. When modeled into the BZI position, indole is properly aligned for nucleophilic attack. The chemically rich structure from NMR-assisted crystallography is key to understanding why BZI does not react, while indole does.

Results and Discussion

X-Ray Crystallography

X-ray crystal structures for the StTS E(A-A) and E(A-A)(BZI) complexes have been reported by our group (40) and others (41). The formation of E(A-A) is characterized by both the α- and β-subunits adopting closed conformations (7, 25, 42-45) . The crystal structures for E(A-A) show three crystallographic waters adjacent to the serine substrate in the β-subunit active site, forming a hydrogen bonded chain extending from the carboxylate of the catalytically essential βGlu109 (Fig. 2A). The position of the central water molecule is particularly striking, as it forms close contacts to both the substrate C^{β} (d_{CO} = 3.2 Å) and the βLys87 ε-amino group (d_{NO} = 3.0 Å). This places the water close to the site that the substrate β-hydroxyl is expected to occupy before it is eliminated. The binding of BZI (Fig. 2B) displaces the waters and causes a slight perturbation of the α-aminoacryloyl group. The

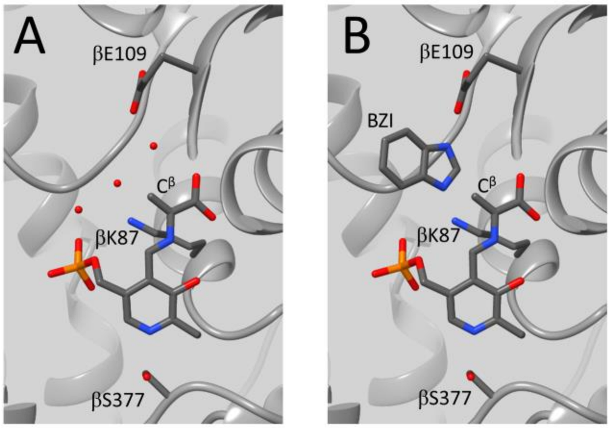


Figure 3.2: The β-subunit active sites for E(A-A) and E(A-A) (BZI) from crystal structures 2J9X and 4HPX, respectively. (A) The E(A-A) intermediate shows three active site waters adjacent to the substrate with the central water forming close contacts to both the substrate Cβ and the βLys87 ε-amino group. (B) The E(A-A)(BZI) complex shows BZI displacing the three waters but otherwise inducing only small changes in the active site structure. Images rendered in UCSF Chimera (83).

binding also induces small movements of the βGlu109 carboxylate and the βLys87 ε-amino groups so that they are within hydrogen bonding distances of the BZI nitrogen atoms. These interactions and the close structural similarity of BZI to indole make the BZI complex a good mimic for the expected alignment of indole for nucleophilic attack at C^{β}. Absent in Fig. 2, however, are the hydrogen atoms, which are critical to identifying the hydrogen bond donors and acceptors and their mechanistic roles.

SSNMR.

SSNMR experiments were performed on microcrystalline protein samples prepared in close analogy to the single crystals used for X-ray crystallography. The single crystals were grown slowly using sitting drop vapor diffusion, while the microcrystals were formed rapidly using a batch approach in which the protein and crystallization buffer solutions were mixed to give the same final reagent concentrations as the crystal plate reservoirs. Microcrystals of TS prepared in this manner have been shown to maintain the same crystal habit as the larger crystals (46), giving high confidence that the NMR and X-ray data can be directly compared.

Both E(A-A) and E(A-A)(BZI) can be observed with SSNMR under conditions that favor the accumulation of the enzyme-bound species while the microcrystals remain catalytically active (46-48). SSNMR spectra of E(A-A) formed by the reaction of L-Ser with microcrystalline TS are shown in Fig. 3. The signals in the cross-polarization magic-angle spinning (CPMAS) spectra correspond to crystalline protein and bound substrate. The spectra were acquired with various combinations of ^{15}N and ^{13}C labels on the PLP cofactor, L-Ser substrate, and protein Lys ε-amino groups. With unlabeled substrate and protein (Fig. 3D), the spectra show only unresolved background signals from the protein, including the

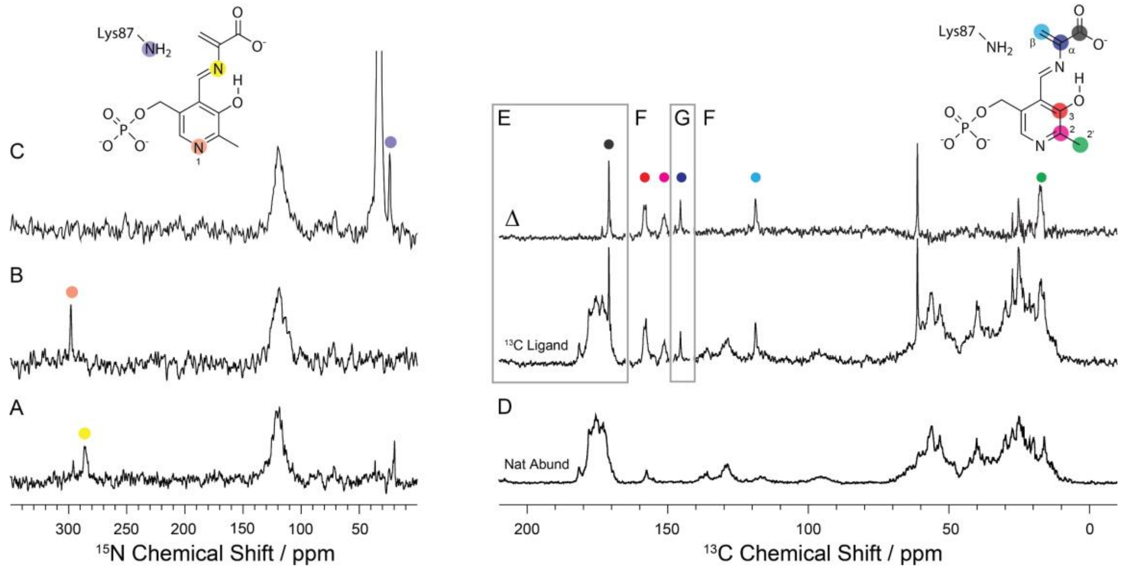


Figure 3.3: 15N and 13C CPMAS SSNMR spectra of microcrystalline TS E(A-A) prepared with the following isotopic labeling: (A) 15N-labeled on the substrate L-Ser; (B) 15N-enriched on the PLP cofactor; (C) 15N-enriched at protein lysine side-chain ε-amino groups; (D) natural abundance isotopomer concentration; (E) 13C-labeled at C′ of the L-Ser substrate; (F) 13C,15N-enriched on the PLP cofactor and Cβ of the substrate L-Ser; and (G) 13C-labeled at Cα of the substrate L-Ser. The top spectra in (E–G) are formed as the difference between the E(A-A) spectra with various cofactor/ligand isotopic labels and the same spectra acquired at natural abundance, emphasizing the resonances for the specific site labels. The large peak at 63.1 ppm is from free serine. Spectra acquired at 9.4 T, −10 °C, and 8 kHz MAS

backbone amide nitrogens (∼120 ppm) and backbone carbonyl (∼170 ppm), aliphatic (∼10 to 70 ppm), and aromatic (∼110 to 160 ppm) carbons.

With the magnification factor of the isotopic enrichment (34), distinct resonances for the active site species become visible. The spectra of the sample prepared with ^{15}N and ^{13}C on the PLP cofactor (Fig. 3 B and F) show a nitrogen resonance at 297.6 ppm and three additional carbon resonances at 17.5, 151.2, and 158.1 ppm that are assigned to the cofactor atoms N1, C2′, C2, and C3, respectively (atom labeling given in Fig. 3). The sample prepared using ^{13}C- and ^{15}N-L-Ser as a substrate (Fig. 3 A, E, F, and G) displays a nitrogen resonance at 286.7 ppm and additional carbon resonances at 170.9, 145.6, and 118.8 ppm that are assigned to the Schiff base nitrogen and carbons that derive from the serine C′, C^{α}, and C^{β}, respectively. The Schiff base nitrogen displays a significant temperature dependence of ∼−0.14 ppm/°C, while the PLP ring nitrogen shows no discernable temperature dependence. . The Schiff base nitrogen chemical shift tensor was measured under slow MAS conditions and the principal axis components, (δ_{11}, δ_{22}, δ_{33}) = (525 ± 11, 305 ± 10, 30 ± 14) ppm, were extracted by fitting the intensity of the spinning sideband manifold (49).

These shifts help establish several key elements of the chemical structure for E(A-A). At this point in the catalytic cycle, the serine substrate has lost its β-hydroxyl, and there is a double bond between C^{α} and C^{β}, which is confirmed by their chemical shifts of 145.6 and 118.8 ppm, respectively. Second, with a shift of 286.7 ppm, the Schiff base nitrogen is found to be neutral, which is further supported by the large span of its chemical shift tensor (50). This isotropic chemical shift, however, is below the limiting value for a fully neutral Schiff base, and the temperature dependence suggests tautomeric exchange involving transient protonation of this site (10, 29). At the same time, the PLP C2 and C3 shifts of 151.2 and 158.1 ppm indicate that the PLP phenolic oxygen is neutral (29). The PLP nitrogen shift of 297.6 ppm also indicates a neutral pyridine ring N1.

The ^{15}N side-chain chemical shift of βLys87 was measured on a TS sample in which all lysine residues were ^{15}N-labeled on the ε-amino group as reported previously (48). In the initial internal aldimine state, E(Ain), this sample shows a peak at 202 ppm, corresponding to the protonated Schiff base linkage between βLys87 and the PLP cofactor(51). Upon addition of L-Ser, this resonance is lost, and a new resonance at 24.2 ppm is observed (Fig. 3C), which is assigned to a neutral ε-amino group on βLys87.

The ^{31}P isotropic chemical shift and chemical shift tensor were also measured for the cofactor's phosphate group The phosphate chemical shift shows a characteristic response in going from the mono- to the dianionic charge state. The phosphate group's isotropic shift of 5.2 ppm and its shift tensor indicate that it is dianionic, in keeping with other TS intermediates (29, 51) and PLP-dependent enzymes (52).

Finally, the ^{17}O chemical shifts of the substrate carboxylate group were measured using ^{17}O quadrupole central transition (QCT) NMR in solution (53-55)

These were the only shifts measured for the intermediate in solution. ^{17}O NMR is not yet considered a standard high-resolution probe for biomolecular NMR, but QCT NMR takes advantage of the unique property of the NMR central transition to narrow as the size of the protein–substrate complex increases. We previously reported preliminary isotropic ^{17}O shifts for both E(A-A) and E(A-A)(BZI)(54). Here, an advantage is taken of additional measurements on the 35.2 T series connect hybrid magnet (56) at the National High Magnetic Field Laboratory to improve the accuracy of the extracted parameters. The ^{17}O chemical shifts of 257 and 289 ppm are consistent with an ionized carboxylate group. The up-field shifted signal is tentatively assigned to the oxygen hydrogen bonded to the side chain of βThr110, which forms the closest contact.

Table 4.1. E(A-A) and E(A-A)(BZI) experimental and first-principles chemical shifts (ppm) for the phenolic (Phen) and protonated Schiff base (PSB) species and their two-site exchange with the following populations: E(A-A) 89.3% Phen,10.7% PSB; E(A-A)(BZI) 89.4% Phen, 10.6% PSB

E(A-A)	Atom	Phen	PSB	Two-Site	Expt	E(A-A) (BZI)	Atom	Phen	PSB	Two-Site	Expt
PLP	N1	302.9	303.3	303.0	297.6	PLP	N1	310.2	307.3	309.9	302.4
	C2	148.7	159.1	149.8	151.2		C2	152.1	162.7	153.2	153.1
	C2'	18.8	19.8	18.9	17.5		C2'	19.4	20.7	19.6	18.2
	C3	154.9	173.2	156.9	158.1		C3	155.2	174.0	157.2	158
L-Ser	SB N	301.4	153.6	285.5	286.7	L-Ser	SB N	307.6	152.0	291.2	292.3
	Cα	144.4	133.1	143.2	145.6		Cα	147.2	136.3	146.0	146
	C'	171.2	168.7	171.0	170.9		C'	168.9	166.5	168.7	169.8
	Cβ	123.3	121.3	123.0	118.8		Cβ	120.6	118.2	120.3	118.7
	O1	262.1	261.4	262.0	257		O1	261.4	260.5	261.3	258
	O2	302.4	294.3	301.5	289		O2	301.4	291.8	300.4	286
βK87	Nζ	20.5	21.1	20.6	24.2	βK87	Nζ	35.6	36.4	35.6	35.6
						BZI	N1	167.5	167.9	167.5	165.5
							N3	231.4	231.7	231.4	227.8
	red-χ^2	3.21	106.1	1.85			red-χ^2	2.19	97.8	0.90	

Chemical shifts for the E(A-A)(BZI) complex were measured analogously and are summarized along with the E(A-A) shifts in Table 1. The spectra for E(A-A)(BZI) show two additional nitrogen resonances at 165.5 and 227.8 ppm that are assigned to N1 and N3 of bound BZI, respectively (atom labeling in Table 1). The spectra are otherwise similar to those for E(A-A) with one notable exception: the loss of the neutral amino resonance at 24.1 ppm and the appearance of an additional charged amino group at 35.6 ppm. The latter was assigned to the active site βLys87 based on $^{15}N\{^{31}P\}$ rotational echo double-resonance (57) experiments that place it in close spatial proximity (<4 Å) to the phosphorus atom of the cofactor Thus, despite significant spectral overlap with the 26 other charged Lys side chains in TS, the catalytic active site residue can be uniquely identified. Similar strategies that make use of distinct chemical shifts, nuclides, and/or mixed labeling to assign and localize atoms within specific regions of proteins have been extensively demonstrated in SSNMR(58, 59).

First-Principles Calculations.

While the preliminary chemical structures shed some light on the chemistry of the active site, they are incomplete in two important respects. First, they lack detailed three-dimensional structure. Second, the chemical shifts do not fully conform to the expected limiting shifts from model compound studies. To interpret the experimental chemical shifts and place the chemistry of the active site in full structural context, we turn to first-principles computational chemistry.

The state-of-the-art for first-principles chemical shift calculations is advanced (27, 60-68), and extensive benchmarking and testing have shown that if the correct structure is known down to the position of every atom, then the NMR chemical shifts can be predicted better than 1.5 ppm RMSD for carbon, 4.3 ppm for nitrogen, and 7.5 ppm for oxygen (66, 68). This expected accuracy enables screening protocols to be established in which proposed structures can be evaluated and ranked for consistency with the experimental chemical shifts. Here, this agreement is quantified using the reduced χ^2 statistic, the squared deviation of the predicted and experimental chemical shifts weighted by the nuclide-dependent mean squared deviations. The benchmark for a solved structure is taken as the identification of a single structure (or a fast-exchange equilibrium between closely related tautomers) that satisfies the 95% confidence limits of the reduced χ^2. To fully distinguish proposed models typically requires 10 or more chemical shifts throughout the active site, supplemented by chemical shift tensor measurements and chemical shift temperature coefficients (30).

To limit bias and broadly consider potential active site chemistries, a pool of candidate structures was generated by systematically varying the protonation states of the cofactor, substrates, and βLys87 ε-amino. The candidate structures were constructed directly as three-dimensional cluster models of the active site, built on the framework of the X-ray crystal structure, and included all residues and crystallographic waters within 7 Å of the substrate and cofactor. The atoms at the exterior of the cluster were fixed at their crystallographic locations, and quantum-mechanical geometry optimization and chemical shift calculations were performed using density functional theory. The clusters contained 703 to 706 atoms (Fig. 4B), a size for which the convergence and accuracy of the chemical shift calculations have been established (68).

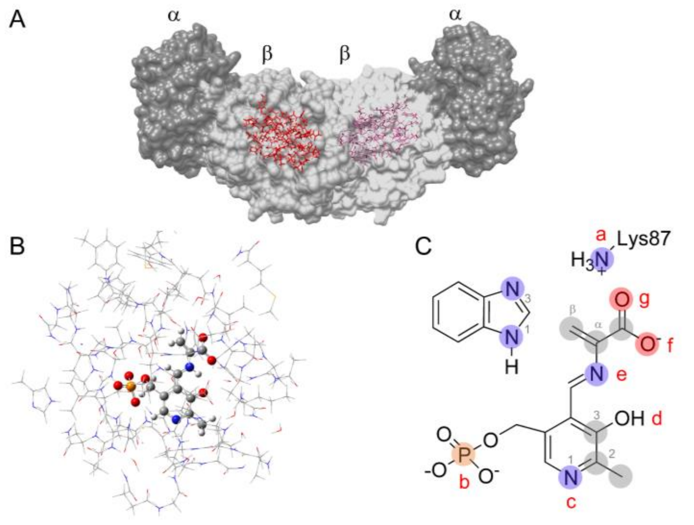


Figure 4.4: Cluster model of the E(A-A) active site. (A) X-ray crystal structure of the TS α2β2 heterodimer with the β-subunit active site highlighted in red. (B) Cluster model of the active site for first-principles geometry optimization and chemical shift calculations with protein side chains displayed in wireframe and cofactor and substrate in ball and stick. (C) Protonation sites on or near the cofactor/substrate complex: [a] the βLys87 side chain, [b] the PLP phosphate group, [c] the PLP pyridine ring nitrogen, [d] the PLP phenolic oxygen, [e] the Schiff-base nitrogen, and [f/g] the substrate carboxylate. Shaded nuclei indicate sites for which experimental NMR chemical shifts are reported.

Thirty-five initial models for E(A-A) and E(A-A)(BZI) were generated (SI Appendix, Schemes S1 and S2), their geometries were optimized, and their chemical shifts were predicted. The structures were then ranked based on their agreement with the experimental chemical shifts as shown in Fig. 5 A and C. There is a clear differentiation of models for both E(A-A) and E(A-A)(BZI), yet the reduced χ^2 values for all candidate structures fall outside of the 95% confidence limits, indicating that none reproduce the experimental chemical shifts with the expected accuracy. One of the largest discrepancies for all models occurs for the Schiff base nitrogen, in which the closest predicted shifts are still greater than three standard errors from the experimental value.

Motivated by the temperature dependence of the Schiff base nitrogen, a two-site, fast-exchange equilibrium model was considered next in which the effective chemical shift is given as the population-weighted average of the individual shifts. To remain unbiased, all structures that differed by the position of a single hydrogen atom were paired, and their populations were optimized for best overall agreement with the experimental data. These models were again ranked and are summarized in Fig. 5 B and D. Multiple models now satisfy or come close to satisfying the 95% confidence limits. To differentiate these models, the predicted principal components for the Schiff base nitrogen chemical shift tensor were also compared to the experimental values. The chemical shift tensor is extremely sensitive to dynamics, as it averages in both magnitude and orientation. For both E(A-A) and E(A-A)(BZI), only a single tautomeric exchange model simultaneously satisfies the isotropic and tensor chemical shift restraints, indicating that the benchmark for a solved structure has been reached.

