Intermolecular and Intramolecular Stable Isotope Studies in Alanine

Thesis by
Brooke H. Dallas

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy in Geology

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2024
Defended May 8, 2024

© 2024

Brooke H. Dallas
ORCID: 0000-0002-1313-3270

All rights reserved

ACKNOWLEDGEMENTS

I would like to thank my advisor, John Eiler, for giving me the opportunity to participate in experimental research at the frontier of stable isotope geochemistry, and allowing me the freedom I needed to explore and orient myself towards the scientific questions I find most compelling. John, thank you for bringing me into the program and seeing me through it!

Several other GPS professors have also played significant roles in the effort that went into this body of work, particularly my thesis committee members: Alex Sessions, whose weekly group meetings I attended for most of my time at Caltech, and who has always been happy to offer valuable research suggestions and approaches for tackling a long series of experimental challenges; Jared Leadbetter, from whom I thoroughly enjoyed learning the concepts, tools and experimental methods unique to microbiology, and with whom I find shared taste not only in scientific problems but in other creative expressions (e.g., photography) as well; and Geoff Blake, who graciously offered to be on my committee, and whom I regret not having visited more often, as I imagine we had much more to discuss than I realized at the time. Joe Kirschvink and the late Jason Saleeby helped me stay true to my roots in geology by providing frequent excuses and means for me to return to the field and visit a wealth of memorably remote and exquisite places that I otherwise likely never would have had the opportunity to experience. In particular, the quarterly Ge 136 trips and traditions therein make GPS a very special department with a unique culture. May they continue on in similar fashion indefinitely. George Rossman has always been excited to share not only his mineral collection and laboratory spaces, but also many enthralling tales of mineralogical adventures around the world.

Fenfang Wu, Nami Kitchen, and Nathan Dalleska supported me in many long days of experimental troubleshooting, often expending unreasonably large amounts of their time toward helping solve my problems. My collaboration with Dave Dixon and Eric Bylaska took the theoretical work of Chapter 2 further than I could have done alone, and Jim Rustad generously offered his valuable expertise in molecular modeling and electronic structure calculations whenever I needed it. In the last years of the experimental work that went into this thesis I discovered the CCE NMR lab and its resourceful director Dave

Vander Velde, whose insights led to a novel NMR application and fruitful collaboration resulting in the hydrogen isotope work presented herein.

I shared my first-year office (The Pit) and many days in the field memorably with Mathieu LaPotre, Max Lloyd, Hayden Miller, Daven Quinn, Lewis Ward, and Jennifer Buz. Discussions with Amy Hoffman, Elise Wilkes, Natalia Solomatova, Brigitte Rooney, Chanel Valiente, Frank Sousa, Cody Finke, Elle Chimiak, Hao Xie, Camilo Ponton, Matthieu Clog, Chris Lamartina, Aditi Chatterjee, Reto Wijker, Daniel Stolper, Alison Piasecki, Maggie Osburn, Ted Present, Zach Erickson, my officemate Paul Magyar, and others whom I will undoubtedly kick myself later for forgetting, enriched my research and my time at Caltech. I particularly appreciated the many conversations I shared with John Magyar, who was my officemate after Paul graduated, and who spent significant time thinking about my research problems and frequently generating ideas and offering solutions that I would not have discovered on my own.

I would like to thank the U.S. Department of Energy for funding this research through the Office of Basic Energy Sciences, Chemical Sciences Division, Award Number DE-SC0016561.

I must also express my deep gratitude to my parents, Peter and Lara, and my sister, Polly, for unconditional love and support; and to those friends who have been my greatest sources of inspiration over the years: Glenn Hair, Victor Lo Forte, Brandon Stone, Tim Fielden-Hosking, Nick Andrzejkiewicz, Stephen Pincus, Greg Pelaez, Kevin Hughes, Alex McCollumn, David Witkowski, Satomi Mitsuya, Mohammad Nazrul, Vic Spahn, Martin Ball, Zev Weinstein, Naila Weinstein, and Pia Malaney. And, finally, to Eric R. Weinstein for always being in my corner, for Geometric Unity (I've got a good feeling about this), and for Truth, Meaning, Fitness, and Grace. Çok teşekkür ederim, en iyisi sensin. Fiziği yeniden canlandıralım!

ABSTRACT

In Chapter 1 of this thesis, we give an introduction to this body of work, providing some background for context.

In Chapter 2, we present a set of theoretical predictions for the carbon isotope distribution between equilibrated carbon sites of alanine and pyruvate. We start with the simplest possible theoretical treatment, and work progressively through higher levels of theory, showing consistency in the direction and magnitude of expected fractionation across these treatments.

In Chapter 3, we present our experimental work to confirm the predictions made in Chapter 2 by measuring the δ^{13}C of the α carbon site in alanine that has undergone equilibration with the analogous carbon site in pyruvate via the alanine transaminase enzyme (ALT).

In Chapter 4, we describe the process that led to our (re)discovery of β-hydrogen-deuterium exchange in amino acids catalyzed by transaminases. We then provide a literature review on the small body of historical work on this system, which took place primarily during the 1960s and 70s. This literature summary provides the background necessary for the reader to appreciate our experimental work presented in the next chapter.

In Chapter 5, we present novel ^{1}H NMR and ^{13}C NMR experimental observations of intermolecular hydrogen isotope exchange between water and the α and β carbon sites of alanine, as well as intramolecular hydrogen isotope exchange between the α and β carbons, all of which is catalyzed by alanine transaminase (ALT). These experiments track the abundances of eight isotopically distinct alanine species varying in their position and/or number of hydrogen isotopes over a series of reactions differing in initial alanine isotopic composition and initial water isotopic composition. With the data collected we are able to determine up to thirteen rate constants and ten equilibrium constants describing the transfer of hydrogen and deuterium amongst these eight isotopic variants and water, as well as the thermodynamic equilibrium constants between them.

PUBLISHED CONTENT AND CONTRIBUTIONS

McNeill, Ashley S, Brooke H Dallas, John M Eiler, Eric J. Bylaska, and David A. Dixon (2020). "Reaction Energetics and ^{13}C Fractionation of Alanine Transamination in the Aqueous and Gas Phases". In: *The Journal of Physical Chemistry A* 124.10, pp. 2077–2089. DOI: 10.1021/acs.jpca.9b11783. B.D. participated in the conception of the project and carried out computational work therein.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

AN INTRODUCTION

> *"It often seems that isotopic fractionations provide too much information about too many processes, combining it all in a package that is unmanageably intricate. In response, investigators keep increasing the complexity of the available data by providing more and more detailed analyses. The proliferation of compound-specific isotopic analyses is a prime example of this phenomenon. Does it increase the information-carrying capacity of the isotopic channel or is it another case of the triumph of entropy? To obtain the preferred result, we will have to understand biosynthetic fractionations... ."*
> -John Hayes, 2001

Over the past decade, investigations in the most fundamental aspects of stable isotope geochemistry have centered around the growing field of "isotomics", which involves several intersecting topics: (1) isotopic clumping, that is, the occurrence of multiple rare, usually adjacent, heavy isotopes in a molecule (e.g., $^{18}O-^{13}C$, $^{2}H-^{2}H$, $^{13}C-^{13}C$, $^{2}H-^{13}C$, $^{15}N-^{15}N$), (2) isotopic equilibration of atoms at specific positions in a molecule, either intermolecularly or intramolecularly, and (3) kinetic fractionation of isotopes between molecular reactants and products under conditions of disequilibrium, and in particular, the coupling of isotopic and chemical equilibrium (e.g., Druhan et al., 2013; Steefel et al., 2014). Most of these efforts have involved small, simple inorganic molecules (e.g., CO_2 (Zhang et al., 1995), and N_2O (Magyar et al., 2016)) and organic molecules, most extensively the lightest alkanes: methane (Stolper et al., 2014a; Stolper et al., 2014b), ethane (Clog et al., 2017), and propane (Piasecki et al., 2016a; Piasecki et al., 2016b; Piasecki et al., 2018; Xie et al., 2018). Such studies have typically been limited to a subset of these topics, primarily due to a combination of technical limitations and the restricted isotopic diversity inherent to these small molecules. The fundamental physical principles governing "mass-dependent" isotopic distributions in molecules are based on the analysis of quantum mechanical oscillators, and have been appreciated since quantum mechanical principles were first applied towards the under-

standing of isotope distributions in natural systems (Urey, 1947; Bigeleisen et al., 1947). Computational predictions of carbon and hydrogen equilibrium isotopic signatures for such molecules (e.g., Richet et al., 1977, Schauble, 2004) have provided benchmarks against which to compare experimental results and have thus encouraged the development of increasingly sensitive analytical tools and methods using an array of new high-mass-resolution mass spectrometers, notably the Thermo Scientific MAT 253 Ultra, Nu Instruments Panorama, Thermo DFS, Q Exactive Orbitrap and related instruments. Improvements in infrared spectroscopic (e.g., Ono et al., 2014) and nuclear magnetic resonance instrumental capabilities (e.g., Liu et al., 2018) have also supported this effort.

Application of the methods developed with these instruments to this small set of molecules was initially focused primarily on probing molecular processes relevant to the natural gas industry and a handful of environmental settings, using isotopes to tease apart molecular sources, sinks, and thermal history, involving both equilibrium and kinetic fractionation processes. For example, as the smallest and simplest alkane, methane (CH_4) on its own has recently yielded an impressively rich set of accessible and useful isotopic information. Considering only the stable isotopes of carbon (^{12}C and ^{13}C) and hydrogen (^{1}H and ^{2}H—sometimes denoted "D"), methane can exist in ten unique isotopic forms, all of which are isotopologues, i.e., molecules that differ only in their isotopic composition. In natural terrestrial contexts these ten isotopologues range in proportional relative abundance from 0.988 for $^{12}CH_4$, to $6.54 * 10^{-18}$ for the rarest of these, $^{13}CD_4$. In conventional mass spectrometric measurements, these ten isotopologues were aggregated through conversion to CO_2 or H_2, erasing the individual contributions of each isotopic species to the overall molecular ratios of $^{13}C/^{12}C$ and D/H, in essence reducing ten distinct quantities down to two. Only recently the ability to measure clumping of $^{13}C-D$ and $D-D$ in methane using, e.g., the Thermo Scientific MAT 253 Ultra and Nu Instruments Panorama, has given us access to five of the ten isotopologues: not only the three most common of the unsubstituted and singly substituted forms ($^{12}CH_4$, $^{13}CH_4$, and $^{12}CH_3D$), but also the two doubly substituted forms, $^{13}CH_3D$, and $^{12}CH_2D_2$. In general, these forms are reported relative to an expected stochastic occurrence based on the known natural abundances of each nuclei, and deviations from these probabilistic quantities are indicative of conditions and processes involved during formation and destruction of the methane molecule. Quantification of these two rarer isotopic species in addition to the more abun-

dant three has proven powerful as a means of differentiating between multiple microbial and abiotic methane sources, as well as the temperatures at which these gases formed (Stolper et al., 2014a; Stolper et al., 2014b). The other five of these isotopologues are so rare that they remain beyond our ability to quantify in natural samples with present-day state-of-the-art instruments.

When considering site-specificity, we see that methane lacks a diversity of distinguishable atomic sites. It is composed of a single carbon atom and four tetrahedrally symmetrical, and therefore indistinguishable, hydrogen atoms. Thus in the case of methane, intramolecular isotopic site preferences are not at play, and yet access to two "clumped" isotope species provides significant novel forensic geochemical leverage. One step up in complexity, ethane (H_3C-CH_3) can exist in 2^8 possible isotopic combinations using the same set of stable carbon and hydrogen isotopes. However, because of its structural symmetry, many of the 2^8 possible isotopic forms are degenerate; that is, there are multiple ways to form isotopically indistinguishable species. Ethane's two carbon atoms are structurally indistinguishable, as are the three hydrogen atoms bound to each, assuming they are not sterically hindered and can rotate freely around the carbon-carbon bond axis. Thus the number of distinguishable, unique species is reduced from 256 to only 36 (see Table 5.7). Nevertheless, ethane's two carbon atoms introduce a site preference component in certain cases, as we will show below.

The probability of forming a given isotopic species (based on the relative abundances of each isotope in the available pools) must be multiplied by the number of ways of forming each species when ethane's symmetries are taken into account. For example, there are three possible ways to make $H_2D^{12}C-{}^{12}CH_3$, all three of which are identical to the three possible ways to make $H_3{}^{12}C-{}^{12}CH_2D$, resulting in a sixfold redundancy. Likewise, $H_3{}^{13}C-{}^{12}CH_3$ is identical to $H_3{}^{12}C-{}^{13}CH_3$, and so on. But some of the possible isotopic species are not redundant in this way, and in these cases, site specificity can apply. In an ethane molecule that is already isotopically asymmetric, e.g., $H_2D^{12}C-{}^{12}CH_3$, swapping in an additional heavy isotope can occur either at one carbon site or the other, making one of four distinct forms, either $HD_2{}^{12}C-{}^{12}CH_3$ or $H_2D^{12}C-{}^{12}CH_2D$ if a D is swapped in for an H, and either $H_2D^{13}C-{}^{12}CH_3$ or $H_2D^{12}C-{}^{13}CH_3$ if a ^{13}C is swapped in for a ^{12}C. It is in this sense, requiring substitution of two or more heavy isotopes, that a kind of site specificity applies

to ethane. This could also be considered a "clumping" of isotopes of ethane, as each addition of a heavy isotope will be "clumped" one, two or three bonds away from any that are preexisting in the molecule. A single study by Clog et al. looks at the occurrence of $^{13}C-^{13}C$ "clumping" in ethane (Clog et al., 2017), relative to an expectation based on natural abundances and temperature, concluding that most of the observed natural variation in the occurrence of this clumped isotopologue reflects kinetic fractionation processes in natural gases of various thermal maturities.

One further step up in complexity, propane is the smallest alkane with multiple structurally distinct atomic sites, having a central carbon with two bound hydrogen atoms, and two identical "terminal" carbon atoms, each bound to the central carbon and to three hydrogen atoms. It is here, still with only two atomic elements, and the same four stable isotopes involved, that we can begin to appreciate the broader range of isotopic forms (216 unique to propane) that are possible in molecules with multiple structurally distinct sites. Many of propane's unique isotopic forms are isotopologues, while some are isotopomers, i.e., structurally identical molecules that have the same number of atoms of each isotope, but in which the isotopes are arranged in different positions. The conventional nomenclature itself dramatically fails to capture the vast array of possible isotopic arrangements, even in a molecule this small. In propane there may be useful geochemical forensic information not only in $^{13}C-^{13}C$, $^{13}C-D$ or D-D clumping, as well as the site-preferences of both ^{13}C and D between the terminal and central positions, but also in the combination of site preference and multiple heavy isotope substitution. For example, starting with the relatively simple context of thermodynamically equilibrated hydrogen isotopes, in a propane molecule with one deuterium atom, does that deuterium atom prefer to be bound to the central carbon atom, or to the terminal carbon atom, assuming the carbon atoms are of either all ^{12}C or all ^{13}C? This question was recently addressed both theoretically and experimentally, in Piasecki et al., 2018 and Xie et al., 2018, with the experimental work suggesting a preference for deuterium at the central propane position of $\sim$80 ‰ at 30 C, broadly consistent with computational predictions. With this information one is bound to wonder how this site preference might shift with each additional deuterium atom on the propane molecule. Might there be a pattern predicting which site they would prefer? Is this pattern measurably altered in propane molecules with mixtures of ^{12}C and ^{13}C? For decades, these questions were generally out

of reach to experimentalists. The dream of isotomics is to quantitatively predict and observe the isotopic "anatomies", or "isotomes" of organic molecules, for with access to such information we would surely have much tighter constraints on the relative contributions of a vast array of chemical and physical processes to the molecules we can sample in nature.

Despite the potential useful information buried in isotomic complexity, there so far exist only a small number of studies reporting variations in even a handful of the numerous possible isotopic species present in organic molecules (e.g., Stolper et al., 2014a, Magyar et al., 2016, Neubauer et al., 2018 and Csernica et al., 2023), and often in these cases the measurements combine contributions from multiple isotopic variants. This is in large part due to the technical challenges of measuring extremely rare isotopic forms in natural abundance contexts. One strategy to deal with this challenge is to use isotopic labeling, that is, artificially "doping" or "spiking" molecules with heavy isotopes significantly (sometimes orders of magnitude) above their natural abundance, usually during stages of methods development and proof of concept experiments. In such studies the normally rare forms are present at higher abundance and are thus easier to measure quantitatively and to higher precision. Still, isotopic labeling has its limitations, particularly beyond early methods development and in the realm of application of these methods to natural samples. Because of this, the isotope geochemistry community tends to consider isotopic labeling experiments a means to an end, employing them only as necessary on the way to applying the techniques to natural samples, where their primary scientific problems lie. There is also always a significant risk of contaminating natural samples or laboratory instruments and equipment with unwanted and hopelessly difficult-to-eradicate heavy isotopes, making such labeling experiments come with an undesirably high potential cost. This is especially true in laboratories relying heavily on mass spectrometry for isotopic discrimination, where measurement involves sample destruction, and thus direct contact between the instrumentation and the molecule under observation. In contrast, nondestructive analyses such as nuclear magnetic resonance, in which sample never comes into contact with the instrument itself, are less vulnerable to this hazard, and routinely use deuterated solvents as part of everyday measurements.

As instruments and techniques emerge that allow us to examine molecules of greater molecular weight and complexity, the number of isotopic variants

(isotopologues and isotopomers) whose relative abundances we can measure and compare increases combinatorically. For a variety of reasons, one class of molecules holds particular appeal for study at the current stage: the amino acids, a set of small organic molecules that each contain an amine and a carboxyl group. The smallest of these (and the only non-chiral amino acid) is glycine, which has just two carbons and the simplest possible R group, a hydrogen atom. Alanine is one carbon atom more complex than glycine, with an R group of CH_3, making it the most simple chiral amino acid. Glycine and alanine belong to a small set of about twenty amino acids that are genetically encoded and make up the proteins in all known life. All of these belong in the category of α-amino acids, i.e., the carboxyl and amine groups are both bound to the α-carbon. Several hundred additional non-proteinogenic amino acids exist in relatively low abundance, primarily outside of earth, some of which have been studied on carbonaceous chondrites. Many of these amino acids are β-, γ-, δ- and ε-amino acids, a feature that distinguishes them from the abundant terrestrial α-amino acids. Alanine was initially chosen for this work in large part because of its simplicity relative to other amino acids, but also because of its central roles in metabolism and relatively high abundance both outside of earth and within terrestrial biology.

Alanine, like propane, is made of three carbon atoms, and both contain at least one methyl group, but the similarities between these molecules end there, and new challenges of analysis begin. Unlike propane, which has two indistinguishable terminal methyl groups, each carbon site of alanine is unique. The molecule includes two oxygen atoms and a nitrogen atom, making it inherently isotopically and structurally more complex than the alkanes, and in turn much trickier to handle. Alkanes lack hydrophilic and reactive groups, and can be manipulated in the gas phase and separated from each other and from contaminants using common laboratory practices involving a Schlenk line and cryogenic trapping, lending themselves cooperatively to gas-source mass spectrometry. In contrast, alanine, and amino acids more generally, contain reactive functional groups, exist as solids at room temperature, and decompose when heated, rather than evaporating. Studying amino acid isotopic structures by mass spectrometry requires laboratory equipment and experimental protocol adaptations that take into account the chemical differences between well-behaved hydrocarbons and tricky amino acids. Generally this involves some kind of derivatization, i.e., addition of functional groups to the amino

acid reactive sites that result in a new, larger molecule with comparatively increased volatility and reduced reactivity, but which still retains the carbon backbone and some functional groups from the original amino acid. In our lab, the introduction of such derivatized amino acid sample material occurs via a prototype gas inlet system, which is contained inside a modified oven designed for gas chromatography. This system uses helium as a carrier gas, which feeds into an electron impact ionization source where the volatilized sample is ionized.

One example of a fundamental proof-of-concept experiment in isotopic studies applied to several n-alkanes (those discussed above) and other small molecules has been to equilibrate the intramolecular isotopic structure at a fixed temperature, either internally (as in the case of N_2O (Magyar et al., 2016), CO_2 (Eiler et al., 2004; Eiler, 2011), and methane (Stolper et al., 2014a; Xie et al., 2018)), with carbon and oxygen in the $CO_2(aq)-HCO_3^-(aq)-CO_3^{2-}(aq)$ system (Zhang et al., 1995; Uchikawa et al., 2021), or with hydrogen/deuterium atoms in H_2O (Xie et al., 2018) or H_2 (Horibe et al., 1995), using fairly straightforward procedures involving catalysts in gas phase. In contrast, alanine requires dissolution in water in the presence of an enzyme in order to isotopically equilibrate any of its carbon or hydrogen sites. Because of this, gas phase experiments on alanine are not possible; instead, any isotopic manipulation of alanine must take place in an aqueous solution, and analysis on gas-source mass spectrometers requires derivatization of alanine's charged groups, meaning that additional atoms must be taken into account for in any mass spectrometric measurement. More recently, measurements using electrospray ionization coupled to Orbitrap-based mass spectrometry have successfully analyzed carbon isotope distributions in alanine without derivatization, with the limitation that the C-2 (α) and C-3 (β) sites are indistinguishable due to lack of fragmentation across the bond between these two carbon atoms (Weiss et al., 2023).

In spite of these challenges, isotopic equilibration of alanine with another molecule is an appealing proof of concept experiment, and an important step to take prior to trying to interpret the isotopic signatures of samples gathered in nature. The answer can be predicted from first principles using various degrees of sophistication, providing theoretical bases against which experimentalists can evaluate results. In the field of stable isotope geochemistry,

opportunities to test specific hypotheses against rigorous theoretical predictions are not abundant in aqueous systems, in part because of the complications introduced by solvation, but also due to the challenge of finding systems that are conducive to investigation with the available tools, and where equilibrium can be unambiguously established (possibly with the aid of enzymes, such as carbonic anhydrase, e.g., Uchikawa et al., 2021). For example, it is only relatively recently that ^{13}C-^{12}C fractionations were available in the $CO_2(aq)-HCO_3^-(aq)-CO_3^{2-}(aq)$ system (Zhang et al., 1995), and theoretical calculations that accurately recovered observed fractionations have only existed within the last two decades (Rustad et al., 2008). The study of site-specific isotope effects associated with amino acid-keto acid transamination reactions is similarly fundamental.

The ALT enzyme catalyzes the transamination reaction below (written stoichiometrically in Equation 1.1 and illustrated in Figure 3.1):

$$\begin{aligned} &HOOC-CO-(CH_2)_2-COOH + CH_3-CHNH_2-COOH \rightleftharpoons \\ &CH_3-CO-COOH + HOOC-CH(NH_2)-(CH_2)_2-COOH \end{aligned} \tag{1.1}$$

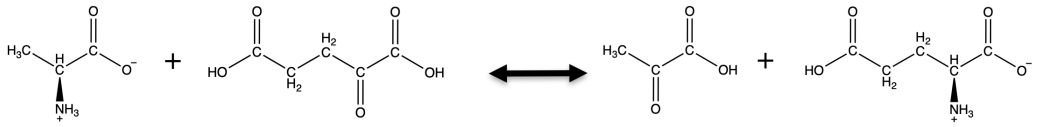

Figure 1.1: The alanine-pyruvate transamination reaction, catalyzed by the alanine transaminase enzyme (ALT), EC number 2.6.1.2.

This transamination reaction was first discovered by Braunstein and Kritzman in the mid-1930s (Braunstein et al., 1937), among the earliest characterizations of biological transformations of amino acids to their corresponding keto acids. The equilibrium constant and kinetics of this reaction were first examined by Lénárd and Straub in the early 1940s (Lénárd et al., 1942). The alanine transamination reaction, mediated by ALT, serves a fundamental and primary central metabolic function required for all life as we know it, carrying out both amino acid synthesis for protein construction, as well as amino acid catabolism for gluconeogenesis, depending on the fluctuating needs of the organism. ALT is ubiquitous across all branches of life, existing in many slightly different

evolved forms known as isoenzymes (e.g., DeRosa et al., 1975). The sequences and structures of three of these isoenzymes (one of several Hordeum vulgare (barley) isoenzymes, one of Escherichia coli K-12, and one of two Homo sapiens isoenzymes) are available on the RCSB Protein Data Bank. In mammals, the enzymatic conversion is key to the Cahill cycle, a sequence of reactions mediating transfer of amino groups and carbon between skeletal muscles and the liver (Felig, 1973). Catabolic degradation of amino acids yields nitrogen that is transaminated from alanine to form pyruvate. The resulting alanine is directed to the liver where the nitrogen enters the urea cycle and the pyruvate is used to produce glucose. The activity of human serum ALT is routinely assayed in standard blood panels for clinical diagnosis of human liver injury and disease, and its role in cancer metabolism is an active area of research (Albers et al., 2008).

For the purpose of stable isotope biogeochemical investigation in small organic molecules, ALT is a uniquely well suited catalyst. While the original intent of this study was to use ALT as a means to isotopically perturb the carbon backbone of alanine by reacting it to form pyruvate and measuring the isotopic offset between product and reactant, in the process of working through complications in that set of experiments, we (re)discovered a useful secondary feature of this enzyme that revealed an opportunity to study hydrogen-deuterium exchange kinetics and equilibrium site preferences for four hydrogen atoms bound to two unique, adjacent carbon atoms. This work yielded unexpectedly fruitful isotopic data, which we present here in addition to our carbon isotope study.

The work of this thesis is presented as follows:

In Chapter 1 (this chapter), we give an introduction to our work, providing some background for context and summarizing each chapter.

In Chapter 2, we present a set of theoretical predictions for the carbon isotope distribution between equilibrated carbon sites of alanine and pyruvate. We start with the simplest possible theoretical treatment, and work progressively through higher levels of theory, showing consistency in the direction and magnitude of expected fractionation across these treatments.

In Chapter 3, we present our experimental work to confirm the predictions made in Chapter 2 by measuring the δ^{13}C of the α carbon site in alanine that has undergone equilibration with the analogous carbon site in pyruvate via

the alanine transaminase enzyme (ALT).

In Chapter 4, we describe the process that led to our (re)discovery of hydrogen deuterium exchange in amino acids catalyzed by transaminases. We then provide a literature review on the small body of historical work on this system, which took place primarily during the 1960s and 70s. This literature summary provides the background necessary for the reader to appreciate our experimental work presented in the next chapter.

In Chapter 5, we present novel ^{1}H NMR and ^{13}C NMR experimental observations of intermolecular hydrogen isotope exchange between water and the α and β carbon sites of alanine, as well as intramolecular hydrogen isotope exchange between the α and β carbons, all of which is catalyzed by alanine transaminase (ALT). These experiments track the abundances of eight isotopically distinct alanine species varying in their position and/or number of hydrogen isotopes over a series of reactions differing in initial alanine isotopic composition and initial water isotopic composition. With the data collected we are able to determine up to thirteen rate constants describing the transfer of hydrogen and deuterium amongst these eight isotopic variants and water, as well as the thermodynamic equilibrium constants between them.

References

Albers, Mark J et al. (2008). "Hyperpolarized ^{13}C lactate, pyruvate, and alanine: noninvasive biomarkers for prostate cancer detection and grading." In: *Cancer Research* 68.20, pp. 8607–8615.

Bigeleisen, Jacob and Maria Goeppert Mayer (1947). "Calculation of Equilibrium Constants for Isotopic Exchange Reactions". In: *The Journal of Chemical Physics* 15.5, pp. 261–267.

Braunstein, A E and M G Kritzmann (1937). "Formation and Breakdown of Amino-acids by Inter-molecular Transfer of the Amino Group". In: *Nature Publishing Group* 140.3542, pp. 503–504.

Clog, Matthieu, Michael Lawson, Brian Peterson, Alexandre A Ferreira, Eugenio V Santos Neto, and John M Eiler (2017). "A reconnaissance study of ^{13}C–^{13}C clumping in ethane from natural gas". In: *Geochimica et Cosmochimica Acta*, pp. 1–53.

Csernica, Timothy, Alex L. Sessions, and John M. Eiler (2023). "High-dimensional isotomics, part 2: Observations of over 100 constraints on methionine's isotome". In: *Chemical Geology* 642, p. 121771. ISSN: 0009-2541. DOI: https://doi.org/10.1016/j.chemgeo.2023.121771. URL: https://www.sciencedirect.com/science/article/pii/S0009254123004722.

DeRosa, G and Robert W Swick (1975). "Metabolic implications of the distribution of the alanine aminotransferase isoenzymes." In: *Journal of Biological Chemistry* 250.20, pp. 7961–7967.

Druhan, Jennifer L., Carl I. Steefel, Kenneth H. Williams, and Donald J. DePaolo (2013). "Calcium isotope fractionation in groundwater: Molecular scale processes influencing field scale behavior". In: *Geochimica et Cosmochimica Acta* 119, pp. 93–116. ISSN: 0016-7037. DOI: https://doi.org/10.1016/j.gca.2013.05.022. URL: https://www.sciencedirect.com/science/article/pii/S0016703713003098.

Eiler, John M (2011). "Paleoclimate reconstruction using carbonate clumped isotope thermometry". In: *Quaternary Science Reviews* 30.25-26, pp. 3575–3588.

Eiler, John M and Edwin Schauble (2004). "18O13C16O in Earth's atmosphere". In: *Geochimica et Cosmochimica Acta* 68.23, pp. 4767–4777. ISSN: 0016-7037. DOI: 10.1016/j.gca.2004.05.035.

Felig, Philip (1973). "The glucose-alanine cycle". In: *Metabolism* 22.2, pp. 179–207.

Hayes, John M. (2001). "Fractionation of carbon and hydrogen isotopes in biosynthetic processes". In: *Stable Isotope Geochemistry* 43, pp. 225–277.

Horibe, Y and H Craig (1995). "DH fractionation in the system methane-hydrogen-water". In: *Geochimica et Cosmochimica Acta* 59.24, pp. 5209–5217.

Lénárd, P and F B Straub (1942). *Aminopherase*. Studies Inst. Med. Chem. Univ. Szeged.

Liu, C, G P McGovern, P Liu, H Zhao, J Horita Chemical Geology, and 2018 (2018). "Position-specific carbon and hydrogen isotopic compositions of propane from natural gases with quantitative NMR". In: *Elsevier* 491, pp. 14–26. DOI: 10.1016/j.chemgeo.2018.05.011. URL: https://www.sciencedirect.com/science/article/pii/S0009254118302316.

Magyar, Paul M, Victoria J Orphan, and John M Eiler (2016). "Measurement of rare isotopologues of nitrous oxide by high-resolution multi-collector mass spectrometry". In: *Rapid Communications in Mass Spectrometry* 30.17, pp. 1923–1940.

Neubauer, Cajetan, Michael J Sweredoski, Annie Moradian, Dianne K Newman, Richard J Robins, and John M Eiler (2018). "Scanning the isotopic structure of molecules by tandem mass spectrometry". In: *International Journal of Mass Spectrometry* 434, pp. 276–286. DOI: 10.1016/j.ijms.2018.08.001.

Ono, Shuhei, David T Wang, Danielle S Gruen, Barbara Sherwood Lollar, Mark S Zahniser, Barry J McManus, and David D Nelson (2014). "Measurement of a Doubly Substituted Methane Isotopologue, 13CH 3D, by Tunable Infrared Laser Direct Absorption Spectroscopy". In: *Analytical Chemistry* 86.13, pp. 6487–6494. ISSN: 0003-2700. DOI: 10.1021/ac5010579. URL: http://pubs.acs.org/doi/abs/10.1021/ac5010579.

Piasecki, Alison, Alex Sessions, Michael Lawson, A A Ferreira, E V Santos Neto, and John M Eiler (2016a). "Analysis of the site-specific carbon isotope composition of propane by gas source isotope ratio mass spectrometer". In: *Geochimica et Cosmochimica Acta* 188.C, pp. 58–72.

Piasecki, Alison, Alex Sessions, Michael Lawson, A A Ferreira, E V Santos Neto, Geoffrey S Ellis, Michael D Lewan, and John M Eiler (2018). "Position-specific ^{13}C distributions within propane from experiments and natural gas samples". In: *Geochimica et Cosmochimica Acta* 220, pp. 110–124.

Piasecki, Alison, Alex Sessions, Brian Peterson, and John M Eiler (2016b). "Prediction of equilibrium distributions of isotopologues for methane, ethane and propane using density functional theory". In: *Geochimica et Cosmochimica Acta* 190.C, pp. 1–12.

Richet, P, Y Bottinga, and M Javoy (1977). "A review of hydrogen, carbon, nitrogen, oxygen, sulphur, and chlorine stable isotope fractionation among gaseous molecules". In: pp. 1–47.

Rustad, James R, Sierra L Nelmes, Virgil E. Jackson, and David A. Dixon (2008). "Quantum-Chemical Calculations of Carbon-Isotope Fractionation in CO 2(g), Aqueous Carbonate Species, and Carbonate Minerals". In: *The Journal of Physical Chemistry A* 112.3, pp. 542–555.

Schauble, Edwin A (2004). "Applying stable isotope fractionation theory to new systems". In: *Geochemistry of Non-Traditional Stable Isotopes*. Ed. by C. M. Johnson, B. L. Beard, and F. Albarade. Vol. 55. Reviews in Mineralogy and Geochemistry, pp. 65–111.

Steefel, Carl I., Jennifer L. Druhan, and Kate Maher (2014). "Modeling Coupled Chemical and Isotopic Equilibration Rates". In: *Procedia Earth and Planetary Science* 10, pp. 208–217. ISSN: 1878-5220. DOI: 10.1016/j.proeps.2014.08.022.

Stolper, Daniel A., A. L. Sessions, A. A. Ferreira, E. V. Santos Neto, A. Schimmelmann, S. S. Shusta, D. L. Valentine, and John M. Eiler (2014a). "Combined ^{13}C–D and D-D clumping in methane: Methods and preliminary results". In: *Geochimica et Cosmochimica Acta* 126.C, pp. 169–191.

Stolper, Daniel A. et al. (2014b). "Formation temperatures of thermogenic and biogenic methane". In: *Science* 344.6191, pp. 1500–1503.

Uchikawa, Joji, Sang Chen, John M. Eiler, Jess F. Adkins, and Richard E. Zeebe (2021). "Trajectory and timescale of oxygen and clumped isotope equilibration in the dissolved carbonate system under normal and enzymatically-catalyzed conditions at 25°C". In: *Geochimica et Cosmochimica Acta* 314, pp. 313–333. ISSN: 0016-7037. DOI: 10.1016/j.gca.2021.08.014.

Urey, Harold C (1947). "The thermodynamic properties of isotopic substances". In: *Journal of the Chemical Society (Resumed)*, pp. 562–581.

Weiss, Gabriella M., Alex L. Sessions, Maxime Julien, Timothy Csernica, Keita Yamada, Alexis Gilbert, Katherine H. Freeman, and John M. Eiler (2023). "Analysis of intramolecular carbon isotope distributions in alanine by electrospray ionization Orbitrap mass spectrometry". In: *International Journal of Mass Spectrometry* 493, p. 117128. ISSN: 1387-3806. DOI: 10.1016/j.ijms.2023.117128.

Xie, Hao, Camilo Ponton, Michael J Formolo, Michael Lawson, Brian K Peterson, Max K Lloyd, Alex L Sessions, and John M Eiler (2018). "Position-specific hydrogen isotope equilibrium in propane". In: *Geochimica et Cosmochimica Acta* 238, pp. 193–207.

Zhang, J, PD Quay, and DO Wilbur (1995). "Carbon isotope fractionation during gas-water exchange and dissolution of CO_2". In: *Geochimica et Cosmochimica Acta* 59.1, pp. 107–114.

Chapter 2

THEORETICAL UNDERPINNINGS: ESTIMATION OF THE INTERMOLECULAR ^{13}C-^{12}C FRACTIONATION BETWEEN ALANINE AND PYRUVATE AT THE α-CARBON AND THE DISTRIBUTION OF ^{1}H AND ^{2}H IN ALANINE EFFECTED THROUGH THE ALANINE TRANSAMINASE REACTION

2.1 Preamble

In this chapter we present key theoretical underpinnings for the experiments performed in Chapters 3 and 5. The development of open-source, freely available, user-friendly quantum chemistry software (such as NWChem developed at Pacific Northwest National Laboratory), as well as an increasing knowledge base about how to use these codes in stable isotope geochemistry (Schauble et al., 2006; James R. Rustad, 2016) enables experimentalists to routinely carry out theoretical estimates of expected fractionation factors. The majority of this chapter deals with estimating ^{13}C-^{12}C preference at the α carbon in converting the $CH_3-C-COOH$ backbone between zwitterionic alanine and pyruvate, which is experimentally addressed in Chapter 3. We also estimate the free energy change on exchanging D and H between the $C(H,D)-C(H,D)_3$ α and β carbons, which is the focus of measurements presented in Chapter 5.

The first part of this chapter, which focuses on α site carbon isotope fractionation between alanine and pyruvate, formed the basis for a collaborative publication with McNeill et al., 2020, listed in Published Content and Contributions. The collaboration with Dave Dixon (U Alabama) and Eric Bylaska (PNNL) was initiated because of the persistent difficulties reconciling our experimental measurements (the subject of Chapter 3) with our theoretical predictions presented here. We envisioned that the work done in this chapter would be combined with the experimental measurements in Chapter 3 in a publication. It does a thorough job of laying the groundwork for theoretical understanding, including empirical techniques, computational calculation of full potential energy surfaces including anharmonic corrections, full vibrational calculations using density functional and molecular orbital methods, a complete treatment of isotopic substitutions at the non-exchanging site, the

potential effects of hydrated and enolated forms of pyruvate, as well as calculations on the α-ketoglutarate/glutamic acid system. Partly because of computational facilities, we were limited to calculations on single molecules without explicit solvation effects. We sought out the collaboration with U Alabama and PNNL to leverage their capabilities in large-scale electronic structure calculations and ab initio molecular dynamics simulations to include explicit water molecules. While including solvation effects did not fundamentally change the ultimate prediction of the carbon isotope fractionation, those capabilities were novel enough to warrant a separate, computationally focused publication because treatment of solvation effects is important in across a wide range of problems in chemical sciences. Dixon's group therefore took the lead in the McNeill et al., 2020 publication, despite the fact that the groundwork for the specific alanine-pyruvate isotope fractionation problem was very completely laid out by our group. In the current scientific culture, work that shows a discrepancy between theoretical predictions and experimental observations is generally not published in a formal academic journal. For this reason we have put off journal publication of this chapter until the discrepancy between theory and experiment can be reconciled.