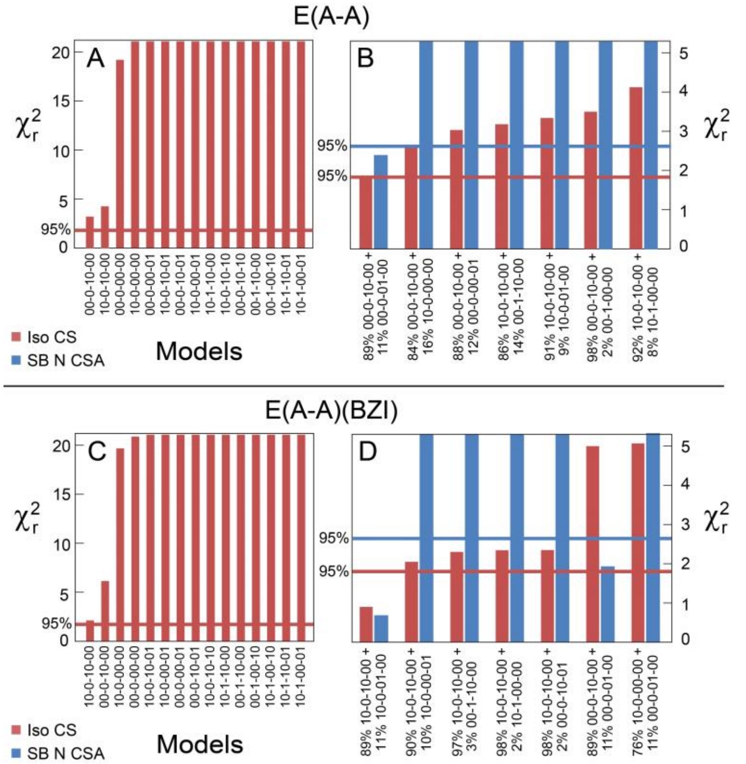


Figure 4.5 Ranking of structural models based on agreement between the experimental and first-principles chemical shifts as quantified by the reduced χ2 statistic. (A) The 15 best geometry-optimized active site models for the E(A-A) intermediate; structures and labeling are given in SI Appendix, Schemes S1 and S2. (B) Rankings of the seven best fast-exchange equilibrium models comparing the experimental and first-principles isotropic chemical shifts (red) and Schiff base nitrogen tensor components (blue). 95% confidence limits are shown as the correspondingly colored horizontal bars. (C and D) Model rankings for the E(A-A)(BZI) complex. For both E(A-A) and E(A-A)(BZI), only a single tautomeric exchange model simultaneously satisfies the isotropic and tensor chemical shift restraints.

For both E(A-A) and E(A-A)(BZI), the best-fit equilibrium is between the phenolic (enolimine; 89%) and protonated Schiff base (ketoenamine; 11%) tautomers, with proton exchange across the internal hydrogen bond (Table 1, center inset). The chemical shifts for the best-fit models are summarized in Table 1 along with the experimental shifts. An analogous exchange was also found for the TS carbanionic intermediate (29) and appears to be a common feature in many PLP-dependent enzymes (2, 9). The only difference in protonation states between the E(A-A) and E(A-A)(BZI) structures is for the βLys87 ε-amino group: for E(A-A) it is neutral, while for E(A-A)(BZI) it is positively charged.

To further test the exchange model, the ^{15}N chemical shift of the Schiff base nitrogen was measured at 95 K under conditions of dynamic nuclear polarization At this temperature, the tautomeric equilibrium is expected to shift predominantly to the phenolic form with less than 1% of the exchange partner present. This was confirmed experimentally, with chemical shifts of 301.4 ppm for E(A-A) and 302.1 ppm for E(A-A)(BZI) at 95 K compared to the predicted shifts for the individual structures of 302.9 and 310.2 ppm, respectively. Thus, both the major tautomers and the dynamic equilibrium of these intermediates can be established.

As a final assessment, the overall confidence in the identification of the experimental structure was quantified using the Bayesian probability analysis following the approach of Engel et al. (69) as described in Methods. Taking into account both the isotropic and anisotropic chemical shifts along with the observed temperature dependence of the Schiff base nitrogen, the proposed exchange model is found to be the most probable experimental structure with 88.5% confidence for E(A-A) and 99.6% confidence for E(A-A)(BZI). The decreased confidence for E(A-A) can be traced back to the lower overall agreement between the predicted and experimental shifts for the best-fit model as reflected in the reduced χ^2. Dynamics of the water molecules adjacent to the substrate may explain why the overall agreement between theory and experiment is lower in this case.

NMR-Assisted Crystallography.

What emerges from the application of NMR-assisted crystallography to the α-aminoacrylate intermediate is a chemically detailed view of the enzyme active site in which the locations of all atoms, including hydrogens, have been specified (Fig. 6). The quality of these structures can be judged by several metrics. First, the refined heavy atom positions are found to agree well with the original X-ray crystal coordinates, with an RMSD of 0.077 Å for E(A-A) and 0.091 Å for E(A-A)(BZI). Second, the positional uncertainties in the structures from NMR-assisted crystallography can be estimated following the method of Hofstetter and Emsley (70). This approach correlates changes in the predicted shifts with displacements in the structural coordinates, defining a positional uncertainty that is consistent with the experimental shifts. Average positional RMSD of 0.11 Å for heavy atoms

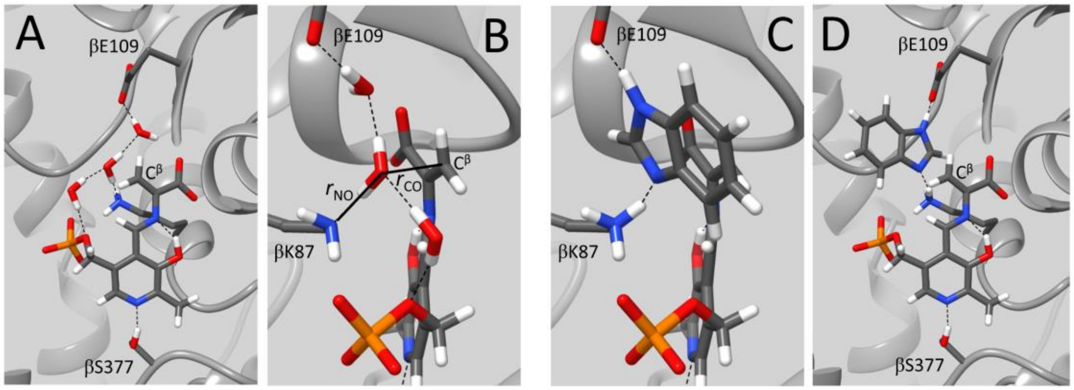


Figure 4.6: Protonation states, hydrogen bonding interactions, and the placement and orientation of structural waters as revealed by NMR-assisted crystallography in the TS β-subunit active site. (A and B) The E(A-A) intermediate and (C and D) the E(A-A)(BZI) complex. For E(A-A), the position and orientation of the central water molecule highlights the reaction coordinate for the loss of the serine β-hydroxyl and confirms βLys87 as the active site acid–base catalytic residue. The E(A-A)(BZI) complex shows BZI bound in the active site with hydrogen bonding interactions to βGlu109 and the charged ε-amino group of βLys87. BZI is poised for, but unable to initiate, the subsequent bond formation step. Images rendered in

and 0.17 Å for hydrogen atoms are found for the cofactor and substrates in both E(A-A) and E(A-A)(BZI). These positional uncertainties are similar in size to those found in organic molecular crystals (39, 70, 71) and are 6.5 times smaller than the heavy atom uncertainties reported in the corresponding protein X-ray crystal structures. While preliminary, these results suggest that positional uncertainties in NMR crystallography are independent of molecular size and that they will depend primarily on the nature and number of the measured NMR observables.

Chemically Rich Crystal Structures and Mechanistic Implications.

Perhaps most striking in the structure of the E(A-A) active site is the position of the central water molecule immediately adjacent to the serine substrate C^{β} and oriented with one hydrogen pointing toward the ε-amino group of βLys87 (Fig. 6B). This water is located 2.9 Å from C^{β} and is perfectly aligned for the reverse nucleophilic attack to regenerate $E(C_1)$. The interaction of this water with βLys87 and its proximity to the C^{β} carbon support the hypothesis that βLys87 is the acid catalyst that protonates the β-hydroxyl during C–O bond scission as $E(C_1)$ is converted to E(A-A) and a water molecule. We postulate this site is the initial binding pocket for the eliminated water molecule.

Critical examination of the crystal structures for E(A-A), however, shows that this central water is not consistently present. The electron density maps for 2J9X and 4HN4 (both with Cs^+ in the monovalent cation-binding site) show significantly lower electron density for this water and likely reflect occupancy less than 50%. To explore this variation, two additional E(A-A) crystal structures for the NH_4^+ and Cs^+ forms were solved at 1.40 and 1.50 Å, respectively (Protein Data Bank identification codes 7MT4 and 7MT5). The diffraction show well-defined electron density for the central water in the NH_4^+ form but an absence of this density in the Cs^+ structure (SI Appendix, Fig. S11). Close inspection of the NMR data for E(A-A) finds that each of the ^{13}C resonances of the bound serine is accompanied

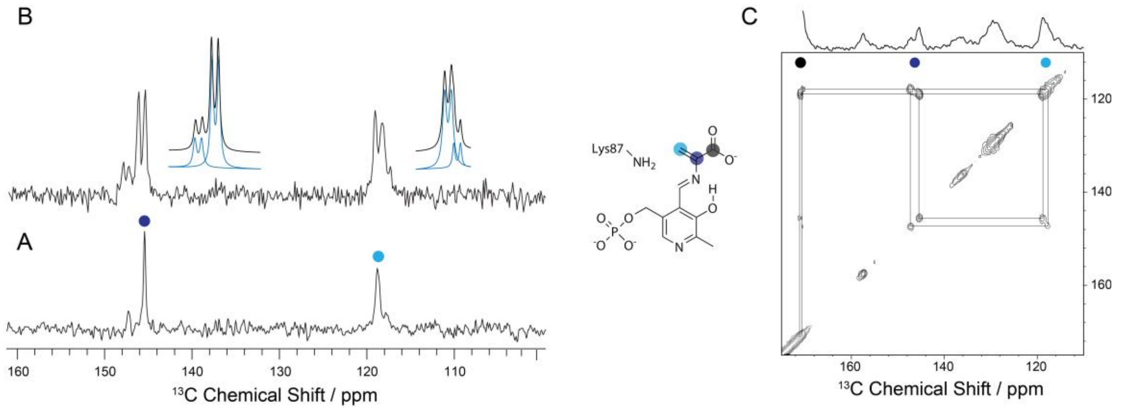


Figure 4.7: 13C SSNMR CPMAS spectra of the microcrystalline TS E(A-A) intermediate prepared with (A) a 50:50 mixture of singly-labeled 2-13C-L-Ser and singly-labeled 3-13C-L-Ser and (B) doubly labeled 2,3-13C2-L-Ser. For both, the natural abundance 13C protein background has been subtracted as in Fig. 3. The Cα and Cβ resonances show a minor peak that, along with the major peak, splits into a doublet upon incorporation of doubly labeled L-Ser substrate (deconvolution shown adjacent to the spectrum). (C) The 2D dipolar driven 13C correlation spectrum of the E(A-A) formed with U-13C3-L-Ser displays distinct crosspeaks for the major and minor peaks, indicating that they belong to two independent E(A-A) species. The small chemical shift changes for the minor peaks correlate with the first-principles predicted shifts in the absence of the central water in E(A-A), consistent with the variable occupancy of this site across X-ray crystal structures. 1D spectra acquired as in Fig. 3; 2D Spectra acquired using CORD mixing (4) on a Bruker BioSolids CryoProbe™ (5) at

by slightly shifted (±1 to 2 ppm) minor species of ~20 to 30% intensity, while those for E(A-A)(BZI) show no such peaks. These resonances are most easily observed with single site ^{13}C serine labels (Fig. 7A) and display the same multiplet pattern as the major resonances upon the incorporation of 2,3-$^{13}C_2$-L-Ser. The minor and major resonances also show distinct sets of cross-peaks (minor to minor, major to major) in the two-dimensional (2D) dipolar-driven correlation experiment acquired with a CPMAS cryoprobe to improve sensitivity (Fig. 7C) (72, 73). We hypothesize that the major and minor resonances belong to two independent E(A-A) species that derive from varying occupancy of the central water position across the macroscopic crystals. To test this, geometry-optimized active site models for the primary phenolic and protonated Schiff base tautomers were constructed in which the water nearest C^β was removed. Their predicted chemical shifts track the experimental chemical shifts of the minor species, and these secondary peaks are therefore tentatively assigned to a minor population in which the central water is no longer bound. Taken together, the SSNMR, X-ray, and computational results are consistent with this site being a weak binding site for water.

The E(A-A)(BZI) complex shows a similarly detailed view of the active site chemistry for the subsequent mechanistic step, nucleophilic addition at the substrate C^β. BZI is an isostere of indole and a potent inhibitor of the β-reaction(40, 74-78). The view from NMR-assisted crystallography (Fig. 6 C and D) shows BZI bound in the active site, displacing the three waters, and poised for nucleophilic attack. Despite being more nucleophilic than indole, BZI does not react to form a covalent bond. A comparison to crystal structures for the carbanionic intermediate analogs formed with indoline and 2-aminophenol (29) makes it clear that BZI is bound to the same subsite. Thus, this intermediate analog appears to be poised close to the transition state between two reactive intermediates in the TS catalytic cycle.

That BZI is not a substrate for TS and does not covalently react with E(A-A) can be attributed to two important factors: the different nucleophilic reaction mechanisms of indole and BZI, and the tight packing and hydrogen bonding interactions within the indole subsite. The mechanisms of nucleophilic reaction for indole and BZI are fundamentally distinct, as BZI reacts via a nitrogen lone pair, and indole reacts via the pi system at C3 (74). NMR-assisted crystallography identifies that it is the nucleophilic N3 that is adjacent to the substrate C^{β} in the E(A-A)(BZI) complex. However, this structure shows that the BZI lone pair does not point toward the E(A-A) C^{β} p-orbital. The tight packing of atoms within the indole subsite, the hydrogen bond between N3 and the charged ε-amino group of βLys87, and the hydrogen bond between the βGlu109 carboxylate and N1 of BZI preclude rearrangement of BZI within the site to allow the reaction.

In contrast, when indole is modeled in place of BZI, the C3 carbon of the five-membered ring is perfectly aligned to form the new C–C bond. The attacking p-orbital on indole points toward the electron-deficient C^{β} carbon and is poised to make orbital overlap as the complex moves along the reaction coordinate to the sp^3 geometry of the product. At the same time, N1 of indole is positioned to hydrogen bond to the carboxylate of βGlu109, which stabilizes positive charge development on N1 as the transition state is approached. The C3 carbon of indole is relatively electron rich but still requires assistance in the attack on C^{β}. The critical role played by βGlu109 can be seen in the βGlu109Asp mutation, which reduces the β reaction rate 27-fold (76). We posit that the E(A-A)(BZI) complex models how indole is bound to E(A-A) just prior to C–C bond formation.

Conclusions

The integrated application of SSNMR spectroscopy, X-ray crystallography, and first-principles computational chemistry offers unprecedented, chemically rich views of the TS E(A-A) and E(A-A)(BZI) active sites. Through the combined determination of heavy atom positions, the identification of protonation states, and the placement and orientation of active site waters, an exceptionally detailed picture of structure and reactivity emerges. Through this lens, the E(A-A) intermediate shows a central water that is well positioned for the reverse nucleophilic attack on C^{β}, with an orientation that points back to the active site acid–base catalytic residue βLys87. This chemically detailed view highlights the reaction coordinate for the elimination of the substrate β-hydroxyl and the initial binding pocket for the eliminated water. Both X-ray crystallography and SSNMR indicate variable occupancy for this site. Upon reaction of E(A-A) with BZI, the three waters adjacent to the substrate are displaced, and NMR-assisted crystallography shows BZI occupying the indole binding pocket but unable to react. Here, the protonation states complete the chemical picture for why BZI is unable to initiate the next step in the reaction: despite being a good nucleophile, BZI is held in the wrong orientation by hydrogen bonds to βGlu109 and the charged ε-amino group of βLys87.

NMR crystallography takes advantage of one of the well-established strengths of NMR spectroscopy—remarkable sensitivity to chemical structure and chemical dynamics. This is, we would argue, where NMR will continue to interface most strongly with the other tools of structural biology, including X-ray and neutron crystallography and cryo-EM. When combined with first-principles computational chemistry, these complementary techniques can build consistent, testable models of structure and reactivity in enzyme active sites. Importantly, this can be accomplished for samples near room temperature and under conditions of active catalysis.

Materials and Methods

First-Principles Calculations

First principles calculations were performed using a cluster-based model of the active site as described previously (29). The clusters for E(A-A) and E(A-A)(BZI) were constructed from the corresponding crystal structures 4HN4 and 4HPX by selecting all atoms within 7 Å of the cofactor-substrate complex (Fig. S1). This selection was expanded to include complete residues and were modified as follows: (1) residues that were not part of continuous backbone segments and with only two atoms within the initial 7 Å cut were deleted (Q142, V201, I238, D383); (2) residues with two backbone atoms and no side chain residues within 7 Å were converted to alanine (N145, K167, D305, S308, V309, L349, N375, L376, R379); (3) residues with two side chain atoms and no backbone atoms within 7 Å were truncated by removing the backbone atoms and replacing Cα with a methyl group (β-site residues F280, H313, K382); (4) N-terminal nitrogen atoms were replaced with a hydrogen atom and C-terminal carbonyls were capped with an –NH2 group; (5) the Cs+ cation was replaced with Na+ ; and (6) hydrogen atoms were added. The final model for E(A-A) included 58 amino acid residues, the PLP-serine complex, and 13 water molecules (waters 511, 515, 516, 519, 535, 538, 559, 570, 583, 604, 744, 844, 914) and had a total of 703 atoms. The final model for E(A-A)(BZI) included the same residues, but also included BZI and only 10 water molecules (waters 505, 511, 513, 516, 520, 541, 549, 683, 688, and 710). The structures were further modified depending on the protonation states of the cofactorsubstrate complex and K87.