2.2 Abstract

In the first part of this chapter, we performed a set of electronic structure calculations to estimate the equilibrium constant, K_1, for the exchange of ^{13}C between the α carbon of alanine and the corresponding carbon site of pyruvate, an isotopic exchange that is possible in biological systems via alanine transaminase (ALT), as given in Reaction 2.1. Along with the alanine-pyruvate couple, we also studied the simpler analogous molecular couples, methylamine-formaldehyde (MA-FA), and isopropylamine-acetone (IA-AC). We used density functional theory (DFT) to calculate the zero-point energies within the local density approximation (LDA), as well as with the hybrid B3LYP functional. Solvent effects in the alanine-pyruvate system were approximated by protonating the pyruvate to pyruvic acid (to mimic hydrogen bond donation into the carboxylic acid group) and by protonating the amine group on alanine to create H–alanine$^+$. The MA-FA, IA-AC, and alanine-pyruvate systems all indicate fractionations of -7 to -12 ‰ (expressed as 1000*ln(K_{eq}), where $K_{eq} = \left(\frac{[^{12}C{=}O][^{13}CHNH_2]}{[^{13}C{=}O][^{12}CHNH_2]}\right)$). The solvent-corrected pyruvic acid-H–alanine$^+$ system predicts fractionations of $\sim$ -10 to -15 ‰, depending on the exchange-

correlation functional used. Benchmark calculations on the MA-FA system at the MP2-aug-cc-pvTZ level give a much lower fractionation, -4.4 ‰, relative to that predicted by DFT (-8 to -10 ‰), indicating the necessity of additional high-level calculations on the larger systems. Anharmonic corrections, estimated by DFT-LDA calculation of the C=O and C–NH_2 potential surfaces are estimated to be less than 0.5 ‰. Calculations on the hydrated from of pyruvate, measured by ^{1}H NMR to be present at 10 to 20% under experimental conditions, predict, unexpectedly, a stronger fractionation (by 2 to 5 ‰, depending on the DFT functional). It is also predicted that the presence of ^{2}H at the α-carbon site (rather than ^{1}H) will decrease the expected fraction by 30%. The fractionation predicted by DFT, even without the use of position-specific carbon isotope analysis (i.e., the isotopic signal is diluted by a factor of three by averaging over the three alanine carbon sites using compound-specific isotopic analysis).

The last section of this Chapter (2.7 presents electronic structure calculations estimating the distribution of ^{1}H and ^{2}H between the α and β sites in alanine, relevant to Chapter 5. We put the calculations in this Chapter to keep the electronic structure calculations in one place.

2.3 Introduction

The measurement of isotopic fractionation at particular sites within molecules, known as site-specific, or position-specific isotope analysis (PSIA), has the potential to reveal diagnostic information about processes that produce molecules observed in natural systems. Notable examples include recent studies of carbon isotopes in (1) propane to reveal formation mechanisms in hydrocarbon reservoirs (Xie et al., 2018), (2) methionine to discern between sources (Neubauer et al., 2018), and (3) amino acids in the Murchison meteorite—a well-studied carbonaceous chondrite—where it was proposed that abiotic alanine recovered from the meteorite may have been produced by the Strecker reaction mechanism (Chimiak et al., 2020). Such measurements may also be useful in understanding the details of chemical reaction mechanisms through measurement of position-specific kinetic isotope effects. Before embarking on such problems, it is important to establish the technique by making measurements on simple systems at equilibrium. While gas-phase systems at high temperatures could be used for this purpose, a large number of stable isotope studies today are focused on relatively low-temperature biogeochemical processes taking

place in the presence of water. In addition to the difficulties of the isotope-ratio mass spectrometry measurements themselves, successful PSIA in these systems depends heavily on quantitative separations that do not impose additional fractionation on the analytes. It is imperative to validate techniques by using the simplest chemical reaction possible that has no difficulties coming to chemical equilibrium at aqueous temperatures. We choose here a simple aqueous biochemical reaction, the alanine transaminase (ALT) reaction as a first foray into applying PSIA in aqueous systems. The ALT enzyme catalyzes the transfer of an amino group from alanine to α-ketoglutarate to produce pyruvate and glutamate (see Figure 2.1). This reaction swaps an NH_2 group for an oxygen at the α-carbon site, thus an equilibrium isotopic fractionation at the α-carbon site is expected for the isotope exchange reaction:

$$\begin{aligned} CH_3{-}{}^{13}\mathbf{C}O{-}COOH + CH_3{-}{}^{12}CHNH_2{-}COOH \rightleftharpoons \\ CH_3{-}{}^{12}CO{-}COOH + CH_3{-}{}^{13}\mathbf{C}HNH_2{-}COOH \end{aligned} \tag{2.1}$$

According to the general rule that heavy isotopes fractionate into shorter, stronger bonds (Schauble, 2004), we expect that this isotope exchange reaction will be driven to the left at chemical equilibrium, i.e., $CH{-}CO{-}COOH$ will be preferentially enriched in ^{13}C because of the strength of the $C{=}O$ bond relative to $C{-}NH_2$.

We have performed compound-specific isotopic analysis of the evolving alanine produced during this reaction (see Chapter 3) and, surprisingly, do not observe whole-molecule depletion of ^{13}C at a precision of < 1 ‰. The position-specific measurement would, *prima facie*, be expected to yield a threefold larger fractionation than the compound-specific measurement, assuming the isotopic composition of the methyl and carboxylic acid groups do not significantly affect isotope exchange at the α-carbon. In this chapter we estimate the magnitude of this fraction using a variety of techniques ranging from fully empirical estimates to fully ab initio electronic structure calculations and also will include attempts to account for the role of solvation in driving the observed fractionation. The hypothesis driving the work is that our chemical intuition based on bond length/bond strength considerations is too simplistic and that the range of values obtained from the various more quantitative and sophisticated estimation methods will indicate the plausibility of a site-specific

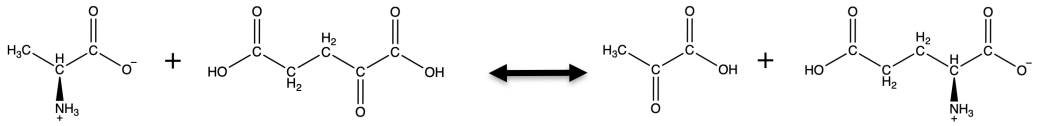

Figure 2.1: The alanine-pyruvate transamination reaction, catalyzed by the alanine transaminase enzyme (ALT), EC number 2.6.1.2.

fractionation of less than 3 ‰, consistent with our inability to measure any significant compound-specific fractionation at < 1 ‰ precision.

2.4 Predicted Isotope Effects on Molecular Free Energies Based on C−N, C−H, and C=O Vibrations Only

In this section we present two semi-quantitative analyses of the vibrational energies of isotopically substituted forms of the molecules of interest to this study, based on application of simplified analytical treatments of bond vibrations and their associated isotope effects. We present these analyses prior to the density functional theory (DFT)-based electronic structure calculations and wavefunction-based first-principles models constructed for this study in order to develop physical intuition regarding the directions and magnitudes and causes of the isotope effects we expect to observe in the more complete (and presumably accurate) DFT and first-principles models in Section 2.5. This exercise, in a sense, quantifies our intuition about the expected fractionation qualitatively based on the relative strengths of the carbonyl and amine bonds and also assess the likely magnitude of anharmonic effects.

2.4.1 Approximate Analysis Based on Canonical Frequencies for Nearest-Neighbor Stretching Modes

One way to estimate the partitioning of ^{13}C between pyruvate and alanine by Reaction 2.1 (and analogous reactions) is to consider the energetics of isotope partitioning in simpler reactions of the form $R_2{}^{13}C{=}O + R_2{}^{12}CH{-}NH_2 \rightleftharpoons R_2{}^{12}C{=}O + R_2{}^{13}CH{-}NH_2$, which define a set of analogs for pyruvate-alanine. The simplest representative system would be methylamine and formaldehyde, where the R groups are all hydrogen. However, it is instructive to perform an analysis of such reactions using vibrational frequencies for two-atom stretching motions in canonical "functional groups", as approximations for the specific frequencies in the particular molecules of interest. This exercise is simply an

attempt to quantify our intuition about the sense of the fractionation. For this calculation, frequencies were taken from (K. Nakanishi, 1962).

1. The carbonyl group C=O stretch: 1750 cm^{-1}
2. The amine group C–N stretch: 1100 cm^{-1}
3. The amine group C–H stretch: 3000 cm^{-1}

These canonical frequencies can be assumed to represent the most abundant isotopic forms of each functional group; $^{12}C{=}^{16}O$, $^{12}C{-}^{14}N$, and $^{12}C{-}^{1}H$, respectively. Table 2.1 presents the reduced masses of diatomic stretches for the ^{12}C and ^{13}C substituted forms of each functional group, under the approximation that each nuclide's mass corresponds to a cardinal number mass in Da (e.g., the reduced mass for the $^{12}C{=}^{16}O$ group = (12*16)/(12+16)). This information can be used to compute the frequency associated with diatomic stretching for the ^{13}C substituted form of each function group, using the relation: $\omega_{13} = \omega_{12} \times \sqrt{(\mu_{12}/\mu_{13}}$ where ω_i and μ_i are the frequency and reduced mass, and i values of 12 or 13 indicate the ^{12}C or ^{13}C substituted forms, respectively (Table 2.1).

The free energy of a quantum harmonic oscillator is given by (see, e.g., K. Denbigh, 1955):

$$A = \sum_i \frac{1}{2} h\omega_i + k_b T \ln(1 - e^{-\frac{hc\omega_i}{k_b T}}) \tag{2.2}$$

For the isotope exchange reaction:

$$^{13}\mathbf{C}{=}O + {}^{12}CHNH_2 \rightleftharpoons {}^{12}C{=}O + {}^{13}\mathbf{C}HNH_2 \tag{2.3}$$

Table 2.1: Reduced masses of diatomic stretches for the ^{12}C and ^{13}C substituted forms of functional groups $^{12}C{=}^{16}O$, $^{12}C{-}^{14}N$, and $^{12}C{-}^{1}H$

	μ_{12}	μ_{13}	$\omega_{12}(cm^{-1})$	$\omega_{13}(cm^{-1})$
C–N	6.4615	6.7407	1100	1076.98
C=O	6.8571	7.1724	1750	1711.11
C–H	0.9231	0.9286	3000	2991.11

The vibrational free energy change ΔA can be calculated from Equation 2.2 summed over the four frequencies for the reactants and products (i.e., the C=O, C–N and two C–H stretches). The equilibrium constant, $K_{eq} = \left(\frac{[^{12}C{=}O][^{13}CHNH_2]}{[^{13}C{=}O][^{12}CHNH_2]}\right)$, is equal to $e^{-\frac{\Delta A}{RT}}$. Table 2.2 provides vibrational energies (joules per mole of stretching bonds) and corresponding isotope exchange energies and equilibrium constants at T= 310 K ("body" temperature is used for these calculations, as the experimental work that will follow is concerned with biological fractionation mediated by enzymes taken from warm-blooded mammals). Note that K_{eq} is less than one for the reaction as written in Equation 2.3 as the ^{13}C prefers the stronger bonding environment in the carbonyl bond. At these temperatures and for the isotope and bonds of interest, the thermal contributions to K_{eq} beyond the zero-point energy are less than 1 ‰—negligible at the level of approximation we make here. From Table 2.2, we see that this simple calculation predicts an approximately 16 ‰ enrichment of ^{13}C in the carbon bound to oxygen, corresponding to pyruvate in Reaction 2.1. This prediction seems reasonably in line with general experience in measuring fractionations, and consistent with our prior assumptions and with our motivation for trying this system as a prototype. For example, the ^{13}C-^{12}C fractionation between organic and inorganic materials is around 40 ‰ (Nier et al., 1939). A 16 ‰ site-specific fractionation would be expected to appear as a $\sim$ 5.3 ‰ fractionation in the compound-specific measurements, well within our ability to detect.

Table 2.2: Calculated free energy (A), ΔA, and 1000*ln(K_eq) for the ^{12}C and ^{13}C substituted forms of functional groups $^{12}C{=}^{16}O$, $^{12}C{-}^{14}N$, and $^{12}C{-}^{1}H$

	A ^{12}C (Joules)			A ^{13}C (Joules)		
	ZPE ^{12}C	Thermal ^{12}C	Total ^{12}C	ZPE ^{13}C	Thermal ^{13}C	Total ^{13}C
C-N	1.0925E-20	-2.6111E-23	1.0899E-20	1.0697E-20	-2.9064E-23	1.0668E-20
C=O	1.7381E-20	-1.2765E-24	1.7380E-20	1.6995E-20	-1.5290E-24	1.6994E-20
C-H	2.9797E-20	-3.8691E-27	2.9797E-20	2.9708E-20	-4.0320E-27	2.9708E-20
	ΔA (Joules)	ΔA (J/mol)	K_{eq}	1000*ln(K_{eq})		
ZPE + Therm	6.6655E-23	40.14	0.984554587	-15.6		
ZPE Only	6.9355E-23	41.77	0.983933818	-16.2		

2.4.2 Computation of Potential Energy Surfaces and Anharmonic Effects

A second way to estimate the expected $^{13}C/^{12}C$ fractionation in the chemical system of interest to this work is to compute the potential energy surfaces for the C−N, C−H, and C=O stretching motions using electronic structure calculations on actual molecules. Taking the methylamine-formaldehyde couple as a model system, and using a "z-matrix" internal coordinate representation of the atomic positions, the C−N, C−H, and C=O bond lengths are scanned over a -0.04Å to 0.04Å range about the equilibrium bond length (obtained from a geometry optimization) of the molecule (100 intervals spaced 0.0008Å). The potential surface is fitted to a Morse function: $E = D(1 - e^{(-a(r-r_0))})^2$. Once this is complete, the zero-point energies, along with anharmonic corrections, can be obtained from the parameters D, a, and r_0. D and a are fitted to the computed potential surface; r_0, the equilibrium bond length, is not fitted, but determined on the optimization step prior to the bond-stretch calculations. The potential surface was computed with NWChem (Valiev et al., 2010), using density functional theory (DFT). We used the local density approximation (LDA) with the Slater exchange functional (Slater, 1951), the VWN correlation functional (Vosko et al., 1980), and the Pople-type triple zeta 6-311++G(2d,2p) basis (Ditchfield et al., 1971). The Morse parameters can be used to calculate the associated harmonic frequency, ω_0:

$$\omega_0 = \sqrt{\frac{2Da^2/\mu}{2\pi}} \tag{2.4}$$

(where μ is the reduced mass). The anharmonic constant, x_e, which represents the degree of anharmonicity of the bond of interest, is given by:

$$x_e = \frac{h\nu_0}{4D} \tag{2.5}$$

Values of ν_0 and x_e are then used to compute the energy levels:

$$E_n = h\nu_0[(n + \frac{1}{2}) - x_e(n + \frac{1}{2})^2] \tag{2.6}$$

For additional background information on these calculations, see, e.g., Richet et al., 1977.

Table 2.3: Morse function parameters for HCl potential energy surface (LDA (SVWN) 6-311++G(2d,2p)).

D_e (Ha)	0.1382
a ($bohr^{-1}$)	1.054
r_0 (bohr)	2.438957
ω_0 (cm^-1)	2879.65
x_e	0.018

We tested the accuracy of our calculations by comparing the results of the method outlined above with the potential energy surface of HCl computed at the LDA/6-311++G(2d,2p) level using NWChem. The optimized dimer has a bond length of 129.064 pm (2.438957 bohr). Evaluating the energy surface in the same manner described above gives the parameters in Table 2.3. The computed value of ω_0 (obtained by using the function fitting capabilities in gnuplot) is very close to the 2877.53 cm^{-1} harmonic frequency computed in NWChem (both of these are computed using NWChem, but ω_0 here is from the anharmonic surface and the value just reported is from a harmonic frequency calculation done with analytic second derivatives within the NWChem code). The computed anharmonic constant of 0.018 is equal to the experimental value given in Richet et al., 1977 (note: the experimental value for ω_0 given in Richet et al., 1977 is 2990.946 cm^{-1}).

The set of gnuplot commands to fit the Morse function to the potential energy surfaces is given in Section 2.9.1.

The C−N, C−H, and C=O potential surfaces were evaluated with NWChem in the same way as the HCl surface, using the same level of theory, LDA/6-311++G(2d,2p). Using the local density approximation (rather than gradient-corrected density functional theory) keeps the theory conceptually straightforward and, for carbon isotope exchange on small molecules with no hydrogen bonding, works as well as any other functional (James R. Rustad, 2016, Figure 5.3). The potential energy surface calculations were done on methylamine and formaldehyde, starting from the fully optimized structures obtained using DFT. When scanning a single bond in a model more complex than a simple dimer, all other internal coordinates are relaxed at each constrained point on the C−N, C−H, and C=O bond length scans. The results of the fitting are given in Table 2.4. Evaluating Equation 2.6 for n=0, gives an estimate of 19.3 ‰ enrichment of ^{13}C for formaldehyde relative to methylamine, which is

Table 2.4: D_e, a, r_0 fitted parameters and derived quantities, ω_0 and x_e (using Equations 2.4 and 2.5).

	Fitted parameters		Derived quantities	
C-N	D_e (Ha)	0.13476	ω_0 ^{12}C (cm^{-1})	1097.2
	a (bohr^{-1})	1.04516	ω_0 ^{13}C (cm^{-1})	1074.2
	r_0 (bohr)	2.73179	x_e	0.009
C-H	D_e (Ha)	0.12410	ω_0 ^{12}C (cm^{-1})	2884.7
	a (bohr^{-1})	1.08609	ω_0 ^{13}C (cm^{-1})	2876.0
	r_0 (bohr)	2.09554	x_e	0.026
C=O	D_e (Ha)	0.29332	ω_0 ^{12}C (cm^{-1})	1800.6
	a (bohr^{-1})	1.19750	ω_0 ^{13}C (cm^{-1})	1760.5
	r_0 (bohr)	2.26688	x_e	0.026

in acceptable agreement with the ZPE-only estimate of 16.2 ‰ presented in Section 2.4.1. If only the harmonic term is kept, the estimate is raised by 0.3 ‰ to 19.6 ‰. As with the thermal contributions above the zero-point energy, the anharmonic contribution is small, so we chose to ignore the anharmonic terms hereafter. More particularly, these calculations indicate that while anharmonic effects do tend to reduce the observed fractionation, they are small and cannot explain the lack of fractionation observed in our compound-specific experiments.

The codes to do the "relaxed" scans (adapted from Ohlin, 2013) are given in Section 2.9.2.

2.5 Full Vibrational Calculations Using Density Functional and Molecular Orbital Electronic Structure Calculations

The analyses presented in sections 2.4.1 and 2.4.2 have the advantage that the dynamics and energetics driving isotope exchange reactions can be evaluated by analytical expressions and their causes can be clearly identified. However, they involve substantial simplifications—most notably, they ignore modes of motion other than the stretching frequencies of nearest neighbor atoms; therefore, they are approximations. In this section, we present a fuller analysis of the energetics of ^{13}C/^{12}C isotope exchange reactions among the chemical species of interest to this work, using a DFT treatment of full molecular vibrational energetics.

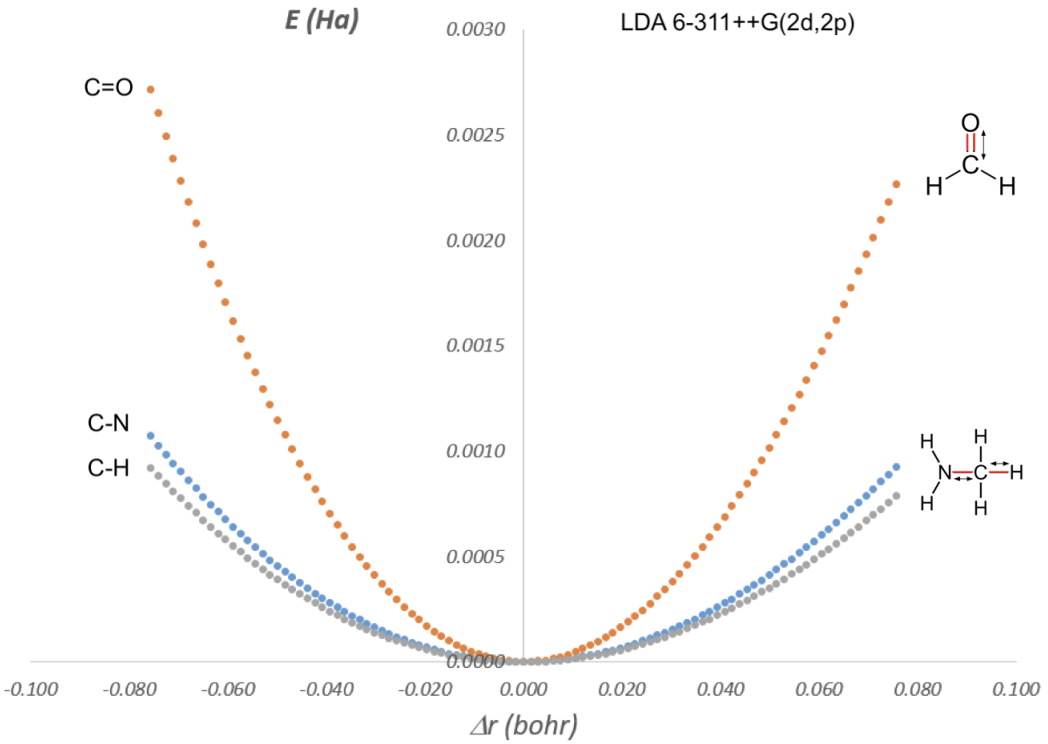

Figure 2.2: Potential energy surfaces for bond stretching calculations (note the use of atomic units).

2.5.1 Density Functional Electronic Structure Calculations

Consider once again ^{12}C–^{13}C isotope exchange between formaldehyde and methylamine:

$$H_2{}^{13}\mathbf{C}O + H_3{}^{12}CNH_2 \rightleftharpoons H_2{}^{12}CO + H_3{}^{13}\mathbf{C}NH_2 \tag{2.7}$$

In the harmonic approximation, the free energy of this reaction can be evaluated by calculating the reduced partition function ratios for methylamine and formaldehyde. The reduced partition function ratio can be obtained from a list of the vibrational frequencies of the molecule with the heavy (h) and light (l) isotopes at the exchanging site.

$$\beta = \left(\frac{Q_h}{Q_l}\right) = \Pi_i \frac{u_{hi}}{u_{li}} \frac{e^{-u_{hi}/2}}{1-e^{-u_{hi}}} \frac{1-e^{-u_{li}}}{e^{-u_{li}/2}} \tag{2.8}$$

where $u = 2\pi\hbar c\omega/k_bT$ (see Urey, 1947, and Bigeleisen et al., 1947). In this expression, the Teller-Redlich rule is used to eliminate the need to evaluate

the rotational contributions, and the frequencies run over the 3N-6 internal vibrations.

This is only strictly applicable to molecules in the gas phase. For the systems of interest here, the molecules are dissolved in water, and the six rotational and translational degrees of freedom are, in part, externally hindered, and become, in part, solvent-coupled vibrations. These vibrations are difficult to measure, and there are no measurements available for the pyruvate/alanine system. They can, in principle, be calculated by including solvent water molecules in quantum mechanical calculations (see James R. Rustad, 2016), or, in classical molecular dynamics calculations (see Bacsik et al., 2002). This level of detail is beyond the scope of this study and is ignored (i.e., we are effectively evaluating the zero-point energy of the isotope exchange equilibria among rotationless vapor phase molecules).

As noted in Section 2.4.1, the energetics of the isotope exchange reactions that we consider ($^{13}C/^{12}C$ exchanges among the compounds of interest at 310 K) can be evaluated to a good approximation considering only the zero-point energy contributions. For a harmonic oscillator with N normal modes the zero-point energy is:

$$\mathrm{ZPE} = \sum_{i}^{N} \frac{1}{2} hc\omega_i \tag{2.9}$$

where h is Planck's constant, c is the speed of light, and ω_i is the frequency of the i^{th} mode. We define ΔZPE to be the difference in ZPE between the ^{13}C-substituted and normal (^{12}C-containing) version of a given compound. Thus, Δ ZPE for the ^{12}C–^{13}C exchange can be defined:

$$\Delta\mathrm{ZPE} = \sum_{i}^{N} \frac{1}{2} hc(\omega_i^{12} - \omega_i^{13}) \tag{2.10}$$

The FORTRAN code used to calculate the ΔZPE at 310 K is given in Section 2.9.3.

Here we define ΔZPE as a positive number that represents the change in zero-point energy upon substitution of ^{12}C for ^{13}C in a given bond environment, (i.e., heavy-isotope bearing molecules have lower vibrational energies

than their light-isotope bearing equivalents). ΔZPE has larger absolute magnitude for molecules with more confining bonding environments (i.e., steeper potential energy surfaces) for the atomic site in question.

To estimate the equilibrium constant for a given isotope exchange reaction, we calculate the difference in ΔZPE for the exchanging molecules (designated ΔΔZPE). For example, to estimate the equilibrium constant for the reaction in Equation 2.7, take ΔZPE(formaldehyde) - ΔZPE(methylamine) (468.11 J/mol–442.8 J/mol) = +25.3 J/mol; K_{eq} = exp(-25.3/RT); 1000*ln(K_{eq}) = -9.8 (see Table 2.5). To keep the signs straight, note that Equation 2.7 has ^{12}C (products) swapped for ^{13}C (reactants) for the formaldehyde. This is "positive", according to the convention described immediately above. On the other hand, Equation 2.7 has ^{13}C (products) swapped for ^{12}C (reactants) in methylamine. This is "negative" according to the convention. Again, K_{eq} is less than 1 as the reaction, as written in Equation 2.7, with ^{12}C in formaldehyde in the products, is unfavorable.

ΔZPE estimates for compounds and substitutions of interest to this study are given in Table 2.5. The main conceptual difference between these values and those given in Sections 2.4.1 and 2.4.2 is that the values in Table 2.5 consider energetic contributions from all of the normal modes, not just the C=O, C–N, and C–H di-atomic stretching vibrations. The analysis presented here is more complete, but also more opaque in molecules of the complexity of those we consider, because the normal modes cannot be obviously associated with specific molecular motions.

Table 2.5: Estimates of ^{12}C–^{13}C fractionation for the model systems, using only ΔZPE terms, as well as Urey-Bigeleisen-Meyer (abbreviated UBM here). Augmented Triple Zeta (ATZ) represents 6-311++G(2d,2p).

	MP2 AUG3	B3LYP ATZ	LDA ATZ	BP86 ATZ
Formaldehyde (ΔZPE) (J/mol)	464.54	468.11	458.67	452.78
Methylamine (ΔZPE) (J/mol)	453.21	442.8	436.89	431.22
ΔΔZPE	11.33	25.31	21.78	21.56
1000*ln(K_{eq})	-4.4	-9.8	-8.4	-8.4
UBM	-7.4	-12.6	-11.5	–
Acetone(ΔZPE) (J/mol)	–	634.61	631.82	613.29
Isopropylamine(ΔZPE) (J/mol)	–	600.76	601.73	583.7
ΔΔZPE	–	33.85	30.09	29.59
1000*ln(K_{eq})	–	-13.1	-11.7	-11.5
UBM	–	-12.2	-10.9	–

We can also use the fundamental modes computed from our DFT analysis of the vibrational dynamics to evaluate the Urey-Bigeleisen-Meyer expression (Equation 2.8), which considers the isotope effects on excited vibrational states. The results of this analysis are presented in Table 2.5 and indicate minor differences between fractionations calculated using the Urey-Bigeleisen-Mayer expressions versus those evaluated using only ΔZPE terms; these differences are possibly significant for reactions involving the smallest molecules (methylamine and formaldehyde), but negligible for the larger molecules (acetone and isopropylamine). In the interests of simplicity and consistency with Sections 2.4.1 and 2.4.2, our subsequent discussion will continue to focus on fractionations computed using ΔZPE terms only.

The fractionations predicted from the full normal mode analysis are lower than predicted in Sections 2.4.1 and 2.4.2 for the methylamine-formaldehyde couple, but generally similar to that simplified analysis for the acetone-isopropylamine couple. This finding reflects the fact that the contributions of C–H stretches for the hydrogen atoms representing the R groups are magnified in importance in smaller molecules (higher frequencies tend to contribute more to the ΔZPE) and cancel less perfectly. To see this, compare the cumulative fractionation plots in Figure 2.3. This figure plots the cumulative sum of the contributions to ΔZPE for each molecule of interest. In the case of the acetone-isopropylamine fractionation, the contributions are fairly clearly delineated: there are significant contributions from the C–N stretch in isopropylamine from 1200 to 1500 cm^{-1}, at 1750 cm^{-1} from the C=O stretch in acetone, and a contribution from the C–H stretch in isopropylamine near 3000 cm^{-1}. For the methylamine-formaldehyde couple, one can see two contributions from C–H stretching frequencies in formaldehyde and three contributions from C–H stretching frequencies in methylamine. Thus, even though the ΔZPE does rise for formaldehyde after the 1750 cm^{-1} carbonyl stretching frequency (unlike acetone), it rises more for methylamine than for isopropylamine (three C–H stretching frequencies for methylamine, as opposed to one for isopropylamine).

The key takeaway here is that even in the case of formaldehyde-methylamine, with a structural change as large as replacing C-C bonds with C-H bonds, the final result does not change beyond a few permil; regardless of some differences in how the ultimate zero point energies are accumulated across the frequency spectrum, we still predict a nearly 10 ‰ enrichment of ^{13}C in the

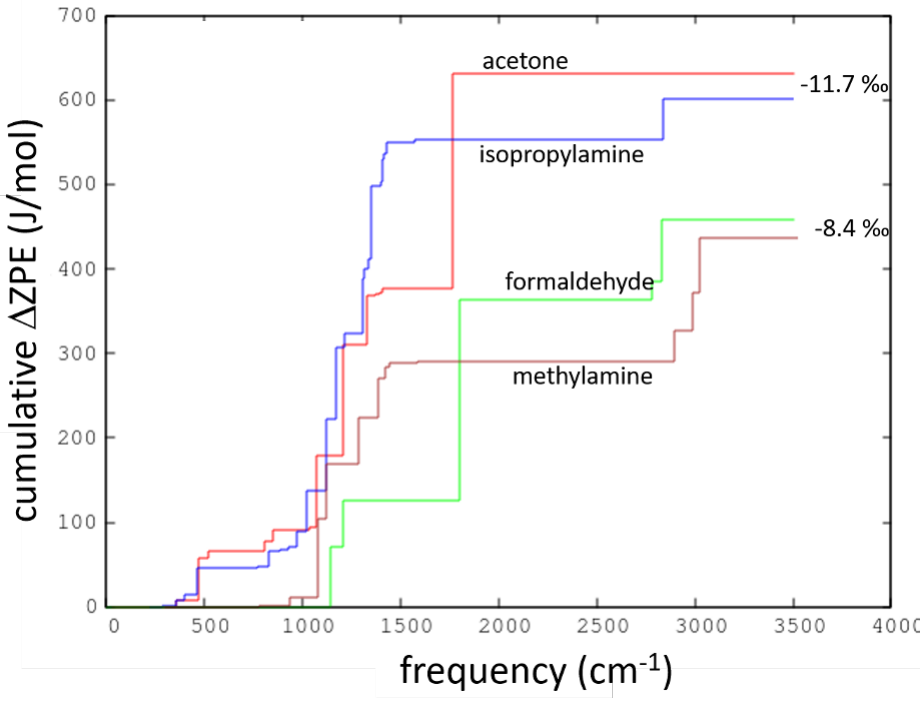

Figure 2.3: Cumulative contributions to ΔZPE as a function of frequency for acetone, isopropylamine, formaldehyde, and methylamine. This style of figure allows visualization and comparison of the frequencies in a given molecule that contribute most significantly to the resulting ΔZPE. In this case, there are large contributions from C–H stretching frequencies coming from the hydrogen atoms representing R groups in methylamine and formaldehyde, occurring around 3000 cm^{-1}., with the result that the ΔΔZPE of the methylamine-formaldehyde couple is smaller than the ΔΔZPE of the acetone/isopropylamine system. These particular frequency contributions are not part of the pyruvate/alanine system, and therefore are not representative of our system of interest.

C=O carbon relative to the C–NH_2 carbon, a result consistent with the "back of the envelope" prediction at the functional group level. Whatever structural perturbations one could make beyond the immediate functional group, such as solvating these structures, these are likely to be smaller perturbations than replacing C–C with C–H. The consistency of results from these sets of calculations argues strongly for a measurable equilibrium fractionation between pyruvate and alanine (see below). NWChem input files for the DFT calculations in this study are given in Section 2.9.4.

2.5.2 Molecular Orbital Calculations on the Methylamine-Formaldehyde System

One advantage to introducing the analog methylamine-formaldehyde is that, even with relatively meager computing resources, this system can be treated at higher levels of theory using molecular orbital methods including many-body perturbation theory approaches (MBPT) such as the 2nd-order MBPT method introduced by Møller and Plesset (MP2) (Møller et al., 1934) and coupled-cluster methods such as CCSD(T) introduced by Bartlett and coworkers (Bartlett et al., 2007). These methods do not rely on an exchange-correlation functional (i.e., the SVWN representation of the local density approximation used above, Slater, 1951, Vosko et al., 1980) and can treat, for example, van der Waals forces that are absent in DFT. They can serve as a check on the performance of DFT methods (this is discussed further below).

The MBPT approaches are expensive, with computational needs growing as the 5^{th} power of the basis set size in MP2 theory, and the 7^{th} power of the basis set size in CCSD(T). Unfortunately these calculations also tend to be much more sensitive to basis set size and design than in DFT. The so-called correlation-consistent (cc-) basis sets by Dunning and co-workers (Kendall et al., 1992) are designed to systematically approach the complete-basis set limit when used with these methods, and the CCSD(T,Q,...) approach when used with the cc-pv(t,q,5,6..)z basis sets represent the current "gold-standard" of quantum chemistry techniques (at least for systems larger than a very few atoms). These computational demands put the coupled-cluster techniques beyond the reach of this study, but we were able to perform vibrational calculations on formaldehyde and methylamine at the MP2-aug-cc-pvTZ level. MP2 can be thought of as an approximation to CCSD when used with a large basis set. It is useful to have one set of "anchor" calculations that tie the DFT calculations here to the world of higher-level methods.

We use the MP2 module of NWChem with the aug-cc-pvTZ basis of Dunning and co-workers (Kendall et al., 1992, which is a triple-zeta basis set. The Pople-type 6-311++G(2d,2p) basis used in our DFT calculations is also an augmented triple-zeta basis set (which we abbreviate ATZ), however the Dunning-type basis is about twice the size of the Pople-type basis.

The results of the MP2 calculations are also shown in Table 2.6. They predict about half the fractionation predicted with density functional theory (-4.4 ‰

Table 2.6: Number of basis functions in the two basis sets used in this work.

	aug-cc-pvTZ	6-311++G(2d,2p)
Methylamine	207	104
Formaldehyde	138	74

at MP2 aug-cc-pvTZ as opposed to -8.4 ‰ at LDA-SVWN 6-311++G(2d,2p)). A fractionation of this modest magnitude is close to the limit of whole-molecule analysis (due to the dilution effect of the other two carbon atoms, for which we expect secondary isotope effects to be negligible) and would likely require position-specific isotope analysis (PSIA) to confirm.

It is important to recognize that these calculations are just a first step towards more accurate quantum chemical calculations. While the subsequent steps, had they been feasible with the computational resources at hand, would lead to a reliable, converged gas-phase fractionation between ^{12}C and ^{13}C in H_2CO and H_3CNH_2, it is important to point out that this initial MP2 aug-cc-pvTZ value cannot serve as any kind of "gold-standard" in itself, or be assumed to be better than the LDA (or, see later B3LYP) 6-311++G(2d,2p) calculations that are the workhorses of this study. It is only after the progressive sequence of wave-function models (e.g., CCSD(T,Q,5,...)), in conjunction with the progressive sequence of basis sets (e.g., aug-cc-pv(t,q,5,6,...)z) have been carried out that a definitively converged value would be obtained. There is also no way to predict whether the calculated fractionation would increase or decrease as these further steps are taken. But this initial step does give some indication of the uncertainties of these calculations, and here it has substantially reduced the fractionation expected for this system. Thus, these calculations might indicate that the fractionations are not as large as the more approximate methods indicate. Note also that, if one were to try to systematically relate/correct the DFT values to the MO values, there is no indication as to whether 4 ‰ should be subtracted from the predicted fractionation for pyruvate-alanine, or that it should instead be multiplied by one half. If, for example, the predicted fractionation is -15 ‰, that difference could be important. Without further study involving additional reaction systems, there is no way to be certain which is more reasonable.

NWChem input files for the MP2 calculations in this study are given in Section 2.9.5.