Within this cluster, candidate structures were systematically generated by varying the protonation of the following seven ionizable sites on or near the PLP-ligand complex (Fig. S1C): the ε-amino group of βLys87, the PLP phosphate group, the pyridine nitrogen, the pyridoxal oxygen, the Schiff base nitrogen, and both carboxylate oxygen atoms; the protonation states of the two BZI nitrogen atoms were also changed for the relevant structures. For both E(A-A) and E(A-A)(BZI), the phosphate group was taken to be dianionic based on its CSA tensor, and models that had more than a single proton placed at either the pyridoxal oxygen or the Schiff base nitrogen were not considered, nor were structures with a doubly protonated carboxylate. Based on its 15N Chemical shifts, BZI was modeled as neutral as shown in Fig. S1C and oriented with the protonated nitrogen adjacent to βGlu109. Initial structures with the protonated nitrogen directed toward a neutral βLys87 with βGlu109 protonated invariably resulted in a redistribution of hydrogen atoms during geometry optimization to give protonated βLys87 and ionized βGlu109. In sum, 35 variations of protonation state were constructed for E(A-A) and E(A-A)(BZI). These models were labeled using a binary code to indicate whether a site was protonated ("1" – yes, "0" – no) in the order: (Nε of βLys87)(phosphate group)-(pyridine nitrogen)-(pyridoxyl phenolic oxygen)(Schiff base nitrogen)-(nearer carboxylate oxygen to the Schiff base nitrogen)(farther carboxylate oxygen); these sites are designated as AB-C-DE-FG in Fig. S1 and shown explicitly in Schemes S1 and S2. Initial proton orientations in the active site clusters were set using molecular dynamics (MD) simulations of the αβ dimer. These simulations were performed using the Amber package with GPU acceleration (79, 80). Coordinates for StTS were obtained from crystal structures 4HN4 for E(A-A) and 4HPX for E(A-A)(BZI) with crystal waters retained (45). Amber force field FF14SB was applied to protein atoms (81). Ligands were parametrized using General Amber Force Field (GAFF) and AM1-bcc charge model (82, 83). Only hydrogens were minimized and the MD simulations were carried out using Generalized Born implicit solvent model

(GBNeck2) (84) at constant temperature of 298 K controlled by Langevin thermostat. Two MD trajectories for each system were collected over 50 ns at 1 ps interval with 2 fs time step. The trajectories were further processed with CPPTRAJ software to contain 5000 frames for analysis (85). The systems were visualized and analyzed using Visual Molecular Dynamics software (86). Dihedral data was collected for residues Thr110, Thr190, Ser235, Ser377, Ser351 and Tyr186 in both systems. In E(A-A), the conformation and orientation for the three water molecules WAT583, WAT604, and WAT914 were also analyzed. The initial conformation and the position of the polar hydrogens in the respective residues and water molecules were selected based on the mean value of the normally distributed dihedral data for each residue. All cluster models were subsequently geometry optimized in Gaussian09 (87) at the DFT B3LYP level of theory using the Grimme D3 empirical dispersion correction (87) and a two-tier, locally-dense basis set (66, 68, 88) with 6- 31G(d,p) for the PLP/substrate complex and 6-31G for all other atoms. Except as noted below, all atoms on or within 4 Å of the PLP/substrate complex and all hydrogen atoms in the cluster could adjust while the remaining atom coordinates were fixed at their crystallographic values. The only exceptions were the ε-amino group nitrogen of βLys87 and the oxygen atoms of the three waters adjacent to the serine moiety for the αaminoacrylate, which were also held fixed at their crystallographic coordinates during the geometry optimizations. S4 For each optimized structure, NMR shieldings were calculated at the DFT B3LYP level of theory, employing a three-tier, locally dense basis set with the PLP/substrate, βLys87, BZI, and proximal water atoms at 6-311+G(2d,p), atoms within 4Å of this subgroup at 6-311G(d,p), and all remaining atoms at 6-31G. NMR shielding values were converted to chemical shifts using the following linear rescaling relationships, derived at the same level of theory and bases (66, 68)

$$^{13}\text{C:}\ \delta\left[\mathrm{TMS}(l)\right] = 173.70 - 0.9685\,\sigma_{\mathrm{calc}}$$

$$^{15}\text{N:}\ \delta\left[\mathrm{NH_3}(l)\right] = \delta\left[\mathrm{NH_4Cl}(s)\right] + 39.27 = 230.45 - 0.9996\,\sigma_{\mathrm{calc}}$$

$$^{17}\text{O:}\ \delta\left[\mathrm{H_2O}(l)\right] = 266.32 - 1.0551\,\sigma_{\mathrm{calc}}$$

Previous benchmark studies across test sets of solid-state structures used to establish these relations, demonstrated root-mean-square errors (RMSE) for isotropic shifts to be 1.5 ppm for 13C, 4.3 ppm for 15N, and 7.5 ppm for 17O (66). RMSE for the chemical shift tensor components were 4.2 ppm and 13.7 ppm for 13C and 15N respectively (29). The a priori determination of the global uncertainties for the predicted shifts allows for the quantitative comparison of the predicted and experimental shifts as described below. Key first-principles chemical shifts for the E(A-A) and E(A-A)(BZI) intermediate candidates are given in Tables S1-S4.

Statistical Methods

Reduced-χ 2 Analysis

The structural models were ranked using the reduced-χ 2 statistic (17), a quantitative measure of the agreement between their first-principles predicted chemical shifts and the experimental NMR parameters,

$$\chi_r^2(\text{model}) = \frac{1}{N-f}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{s_i^2}$$

In this expression, the summation i runs over all N measured NMR shifts, f is the number of adjustable model parameters (0 for the direct ranking of models; 1 for the exchange model with optimized populations), $\delta_\mathrm{i}^{\mathrm{exp}}$ is the experimental chemical shift, $\delta_\mathrm{i}^{\mathrm{model}}$ is the corresponding predicted shift for a given model, and s_i^2 is the nuclide-specific weighting derived by setting s_i to the corresponding root-mean-square error derived from benchmark studies described above. Importantly, the reduced-χ 2 allows for chemical shift data from multiple nuclides (13C, 15N, 17O, etc.) and of different types (isotropic, tensor principal components) to be combined into a single statistic. For a priori assigned weights, the reduced-χ 2 allows for rigorous evaluation of the consistency of a model with the experimental data (89).

Bayesian Probability Analysis The overall confidence in the identification of the experimental structure was also quantified using Bayesian probability analysis following the approach of Engel et al. (69). In the context of NMR-assisted crystallography, S5 the application of Bayes' Theorem gives the following expression for the posterior probability, $p\left(M \middle| y^{*}\right)$, that M is the experimental structure given the measured set of chemical shifts y

$$p\left(M \middle| y^{*}\right) = \frac{p\left(y^{*} \middle| M\right) p(M)}{p\left(y^{*}\right)} = \frac{p\left(y^{*} \middle| M\right) p(M)}{\sum_{M'} p\left(y^{*} \middle| M'\right) p(M')}$$

Here, $p\left(y^{*} \middle| M\right)$ M is the probability of observing chemical shifts y* for a given structure M, and $p(M)$ is the prior probability of structure M matching the experimental structure. The probabilities $p\left(y^{*} \middle| M\right)$ M were determined relative to the probability of the best-fit (BF) model based on the one-sided p-value (z) from the corresponding F-distribution evaluated at the critical value set by the ratio of the reduced-χ 2 for model M and the best fit model (89)

$$\alpha = \frac{\chi_{M}^{2}}{\chi_{BF}^{2}}$$

The p-value was then converted into a relative probability for model M according to

$$\frac{p\left(y^{*} \middle| M\right)}{p\left(y^{*} \middle| M\right) + p\left(y^{*} \middle| M^{BF}\right)} = z$$

using a fixed relative probability $p\left(y^{*} \middle| M^{BF}\right) = 1$. Reduced-χ 2 for the 10 best exchange models for E(A-A) and E(A-A)(BZI) are given in Tables S6 and S7, along with their relative probabilities, $p\left(y^{*} \middle| M\right)$. Following Engel et al. (69), we also incorporated into our analysis the possibility that the experimental structure was not part of our ensemble through the introduction of a virtual structure, M , whose shifts were taken as the mean of the shifts for the other candidate structures. In this case, the virtual structure was found to have negligible probability of matching the experimental structure for both E(A-A) and E(A-A)(BZI) (Tables S6, S7). The prior probability p M( ) was used to introduce consistency with the observed temperature dependence of the Schiff base nitrogen chemical shift. Temperature dependence at this site requires exchange models that include tautomeric exchange at the Schiff base nitrogen; such models were assigned relative probabilities $p\left(M\right)=1$while all other models were assigned $p\left(M\right)=0$. Taking into account both the isotropic and anisotropic chemical shift data, along with the observed temperature dependence of the Schiff base nitrogen, the proposed exchange model was found to be the most probable experimental structure with 88.5% confidence for E(A-A), and 99.6% confidence for E(A-A)(BZI). All probabilities are summarized in Tables S6 and S7.

Reported ADPs

In this preliminary analysis, ADPs are reported only for atoms with measured chemical shifts and their covalently- or hydrogen-bonded neighbors. This restriction can likely be relaxed, a subject of ongoing work. The current restriction recognizes a fundamental difference in the application of NMR crystallography to an active site compared to its application in organic molecular crystals(41, 90): active sites are typically modeled using a cluster-based approach in which the measured chemical shifts are for atoms concentrated at the center of the model, while molecular crystals are typically modeled in a unit-cell based approach and have measured chemical shifts distributed throughout. Displacements of atoms at the periphery of cluster models are expected to have little influence on the measured chemical shifts at the center of the cluster, and therefore are expected to be unconstrained. To better quantify this, the sensitivity of the reduced-χ 2 to atomic displacements was assessed by forming its gradient with respect to single atom displacements using finite differences. It is known that single atom displacements lead to unphysical models and unrealistically large changes in chemical shifts within a covalent network (19), but this is expected to be a less severe problem for displacements outside of the shift region. These simulations confirm that atoms whose ADP have been reported show the largest gradients – more than 3 orders of magnitude larger than for atoms at the edge of the model.

Protein Preparation

Tryptophan synthase was prepared by overexpression of StTS in E. coli BL21 cells as previously described(48, 51). Samples were prepared with the following isotopic labeling schemes for the cofactor and protein components: 1. Protein and cofactor unlabeled/natural abundance isotopomer concentration; 2. Protein 15Nlabeled at lysine ε-nitrogen sites (ε15N-Lys TS); and 3. Protein with the PLP cofactor selectively 13C enriched at carbon sites C2, C2', and C3 and 15N enriched at the pyridine ring nitrogen (2,2',3- 13C3, 15N-PLP; TS). Isotopically labeled PLP was prepared as described below and exchanged into the β-subunit active site as detailed in (51)

Microcrystalline Protein Samples for Solid-State NMR

Microcrystalline samples of TS were prepared by diluting enzyme solution 1:1 with 50 mM Cs+ -bicine buffer, pH 7.8, containing 14% PEG-8000 and 3.0 mM spermine (51). Microcrystals were collected and washed with 50 mM Cs+ -bicine, pH 7.8, containing 8% PEG-8000, 1.8 mM spermine, and 3 mM N-(4'-trifluoromethoxybenzenesulfonyl)-2-aminoethyl phosphate (F9; a high affinity alpha site ligand and analogue of the natural α-site substrate 3-idole-D-glycerol-3'-phosphate (IGP) (91). The crystals were packed at 10,000 rpm into a Bruker 4 mm magic-angle spinning (MAS) rotor with an approximate volume of 80 μL; each rotor contained 25-30 mg of protein. To form the α-aminoacrylate intermediate, serine was introduced by direct addition of 5 μL of 1.2 M L-Ser to the packed MAS rotor. Stabilization of the α-aminoacrylate species is enhanced by low temperature (-5 °C), the use of the tight binding α-subunit ligand F9, and the presence of Cs+ (91), which binds to the monovalent cation site in the β-subunit.

13C and 15N Solid-State NMR Spectroscopy

13C and 15N cross-polarization (CP) magic-angle-spinning (MAS) experiments were performed at 9.4 T (400.37 MHz 1H, 100.69 MHz 13C, 40.57 MHz 15N) on a Bruker AVIII spectrometer equipped with a double resonance, 4 mm MAS probe, spinning at MAS rates of 8 kHz; the bearing gas was cooled to -15 °C, giving an effective sample temperature of -10 °C. Cross-polarization was accomplished at a 1H spin-lock field of 45 kHz and a 13C/15N spin-lock of 54 kHz (13C) and 37 kHz (15N) (ramped +/- 2 kHz); 85 kHz Spinal64 1H decoupling (92) was used throughout. 13C spectra consist of the sum of 16,384 transients acquired with a relaxation delay of 4 s, for a total acquisition time of 18.3 h. 13C chemical shifts were referenced indirectly to neat TMS via an external solid-state sample of adamantane with the downfield-shifted peak set to 38.48 ppm (93, 94). 15N spectra consist of the sum of 81,920 transients acquired with a relaxation delay of 4 s, for a total acquisition time of 3 d 19 h. 15N chemical shifts were referenced indirectly to liq-NH3 (25 °C) via an external solid-state sample of 15NH4Cl, in which the resonance frequency was set to 39.27 ppm (93).

The acquisition of solid-state NMR spectra was interleaved with single pulse, low-power decoupling experiments (64 scans 13C, 256 scans 15N) reporting predominantly on free ligand and reaction products in solution (mother liquor) (Fig. S5). Acquisition of solid-state NMR spectra for the intermediate was halted before reactant concentrations in solution fell to zero.

31P Solid-State NMR Spectroscopy

31P CPMAS experiments on the α-aminoacrylate species were performed at 14.1 T (600.01 MHz 1H, 242.89 MHz 31P) on a Bruker NEO spectrometer equipped with an 1H-X double resonance 4 mm MAS probe, spinning at a MAS rate of 10 kHz. The bearing gas was cooled to -15 °C, giving an effective sample temperature of -5 °C. Cross-polarization was accomplished with a 1H spin-lock field of 45 kHz and a 31P spin-lock of 47 kHz (ramped +/- 5 kHz); 58 kHz Spinal64 1H decoupling (36) was used during detection. The 31P spectra consist of the sum of 8,192 transients acquired with a relaxation delay of 4 s, for a total acquisition time of 9.1 h. 31P chemical shifts were indirectly referenced to 85% H3PO4 (MAS). For comparison to measurements in solution, δ[85% H3PO4 (capillary)] = δ[85% H3PO4 (sphere/MAS)] + 0.36 ppm (95).

15N Chemical Shift Anisotropy Measurements

Slow spinning 15N CPMAS experiments on E(A-A) and E(A-A)(BZI) were performed at 14.1 T (600.01 MHz 1H, 60.82 MHz 15N) on a Bruker NEO spectrometer equipped with an 1H-X double resonance 4 mm MAS probe, S12 spinning at a MAS rate of 2.3 kHz. The bearing gas was cooled to -15 °C, giving an effective sample temperature of -10 °C. Cross-polarization was accomplished with a 1H spin-lock field of 45 kHz and a 15N spinlock of 47 kHz (ramped +/- 5 kHz); 83 kHz Spinal64 1H decoupling (92) was used during detection. 15N spectra consist of the sum of 81,920 transients acquired with a relaxation delay of 4 s, for a total acquisition time of 3 d 19 h. Chemical shift tensor principal axis components were determined by a fit of the sideband intensities in the slow MAS spectra using Herzfeld-Berger analysis (49), implemented within Bruker BioSpin's Topspin 3.6 processing software. Standard errors were determined by a numerical calculation of the covariance matrix (96) in Mathematica, using spectra that were zero-filled once and processed without line broadening; this ensures that extrapolated data points (from zero-filling more than once) were not included in the calculation of the covariance matrix and that an accurate estimate of uncorrelated RMS noise was obtained.

2D 13C Correlation Spectroscopy on a CPMAS CryoProbe

2D 13C13C CORD (44) were collected at 14.1 T (600.03 MHz 1H, 150.9 MHz 13C) on a standard bore Bruker NEO spectrometer equipped with a 3.2 mm CPMAS cryoprobe (Bruker Biosolids CryoProbe™) in which the detection coil and preamplifier operate at cryogenic temperatures to increase sensitivity while the sample remains at lab temperature (73). Experiments were performed at a MAS rate of 13.5 kHz. Cross Polarization was accomplished with a 13C spin lock of 55 kHz and 1H spin lock of 68 kHz (linearly ramped ±15%). 62.5 kHz 13C and 100 kHz 1H pulses were used throughout and 100 kHz Spinal64 1H decoupling (36) was used during detection. The CORD mixing time was 50ms. In the indirect dimension, 160 complex points with a dwell of 37.04 µs (spectral width 27 kHz, total acquisition time 5.93 ms) were acquired. In the direct dimension, 1024 complex points with a dwell of 22 µs (spectral width 45.45 kHz, total acquisition time 22.52 ms) were acquired. 16 transients at each t1 point were coadded with a relaxation delay 2 s for a total experiment time of 2.84 hr. The sample temperature was regulated with Bruker variable temperature controller set to -21 °C, giving an effective sample temperature of -10 °C

Phosphorylation of ^{15}N pyridoxal to ^{15}N PLP

A 5ml solution composed of 32 µM of K229Q PL Kinase, 32 mM ATP, 50 mM KCl, 200 mM HEPES•K pH 7.5, 0.2 mM MgCl, and 5 mM (15N)-pyridoxal (15.7

mM) was stirred and the reaction run overnight at room temperature. A 1mL aliquot of the solution was monitored by UV-Vis at 388 nm, showing the buildup of PLP. The crude solution was used without further purification for exchange into the apoenzyme as previously described (51). Spectra for the holoenzyme were consistent with literature. The K229Q mutant ePL kinase was expressed and purified according to (97).

References

1. H. Hayashi, Pyridoxal Enzymes: Mechanistic Diversity and Uniformity. The Journal of Biochemistry **118**, 463-473 (1995).
2. C. Walsh, Enzymatic Reaction Mechanisms (W. H. Freeman and Company, San Francisco, 1979), pp. 978.
3. R. Percudani, A. Peracchi, A genomic overview of pyridoxal-phosphate-dependent enzymes. EMBO reports **4**, 850-854 (2003).
4. E. F. Pettersen et al., UCSF chimera - A visualization system for exploratory research and analysis. J Comput Chem **25**, 1605-1612 (2004).
5. M. D. Toney, Controlling reaction specificity in pyridoxal phosphate enzymes. BBA - Protein Proteom **1814**, 1407-1418 (2011).
6. E. W. Miles, Tryptophan synthase: structure, function, and subunit interaction. Adv Enzymol Relat Areas Mol Biol **49**, 127-186 (1979).
7. C. Yanofsky, I. P. Crawford, "Tryptophan Synthetase" in The Enzymes Third Edition, P. D. Boyer, Ed. (Academic Press, Inc., New York, NY, 1972), vol. 7, chap. 1, pp. 1-32.
8. M. F. Dunn, D. Niks, H. Ngo, T. R. Barends, I. Schlichting, Tryptophan synthase: the workings of a channeling nanomachine. Trends Biochem Sci **33**, 254-264 (2008).
9. H. C. Dunathan, Conformation and Reaction Specificity in Pyridoxal Phosphate Enzymes. P Natl Acad Sci USA **55**, 712 (1966).
10. M. Chan-Huot et al., 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. **135**, 18160-18175 (2013).
11. P. S. Langan et al., Substrate binding induces conformational changes in a class A β-lactamase that prime it for catalysis. ACS Catalysis **8**, 2428-2437 (2018).
12. S. Dajnowicz et al., Direct visualization of critical hydrogen atoms in a

pyridoxal 5′-phosphate enzyme. Nature communications **8**, 1-9 (2017).

13. K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, Atomic-resolution protein structure determination by cryo-EM. Nature **587**, 157-161 (2020).
14. A. T. Petkova et al., A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. Proceedings of the National Academy of Sciences **99**, 16742-16747 (2002).
15. C. Luchinat, G. Parigi, E. Ravera, M. Rinaldelli, Solid-State NMR Crystallography through Paramagnetic Restraints. J. Am. Chem. Soc. **134**, 5006-5009 (2012).
16. J.-P. Demers et al., High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. Nature Communications **5**, 4976 (2014).
17. L. Sborgi et al., Structure and assembly of the mouse ASC inflammasome by combined NMR spectroscopy and cryo-electron microscopy. Proceedings of the National Academy of Sciences **112**, 13237-13242 (2015).
18. R. Michalczyk et al., Joint neutron crystallographic and NMR solution studies of Tyr residue ionization and hydrogen bonding: Implications for enzyme-mediated proton transfer. Proceedings of the National Academy of Sciences **112**, 5673-5678 (2015).
19. A. Carlon et al., Improved Accuracy from Joint X-ray and NMR Refinement of a Protein–RNA Complex Structure. J. Am. Chem. Soc. **138**, 1601-1610 (2016).
20. L. Gremer et al., Fibril structure of amyloid-β(1–42) by cryo–electron microscopy. Science **358**, 116-119 (2017).
21. J. R. Perilla et al., CryoEM Structure Refinement by Integrating NMR Chemical Shifts with Molecular Dynamics Simulations. The Journal of Physical Chemistry B **121**, 3853-3863 (2017).
22. K. Zhang et al., Structure of the 30 kDa HIV-1 RNA Dimerization Signal by a Hybrid Cryo-EM, NMR, and Molecular Dynamics Approach. Structure **26**, 490-498.e493 (2018).
23. L. Cerofolini et al., Characterization of PEGylated Asparaginase: New Opportunities from NMR Analysis of Large PEGylated Therapeutics. Chemistry **25**, 1984-1991 (2019).

24. D. F. Gauto et al., Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. Nat Commun **10**, 2697 (2019).

25. J. Felix et al., Mechanism of the allosteric activation of the ClpP protease machinery by substrates and active-site inhibitors. Science Advances **5**, eaaw3818 (2019).

26. S. Kosol et al., Structural basis for chain release from the enacyloxin polyketide synthase. Nature Chemistry **11**, 913-923 (2019).

27. J. A. Gascon, E. M. Sproviero, V. S. Batista, QM/MM Study of the NMR Spectroscopy of the Retinyl Chromophore in Visual Rhodopsin. J Chem Theory Comput **1**, 674-685 (2005).

28. J. Lai et al., X-ray and NMR crystallography in an enzyme active site: the indoline quinonoid intermediate in tryptophan synthase. J. Am. Chem. Soc. **133**, 4-7 (2011).

29. B. G. Caulkins et al., NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. J. Am. Chem. Soc. **138**, 15214-15226 (2016).

30. L. J. Mueller, M. F. Dunn, NMR crystallography of enzyme active sites: probing chemically detailed, three-dimensional structure in tryptophan synthase. Accounts of Chemical Research **46**, 2008-2017 (2013).

31. H. L. Zhang et al., HIV-1 Capsid Function Is Regulated by Dynamics: Quantitative. Atomic-Resolution Insights by Integrating Magic-Angle-Spinning NMR, QM/MM, and MD. J. Am. Chem. Soc. **138**, 14066-14075 (2016).

32. H. Singh et al., Fast Microsecond Dynamics of the Protein-Water Network in the Active Site of Human Carbonic Anhydrase II Studied by Solid-State NMR Spectroscopy. J. Am. Chem. Soc. **141**, 19276-19288 (2019).

33. A. Bertarello et al., Picometer Resolution Structure of the Coordination Sphere in the Metal-Binding Site in a Metalloprotein by NMR. J. Am. Chem. Soc. **142**, 16757-16765 (2020).

34. J. C. Facelli, D. M. Grant, Determination of molecular symmetry in crystalline naphthalene using solid-state NMR. Nature **365**, 325-327 (1993).

35. M. Baias et al., De novo determination of the crystal structure of a large

drug molecule by crystal structure prediction-based powder NMR crystallography. J. Am. Chem. Soc. **135**, 17501-17507 (2013).

36. D. H. Brouwer et al., A general protocol for determining the structures of molecularly ordered but noncrystalline silicate frameworks. J. Am. Chem. Soc. **135**, 5641-5655 (2013).

37. C. Martineau, NMR crystallography: Applications to inorganic materials. Solid State Nuclear Magnetic Resonance **63**, 1-12 (2014).

38. D. Bryce, NMR crystallography: structure and properties of materials from solid-state nuclear magnetic resonance observables. IUCrJ **4**, 350-359 (2017).

39. P. Hodgkinson, NMR crystallography of molecular organics. Progress in Nuclear Magnetic Resonance Spectroscopy **118-119**, 10-53 (2020).

40. D. Niks et al., Allostery and substrate channeling in the tryptophan synthase bienzyme complex: evidence for two subunit conformations and four quaternary states. Biochemistry **52**, 6396-6411 (2013).

41. M. Cordova et al., Structure determination of an amorphous drug through large-scale NMR predictions. Nature Communications **12**, 2964 (2021).

42. M. F. Dunn, Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. Archives of Biochemistry and Biophysics **519**, 154-166 (2012).

43. M. A. Maria-Solano, J. Iglesias-Fernández, S. Osuna, Deciphering the allosterically driven conformational ensemble in tryptophan synthase evolution. J. Am. Chem. Soc. **141**, 13049-13056 (2019).

44. H. Ngo et al., Allosteric regulation of substrate channeling in tryptophan synthase: modulation of the L-serine reaction in stage I of the beta-reaction by alpha-site ligands. Biochemistry **46**, 7740-7753 (2007).

45. D. Niks et al., Allostery and substrate channeling in the tryptophan synthase bienzyme complex: evidence for two subunit conformations and four quaternary States. Biochemistry **52**, 6396-6411 (2013).

46. S. A. Ahmed, C. C. Hyde, G. Thomas, E. W. Miles, Microcrystals of tryptophan synthase $\alpha_2\beta_2$ complex from Salmonella typhimurium are catalytically active. Biochemistry **26**, 5492-5498 (1987).

47. L. M. McDowell, M. S. Lee, J. Schaefer, K. S. Anderson, Observation of

an Aminoacrylate Enzyme Intermediate in the Tryptophan Synthase Reaction by Solid-State NMR. J. Am. Chem. Soc. **117**, 12352-12353 (1995).

48. B. G. Caulkins et al., Catalytic roles of βLys87 in tryptophan synthase: ^{15}N solid state NMR studies. BBA - Protein Proteom **1854**, 1194-1199 (2015).

49. J. Herzfeld, A. E. Berger, Sideband intensities in NMR spectra of samples spinning at the magic angle. Journal of Chemical Physics **73**, 6021-6030 (1980).

50. V. Copie, W. S. Faraci, C. T. Walsh, R. G. Griffin, Inhibition of alanine racemase by alanine phosphonate: detection of an imine linkage to pyridoxal 5'-phosphate in the enzyme-inhibitor complex by solid-state nitrogen-15 nuclear magnetic resonance. Biochemistry **27**, 4966-4970 (1988).

51. B. G. Caulkins et al., 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. **136**, 12824-12827 (2014).

52. K. D. Schnackerz, B. Andi, P. F. Cook, ^{31}P NMR spectroscopy senses the microenvironment of the 5'-phosphate group of enzyme-bound pyridoxal 5'-phosphate. BBA - Protein Proteom **1814**, 1447-1458 (2011).

53. J. Zhu, G. Wu, Quadrupole central transition ^{17}O NMR spectroscopy of biological macromolecules in aqueous solution. J. Am. Chem. Soc. **133**, 920-932 (2011).

54. R. P. Young et al., Solution-State ^{17}O Quadrupole Central-Transition NMR Spectroscopy in the Active Site of Tryptophan Synthase. Angewandte Chemie International Edition **55**, 1350-1354 (2016).

55. H. C. Lee, K. Cummings, K. Hall, L. P. Hager, E. Oldfield, Oxygen-17 nuclear magnetic resonance spectroscopic studies of carbonmonoxyperoxidases. J Biol Chem **263**, 16118-16124 (1988).

56. Z. Gan et al., NMR spectroscopy up to 35.2 T using a series-connected hybrid magnet. Journal of Magnetic Resonance **284**, 125-136 (2017).

57. T. Gullion, J. Schaefer, Rotational-Echo Double-Resonance NMR. Journal of Magnetic Resonance **81**, 196-200 (1989).

58. A. K. Schütz, Solid-state NMR approaches to investigate large enzymes in

complex with substrates and inhibitors. Biochemical Society Transactions **49**, 131-144 (2020).

59. J. Kraus, S. Sarkar, C. M. Quinn, T. Polenova, "Solid-state NMR spectroscopy of microcrystalline proteins" in Annual Reports on NMR Spectroscopy, G. A. Webb, Ed. (Academic Press, 2021), vol. 102, pp. 81-151.

60. A. C. de Dios, J. G. Pearson, E. Oldfield, Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. Science **260**, 1491-1496 (1993).

61. Q. Cui, M. Karplus, Molecular Properties from Combined QM/MM Methods. 2. Chemical Shifts in Large Molecules. The Journal of Physical Chemistry B **104**, 3721-3743 (2000).

62. L. B. Casabianca, A. C. de Dios, Ab initio calculations of NMR chemical shifts. The Journal of Chemical Physics **128**, 052201 (2008).

63. J. C. Facelli, Chemical shift tensors: theory and application to molecular structural problems. Prog Nucl Magn Reson Spectrosc **58**, 176-201 (2011).

64. T. E. Exner, A. Frank, I. Onila, H. M. Möller, Toward the Quantum Chemical Calculation of NMR Chemical Shifts of Proteins. 3. Conformational Sampling and Explicit Solvents Model. Journal of Chemical Theory and Computation **8**, 4818-4827 (2012).

65. J. Swails, T. Zhu, X. He, D. A. Case, AFNMR: automated fragmentation quantum mechanical calculation of NMR chemical shifts for biomolecules. Journal of Biomolecular NMR **63**, 125-139 (2015).

66. J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller, G. J. Beran, Benchmark fragment-based ^{1}H, ^{13}C, ^{15}N and ^{17}O chemical shift predictions in molecular crystals. Physical Chemistry Chemical Physics **18**, 21686-21709 (2016).

67. S. K. Chandy, B. Thapa, K. Raghavachari, Accurate and cost-effective NMR chemical shift predictions for proteins using a molecules-in-molecules fragmentation-based method. Physical Chemistry Chemical Physics **22**, 27781-27799 (2020).

68. J. D. Hartman, T. J. Neubauer, B. G. Caulkins, L. J. Mueller, G. J. Beran, Converging nuclear magnetic shielding calculations with respect to basis and system size in protein systems. J. Biomol. NMR **62**, 327-340 (2015).

69. E. A. Engel et al., A Bayesian approach to NMR crystal structure determination. Physical Chemistry Chemical Physics **21**, 23385-23400 (2019).

70. A. Hofstetter, L. Emsley, Positional Variance in NMR Crystallography. J Am Chem Soc **139**, 2573-2576 (2017).

71. L. Wang, F. J. Uribe-Romo, L. J. Mueller, J. K. Harper, Predicting anisotropic thermal displacements for hydrogens from solid-state NMR: a study on hydrogen bonding in polymorphs of palmitic acid. Physical Chemistry Chemical Physics **20**, 8475-8487 (2018).

72. G. Hou, S. Yan, J. Trebosc, J. P. Amoureux, T. Polenova, Broadband homonuclear correlation spectroscopy driven by combined R2(n)(v) sequences under fast magic angle spinning for NMR structural analysis of organic and biological solids. J Magn Reson **232**, 18-30 (2013).

73. A. Hassan et al., Sensitivity boosts by the CPMAS CryoProbe for challenging biological assemblies. J Magn Reson **311**, 106680 (2020).

74. M. Roy, S. Keblawi, M. F. Dunn, Stereoelectronic control of bond formation in Escherichia coli tryptophan synthase: substrate specificity and enzymatic synthesis of the novel amino acid dihydroisotryptophan. Biochemistry **27**, 6698-6704 (1988).

75. M. F. Dunn et al., The tryptophan synthase bienzyme complex transfers indole between the alpha- and beta-sites via a 25-30 Å long tunnel. Biochemistry **29**, 8598-8607 (1990).

76. P. S. Brzovic, A. M. Kayastha, E. W. Miles, M. F. Dunn, Substitution of glutamic acid 109 by aspartic acid alters the substrate specificity and catalytic activity of the β-subunit in the tryptophan synthase bienzyme complex from Salmonella typhimurium. Biochemistry **31**, 1180-1190 (1992).

77. R. M. Harris, H. Ngo, M. F. Dunn, Synergistic effects on escape of a ligand from the closed tryptophan synthase bienzyme complex. Biochemistry **44**, 16886-16895 (2005).

78. E. Hilario et al., Visualizing the tunnel in tryptophan synthase with crystallography: Insights into a selective filter for accommodating indole and rejecting water. BBA - Protein Proteom **1864**, 268-279 (2016).

79. D. Case et al., AMBER 2018; 2018. University of California, San Francisco.

80. R. Salomon-Ferrer, A. W. Gotz, D. Poole, S. Le Grand, R. C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. Journal of chemical theory and computation **9**, 3878-3888 (2013).

81. J. A. Maier et al., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. Journal of chemical theory and computation **11**, 3696-3713 (2015).

82. A. Jakalian, D. B. Jack, C. I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem **23**, 1623-1641 (2002).

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

84. H. Nguyen, D. R. Roe, C. Simmerling, Improved Generalized Born Solvent Model Parameters for Protein Simulations. J Chem Theory Comput **9**, 2020-2034 (2013).

85. D. R. Roe, T. E. Cheatham, 3rd, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput **9**, 3084-3095 (2013).

86. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. J Mol Graph **14**, 33-38, 27-38 (1996).

87. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys **132**, 154104 (2010).

88. S. Moon, D. A. Case, A comparison of quantum chemical models for calculating NMR shielding parameters in peptides: mixed basis set and ONIOM methods combined with a complete basis set extrapolation. J Comput Chem **27**, 825-836 (2006).

89. D. P. Shoemaker, C. W. Garland, Experiments in physical chemistry (ed. 2, 1967).

90. A. Hofstetter, L. Emsley, Positional Variance in NMR Crystallography. J. Am. Chem. Soc. **139**, 2573-2576 (2017).

91. H. Ngo et al., Synthesis and characterization of allosteric probes of substrate channeling in the tryptophan synthase bienzyme complex.

Biochemistry **46**, 7713-7727 (2007).

92. B. M. Fung, A. K. Khitrin, K. Ermolaev, An improved broadband decoupling sequence for liquid crystals and solids. J Magn Reson **142**, 97-101 (2000).

93. R. K. Harris et al., Further conventions for NMR shielding and chemical shifts (IUPAC Recommendations 2008). Pure Appl Chem **80**, 59-84 (2008).

94. C. R. Morcombe, K. W. Zilm, Chemical shift referencing in MAS solid state NMR. J Magn Reson **162**, 479-486 (2003).

95. T. Maurer, H. R. Kalbitzer, Indirect Referencing of 31P and 19F NMR Spectra. J Magn Reson Ser B **113**, 177-178 (1996).

96. W. H. Press, Numerical recipes in C : the art of scientific computing (Cambridge University Press, Cambridge ; New York, ed. 2nd, 1992), pp. xxvi, 994 p.

97. M. S. Ghatge et al., Pyridoxal 5′-Phosphate Is a Slow Tight Binding Inhibitor of E. coli Pyridoxal Kinase. PLOS ONE **7**, e41680 (2012).

Chapter 5 – Anisotropic Displacement Parameters in Enzyme Active Sites

Introduction

Structure plays a critical role in shaping our understanding of the chemistry in enzyme active sites. Recently, our group and others have developed integrative techniques that couple solid-state NMR with X-ray diffraction and computational chemistry to determine the structure of enzyme active sites with atomic resolution(1-4). A particular emphasis has been placed on the determination of hydrogen atom locations, which are challenging to extract using the primary tools of structural biology: X-ray diffraction and cryoEM. Accurate knowledge of proton locations is critical to establishing the mechanism of acid-base catalysis and as input for molecular dynamics simulations, molecular docking routines, and structure-based drug design.

The integrative approach for enzyme active sites is typically three-fold: first, X-ray crystallography is used to provide a coarse structural framework upon which, second, chemically-detailed models of the active site are built using computational chemistry, and various active site chemical and structural models explored; third, these models are quantitatively distinguished by comparing their predicted NMR chemical shifts with the results from solid-state NMR experiments. Provided enough chemical shift restraints are measured throughout the active site, NMR-assisted crystallography can identify a uniquely consistent structure or, equally important, determine that none of the candidates is consistent with the experimental data. For the case of the “quinonoid” intermediate in tryptophan synthase (TS), this approach demonstrated that the intermediate is a carbanionic species with a deprotonated pyridine ring nitrogen – a structure that we propose is fundamental to understanding reaction specificity in TS(2). For the α-aminoacrylate intermediate, NMR-assisted crystallography highlighted the reaction coordinate for the loss of the substrate L-Ser hydroxyl and the protonation states that are critical to why the inhibitor benzimidazole (BZI) does not react while the natural substrate indole does(1).

This general integrative approach is often referred to as NMR crystallography and was initially developed for structure determination in molecular organic and inorganic solids(5, 6). NMR crystallography merges the sensitivity of diffraction to long range order with that of NMR to local structure and chemistry. It has recently been recognized that an important ingredient to NMR crystallography should be a quantitative description of the quality of the structure, and several groups have delineated approaches to quantifying positional uncertainties in the reported coordinates NMR crystal structures(7, 8). In various ways, these approaches define sets of structures that are fully consistent with the NMR restraints, and those which are not. The boundary between these sets is used to define anisotropic displacement parameters at each atom: coordinates within the boundary are (to a stated confidence) are fully consistent with the NMR chemical shifts.