2.6 Density Functional Theory Calculations on the Pyruvate-Alanine System

2.6.1 Representation of the Aqueous Ions and Effects of Solvation

With the intuition built through the preceding calculations, we now consider electronic structure calculations of the ΔZPE for pyruvate and alanine. In addition to LDA 6-311++G(2d,2p), we also show results for B3LYP (Becke, 1988; Lee et al., 1988) and BP86 (Becke, 1988; Perdew, 1986) functionals, the former being one of the canonical "hybrid" functionals which includes a component of Hartree-Fock exchange. BP86 is one of the early Generalized Gradient Approximation functionals. These GGA and hybrid functionals are included because it is well known that hydrogen bonds are particularly poorly described with LDA (Xantheas, 1995). While we do not consider explicit waters of solvation in these calculations, it is not always completely clear whether some intramolecular hydrogen bonding might be contributing to the fractionation (in fact, the larger α-ketoglutarate/glutamic acid system definitively shows internal hydrogen bonding, as discussed below in Section 2.6.4). Results of these calculations are given in Table 2.7 (also, see discussion below, as there are several molecules listed in Table 2.7 that we have not yet covered). There are not alarming differences between LDA and B3LYP for these systems, although in both cases the B3LYP functional predicts somewhat larger amplitude fractionation than LDA. Although in many cases over many types of quantum chemistry calculations, particularly those involving water, hybrid functionals work significantly better than LDA (Xantheas, 1995), for vibrational frequencies on small molecules there is no indication that this is true. Again, as mentioned above, work by Zarzycki and Rustad (as reported in James R. Rustad, 2016) comparing the performance of many functional/basis set combinations on a set of small molecules taken from Richet et al., 1977, indicates no dramatic improvement of B3LYP over LDA in this context. Here we assume they are equally likely to be correct.

A few questions arise here concerning the exact forms of the molecules to use in the calculations. First, at the pHs of interest, pyruvic acid would be deprotonated, indicating that the calculations should be conducted on the pyruvate anion. To do this, we removed the proton from the carboxylic acid group and ran the calculations with an overall charge of -1e. The fractionation calculated between pyruvate and alanine is somewhat smaller than the 13 to 15 ‰ calculated for the analog compounds, yielding 9.4 ‰ (B3LYP) and 7.2 ‰

(LDA). However, this estimate is likely to be in error due to solvation effects. In solution, the carboxylate group would be accepting hydrogen bonds from surrounding water molecules. Also, in solution, the alanine would be zwitterionic. This would be expected to increase the fractionation with a longer C–NH_3^+ bond relative to C–NH_2. To account for hydrogen bond donation to pyruvate, we ran the pyruvate as pyruvic acid. It is reasonable to expect that the ΔZPE for solvated pyruvate, with hydrogen bonds donated from surrounding water molecules into the carboxylate group, would lie between the gas-phase pyruvate anion and gas-phase pyruvic acid. We attempted to run the alanine as a zwitterion by removing the carboxylic acid proton and moving it to the amine, however the zwitterion was not stable in the gas phase, and the proton moved back onto the carboxylate group from the NH_3^+ group upon optimization of the molecule; i.e., the zwitterion is solvent-stabilized due to hydrogen bonding into the deprotonated carboxylate group. One possible solution to this problem is to protonate both the amine and the carboxylate group, running the molecule as H–alanine$^+$ with the C–NH_2 group protonated to C–NH_3^+. This representation is consistent with doing the vibrational calculation on pyruvic acid as opposed to pyruvate to approximate the solvent effect.

The predicted fractionation between pyruvic acid and H–alanine$^+$ is, as expected, significantly larger than the pyruvate-alanine fractionation, but, interestingly, not necessarily for the expected reason (i.e., the longer C–N bond distance and lower stretching frequency expected for the protonated amine). In fact, the ΔZPE value for H–alanine$^+$ is only about 5 J/mol lower than that for alanine, while the increase in ΔZPE in going from pyruvate to pyruvic acid is 10 J/mol. This is counter-intuitive because adding a proton to the carboxylate group, not directly attached to the isotope exchange site, would not be expected to have as much of an effect as putting a proton on the exchanging site (as was done with alanine). What appears to be happening here is a general stiffening of the pyruvate monoanion when it becomes pyruvic acid. The extent to which that would actually happen in a real solvated environment is not known, but the increase seen for the pyruvic acid is probably an upper limit. Figure 2.4 shows that while the contributions from alanine, H–alanine$^+$, and isopropylamine all give a similar ΔZPE, pyruvic acid, pyruvate anion, and acetone give, by comparison, quite different ΔZPE values. Again, it is surprising that the characteristics of the keto molecule has a stronger effect

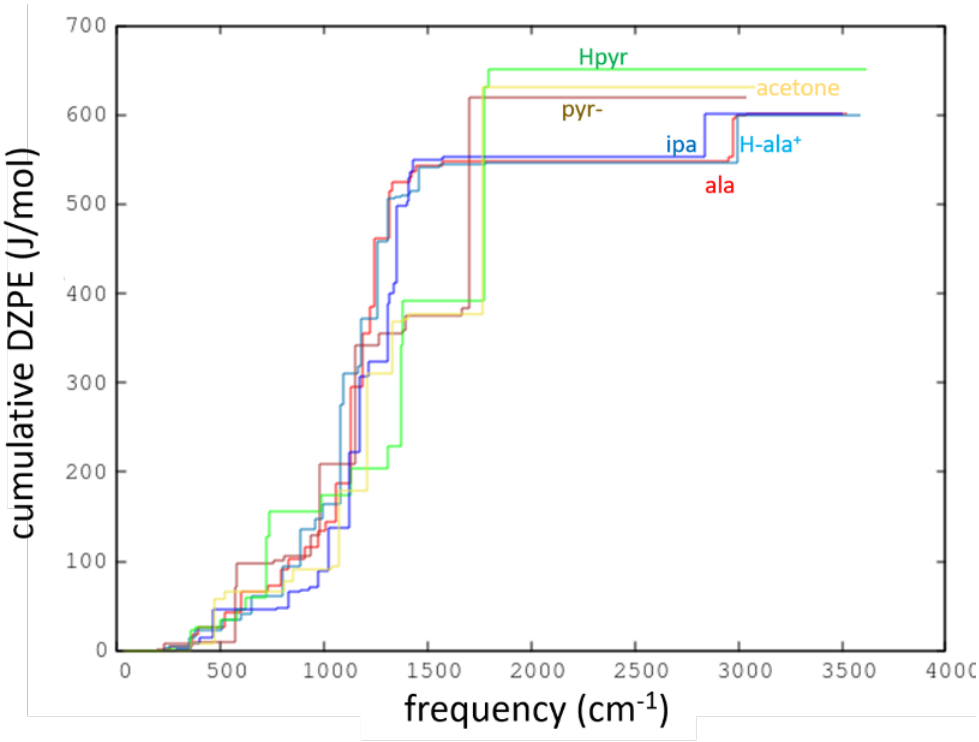

Figure 2.4: Frequency contributions to ΔZPE of Alanine LDA (ala lda), alz +, pyruvate (pyr), pyruvic acid (pyrh), isopropylamine (ipa), and acetone.

than the amine.

NWChem input files for LDA DFT calculations on pyruvic acid and H–alanine$^+$, are given in Section 2.9.6.

2.6.2 Isotopic Substitutions at the Non-Exchanging Site

We next considered the effects of heavy isotope substitutions on the equilibrium constants of the reactions we consider, in atomic sites that are not directly involved in those reactions (i.e., where no bonds to the α-carbon are made or broken over the course of the reaction). In a typical implementation of an electronic structure model calculation, such as in the NWChem code, the program stores masses for each atom and uses them to construct the dynamical matrix $dF_{(\alpha_i)}/d\beta_j/\sqrt{(m_i m_j)}$, the eigenvalues of which are the squares of the frequencies. Within a given code, these masses may be set equal to the abundance-weighted mean atomic weights, or using the atomic weight of the most abundant isotope. For the world outside isotope geochemistry, such choices are not important for most problems. The most complete analysis

Table 2.7: ΔZPE (J/mol), ΔΔZPE (J/mol), and 1000*ln(K_{eq}) for pyruvate (PYR), alanine (ALA), pyruvic acid (H-PYR), H–alanine$^+$ (H-ALA+), 2,2-dihydroxoproanoic acid (DHPA), enolpyruvate (EPYRH), α-ketoglutarate (AKG), and H-glutamic acid$^+$ (H-GLU+). Augmented Triple Zeta (ATZ) represents the 6-311++G(2d,2p) basis.

	B3LYP ATZ	LDA ATZ	BP86 ATZ
PYR ΔZPE (J/mol)	626.19	619.28	597.54
ALA ΔZPE (J/mol)	601.92	600.84	580.27
ΔΔZPE	24.27	18.44	17.27
1000*ln(K_{eq})	-9.4	-7.2	-6.7
H-PYR ΔZPE (J/mol)	634.71	631.98	609.71
H-ALA$^+$ ΔZPE (J/mol)	596.6	600.26	576.86
ΔΔZPE	38.11	31.72	32.85
1000*ln(K_{eq})	-14.8	-12.3	-12.7
DHPA ΔZPE (J/mol)	653.7 (655.4)	656.0 (656.0)	625.9
ΔΔZPE (wrt H-ALA$^+$)	57.1	55.7	49.04
1000*ln(K_{eq})	-22.2	-21.6	-19.3
EPYRH ΔZPE (J/mol)	624.3 (624.9)	627.1 (627.1)	616.3
ΔΔZPE (wrt H-ALA$^+$)	27.7	26.9	39.44
1000*ln(K_{eq})	-10.7	-10.4	-15.6
AKG ΔZPE (J/mol)	645.76	638.88	619.11
H-GLU$^+$ ΔZPE (J/mol)	613.82	620.11	592.09
ΔΔZPE	31.94	18.77	27.02
1000*ln(K_{eq})	-12.4	-7.3	-10.5

would specify the isotopic identity of all atoms in the molecule of interest, including those that are not part of the site of reaction and possible isotope exchange.

We calculated ΔΔZPE values for all 576 isotopologues of $H_3C-C-COOH$ (the base common to both pyruvic acid and H–alanine$^+$) using B3LYP/ATZ. This number is based on the fact that there are two isotopes for H, two for C, and three for O (^{1}H, ^{2}H, ^{12}C, ^{13}C, ^{16}O, ^{17}O, ^{18}O). 2^6=64 for the H, C combinations, (the α-carbon is not included because the ^{12}C, ^{13}C exchange at that site, of course, defines the ΔZPE value). The three choices for each of the two oxygen atoms gives 3^2x64 = 576. We wrote a code (given in Section 2.9.7) to rerun the frequency calculation for all 576 possible isotopologues, after which the

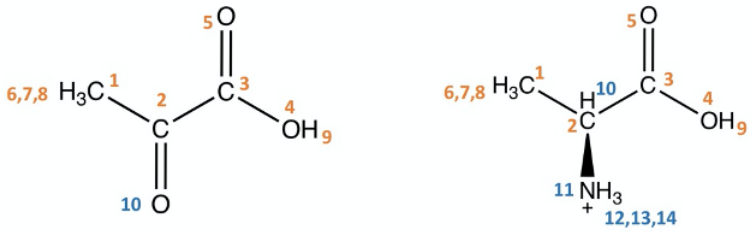

Figure 2.5: Numbered atoms in pyruvic acid (left) and H–alanine$^+$ (right). Atoms in the $H_3C-C-COOH$ moiety common to both pyruvic acid and H–alanine$^+$ are labeled in orange; those unique to either pyruvate or alanine are labeled in blue.

$\Delta\Delta$ZPE values were calculated between H–alanine$^+$ and pyruvic acid, with the same isotopic distribution in the non-exchanging atoms. Consider, for example, the reaction:

$$\begin{aligned} H_3{}^{13}\underline{C}H(^{12}\mathbf{C})NH_3{}^{12}C^{16}O_2H + H_3{}^{13}\underline{C}H(^{13}\mathbf{C})^{16}O^{12}C^{16}O_2H \rightleftharpoons \\ H_3{}^{13}\underline{C}H(^{13}\mathbf{C})NH_3{}^{12}C^{16}O_2H + H_3{}^{13}\underline{C}H(^{12}\mathbf{C})^{16}O^{12}C^{16}O_2H \end{aligned} \tag{2.11}$$

This equation, which represents an exchange reaction at the α-carbon site (in boldface) between H–alanine$^+$ and pyruvate, with ^{13}C substituted at the methyl carbon site (underlined), would be one of these 576 reactions. Note that in Table 2.7, the methyl carbon was taken as ^{12}C. Also note that this procedure does *not* recompute the Hessian matrix (a computationally demanding calculation); the Hessian matrix is re-used, through the "reuse" command in NWChem, and the isotopic substitutions are then used to create the dynamical matrix. We also calculated $\Delta\Delta$ZPE values for reactions in which the exchanging atoms O-10 (pyruvic acid) and H-(10, 12, 13, 14), N-11 (H–alanine$^+$) were substituted with heavy isotopes. Table 2.8 gives fractionations for several choices of isotopes for the exchanging atoms. In this table, pyruvic acid is labeled with ^{16}O, ^{18}O, depending on which oxygen isotope is present at the carbonyl site. For H–alanine$^+$, we show results for "base" (all H as ^{1}H, N as ^{14}N), as well as the indicated isotopic substitutions within the exchanging $HCNH_3$ moiety.

Table 2.8: Variation in ^{13}C–^{12}C fractionation between pyruvate and alanine with additional isotopic substitutions. For each row, we calculated ΔΔZPE values for 576 isotope exchange reactions that include a heavy isotope substitution at one atomic position in the $H_3C-C-COOH$ base common to both alanine and pyruvate. We summarize this by taking the minimum, maximum and average 1000*ln(K) for each set of 576 reactions (i.e., this table represents the results of 12 * 576 = 6912 fractionation calculations). For example, row 1 gives these values for the "base", where all exchanging atoms are present in their most abundant isotopic form (i.e., ^{1}H, ^{13}C, ^{14}N, ^{16}O). Each subsequent row also has one of the exchanging sites substituted with a heavy isotope.

Isotopic Substitution	Average (‰)	Min (‰)	Max (‰)
^{16}O, $HCNH_3$ (base)	-15.1	-14.9	-15.3
^{16}O, $HC^{15}NH_3$	-14.8	-14.6	-15.0
^{16}O, $HCNDH_2$	-14.7	-14.6	-14.9
^{16}O, $DCNH_3$	-10.0	-9.8	-10.1
^{16}O, $DC^{15}NH_3$	-9.6	-9.5	-9.8
^{16}O, $DC^{15}NDH_2$	-9.3	-9.1	-9.4
^{18}O, $HCNH_3$	-16.7	-16.5	-16.9
^{18}O, $HC^{15}NH_3$	-16.4	-16.2	-16.5
^{18}O, $HCNDH_2$	-16.3	-16.1	-16.5
^{18}O, $DCNH_3$	-11.5	-11.4	-11.7
^{18}O, $DC^{15}NH_3$	-11.2	-11.0	-11.4
^{18}O, $DC^{15}NDH_2$	-10.8	-10.7	-11.0

It is apparent from Table 2.8 that none of these substitutions makes a significant difference to the predicted fractionation between the α-carbon site of alanine and the corresponding carbon site of pyruvate, *except* when deuterium is present at the hydrogen attached to the α-carbon, in which case our calculations indicate that the fractionation is reduced by about $\frac{1}{3}$. If it becomes possible to make these measurements with the necessary degree of precision, an interesting check would be to run the experiments with D at the α-carbon site to see if the equilibrium fractionation is substantially reduced. Another way to say this is that these calculations imply a ^{13}C-^{2}H clumping preference in alanine at the α-site.

2.6.3 Fractionation Involving Hydrated and Enolated forms of Pyruvate/Pyruvic Acid

A complication with pyruvate/pyruvic acid is that in solution it is present in at least two additional forms, depending on pH, temperature, and water activity

(Pocker et al., 1969). Around neutral pH, some fraction exists in its hydrated form, 2,2-dihydroxopropanoic acid (DHPA). A much smaller fraction exists in its enolated form (EPYRH) (Muller et al., 1939; Damitio et al., 1992). It can also condense to form 6-carbon structures that can similarly exist in keto, enol and hydrated forms, though these are of insignificant abundance at near-neutral pH and are not considered in this study. Estimates can be made for the expected differences in fractionation for the hydrated and enolated forms.

Intuitively it might be expected that both DHPA and EPYRH would have lower ΔZPE values than pyruvic acid due to loss of the stiff carbonyl bond. While this is observed for EPYRH, electronic structure calculations predict a significantly higher ΔZPE for DHPA than for pyruvic acid (see Table 2.7). As shown in the cumulative ΔZPE plots in Figure 2.4, the C–OH contributions from both C–OH groups, while coming in at lower frequencies than the carbonyl stretch, do collectively raise the total ΔZPE for DHPA well above those for pyruvic acid. This is a good reminder that intuition based on bond strengths is not always a good guide, and that coupling of vibrations at the fractionating site to higher-frequency vibrations at nearby sites can also have a strong influence (see James R. Rustad et al., 2009). In any case, Table 2.7 shows that while a EPYRH shows a little reduction in fractionation, depending on the exchange-correlation functional used, DHPA is predicted to give rise to an equally large, if not larger increase in fractionation. It seems unlikely that these species would have a dramatic effect on the fractionation, even if all the pyruvate component were EPYRH.

NWChem input files used to calculate fractionation involving hydrated and enolated forms of pyruvic acid are given in Section 2.9.8.

2.6.4 The α-Ketoglutarate/Glutamic Acid System

As a final illustrative calculation, we also consider the fractionation between glutamic acid (GLU) and α-ketoglutarate (AKG), the mirror of the pyruvate-alanine couple in the ALT reaction. In keeping with the calculations above, GLU is represented by H–GLU$^+$ (both carboxylic acids protonated, as well as the amine), and glutamate is run as glutamic acid. The predicted fractionation (see Table 2.7) of 12.4 ‰ (B3LYP/ATZ) and 7.3 ‰ (LDA) is smaller than the predictions for pyruvic acid and H–alanine$^+$. This may be caused by the strong internal hydrogen bonding present in the H–GLU$^+$ (see Figure 2.7).

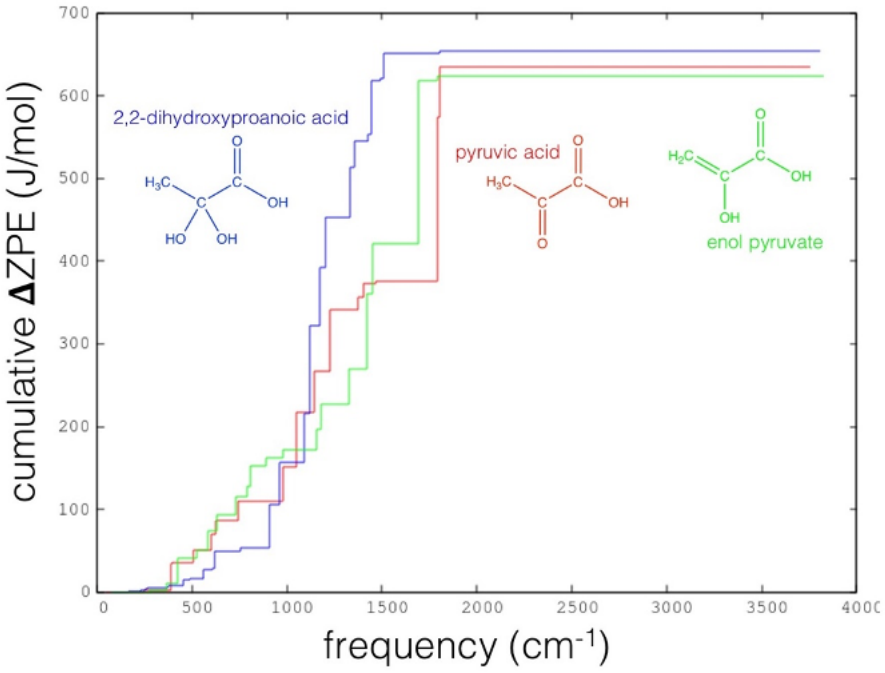

Figure 2.6: Cumulative contributions to ΔZPE of pyruvic acid, dihydroxopropanoic acid, and enol pyruvate.

In the gas phase, H-GLU$^+$ is large enough to have conformational flexibility allowing it to fold over on itself to form internal hydrogen bonds, causing it to stiffen. In solution, GLU would be hydrogen bonded to the surrounding solvent molecules, i.e., the internal hydrogen bonding that we see here in H–GLU$^+$ is essentially an artifact of the gas phase calculation; and not likely representative of the species in solution. The lowest energy conformer of H–GLU$^+$ is thus likely to be different in the gas phase than in solution because of this conformational flexibility. Alanine does not experience an analogous effect, however, because it is too small and thus lacks the conformational flexibility necessary for internal hydrogen bonding. This stiffening of H–GLU$^+$ likely accounts for the high ΔZPE relative to H–alanine$^+$ (see cumulative plot in Figure 2.8, and note that AKG and PYR are closer to each other than H–ALA$^+$ and H–GLU$^+$). For this larger system, it would be a good idea to explore conformational space to some extent, as in James R Rustad, 2009, where molecular dynamics simulations of the amino acids using forces from semi-empirical methods were used to generate conformers. However, for our purposes here, the effort required for such calculations probably would not be justifiable. The main point is that the AKG-GLU system is in broad agreement

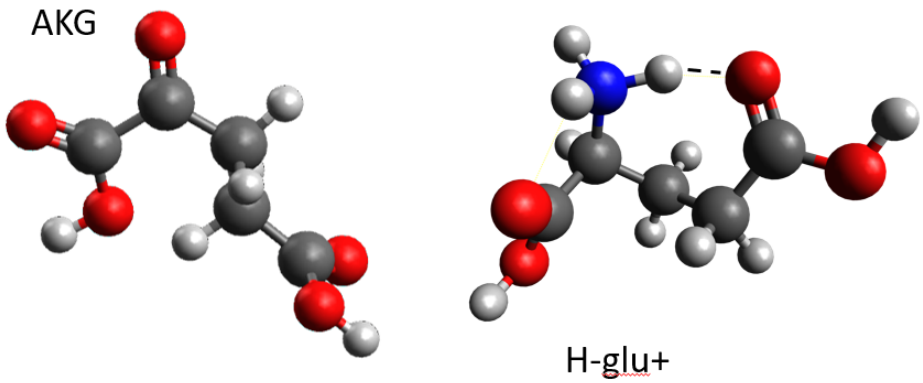

Figure 2.7: Ball and stick representation of α-ketoglutarate and H-GLU$^+$, illustrating internal hydrogen bonding in H-GLU$^+$, which artificially stiffens it.

with the PYR-ALA system.

NWChem input files for AKG and GLU are given in Section 2.9.9.

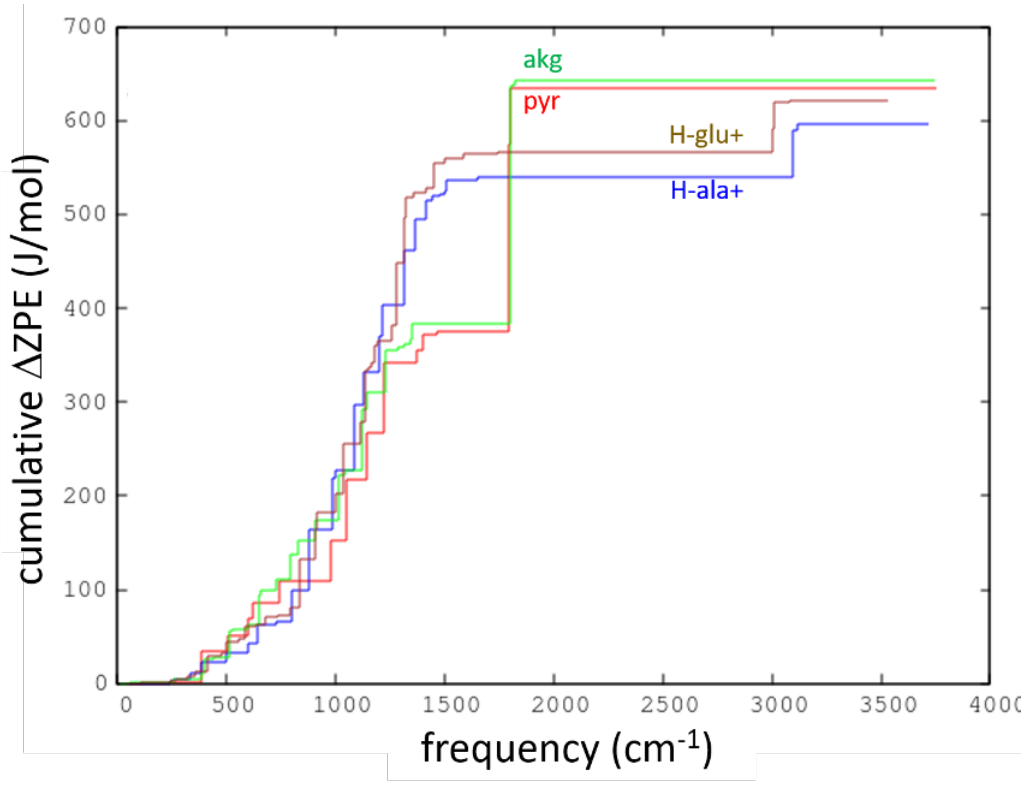

Figure 2.8: Cumulative fractionation for α-ketoglutarate (akg) and H-GLU$^+$ relative to pyruvic acid (pyr) and H-ALA$^+$. The stiffening of H-GLU$^+$ results from internal hydrogen bonding, as seen in Figure 2.7. This is the likely reason for raising of the H-GLU$^+$ above H-ALA$^+$.

2.7 Computational Prediction of Hydrogen Isotope α, β Site Preference at Equilibrium

In anticipation of the experimental work we present in Chapter 5, which involves hydrogen isotope preferences between the α and β carbon sites of alanine, we now turn our attention briefly towards predictions for this D/H site preference. For this set of calculations we use the same methods as outlined in Piasecki et al., 2016, which in that context were used to calculate D/H preferences for the -CH_2- and -CH_3 sites in propane. This involves calculation of the vibrational frequencies using density functional methods, as described in Section 2.5. In keeping with Piasecki et al., 2016, we use the Urey-Bigeleisen-Mayer (UBM) equations based on the reduced partition function ratio and the B3LYP exchange-correlation functional. Instead of using Gaussian software as was used in Piasecki et al., 2016, we continue to use NWChem (as we have done throughout this chapter), which is open source and freely available. Because the B3LYP functional has slightly different formulations between NWChem and Gaussian, we re-ran the Piasecki et al., 2016 calculations on the propane system with DFT using B3LYP/6-311G**, obtaining the following reduced partition function ratios (at 300K) for the three unique H sites: 13.152386 (–CH_2–), 12.114795 (–CH_3(a)), and 12.043858 (–CH_3(b)). This gives $1000\ln(RPFR_{CH_2}/RPFR_{CH_3})$ = 82 ‰ and 88 ‰, respectively, for the two unique hydrogen atoms in the methyl group for propane in C2V symmetry. These results are in acceptable agreement with the value of 93.20 ‰ reported in Piasecki et al., 2016 (it is not specified which hydrogen this value refers to in that paper). The minor differences between our results and those in Piasecki et al., 2016, could be due to the DFT integration grid (we use "xfine", i.e., extra fine), or the precise implementation of the B3LYP functional.

For the calculations on alanine, we use the same H-alanine$^+$ model as described earlier in this Chapter, Section 2.6.1. In fact, using the UBM technique, we are able to re-use the Hessian matrix ($H_{ij} = d^2E/d\alpha_i d\beta_j$) and substitute the mass of ^{2}H (D) in the appropriate site to construct the dynamical matrix ($H_{ij}/\sqrt{(m_i m_j)}$), so no new electronic structure calculations are required. This is done with the "reuse" option in NWChem frequency calculations, as described in Section 2.6.2.

Table 2.9 gives the calculated RPFR values and equilibrium constants at 303 K for a series of H $\rightarrow$ D substitutions using the following notation: the α

hydrogen atom is designated as "α" and the three β hydrogen atoms as "β_1", "β_2" and "β_3". Recall that RPFR values are calculated through computing the change in vibrational frequency that results by changing the mass of an atom at a particular site, in this case, changing ^{1}H to ^{2}H. In Table 2.9 we indicate the substituting site by "HD", and also indicate any deuterium substitutions at other sites, held fixed as D in both ^{1}H to ^{2}H vibrational calculations. For example, "α-HD; β_1-D" refers to an RPFR calculated using vibrational frequencies from "$CDCH_2D$" or "α-D; β_1-D" and "$CHCH_2D$" or "α-H; β_1-D" isotopologues.

Consider the reaction:

$$\alpha_H + \beta_D \rightleftharpoons \alpha_D + \beta_H \tag{2.12}$$

The equilibrium constant:

$$K_{\alpha\beta} = \frac{[\alpha_D][\beta_H]}{[\alpha_H][\beta_D]} \tag{2.13}$$

is given by the ratio of two RPFRs:

$$K_{\alpha\beta_i} = [\alpha_{HD}]/[\beta_{HD_i}] \tag{2.14}$$

where i refers to the three possible choices of substitution at the β site. With ^{12}C at both the α and β (methyl) carbons, we predict that $1000\ln(K_{\alpha\beta})$ lies between 171 ‰ and 179 ‰, depending on which β site is substituted. Preference is greatest for site 2 (trans to the α H), and least for site 3 (cis on the carboxyl side). Site 1 (cis on the amine side) is intermediate between site 2 and site 3, and close to their average. This relative preference for D among the β sites (1,2,3) is preserved throughout the calculations. We assume the β group is rotating at room temperature, washing out the differences in preference among the β hydrogen sites, resulting in a $1000\ln(K_{\alpha\beta})$ of 175 per mil (174.8 in Table 2.9

Note: We abbreviate alanine (and its isotopic variants) as "$CHCH_3$", "$CDCH_3$", etc. for readability and clarity, leaving out the carboxyl and amine functional groups, which do not play a role in the following discussion.

	RPFR α,β-^{12}C	1000*ln(K)	RPFR α-^{13}C	1000*ln(K)	RPFR β-^{13}C	1000*ln(K)
α-HD	14.404803		14.46986		14.409580	
β_1-HD	12.105544	173.9	12.10914	178.1	12.169112	169.0
β_2-HD	12.038915	179.4	12.042933	183.6	12.102699	174.5
β_3-HD	12.140067	171.1	12.142889	175.3	12.203561	166.2
AVG 1000*ln($\frac{\alpha-HD}{\beta-HD}$) ($0\beta$-D)		**174.8**		**179.0**		**169.9**
α-HD; β_1-D	14.405662	163.9	14.470609	168.1	14.410414	159.1
		154.7		159.0		149.9
α-HD; β_2-D	14.417122	159.2	14.481793	163.4	14.421722	154.4
		155.4		159.6		150.6
α-HD; β_3-D	14.405483	157.5	14.470389	161.8	14.410212	152.8
		163.0		167.1		158.1
AVG 1000*ln($\frac{\alpha-HD}{\beta-HD}$) ($1\beta$-D)		**159.0**		**163.2**		**154.2**
α-HD; β_1,β_2-D	14.417963	139.4	14.482628	143.6	14.422388	134.7
α-HD; β_1,β_3-D	14.406402	147.8	14.47111	151.9	14.411058	143.1
α-HD; β_2,β_3-D	14.417770	143.2	14.482334	147.3	14.422164	138.5
AVG 1000*ln($\frac{\alpha-HD}{\beta-HD}$) ($2\beta$-D)		**143.5**		**147.6**		**138.8**
β_1-HD; β_2-D	12.295190	15.5	12.298824	15.5	12.358192	15.4
β_1-HD; β_3-D	12.305656	16.4	12.309223	16.4	12.368870	16.3
β_2-HD; β_1-D	12.227517	15.5	12.23158	15.5	12.290747	15.4
β_2-HD; β_3-D	12.239371	16.5	12.243374	16.5	12.302840	16.4
β_3-HD; β_1-D	12.340749	16.4	12.343529	16.4	12.403887	16.3
β_3-HD; β_2-D	12.342207	16.5	12.344993	16.5	12.405370	16.4
AVG 1000*ln($\frac{\beta-HD}{\beta-HD}$) ($1\beta$-D)		**16.2**		**16.1**		**16.0**
β_1-HD; α, β_2-D	12.295908	15.6	12.299534	15.6	12.358763	15.5
β_1-HD; α, β_3-D	12.306441	16.5	12.309837	16.4	12.369599	16.3
β_2-HD; α, β_1-D	12.237959	16.4	12.241739	16.4	12.300960	16.2
β_2-HD; α, β_3-D	12.249810	17.4	12.25348	17.3	12.313044	17.2
β_3-HD; α, β_2-D	12.342761	16.6	12.345454	16.5	12.405749	16.4
β_3-HD; α, β_1-D	12.341383	16.4	12.343957	16.4	12.404441	16.3
AVG 1000*ln($\frac{\beta-HD}{\beta-HD}$) ($\alpha$-D,$1\beta$-D)		**16.5**		**16.5**		**16.3**
β_1-HD; β_2,β_3-D	12.494370	31.6	12.498142	31.6	12.556600	31.3
β_2-HD; β_1,β_3-D	12.427076	31.7	12.431282	31.7	12.489565	31.5
β_3-HD; β_1,β_2-D	12.542156	32.6	12.545059	32.6	12.604535	32.3
AVG 1000*ln($\frac{\beta-HD}{\beta-HD}$) ($2\beta$-D)		**32.0**		**32.0**		**31.7**
β_1-HD; α, β_2,β_3-D	12.495142	31.7	12.498835	31.7	12.557374	31.4
β_2-HD; α, β_1,β_3-D	12.437643	32.6	12.441613	32.6	12.499960	32.3
β_3-HD; α, β_1,β_2-D	12.542755	32.6	12.5455	32.6	12.605116	32.4
AVG 1000*ln($\frac{\beta-HD}{\beta-HD}$) ($2\beta$-D)		**32.3**		**32.3**		**32.0**
α-HD; β_1,β_2,β_3-D	14.418652	1.0	14.483137	0.9	14.423053	0.9
β_1-HD; β_2,β_3-D	12.106265	0.1	12.109767	0.1	12.169817	0.1
β_2-HD; β_1,β_3-D	12.049210	0.9	12.052865	0.8	12.112897	0.8
β_3-HD; β_1,β_2-D	12.140639	0.0	12.143332	0.0	12.204096	0.0

Table 2.9: Calculated alanine RPFR values and equilibrium constants at 303 K for a series of H $\rightarrow$ D substitutions using the following notation: the α hydrogen atom is designated as "α" and the three β hydrogen atoms as "β_1", "β_2" and "β_3". We indicate the substituting site by "HD", and also indicate any deuterium substitutions at other sites, held fixed as D in both ^{1}H to ^{2}H vibrational calculations. For example, "α-HD; β_1-D" refers to an RPFR calculated using vibrational frequencies from "$CDCH_2D$" or "α-D; β_1-D" and "$CHCH_2D$" or "α-H; β_1-D" isotopologues.

The findings in Table 2.9 can be summarized as follows:

1. We predict a D preference of ∼180 ‰ α site relative to the β site for the reaction $CHCH_2D \rightleftharpoons CDCH_3$.

2. We predict a ∼16 ‰ clumping effect stabilizing β D substitution for each D already present in the β group (i.e., if there are already two deuterium atoms present in the β group the α site preference for a third deuterium atom is reduced by ∼32 ‰). Thus $1000ln(\mathrm{K})$ for $CHCHD_2 \rightleftharpoons CDCH_2D$ is reduced to 159 ‰ and $1000ln(\mathrm{K})$ for $CHCD_3 \rightleftharpoons CDCHD_2$ is reduced to 143 ‰.

3. We predict that the presence of D vs. H at the α position has negligible effect (1 ‰ or less) on substitutions at the β position (and vice versa). (i.e., the $1000ln(\mathrm{K})$ for a reaction such as $CHCD_3 + CDCH_3 \rightleftharpoons CDCD_3 + CHCH_3$ is less than 1 ‰).

4. We predict a ∼5 ‰ clumping effect stabilizing β-D substitutions if ^{13}C is present in the β position. Thus, all of the equilibrium constants for hydrogen-deuterium exchange ("HDX") reactions are reduced by ∼5 ‰ if ^{13}C is present in the β position. Likewise, there is a corresponding ∼5 ‰ clumping effect stabilizing α-D substitutions if ^{13}C is present in the α position. In this scenario, all of the equilibrium constants for HDX reactions are increased by ∼5 ‰.

2.8 Conclusions

Our findings can be summarized briefly as follows:

- Our electronic structure calculations are consistent with empirical estimates based on measured C–N, C=O, and C–H vibrational frequencies.
- Predicted PYR-ALA fractionations are similar to the analogous formaldehyde-methylamine and acetone-isopropylamine systems.
- Anharmonic effects are unlikely to influence the results.
- Isotopic clumping effects are negligible except for H/D at the α-carbon.
- Different forms of pyruvate are unlikely to result in significantly different fractionation factors.

The predicted 10–15 ‰ depletion of ^{13}C in the α-carbon site of alanine relative to the corresponding carbon site in pyruvate, predicted by the empirical estimates and DFT calculations, is large enough to be observable even at the whole-molecule level. We should expect to measure a $\sim$3–5 ‰ fractionation with whole-molecule analysis, $\frac{1}{3}$ of the predicted fractionation, due to signal dilution from the other two carbon atoms, whose isotope ratios remain constant assuming negligible carbon-carbon clumping. A signal of this magnitude is well within the measurement capabilities of, e.g., EA/IRMS. The original hypothesis motivating this work was that our chemical intuition based on bond length/bond strength considerations would likely be too simplistic to make accurate predictions, and that the range of values obtained from the various more quantitative and sophisticated estimation methods will indicate the plausibility of a site-specific fractionation of less than 3 ‰, consistent with our inability to measure any significant compound-specific fractionation at < 1 ‰ precision. Results from these calculations (as well as the calculations on solvent effects reported in McNeill et al., 2020) effectively disprove this hypothesis–there is no way to reconcile our GC-IRMS and EA/IRMS measurements (documented in Chapter 3) with these calculations.