While the spirit of NMR crystallography in organic molecular solids and enzyme active sites are similar, there are important distinctions due to the different relative sizes of the region of interest to the crystal unit cell. For most organic molecular crystals, the region of interest and the unit cell are typically the same, and chemical shift restraints are dispersed across the entire unit cell. In enzyme active sites, the region of interest is restricted by the modeling constraints to be only a few percent of the (albeit quite large) protein crystal unit cell, which shifts concentrated within an even smaller part of that region. Fig. 1 highlights the various size scales for the α-aminoacrylate intermediate in TS, including the cluster model region for first-principles computational chemistry and those positions at the center of the cluster for which chemical shifts are measured.

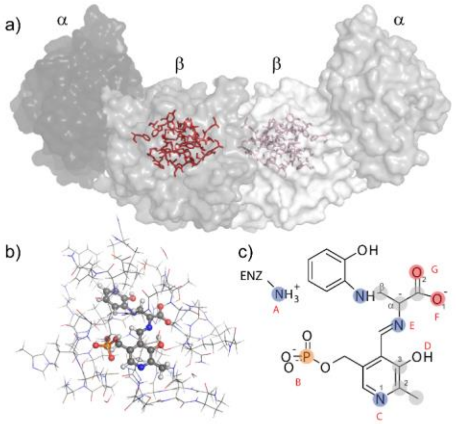


Fig. 5.1. Cluster model of the E(C)$_{2AP}$ active site. (A) X-ray crystal structure of the TS $\alpha_2\beta_2$ heterodimer with the β-subunit active site in red. (B) Cluster model of the active site for first-principles geometry optimization and chemical shift calculations; protein side chains displayed in wireframe and cofactor and substrate in ball-and-stick. (C) Protonation sites on and near the cofactor/substrate complex: A, the βLys87 side chain; B. the PLP phosphate group; C, the PLP pyridine ring nitrogen; D, the PLP phenolic oxygen; E, the Schiff-base nitrogen; and F/G, the substrate carboxylate. Shaded nuclei indicate sites for which experimental NMR chemical shifts are reported.

Here, we focus on setting a methodology for determining positional uncertainties that is applicable to NMR crystallography of enzyme active sites. We expand on the framework set by Hofstetter and Emsley for calculating the positional uncertainties in NMR crystallography of molecular organic crystals by explicitly dividing our system into shift-rich and shift-lacking regions. For TS, these correspond to the cofactor/substrate and the surrounding residues, respectively (Fig. 1). The positional uncertainties are determined by examining the chemical shift perturbation caused by molecular motions in two. We consider three motional schemes derived from the same MD simulations. First, the fully perturbed ensemble in which both the cofactor/substrate and surrounding residues or lattice are free to move. Second, an ensemble in which the cofactor/substrate are free to move and the lattice is held fixed. Last, the inverse of ensemble two where the cofactor/substrate is held fixed and the lattice is free to move.

Positional Uncertainties in Enzyme Active Sites

Molecular Dynamics Simulations

Low-temperature MD for three structures (PDB IDs: 4HPJ with 2-aminophenol quinonoid (9)and 3PR2 with indoline quinonoid (10)) were performed on the TS αβ dimer using AMBER package.(11-14) Structure coordinates for the StTS were downloaded from the protein data bank with crystal water retained. (15) System protonation states and the active site were determined and refined by NMR crystallography. Amber force field FF14SB and General Amber Force Field 2 (GAFF2) and AM1-BCC charge model were applied to protein and ligand parametrization, respectively. (13, 16, 17)

All three systems were subjected to a three-step minimization process: hydrogens, site chains and full protein, with the refined region within 5Å of the β subunit ligand (PLP cofactor and covalently attached ligands) restrained. A standard AMBER simulation protocol was applied; each system was solvated in a TIP3P water box(18)with a padding of 12Å from the edge of the box to the protein atoms. Appropriate Na^+ and Cl^- ions were added to neutralize the system in the water box. Long-range electrostatic interactions were computed by the particle mesh Ewald method. (19) The SHAKE constraint was applied all atoms including hydrogens. (20) The solvated systems were minimized again in two steps: water only followed by the whole system.

Next, each system was gradually heated at constant pressure from 1K to 10K at 1K interval, and subsequently to 150K at 5K interval. 300 ps MD simulations were performed at each temperature with a time step of 2 fs using the thermal fluctuation configuration ensemble. Trajectories were processed to contain 300 frames (1ps per frame) at each temperature interval for each system.

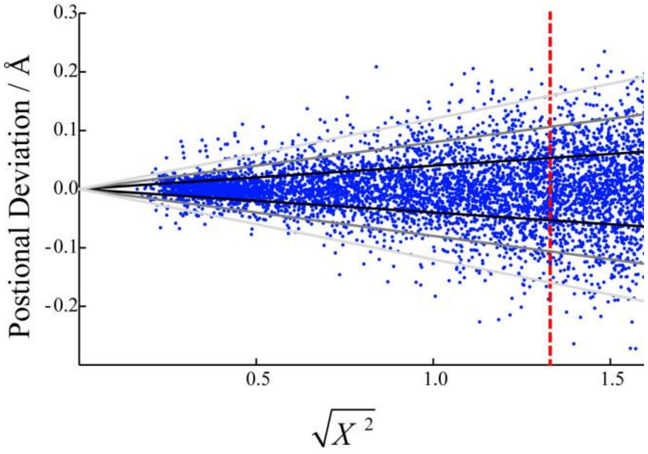


Fig. 5.2: Scatter plot of $\sqrt{\chi_r^2}$ and positional deviation along one principal axis of the anisotropic displacement tensor for C^α of E(AA)(2AP). The vertical line at $\sqrt{\chi_r^2} = \sqrt{1.79}$ shows the 95% confidence cutoff in the reduced-χ^2. The black and gray lines indicate ±1, 2, and 3 times the width of the Gaussian distribution at a given reduced-χ^2.

First Principles NMR Calculations

For each of the 11400 structures in the ensemble, the active site region consisting of the PLP-substrate covalent complex, any ligands and in the case of

E(A-A) the adjacent 3 water, and residues βSer377(C^{α}), βHis86(C^{β}), βLys87(C^{γ}), βGlu109, βThr110, βGly111, βAla112, βGly232, βGly233, βGly234, βGly235, βAsn236, βGln114(bb), and βHis115(bb) were extracted for NMR chemical shift calculations. Only the sidechain atoms of residues βSer377, βHis86, and βLys87 were included, truncated at the indicated atom and capped with a hydrogen. Only the backbone atoms and sidechains through C^{β} were included for βGln114 and βHis115. Terminal residues in contiguous chains were capped with hydrogen atoms. E(A-A) structures contained 192 atoms, E(A-A)(BZI) structures contained 199, E(Q3)(2AP) structures contained 198, and E(Q3)(Indoline) structures contained 202.

For each structure, first principles chemical shifts were calculated at the DFT level of theory (B3LYP/6-31G). These shifts were used to calculate a reduced-χ^2 relative to a reference structure. For the structures in which the cofactor and surroundings are perturbed (full perturbation) internal referencing at each temperature is used. For each temperature, the structure which minimizes reduced-χ^2 of the ensemble is used as the reference structure. This reference structure is also used to re-refence coordinates so that a structure whose reduced-χ^2 is 0 also has an RMSD of 0 Å

Anisotropic Displacement Parameters

The positional deviations of each atom relative to its temperature-specific reference structure were used to define a covariance matrix of displacements and an atom-specific principal axis system (7). The correlation of positional deviations along a PAS axes (r_i) and the corresponding reduced-χ^2 was used to fit a 2D probability density at each atom (l)

$$G(r_{i,l},\chi_r^2) = \frac{A_{i,l}\left(\chi_r^2\right)}{\sqrt{2\pi\sigma_{i,l}^2\chi_r^2}} \exp\left\{-\frac{r_{i,l}^2}{2\sigma_{i,l}^2\chi_r^2}\right\}$$

Here $A_{i,l}\left(\chi_r^2\right)$ is an arbitrary normalization that accounts for different numbers of samples at the various reduced-χ^2 values. The equation above describes a Gaussian distribution with a width that is linearly dependent on $\sqrt{\chi_r^2}$; the dependence of width on χ_r^2 (or in their case RMSD) is a key result from Hofstetter and Emsley (7).

Fig. 5.2 shows an example of the scatter plot of positional variance and reduced-χ^2. The NMR-derived anisotropic displacement parameters (ADP) were defined as the Gaussian widths in the principal axis system positional distributions drawn at the 95% confidence limit of the chemical shift reduced-χ^2

$$U_{ii,l}^{PAS} = \sigma_{i,l}^2\chi_{r,95\%}^2$$

These were transformed back to the coordinate frame of the crystal structure using standard tensor methods (21, 22), and were incorporated into an edited version of the protein crystal structure file along with the refined active site coordinates from NMR-assisted crystallography. The active sites with positional variances drawn at the 95% probability level are shown in Fig. 5.3.

In the text, the average positional RMSD for a given atom type is reported as

$$\langle r_{av} \rangle = \sqrt{3U_{eq}}$$

where

$$U_{eq} = \frac{1}{N}\sum_{l=1}^{N} U_{eq}^{l}$$

and

$$U_{eq}^{l} = \frac{1}{3}\left(U_{11,l}^{PAS} + U_{22,l}^{PAS} + U_{33,l}^{PAS}\right)$$

Displacement Parameters of the Cofactor and Substrate

The anisotropic parameters for E(Q3)(2AP) from MD simulations show an average displacement parameter of 0.17 and 0.16 Å, with and without DMSO PCM. Shown in figure 5.3, are the ADP ellipsoids for all atoms. Surprisingly the displacement for atoms in the exterior are like those in on the cofactor even those atoms which do not have direct interaction with the cofactor. To understand this interaction we separate the cluster into two regions.

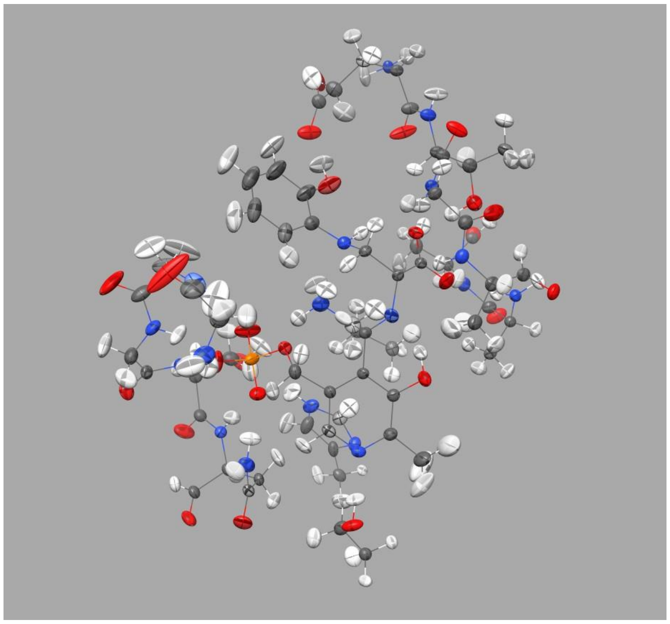


Figure 5.3: The ellipsoids represent the displacement parameters for each residue. This is determined from the full perturbation scheme with the bounding surface shown at 95% probability.

The region containing all measure chemical shifts, the shift-rich region, and the residues surround the cofactor with no measure shift, the shift-deficient-region.

Separating Shift-Rich and Shift-Deficit Regions

The interactions which contribute to the positional uncertainty of the cofactor can be thought of in three parts. First, the intramolecular interactions of the cofactor/substrate which have experimental measurements. Second, the movement of the residues surrounding the cofactor. Third, the motion of the cofactor in the active site or the concerted motion of several residues. These three factors contribute unequally to a change in the chemical shift of the cofactor, Δδ(Substrate(S), Lattice(L)).

$$\Delta\delta = \Delta\delta(\mathrm{S}) + \Delta\delta(\mathrm{L}) + \Delta\delta(\mathrm{S}, \mathrm{L})$$

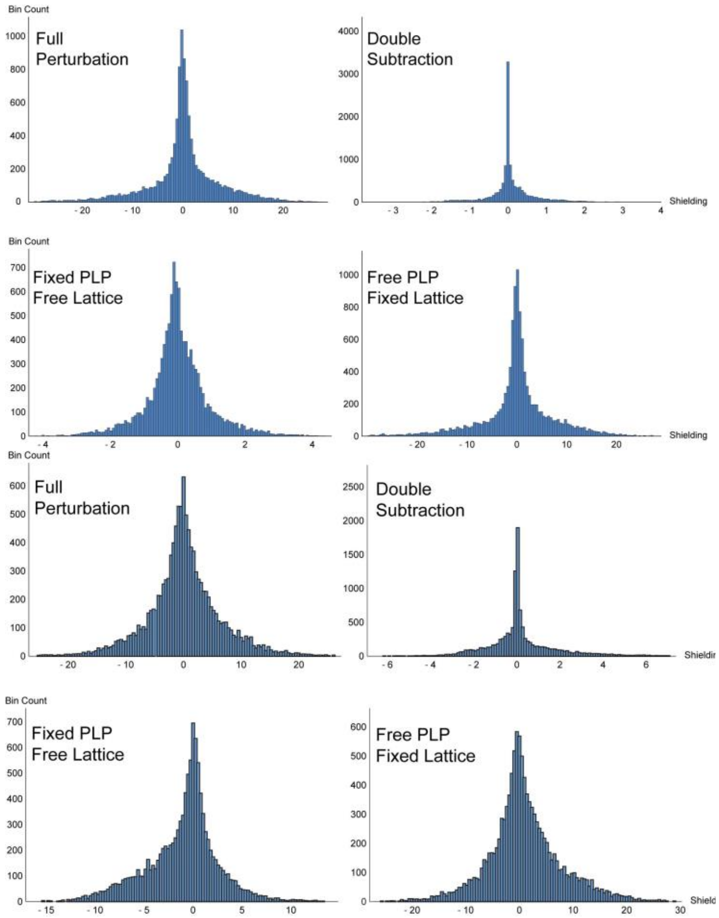


Figure 5.4: Distribution of reference calculated isotropic shielding for Ca of substrate L-serine for E(Q3)(2AP) in left four squares. Distribution of reference calculated isotropic shielding for pyridine nitrogen of PLP Cofactor for E(Q3)(2AP) in right four squares.

Where Δδ(S) is the uncertainty from intramolecular motions of the substrate, Δδ(L) is the chemical shift perturbation of the cofactor from motions of surrounding residues, and Δδ(S,L) come from the concerted motion of residues and the cofactor. An ensemble containing only PLP and no surroundings was used to estimate the displacement parameters of from only intramolecular motion, Δδ(S). Average uncertainties for measured atoms are similar compared to the full perturbation with and without a solvent model meaning that ADP values from the full perturbation are dominated by the intramolecular motion of the cofactor. This is not a surprising result as local geometry is a driving force for chemical shifts. Turing to the surrounding residues in the full perturbation, the average isotropic displacement parameter is similar to the cofactor uncertainties shown in Table A.1 and A.2. This, however, is an unexpected result which would imply motion surrounding the cofactor plays an equal role in chemical shift perturbation. To test this result, we look to additional perturbation scheme with either the shift-rich region being held fixed or the shift-deficient region held fixed.

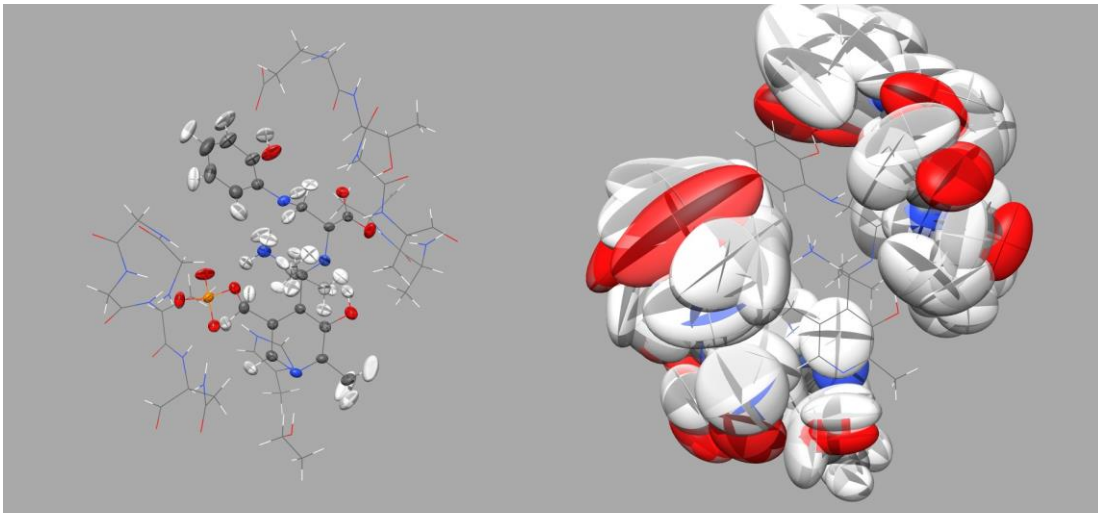


Figure 5.5: The ellipsoids represent the positional uncertainties with ellipsoids are shown with a 95% bounding surface. The left panel represent the displacements from the ensemble with a fixed lattice. The right panel shows the displacements from the fixed cofactor ensemble.

To understand contributions to positional uncertainty from both shift regions, two additional sets of ensembles were generated with alterations to the MD simulations. First is an ensemble with the cofactor and substrate are held at a constant position from an arbitrary structure within the temperature set and the Lattice perturbed (fixed PLP and free Lattice). Second is the inverse, where the Lattice is held at a constant position of an arbitrary structure within the temperature set and the cofactor is perturbed (free PLP and fixed Lattice). All ensembles are generated from the same simulations.

In Figure 5.4, the calculated shielding distributions show the impact of the regions with free movement on c^{α} in the cofactor. These distributions generally show that the motion of the cofactor have a much stronger impact on the chemical shift than the surrounding residues. By separating the regions, we can now qualitatively access the impact of the surrounding resides as a displacement parameter. The smaller the positional uncertainty, the larger of an impact the atom has on the shift-rich region. We also use a double subtraction in which we subtract the perturbed ensemble from the full ensemble to estimate potential error.

Quantifying the effects of the alteration, the shielding values of the altered ensembles were subtracted from the shielding values of the full perturbation structure. The uncertainty for measured shifts is reported in the table below. By using this method, the aim is to remove or greatly reduce specific factors which contribute to the reduced-χ^2. For validation of this method, the fixed PLP with free Lattice ensemble and free PLP with fixed Lattice ensemble were both subtracted from full perturbation to see if any remaining correlation was present. The uncertainties are well above 2 Å and indicate that no correlation between shifts and positional deviation remains. The two ensembles are also added together which resulted in the same isotropic uncertainties as those from the full perturbation ensemble.

With the surrounding lattice, the shift-deficient region, held constant we can again see similar uncertainties to those of the full perturbation and cofactor only models. This small differences in uncertainty between the full perturbation and the cofactor only schemes could be attributed to the lack of surrounding residues even in with a solvent model. In this case the, the fixed lattice should correct this but still the uncertainties are different compared to the full perturbation scheme. This subtle difference is the effect of the cofactor being “misaligned” compared to the native frame and represents the contribution of Δδ(S,L). The implications is that Δδ(S,L) is a corrective term for when the cofactor is perturbed from an equilibrium position based on the position of the surrounding residues.

Constraining the cofactor/substrate positions in the presence of motion from the surrounding amino acids gives an ability to estimate the positional uncertainties of atoms in the shift-deficient region. Using this perturbation scheme, we see an average uncertainty of about 1 Å for E(Q3)(2AP). This result is more inline with expectations of uncertainties outside of the covalent network of the shift-rich region. One downside to using the reduced-χ^2 to measure uncertainties is the loss of information for individual atoms on the cofactor. In Figure 5.4, the distribution of shifts for the pyridine nitrogen of PLP is shown. Unlike with C^α of the substrate serine, the pyridine nitrogen appears more sensitive to motion of βSer377. The β-hydroxyl and the nitrogen form a strong hydrogen bond and so it would be expected that any motion differing from the equilibrium structure would impact the nitrogen chemical shift significantly. This shift distribution also provides evidence that we can include this hydrogen in our shift-rich region. Conservatively, the heavy atoms of βSer377 are left in shift-deficient region. Additionally, we can look at these shift distributions and add some commentary about broad signals at lower temperatures. Its reasonable to assume at or near room temperature, the motions sufficiently average to the equilibrium chemical shift to give smaller line shapes. At low temperatures, the residue motions will be trapped in higher energy equilibriums causing a wider distribution of shifts while the cofactor is expected to have a high degree of freedom.

The motion of surrounding residues, $\Delta\delta(L)$, may be weakly expressed in low temperature simulations and will not have a strong influence on the change of the reduced-χ^2 for the shift-rich region. At higher temperatures, motions such as ring flipping, conformer changes, and larger concerted motions may increase the magnitude of change but is expected to remain relatively small compared to perturbations caused by intramolecular motion.

Conclusion

We expand on a previous method applied to small molecules in order to determine positional uncertainties for ligands in an enzyme active site. Importantly, it provides a way to determine positional uncertainty for hydrogen placed in the active site, critical atoms for understanding enzyme mechanism. We show in various intermediates of tryptophan synthase that heavy atoms in the shift rich region are less than 0.2 Å and that heavy atoms in the lattice are larger , as expected, contrary to the result from the full perturbation ensemble. This method can be further modified for determining positional uncertainties of amino acids from NMR-refined protein structures. It is understood for residues with a majority of chemical shift assigned for heavy atoms, the positional uncertainty is determined by intramolecular motion and calculations can be reasonably estimated in implicit solvent models. It is also possible to invert our defined shift regions using the same methodology described in this work. The shift rich region now becomes the assigned residues surrounding a ligand with few to no chemical shift. The positional uncertainties of the ligand in this case are limited by the number direct interaction with surrounding residues. As a last note, in these studies, we over sampled the number of positions to fit the data. It is possible to fewer points and converge to a similar answer although with less precision.

References

1. J. B. Holmes et al., Imaging active site chemistry and protonation states: NMR crystallography of the tryptophan synthase alpha-aminoacrylate intermediate. Proc Natl Acad Sci U S A **119** (2022).

2. B. G. Caulkins et al., NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. J. Am. Chem. Soc. **138**, 15214-15226 (2016).

3. H. L. Zhang et al., HIV-1 Capsid Function Is Regulated by Dynamics: Quantitative. Atomic-Resolution Insights by Integrating Magic-Angle-Spinning NMR, QM/MM, and MD. J. Am. Chem. Soc. **138**, 14066-14075 (2016).

4. H. Singh et al., Fast Microsecond Dynamics of the Protein-Water Network in the Active Site of Human Carbonic Anhydrase II Studied by Solid-State NMR Spectroscopy. J. Am. Chem. Soc. **141**, 19276-19288 (2019).

5. M. Cordova et al., Structure determination of an amorphous drug through large-scale NMR predictions. Nature Communications **12**, 2964 (2021).

6. M. Baias et al., De novo determination of the crystal structure of a large drug molecule by crystal structure prediction-based powder NMR crystallography. J. Am. Chem. Soc. **135**, 17501-17507 (2013).

7. A. Hofstetter, L. Emsley, Positional Variance in NMR Crystallography. J. Am. Chem. Soc. **139**, 2573-2576 (2017).

8. L. Wang, F. J. Uribe-Romo, L. J. Mueller, J. K. Harper, Predicting anisotropic thermal displacements for hydrogens from solid-state NMR: a study on hydrogen bonding in polymorphs of palmitic acid. Physical Chemistry Chemical Physics **20**, 8475-8487 (2018).

9. D. Niks et al., Allostery and substrate channeling in the tryptophan synthase bienzyme complex: evidence for two subunit conformations and four quaternary States. Biochemistry **52**, 6396-6411 (2013).

10. J. Lai et al., X-ray and NMR crystallography in an enzyme active site: the indoline quinonoid intermediate in tryptophan synthase. J. Am. Chem. Soc. **133**, 4-7 (2011).

11. R. Salomon-Ferrer, A. W. Gotz, D. Poole, S. Le Grand, R. C. Walker, Routine microsecond molecular dynamics simulations with AMBER on

GPUs. 2. Explicit solvent particle mesh Ewald. Journal of chemical theory and computation **9**, 3878-3888 (2013).

12. D. Case et al., AMBER 2018; 2018. University of California, San Francisco.

13. X. He, V. H. Man, W. Yang, T.-S. Lee, J. Wang, A fast and high-quality charge model for the next generation general AMBER force field. The Journal of Chemical Physics **153**, 114502 (2020).

14. D. A. Case et al. (2018) Amber 2018. (University of California, San Francisco).

15. S. K. Burley et al., RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic acids research **49**, D437-D451 (2021).

16. J. A. Maier et al., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. Journal of chemical theory and computation **11**, 3696-3713 (2015).

17. A. Jakalian, D. B. Jack, C. I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. Journal of computational chemistry **23**, 1623-1641 (2002).

18. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics **79**, 926-935 (1983).

19. U. Essmann et al., A smooth particle mesh Ewald method. The Journal of chemical physics **103**, 8577-8593 (1995).

20. J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of computational physics **23**, 327-341 (1977).

21. L. J. Mueller, Tensors and Rotations in NMR. Concept Magn Reson A **38A**, 221-235 (2011).

22. R. P. Young et al., TensorView: A software tool for displaying NMR tensors. Magn Reson Chem **57**, 211-223 (2019).

Chapter 6 – Future Work: NMR-Assisted Refinement of Toho-1-β-Lactamase

Introduction

We have shown remarkable success applying NMR crystallography to the β-subunit active site of tryptophan synthase for several intermediates. With the aminoacrylate intermediate we provide chemical and structural context to understand how benzimidazole (BZI) acts as an inhibitor. In this case, the amino group of βlys87 is positively charged and forms a hydrogen bond with a nitrogen while the other nitrogen is hydrogen bonded to βglu109. This hydrogen bond network prevents BZI, which reacts via the lone pair electrons, from orienting to form a bond with C^{β} of the aminoacrylate. Another finding is weak binding pocket for the newly eliminated β-hydroxyl from substrate serine. This finding is consistent with a minor peak observed in NMR for C^{α} and C^{β} of the aminoacrylate as well as inconsistent electron density for the active site water. We proposed the minor peak comes from a second population of the aminoacrylate where the active site water closest to C^{β} is missing. From all this we are also able to determine anisotropic displacement parameters in the active site for this intermediate and previously studied intermediates. The aminoacrylate was found to have a positional uncertainty of 0.11 Å for heavy atoms and 0.17 Å for hydrogen placed through NMR crystallography. The gains in structural certainty and high confidence in the proton positions are important for understanding the chemistry and reactivity of any active sites. These NMR-refined structures show the significance of integrating techniques for determining 3D-chemical structures which elucidate enzymatic mechanism not able be determined from a single technique. This powerful framework provided can be applied to other intermediates of tryptophan synthase such as the serine external aldimine, the tryptophan external aldimine, and the α-subunit active site. Beyond tryptophan synthase other PLP-fold types should be investigated so that the role of protonation states in reaction specificity can be better understood.

The other advance to NMR crystallography is the use of an “outside-in” approach undertaken with the internal aldimine. It provides a framework for using unambiguous assignments of mechanistically important residues in active sites to understand their catalytic role. Where this approach is most useful is when ligands are not readily available with isotopic enrichment. In Toho-1-β-lactamse nearly 80% of side chains have been assigned and this provides an opportunity to combine the plethora of shifts with NMR-crystallography to refine the entire protein. In the next sections, the approaches to these refinements are discussed.

NMR Crystallography of Toho-1- β-lactamase

Toho-1- β -lactamase is part of a class of enzymes which provide antibiotic resistance to β-lactams such as penicillin (1-3). This resistance poses a threat for harmful bacterial to resist treatment. β-lactamase function by using serine in a hydrolysis reaction to open the ring of the β-lactam antibiotic causing it to be inactivated. The underlying proton transfer in this reaction is still not well understood although key residues, Ser 70, Lys 73, and Glu 166, have been identified. A neutron structure (PDB ID 2XQZ) has been solved for the apo enzymes but a neutron structure has yet to be deposited. NMR crystallography can be used here to identify the protonation states of residues in the active site(4, 5). In this approach we are using an “outside-in” approach which using chemical shift from surrounding residues instead of and a ligand.

For the apo enzyme, a neutron structure of has been solved (PDB ID: 2XQZ) and acts as a framework for our initial investigation. To do this we employ the integrated approach of NMR crystallography to build “chemically-rich” structures. For enzyme active sites, traditionally NMR crystallography has been the application of solid-state NMR, X-ray crystallography, and computational chemistry to determine a structure with atomic resolution and determine protonation states. The X-ray structure severed as the heavy atom scaffolding used to build testable models. Various protonation states can be explored via first principles and tested against the chemical shifts determined from NMR. Given enough chemical shifts in the modeled area, significantly lower positional uncertainties are possible as shown with the aminoacrylate intermediate of tryptophan synthase.

In previous studies, chemical shifts for a cofactor and substrate were used as the experimental shifts against which our models are compared. In the case of Toho-1-β-lactamase, avibactam, the substrate of interest, is not readily available with ^{13}C- or ^{15}N enrichment posing an issue for modeling the active site. Instead, an inverted enrichment scheme is proposed where the surrounding residues are U-^{13}C,^{15}N-enriched. As discussed in chapter 1, this labeling strategy increases spectral complexity but is remedied using 2D and 3D NMR experiments where atom correlations are used to resolve individual residues. Additionally, the challenge is that side chain assignments be made for atoms that are most sensitive to the changes in the active site. Examples are the ε-amino group and the Cε of Lys73 as well as the C^{δ} of Glu166 which are mechanistically relevant residues. Fortunately, an impressive degree of unambiguous assignment has been made for Toho-1-β-lactamase. The solid-state NMR assignments are unpublished at this time but high confidence in assignments for residues Ser70, Lys73, and Glu166 allow models of the active site to be tested. Since no

Table 6.1: Experimental and First Principles shifts (ppm) to Toho-1-β-lactamase.

Residues	Atom	Experimental	Calculated
Ser 70	C^{α}	62.7	59.5
Lys 73	C^{α}	60.2	57.9
	C^{β}	32.4	33.1
	C^{γ}	24.1	25.3
	C^{δ}	29.6	28.5
	C^{ε}	41.9	39.6
	N^{ζ}	32.6	32.6
Glu 166	C^{α}	55.4	55.0
	C^{β}	28.8	28.5
	C^{γ}	36.1	33.8
	C^{δ}	182.8	181.0
Reduced-χ2	-	-	1.32

direct measurements can be made for avibactam, it is critical to model the active site before and after the addition to understand each residue's acid/base roles. The approach can be described as an “outside-in” variation of previously studied active sites.

Preliminary chemical shifts have been calculated using a cluster-based model approach for the apo enzyme using the neutron structure 2XQZ. Table 6.1 shows the reduced-χ^2 statistics for a model following the protonation states determined in the neutron structure has good agreement with experimental data. The result is Lys73 starts as an acid in the catalytic cycle which has been proposed to act as a base in the acetylation mechanism. Currently, no neutron structure is available for Toho-1-β-lactamase with avibactam bound so various protonation states of residues in the active site will need to be modeled against experimental shifts to determine the protonation state in the proceeding step. This “outside-in” approach to NMR-crystallography will help understand the reaction mechanism and determine the acid or base roles of each residue or, of equal importance, those models which do not fit experimental data.

Full Structure Refinement

NMR has shown to be a useful technique for the structure refinement of proteins because it is not bound to some of the restraints imposed for high-resolution crystallography. CryoEm offers a solution for systems that cannot be crystallized or do not have a high degree of order. NMR can provide an additional set of distance restraints and torsional angles to assist in resolution improvement even in high-resolution structures. Additionally, the chemical shift of atoms can be used as a restraint to determine the quality of a structure. Many programs offer a way to predict the chemical shift such as AFNMR, SHIFTX2, Sparta+, and others. Some of these prediction methods have an RMSE of less than 1 ppm for backbone chemical shifts and C^β chemical which is remarkable precision but the prediction of atoms further down the side chain has less precision. From a mechanistic standpoint, these chemical shifts are most important for understanding interactions made by the protein or biomolecule and therefore critical that accurate predictions can be made to test models. We propose a method for the full refinement of Toho-1-β-lactamase using a solid-state NMR, first-principles calculations at a DFT level of theory, and the reported neutron structure (2XQZ). While this is an ambitious goal given that first-principles calculations are typically a bottleneck for structure calculation given the high computational cost which grows severely as the number of atoms increases. Given current computational resources, it is not feasible to geometry optimize an entire protein structure at a DFT level of theory but the amino acids could be segmented into small clusters and iteratively refined. The question is whether such an undertaking is necessary for understanding the chemistry of residues. As reported before, structures refined using NMR, first principles, and X-ray have offered incredible resolution down to the orientation of water molecules. In a qualitative assessment of the neutron structure, 5 Å clusters were made for residues Ser70, Lys73, and Glu166. For these clusters, only a hydrogen geometry optimization using AM1 level of theory was performed with the heavy atoms in their crystallographic location. After geometry optimization, NMR shielding calculations are performed in a two-tier, locally dense basis set. The calculated shifts are compared against experimental shifts using the reduced-χ^2 statistic. As shown in Table 6.2 is that there is very little agreement between experiment and the model. Comparing the crystal structure to the refined model for the “outside-in” approach we see atomic movements are generally only tenths of angstroms and yet these adjustments led to a structure with very good agreement. It is worth noting that NMR is a sensitive reporter of the local environment and so this result is not unexpected. Knowing that a structure can be proposed that agrees with the electron density and NMR chemical shifts would be a remarkable feat.

Table 6.2: Reduced-χ2 of Predicted Chemical Shifts using Heavy Atom Crystal Coordinates

Residue	*Atom*	*Experimental*	*Calculated*
Ser 70	C	176.0	164.7
	N	119.0	111.0
	CA	62.7	52.1
Reduced-χ2			36.8
Lys 73	C	179.2	172.3
	CD	29.6	37.9
	CG	24.1	28.4
	NZ	32.6	38.3
	CB	32.4	38.5
	CE	41.9	44.0
	CA	60.2	58.4
	N	122.6	112.6
Reduced-χ2			10.8
Glu 166	C	177.9	177.0
	CG	36.1	35.1
	CD	182.7	173.9
	CB	28.8	33.8
	N	116.9	110.3
	CA	55.3	55.0
Reduced-χ2			8.1

To accomplish a full structure refinement, we propose a" bead-like" refinement that takes 5 Å clusters around a residue using atom coordinates from a crystal structure. The target residue and all hydrogens in the cluster are geometry optimized, initially using AM1 level of theory. Once the optimization completes, the PDB file is updated with the optimized coordinates and the subsequent amino acid is optimized in the same fashion. If there are any potential crystal contacts, those are also included in the model. After an initial refinement using AM1, the level of theory can be changed to DFT. While the refinement is still in the early stages, we expect promising results from this refinement and the prospects of working with such a structure. It would be possible to fit a linear rescaling specifically to proteins that could reduce the RMSE of shift prediction, reveal critical residue interactions, and allows for ligands to be modeled along with the protein structure.

Methods

Active Site Models

First Principles calculations were performed as previously described (4, 5) using a cluster-based model approach to the residues of interest. The cluster was constructed from the Neutron structure 2XQZ and selected all atoms within 8 Å of Ser 70, Lys 73, and Glu 166. Prior to atom selection, the following changes were made using Chimera Version 1.16(6). The model was solvated using TIP3PBOX with a box size of 5. Crystallographic waters were not removed during the solvation process. Waters were then minimized, within the chimera interface, with the following restrictions: All atoms in the protein structure we held fixed to their crystallographic location as well as the oxygen in crystal waters and automatically set a periodic boundary. Within the 8 Å selection of atoms, the following modifications are made: (1) residues with two or more backbone atoms are mutated to alanine (Asp 233, Tyr 129, Leu 122, Gln 128) ; (2) residues with two or more side chain atoms and no backbone were modified by removing backbone atom and replacing C^{α} with a methyl group (Leu 138,Val 148, Phe, 160, Leu 162, Met 186, Trp 210, Thr 216, Ile 247); (3) residues at terminating end of the 8 Å cut were modified either with the N-terminal nitrogen replaced as a hydrogen atom or amide group from the neighboring residue. The final model included, 8 crystal waters, and a total of 881 atoms. Protonation states in the initial model follow those found in the neutron structure with Ser 70 having a β-hydroxyl, Lys 73 being a charge ε-amino group, and Glu 166 side chain carboxy group being negatively charged.