Our MP2 calculations, which could only be carried out on the smallest representative system (the methylamine-formaldehyde system) do predict a smaller fractionation and may point towards a resolution of the questions posed by the experimental findings, but without a full treatment, perhaps with coupled cluster methods and more complete basis sets, it is really not possible to come to any conclusions as to whether the MP2 calculations are any better than the DFT calculations. The MP2 calculations do suggest, however, that it would be worthwhile doing exhaustive high-level calculations (e.g., large-basis set coupled-cluster calculations) on the formaldehyde-methylamine fractionation; these molecules are small enough that the high-level calculations could be attempted with relatively modest resources and provide indicative values of what to expect for alanine-pyruvate fractionation. Another avenue for further exploration is the treatment of solvent effects. Although we believe our approximations would be more likely to overestimate the true solvent effects, this needs to be further tested.

References

Bacsik, Z, J N Canongia Lopes, M F Costa Gomes, G Jansco, J Mink, and A A H Pádua (2002). "Solubility isotope effects in aqueous solutions of methane". In: *The Journal of Chemical Physics* 116.24, pp. 10816–10824.

Bartlett, Rodney J and Monika Musiał (2007). "Coupled-cluster theory in quantum chemistry". In: *Reviews of Modern Physics* 79.1, pp. 291–352.

Becke, AD (1988). "Density-functional exchange-energy approximation with correct asymptotic behavior." In: *Physical review. A, General physics* 38.6, pp. 3098–3100.

Bigeleisen, Jacob and Maria Goeppert Mayer (1947). "Calculation of Equilibrium Constants for Isotopic Exchange Reactions". In: *The Journal of Chemical Physics* 15.5, pp. 261–267.

Chimiak, L, J E Elsila, B Dallas, J P Dworkin, J C Aponte, A L Sessions, and John M Eiler (2020). "Carbon isotope evidence for the substrates and mechanisms of prebiotic synthesis in the early solar system". In: *Geochimica et Cosmochimica Acta*, pp. 1–37. DOI: 10.1016/j.gca.2020.09.026.

Damitio, J, G Smith, J E Meany, and Y Pocker (1992). "A comparative study of the enolization of pyruvate and the reversible dehydration of pyruvate hydrate". In: *Journal of the American Chemical Society* 114.8, pp. 3081–3087.

Ditchfield, R, W J Hehre, and J A Pople (1971). "Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules". In: *The Journal of Chemical Physics* 54.2, pp. 724–728.

K. Denbigh (1955). *Principles of Chemical Equilibrium*. Cambridge University Press, Cambridge, UK.

K. Nakanishi (1962). *Infrared Absorption Spectroscopy, Practical*. Holden-Day, San Francisco, Calif.

Kendall, Rick A, Thom H Dunning Jr., and Robert J Harrison (1992). "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions". In: *The Journal of Chemical Physics* 96.9, pp. 6796–6806.

Lee, Chengteh, Weitao Yang, and Robert G Parr (1988). "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density". In: *Physical Review B* 37.2, pp. 785–789.

McNeill, Ashley S, Brooke H Dallas, John M Eiler, Eric J. Bylaska, and David A. Dixon (2020). "Reaction Energetics and ^{13}C Fractionation of Alanine Transamination in the Aqueous and Gas Phases". In: *The Journal of Physical Chemistry A* 124.10, pp. 2077–2089. DOI: 10.1021/acs.jpca.9b11783.

Møller, Christian and Milton S Plesset (1934). "Note on an Approximation Treatment for Many-Electron Systems". In: *Physical Review* 46.7, pp. 618–622.

Muller, Otto H and J Percy Baumberger (1939). "The Keto-Enol Tautomerism of Pyruvate Ion Studied Polarographically". In: *Journal of the American Chemical Society* 61.3, pp. 590–596.

Neubauer, Cajetan, Michael J Sweredoski, Annie Moradian, Dianne K Newman, Richard J Robins, and John M Eiler (2018). "Scanning the isotopic structure of molecules by tandem mass spectrometry". In: *International Journal of Mass Spectrometry* 434, pp. 276–286. DOI: 10.1016/j.ijms.2018.08.001.

Nier, A O and E A Gulbransen (1939). "Variations in the relative abundance of the carbon isotopes". In: *Journal of the American Chemical Society* 61, pp. 697–698.

Ohlin, Andy (2013). *503. (relaxed) PES scanning in Nwchem revisited.* URL: http://verahill.blogspot.com/2013/08/503-relaxed-pes-scanning-in-nwchem.html.

Perdew, JP (1986). "Density-functional approximation for the correlation energy of the inhomogeneous electron gas." In: *Physical review. B, Condensed matter* 33.12, pp. 8822–8824.

Piasecki, Alison, Alex Sessions, Brian Peterson, and John M Eiler (2016). "Prediction of equilibrium distributions of isotopologues for methane, ethane and propane using density functional theory". In: *Geochimica et Cosmochimica Acta* 190.C, pp. 1–12.

Pocker, Y, J E Meany, B J Nist, and C Zadorojny (1969). "Reversible hydration of pyruvic acid. I. Equilibrium studies". In: *ACS Publications* 73.9, pp. 2879–2882.

Richet, P, Y Bottinga, and M Javoy (1977). "A review of hydrogen, carbon, nitrogen, oxygen, sulphur, and chlorine stable isotope fractionation among gaseous molecules". In: pp. 1–47.

Rustad, James R (2009). "Ab initio calculation of the carbon isotope signatures of amino acids". In: *Organic Geochemistry* 40.6, pp. 720–723.

– (2016). "Computational isotope geochemistry". In: *Molecular modeling of geochemical reactions: An Introduction*. Ed. by JD Kubicki. Wiley and Sons.

Rustad, James R. and Qing-Zhu Yin (2009). "Iron isotope fractionation in the Earth's lower mantle". In: *Nature Geoscience* 2.7, pp. 514–518.

Schauble, Edwin A (2004). "Applying stable isotope fractionation theory to new systems". In: *Geochemistry of Non-Traditional Stable Isotopes*. Ed. by C. M. Johnson, B. L. Beard, and F. Albarade. Vol. 55. Reviews in Mineralogy and Geochemistry, pp. 65–111.

Schauble, Edwin A, Prosenjit Ghosh, and John M. Eiler (2006). "Preferential formation of 13C–18O bonds in carbonate minerals, estimated using first-principles lattice dynamics". In: *Geochimica et Cosmochimica Acta* 70.10, pp. 2510–2529. ISSN: 0016-7037. DOI: 10.1016/j.gca.2006.02.011.

Slater, J C (1951). "A Simplification of the Hartree-Fock Method". In: *Physical Review* 81.3, pp. 385–390.

Urey, Harold C (1947). "The thermodynamic properties of isotopic substances". In: *Journal of the Chemical Society (Resumed)*, pp. 562–581.

Valiev, M et al. (2010). "NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations". In: *Computer Physics Communications* 181.9, pp. 1477–1489.

Vosko, S H, L Wilk, and M Nusair (1980). "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis". In: *Canadian Journal of Physics* 58.8, pp. 1200–1211.

Xantheas, Sotiris S. (1995). "Ab initio studies of cyclic water clusters $(H_2O)_n$, n=1–6. III. Comparison of density functional with MP2 results". In: *The Journal of Chemical Physics* 102.11, pp. 4505–4517.

Xie, Hao, Camilo Ponton, Michael J Formolo, Michael Lawson, Brian K Peterson, Max K Lloyd, Alex L Sessions, and John M Eiler (2018). "Position-specific hydrogen isotope equilibrium in propane". In: *Geochimica et Cosmochimica Acta* 238, pp. 193–207.

2.9 Supplementary Materials

2.9.1 Appendix A: GNUPLOT Session for Fitting Parameters for the HCl Surface

```
gnuplot> f(x)=d*(1-exp(-a*x))**2
gnuplot> a=1;d=0.2
gnuplot> fit f(x) 'hcl_lda_atz' via d,a

FIT:  data read from 'hcl_lda_atz'
      format = z
      #datapoints = 101
      residuals are weighted equally (unit weight)

function used for fitting: f(x)
f(x)=d*(1-exp(-a*x))**2
fitted parameters initialized with current variable values

iter      chisq     delta/lim  lambda   d           a
   0 1.4791654077e-06   0.00e+00  8.32e-04  2.000000e-01
   1.000000e+00
  63 2.9512036918e-10  -8.41e-07  8.32e-09  1.382429e-01
   1.054443e+00

After 63 iterations the fit converged.
final sum of squares of residuals : 2.9512e-10
rel. change during last iteration : -8.41313e-12

degrees of freedom    (FIT_NDF)                        : 99
rms of residuals      (FIT_STDFIT) = sqrt(WSSR/ndf) : 1.72656e
   -06
variance of residuals (reduced chisquare) = WSSR/ndf   :
   2.98101e-12

Final set of parameters          Asymptotic Standard Error
=======================          ==========================
d               = 0.138243      +/- 0.001757     (1.271%)
a               = 1.05444       +/- 0.006659     (0.6315%)

correlation matrix of the fit parameters:
                d   a
d               1.000
a              -0.999  1.000
```

2.9.2 Appendix B: "Relaxed" Scans

Adapted from http://verahill.blogspot.com/2013/08/503-relaxed-pes-scanning-in-nwchem.html

2.9.2.1 Formaldehyde:

```
start formaldehyde
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

basis "ao basis" spherical print
o  library "6-311++G(2d,2p)"
h  library "6-311++G(2d,2p)"
c  library "6-311++G(2d,2p)"
end

charge 0
dft
 mult 1
 iterations 200
 grid xfine
 tolerances tight
 tolerances tol_rho 1.e-12
 tolerances accCoul 12
end

driver; tight; end

geometry formaldehyde
c    -0.00000000 -0.00000000 -0.53241468
o    -0.00000001   0.00000000  0.66717002
h    0.95226353  0.00000000  -1.12354372
h    -0.95226352   0.00000000  -1.12354371
end

set geometry formaldehyde

driver
  default
  maxiter 100
end
```

```
39
40 python
41 from nwgeom import *
42 geom = '''
43   geometry adjust
44         zcoord
45           bond 1 2 %f cccc constant
46       end
47   end
48 '''
49 results=scan_input(geom,[1.19878],[1.24038],51,'dft',
     task_optimize)
50 for i in range(0,len(results)):
51   print results[i][0][0],results[i][1]
52 end
53
54 task python
```

2.9.2.2 Methylamine

```
1 basis "ao basis" spherical print
2 N  library "6-311++G(2d,2p)"
3 H  library "6-311++G(2d,2p)"
4 C  library "6-311++G(2d,2p)"
5 end
6
7 charge 0
8 dft
9  mult 1
10 iterations 200
11 grid xfine
12 tolerances tight
13 tolerances tol_rho 1.e-12
14 tolerances accCoul 12
15 end
16
17 driver; tight; end
18
19 geometry methylamine
20 H    -0.64179809 -0.00016870 -1.37980579
21 C    -0.64744677   0.00000550  -0.27090590
22 N    0.64985908  -0.00009222   0.36686769
23 H    1.19623587  0.82132760  0.10383691
24 H    1.19599562  -0.82174859   0.10407113
```

```
H   -1.21884250   0.88155470  0.05774748
H   -1.21910285 -0.88127395   0.05801975
end

set geometry methylamine

driver
  default
  maxiter 100
end

python
from nwgeom import *
geom = '''
  geometry adjust
        zcoord
          bond 2 3 %f cccc constant
      end
  end
'''
results=scan_input(geom,[1.44480],[1.48640],51,'dft',
   task_optimize)
for i in range(0,len(results)):
  print results[i][0][0],results[i][1]
end

task python
```

2.9.3 Appendix C: FORTRAN Code Used to Calculate ΔZPE at 310 K

```
      implicit  real*8 (a-h,o-z)
      parameter (planck=6.62607004D-34) !Joule Seconds
      parameter (clight=29979245800.0d0) !cm/second
      parameter (gas=8.31445980d0) !Joule/mole
      parameter (avogadro=6.022140857D23)
      parameter (temperature=310.150d0)
      real*8 freqh(100)
      real*8 freql(100)
      real*8 zpeh(100),zpel(100)
        read(*,*)n
      do i=1,n
           read(*,*) freqh(i),freql(i)
        enddo
        zpeht=0.0
      do i=1,n
           zpei=0.5*planck*clight*freqh(i)
           zpeht=zpeht+zpei
           zpeh(i)=zpei
        enddo
        zpelt=0.0
      do i=1,n
           zpei=0.5*planck*clight*freql(i)
           zpelt=zpelt+zpei
           zpel(i)=zpei
        enddo
        summ=0
      do i=1,n
           summ=summ+(zpel(i)-zpeh(i))*avogadro
c       print *,freql(i),(zpel(i)-zpeh(i))*avogadro,summ
        enddo
      print *, (zpelt-zpeht)*avogadro
      stop
      end
```

2.9.4 Appendix D: NWChem Input Files for DFT Calculations

Note: these input files will do the DFT calculations in the local density approximation (LDA). This is done by default in NWChem. If another exchange-correlation functional is required, the appropriate line needs to be inserted into the dft block.

2.9.4.1 Formaldehyde

```
start formaldehyde
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

geometry formaldehyde
zmatrix
c
o   1  B1
h   1  B2     2  A1
h   1  B3  2  A2  3  D1
variables
B1  1.20651
B2  1.12578
A1   122.43969
B3  1.12578
A2   122.43969
D1  -180.00000
end
symmetry c1
end

basis "ao basis" spherical print
o  library "6-311++G(2d,2p)"
h  library "6-311++G(2d,2p)"
c  library "6-311++G(2d,2p)"
end

set geometry formaldehyde
charge 0
dft
 mult 1
 iterations 200
```

```
 grid xfine
 tolerances tight
 tolerances tol_rho 1.e-12
 tolerances accCoul 12
end

driver; tight; maxiter 500; end
task dft optimize
task dft frequencies numerical
```

2.9.4.2 Methylamine

```

start methylamine
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

geometry methylamine
zmatrix
H
C   1     B1
N   2     B2  1   A1
H   3     B3  2   A2  1   D1
H   3     B4  2   A3  1   D2
H   2     B5  3   A4  4   D3
H   2     B6  3   A5  4   D4
variables
B1  1.11711
B2  1.44512
A1   116.80296
B3  1.02633
A2   110.35031
D1  58.58089
B4  1.02633
A3   110.34873
D2   -58.65303
B5  1.10619
A4   109.45445
D3   -63.32240
B6  1.10618
A5   109.45268
```

```
D4  -179.51649
end
symmetry c1
end

basis "ao basis" spherical print
h  library "6-311++G(2d,2p)"
n  library "6-311++G(2d,2p)"
c  library "6-311++G(2d,2p)"
end

set geometry methylamine
charge 0
dft
 mult 1
 iterations 200
 grid xfine
 tolerances tight
 tolerances tol_rho 1.e-12
 tolerances accCoul 12
end

driver; tight; maxiter 500; end
task dft optimize
task dft frequencies numerical
```

2.9.4.3 Acetone

```
start acetone
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

basis "ao basis" spherical print
h  library "6-311++G(2d,2p)"
c  library "6-311++G(2d,2p)"
o  library "6-311++G(2d,2p)"
end

geometry acetone
c   0.10391590  -1.41324822   0.00000179
c   0.15370320  0.08864386  0.00000034
```

```
c    -1.17208069    0.79641916   -0.00000221
o    1.20843185   0.69784947   0.00000078
h    1.12667483   -1.82244935  -0.00000098
h    -0.44658558  -1.78508048    0.88588891
h    -0.44659241  -1.78508356  -0.88587964
h    -1.01437298    1.88667371   -0.00000001
h    -1.76958807    0.50589605   -0.88590447
h    -1.76959356    0.50589310   0.88589550
symmetry c1
end

set geometry acetone
charge 0
dft
 mult 1
 iterations 200
 grid xfine
 tolerances tight
 tolerances tol_rho 1.e-12
 tolerances accCoul 12
end

driver; tight; maxiter 500; xyz pyr; end
task dft optimize
task dft frequencies numerical
```

2.9.4.4 Isopropylamine

```
start isopropylamine
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

basis "ao basis" spherical print
h  library "6-311++G(2d,2p)"
c  library "6-311++G(2d,2p)"
n  library "6-311++G(2d,2p)"
end

geometry isopropylamine noautoz
c    0.07143839   0.04160596   0.35262611
c    0.00039146   -1.43804581    0.04680482
```

```
c    -1.24511285    0.71976106   0.04507680
n    1.09920366   0.63341572   -0.48369002
h    0.96120058   -1.94316098    0.26565061
h    -0.78575412 -1.93318157    0.64567004
h    -0.22202063 -1.58224590 -1.02739916
h    -1.20351874    1.80402998   0.26592880
h    -1.47883927    0.60021568   -1.02975094
h    -2.06793978    0.28511933   0.64153433
h    1.18659719   1.63762714   -0.27954955
h    2.01201251   0.20798384   -0.27508607
h    0.26925573   0.15678439   1.44983128
symmetry c1
end

set geometry isopropylamine
charge 0
dft
 mult 1
 iterations 200
 grid xfine
 tolerances tight
 tolerances tol_rho 1.e-12
 tolerances accCoul 12
end

driver; tight; maxiter 500; xyz pyr; end
task dft optimize
task dft frequencies numerical
```

2.9.5 Appendix E: NWChem Input Files for MP2 Calculations

2.9.5.1 Formaldehyde

```
start formaldehyde
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

geometry formaldehyde
zmatrix
C
O   1  B1
H   1  B2     2  A1
H   1  B3  2  A2  3  D1
variables
B1        1.20875
B2        1.09644
A1      121.75464
B3        1.09644
A2      121.75464
D1      180.00000
end
symmetry c1
end

basis "ao basis" spherical print
h  library aug-cc-pvtz
o  library aug-cc-pvtz
c  library aug-cc-pvtz
end

set geometry formaldehyde

mp2
freeze core
end

driver; tight; end

task mp2 optimize
task mp2 frequencies numerical
```

2.9.5.2 Methylamine

```
start methylamine
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

geometry methylamine
zmatrix
H
C   1     B1
N   2     B2  1   A1
H   3     B3  2   A2  1   D1
H   3     B4  2   A3  1   D2
H   2     B5  3   A4  4   D3
H   2     B6  3   A5  4   D4
variables
B1    1.09145
B2    1.45779
A1  114.83454
B3    1.00929
A2  110.45566
D1    58.50664
B4    1.00928
A3  110.43179
D2  -59.07348
B5    1.08574
A4  109.06552
D3  -63.10342
B6    1.08579
A5  109.02048
D4  -179.94423
end
symmetry c1
end

basis "ao basis" spherical print
h  library aug-cc-pvtz
n  library aug-cc-pvtz
c  library aug-cc-pvtz
end

```

```
43 set geometry methylamine
44
45 mp2
46 freeze core
47 end
48
49 task mp2 optimize
50 task mp2 frequencies numerical
```

2.9.6 Appendix F: NWChem Input Files for B3LYP DFT Calculations on Pyruvic Acid and H–Alanine$^+$

2.9.6.1 PYRH

```
start pyrh
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx


basis "ao basis" spherical print
h  library "6-311++G(2d,2p)"
o  library "6-311++G(2d,2p)"
c  library "6-311++G(2d,2p)"
end

geometry pyrh noautoz
c    -0.97700620 -1.66191644 -0.00881858
c    0.21191777  -0.74424711   0.00073269
c    -0.13147017   0.76770984  -0.00362319
o    0.95942201  1.54253777  -0.00068086
o    -1.26308932   1.18031058  -0.00874912
h    -1.60832330 -1.46544766   0.85884435
h    -0.63716461 -2.69292907 -0.00609885
h    -1.59461680 -1.46540902 -0.88627568
h    0.65860209  2.46313112  -0.00335730
o    1.35899033  -1.10592399   0.01169561
symmetry c1
end

set geometry pyrh
charge 0
dft
 xc b3lyp
 mult 1
 iterations 200
 grid xfine
 tolerances tight
 tolerances tol_rho 1.e-12
 tolerances accCoul 12
end

```

```
driver; maxiter 500; xyz pyr; end
task dft optimize
task dft frequencies numerical
```

2.9.6.2 ALZ+

```
start alz+
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

geometry alz+ noautoz
c    -1.37149519    1.18795415   0.41321607
c    -0.21713344    0.61316909   -0.39939848
c    0.14535728   -0.83167579  -0.03265822
o    -0.86409872  -1.65788215  -0.22574359
o    1.24509526   -1.11888637    0.37530983
h    -1.57775557    2.21585419   0.11772297
h    -1.16125700    1.15327643   1.48280735
h    -2.26653876    0.59969331   0.22617989
h    -0.60430122  -2.56473420    0.00782131
h    1.76136017   0.68199004   0.16903325
h    0.96166455   2.10725527   0.54928574
h    1.42919052   1.83015038   -1.01176400
h    -0.43448424    0.65669208   -1.46523175
n    1.07346038   1.38749894   -0.16516138
end

basis "ao basis" spherical print
h  library "6-311++G(2d,2p)"
o  library "6-311++G(2d,2p)"
c  library "6-311++G(2d,2p)"
n  library "6-311++G(2d,2p)"
end

set geometry alz+
charge 1
dft
 xc b3lyp
 mult 1
 iterations 200
 grid xfine
```

```
 tolerances tight
 tolerances tol_rho 1.e-12
 tolerances accCoul 12
end

driver; tight; maxiter 1; xyz alz+; end
task dft optimize
task dft frequencies
```

2.9.7 Appendix G: Frequency Calculation for All 576 Possible Isotopologues

```
      parameter(nfreq=24)
      real*8 freqh(nfreq),freql(nfreq)
      parameter (ntopemax=1000000)
      real*8 dzpe(ntopemax)
      real*8 c12,c13,h1,h2,o16,o17,o18,n14,n15
      integer nmass(natoms),prod
      character*2 label(natoms)
      character*1 digit(3)
      character*14 code(ntopemax)
      character*120 aline
      real*8 mass(3,natoms)
      integer i,j
        digit(1)='1'
        digit(2)='2'
        digit(3)='3'
        nmassc=2
        nmassh=2
        nmasso=3
        nmassn=2
      C12=12.00000000000
      C13=13.00335483507
      H1= 1.00782503223
      H2= 2.01410177812
      O16=  15.99491461957
      O17=  16.99913175650
      O18=  17.99915961286
      N14=  14.00307400443
      N15=  15.00010889888
        label(1)='C '
        label(2)='C '
        label(3)='C '
        label(4)='O '
        label(5)='O '
        label(6)='H '
        label(7)='H '
        label(8)='H '
        label(9)='H '
        label(10)='O '

```

```
40      do i=1,natoms
41        if(label(i).eq.'C ') then
42              nmass(i)=nmassc
43            mass(1,i)=C12
44            mass(2,i)=C13
45        elseif(label(i).eq.'H ')then
46              nmass(i)=nmassh
47            mass(1,i)=H1
48            mass(2,i)=H2
49        elseif(label(i).eq.'O ')then
50              nmass(i)=nmasso
51            mass(1,i)=O16
52            mass(2,i)=O17
53            mass(3,i)=O18
54        elseif(label(i).eq.'N ')then
55              nmass(i)=nmassn
56            mass(1,i)=N14
57            mass(2,i)=N15
58        endif
59        enddo
60
61
62        ncode=0
63        nvals=0
64  10    format(a51)
65  12    format(a56)
66  11      format(a7,2x,f15.10)
67      do i1=1,nmass(1)
68      do i3=1,nmass(3)
69      do i4=1,nmass(4)
70      do i5=1,nmass(5)
71      do i6=1,nmass(6)
72      do i7=1,nmass(7)
73      do i8=1,nmass(8)
74      do i9=1,nmass(9)
75      do i10=1,nmass(10)
76      open(55,file='freq.nw')
77      rewind 55
78        write(55,10)'restart pyrh                                    '
79        write(55,10)'echo                                            '
80        write(55,10)'memory 2000 mb                                  '
81        write(55,10)'scratch_dir /tmp                                '
```

```
   change scratch
      write(55,12)'permanent_dir /home/PYRH/ATZ/REORDER/p
  &erm'  change permanent directory
c234567891123456789212345678931234567894123456789512345678961234567897123456789
      write(55,10)'freq                                                 '
      write(55,10)'reuse                                                '
      write(55,11)'mass  1',mass(i1,1)
      write(55,11)'mass  2',mass(2,2)
      write(55,11)'mass  3',mass(i3,3)
      write(55,11)'mass  4',mass(i4,4)
      write(55,11)'mass  5',mass(i5,5)
      write(55,11)'mass  6',mass(i6,6)
      write(55,11)'mass  7',mass(i7,7)
      write(55,11)'mass  8',mass(i8,8)
      write(55,11)'mass  9',mass(i9,9)
      write(55,11)'mass 10',mass(1,10)
      write(55,10)'end                                                  '
      write(55,10)'task dft frequencies                                 '
      close(55)
    call system ('nwchem ./freq.nw > freq_out')
    open(55,file='freq_out')
      rewind(55)
    do i=1,100000
    read(55,'(a)',err=100,end=100)aline
    if(aline(26:60).eq.'Projected Derivative Dipole Moments') then
    do j=1,8
    read(55,'(a)',err=100,end=100)aline
      enddo
    do j=1,nfreq
    read(55,'(a)')aline
    read(unit=aline(11:18),fmt=*)freqh(j)
      enddo
    endif
      enddo
100   continue
      close(55)
    open(55,file='freq.nw')
    rewind 55
      write(55,10)'restart pyrh                                          '
      write(55,10)'echo                                                  '
      write(55,10)'memory 2000 mb                                        '
```

```
122       write(55,10)'scratch_dir /tmp                                    '
    change scratch directory
123       write(55,12)'permanent_dir /home/PYRH/ATZ/REORDER/perm           '
    change perm
124 c234567891123456789212345678931234567894123456789512345678961234567897123456789

125       write(55,10)'freq                                                '
126       write(55,10)'reuse                                               '
127       write(55,11)'mass  1',mass(i1,1)
128       write(55,11)'mass  2',mass(1,2)
129       write(55,11)'mass  3',mass(i3,3)
130       write(55,11)'mass  4',mass(i4,4)
131       write(55,11)'mass  5',mass(i5,5)
132       write(55,11)'mass  6',mass(i6,6)
133       write(55,11)'mass  7',mass(i7,7)
134       write(55,11)'mass  8',mass(i8,8)
135       write(55,11)'mass  9',mass(i9,9)
136       write(55,11)'mass 10',mass(1,10)
137       write(55,10)'end                                                 '
138       write(55,10)'task dft frequencies                                '
139       close(55)
140     call system ('nwchem ./freq.nw > freq_out')
141     open(55,file='freq_out')
142       rewind(55)
143     do i=1,100000
144     read(55,'(a)',err=200,end=200)aline
145     if(aline(26:60).eq.'Projected Derivative Dipole Moments') then
146     do j=1,8
147     read(55,'(a)',err=100,end=100)aline
148       enddo
149     do j=1,nfreq
150     read(55,'(a)')aline
151     read(unit=aline(11:18),fmt=*)freql(j)
152       enddo
153     endif
154       enddo
155 200   continue
156     open(55,file='freqlist')
157       rewind(55)
158       write(55,*)nfreq
159     do j=1,nfreq
160       write(55,*)freqh(j),freql(j)
```

```
      enddo
      close(55)
    call system('./vg2.x < freqlist > dzpe')
      nvals=nvals+1
    open(55,file='dzpe')
    read(55,*)dzpe(nvals)
      close(55)
      ncode=ncode+1
    code(ncode)=digit(i1)//'*'//digit(i3)//digit(i4)
c &//digit(i5)
  &//digit(i5)//digit(i6)//digit(i7)//digit(i8)//digit(i9)
  &//digit(i10)
    print *,nvals,dzpe(nvals),code(ncode)
      enddo
      enddo
      enddo
      enddo
      enddo
      enddo
      enddo
      enddo
      enddo
    open(55,file='dzpevals')
23    format(f12.6,2x,a14)
    do i=1,nvals
    write(55,23)dzpe(i),code(i)
      enddo
      close(55)
    stop
c234567891123456789212345678931234567894123456789512345678961234567897123
    end
```

2.9.8 Appendix H: NWChem Input Files Used to Calculate Fractionation Involving Hydrated and Enolated Forms of Pyruvic Acid

2.9.8.1 DHPA

```
    start dhpa
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

basis "ao basis" spherical print
h  library "6-311++G(2d,2p)"
c  library "6-311++G(2d,2p)"
o  library "6-311++G(2d,2p)"
end

geometry dhpa noautoz
c    0.08532599   -0.58627457    0.01094416
c    0.80022449   -1.05050869    1.27002698
c    -0.09243680    0.94518502   -0.07599580
o    -1.17689025  -1.18235961  -0.02216110
o    0.83341017   -0.89019824  -1.16270697
o    1.01805670   1.62751381   0.23086103
o    -1.12946844    1.46037453   -0.40742409
h    0.83230901   -2.13989412    1.26874394
h    0.26251230   -0.72560414    2.15951627
h    1.81404360   -0.66167650    1.29386909
h    -1.76686467  -0.59434114  -0.51521472
h    0.71570378   -1.83157587  -1.33198340
h    0.82705522   2.56867893   0.10467300
end

set geometry dhpa

dft
 mult 1
 iterations 200
 xc b3lyp
 grid xfine
 tolerances tight
 tolerances tol_rho 1.e-12
 tolerances accCoul 12
```

```
end

task dft optimize

task dft frequencies
```

2.9.8.2 EPYRH

```
start epyrh
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

basis "ao basis" spherical print
h  library "6-311++G(2d,2p)"
c  library "6-311++G(2d,2p)"
o  library "6-311++G(2d,2p)"
end

geometry epyrh noautoz
c   0.14006052  0.73918552  0.13099020
c   1.26277468  1.37715814  -0.19584448
c   0.15277532  -0.74424935   0.33370661
o   -1.08423524   1.30846306  0.31223832
o   -1.05864293 -1.23638069   0.65932957
o   1.13621016  -1.43275277   0.22003276
h   1.28447993  2.44720157  -0.34862018
h   2.17510532  0.81677380  -0.31205541
h   -1.01931395   2.25780955  0.16700242
h   -0.94505707 -2.19139706   0.76743280
end

set geometry epyrh

dft
 mult 1
 iterations 200
 xc b3lyp
 grid xfine
 tolerances tight
 tolerances tol_rho 1.e-12
 tolerances accCoul 12
```

```
37 end
38
39 driver; tight; maxiter 100; end
40
41 task dft optimize
42
43 task dft frequencies
```

2.9.9 Appendix I: NWCHEM Input Files for α-Ketoglutarate and GLU^{+}

2.9.9.1 α-Ketoglutarate

```
start akg
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

basis "ao basis" spherical print
h  library "6-311++G(2d,2p)"
c  library "6-311++G(2d,2p)"
o  library "6-311++G(2d,2p)"
end

geometry akg noautoz
c   0.57545076  -2.09075453   0.07840194
c   -0.77800957 -1.34932147 -0.11670750
c   -0.74788090   0.09703968  -0.55267198
c   -0.14451563   1.00537800  0.52478286
c   -0.02318887   2.43734807  0.05641988
o   1.54916757  -1.56841597 -0.70357736
o   0.73323409  -3.01373225   0.82245991
o   0.37917568  3.24849965  1.06258728
o   -0.23906928   2.83613476  -1.05775226
h   2.35180443  -2.09198151 -0.55823499
h   0.44773419  4.14285504  0.69829877
h   -1.77031489   0.39365403  -0.77346702
h   -0.16008772   0.18758596  -1.46618265
h   -0.74306140   0.99429633  1.43684098
h   0.85675617  0.67795307  0.81364844
o   -1.78640787 -1.96568069   0.10064432
symmetry c1
end

set geometry akg

dft
 mult 1
 iterations 200
 xc b3lyp
 grid xfine
```

```
 tolerances tight
 tolerances tol_rho 1.e-12
 tolerances accCoul 12
end

driver; tight; maxiter 100; end

task dft optimize

task dft frequencies numerical
```

2.9.9.2 GLU+

```
start glu+
echo

memory 2000 mb
scratch_dir /xxxx
permanent_dir /xxx

basis "ao basis" spherical print
h  library "6-31G*"
c  library "6-31G*"
o  library "6-31G*"
n  library "6-31G*"
end

geometry glu+ noautoz
O    1.50901187  -1.81976598 -0.81036382
O    0.72443926  3.04662364  1.01339558
O    0.42276077  -2.85429868   0.86161844
O    -0.81218956   2.93845222  -0.61786210
N    -1.75376811 -1.35025158   0.54631437
C    -0.44179313   0.14316859  -0.88762442
C    -0.79027860 -1.29595713 -0.52158421
C    0.05563540  0.89553179  0.32252902
C    0.42187292  -2.06738583 -0.06661114
C    -0.07051619   2.37090761  0.16266915
H    -1.34741470   0.65752605  -1.26409737
H    0.29992861  0.15119238  -1.70769061
H    -1.11426948 -1.81318234 -1.46048522
H    -0.58840067   0.57948134  1.17732063
H    1.09651786  0.63485323  0.59483506
H    -1.59198405 -2.20350042   1.09892231
```

```
H    -2.71556526 -1.33542001   0.19633209
H    2.21512619  -2.40133232 -0.43693819
H    0.52372009  3.99792421  0.83695776
H    -1.75376811 -0.35025158   0.54631437
symmetry c1
end

dft
xc b3lyp
end


set geometry glu+

charge 1

task dft optimize

task dft frequencies numerical
```

Chapter 3

MEASURING EQUILIBRIUM AND KINETIC SITE-SPECIFIC INTERMOLECULAR CARBON ISOTOPE FRACTIONATION BETWEEN ALANINE AND PYRUVIC ACID AT THE α-CARBON SITE THROUGH THE ALANINE TRANSAMINASE REACTION

3.1 Introduction

A central goal of many stable isotope studies is to measure the isotopic fractionation between two or more molecules. The focus of this chapter is the carbon isotopic fractionation between equilibrated alanine and pyruvate. In recent years, an overarching goal of many in the community, and the Eiler group in particular, has been to evaluate the efficacy of newly developed high-mass-resolution isotope ratio mass spectrometers to measure *site-specific* isotopic compositions in molecules of increasing molecular weight and complexity, with the hope that this will open such inquiries to a broader set of problems. In particular, Thermo Scientific's Orbitrap-based mass spectrometers hold promise to launch a new era of high-precision, site-specific stable isotope measurements in complex organic molecules, a class of measurements that has generally eluded the field in the many decades since Abelson and Hoering first examined site-specific carbon isotopes of amino acid carboxyl groups using wet chemical cleavage methods in 1961 (Abelson et al., 1961).

The initial framing of this study was to use the relatively simple alanine transamination reaction, as defined in Equation 3.1 and illustrated in Figure 3.1, to work through the various hurdles and stages such a measurement might entail, as much as to determine an experimental value for the ALA/PYR equilibrium fractionation factor. In other words, we initially expected the Orbitrap mass spectrometric analysis to be the most "expensive" aspect of this work, and our goal, setting out, was to analyze the site-specific isotopic structure of alanine, and to be able to relate this to the carbon isotopic composition of pyruvate, the product of the reaction of interest. In reality, however, a series of complications and challenges along the way prevented us from analyzing the samples from this experiment using the Q Exactive instrument for a truly

site-specific measurement. Nevertheless, the data gathered through the experimental methods described in this chapter is sufficient to report site-specific values for the carbon position of interest, C2 (i.e., the α-carbon).

$$\mathrm{CH_3{-}{}^{13}\mathbf{C}O{-}COOH + CH_3{-}{}^{12}CHNH_2{-}COOH} \rightleftharpoons \mathrm{CH_3{-}{}^{12}CO{-}COOH + CH_3{-}{}^{13}\mathbf{C}HNH_2{-}COOH} \tag{3.1}$$

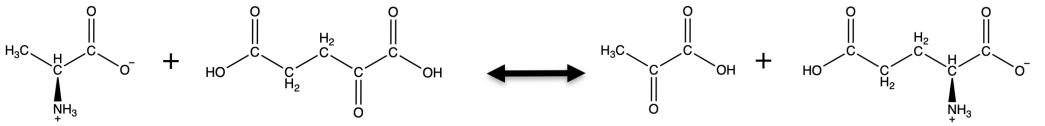

Figure 3.1: The alanine-pyruvate transamination reaction, catalyzed by the alanine transaminase enzyme (ALT), EC number 2.6.1.2.

The $\sim$10 ‰ fractionation at the α-carbon site between alanine and pyruvate predicted from theory in the previous chapter is large enough that the fractionation should be detectable, even without position-specific analysis capabilities. If the isotopic composition of the C1 (carboxyl-carbon) and C3 (methyl-, or β-carbon) sites has a negligible effect on fractionation observed at the α-carbon site, as we expect from calculations described in the previous chapter, then, since there are three carbons in alanine, the site-specific $\sim$10 ‰ preference for ^{13}C in pyruvate should be manifested as a $\sim$3 ‰ whole-molecule (i.e., not site-specific) fractionation, a signal that is easily observed using conventional carbon stable isotope analysis. We are also assuming carbon isotopic preference at the carboxyl-carbon and β-carbon sites is negligible, as predicted in the previous chapter, due to the fact that the bonding environments for these sites are conserved between alanine and pyruvate to a distance of one bond, i.e., the differences between these molecules are two bonds away from these sites, and therefore exert only secondary influence on isotopic preference at these sites. Thus, for the purposes of this study, we attribute any change in carbon isotope composition of the whole molecule solely to isotopic fractionation at the α-carbon site, where bond-breaking occurs.

Given that we expect ^{12}C to preferentially fractionate into the alanine α-carbon site, when we initiate the transamination reaction by addition of ALT enzyme to a solution containing alanine and α-ketoglutarate, absent pyruvate, such

that any pyruvate that forms is derived directly from the alanine in the initial solution, we expect the carbon isotopic composition of the alanine to evolve toward more negative values and approach some limiting $\delta^{13}C$, as alanine is converted to pyruvate and the reaction comes to equilibrium. If there is a normal kinetic isotope effect, we may see a transient rise in $\delta^{13}C$ in alanine as ^{12}C is initially preferentially transmitted into the pyruvate, but eventually the isotopic composition of the alanine will become more depleted in ^{13}C than it started out.

In this chapter we present carbon isotopic measurements on alanine over the course of reaction with α-ketoglutaric acid to produce pyruvate and glutamic acid, mediated by alanine transaminase (ALT). We first determine the chemical equilibrium constant and forward and backward rate constants using ALT derived from porcine heart, by measuring the concentration of alanine over the course of this reaction using ^{1}H NMR and fitting the data using a second-order rate law. We then run a larger volume of the reaction, collecting aliquots from the reaction mixture at intervals on the way to equilibrium, and recovering and isolating the alanine to measure its evolving carbon isotopic composition. Note that we take a somewhat indirect approach here as we are not measuring the carbon isotope composition of both the product (pyruvate) and the reactant (alanine). With our knowledge of reaction progress, and thus mass balance, at the time each sample is recovered, we can determine the carbon isotope fractionation associated with this reaction. From knowledge of the chemical rate law and the equilibrium fraction factor (and hence the ratio of the forward and backward isotopic exchange rates) we can, with a model that couples the isotopic exchange to the chemical reaction rate, predict the isotopic composition of the alanine as a function of time. Finally, we compare our results to our theoretical predictions presented in the previous chapter of this thesis and to models for isotopic evolution presented here.

3.2 Materials and Methods

3.2.1 Experimental Materials

We used unlabeled alanine expected to have isotopic signatures close to natural abundance from Alfa Aesar (L-alanine, 99% purity, Lot # 10167391). Ultrapure water was obtained from a Millipore ultrahigh-purity (18.2 MΩ cm) water system at Caltech. For ^{1}H NMR we also used D_2O purchased from Cambridge Isotope Laboratories, Inc. (D, 99.9% Lot # M3063).

Alanine transaminase (abbreviated ALT, EC number 2.6.1.2) derived from porcine heart was used to catalyze the alanine-glutamate transamination reaction as given in Figure 3.1, thus facilitating exchange of carbon isotopes between alanine and pyruvate pools. ALT (∼7.65 mg protein/mL) suspended in 1.2 M ammonium sulfate was purchased from LeeBiosolutions (Lot # 07B3133). The lot used in the experiments presented here was measured by the vendor to have an activity of 1990 Units per milliliter ammonium sulfate suspension, or 260 U/mg protein, based on an assay run at 310 K and pH 7.4, with initial concentrations of alanine and α-ketoglutarate at 260 mM and 20 mM, respectively, tris buffer at 100 mM, pyridoxal-5'-phosphate at 0.15 mM, and enzyme diluted to <1 U/mL. The protein exists as a dimer in solution, with a molecular weight of ∼100,000 g/mol dimeric ALT (i.e., ∼50,000 g/mol per monomer).

Additional reaction components included pyruvic acid from Sigma Aldrich (purity 98%, Lot # SHBH2749V), α-ketoglutaric acid from ChemCruz (purity 99%, Lot A1717), pyridoxal-5'-phosphate (P5P) from Sigma Aldrich (purity 98%, Lot # SLBM9225), and potassium phosphate dibasic from Sigma Aldrich (purity ¿98%, Batch # 126K0747). For separations we used 30 kDa Amicon Ultra-15 mL centrifugal filters, Dowex 50WX8 hydrogen form, and a Primesep A HPLC column (10x250 mm, Part # A-100.250.051C, Particle 5 μm 100 A).

3.2.2 Experimental Procedures

We prepared reaction media by weighing 331.9 mg of alanine and 5373.25 mg α-ketoglutaric acid, which was dissolved in 670 mL ultrapure H_2O in a 1 L bottle. To this we added 20 mL 6 mM pyridoxal-5'-phosphate (the co-enzyme for this reaction) and 100 mL of 400 mM dipotassium phosphate, and raised the pH to 7.5 by addition of ∼10 mL 5 M NaOH. This brought the solution volume to a total of ∼800 mL, resulting in a concentration of 4.6 mM alanine, 40.6 mM α-ketoglutaric acid, 0.15 mM pyridoxal-5'-phosphate, and 200 mM dipotassium phosphate. We chose a tenfold higher initial concentration for α-ketoglutaric acid relative to alanine, with the intention of pushing the reaction toward a very low alanine concentration (i.e., high pyruvate concentration) relative to initial conditions. This maximizes the expected observable isotopic fractionation in alanine, which is preferred, given that we are not able to measure the carbon isotopic composition of pyruvate to compare directly to alanine. The reaction volume was optimized for recovery of sufficient alanine

for GC-IRMS and EA/IRMS analysis. We set aside 560 μL of this 800 mL solution to measure chemical kinetics as described below, while the rest we reacted separately, collecting samples at intervals throughout the reaction for isotopic analysis of remaining alanine, also described below.

To acquire chemical kinetics data, we filled an NMR tube with 560 μL reaction media set aside as described above, to which was added 30 μL D_2O to serve as a frequency lock. To begin the reaction, 7.5 μL of 50x diluted enzyme solution was added to give an expected enzyme activity of 0.5 U/mL, and the NMR tube vortexed for several seconds and then quickly transferred to an NMR spinner. Quantitative ^{1}H NMR spectra were recorded on a Varian 600 MHz spectrometer with 5 mm inverse triple resonance probe (^{1}H, ^{2}H, ^{13}C, ^{15}N, ^{31}P) set to 303 K. We acquired 8-scan spectra at 4-min intervals spanning 9 hours, resulting in an array of 128 spectra. Using MestReNova software, we stacked these spectra and integrated peak area over the upfield peak of the alanine doublet, tracking this value as change in alanine concentration over the duration of the experiment. We avoided integrating over the downfield peak of the alanine doublet as it suffered interference with the pyruvate hydrate methyl group singlet, which increases over the course of the experiment, while the upfield alanine doublet peak was resolved and sufficient for peak area integration.

To initiate the reaction for isotopic analysis, 200 μL ALT enzyme stock solution was added to the $\sim$800 mL reaction media described above, for an expected enzyme activity of 0.5 U/mL, as above. The reaction was incubated at 303 K, and samples of the solution were removed at 1, 2, 3, 4, 5, 6, 9, 12, and 24 hours. Upon sample removal, ALT reaction was quenched with liquid nitrogen, followed by filtration of the sample solution through 30 kDa Amicon Ultra-15 mL centrifugal filters to remove ALT, thus ensuring the reaction could not continue during the following steps. Each sample solution was eluted through a cation exchange column using Dowex 50WX8 hydrogen form and 2M ammonium hydroxide to remove buffer, pyruvic acid, and α-ketoglutaric acid, recovering a mixture of alanine and glutamate. Collected fractions were tested for presence of amino acids using ninhydrin on TLC plates, followed by ^{1}H NMR, and all fractions testing positive by either method were combined and dried down in a TurboVap. Each of these recovered alanine and glutamate mixtures was then split into two aliquots; one aliquot was run through HPLC

to separate alanine from glutamate using a Primesep A column. Using ^{1}H NMR we verified the purity of each alanine sample recovered to be ¿ 99 % pure of organic contaminants, before sending to UCD SIF for elemental analysis (EA). The second aliquot (mixture of alanine and glutamate) was sent to UCD SIF for derivatization followed by GC-IRMS analysis.

3.2.3 Determination of ALT Reaction Equilibrium and Kinetics

Knowledge of the chemical equilibrium constant, K_{eq}, given in Equation 3.2 for the ALT reaction is necessary, for our experiment is designed to recover samples of alanine while the reaction is in progress, as well as once it has reached equilibrium. Several research groups measured this K_{eq} in the 1940s and 50s: Lénárd in 1942 reporting 0.699 (Lénárd et al., 1942), Darling in 1947 reporting 0.444 (Darling, 1947), and Krebs in 1953 reporting 0.658 (Krebs, 1953). Two more publications on ALT kinetics during the 1960s reported additional values: Hopper and Segal reporting 0.38 and 0.63 (acquired thermodynamically and kinetically, respectively) (Hopper et al., 1962) and Bulos and Handler reporting 0.45 (Bulos et al., 1965). (Note: All of these authors wrote Reaction 1 in the opposite direction, so they actually report reciprocal values in their papers. Oddly, in their discussion, Bulos and Handler (Bulos et al., 1965) somehow confused themselves and compared their K_{eq} of 2.2 (i.e., 1/0.45) with the reciprocals of Krebs' and Hopper and Segal's published values, presenting apparent discrepancies much larger than the actual discrepancies, which on their own are already significant. It is disconcerting that the magnitude of the discrepancy did not cause enough concern to correct the error in the reaction direction, especially since it must have been clear that they were comparing a K_{eq} greater than unity to one less than unity. In their words, "No explanation can be offered for the relatively small difference between K_{eq} as determined by [Krebs (0.658) and Hopper and Segal (0.63)] and the value of 2.2 reported herein.") It is somewhat surprising to see this degree of variation in reported values, with no obvious convergence toward a single value as the number of studies increased, and no meta discussion on the potential explanations for these discrepancies. It is not clear from these literature values which to take as the correct one, so we use our chemical kinetics data to determine K_{eq} for ourselves and compare this to the scattered literature values.

$$K_{eq} = \frac{[\mathrm{PYR}][\mathrm{GLU}]}{[\mathrm{ALA}][\alpha\mathrm{KG}]} \tag{3.2}$$

It is also important that we have some sense of the expected reaction rate, to optimize sampling frequency for isotopic analysis, as well as to determine the length of time that the reaction needs to run to establish equilibrium.

Given the alanine transamination reaction shown in Figure 3.1, we assume a second-order rate law with two parameters, a forward rate, k_f, and a backward rate, k_b:

$$\frac{d[\text{ALA}]}{dt} = -k_f[\text{ALA}][\alpha\text{KG}] + k_b[\text{PYR}][\text{GLU}] \tag{3.3}$$

The negative sign on the first term of the right hand side of the reaction indicates that [ALA] decreases as the reaction proceeds from left to right. It is most convenient to fit k_f and K_{eq}, with k_b taken as k_f/K_{eq}.

These parameters are optimized to obtain the best fit of the numerically integrated equation to the ^{1}H NMR data on [ALA] as described in the previous section. Numerical integration was done with Euler's method, and fitting was performed with the downhill simplex method of Nelder and Meade (Nelder et al., 1965; note: there is more discussion of this technique in Chapter 5 where it is used more extensively). Codes are given in 3.5.2.

The best fit to our data, as shown in Figure 3.2, gave K_{eq} as 0.431, k_f=9.89×10^{-5} mM^{-1} min^{-1} and k_b=k_f/0.431 = 2.32×10^{-4} mM^{-1} min^{-1}. Interestingly, our K_{eq} is closest to Darling's value of 0.444 (Darling, 1947) and second closest to Bulos and Handler's value of 0.45 (Bulos et al., 1965). In Figure 3.2 we compare the fit using these two parameters to a single-parameter fit constraining k_b to k_f/0.658 to match Krebs' equilibrium constant. In the case of the constrained single-parameter fit, k_f=9.40×10^{-5} mM^{-1}min^{-1}. It is visually obvious that our two-parameter fit to the data is significantly better than the constrained single-parameter fit. However, it turns out that choosing either the constrained single-parameter model or the two-parameter model does not affect the conclusions we draw in this paper. Were we to show similar constrained single-parameter fits using the other four K_{eq} values reported in the papers referenced above, all but Lénárd's would be closer to our two parameter fit than Krebs'.

Given our fitted K_{eq} of 0.431 and the starting concentrations of our experiment as described earlier, we expect equilibrated substrate concentrations to be 0.83 mM alanine, 36.8 mM α-ketoglutarate, 3.6 mM pyurate, and 3.6 mM

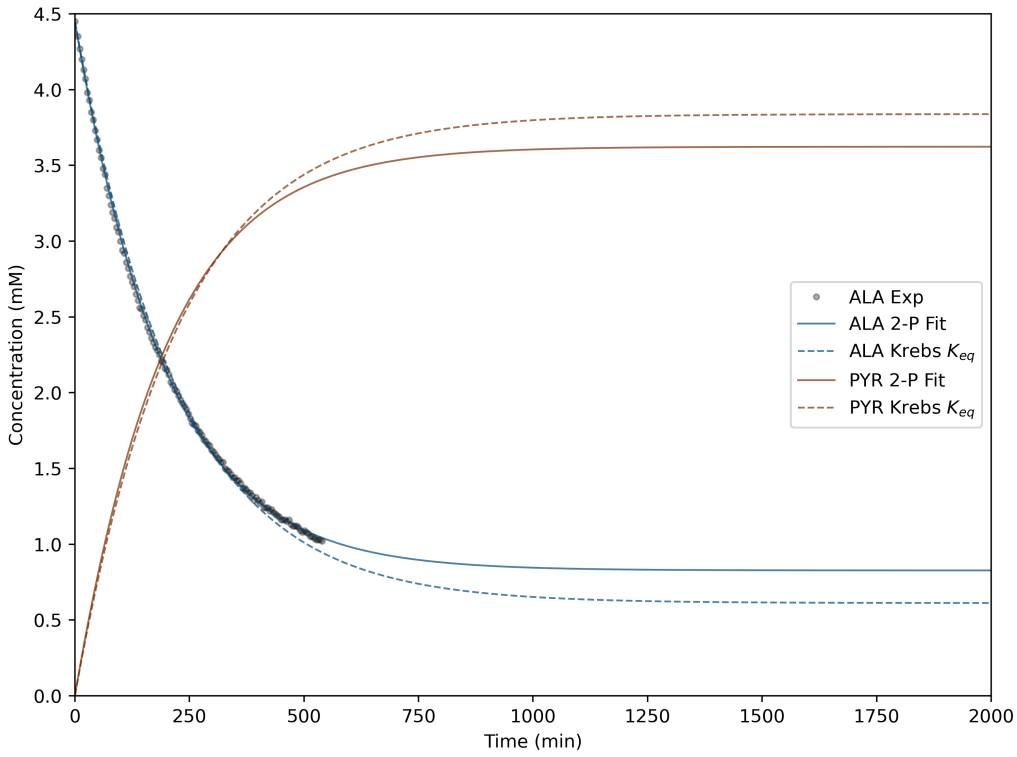

Figure 3.2: Concentration of alanine as a function of time from ^{1}H NMR (gray circles) and the fitted model concentrations of alanine, fitting both K_{eq} and k_f (blue solid line), and fitting k_f with k_b fixed by Krebs' equilibrium constant (blue dashed line). Pyruvate concentration also shown (brown solid line and brown dashed line, respectively).

glutamate. Accordingly, approximately 81.5 % of the initial alanine will be converted to pyruvate, giving an apparent equilibrium isotope fractionation that is 81.5 % of the theoretical value calculated in the previous chapter.

The discrepancy between the K_{eq} determined kinetically and the thermodynamic value given by Krebs is apparently typical of studies on ALT kinetics (0.45 kinetically determined in (Bulos et al., 1965); 0.38 kinetically determined in Hopper et al., 1962). Hopper and Segal report a thermodynamic K_{eq} of 0.63 (Hopper et al., 1962), close to Krebs' value of 0.658, determined in some of their earlier work (Segal et al., 1962). They do not comment on possible reasons for the discrepancy. Here we focus on the kinetics and take the k_f and k_b as two parameters that describe the kinetics that we measured as a basis to understand the isotope exchange kinetics.

Enzyme-catalyzed reactions are not typically represented by rate laws like

Equation 3.3 even for the simplest single-substrate reactions, as enzyme-substrate complexes can introduce additional substrate dissociation parameters. The ALT reaction is relatively complex, operating via a "bi-bi ping-pong" mechanism with the enzyme serving as a reaction target that alternately converts between reactants and products of the two half-reactions (ALA $\rightleftharpoons$ PYR and αKG $\rightleftharpoons$ GLU) (Hopper et al., 1962; Bulos et al., 1965). For our purposes, however, given that we care principally about the carbon kinetic and equilibrium isotope affects, and are not directly concerned with broader enzymology topics such as inhibition, it is sufficient that our "naive" rate law fits our experimental concentration data and successfully predicts the amount of alanine converted to pyruvate with time. It should not be blindly assumed to apply to the alanine transaminase reaction under different reaction conditions, and even under the same conditions using a different ALT isoform (further discussed below). It is also worth noting the that the enzyme concentration used in our experiment is many orders of magnitude lower than the alanine concentration and does not serve as a significant pool of bound alanine or pyruvate during the NH_2/O exchange and thus should not significantly influence the isotopic compositions of alanine and pyruvate.

The porcine-heart-derived ALT used in this study came with documentation from the vendor giving a specific activity of 260 U/mg, from which we can derive a rate constant. Measured specific activity for alanine transaminase ranges over several orders of magnitude, with values ranging from $\sim$1 U/mg for Zea mays (corn) to 2231 U/mg for Hordeum vulgare (barley), depending largely on the type of organism producing it (Good et al., 1992; Schomburg et al., 2007). Note that we are not determining K_m or other Michaelis-Menten kinetic parameters in this study, which would require running the reaction over multiple ranges of substrate concentrations. In 1967, Saier and Jenkins reported K_m values for porcine heart alanine transaminase as 28 mM for alanine and 0.4 mM for α-ketoglutarate (Saier Jr. et al., 1967). K_m values can vary greatly depending on the isoenzyme across organisms, and even within the same organism. More recent studies have shown that distinct isoenzymes of alanine transaminase can have radically differing affinities for substrates, in some cases favoring opposite reaction directions (DeRosa et al., 1975; Duff et al., 2012; McAllister et al., 2013).

We calculate the actual initial reaction rate based on our fitted rate constant,

k_f, and the initial concentrations of alanine and α-ketoglutarate, and compare this value to the expected enzyme activity given the amount of enzyme added to our reaction solution and the activity of the enzyme as assayed by the vendor prior to our purchase:

$$\frac{d[\text{ALA}]}{dt} = -k_f[\text{ALA}][\alpha\text{KG}] + k_b[\text{PYR}][\text{GLU}]$$

$$= -9.89 * 10^{-5}\text{mM}^{-1}\text{min}^{-1} * 4.6\text{mM ALA} * 40.6\text{mM}\alpha\text{KG} \quad (3.4)$$

$$= 0.018\frac{\text{mmol ALA}}{\text{L*min}} = 0.018\frac{\text{U}}{\text{mL}}$$

Initial substrate concentrations for the vendor assay are 260 mM alanine and 20 mM α-ketoglutarate, and the assay is performed with diluted enzyme at <1 U/mL, which is then concentrated to 1990 U/mL in the commercial product. As described in the experimental section above, we diluted the concentrated enzyme solution by a factor of ∼4000, to a targeted value of 0.5 U/mL, assuming substrate concentrations identical to the assay conditions. Using our fitted k_f and the vendor assay substrate concentrations, we calculate an initial reaction rate of 0.51 U/mL, matching the target value. That is to say, our rate constant gives us the initial reaction rate we would expect, had we used the assay substrate concentrations, and otherwise identical conditions. In reality, experimental temperature (304 K) and buffer system (phosphate) were also different from the assay conditions (310 K and tris buffer), but they appear not to have made a significant difference here. It is actually somewhat surprising that this calculation results in an answer so close to the target value, since the assumption that the observed reaction rate is directly proportional to substrate concentrations over a wide range in a multi-substrate enzyme-catalyzed reaction is not necessarily justified. We conducted this check simply to gauge whether the enzyme was behaving roughly as expected based off of its documentation.

3.2.4 Expected Isotopic Evolution of Alanine

We used the rate law in Equation 3.3 as a basis for generating models of isotopic evolution as a function of different magnitudes of kinetic isotope effect in the following manner: we begin with the concentration of ^{13}C at -19.83 ‰, the

composition of our Alfa Aesar alanine standard on the VPDB scale, as measured by GC-IRMS. An element of alanine, d[ALA], is drawn from the alanine pool with ratio $^{12}C/^{13}C$ (r hereafter), to be converted to pyruvate. Note that in the context of creating the code, we find it easier to think of $^{12}C/^{13}C$ rather than the more usual inverse because the transmission factor (see immediately below), representing the kinetic isotope effect is then slightly greater than 1 for a normal kinetic isotope effect, showing preferential transmission of ^{12}C. However, wherever we report an isotopic composition in terms of δ notation we use the conventional $^{13}C/^{12}C$ ratio. We postulate a transmission factor, k_{i_f}, giving the amount of ^{12}C or ^{13}C that preferentially converts from reactants to products, as well as a similar factor, k_{i_b}, giving the amount of ^{12}C or ^{13}C that preferentially converts from products to reactants in the backward reaction. If there is no preference for the forward reaction, k_{i_f} would equal 1.000. However, from theory we expect that the pyruvate will preferentially incorporate ^{13}C by ∼10 ‰, and so expect, for the backward reaction, that k_{i_b} = 1.010, that is, the pyruvate preferentially binds ^{13}C, so the transmission of ^{12}C is enhanced. Note that we use the expected site-specific ∼10 ‰ fractionation here in simulating the reaction; the ∼3 ‰ signature is a result of the analytical methods we used in the study, which involve combustion to CO_2 resulting in dilution of the α-carbon with the carboxyl carbon and β-carbon, which we assume are not subjected to significant isotopic fractionation on the $CH_3-C-COOH$ backbone that is conserved between alanine and pyruvate. Note also that there is a difference between these "transmission factors" as we have defined them here and the actual isotope exchange rates that we would measure, for example, under conditions of chemical equilibrium. These transmission factors, as written, are unitless numbers that couple the isotopic exchange to the chemical reaction rates. Going a bit further, we could think of a normal kinetic isotope effect (where "normal" means ^{12}C is preferentially transmitted) being represented by a k_{i_f} that is greater than one. That is, there is a preferential transmission of ^{12}C as we convert the products to reactants at a rate determined by the chemical kinetics. For example, a 5 ‰ normal kinetic isotope effect would have a k_{i_f} = 1.005, and, given that we expect a ∼10 ‰ fractionation at equilibrium, k_{i_b} would then be set to 1.015. The isotopic flux is incorporated into the chemical rate law as follows: let an amount d[ALA] of alanine, with isotope ratio r=$d[^{12}C_{ALA}]/d[^{13}C_{ALA}]$, be converted from product to reactant. For the forward reaction in a normal KIE, a certain amount, ε, of

$^{13}C_{ALA}$ is removed from the transmitted element of alanine and a corresponding amount of $^{12}C_{ALA}$ is added. It is evident that the amount removed/added satisfies the equation:

$$k_{i_f} r = \frac{d[^{12}C_{ALA}] + \varepsilon}{d[^{13}C_{ALA}] - \varepsilon} \tag{3.5}$$

Solving this equation for ε yields:

$$\varepsilon = \frac{d[^{13}C_{ALA}]k_{i_f} r - d[^{12}C_{ALA}]}{1 + k_{i_f} r} \tag{3.6}$$

Exactly the same considerations apply to the reverse reaction, but now choosing k_{i_b} as the rate constant. At each timestep in the chemical reaction, equation 3.6 (as well as its equivalent in the reverse direction) is solved, which determines the evolution of the isotopic composition of the alanine and pyruvate.

A Python script (given in 3.5.3) was written to follow the isotopic composition of alanine for ∼0, 5, 10, and 15 ‰ kinetic isotope effects. The expected evolution is given in Figure 3.3. The magnitude of the KIE is adjusted by increasing k_{i_f} up from 1.000 (0 ‰ KIE) to 1.015 (∼15 ‰ KIE).

The model behaves as expected, with the initial difference in alanine and pyruvate isotopic compositions reflecting the kinetic isotope effect, with a transient enrichment of ^{13}C in alanine that rises with the magnitude of the kinetic isotope effect, eventually reaching the equilibrium value of a ∼3 ‰ whole-molecule enrichment of ^{13}C in pyruvate at equilibrium. The model shows that isotopic equilibrium should be obtained within approximately a day (∼1500 minutes), which it is forced to by the one-to-one coupling with the chemical evolution of the system. However, our experimental data are obviously inconsistent with the theoretical predictions.

It is worth noting also the sensitivity of the isotopic evolution to the chemical kinetics. While the 2-parameter model is a better fit to our chemical kinetics data, the 1-parameter model, with K_{eq} fixed to Krebs' value, is not terribly far off (as depicted in Figure 3.2). The difference between them has a surprisingly large effect on the prediction of the time dependence of the isotopic equilibrium with a significantly larger transient enrichment of alanine in ^{13}C, especially for large KIEs (as shown in Figure 3.3). Neither model, however, predicts our isotopic measurements.

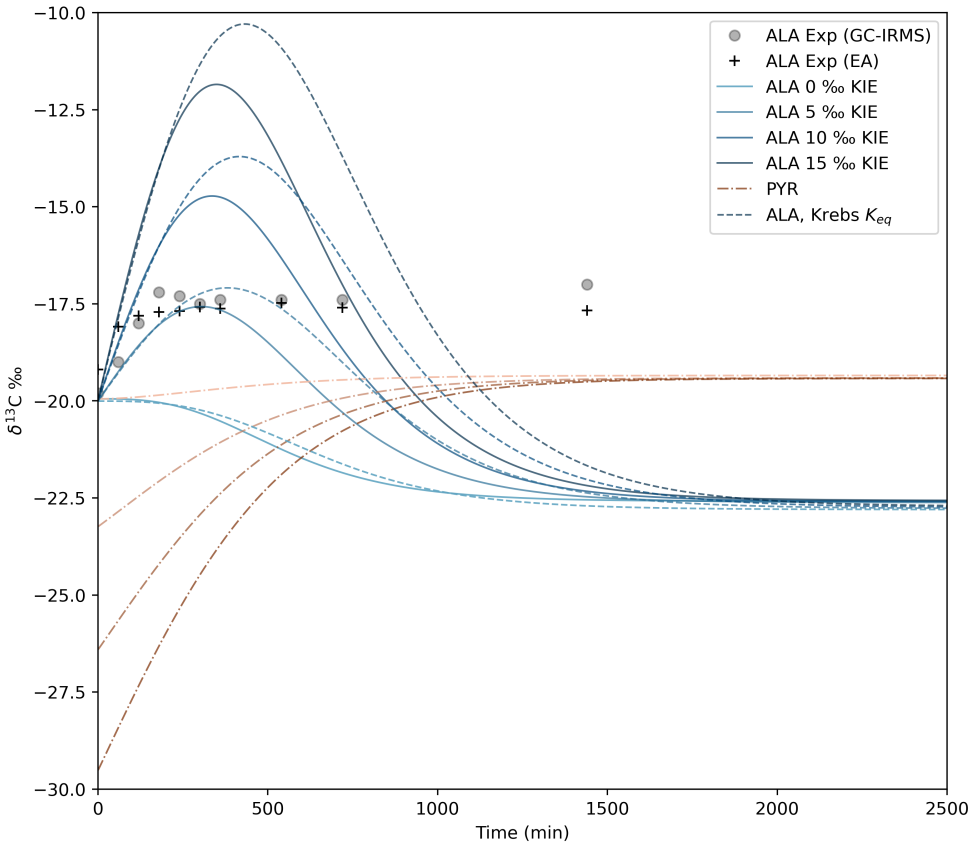

Figure 3.3: Predicted whole-molecule carbon isotopic composition of alanine (dashed lines) and pyruvate (dash-dot lines) as a function of time, with the kinetic parameters used in Figure 3.2. Blue and black dots are the GC-IRMS and EA/IRMS data, respectively. Thin lines for alanine are calculated from model constrained to Krebs' chemical K_{eq}.

Only by reversing the direction of the equilibrium isotope affect can we reproduce the data as it approaches equilibrium. That is, if:

$$1000 * \ln(k_{i_f}/k_{i_b}) = 10 \tag{3.7}$$

Thus, our data indicate that there is a ∼10 ‰ site-specific enrichment (∼3 ‰ whole-molecule) in the other direction, with alanine being enriched in ^{13}C relative to pyruvate. In Figure 3.4 we show predictions for a range of possible kinetic isotope effects. Not only do the data indicate the opposite sense of fractionation from what was predicted, but the time dependence is not consistent with any single kinetic fractionation factor, even with the reversed fractiona-

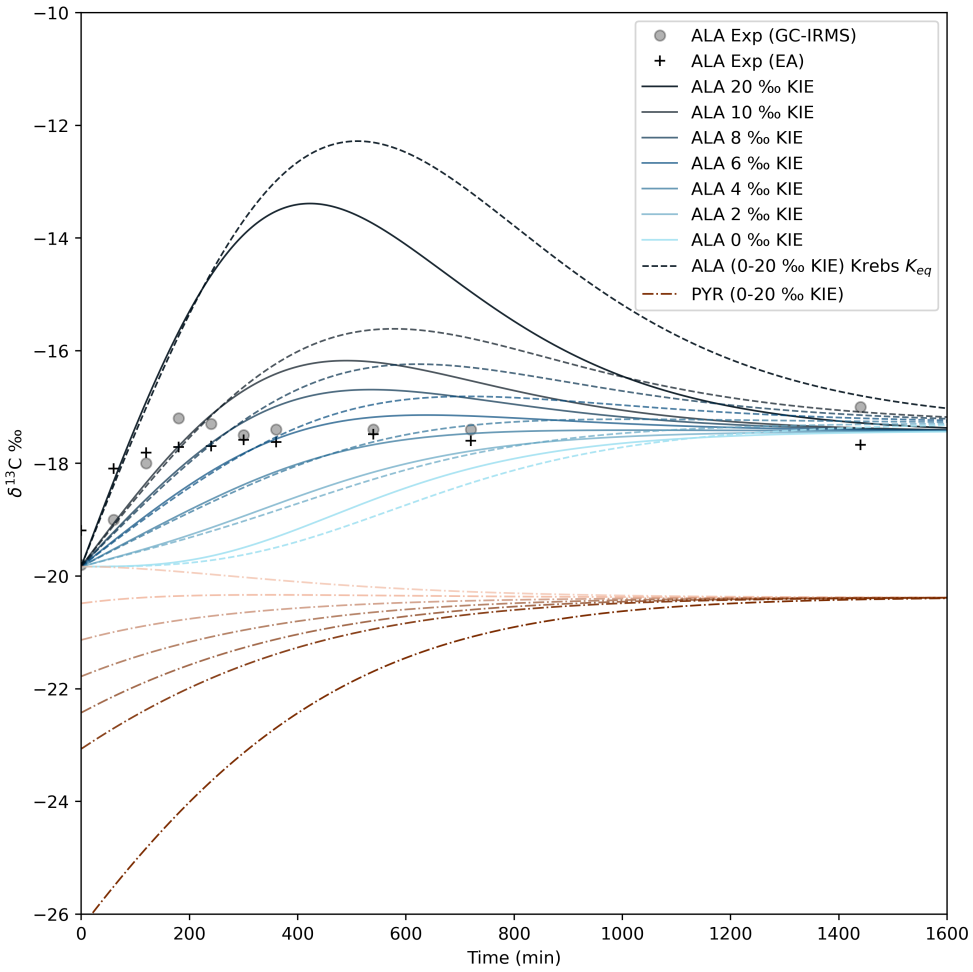

Figure 3.4: Predicted whole-molecule carbon isotopic composition of alanine (solid blue lines) and pyruvate (brown dash-dot lines) as a function of time, with the kinetic parameters used in Figure 3.2, but with a 10 ‰ equilibrium fractionation (alanine enriched in ^{13}C) imposed. Gray circles and black crosses are the GC-IRMS and EA/IRMS data, respectively. Dashed blue lines for alanine are calculated from model constrained to Krebs' chemical K_{eq}.

tion put into the model. The observed behavior is more complex, crossing a range of isokinetic fractionation curves (with different KIEs) .

3.3 Discussion

Comparing these experimental results with our theoretical predictions in Chapter 2, the obvious question is what could account for the discrepancy not only in degree of fractionation, but, more strikingly, the direction of fractionation. That is, we observed a ^{13}C isotope enrichment in alanine over the course of this

reaction, in contrast to our prediction that the reaction, once it had reached equilibrium, would result in alanine depleted in ^{13}C relative to its starting composition. It is sensible to evaluate first whether an experimental methodology issue could be lurking in such unexpected results. Separation procedures such as those that we use in this study (i.e., ion exchange column to purify alanine and glutamate from the other reaction components, and HPLC to separate alanine from glutamate) can introduce fractionation if analyte recovery is not quantitative (e.g., Caimi et al., 1997). We control for this in part by analyzing using two independent methods: EA/IRMS, which requires HPLC, and GC-IRMS, which does not require HPLC as a purification step prior to analysis. If HPLC introduces a fractionation, it will be observable in the difference between the results of the two analytical methods. While there is some scatter between these two analytical methods, it is visibly insignificant with respect to our conclusions. In addition, any fractionation introduced by our separation methods is unlikely to result in the time-dependent trend and regularity our data displays.

From our chemical kinetics data, it is easy to see that our reaction would have reached chemical equilibrium by the time the last aliquot was collected, but can we be certain that the reaction had also reached isotopic equilibrium? A previous study by Rishavy and Cleland investigated a set of enzyme-mediated carbon equilibrium isotope fractionation measurements, with the alanine/pyruvate fractionation among them (Rishavy et al., 1999). They reported a site-specific value of 12 ‰, with ^{13}C preferring pyruvate, which is similar to what we predicted theoretically in the previous chapter. Their study employed a chemical degradation method involving decomposition of pyruvate into acetate, avoiding detection and measurement of alanine altogether. Unfortunately, there is not enough detail reported in Rishavy and Cleland's paper to replicate their work. They do not provide any time-dependent chemical or isotopic equilibration data, such as that which we give in Figures 3.2 and 3.3 here. There are no error bars included in their data, nor discussion of potential sources of error. Rishavy and Cleland casually mention that they "usually allow 10 times longer" than the characteristic chemical equilibration time "to ensure that true isotopic equilibrium has been reached", but they do not explain this claim or give any indication where to find further discussion on the topic. We were not able to determine whether this question of decoupling between chemical and isotopic equilibration has been rigorously examined or

explained elsewhere.

In the context of the discussion above, modeling a decoupled chemical or isotopic equilibrium would require a time-dependent $K_i = k_{i_f}/k_{i_b}$, and, more specifically, knowledge of a functional form for such. At each time step, our code converts a specified amount of alanine to pyruvate; we would need a recipe to replace the simpler idea above of having a kinetic fractionation represented by the deviation of the k_{i_f} transmission factor from unity and an equilibrium fractionation expressed by the k_{i_f}/k_{i_b} ratio. While the decoupling of isotopic and chemical equilibrium is not an unheard of concept, there is little guidance from the literature in how to think about this decoupling mechanistically in a way that could be used rigorously, as in a computational algorithm implemented in our code. One idea is to postulate that there is some systematic relationship between the expressed fractionation and the reaction flux (i.e., the rate constant times the concentration), being closer to unity when the reaction flux is high and closer to the equilibrium fractionation when the flux is low. We can test this by running a simulation with k_{i_f}/k_{i_b} set to unity for various KIEs, for example, both k_i at 1.002 for a 2 ‰ KIE, 1.005 for a 5 ‰ KIE, etc. This would correspond to an "infinite" time to reach equilibrium. If such a model cannot reproduce the experimental data, then there is little point in thinking about the precise relationship between reaction flux and the expressed fractionation, at least in the context of a "fast-flux-implies-lower-fractionation" model. It is clear from the results shown in Figure 3.5 that a flux-dependent fractionation approach, at least one having zero fractionation at high flux and equilibrium fractionation at low flux, would not be able to fit the data.

The variability in specific activity for alanine transaminase isolated from various organisms as discussed earlier could in fact be an advantage in pursuing the secondary goal of understanding how the rates of chemical and isotopic equilibrium are related. For example, it is possible the chemical/isotopic coupling depends on reaction flux, with negligible fractionation at fast conversion rates and becoming closer to equilibrium fractionation at slow conversion rates. Having a suite of enzymes with a large variation in activity could be useful in exploring this issue in future work.