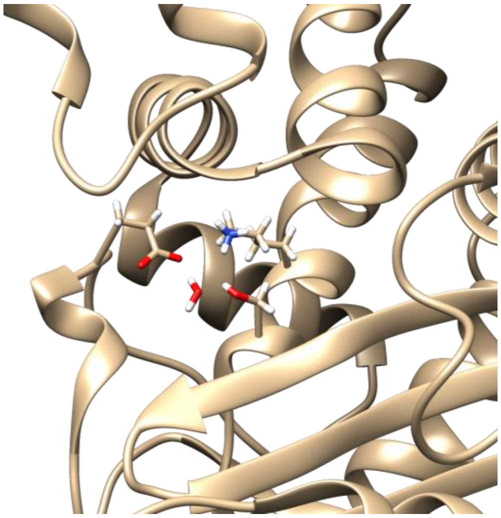


Figure 6.1: The active site of Toho showing Ser 70, Lys 73, Glu 116 and water 354.

The cluster model was geometry optimized in Gaussian 16 (7) at the DFT level of theory using the Grimme D3 (8) empirical dispersion correction and a two-tier, locally-dense basis set with 6-31G(d,p) for side chain starting at C^{α} for Ser 70, Lys 73, and Glu 166 and for crystal water 357 and 6-31G for the remaining atoms. Atoms within 4 Å of the listed residues and all hydrogen atoms were geometry optimized without constraint except water 357. Heavy atoms outside this region were held fixed at their crystallographic location along with the oxygen of water 357.

NMR shielding calculations were performed on the geometry optimized structure at the DFT level of theory with a three-tier, locally dense basis set for Ser 70, Lys 73, Glu 166, and water 357 at 6-311++G(2d,p), atoms within 4 Å of these at 6-311G(d,p), and all remaining atoms at 6-31G. The NMR shielding values were converted to chemical shifts using the linear rescaling described (9, 10)in chapter 4 with the respective root-mean-square errors. The reduced-χ^2 statistic (11), also described in chapter 4, is used to measure agreement between experimental shifts and models.

References

1. S. J. Tomanicek et al., Neutron Diffraction Studies of a Class A beta-Lactamase Toho-1 E166A/R274N/R276N Triple Mutant. J Mol Biol **396**, 1070-1080 (2010).
2. P. S. Langan et al., Substrate Binding Induces Conformational Changes in a Class A beta-lactamase That Prime It for Catalysis. Acs Catalysis **8**, 2428-2437 (2018).
3. V. V. Sakhrani et al., Toho-1 beta-lactamase: backbone chemical shift assignments and changes in dynamics upon binding with avibactam. Journal of Biomolecular Nmr **75**, 303-318 (2021).
4. J. B. Holmes et al., Imaging active site chemistry and protonation states: NMR crystallography of the tryptophan synthase alpha-aminoacrylate intermediate. Proc Natl Acad Sci U S A **119** (2022).
5. B. G. Caulkins et al., NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. J. Am. Chem. Soc. **138**, 15214-15226 (2016).
6. E. F. Pettersen et al., UCSF chimera - A visualization system for exploratory research and analysis. J Comput Chem **25**, 1605-1612 (2004).
7. M. J. Frisch et al. (2016) Gaussian 16 Rev. C.01. (Wallingford, CT).
8. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys **132**, 154104 (2010).
9. J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller, G. J. Beran, Benchmark fragment-based ^{1}H, ^{13}C, ^{15}N and ^{17}O chemical shift predictions in molecular crystals. Physical Chemistry Chemical Physics **18**, 21686-21709 (2016).
10. J. D. Hartman, T. J. Neubauer, B. G. Caulkins, L. J. Mueller, G. J. Beran, Converging nuclear magnetic shielding calculations with respect to basis and system size in protein systems. J. Biomol. NMR **62**, 327-340 (2015).
11. D. P. Shoemaker, C. W. Garland, Experiments in physical chemistry (ed. 2, 1967).

Appendix

Anisotropic displacement parameters
Table of displacement parameters.

Tables

Table A.1: Positional uncertainties in Å of E(Q3)(2AP) / PDB 4HPJ

		Full Perturbation	Free Cofactor Fixed Surrounding	Fixed Cofactor Free Surround	PLP Only
L-Serine	Co	0.135	0.148	-	0.149
	Ca	0.134	0.146	-	0.143
	Cb	0.14	0.15	-	0.148
	Schiff Base Nitrogen	0.176	0.177	-	0.189
	O1	0.206	0.207	-	0.202
	O2	0.166	0.175	-	0.17
PLP	PN	0.157	0.163	-	0.174
	C2	0.156	0.16	-	0.157
	C2'	0.214	0.211	-	0.22
	C3	0.148	0.166	-	0.151
bLys87	Ne	0.189	0.19	-	0.193
Ligand	N-2AP	0.154	0.162	-	0.163
Average Lattice RMSD		-			
Average Cofactor RMSD		0.164583	-		

Table A.2: Positional uncertainties in Å of E(Q3)(2AP) / PDB 4HPJ using DCM PCM

		Full Perturbation	Free Cofactor Fixed Surrounding	Fixed Cofactor Free Surround	PLP Only
L-Serine	Co	0.135	0.148	-	0.149
	Ca	0.134	0.146	-	0.143
	Cb	0.14	0.15	-	0.148
	Schiff Base Nitrogen	0.176	0.177	-	0.189
	O1	0.206	0.207	-	0.202
	O2	0.166	0.175	-	0.17
PLP	PN	0.157	0.163	-	0.174
	C2	0.156	0.16	-	0.157
	C2'	0.214	0.211	-	0.22
	C3	0.148	0.166	-	0.151
bLys87	Ne	0.189	0.19	-	0.193
Ligand	N-2AP	0.154	0.162	-	0.163
Average Lattice RMSD		-			
Average Cofactor RMSD		0.164583	-		

Table A.3: Positional uncertainties in Å of E(AA) / PDB 4HN4

		Full Perturbation	Free Cofactor Fixed Surrounding	Fixed Cofactor Free Surround	PLP Only
L-Serine	Co	0.135	0.148	-	0.149
	Ca	0.134	0.146	-	0.143
	Cb	0.14	0.15	-	0.148
	Schiff Base Nitrogen	0.176	0.177	-	0.189
	O1	0.206	0.207	-	0.202
	O2	0.166	0.175	-	0.17
PLP	PN	0.157	0.163	-	0.174
	C2	0.156	0.16	-	0.157
	C2'	0.214	0.211	-	0.22
	C3	0.148	0.166	-	0.151
bLys87	Ne	0.189	0.19	-	0.193
Ligand	N-2AP	0.154	0.162	-	0.163
Average Lattice RMSD		-			
Average Cofactor RMSD		0.164583	-		

Beaded Refinement

I present bash script for the beaded refinement of toho which can be expanded to and system. There are also depiction of the cluster and their relative size.

```
#!/bin/bash

cd ~/Toho_full_opt/
```

```
#residue incrementer thank the forum boards for all of
this################################################################
#note set n=1 if restarting

n=139;#the variable that I want to be incremented
next_n=$[$n+1]
sed -i "/#the variable that I want to be incremented$/s/=.*#/=$next_n;#/" ${0}
echo $n >> tracker_NMR_Hopt_NoHVopt.txt

rm ~/Toho_full_opt/ala_NMR_Hopt_NoHVopt.pdb

awk '$5=='$n' {print}' ~/Toho_full_opt/2xqz_XCC_3_7wshell_read_og.pdb >>
~/Toho_full_opt/ala_NMR_Hopt_NoHVopt.pdb


rm ~/Toho_full_opt/test2_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/ala2_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/ala3_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt


#find residues with 5A of selected
residued#################################################################
###########


while read num1 num2 num3 num4 num5 num6 num7 num8 num9
do
        echo "$num1" >> ~/Toho_full_opt/test2_NMR_Hopt_NoHVopt.txt
        awk '(('$num6'-($6))^2 + ('$num7'-($7))^2 +('$num8'-($8))^2)^0.5 < 5 {print
$5}' ~/Toho_full_opt/2xqz_XCC_3_7wshell_read_og.pdb >>
~/Toho_full_opt/ala2_NMR_Hopt_NoHVopt.txt
done < ~/Toho_full_opt/ala_NMR_Hopt_NoHVopt.pdb

awk '!seen[$0]++' ~/Toho_full_opt/ala2_NMR_Hopt_NoHVopt.txt >
~/Toho_full_opt/ala3_NMR_Hopt_NoHVopt.txt

while read num1
do
        echo "$num1" >> ~/Toho_full_opt/test2_NMR_Hopt_NoHVopt.txt
        awk '$5=='$num1' {print}'
~/Toho_full_opt/2xqz_XCC_3_7wshell_read_og.pdb >>
~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt
done < ~/Toho_full_opt/ala3_NMR_Hopt_NoHVopt.txt
```

```
#create gaussuan
file##################################################################
################################

rm ~/Toho_full_opt/termN_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/termCO_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/Nlist_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/Nlist2_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/test4_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/N_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/test100_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/test200_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/termCO3_NMR_Hopt_NoHVopt.txt



awk '$3=="N" {print}' ~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/Nlist_NMR_Hopt_NoHVopt.txt
awk '$3=="H" {print}' ~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/Nlist_NMR_Hopt_NoHVopt.txt



while read num1 num2 num3 num4 num5 remainder
do
        if [[ $num4 == "PRO" ]]
                then
                        awk '$3=="C" && $5=='$num5' - 1 {print}'
~/Toho_full_opt/2xqz_XCC_3_7wshell_read_og.pdb >>
~/Toho_full_opt/termCO3_NMR_Hopt_NoHVopt.txt

                else

                        #add second H to terminal NH2
                        awk '$3=="C" && $5=='$num5' - 1 {print}'
~/Toho_full_opt/2xqz_XCC_3_7wshell_read_og.pdb >>
~/Toho_full_opt/termN_NMR_Hopt_NoHVopt.txt
        fi

        #add NH2 to terminal CO
```

```
	awk '$3=="N" && $5=='$num5' + 1 {print}'
~/Toho_full_opt/2xqz_XCC_3_7wshell_read_og.pdb >>
~/Toho_full_opt/termCO_NMR_Hopt_NoHVopt.txt
	awk '$3=="H" && $5=='$num5' + 1 {print}'
~/Toho_full_opt/2xqz_XCC_3_7wshell_read_og.pdb >>
~/Toho_full_opt/termCO_NMR_Hopt_NoHVopt.txt
	awk '$3=="CA" && $5=='$num5' + 1 {print}'
~/Toho_full_opt/2xqz_XCC_3_7wshell_read_og.pdb >>
~/Toho_full_opt/termCO_NMR_Hopt_NoHVopt.txt

done < ~/Toho_full_opt/Nlist_NMR_Hopt_NoHVopt.txt


while read num1 num2 num3 num4 num5 remainder
do
if [[ $num4 == "PRO" ]]
	then
		awk '$3=="CD" && $5=='$num5' {print}'
~/Toho_full_opt/2xqz_XCC_3_7wshell_read_og.pdb >>
~/Toho_full_opt/termCO3_NMR_Hopt_NoHVopt.txt

fi
done < ~/Toho_full_opt/termCO_NMR_Hopt_NoHVopt.txt

cat ~/Toho_full_opt/termCO_NMR_Hopt_NoHVopt.txt
~/Toho_full_opt/termCO3_NMR_Hopt_NoHVopt.txt >
~/Toho_full_opt/termCO4_NMR_Hopt_NoHVopt.txt

awk '!seen[$0]++' ~/Toho_full_opt/termN_NMR_Hopt_NoHVopt.txt >
~/Toho_full_opt/termN2_NMR_Hopt_NoHVopt.txt
awk '!seen[$0]++' ~/Toho_full_opt/termCO4_NMR_Hopt_NoHVopt.txt >
~/Toho_full_opt/termCO2_NMR_Hopt_NoHVopt.txt

echo none >>test4_NMR_Hopt_NoHVopt.txt

grep -F -x -v -f ~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt
~/Toho_full_opt/termN2_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/test4_NMR_Hopt_NoHVopt.txt

while read num1 num2 num3 num4 num5 remainder
do
	awk '$3=="N" && $5=='$num5' + 1 {print}'
~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/N_NMR_Hopt_NoHVopt.txt
```

```
done < ~/Toho_full_opt/test4_NMR_Hopt_NoHVopt.txt

while read num1 num2 num3 num4 num5 remainder
do
	awk '$3=="H" && $5=='$num5' + 1 {print}'
~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/N_NMR_Hopt_NoHVopt.txt
done < ~/Toho_full_opt/test4_NMR_Hopt_NoHVopt.txt

grep -F -x -v -f ~/Toho_full_opt/N_NMR_Hopt_NoHVopt.txt
~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.txt

grep -F -x -v -f ~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt
~/Toho_full_opt/termCO2_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt

while read num1 num2 num3 num4 num5 num6 num7 num8 num9
do
	#echo $num2
	#echo $num9
	awk '$2=='$num2' && $NF=="N" {print}'
~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.txt
 done < test101_NMR_Hopt_NoHVopt.txt

while read num1 num2 num3 num4 num5 num6 num7 num8 num9
do
	#echo $num2
	#echo $num9
	awk '$2=='$num2' && $NF=="H" {print}'
~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.txt
 done < ~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt

while read num1 num2 num3 num4 num5 num6 num7 num8 num9
do
	#echo $num2
	#echo $num9
	awk -v OFS=' ' '$2=='$num2' && $3=="CA" {print gensub (/[^[:blank:]]+/,
"H", 9)}' ~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.txt
 done < ~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt
```

```
while read num1 num2 num3 num4 num5 num6 num7 num8 num9
do
	#echo $num2
	#echo $num9
	awk -v OFS=' ' '$2=='$num2' && $3=="CD" {print gensub (/[^[:blank:]]+/,
"H", 9)}' ~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.txt
 done < ~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt

while read num1 num2 num3 num4 num5 num6 num7 num8 num9
do
	#echo $num2
	#echo $num9
	awk -v OFS=' ' '$2=='$num2' && $3=="C" {print gensub (/[^[:blank:]]+/, "H",
9)}' ~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.txt
 done < ~/Toho_full_opt/test101_NMR_Hopt_NoHVopt.txt


while read num1 num2 num3 num4 num5 num6 num7 num8 num9
do
	#echo $num2
	#echo $num9
	awk -v OFS=' ' '$2=='$num2' && $3=="N" {print gensub (/[^[:blank:]]+/, "H",
9)}' ~/Toho_full_opt/N_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.txt
 done < ~/Toho_full_opt/N_NMR_Hopt_NoHVopt.txt



rm ~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.pdb

cp ~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.txt
~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.pdb




rm ~/Toho_full_opt/charge_NMR_Hopt_NoHVopt.txt
rm ~/Toho_full_opt/charge2_NMR_Hopt_NoHVopt.txt
```

```
awk -v OFS=' ' ' $5==260 {print $4 "end  " $5 "             -1"}'
~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/charge_NMR_Hopt_NoHVopt.txt
awk -v OFS=' ' ' $4=="ASP" || $4=="GLU" {print $4 "    " $5 "             -1"}'
~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/charge_NMR_Hopt_NoHVopt.txt
awk -v OFS=' ' '$4=="PO4" {print $4 "      " $5 "             -3"}'
~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/charge_NMR_Hopt_NoHVopt.txt

awk -v OFS=' ' '$4=="ARG" || $4=="LYS" || $5==1 {print $4 "  " $5 "             1"}'
~/Toho_full_opt/ala4_NMR_Hopt_NoHVopt.txt >>
~/Toho_full_opt/charge_NMR_Hopt_NoHVopt.txt

awk '!seen[$0]++' ~/Toho_full_opt/charge_NMR_Hopt_NoHVopt.txt >
~/Toho_full_opt/charge2_NMR_Hopt_NoHVopt.txt

crg="0"

while read num1 num2 num3
do
        crg=$(($crg + $num3))
        echo $crg
done < ~/Toho_full_opt/charge2_NMR_Hopt_NoHVopt.txt

echo "%nprocshared=32" > ~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
echo "%mem=64GB" >> ~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
echo "# opt am1 nosymm integral=(grid=99590,acc2e=12)" >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
echo "scf=(tight,xqc,maxcycles=256,maxconventionalcycles=50)" >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
echo -en '\n' >> ~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
echo "whole refinement" >> ~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
echo -en '\n' >> ~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
echo "$crg 1" >> ~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf

while read num1 num2 num3 num4 num5 num6 num7 num8 num9
do

        ####if  [ $num5 == "$n" ]
        #####then
                ###printf  "$num9  -1  " >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
```

```
                    ####LC_NUMERIC="en_US.UTF-8" printf " %.3f " $num6 >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
                    ####LC_NUMERIC="en_US.UTF-8" printf " %.3f " $num7 >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
                    ####LC_NUMERIC="en_US.UTF-8" printf " %.3f \n" $num8 >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
          ####el
if  [[ $num9 =~ H ]]
          then
                    printf  "$num9  0  " >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
                    LC_NUMERIC="en_US.UTF-8" printf " %.3f " $num6 >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
                    LC_NUMERIC="en_US.UTF-8" printf " %.3f " $num7 >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
                    LC_NUMERIC="en_US.UTF-8" printf " %.3f \n" $num8 >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
          else
                    printf  "$num9  -1  " >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
                    LC_NUMERIC="en_US.UTF-8" printf " %.3f " $num6 >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
                    LC_NUMERIC="en_US.UTF-8" printf " %.3f " $num7 >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf
                    LC_NUMERIC="en_US.UTF-8" printf " %.3f \n" $num8 >>
~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf

fi
done < ~/Toho_full_opt/ala5_NMR_Hopt_NoHVopt.txt

echo -en '\n\n' >> ~/Toho_full_opt/AA_${n}_NMR_Hopt_NoHVopt.gjf


sbatch -p batch --nodes=1 --ntasks=32 --mem=64gb --time=10-00:00:00
~/Toho_full_opt/AAT_Aex_2_NMR_Hopt_NoHVopt.sh

exit

sleep 120

until [ -f ~/Toho_full_opt/current_NMR_Hopt_NoHVopt.log ]
do
          squeue -u jholm0l0
          sleep 900
```

```
done
echo "File found"

tail -f ~/Toho_full_opt/current_NMR_Hopt_NoHVopt.log | sed '/Normal termination of Gaussian/ q;/Error termination via Lnk1e/ q'


#while read i
#do
#	if grep -q "Normal termination of Gaussian" ~/Toho_full_opt/current_NMR_Hopt_NoHVopt.log
#	then
#	break
#fi
#done < ~/Toho_full_opt/current_NMR_Hopt_NoHVopt.log

bash ~/Toho_full_opt/updatepdb_NMR_Hopt_NoHVopt.sh
```

```
#!/bin/bash

cd ~/Toho_full_opt/

n=11;#the variable that I want to be incremented
next_n=$[$n+1]
sed -i "/#the variable that I want to be incremented$/s/=.*#/=$next_n;#/" ${0}
echo $n

rm ala6_631g.txt
rm ala7_631g.txt
rm ala8_631g.txt
rm update1_631g.pdb
rm update2_631g.pdb
rm update3_631g.pdb


var=$(wc -l ala5_631g.txt)
nb_lines=${var%% *}


echo $nb_lines

sed -n '/Input orientation/,/Rotational constants/p'
~/Toho_full_opt/current_631g.log > ~/Toho_full_opt/test500_631g.txt

head -n -2 ~/Toho_full_opt/test500_631g.txt > ~/Toho_full_opt/test501_631g.txt

var1=$(wc -l ~/Toho_full_opt/test501_631g.txt)
nb_lines2=${var1%% *}

echo $nb_lines
echo $nb_lines2

nb_lines3=$(expr $nb_lines2 - $nb_lines)

echo $nb_lines3


awk 'NR > '$nb_lines3' { print }' ~/Toho_full_opt/test501_631g.txt >
~/Toho_full_opt/test502_631g.txt


while read num1 num2 num3 num4 num5 num6
```

```
do
        awk -v OFS=' ' 'NR=='$num1' {print gensub (/[^[:blank:]]+/, '$num6', 8)}'
~/Toho_full_opt/ala5_631g.txt >> ~/Toho_full_opt/ala6_631g.txt
        awk -v OFS=' ' 'NR=='$num1' {print gensub (/[^[:blank:]]+/, '$num5', 7)}'
~/Toho_full_opt/ala6_631g.txt >> ~/Toho_full_opt/ala7_631g.txt
        awk -v OFS=' ' 'NR=='$num1' {print gensub (/[^[:blank:]]+/, '$num4', 6)}'
~/Toho_full_opt/ala7_631g.txt >> ~/Toho_full_opt/ala8_631g.txt

done < ~/Toho_full_opt/test502_631g.txt

var4=$(wc -l ~/Toho_full_opt/test101_631g.txt)
nb_lines4=${var4%% *}
var5=$(wc -l N_631g.txt)
nb_lines5=${var5%% *}

echo $nb_lines4
echo $nb_lines
echo $nb_lines5

nb_lines5=$(expr $nb_lines5 / 2)

echo $nb_lines5

nb_lines6=$(expr $nb_lines - $nb_lines4 - $nb_lines5 + 1)

echo $nb_lines6

awk 'NR < '$nb_lines6' { print }' ~/Toho_full_opt/ala8_631g.txt >
~/Toho_full_opt/ala9_631g.txt

rm ~/Toho_full_opt/testpdb_631g.txt
rm ~/Toho_full_opt/ala10_631g.txt


while read num1 num2 num3 num4 num5 num6 num7 num8 num9
do
        sed -i 's/\('$num1'[[:space:]]*\)\('$num2'[[:space:]]*\)'$num3'.*/'$num1'
'$num2'  '$num3'  '$num4'  '$num5'  '$num6'  '$num7'  '$num8'  '$num9'/g'
2xqz_XCC_3_7wshell_read.pdb
done < ala9_631g.txt

cp ~/Toho_full_opt/current_631g.log ~/Toho_full_opt/AA_${n}_631g.log
rm ~/Toho_full_opt/current_631g.log
```

```
if [ $n == 260 ]
	then
		exit
	else
		echo "not finished yet"
fi

bash gaussgen_631g.sh
```