Given that this experiment involved equilibration of the α carbon isotopes between alanine and pyruvate, it was of interest to us to determine whether

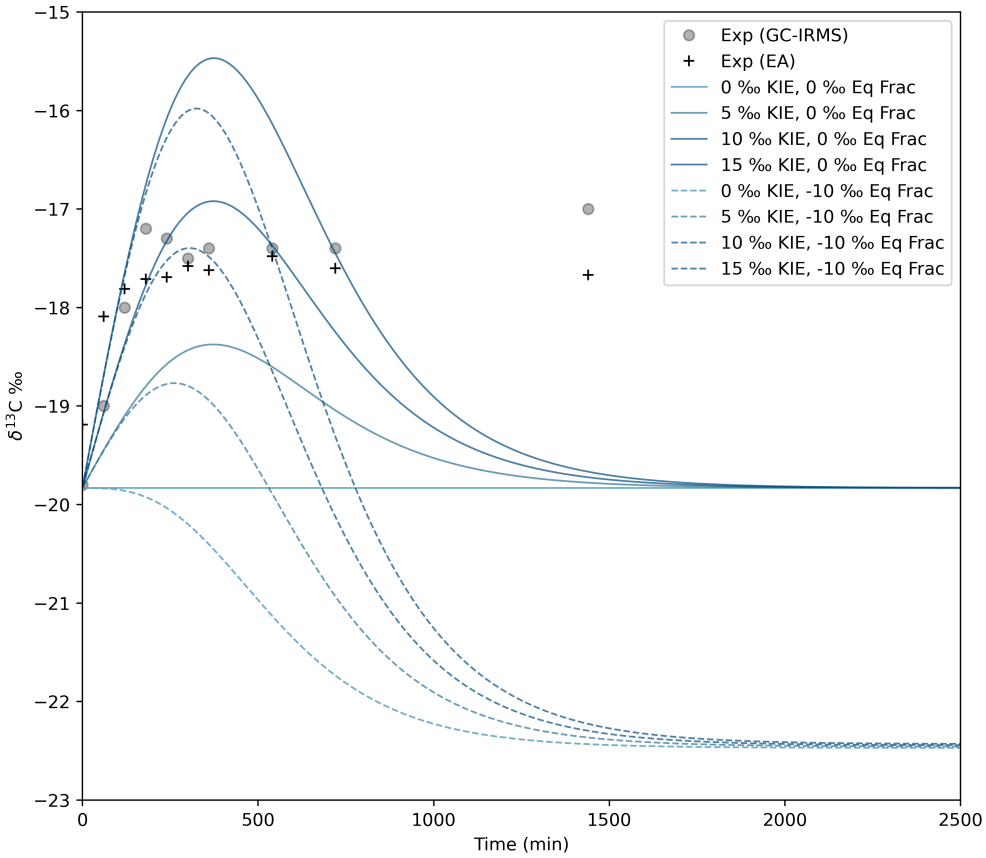

Figure 3.5: Predicted carbon isotopic composition of alanine as a function of time with the kinetic parameters used in Figure 3.2 and a 3 ‰ enrichment of ^{13}C in pyruvate (dashed lines) and with no fractionation (solid lines) for a range of kinetic isotope effects. The zero-fractionation might be appropriate for fast conversion rates. Gray circles and black crosses are the GC-IRMS and EA/IRMS data, respectively. The figure shows that even if zero fractionation persisted throughout the progress of the reaction, this could not explain the experimental data.

there was a straightforward way to measure the δ^{13}C of pyruvate along with that of alanine. If this was an option, it would provide an independent way to observe the isotopic evolution of the system, as the isotopic compositions of pyruvate and alanine evolve interrelatedly and the offset between them at equilibrium would provide an independent approach to determining their equilibrium fractionation. Unfortunately for our purposes, dealing with pyruvate in this way was not feasible due to a number of complications it would introduce. One major challenge is dealing with the complicated behavior of pyruvate in solution, whether in an effort to isolate pyruvate from the other reaction components, or to measure its δ^{13}C in a mixture. In water, pyruvic acid is known to equilibrate with pyruvate and pyruvate hydrate, and can also exist in its enol form, as well as undergoing aldol condensation to form dimers and higher polymers (see Supplementary Figure 3.6 for a ^{1}H NMR spectrum showing several species we identified in a solution of pyruvate at neutral pH). The relative abundances of these species are very sensitive to temperature and pH (Damitio et al., 1992). Quantitatively recovering these species and separating them from the other reaction components without trading off quantitative recovery of alanine is a project we leave for someone else. While others (e.g., Melzer and Schmidt 1987) have analyzed pyruvate ^{13}C by chemical degradation in the dehydrogenase reaction, those papers did not discuss the challenges of pyruvate speciation, so it is not clear whether they were aware of this complication, and how they dealt with it. In their reaction of interest, pyruvate was a reactant (rather than a product, as in our case) of dehydrogenase reactions that were run for relatively brief intervals, which we can only speculate might have limited the degree to which pyruvate was converted into its other forms. Regardless, procedures involving chemical degradation for isotopic analysis were outside the scope of our project.

Pyruvate speciation (again, see Supplementary Figure 3.6) also raises another question: could these species introduce additional carbon fractionation factors that lead to a delayed or altered equilibrium between alanine and pyruvate? Our theoretical predictions in the previous chapter included calculations for some of these species (pyruvic acid, enol-pyruvate), and the equilibrium fractionation between them and alanine was not significantly different from that of pyruvate. Intuitively, because the carbon bonding environments of these pyruvate species are so similar to each other, we would expect fractionation factors amongst these species to be small compared to the fractionation factor

between any one of them and alanine. Nevertheless we cannot exclude the possibility that an explanation for the disagreement between our theoretical predictions for alanine and our actual observations has something to do with pyruvate speciation. Similarly, it is possible that irreversible reactions occur through pyruvate dimerization and polymerization, as mentioned above, which could act as sinks for pyruvate, thus leading to a situation where steady state is never truly reached in our reaction. If so, this could also potentially account for the mismatch between our experimental results and theoretical predictions. Further investigation would be required to resolve these questions.

3.4 Conclusions

Here we present measurements of carbon isotopic (^{12}C and ^{13}C) preference between the α carbon site of alanine and the corresponding (central) carbon site in pyruvate, which differs from alanine only in that it is double-bonded at this carbon site to an oxygen atom, while alanine is bound to a proton and an amine group. The alanine transamination reaction, an enzyme-mediated biochemical reaction converting between two amino acid, keto-acid pairs (simultaneously between alanine and pyruvate, and between α-ketoglutarate and glutamate) mediated by alanine transaminase (ALT), facilitates carbon isotopic equilibration between alanine and pyruvate. Using GC-IRMS and elemental analysis (EA), we measured the δ^{13}C of alanine recovered from regular intervals along this reaction and compare the measured value to the original isotopic composition. We find a clear discrepancy between our experimental measurements and theoretical estimates (and chemical intuition) that predict pyruvate, with its stiff carbonyl bond, should be enriched in ^{13}C. Instead of observing that the isotopic composition of alanine becomes more depleted in ^{13}C as equilibrium is approached under conditions where it is driven down to $\sim$20 % of its initial concentration, we find it becomes more enriched in ^{13}C with time. Our model shows that the observed time dependence isotopic composition of the alanine does not correspond to any simple combination of kinetic and equilibrium fractionation. One possible way to explain this discrepancy would be that the timescale for isotopic equilibration was much longer than the chemical equilibration. To evaluate this possibility, we considered a model involving a flux-dependent fractionation where greater fluxes implied a lower degree of fractionation, but even a model with zero kinetic fractionation across the range of fluxes calculated in the model was unable to account for the observed time

dependence of the alanine isotopic composition, implying that the thermodynamic fractionation factor itself would actually have to be reversed.

References

Abelson, P H and T C Hoering (1961). "Carbon Isotope Fractionation in Formation of Amino Acids by Photosynthetic Organisms". In: *Proceedings of the National Academy of Sciences of the United States of America* 47.5, pp. 623–632.

Bulos, Bernard and Philip Handler (1965). "Kinetics of Beef Heart Glutamic-Alanine Transaminase". In: *Journal of Biological Chemistry* 240.8, pp. 3283–3294. ISSN: 0021-9258. DOI: 10.1016/s0021-9258(18)97216-9.

Caimi, Richard J and J Thomas Brenna (1997). "Quantitative evaluation of carbon isotopic fractionation during reversed-phase high-performance liquid chromatography". In: *Journal of Chromatography A* 757.1–2, pp. 307–310. ISSN: 0021-9673. DOI: 10.1016/s0021-9673(96)00694-2.

Damitio, J, G Smith, J E Meany, and Y Pocker (1992). "A comparative study of the enolization of pyruvate and the reversible dehydration of pyruvate hydrate". In: *Journal of the American Chemical Society* 114.8, pp. 3081–3087.

Darling, S (1947). "Computation of the Free Energy of α-Ketoglutarate and Pyruvate from Constants of the Transamination Process". In: *Nature* 160.4064, pp. 403–404.

DeRosa, G and Robert W Swick (1975). "Metabolic implications of the distribution of the alanine aminotransferase isoenzymes." In: *Journal of Biological Chemistry* 250.20, pp. 7961–7967.

Duff, Stephen M G et al. (2012). "The Enzymology of alanine aminotransferase (AlaAT) isoforms from Hordeum vulgare and other organisms, and the HvAlaAT crystal structure". In: *Archives of Biochemistry and Biophysics* 528.1, pp. 90–101. DOI: https://doi.org/10.1016/j.abb.2012.06.006.

Good, Allen G and Douglas G Muench (1992). "Purification and characterization of an anaerobically induced alanine aminotransferase from barley roots". In: *Plant Physiology* 99.4, p. 1520.

Hopper, Sarah and Harold L. Segal (1962). "Kinetic Studies of Rat Liver Glutamic-Alanine Transaminase". In: *Journal of Biological Chemistry* 237.10, pp. 3189–3195. ISSN: 0021-9258. DOI: 10.1016/s0021-9258(18)50142-3.

Krebs, H A (1953). "Equilibria in transamination systems." In: *Biochem. J* 54.1, pp. 82–86.

Lénárd, P and F B Straub (1942). *Aminopherase*. Studies Inst. Med. Chem. Univ. Szeged.

McAllister, Chandra H, Michelle Facette, Andrew Holt, and Allen G Good (2013). "Analysis of the enzymatic properties of a broad family of alanine aminotransferases". In: *PloS one* 8.2, e55032.

Nelder, John A and R Mead (1965). "A Simplex Method for Function Minimization." In: *Comput. J.*

Rishavy, Mark A and WW Cleland (1999). "^{13}C, ^{15}N, and ^{18}O equilibrium isotope effects and fractionation factors1". In: *Canadian Journal of Chemistry* 77.5-6, pp. 967–977.

Saier Jr., Milton H and W Terry Jenkins (1967). "Alanine aminotransferase: I. Purification and properties". In: *Journal of Biological Chemistry* 242.1, pp. 91–100.

Schomburg, Dietmar, Ida Scomburg, and Antje Chang (2007). *Springer Handbook of Enzymes: Class 2• Transferases X EC 2.7. 1.113–2.7. 5.7*. Springer.

Segal, Harold L, Diana S Beattie, and Sarah Hopper (1962). "Purification and properties of liver glutamic-alanine transaminase from normal and corticoid-treated rats". In: *Journal of Biological Chemistry* 237.6, pp. 1914–1920.

3.5 Supplementary Materials

3.5.1 Supplementary Figures

To deconvolve the pyruvate-related peaks, we used NMR to investigate the molecular species present in a sample of pyruvic acid purchased from Aldrich Chemistry (product # 107360-100G, Lot # SHBH2749V). In water, pyruvic acid is known to exist in equilibrium with pyruvate and pyruvate hydrate, and can also exist in its enol form, while the abundances of these species will depend largely on pH (Damitio et al., 1992). Pyruvate can also undergo aldol condensation to form dimers. We found that in our reaction mixture, dimers were not abundant enough to detect easily with ^{13}C NMR, indicating that they make up <1% of the total pyruvate species. However, in an aqueous solution of pyruvate prepared from pyruvic acid, three dimers were found to be abundant. This solution, prepared in a 2 mL GC vial, was composed of 10 μL pyruvic acid in 400 μL 0.4 M phosphate buffer at pH 7.2, to which was added 200 μL D_2O for the lock frequency. Before transferring to an NMR tube, the pH of this solution was measured with thin strips of pH paper and verified to be 6–7.

Figure 3.6 shows the methyl region of the ^{1}H NMR spectrum taken on the pyruvate solution described above. Annotations indicate the methyl protons from which each peak arises. A ^{13}C spectrum and two 2D spectra (HSQC and HMBC), provided in the Supplementary Attachments, were also obtained to aid in identifying the species present in this sample. Five distinct major species were identified, including pyruvate (most abundant), pyruvate hydrate, and three condensation products. Surprisingly, no enol monomer was found present in this solution, but an enol dimer was present. Integrals were drawn on each of these five species and are shown relative to pyruvate (100%). The four less abundant species range in relative abundance from ∼10–20%.

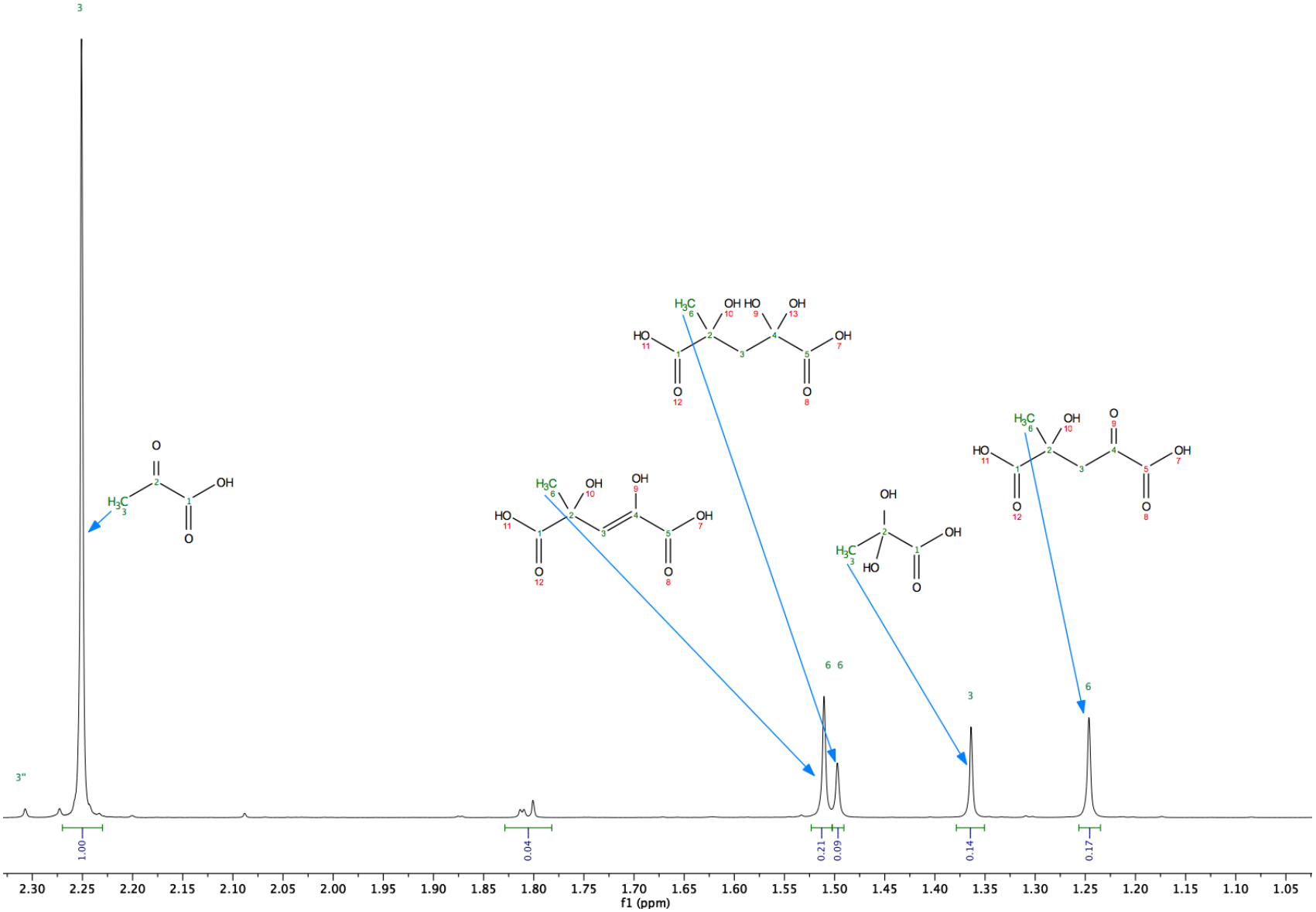

Figure 3.6: ^{1}H NMR spectrum showing signals arising from methyl groups on at least five variants of pyruvate, hydrated pyruvic acid, and related dimers that naturally arise around neutral pH. The relative abundances of these species is very sensitive to pH.

3.5.2 Python Code for Chemical Kinetics Model

```
import numpy as np
from numpy import array
import pandas as pd
import matplotlib.pyplot as plt

pd.set_option('display.max_rows', None)

df1=pd.read_csv('chem.dat',names=['time','A'],sep="\s+")

# 2-parameter model
maxiter=10001 # chemical rate constant in forward direction in
    units of mMol-1 min-1
kf=9.89E-05 # chemical rate constant in backward direction in
    units of mMol-1 min-1
kb=2.32*kf
dt=0.25000 # time step in minutes
a=4.45 # starting alanine concentration in mMol
b=40.45 # starting aKG concentration in mMol
c=0.000010 # starting pyruvate concentration in mMol
d=0.000010 # starting glutamic acid concentration in mMol
t=0.00 # starting time
plist = []
for i in range(1,maxiter):
    daf=kf*a*b*dt
    dab=kb*c*d*dt
    da=-daf+dab
    dc=-da
    db=da
    dd=dc
    a=a+da
    b=b+db
    c=c+dc
    d=d+dd
    t=t+dt
    ratio=(c*d)/(a*b)
    plist.append(array([t,a,b,c,d,ratio]))
df2=pd.DataFrame(plist,columns=['time','A','B','C','D','K'])

# 1-parameter model using Krebs' Keq
kf=9.40E-05 # chemical rate constant in forward direction in
    units of mMol-1 min-1
kb=1.52*kf # chemical rate constant in backward direction in
```

```
   units of mMol-1 min-1
dt=0.25000 # time step in minutes
a=4.45 # starting alanine concentration in mMol
b=40.45 # starting aKG concentration in mMol
c=0.000010 # starting pyruvate concentration in mMol
d=0.000010 # starting glutamic acid concentration in mMol
t=0.00 # starting time
plist = []
for i in range(1,maxiter):
    daf=kf*a*b*dt
    dab=kb*c*d*dt
    da=-daf+dab
    dc=-da
    db=da
    dd=dc
    a=a+da
    b=b+db
    c=c+dc
    d=d+dd
    t=t+dt
    ratio=(c*d)/(a*b)
    plist.append(array([t,a,b,c,d,ratio]))
df3=pd.DataFrame(plist,columns=['time','A','B','C','D','K'])

fig1, ax=plt.subplots(figsize=(8,6))
ax.plot(df1['time'],df1['A'],'.',color='k',alpha=0.3,label=r'
   ALA Exp')
ax.plot(df2['time'],df2['A'],'-',color='#01497c',lw=1,alpha
   =0.7,label=r'ALA 2-P Fit')
ax.plot(df3['time'],df3['A'],'--',color='#01497c',lw=1,alpha
   =0.7,label=r'ALA Krebs $K_{eq}$')
ax.plot(df2['time'],df2['C'],'-',color='#7c2c01',lw=1,alpha
   =0.7,label=r'PYR 2-P Fit')
ax.plot(df3['time'],df3['C'],'--',color='#7c2c01',lw=1,alpha
   =0.7,label=r'PYR Krebs $K_{eq}$')
ax.set_xlabel('Time (min)')
ax.set_ylabel('Concentration (mM)')
ax.set_xlim(0,2000)
ax.set_ylim(0,4.5)
ax.legend()
fig1.tight_layout()
fig1.savefig('Ch3_Chem-Model.png',dpi=400,transparent=False)
```

Listing 3.1: `Ch3_Chem-Model.py`

3.5.3 Python Code to Fit Experimental Data

```
import numpy as np
from numpy import array
import pandas as pd
import matplotlib.pyplot as plt

pd.set_option('display.max_rows', None)

maxiter=10001
peedee=0.011180 # fraction of 13C in PDB standard
alaf=0.0108395 # starting fraction of 13C in alanine (d13C =
    -19.83 \permil)
kf=9.89E-05 # chemical rate constant in forward direction in
    units of mMol-1 min-1
kb=2.32*kf # chemical rate constant in backward direction in
    units of mMol-1 min-1
dt=0.25000 # time step in minutes

df_isotope1 = pd.read_csv('davis_GCIRMS.dat',names=['time','
    d13C_ala'])
df_isotope2 = pd.read_csv('davis_EA.dat',names=['time','
    d13C_ala'])

factorkf=1.000 # isotope factor in forward direction
factorkb=1.010 # isotope factor in backward direction
a=4.45 # starting alanine concentration in mMol
b=40.45 # starting aKG concentration in mMol
c=0.000010 # starting pyruvate concentration in mMol
d=0.000010 # starting glutamic acid concentration in mMol

a13=a*alaf # alanine 13C concentration in mMol
a12=a*(1-alaf) # alanine 12C concentration in mMol
c13=c*alaf # pyruvate 13C concentration in mMol
c12=c*(1-alaf) # pyruvate 12C concentration in mMol

t=0.00 # starting time
plist = []
for i in range(1,maxiter):
    daf=kf*a*b*dt
    dab=kb*c*d*dt
    da=-daf+dab
    dc=-da

```

```
38      daf13=daf*a13/(a12+a13)
39      daf12=daf*a12/(a12+a13)
40      rf=daf12/daf13
41      epsilon=(factorkf*rf*daf13-daf12)/(factorkf*rf+1.0)
42      daf13=daf13-epsilon
43      daf12=daf12+epsilon
44
45  # same thing for pyruvate (c) backwards
46      dab13=dab*c13/(c12+c13)
47      dab12=dab*c12/(c12+c13)
48      rb=dab12/dab13
49      epsilon=(factorkb*rb*dab13-dab12)/(factorkb*rb+1.0)
50      dab13=dab13-epsilon
51      dab12=dab12+epsilon
52
53      da13=-daf13+dab13
54      da12=-daf12+dab12
55      dc13=-da13
56      dc12=-da12
57
58      db=da
59      dd=dc
60
61      a=a+da
62      b=b+db
63      c=c+dc
64      d=d+dd
65      a13=a13+da13
66      a12=a12+da12
67      c13=c13+dc13
68      c12=c12+dc12
69      t=t+dt
70      site_1_3_a13=2*a*0.01083760
71      site_1_3_a12=2*a*(1-0.01083760)
72      site_1_3_c13=2*c*0.01083760
73      site_1_3_c12=2*c*(1-0.01083760)
74      delta13Ca=(((a13+site_1_3_a13)/(a12+site_1_3_a12))/peedee
        -1.0)*1000
75      delta13Cc=(((c13+site_1_3_c13)/(c12+site_1_3_c12))/peedee
        -1.0)*1000
76      ratio=(c*d)/(a*b)
77      plist.append(array([i,t,a,delta13Ca,c,delta13Cc,ratio]))
78
```

```
79 df1=pd.DataFrame(plist,columns=['iter','time','alanine','
       d13C_ala','pyruvate','d13C_pyr','cd/ab'])
80
81
82 factorkf=1.010
83 factorkb=1.020
84 a=4.45
85 b=40.45
86 c=0.000010
87 d=0.000010
88 a13=a*0.01083760
89 a12=a*(1-0.01083760)
90 c13=c*0.01083760
91 c12=c*(1-0.01083760)
92 t=0.00
93 plist = []
94 for i in range(1,maxiter):
95     daf=kf*a*b*dt
96     dab=kb*c*d*dt
97     da=-daf+dab
98     dc=-da
99     daf13=daf*a13/(a12+a13)
100    daf12=daf*a12/(a12+a13)
101    rf=daf12/daf13
102    epsilon=(factorkf*rf*daf13-daf12)/(factorkf*rf+1.0)
103    daf13=daf13-epsilon
104    daf12=daf12+epsilon
105    dab13=dab*c13/(c12+c13)
106    dab12=dab*c12/(c12+c13)
107    rb=dab12/dab13
108    epsilon=(factorkb*rb*dab13-dab12)/(factorkb*rb+1.0)
109    dab13=dab13-epsilon
110    dab12=dab12+epsilon
111    da13=-daf13+dab13
112    da12=-daf12+dab12
113    dc13=-da13
114    dc12=-da12
115    db=da
116    dd=dc
117    a=a+da
118    b=b+db
119    c=c+dc
120    d=d+dd
```

```
121     a13=a13+da13
122     a12=a12+da12
123     c13=c13+dc13
124     c12=c12+dc12
125     t=t+dt
126     site_1_3_a13=2*a*0.01083760
127     site_1_3_a12=2*a*(1-0.01083760)
128     site_1_3_c13=2*c*0.01083760
129     site_1_3_c12=2*c*(1-0.01083760)
130     delta13Ca=(((a13+site_1_3_a13)/(a12+site_1_3_a12))/peedee
    -1.0)*1000
131     delta13Cc=(((c13+site_1_3_c13)/(c12+site_1_3_c12))/peedee
    -1.0)*1000
132     ratio=(c*d)/(a*b)
133     plist.append(array([i,t,a,delta13Ca,c,delta13Cc,ratio]))
134 
135 df2=pd.DataFrame(plist,columns=['iter','time','alanine','
    d13C_ala','pyruvate','d13C_pyr','cd/ab'])
136 
137 factorkf=1.020
138 factorkb=1.030
139 a=4.45
140 b=40.45
141 c=0.000010
142 d=0.000010
143 a13=a*0.01083760
144 a12=a*(1-0.01083760)
145 c13=c*0.01083760
146 c12=c*(1-0.01083760)
147 t=0.00
148 plist = []
149 for i in range(1,maxiter):
150     daf=kf*a*b*dt
151     dab=kb*c*d*dt
152     da=-daf+dab
153     dc=-da
154     daf13=daf*a13/(a12+a13)
155     daf12=daf*a12/(a12+a13)
156     rf=daf12/daf13
157     epsilon=(factorkf*rf*daf13-daf12)/(factorkf*rf+1.0)
158     daf13=daf13-epsilon
159     daf12=daf12+epsilon
160     dab13=dab*c13/(c12+c13)
```

```
161     dab12=dab*c12/(c12+c13)
162     rb=dab12/dab13
163     epsilon=(factorkb*rb*dab13-dab12)/(factorkb*rb+1.0)
164     dab13=dab13-epsilon
165     dab12=dab12+epsilon
166     da13=-daf13+dab13
167     da12=-daf12+dab12
168     dc13=-da13
169     dc12=-da12
170     db=da
171     dd=dc
172     a=a+da
173     b=b+db
174     c=c+dc
175     d=d+dd
176     a13=a13+da13
177     a12=a12+da12
178     c13=c13+dc13
179     c12=c12+dc12
180     t=t+dt
181     site_1_3_a13=2*a*0.01083760
182     site_1_3_a12=2*a*(1-0.01083760)
183     site_1_3_c13=2*c*0.01083760
184     site_1_3_c12=2*c*(1-0.01083760)
185     delta13Ca=(((a13+site_1_3_a13)/(a12+site_1_3_a12))/peedee
    -1.0)*1000
186     delta13Cc=(((c13+site_1_3_c13)/(c12+site_1_3_c12))/peedee
    -1.0)*1000
187     ratio=(c*d)/(a*b)
188     plist.append(array([i,t,a,delta13Ca,c,delta13Cc,ratio]))
189
190 df3=pd.DataFrame(plist,columns=['iter','time','alanine','
    d13C_ala','pyruvate','d13C_pyr','cd/ab'])
191
192 factorkf=1.030
193 factorkb=1.040
194 # this is not quite right, but close enough (i.e 1.018/1.015 is
     not exactly 1.003)
195 a=4.45
196 b=40.45
197 c=0.000010
198 d=0.000010
199 a13=a*0.01083760
```

```
a12=a*(1-0.01083760)
c13=c*0.01083760
c12=c*(1-0.01083760)
t=0.00
plist = []
for i in range(1,maxiter):
    daf=kf*a*b*dt
    dab=kb*c*d*dt
    da=-daf+dab
    dc=-da
    daf13=daf*a13/(a12+a13)
    daf12=daf*a12/(a12+a13)
    rf=daf12/daf13
    epsilon=(factorkf*rf*daf13-daf12)/(factorkf*rf+1.0)
    daf13=daf13-epsilon
    daf12=daf12+epsilon
    dab13=dab*c13/(c12+c13)
    dab12=dab*c12/(c12+c13)
    rb=dab12/dab13
    epsilon=(factorkb*rb*dab13-dab12)/(factorkb*rb+1.0)
    dab13=dab13-epsilon
    dab12=dab12+epsilon
    da13=-daf13+dab13
    da12=-daf12+dab12
    dc13=-da13
    dc12=-da12
    db=da
    dd=dc
    a=a+da
    b=b+db
    c=c+dc
    d=d+dd
    a13=a13+da13
    a12=a12+da12
    c13=c13+dc13
    c12=c12+dc12
    t=t+dt
    site_1_3_a13=2*a*0.01083760
    site_1_3_a12=2*a*(1-0.01083760)
    site_1_3_c13=2*c*0.01083760
    site_1_3_c12=2*c*(1-0.01083760)
    delta13Ca=(((a13+site_1_3_a13)/(a12+site_1_3_a12))/peedee
   -1.0)*1000
```

```
242     delta13Cc=(((c13+site_1_3_c13)/(c12+site_1_3_c12))/peedee
        -1.0)*1000
243     ratio=(c*d)/(a*b)
244     plist.append(array([i,t,a,delta13Ca,c,delta13Cc,ratio]))
245 
246 df4=pd.DataFrame(plist,columns=['iter','time','alanine','
        d13C_ala','pyruvate','d13C_pyr','cd/ab'])
247 
248 kf=9.4E-05  # these are units of mmol-1 min-1
249 kb=1.52*kf
250 
251 factorkf=1.000
252 factorkb=1.010
253 
254 a=4.45
255 b=40.45
256 c=0.000010
257 d=0.000010
258 
259 a13=a*0.01083760
260 a12=a*(1-0.01083760)
261 c13=c*0.01083760
262 c12=c*(1-0.01083760)
263 
264 t=0.00
265 plist = []
266 for i in range(1,maxiter):
267     daf=kf*a*b*dt
268     dab=kb*c*d*dt
269     da=-daf+dab
270     dc=-da
271 
272     daf13=daf*a13/(a12+a13)
273     daf12=daf*a12/(a12+a13)
274     rf=daf12/daf13
275     epsilon=(factorkf*rf*daf13-daf12)/(factorkf*rf+1.0)
276     daf13=daf13-epsilon
277     daf12=daf12+epsilon
278 
279 # same thing for pyruvate (c) backwards
280     dab13=dab*c13/(c12+c13)
281     dab12=dab*c12/(c12+c13)
282     rb=dab12/dab13
```

```
283     epsilon=(factorkb*rb*dab13-dab12)/(factorkb*rb+1.0)
284     dab13=dab13-epsilon
285     dab12=dab12+epsilon
286 
287     da13=-daf13+dab13
288     da12=-daf12+dab12
289     dc13=-da13
290     dc12=-da12
291 
292     db=da
293     dd=dc
294 
295     a=a+da
296     b=b+db
297     c=c+dc
298     d=d+dd
299     a13=a13+da13
300     a12=a12+da12
301     c13=c13+dc13
302     c12=c12+dc12
303     t=t+dt
304     site_1_3_a13=2*a*0.01083760
305     site_1_3_a12=2*a*(1-0.01083760)
306     site_1_3_c13=2*c*0.01083760
307     site_1_3_c12=2*c*(1-0.01083760)
308     delta13Ca=(((a13+site_1_3_a13)/(a12+site_1_3_a12))/peedee
    -1.0)*1000
309     delta13Cc=(((c13+site_1_3_c13)/(c12+site_1_3_c12))/peedee
    -1.0)*1000
310     ratio=(c*d)/(a*b)
311     plist.append(array([i,t,a,delta13Ca,c,delta13Cc,ratio]))
312 
313 df5=pd.DataFrame(plist,columns=['iter','time','alanine','
    d13C_ala','pyruvate','d13C_pyr','cd/ab'])
314 
315 
316 factorkf=1.010
317 factorkb=1.020
318 a=4.45
319 b=40.45
320 c=0.000010
321 d=0.000010
322 a13=a*0.01083760
```

```
a12=a*(1-0.01083760)
c13=c*0.01083760
c12=c*(1-0.01083760)
t=0.00
plist = []
for i in range(1,maxiter):
    daf=kf*a*b*dt
    dab=kb*c*d*dt
    da=-daf+dab
    dc=-da
    daf13=daf*a13/(a12+a13)
    daf12=daf*a12/(a12+a13)
    rf=daf12/daf13
    epsilon=(factorkf*rf*daf13-daf12)/(factorkf*rf+1.0)
    daf13=daf13-epsilon
    daf12=daf12+epsilon
    dab13=dab*c13/(c12+c13)
    dab12=dab*c12/(c12+c13)
    rb=dab12/dab13
    epsilon=(factorkb*rb*dab13-dab12)/(factorkb*rb+1.0)
    dab13=dab13-epsilon
    dab12=dab12+epsilon
    da13=-daf13+dab13
    da12=-daf12+dab12
    dc13=-da13
    dc12=-da12
    db=da
    dd=dc
    a=a+da
    b=b+db
    c=c+dc
    d=d+dd
    a13=a13+da13
    a12=a12+da12
    c13=c13+dc13
    c12=c12+dc12
    t=t+dt
    site_1_3_a13=2*a*0.01083760
    site_1_3_a12=2*a*(1-0.01083760)
    site_1_3_c13=2*c*0.01083760
    site_1_3_c12=2*c*(1-0.01083760)
    delta13Ca=(((a13+site_1_3_a13)/(a12+site_1_3_a12))/peedee
   -1.0)*1000
```

```
365     delta13Cc=(((c13+site_1_3_c13)/(c12+site_1_3_c12))/peedee
        -1.0)*1000
366     ratio=(c*d)/(a*b)
367     plist.append(array([i,t,a,delta13Ca,c,delta13Cc,ratio]))
368 
369 df6=pd.DataFrame(plist,columns=['iter','time','alanine','
        d13C_ala','pyruvate','d13C_pyr','cd/ab'])
370 
371 factorkf=1.020
372 factorkb=1.030
373 a=4.45
374 b=40.45
375 c=0.000010
376 d=0.000010
377 a13=a*0.01083760
378 a12=a*(1-0.01083760)
379 c13=c*0.01083760
380 c12=c*(1-0.01083760)
381 t=0.00
382 plist = []
383 for i in range(1,maxiter):
384     daf=kf*a*b*dt
385     dab=kb*c*d*dt
386     da=-daf+dab
387     dc=-da
388     daf13=daf*a13/(a12+a13)
389     daf12=daf*a12/(a12+a13)
390     rf=daf12/daf13
391     epsilon=(factorkf*rf*daf13-daf12)/(factorkf*rf+1.0)
392     daf13=daf13-epsilon
393     daf12=daf12+epsilon
394     dab13=dab*c13/(c12+c13)
395     dab12=dab*c12/(c12+c13)
396     rb=dab12/dab13
397     epsilon=(factorkb*rb*dab13-dab12)/(factorkb*rb+1.0)
398     dab13=dab13-epsilon
399     dab12=dab12+epsilon
400     da13=-daf13+dab13
401     da12=-daf12+dab12
402     dc13=-da13
403     dc12=-da12
404     db=da
405     dd=dc
```

```
406     a=a+da
407     b=b+db
408     c=c+dc
409     d=d+dd
410     a13=a13+da13
411     a12=a12+da12
412     c13=c13+dc13
413     c12=c12+dc12
414     t=t+dt
415     site_1_3_a13=2*a*0.01083760
416     site_1_3_a12=2*a*(1-0.01083760)
417     site_1_3_c13=2*c*0.01083760
418     site_1_3_c12=2*c*(1-0.01083760)
419     delta13Ca=(((a13+site_1_3_a13)/(a12+site_1_3_a12))/peedee
    -1.0)*1000
420     delta13Cc=(((c13+site_1_3_c13)/(c12+site_1_3_c12))/peedee
    -1.0)*1000
421     ratio=(c*d)/(a*b)
422     plist.append(array([i,t,a,delta13Ca,c,delta13Cc,ratio]))
423 
424 df7=pd.DataFrame(plist,columns=['iter','time','alanine','
    d13C_ala','pyruvate','d13C_pyr','cd/ab'])
425 
426 factorkf=1.030
427 factorkb=1.040
428 # this is not quite right, but close enough (i.e 1.018/1.015 is
     not exactly 1.003)
429 a=4.45
430 b=40.45
431 c=0.000010
432 d=0.000010
433 a13=a*0.01083760
434 a12=a*(1-0.01083760)
435 c13=c*0.01083760
436 c12=c*(1-0.01083760)
437 t=0.00
438 plist = []
439 for i in range(1,maxiter):
440     daf=kf*a*b*dt
441     dab=kb*c*d*dt
442     da=-daf+dab
443     dc=-da
444     daf13=daf*a13/(a12+a13)
```

```
445     daf12=daf*a12/(a12+a13)
446     rf=daf12/daf13
447     epsilon=(factorkf*rf*daf13-daf12)/(factorkf*rf+1.0)
448     daf13=daf13-epsilon
449     daf12=daf12+epsilon
450     dab13=dab*c13/(c12+c13)
451     dab12=dab*c12/(c12+c13)
452     rb=dab12/dab13
453     epsilon=(factorkb*rb*dab13-dab12)/(factorkb*rb+1.0)
454     dab13=dab13-epsilon
455     dab12=dab12+epsilon
456     da13=-daf13+dab13
457     da12=-daf12+dab12
458     dc13=-da13
459     dc12=-da12
460     db=da
461     dd=dc
462     a=a+da
463     b=b+db
464     c=c+dc
465     d=d+dd
466     a13=a13+da13
467     a12=a12+da12
468     c13=c13+dc13
469     c12=c12+dc12
470     t=t+dt
471     site_1_3_a13=2*a*0.01083760
472     site_1_3_a12=2*a*(1-0.01083760)
473     site_1_3_c13=2*c*0.01083760
474     site_1_3_c12=2*c*(1-0.01083760)
475     delta13Ca=(((a13+site_1_3_a13)/(a12+site_1_3_a12))/peedee
    -1.0)*1000
476     delta13Cc=(((c13+site_1_3_c13)/(c12+site_1_3_c12))/peedee
    -1.0)*1000
477     ratio=(c*d)/(a*b)
478     plist.append(array([i,t,a,delta13Ca,c,delta13Cc,ratio]))
479
480 df8=pd.DataFrame(plist,columns=['iter','time','alanine','
    d13C_ala','pyruvate','d13C_pyr','cd/ab'])
481
482 fig1, ax=plt.subplots(figsize=(8,7))
483 ax.plot(df_isotope1['time'],df_isotope1['d13C_ala'],'o',color='
    k',alpha=0.3,label=r'ALA Exp (GC-IRMS)')
```

```
484 ax.plot(df_isotope2['time'],df_isotope2['d13C_ala'],'+',color='
    k',label=r'ALA Exp (EA)')
485 ax.plot(df1['time'],df1['d13C_ala'],'-',color='#61a5c2',alpha
    =0.9,lw=1,label=u'ALA 0 \permil KIE')
486 ax.plot(df2['time'],df2['d13C_ala'],'-',color='#2a6f97',alpha
    =0.7,lw=1,label=u'ALA 5 \permil KIE')
487 ax.plot(df3['time'],df3['d13C_ala'],'-',color='#01497c',alpha
    =0.7,lw=1,label=u'ALA 10 \permil KIE')
488 ax.plot(df4['time'],df4['d13C_ala'],'-',color='#012c4a',alpha
    =0.7,lw=1,label=u'ALA 15 \permil KIE')
489 ax.plot(df1['time'],df1['d13C_pyr'],'-.',color='#f2baa2',alpha
    =0.9,lw=1)
490 ax.plot(df2['time'],df2['d13C_pyr'],'-.',color='#c27f61',alpha
    =0.7,lw=1)
491 ax.plot(df3['time'],df3['d13C_pyr'],'-.',color='#97502a',alpha
    =0.7,lw=1)
492 ax.plot(df4['time'],df4['d13C_pyr'],'-.',color='#7c2c01',alpha
    =0.7,lw=1,label=u'PYR')
493 ax.plot(df5['time'],df5['d13C_ala'],'--',color='#61a5c2',alpha
    =0.9,lw=1)
494 ax.plot(df6['time'],df6['d13C_ala'],'--',color='#2a6f97',alpha
    =0.7,lw=1)
495 ax.plot(df7['time'],df7['d13C_ala'],'--',color='#01497c',alpha
    =0.7,lw=1)
496 ax.plot(df8['time'],df8['d13C_ala'],'--',color='#012c4a',alpha
    =0.7,lw=1,label=u'ALA, Krebs $K_{eq}$')
497 ax.set_xlabel('Time (min)')
498 ax.set_ylabel(u'$\delta^{13}$C \permil')
499 ax.set_xlim(0,2500)
500 ax.set_ylim(-30,-10)
501 ax.legend()
502 fig1.tight_layout()
503 fig1.savefig('Ch3_Fig3.png',dpi=400,transparent=False)
504 plt.show()
```

Listing 3.2: `Ch3_Fit-Exp.py`

Chapter 4

BACKGROUND ON HYDROGEN-DEUTERIUM EXCHANGE CATALYZED BY TRANSAMINASES

In the previous chapter of this thesis, we described an investigation of carbon isotope kinetics and equilibrium during alanine transamination via alanine transaminase (ALT). This investigation involved a set of ^{1}H NMR experiments, which in turn led to an unanticipated tangential observation: while incubating this reaction with 10% $^{2}H_2O$ (i.e., D_2O, which, when added to H_2O, quickly mixes to form a solution of H_2O, HDO, and D_2O) serving as the frequency lock, a set of multiplets appeared adjacent to and slightly upfield of the alanine methyl (β) doublet. These multiplets increased with time and showed splitting in accordance with a spin=1 nucleus (a 1:1:1 triplet, in the case of the multiplet just upfield of the β doublet), indicating ^{1}H-^{2}H coupling (Figure 4.1). Their growth coincided with a decrease in the peak area of the methyl doublet. With these observations we determined that protons on the alanine β carbon were exchanging with deuterium in the solvent.

It is obvious upon inspection of the alanine transamination reaction (Figure 3.1) that removal of the hydrogen atom at the α carbon site occurs by necessity during transamination; that is, the α-C–H bond must be broken for the forward reaction to occur, because pyruvate's respective carbon site is attached by a double bond to an oxygen in place of the hydrogen and amine on alanine. When the reaction occurs in the reverse direction, an amine group from the pyridoxal phosphate (PLP) coenzyme and a hydrogen atom from the solvent water will replace the oxygen on the central pyruvate carbon atom to form alanine. Thus in going from alanine to pyruvate and back again to alanine, hydrogen exchange will have occurred between water and the α carbon of alanine. However, it is not at all clear from such an inspection that hydrogen exchange will also take place at alanine's β carbon. In standard biochemistry textbook discussions of the transaminase mechanism (e.g., Voet et al., 2016), amino acid R-groups, such as the methyl group on alanine, are not discussed as participants in transamination. Given the generality of transamination, it was a surprise to us that transaminases can indeed in some cases also catalyze hydrogen-deuterium exchange ("HDX") on amino acid β carbon sites, i.e., not

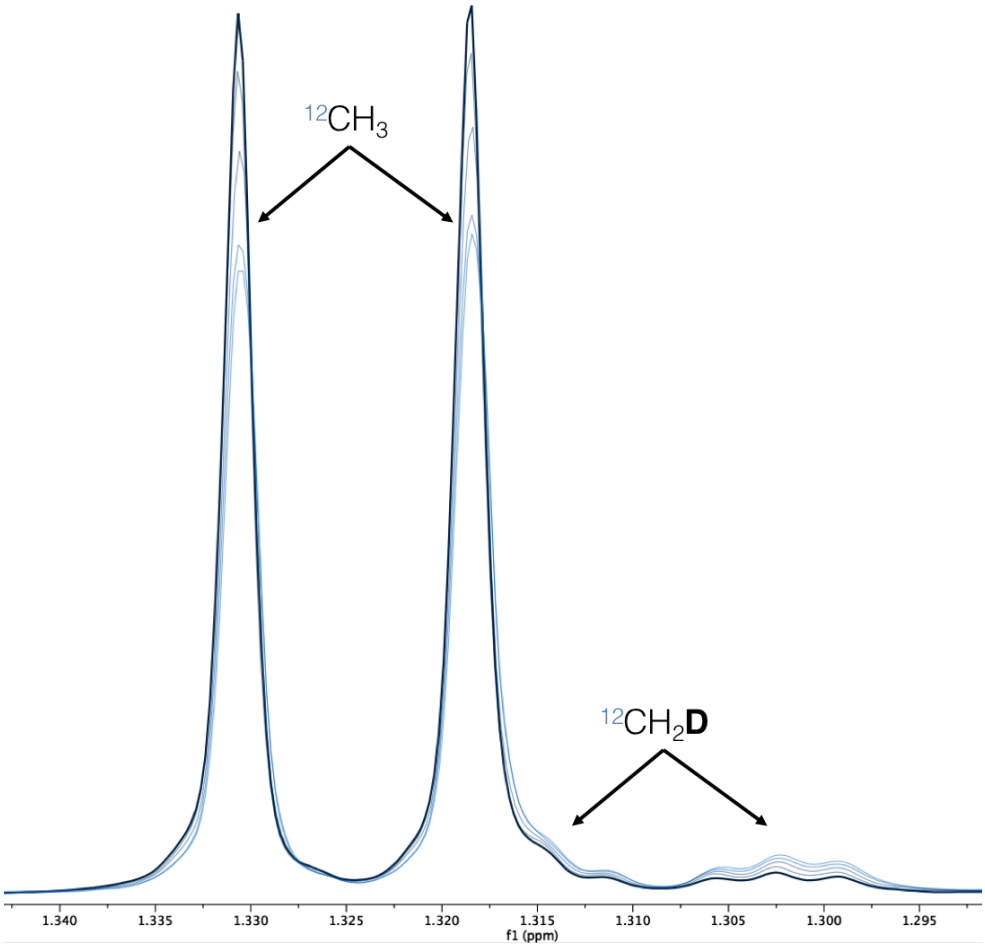

Figure 4.1: ^{1}H NMR superimposed spectra illustrating decay of the alanine $^{12}CH_3$ doublet and rise of the $^{12}CH_2D$ triplets observed during initial incubation of alanine with ALT, in 10 % D_2O. Spectra evolve forward in time, represented by color gradient from dark blue (earlier) to light blue (later).

just alanine's β carbon, but, as we will review below, on at least one other documented amino acid (glutamate) as well.

A literature search revealed that HDX on the β carbon site of alanine had been documented before, beginning with Oshima et al., 1961. And yet, after a modest collection of studies by several independent research groups in the mid-1970s (Cooper, 1976; Golichowski et al., 1977; Walter et al., 1975; Babu et al., 1974), the phenomenon has seen little interest in the decades since, and seems to have been largely unnoticed or forgotten, with a few exceptions in the enzymology community (e.g., Barb et al., 2013; Funk et al., 2019). Amino acid R-groups do not appear to be involved in any necessary step in the transamination reaction pathway, and, while some have characterized β

exchange as an "off-pathway" or "side" reaction (Barb et al., 2013), it remains an open question whether it is purely coincidental, or whether there might be some mechanistic advantage that such exchange lends to the catalysis of the subset of transamination reactions in which it occurs.

Our colleagues studying stable isotope distributions in organic molecules were, like us, unaware that this natural enzyme system facilitated an interesting isotopic exchange. It quickly became apparent that we had found a biological context, largely unexplored with modern analytical tools, in which it was possible to investigate hydrogen isotope equilibrium fractionation as well as kinetic isotope effects at multiple distinct sites in an amino acid. Like other biocatalyzed reactions, such as H_2CO_3-carbonic anhydrase, one key advantage is that the system can achieve equilibrium in low-temperature aqueous environments, here with respect to H/D exchange in otherwise inert hydrogen atoms. In addition, here it is possible to preserve a hydrogen isotopic signature into the alanine β site (or the pyruvate methyl site) that is inherited from the local water pool at the moment the molecule undergoes transamination through the ALT enzyme, but which is otherwise non-exchangeable.

After the many challenges we encountered and unresolved questions we were left with in our study of carbon isotope fractionation within this same enzyme-mediated reaction, it is hard to overstate how rewarding this opportunity turned out to be.

With the combination of α and β hydrogen exchange on alanine, it becomes possible to study the intermolecular exchange kinetics and equilibrium preference of H and D between water and these two carbon-bound hydrogen sites in alanine, as well as the intramolecular site preference of H and D within alanine, independent of water. The closest comparable work of which we are aware is the recent study from the Eiler group by Xie et al., in which the authors measured Ni- and Pd-catalyzed hydrogen isotope exchange between the two carbon positions in propane, i.e., the single central carbon versus the two identical terminal carbons. With kinetics experiments using water enriched in deuterium, they were able to deduce that exchange was occurring by at least two different mechanisms, because, all else being equal, the rates at which deuterium from the water exchanged with the two propane sites were roughly an order of magnitude faster for Pd than for Ni. Once samples reached isotopic equilibrium using these metal catalysts, they were measured to determine

their equilibrium site preference, which was found to be an approximately 80 ‰ preference for deuterium at the central carbon site at 303 K (Xie et al., 2018).

To the best of our knowledge, no analogous studies have been attempted on an organic molecule in aqueous solution. Many reactions of interest to biogeochemists occur in the liquid phase, with substrates typically dissolved in water. Of course, biological reactions occur inside cells, where solvating water molecules are abundant, and the interactions between solvent, enzyme active site, co-enzyme, and substrates create a reaction environment that is fundamentally different from that of gas phase reactions. Thus, the alanine transamination reaction provides a unique opportunity to study HDX in a solvated and metabolically central molecule at two distinct, adjacent carbon sites over time scales and chemical shifts that are easily accessible by NMR. In the next chapter, we present ^{1}H and ^{13}C NMR data that provide new insights into the kinetics and equilibrium site-preference of deuterium and hydrogen between the α and β sites in alanine.

For the remainder of this chapter we provide a literature review as background for that work. As mentioned above, this small body of literature on enzyme-catalyzed hydrogen-deuterium exchange in amino acids has been largely forgotten and is not widely appreciated, particularly in the biogeochemistry community, receiving a small number of citations in the 4+ decades since the research was performed on this system during the 1960s and 1970s. Most of these citations are from within the enzymology and medical communities, where isotopes tend to be employed in the service of some practical application, rather than taken as a target for basic research in their own right; none of these citations are from within the biogeochemistry or stable isotope geochemistry community. Thus we believe it is worth reviewing the prior work on this topic.

4.1 Literature Review

Hydrogen-deuterium exchange in alanine was first documented in 1947 by Konikova et al. (Konikova et al., 1947), who recognized that the removal of the α-hydrogen was necessary for conversion into the corresponding α-keto acid during transamination. They confirmed this expectation using a series of amino acids labeled with deuterium at the α-carbon site and measuring the

dissociation of this deuterium by densimetric estimation of excess deuterium coming into the aqueous medium as the transamination reactions progressed. It was not discovered until fourteen years later that the β hydrogen atoms in alanine and a few other amino acids were also exchangeable with water in the presence of certain transaminases.

The first observation of enzymatically catalyzed hydrogen exchange on the β carbon of an amino acid was presented by Oshima and Tamiya, who in 1961 used mass spectrometry to show the presence of 3.76 atoms of deuterium per molecule of alanine (translating to 94% of the enzyme-catalyzed exchangeable hydrogen atoms) after incubation for 38 hours with ALT at 310 K in the presence of 99.8% D_2O. After incubation, prior to analysis, alanine's three exchangeable amine deuterium atoms were allowed to re-equilibrate with H_2O by filtering out the enzyme and redissolving the alanine in a minimal amount of H_2O. Infrared spectra taken on the D_2O-plus-enzyme-equilibrated alanine, compared with spectra from fully hydrogenated alanine and selectively deuterated α-deutero-DL-alanine, confirmed HDX in both the β and the α alanine sites. Oshima and Tamiya further confirmed this result by running the same experiment in reverse, i.e., taking the deuterated alanine and converting it back into hydrogenated alanine and observing that its infrared spectrum changed in reverse, as expected. They also recovered alanine after short (tens of minutes, rather than hours) incubation periods, and concluded through integration of infrared spectra on CH and CH_3 peaks that the rates of α and β hydrogen exchange were roughly equivalent. They argued that these similar exchange rates indicated that the dissociation of a β proton *must be an essential step* in the transaminase mechanism, and therefore a Schiff base intermediate would be required, in which a β (rather than α) proton was dissociated (Oshima et al., 1961).

These experiments are primitive in comparison to what is possible today, and it is impressive that the authors were able to assess the β and α exchange rates, in spite of the fact that they did not report any evaluation of experimental precision. Their interpretation that similar exchange rates indicated β exchange as a necessary part of transamination (Figure 4.2) was perhaps reasonable based on the limited information available at the time, though it has since become widely known that transamination mechanisms are similar across all amino acids–all of them rely on common intermediates formed between a

lysine residue and the pyridoxal phosphate coenzyme (abbreviated PLP, also known as the active form of vitamin B6)–and β exchange occurs in only a subset of these. At that time, it was recognized that PLP played a co-enzymatic role in transamination, but little was known about the corresponding role of transaminase enzyme active sites stabilizing reaction intermediates. Research in subsequent decades, particularly resolving the structures of transaminases through x-ray crystallography beginning in the 1980s, has been crucial in developing a detailed mechanistic understanding of transamination (Toney, 2005).

The alanine transaminase's ability to catalyze the exchange of amino acid β hydrogen is apparently not unique to this enzyme, although the rate of alanine β-exchange catalyzed by ALT appears to be faster than amino acid β-exchange catalyzed by other transaminases. In 1969, Whelan and Long used 60 MHz ^{1}H NMR, developed earlier that decade, to show that porcine-heart glutamate-aspartate aminotransferase (now known as aspartate transaminase, or AST) catalyzed slow hydrogen exchange on the β carbon of glutamic acid, and much faster exchange at the α carbon, but ALT catalyzed exchange only at the glutamic acid α carbon (Whelan et al., 1969). Hydrogen-deuterium exchange on glutamate had been reported about 15 years earlier by (Grisolia et al., 1954) and (Hilton et al., 1954), but, having incubated their reaction for a relatively short time, only a small amount of exchange had taken place, compared to the amount of material available. Lacking any means to resolve site-specificity at that time, they assumed that exchange was taking place only at the α carbon.

In 1975, Walter et al. conducted a series of incubation experiments with either glutamic-pyruvic transaminase (a deprecated name for ALT) or glutamic-oxaloacetic transaminase (another deprecated name for AST) and either alanine, glutamate, or aspartate, all of which were labeled with tritium (^{3}H) at their β-carbon site (Walter et al., 1975). They found that hydrogen exchange occurred at the β-carbons of both alanine and glutamate in the presence of ALT, however in the case of alanine, the presence of α-ketoglutarate (the corresponding 2-oxo acid) was not required for exchange to occur, while for glutamate no exchange was observed unless pyruvate was present. They also saw exchange at the β-carbon of glutamate in the presence of AST, regardless of whether any oxaloacetate was present, but did not observe exchange at the β-carbon of aspartate under either condition. The findings of Walter et al.,

that HDX occurs at the β-carbon position of glutamate in the presence of both AST and ALT and the corresponding 2-oxo acid, appear to contradict the findings of Whelan and Long, that this is the case only for glutamate with AST, as described above. We were unable to determine whether this discrepancy was resolved in later work.

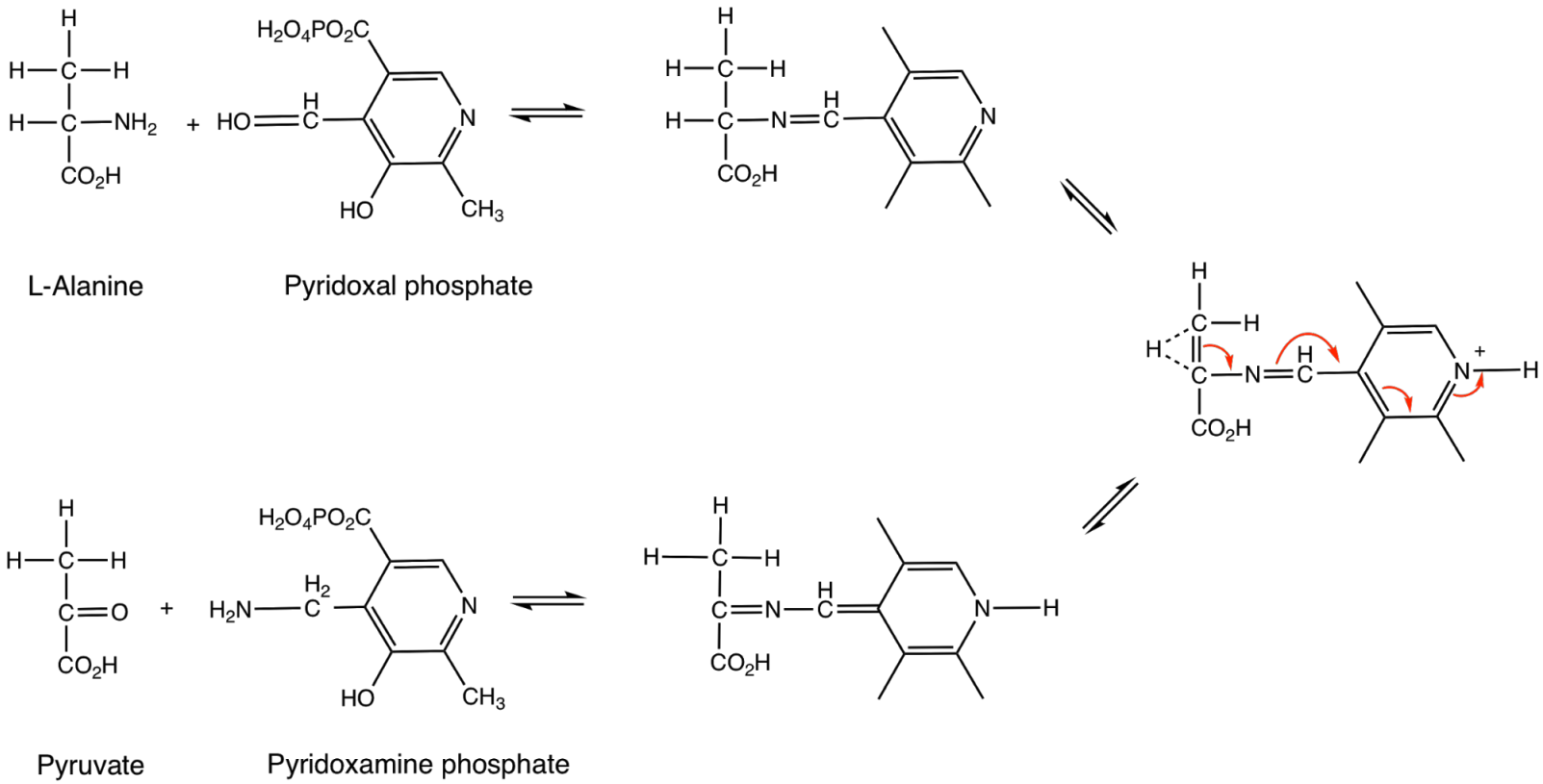

Figure 4.2: The mechanism for alanine β-exchange proposed by Oshima and Tamiya (Oshima et al., 1961), in which a Schiff base forms between the amino acid and the co-enzyme, pyridoxal phosphate, allowing dissociation of a proton primarily from the β-carbon rather than from the α-carbon. This mechanistic representation is oversimplified—it does not depict any interaction of the substrate and co-enzyme with the enzyme itself. At the time of Oshima and Tamiya's research, the role of the ALT enzyme in mediating this reaction was unknown. This mechanistic depiction is outdated, but is presented here to provide historical context.

In 1976 and 1977, two independent researchers, Cooper and Golichowski et al., respectively, used ^{1}H NMR to study the kinetics of alanine β- and α-hydrogen exchange. ^{1}H NMR made it possible to monitor the reaction in real-time, enabling a much richer study of this exchange than previously possible. Figure 4.3 shows the earliest published kinetics data for this reaction, from Cooper, who interpreted these data as showing a $\sim$3.3x faster exchange at the β position than at the α position (this is likely not accounting for multiplicity, although there is no discussion of this in the paper).

One of the hypotheses that Cooper attempted to test was the possibility of a "scrambling mechanism", in which exchange at the α and β positions are coupled—i.e., a proton or deuteron on the β carbon is removed via one mechanism and transferred rapidly to the α carbon, while a second mechanism transfers the α proton or deuteron away from the molecule. This hypothesis was tested by monitoring the kinetics of the reaction starting with L-alanine-2-d (that is, alanine with deuterium at the α position) in a solution of D_2O. In this scenario, the α deuteron is already equilibrated with the solvent, such that any resolvable proton signal that arises in the α region must come from the β position. They did not see any signal arise in the α region, and therefore concluded that a "scrambling mechanism" was not involved, at least not to an extent greater than their lower limit of detection, which was estimated to be 5% incorporation at the α position of protons coming from the β position. We replicated this experiment with the additional capability of finer time resolution and a lower limit of detection, and show that, contrary to Cooper's findings, there is a significant degree of hydrogen and deuterium mixing between the α and β positions during reaction (see Section 5.4.1.3 below for details).

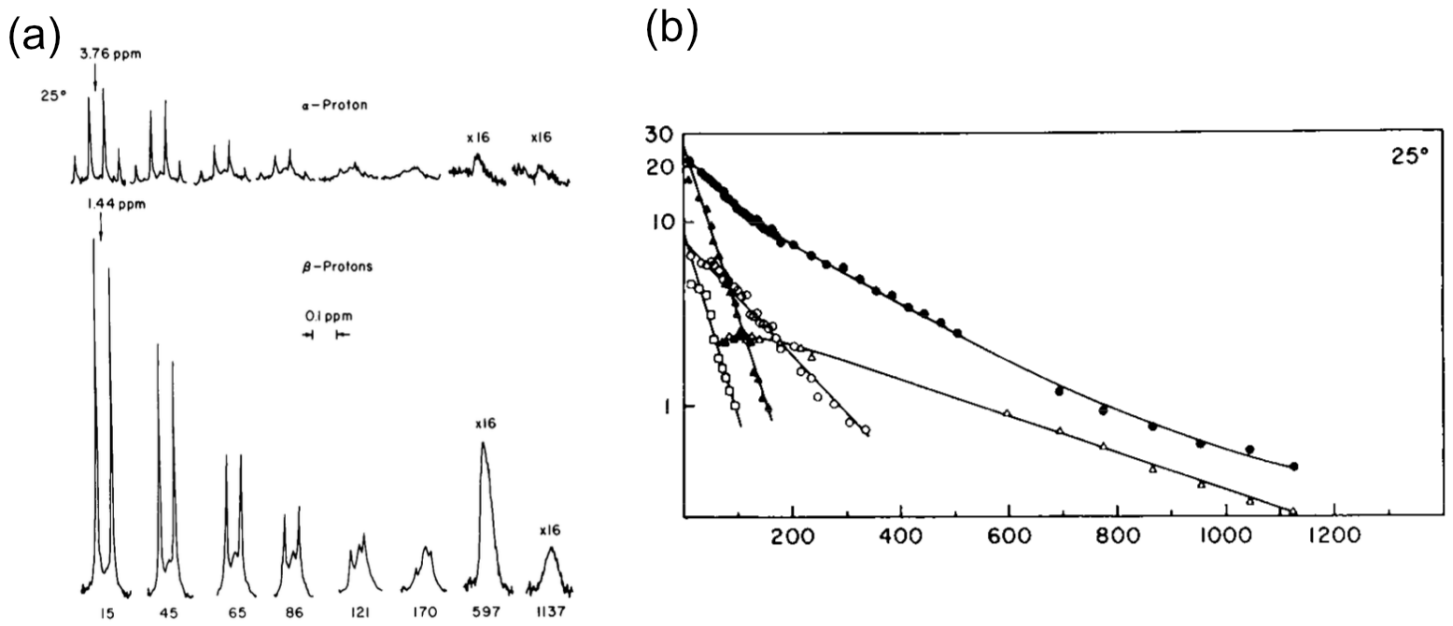

Figure 4.3: (a) ^{1}H NMR data published by (Cooper, 1976), showing exponential decay of the alanine α proton quartet (top) and β proton doublet (bottom) at 298 K. Time is given in minutes on the bottom. As the α and β hydrogen atoms are exchanged for deuterium, not only do the quartet and doublet peak areas decay, they also evolve into more complex signals due to changing proton and deuterium splitting. (b) Exponential decay of signal area from the entire β proton region (closed circles) and α proton region (open circles) using the data shown on the left. Cooper also attempted to break the multiplets into signal coming specifically from the "β singlet" (open triangle), "β doublet" (closed triangle) and "α quartet" (open square). These signals proved more difficult to deconvolve than the author recognized at the time, making the data impossible to interpret, and probably not meaningful. Cooper used these data to argue that the loss of β protons was $\sim$3.3 times faster than that of the α proton. Vertical axis is Log10 Signal Area (arbitrary units); horizontal axis is time in minutes.

Cooper also investigated the uniqueness of the alanine transaminase's ability to catalyze β HDX by measuring exchange in L-alanine, L-glutamate, and L-aspartate using another transaminase enzyme, glutamate-aspartate transaminase (AST). Alanine is able to bind to AST, but AST does not catalyze alanine-glutamate transamination. The HDX observed on the alanine β position using AST was so slow that Cooper attributed it to enolization of pyruvate, rather than enzymatic catalysis, in disagreement with previous work by Babu and Johnston (Babu et al., 1974). Cooper posited that Babu and Johnston's AST enzyme preparation likely contained contaminating ALT, which would have given rise to mis-attributed HDX. Cooper saw no significant change in glutamate β proton signal after incubation with AST in D_2O for 24 hours, while the glutamate α proton signal completely disappeared in this same period. On the other hand, aspartate incubated with AST saw complete α proton signal loss within 3 hours, with 15% β proton signal loss in the same period, and 75% loss after 24 hours.

In recent years, tracking pyruvate through metabolic pathways *in vivo* has become useful in clinical settings to investigate cancer metabolism and ALT activity in diseased tissues (e.g., Albers et al., 2008, Witney et al., 2010). Pyruvate metabolism can be tracked in tissues, and even living animals (Xu et al., 2011), with NMR over timescales of seconds by employing hyperpolarization techniques, which increase molecular ^{13}C NMR signal intensity, giving a temporary sensitivity boost of $> 10,000$ (Ardenkjaer-Larsen et al., 2003). Barb et al. used hyperpolarized $^{13}C_3D_3$-pyruvate to investigate the alanine transaminase mechanism over timescales of seconds (an experiment that is only possible via the sensitivity boost provided by hyperpolarization). They demonstrated that ALT introduced approximately 0.8 new protons into the β group of alanine for every molecule converted from perdeuterated pyruvate. Multiple alanine isotopologues appeared after just a single catalytic cycle, indicating a small probability of converting $^{13}C_3D_3$-pyruvate to $^{13}C_3H_2D$-alanine in a single encounter with the enzyme. These experiments were challenging, and precision was limited such that the formation of the isotopologues resulting from this reaction could not be quantified, however an interesting conclusion that Barb et al. report is that hydrogen isotope exchange at the β site occurs in alanine transamination not as an integral part of the transamination mechanism, but as an off-pathway reaction involving an enamine intermediate (Figure 4.4, Barb et al., 2013). In their proposed side-reaction, there is a finite

chance of hydrogen exchange occurring not just once, but twice or even three times before the product alanine is released from the enzyme active site (i.e., a single catalytic cycle will result in some distribution of the four possible alanine isotopologues, $^{13}C_3D_3$-, $^{13}C_3HD_2$-, $^{13}C_3H_2D$-, or $^{13}C_3H_3$-alanine from $^{13}C_3D_3$-pyruvate). It is not clear to us whether Barb et al. are correct in concluding from their observations that the existence of β exchange is necessarily a coincidental feature of this transamination reaction. We think it possible that there is some advantage to transamination rates conferred by such β exchange, and that further exploration of such questions may provide some understanding of natural variation in transamination mechanisms and rates, even within just the various isoforms of ALT, and the biological advantages to having such flexibility.

The effectiveness of hyperpolarization is dependent on the T_1 relaxation of the molecule of interest—longer relaxation times lead to longer-lasting hyperpolarization, thus it is desirable to maximize T_1 for experiments in which hyperpolarization is used. The T_1 relaxation of a given molecule can be prolonged by replacing its hydrogen atoms with deuterium. Doing so, however, exposes the experiment to a new question—does perdeuteration of pyruvate change the rate at which it is metabolized? A 2019 NMR study by Funk et al. demonstrates that the rate of pyruvate metabolism in heart muscle is not significantly affected by deuteration of the pyruvate methyl group, such that this technique can be used without concern that the metabolic rates being studied are influenced by this factor (Funk et al., 2019). This finding is in agreement with Cooper, who documented no difference in transamination rate between $^{13}C_3H_3$-alanine and $^{13}C_3D_3$-alanine (Cooper, 1976).

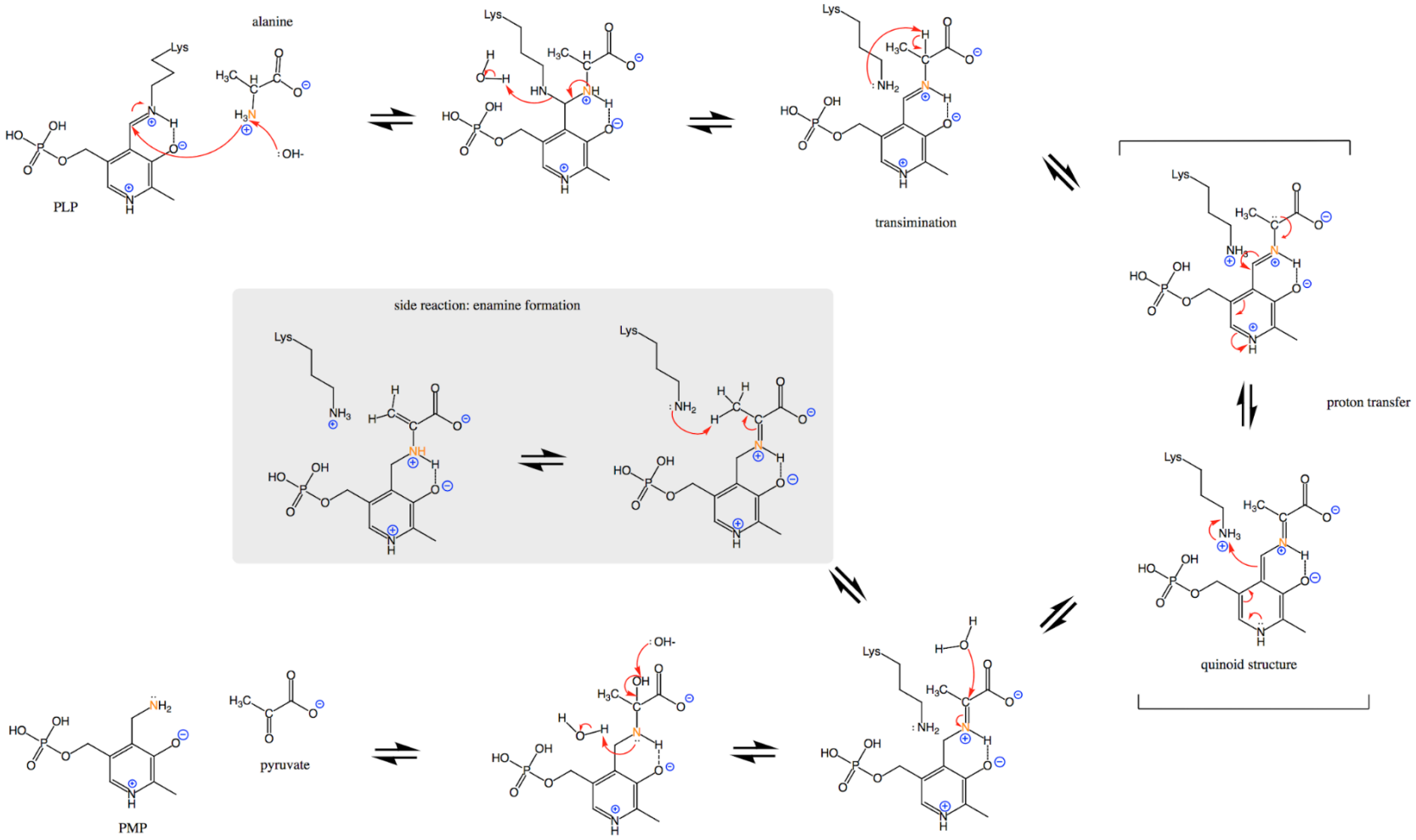

Figure 4.4: Barb et al., 2013's proposed Alanine Transaminase (ALT) catalytic mechanism. An excursion from the main pathway, as highlighted by a gray box, would explain the observed D/H exchange (figure adapted from Barb et al., 2013).

References

Albers, Mark J et al. (2008). "Hyperpolarized ^{13}C lactate, pyruvate, and alanine: noninvasive biomarkers for prostate cancer detection and grading." In: *Cancer Research* 68.20, pp. 8607–8615.

Ardenkjaer-Larsen, Jan H, Björn Fridlund, Andreas Gram, Georg Hansson, Lennart Hansson, Mathilde H Lerche, Rolf Servin, Mikkel Thaning, and Klaes Golman (2003). "Increase in signal-to-noise ratio of ¿ 10,000 times in liquid-state NMR." In: *Proceedings of the National Academy of Sciences of the United States of America* 100.18, pp. 10158–10163.

Babu, U M and R B Johnston (1974). "D_2O-Alanine exchange reactions catalyzed by Alanine Racemase and Glutamic Pyruvic Transaminase". In: *Biochemical and Biophysical Research Communications* 58.2, pp. 460–466.

Barb, A W, S K Hekmatyar, J N Glushka, and J H Prestegard (2013). "Probing alanine transaminase catalysis with hyperpolarized $^{13}CD_3$-pyruvate". In: *Journal of Magnetic Resonance* 228.Supplement C, pp. 59–65.

Cooper, Arthur J L (1976). "Proton magnetic resonance studies of glutamate-alanine transaminase-catalyzed deuterium exchange. Evidence for proton conservation during prototropic transfer from the alpha carbon of L-alanine to the C4-position of pyridoxal 5'-phosphate." In: *Journal of Biological Chemistry* 251.4, pp. 1088–1096.

Funk, Alexander M, Xiaodong Wen, Thomas Hever, Nesmine R Maptue, Chalermchai Khemtong, A D Sherry, and Craig R Malloy (2019). "Effects of deuteration on transamination and oxidation of hyperpolarized ^{13}C-Pyruvate in the isolated heart". In: *Journal of Magnetic Resonance* 301, pp. 102–108.

Golichowski, Alan, Richard C Harruff, and W Terry Jenkins (1977). "The effects of pH on the rates of isotope exchange catalyzed by alanine aminotransferase". In: *Archives of Biochemistry and Biophysics* 178.2, pp. 459–467.

Grisolia, Santiago and R H Burris (1954). "Preparation Of Glutamate And Carbamyl Glutamate Selectively Labeled With Deuterium". In: *Journal of Biological Chemistry* 210.1, pp. 109–117.

Hilton, Mary A, Frederick W Barnes Jr., Sarah S Henry, and Theodore Enns (1954). "Mechanisms In Enzymatic Transamination". In: *Journal of Biological Chemistry* 209.2, pp. 743–754.

Konikova, A S, N N Dobbert, and A E Braunstein (1947). "Labilization of the α-Hydrogen of Amino-acids in the Presence of Aminopherase". In: *Nature Publishing Group* 159.4028, pp. 67–68.

Oshima, T and N Tamiya (1961). "Mechanism of transaminase action." In: *Biochem. J* 78.1, pp. 116–119.

Toney, Michael D (2005). "Reaction specificity in pyridoxal phosphate enzymes". In: *Archives of Biochemistry and Biophysics* 433.1, pp. 279–287.

Voet, Donald, Judith G Voet, and Charlotte W Pratt (2016). *Fundamentals of Biochemistry: Life at the Molecular Level, 5th Edition*. Life at the Molecular Level. Wiley Global Education.