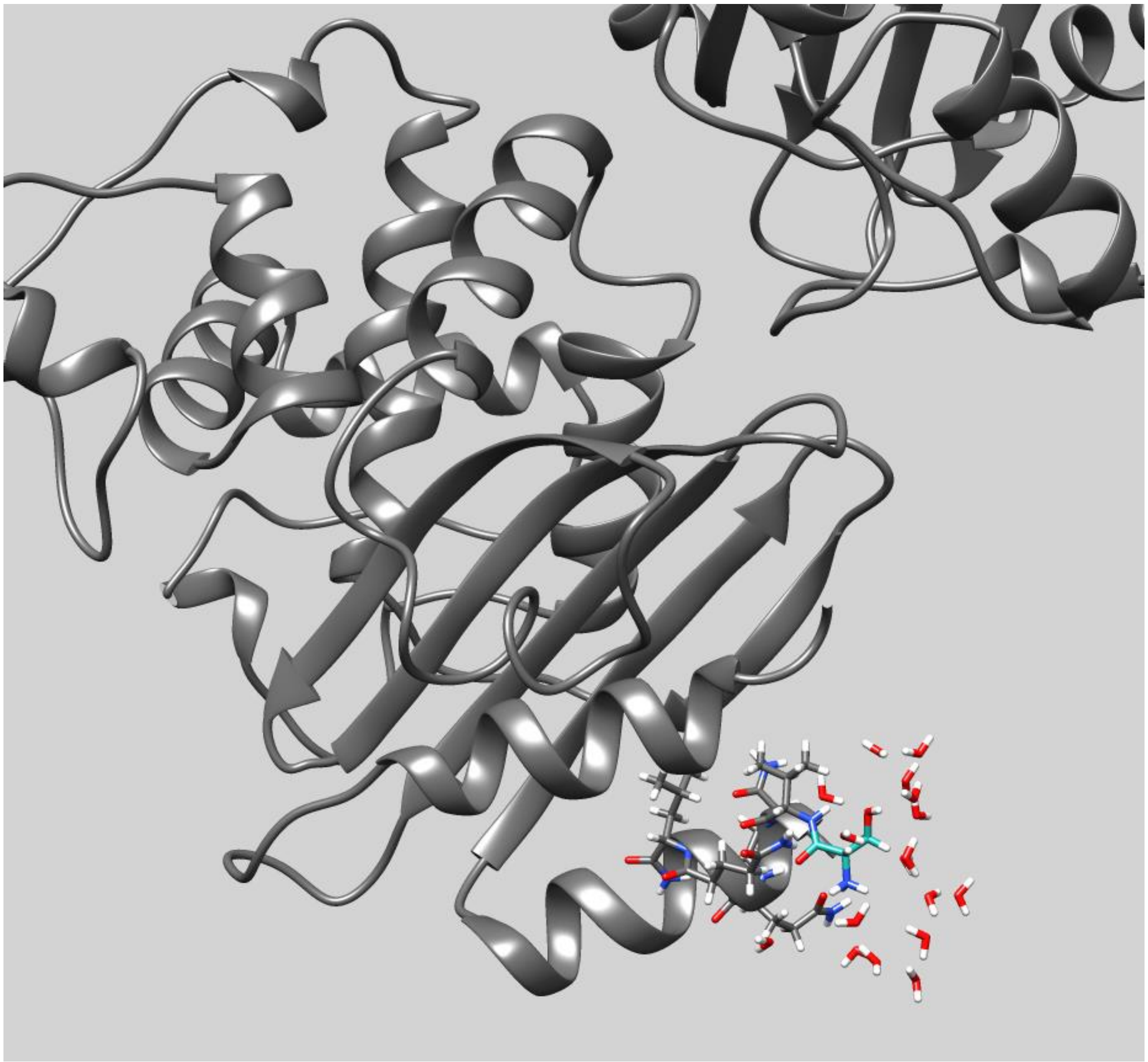


Figure A.1 Beaded refinement of residues 28

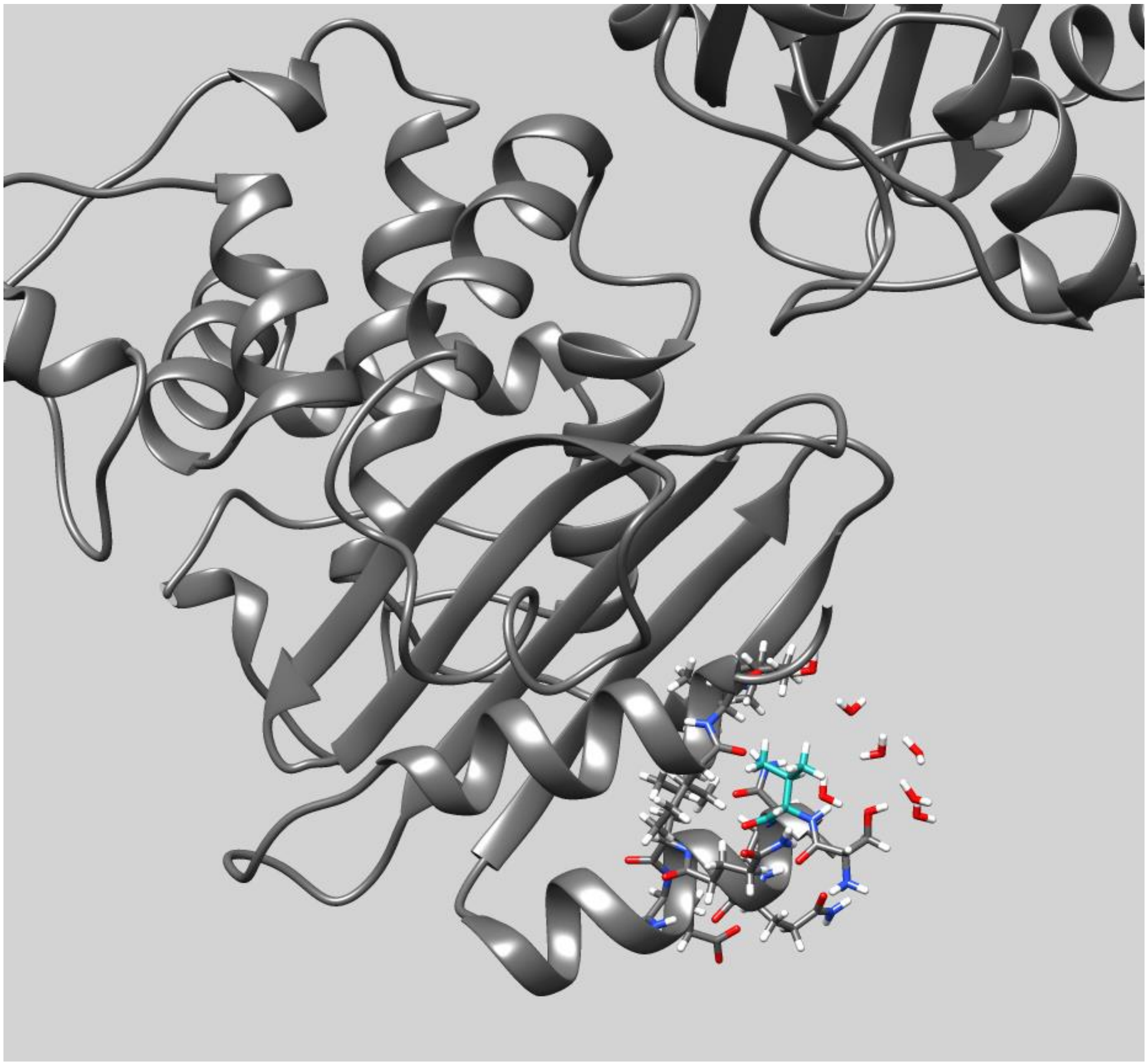


Figure A.2:Beaded refinement of residues 29

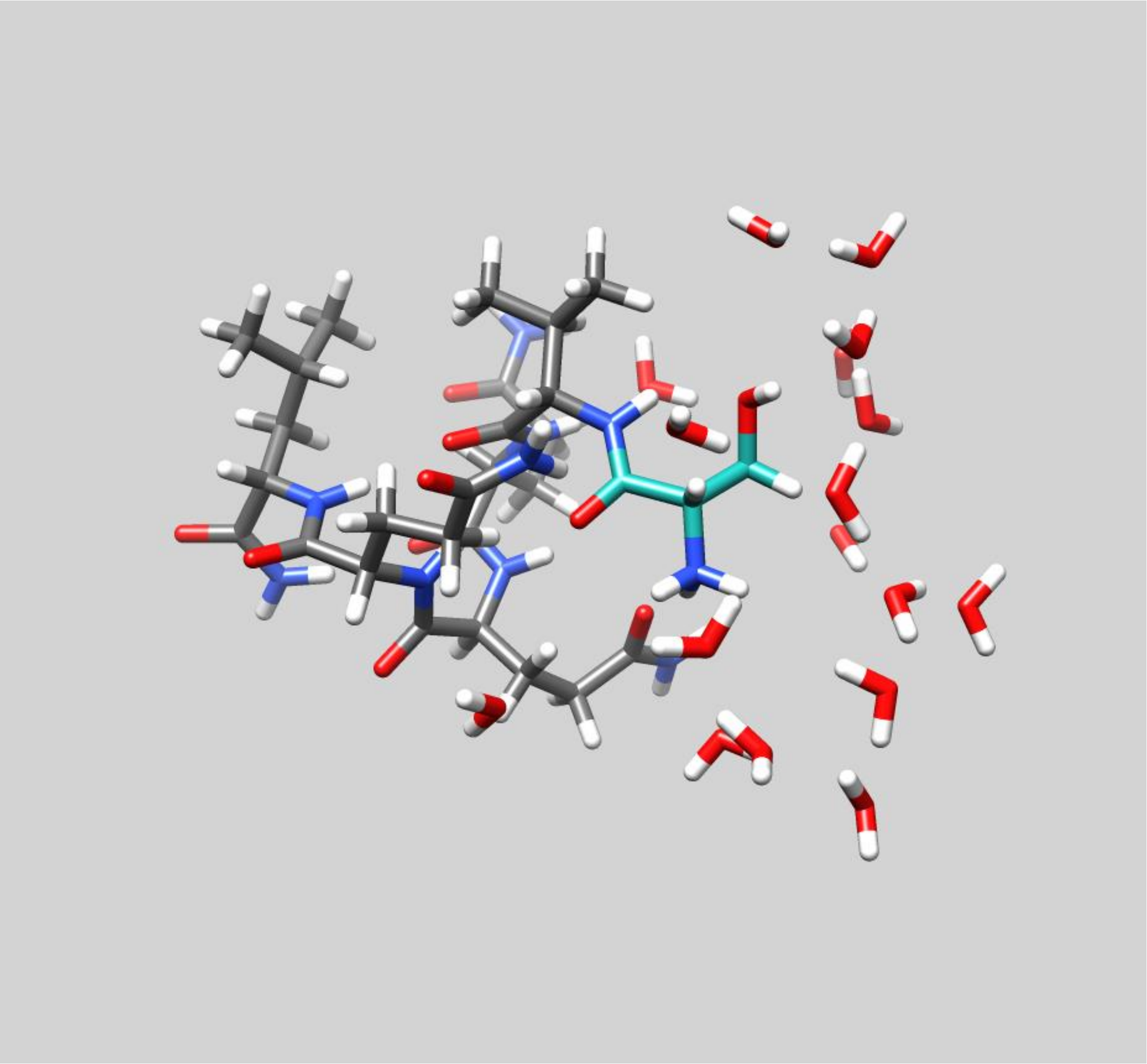


Figure A.4 Refinement Cluster of residues 28

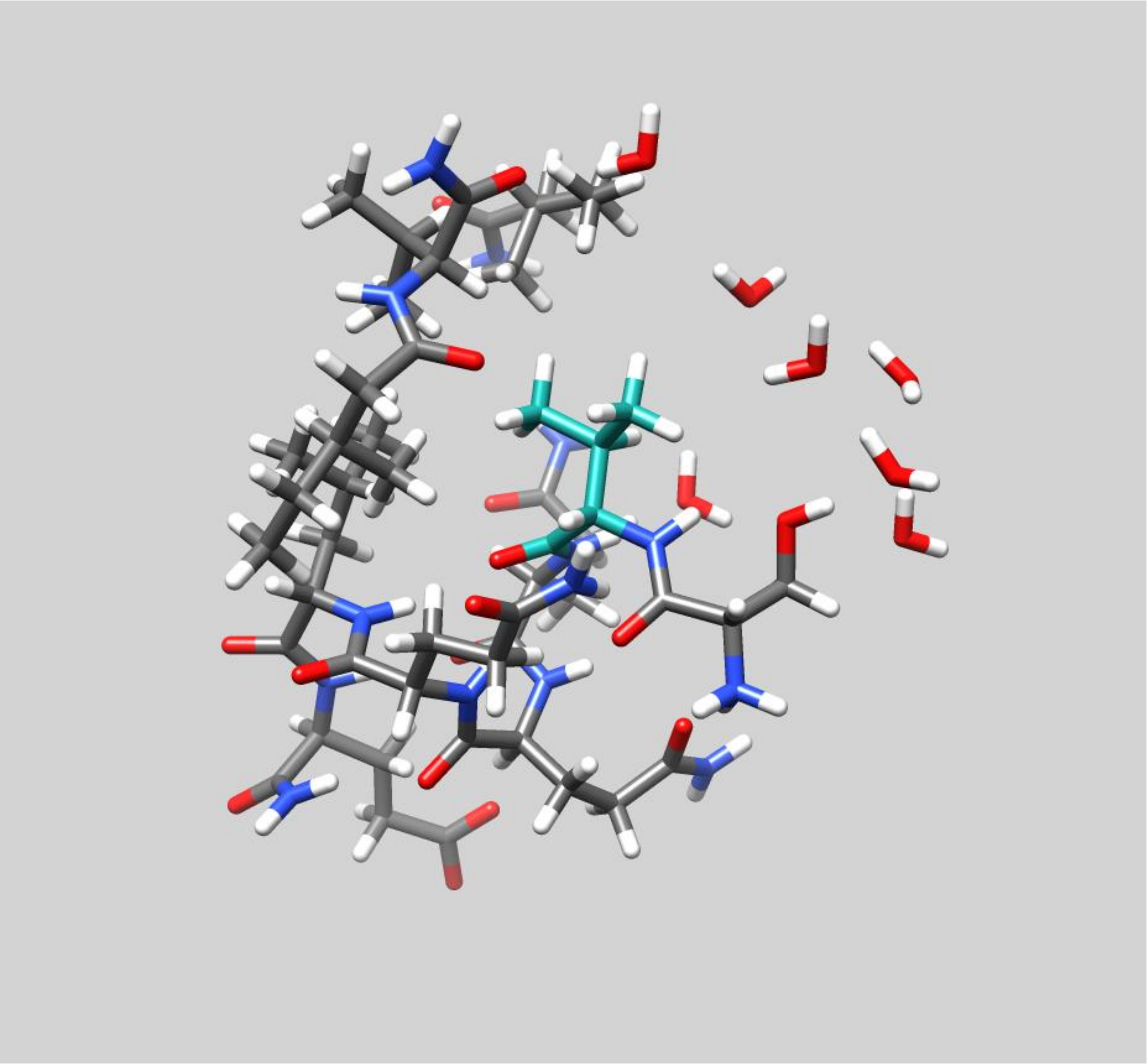


Figure A.5 Refinement Cluster of residues 29