Walter, Ulrich, Hilmar Luthe, Hans Dieter S"oling, and Fritz Gerhart (1975). "Hydrogen Exchange at the β-Carbon of Amino Acids during Transamination". In: *European Journal of Biochemistry* 59.2, pp. 395–403.

Whelan, D J and G J Long (1969). "Enzymatic transamination: The synthesis of α,β,β-Trideutero-L-glutamic acid and α-Dentero-L-glutamic acid". In: *Australian Journal of Chemistry* 22.8, pp. 1779–1782.

Witney, Timothy H and Kevin M Brindle (2010). "Imaging tumour cell metabolism using hyperpolarized ^{13}C magnetic resonance spectroscopy". In: *Biochemical Society Transactions* 38.5, pp. 1220–1224.

Xie, Hao, Camilo Ponton, Michael J Formolo, Michael Lawson, Brian K Peterson, Max K Lloyd, Alex L Sessions, and John M Eiler (2018). "Position-specific hydrogen isotope equilibrium in propane". In: *Geochimica et Cosmochimica Acta* 238, pp. 193–207.

Xu, Tao, Dirk Mayer, Meng Gu, Yi-Fen Yen, Sonal Josan, James Tropp, Adolf Pfefferbaum, Ralph Hurd, and Daniel Spielman (2011). "Quantification of in vivo metabolic kinetics of hyperpolarized pyruvate in rat kidneys using dynamic ^{13}C MRSI". In: *NMR in Biomedicine* 24.8, pp. 997–1005.

Chapter 5

HYDROGEN-DEUTERIUM ISOTOPIC DYNAMICS AND SITE PREFERENCES IN ALANINE IN THE PRESENCE OF ALT ENZYME, ABSENT TRANSAMINATION

5.1 Abstract

We have determined by ^{13}C and ^{1}H NMR the equilibrium constants and rate constants characterizing the conversion between eight isotopically distinct species of alanine that vary in hydrogen isotope composition at the α-carbon and β-carbon sites. The hydrogen exchange between these two sites and solvent water is mediated by alanine transaminase (ALT), a ubiquitous enzyme involved in central metabolism. Our experiments track the abundances of these eight alanine species over a series of reactions differing in initial alanine isotopic composition and/or initial water isotopic composition. These eight isotopic species (involving both "clumped", or multiply substituted, isotopologues and site-specifically distinct isotopomers) are the only eight species possible considering only variation in hydrogen and deuterium at these two carbon sites, meaning that our measurements have the potential to characterize this isotopically exchangeable system completely. This study involves high density observations (measurements are frequent relative to the timescale of the reactions) during exchange kinetics, providing an unprecedentedly high-temporal-resolution look at enzymatically-mediated isotopic exchange, with observations taking place with frequency of tens of seconds. This is made possible by novel ^{13}C-NMR applications that we developed in the Caltech CCE NMR lab and present here for the first time. We built a kinetic model that allows determination of up to thirteen rate constants (three reactions involving intramolecular alanine exchange, and ten involving intermolecular solvent-alanine exchange) and ten equilibrium constants characterizing this system by fitting to our ^{13}C and ^{1}H NMR dataset, and evaluated the robustness of our model by iterating through a series of reduced parameter versions of the model and tracking the consistency of parameter distributions as we increased degrees of freedom.

We find evidence of a set of ten solvent exchange rate constants and three

intramolecular exchange rates that describe the evolution of the system. The intramolecular exchange rate constants are approximately an order of magnitude larger than those for solvent exchange. We find rate-constant variations of approximately a factor of four within the solvent exchange rate constants. Generally, the β-exchange rate constants are faster than those for α, but they do overlap; the fastest rate constant for α-exchange is very close to the slowest rate for β-exchange. There is a variation of about a factor of three within the rate constants for β exchange, and about a factor of two across those for α exchange.

5.2 Introduction

The idea that the equilibrium isotopic composition of a particular site in a molecule should, through the principles of mass-dependent isotope fractionation, depend on the isotopic composition of other atoms within the molecule has been appreciated since the early days of applying quantum mechanical principles to isotope fractionation (e.g., Urey, 1947; Bigeleisen et al., 1947). However, to our knowledge it was not until the early 2000s that explicit calculations were conducted to estimate the magnitude of this effect (Schauble et al., 2006 and Wang et al., 2004). These theoretical estimates were followed by development of analytical techniques capable of measuring these correlations using what is now called "clumped isotope" geochemistry. Advances in mass spectrometry in the last two decades have established the field of "clumped isotope" chemistry, which has been useful in a wide variety of geochemical contexts (Eiler, 2007).

Given the success of this recent work on isotopically equilibrated systems, the question naturally arises of whether to expect *kinetic*, in addition to thermodynamic, "clumping" effects. That is, does the rate of an isotope exchange reaction occurring at a particular site in a molecule also depend on the isotopic composition at other sites within that molecule? Because free energy differences, as reflected in equilibrium constants, drive rates, the answer to this question is expected to be affirmative, but there have been no studies demonstrating the extent to which the rate of isotope exchange at one site in a molecule depends on isotopic composition of nearby sites. This idea is distinct from the familiar kinetic isotope effect: it is not just a question of determining the difference between rates of H exchange and D exchange from a particular site on a molecule, but how these rates depend on the H/D pop-

ulation at neighboring sites on that particular molecule.

As a first step into understanding kinetic clumping, here we investigate H/D exchange in alanine, facilitated by the alanine transaminase enzyme (ALT). ALT catalyzes the transfer of the amine group in L-alanine to α-ketoglutarate, forming pyruvate and L-glutamate using pyridoxal phosphate (PLP) as a coenzyme. ALT is ubiquitous across all branches of life, existing as isoenzymes in many slightly different evolved forms (e.g., DeRosa et al., 1975). While the conversion of alanine to pyruvate obviously requires the removal of hydrogen bound to alanine's α-carbon, what makes this enzyme particularly useful and interesting in this study is the fact that, while alanine is bound to the active site of ALT, the enzyme also catalyzes an off-pathway exchange of the three hydrogen atoms bound to alanine's β-carbon (alternately denoted the "methyl-carbon"). The adjacency of hydrogen exchange is crucial to studying the system, that is, if there were more bonds separating the two exchanging sites, in general, we would not predict a significant exchange rate dependence. This behavior of two adjacent, unique carbon sites simultaneously exchanging hydrogen offers a serendipitous opportunity to study isotopic dynamics at a level of detail that to our knowledge has never been attempted previously (see Chapter 1 for background discussion).

Another aspect of this reaction that enables this study is the fact that both of these site-specific hydrogen isotope exchanges occur in the absence of glutamate and α-ketoglutarate. The "bi-bi ping-pong" nature of transaminase reactions is actually the coupling of two independent "half-reactions" between two amino-acid-keto-acid pairs, in the case of ALT, one between alanine and pyruvate, and another between glutamate and α-ketoglutarate (see Figure 3.1). Only when these two half-reactions are coupled can the chemical reaction proceed. In contrast, for the hydrogen isotope reaction to proceed at the two sites of interest, there need only be some minimal amount of alanine and pyruvate exchanging amino and keto groups back and forth via the enzyme, while the quantities of the two chemical components remains constant (upon equilibration with the enzyme itself, which is generally many orders of magnitude lower than the amino acid and corresponding keto acid concentrations, and is therefor negligible). The experiments we present in this chapter take advantage of this "half-reaction" involving only alanine and pyruvate, and the concentrations of these two chemical species are essentially fixed throughout

the reaction, i.e., the change occurring during reactions is limited to hydrogen exchange between the two exchangeable sites of alanine and water. Constructing our experiments in this way allows us to focus completely on the isotopic state of the system, eliminating the complication of chemical equilibration, which we considered for carbon isotopic exchange in Chapter 3. This is a convenient feature of this system as systems where isotopic and chemical equilibrium can be decoupled at will are not common. This ability could be important in understanding coupling of isotopic and chemical equilibrium (Druhan et al., 2013; Steefel et al., 2014).

We use ^{13}C NMR to establish the equilibrium populations and rates of conversion of the eight distinct enzyme-mediated H/D isotopic variants of alanine, i.e., from the slowly exchanging sites on the α- and β-carbon sites (notationally simplified as variations on $CHCH_3$), excluding the non-catalytic, very rapidly exchanging sites on the amine and carboxyl groups. We then fit these populations to a phenomenological kinetic model which allows determination of the reaction rates that convert one isotopic variant into another. Through this work we want to establish:

1. What are the rates of conversion amongst the eight isotopic variants? How much variation is there in theses rates? Are the magnitudes of such variations small compared to, for example, differences between exchange rates at the α- and β-carbon sites in isotopic tracing experiments, e.g., Cooper, 1976; Golichowski et al., 1977?

2. Assuming they are large enough to be observed, is there any pattern in the conversion rates that would give some insights into the mechanism of ALT-induced H/D exchange in alanine, and therefore further insight into the mechanism of ALT itself? For example, perhaps the H/D exchange rates at the α-carbon are similar, but the rates at the β-carbon depend systematically on the number of H or D on the β site; or maybe the H/D exchanges at the β-carbon all have similar rates and the exchanges at the α-carbon depend on the H/D ratio at the β-carbon.

It is important to recognize that the experiments and results we present in this chapter are unusual and even unprecedented in a number of ways. In fact, the NMR capabilities to conduct the key experiment in our study have existed

since the mid-1990s; the instrument employed for our experiments (including its 3-channel probe, the essential component for our key ^{13}C NMR experiment) has resided in the central NMR facility on our campus since 1998. That over two decades elapsed between the emergence of this capability and its use for this set of experiments begs for explanation.

We suggest that there are several factors contributing to this delay. First, fundamental experimental work in isotope geochemistry has not historically made broad use of NMR as the analytical tool of choice. To the extent that NMR is employed toward isotopic studies today, the vast majority of this work is heavily driven by application (e.g., food forensics, fossil fuel industry, medicine), and would rarely be considered fundamental chemistry. This is, in part, a natural consequence of the historical developmental paths of analytical tools available to be employed in such work. The isotope geochemist's instrument of choice, the isotope-ratio mass spectrometer (IRMS), is used for the vast majority of isotopic investigations, and justifiably so. Mass spectrometry was developed significantly earlier than NMR as a tool for observation of isotopic differences, and so, in part, the scientific community that was focused on isotopic partitioning in natural materials became habituated to designing experiments around what was possible with this tool. This had the benefit of encouraging further development in mass spectrometry toward greater capability in, e.g., distinguishing smaller mass differences and handling a larger variety of molecules. State of the art mass spectrometers are extremely capable instruments, highly sensitive to tiny differences in relative abundances of distinct species as well as to differences in mass as low as millidaltons (mDa), meaning that many ionized molecular species and fragments having the same number of nucleons (e.g., methane species $^{13}CH_3D$ and $^{12}CH_2D_2$, which differ in mass by $\sim$3 mDa) can be distinguished.

Nevertheless, we are not aware of any study using mass spectrometry to measure the intensities of multiple isotopic species (in our case, eight species, three pairs of which share the same number of nucleons) over the course of an isotopic exchange reaction at a fine temporal resolution, during which the compositions and relative abundances of these species change dramatically, resulting in a data set that can be used to extract both a set of rate constants and a set of equilibrium constants for the observed reactions. There are several barriers to the feasibility of a mass spectrometric study of this kind. At the time of our

experimental work, there were neither commercially available nor prototype mass spectrometers that offered all the features required for this experiment: first, in order to differentiate and quantify all eight species, it would be necessary to have the ionization source generate both the molecular ion *and* a specific set of ionized fragments of the molecule in order to differentiate between variants having the same mass—this would require an electron impact ionization source, rather than a softer ionization method such as electrospray; second it would need to simultaneously inject a temperature-controlled reaction solution for continuous observation of the masses of interest as they evolve over time. In addition, because it is a destructive method in which the molecules under observation make direct contact with the surfaces of the instrument, mass spectrometry is sensitive to isotopic contaminants. Research in stable isotope geochemistry is generally performed using mass spectrometers to observe isotopic systems having natural or near-natural isotopic abundances, with small amounts of heavy isotope doping used occasionally during initial stages of methods development. An investigation of the sort performed in this study requires unnatural isotopic abundances; that is, while terrestrial stable isotopes of hydrogen include only ∼0.1% deuterium, our study includes experiments with as much as ∼85% deuterium in water. This high level of deuterium would cause severe contamination to any mass spectrometer used for measurements on substances containing natural isotopic abundances.

In contrast, NMR is an independent technology that can be used to observe mass differences in molecules. While NMR has significant constraints that mass spectrometry does not share, it does not suffer from some of the limitations of mass spectrometry as described above, making it well suited to certain tasks to which mass spectrometry is poorly suited (and, of course, vice versa). One of the strengths of NMR is the fact that it is a non-destructive analytical method, which in our case is indirectly useful in the sense that the sample never comes directly into contact with the instrument, and therefore isotopic contamination is not an issue. While many isotope geochemistry investigations deal with a small, precious volume of sample extracted laboriously from some natural source, our experiment does not involve natural samples, and is therefore limited in scale only by cost of isotopically labeled reagents and the volume of sample admitted to the NMR probe (∼600 µL), so we are able to scale up our reaction to a volume and concentration suitable for this analytical technique, which requires a large amount of sample relative to the

amount that would be required for measurement by mass spectrometry. The non-destructiveness of NMR also implies that there is no ionization or fragmentation of the molecules under observation, which means this method is free from the complications introduced by ionization required for observation by mass spectrometry.

In essence, our study leverages several features unique to NMR, without compromising on experimental design, which, when applied to the alanine transaminase system, result in meaningful data that could not be acquired by other existing means, and is not easily compared with any preexisting work. The most uniquely distinctive feature of our approach was our use of "double decoupling", that is, simultaneously decoupling ^{1}H from ^{13}C as well as ^{2}H from ^{13}C in our ^{13}C NMR experiments presented in this chapter. This "double decoupling" process serves to simplify our spectra such that each of the eight unique isotopic variants of interest corresponds to a unique ^{13}C NMR peak. This is an essential feature of our work, as it allows us to track the quantity of each variant over time, making it possible to model the system phenomologically, as we present below. While "double decoupling" of these nuclei is not unheard of, it is not frequently used outside of NMR protein structure studies conducted at high field strengths. Naturally it is useful only in contexts where there are significant proportions of both protium and deuterium. In the protein structure context, proteins are deuterated to increase T_1; that is, signal arising from carbon bound to deuterium suffers less from loss due to pulse sequences than does carbon bound to protium, thus giving rise to stronger carbon signal.

It is our belief that the isotope geochemistry community would benefit from intentional consideration of a subset of fundamental studies in isotope partitioning that would lend themselves particularly well to NMR, complementing the strengths of mass spectrometry.

5.3 Materials and Methods

5.3.1 Experimental Materials

We used L-alanine from Alfa Aesar (99% purity, Lot # 10167391) with isotopic signatures close to natural abundance. We also used L-alanine purchased from Sigma Aldrich, labeled selectively with either ^{2}H or ^{13}C: (1) 99 atom % ^{13}C label at C-3 (Lot # EB2211V), (2) 98 atom % D L-alanine-2-d (Lot #

MBBB5282V), and (3) 99 atom % D L-alanine-3,3,3-d3 (Lot # SN1346V).

Ultrapure water was obtained from a Millipore ultrahigh-purity water (18.2 MΩ cm) system at Caltech. Most experiments were conducted with deuterium enriched water. We used D_2O purchased from Cambridge Isotope Laboratories, Inc. (D, 99.9%, Lot # M3063). Water solutions of varying D/H were prepared by mixing volumes of D_2O and H_2O to the desired ratio. This labeled water was used to examine the kinetics of hydrogen isotope exchange between water and alanine.

Alanine transaminase (abbreviated ALT, EC number 2.6.1.2) derived from porcine heart was used to catalyze hydrogen exchange on alanine α and β carbons. ALT (∼7.65 mg protein/mL) suspended in 1.2 M ammonium sulfate was purchased from LeeBiosolutions (Lot # 07B3133). The lot used in the experiments presented here was measured by the vendor to have an activity of 1990 Units/mL Ammonium Sulfate suspension, or 260 U/mg protein, based on an assay run at 310 K and pH 7.4, with initial concentrations of alanine and α-ketoglutarate at 260 mM and 20 mM, respectively, tris buffer at 100 mM, pyridoxal-5'-phosphate at 0.15 mM, and enzyme diluted to <1 U/mL. The protein exists as a dimer in solution, with a molecular weight of ∼100,000 g/mol dimeric ALT (i.e., ∼50,000 g/mol per monomer).

Additional reaction components included pyruvic acid from Sigma Aldrich (purity 98%, Lot # SHBH2749V). Gadolinium(III) Chloride hexahydrate from Sigma Aldrich (purity 99.999%, trace metals basis, Lot # MKCC3867) was used to reduce relaxation times so that more frequent scans could be acquired during kinetics without compromising quantitative signal.

5.3.2 Experimental Procedures

Hydrogen-Deuterium exchange ("HDX") experiments were conducted by incubating alanine at 303 K in the presence of ALT. Reaction media were prepared by weighing 53 mg of alanine for each experiment, which was dissolved in 600 μL of solution, resulting in a concentration of alanine of 1 molar. To this solution was added 1.5 mM pyridoxal-5'-phosphate (the co-enzyme for this reaction), and 5 mM pyruvate (except where otherwise noted). Stock solutions of pyruvate were prepared by adding 35 μL pyruvic acid to 5 mL of either ultrapure H_2O or 99.8% D_2O, and raising the pH to 7.5 by addition of sodium hydroxide. 45 μL ALT enzyme stock solution was added (for an

enzyme concentration of 150 Units/mL) to initiate each reaction. Reactions were conducted at a pH of 7. To avoid possible damage to the enzyme that might result in reduced enzyme activity, we left the enzyme in its ammonium sulfate suspension (rather than removing it from the suspension via centrifugation) and added this solution directly to the reaction mixtures, resulting in a concentration of ammonium sulfate approximately 50 mM. Upon addition of ALT enzyme to the solution, the tube containing the reaction solution was vortexed for several seconds and then quickly transferred to an NMR spinner. Because NMR probe tuning is sensitive to salt concentrations, we manually tuned the probe as soon as each kinetics sample was inside the magnet, i.e., immediately prior to shimming, followed by initiating data collection for each kinetics experiment. The time delay between adding ALT enzyme to the prepared solution and the start of kinetics data acquisition was typically 5–7 minutes. We have identified a way of correcting some experiments for this delay, as detailed in the Supplementary Materials Section 5.8.4.2 Table 5.1 lists additional concentrations and conditions for each experiment.

5.3.3 NMR Procedures

Data collection for this study was performed on a Varian 600 MHz spectrometer with a 5mm inverse triple resonance probe (^{1}H, ^{2}H, ^{13}C, ^{15}N, ^{31}P), using combined ^{1}H-^{13}C and ^{2}H-^{13}C decoupling for some experiments (as indicated in Table 5.1). The temperature of the probe was set to 303 K, following the experimental conditions in Golichowski et al. (Golichowski et al., 1977). For each kinetics experiment, reaction mixtures were prepared without enzyme, and a spectrum acquired to get a pre-kinetics data point. Enzyme addition took place just prior to beginning observation of kinetics by ^{1}H or ^{13}C NMR. A delay of several minutes between enzyme addition and the beginning of data acquisition was necessary for shimming and manual probe tuning; these first several minutes of each reaction are therefore absent from our data sets. All original spectra and extracted data are included as supplementary files.

We determined the final D:H of water in each kinetics experiment (^{1}H NMR and ^{13}C NMR) by measuring the ^{1}H NMR water signal in each sample once reactions had reached equilibrium and comparing it to a standard calibration curve. These measured water composition values were then used as constraints in determining the intramolecular K_{eq} values in Section 5.4.3. Because NMR signal intensity scales linearly with concentration of the observed nucleus in so-

lution (given a constant parameter set and constant volume of solution within the receiver coil), a measurement of a single sample with known water D:H is the minimum information necessary to calibrate D:H of water to signal intensity. However, in practice, the very high signal intensity of D_2O (or H_2O, depending on the sample and nucleus of interest) causes radiation damping in some of our samples, thus distorting the baseline, which could introduce some nonlinearity. By using a set of prepared Milli-Q H_2O and 99.8% D_2O mixtures, we made a standard calibration curve, intended to account for any minor nonlinear effects. These mixtures had Milli-Q H_2O/99.8% D_2O ratios of 25/75, 50/50, 75/25 and 100/0. With the y-intercept set to 0, a linear fit to these data gave $r^2 \geq 0.998$.

5.3.3.1 ^{1}H NMR HDX Kinetics Experiments

For ^{1}H NMR kinetics experiments, we acquired single-scan spectra at 1-min intervals, typically spanning six hours, resulting in arrays of 360 spectra. Deuterium decoupling was used in some experiments (see Table 5.1). Deuterium decoupling collapses the small J-coupling of protons three bonds away from a deuterium into singlets. Earlier experiments were not deuterium decoupled, but we eventually found broadband deuterium decoupling to be advantageous, as it simplifies spectra in which carbon-bound deuterium is present. In this series of experiments, the four possible isotopic variants of interest that contain an H on the α carbon site (i.e., $^{12}CH^{13}CH_3$, $^{12}CH^{13}CH_2D$, $^{12}CH^{13}CHD_2$, and $^{12}CH^{13}CD_3$) appear in the resulting NMR spectra as a group of overlapping peaks in the $\sim$4 ppm region, while the seven isotopic variants containing one or more H atoms on the β carbon site (i.e., the four α_H variants listed above, plus $^{12}CD^{13}CH_3$, $^{12}CD^{13}CH_2D$, and $^{12}CD^{13}CHD_2$) appear as a group of overlapping peaks in the $\sim$1 ppm region. The eighth possible variant, $^{12}CD^{13}CD_3$, which lacks any hydrogen atoms, is dark to this method. Note: We simplify notation by leaving out the amine and carboxyl groups in these formulas, as those remain constant across all variants.

We do not attempt to quantitatively monitor HDX on the alanine amine group (NH_3^+) over the course of our HDX kinetics experiments, as the hydrogen atoms on this group equilibrate with water effectively immediately (independent of enzyme), upon alanine dissolution. Because of this rapid exchange, the amine hydrogens appear in ^{1}H and ^{2}H NMR spectra as a very broad peak (>100 Hz) that is not easily quantifiable. While these hydrogen atoms are

a significant pool of hydrogen in the system (given the 1 M alanine concentration), for our purposes their interaction with HDX at the α and β carbon positions is expected to be insignificant given that they are isotopically in equilibrium with water on relevant experimental timescales. The total contribution of exchangeable hydrogen coming from the enzyme itself is also assumed to be negligible, based on a concentration of 0.35 mg enzyme per 600 μL experiment (compared to the experimental concentration of alanine, 1 M, i.e., 54 mg per 600 μL experiment).

5.3.3.2 ^{13}C NMR HDX Kinetics Experiment

While ^{1}H NMR, used for the kinetics experiments described above, provides the ability to quantitatively track the abundance of protons bound to the α and β carbon sites, it does not allow distinction between the isotopic variants (a set of isotopologues and isotopomers) that result. Mixtures of deuterated and hydrogenated alanine species give ^{1}H NMR spectra in which the chemical shifts are not large enough relative to the peak widths to be able to resolve all eight isotopic species, even with deuterium decoupling. In addition, the fully deuterated species has no ^{1}H NMR signal. In contrast, ^{13}C NMR gives narrow peaks with large chemical shifts, resulting in widely dispersed, well-resolved signals. When combined with decoupling procedures, each peak represents a single isotopic species, allowing us to quantitatively track the abundances of each of the eight isotopic variants as they evolve through time.

Because the ^{12}C nucleus is spin 0, it is NMR-silent, and thus cannot be directly observed; ^{13}C, however, is spin $\frac{1}{2}$, and therefore NMR active, meaning it can be used as a means of indirectly observing bound hydrogen and deuterium to determine the abundance of the isotopic variants that arise via the HDX reaction. Quantifying isotopic variants with the natural abundance ^{13}C present in our ^{1}H NMR experiments (described in the previous section) proved challenging due to ^{13}C's relatively low natural abundance ($\sim$1% of all carbon), and its low NMR sensitivity relative to the ^{1}H nucleus. However, by running the same HDX reaction using alanine labeled with ^{13}C in the β site, we were able to determine the relative abundances of these isotopic variants with precision of several permil.

The ^{13}C NMR spectrum of deuterated molecules is determined by multiple factors. The ^{2}H nucleus is quadrupolar and has a spin = 1 (in contrast to

the ^{1}H dipolar nucleus spin of $\frac{1}{2}$), which increases the complexity of ^{13}C NMR spectra for carbon atoms with one or more directly attached deuterium atoms. In specific, ^{2}H-^{13}C coupling forms 1:1:1 triplets for a single coupled deuteron, 1:2:3:2:1 quintets for two coupled deuterons and 1:3:6:7:6:3:1 septets for three, whereas ^{1}H-^{13}C coupling forms 1:1 doublets for a single coupled proton, 1:2:1 triplets for two, and 1:3:3:1 quartets for three. In the absence of decoupling, the multiplets resulting from all of the alanine species present in our experiments are complicated and their signal split over a wide area only to be drowned by noise, making quantification impractical, if not impossible. ^{1}H-^{13}C decoupling is routinely used in ^{13}C NMR to simplify ^{13}C spectra, while ^{2}H-^{13}C decoupling is less routine and requires a probe with three or more channels. NMR instruments used for protein analysis typically have these capabilities, and serendipitously one of the Varian instruments in our campus facility was designed for such work, and was supplied with the necessary probe. Thus, we were able to use simultaneous ^{1}H-^{13}C and ^{2}H-^{13}C decoupling to simplify the spectra into eight baseline-separated peaks that arise from the isotopically distinct species of interest in this study. The mass difference between ^{1}H and ^{2}H also induces a chemical shift in the ^{13}C NMR spectrum.

Because spectra of these alanine isotopic species have not, to our knowledge, been documented previously in the literature, we verified their identities by preparing mixtures of fully hydrogenated alanine and alanine deuterated selectively at either the α or the β carbon positions, and measuring the chemical shift between these species at their β carbon peaks using ^{13}C NMR. As expected, the chemical shift decreases with increasing deuteration, i.e., alanine containing more hydrogen atoms appears downfield of alanine containing deuterium atoms in place of hydrogen. In the case of alanine, for each additional ^{2}H on the β carbon, we observe a chemical shift of the β carbon signal for that species of 0.25 ppm, while the two-bond shift that results from adding a ^{2}H at the α carbon position is 0.109 ppm. Supplementary Figure 5.17 gives example spectra illustrating this exercise.

Once these checks were completed, we observed one H/D isotopic exchange reaction using ^{13}C NMR. This reaction began with $> 99\%$ of a single isotopic species, alanine-3-^{13}C (alanine labeled with ^{13}C at its β-carbon site, i.e., alanine with an α- and β-carbon composition of $^{12}CH^{13}CH_3$), in a $\sim$1:1 mixture of D_2O:H_2O (measured to be 0.483:0.517). This kinetics experiment, designated

Experiment "Q", follows the time evolution of the eight isotopic variants of interest, and was conducted using the same acquisition parameters and experimental conditions as for the ^{1}H NMR kinetics experiments described above except as otherwise noted. The eight isotopic species we observe are listed here as follows: $^{12}CH^{13}CH_3$, $^{12}CD^{13}CH_3$, $^{12}CH^{13}CH_2D$, $^{12}CD^{13}CH_2D$, $^{12}CH^{13}CHD_2$, $^{12}CD^{13}CHD_2$, $^{12}CH^{13}CD_3$, $^{12}CD^{13}CD_3$. (Recall that we simplify notation by leaving out the amine and carboxyl groups in these formulas, as those remain constant across all variants.) Each of these eight isotopic variants gives rise to a unique ^{13}C NMR peak in the β carbon resonance region, in the range of 15–16 ppm.

Note that the eight isotopic variants also generate corresponding ^{13}C NMR peaks in the α-carbon resonance region, around 50 ppm (e.g., $^{13}CH^{12}CH_3$, $^{13}CH^{12}CH_2D$, etc.). We do not attempt to quantify these peaks in this study, in part because they are clustered more densely than the eight in the β region. The chemical shifts induced by the three hydrogen/deuterium atoms on the β-carbon are smaller because they are two bonds away, rather than one, and therefore have less influence on the α-carbon's chemical shift. In addition, any attempt at quantification of a similar nature to that used here in the β-carbon resonance region would have required an independent experiment with ^{13}C label at the α carbon site. Our observations do not enable us to quantify ^{13}C–^{13}C clumped species, (e.g., $^{13}CH^{13}CH_3$, $^{13}CH^{13}CH_2D$, etc.), because the signal coming from these rare species is too weak to be measured with the methods presented here. The error arising from the presence of some small amount of these species is expected to be negligible.

We used the same experimental methods as described for the ^{1}H NMR kinetics, except for the difference that the single-scan acquisitions were 2 minutes apart, rather than 1. Unlike proton relaxation times, which rarely exceed several seconds, carbon relaxation times can be significantly longer and must be accounted for when quantitative data is desired. We measured Carbon T_1 relaxation times of each of the eight isotopic species, and found they ranged from ∼2 seconds ($^{12}CH^{13}CH_3$) to ∼25 seconds ($^{12}CD^{13}CD_3$). Thus, to acquire quantitative signal on all eight species, we used a D1 of 120 seconds, approximately five times the longest T_1, resulting in an array of 270 spectra acquired over 9 hours.

The relative abundances of the eight species are determined by integrating the

area of each of the eight distinct peaks in the ^{13}C NMR spectrum, and dividing each one by the total integrated area for all eight. This procedure is done for all 270 spectra belonging to this experiment. An example spectrum is given in Figure 5.1.

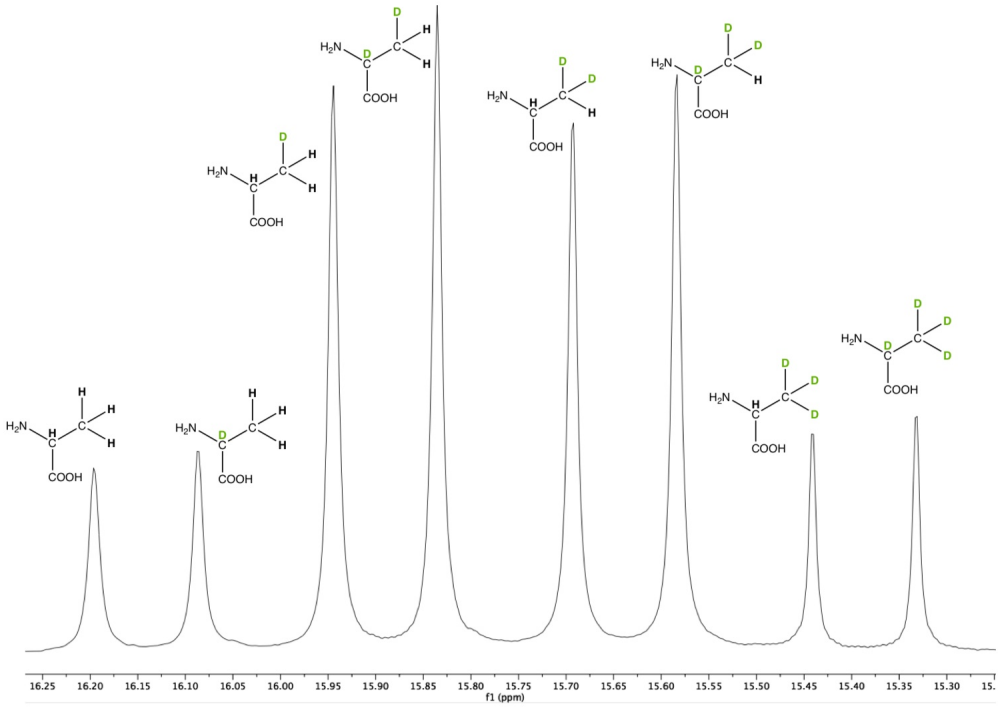

Figure 5.1: An example ^{13}C NMR spectrum (using simultaneous ^{1}H-^{13}C and ^{2}H-^{13}C decoupling) of L-alanine in isotopic equilibrium with water having a composition of 51.7 % H, 48.3 % D, showing the eight isotopic species of alanine quantified in this study. All eight isotopic variants are easily distinguishable from one another, and present in significant abundance in this solution.

5.3.3.3 ^{13}C NMR Equilibrium Acquisitions

Once the isotopic exchange reaction in Experiment "Q" reached equilibrium with water, we quantified the eight alanine isotopic variants over the following two days at significantly higher precision than was used for the kinetics portion of Experiment "Q" by acquiring sets of 16-scan ^{13}C NMR spectra on the equilibrated sample, keeping all other ^{13}C NMR acquisition parameters identical to those used in the ^{13}C NMR kinetics given above. Integrating multiple scans per acquisition increases signal to noise (and therefore precision), which is advantageous in an equilibrated solution, where the relative and absolute proportions of the eight isotopic variants are constant over time. With this method, we were able to reach a measurement precision of several permil,

which we expect to be more than sufficient to determine equilibrium hydrogen isotopic site preferences. We will use these equilibrium constants to reduce the parameter space in our modeling work presented in Section 5.5. As shown in Figure 5.1, each of the eight alanine isotopic variants gives rise to a unique, isolated and quantifiable ^{13}C NMR peak in the β-carbon resonance region, in the range of 15–16 ppm.

We processed the spectra using the autophasing and 3rd-order Bernstein Polynomial baseline correction features built into MestReNova 14.0, prior to fitting the peaks using the MestReNova line-fitting feature using Generalized Lorenzian peak shape, Simulated Annealing, and otherwise default settings. This processing is distinct from taking the peak area integrals directly, in the case of the ^{13}C NMR kinetics data described in Section 5.3.3.2.

5.4 Experimental Results

5.4.1 H/D Exchange Kinetics by ^{1}H NMR

As explained in Section 5.3.3.2, we found that ^{13}C NMR, when used in conjunction with alanine labeled with ^{13}C at the β carbon position and simultaneous proton and deuterium decoupling, provides a far superior method of observing the alanine isotopic system compared with ^{1}H NMR. While ^{1}H NMR can differentiate hydrogen bound to the α carbon from hydrogen bound to the β carbon, it is unable to distinguish the eight isotopic variants of interest to us here (proton signal from the β carbon cannot be resolved into distinct species), and cannot detect the $CDCD_3$ isotopic variant at all, because it contains no bound protons at either site.

Nevertheless, the majority of our experiments were conducted using ^{1}H NMR, as it took some time and method development for us to discover the novel ^{13}C NMR approach and determine that it was a viable option, and shortly after conducting our first ^{13}C NMR kinetics experiment, our window for further experimentation had passed. In spite of this, the ^{1}H NMR experiments we performed did produce useful data, which are important to present here. As presented in later in this chapter, a subset of our ^{1}H NMR kinetics data provided additional constraints on the 8-box model we constructed based on our ^{13}C NMR kinetics experiment. In addition to this we were able to build on investigations published the 1970s out of several research groups (e.g., Cooper, 1976, and Golichowski et al., 1977 as reviewed in the previous chapter).

Table 5.1 lists all kinetics experiments both ^{1}H NMR and ^{13}C NMR) and their corresponding parameters.

EXP	Date	Length (min)	PYR (M)	Final water D/H	α starting D/H	β starting D/H	β ^{13}C	Nucleus Observed	D Decoupling
A	06/25/2019	90	0	77/23	natural abundance	natural abundance	n	^{1}H	n
B	06/25/2019	90	0.001	77/23	natural abundance	natural abundance	n	^{1}H	n
C	06/25/2019	90	0.005	77/23	natural abundance	natural abundance	n	^{1}H	n
D	06/28/2019	54	0.005	77/23	natural abundance	natural abundance	n	^{1}H	n
E	06/29/2019	180	0.005	77/23	natural abundance	natural abundance	n	^{1}H	n
F	07/03/2019	360	0.005	77/23	natural abundance	natural abundance	n	^{1}H	n
G	07/05/2019	360	0.005	47/53	natural abundance	natural abundance	n	^{1}H	n
H	07/07/2019	360	0.005	62/38	natural abundance	natural abundance	n	^{1}H	n
I	07/08/2019	360	0.005	55/45	natural abundance	natural abundance	n	^{1}H	n
J	07/11/2019	360	0.005	85/15	99 % D	natural abundance	n	^{1}H	n
L	07/13/2019	360	0.005	18/82	natural abundance	natural abundance	n	^{1}H	n
M	07/18/2019	360	0.005	87/13	natural abundance	99 % D	n	^{1}H	y
N	07/19/2019	360	0.005	1/99	99 % D	natural abundance	n	^{1}H	y
O	07/23/2019	360	0.005	2/98	natural abundance	99 % D	n	^{1}H	y
P	08/15/2019	540	0.005	85/15	natural abundance	natural abundance	n	^{1}H	y
Q	09/13/2019	540	0.005	48/52	natural abundance	natural abundance	y	^{13}C	y

Table 5.1: ^{1}H and ^{13}C NMR Kinetics Experimental Parameters and Conditions. All experiments were performed at 304 K and with 1 M alanine. Original ^{1}H and ^{13}C NMR data are given in the Supporting Information.

5.4.1.1 Pyruvate Concentration Dependence

Our first set of ^{1}H NMR kinetics experiments tested the reaction rate-dependence on pyruvate concentration, as shown in Figure 5.2. Golichowski et al. reported that the rate of H-D exchange increased with increasing pyruvate concentration, but did not discuss the implications of this dependence, nor did they report any experimental data for the reaction run in the *absence* of pyruvate (Golichowski et al., 1977). In our initial tests to confirm our observation of hydrogen exchange at the alanine β carbon position, samples were composed of alanine in 10% D_2O solution in the presence of ALT, absent any other components of the alanine transamination reaction (i.e., no pyruvate, alpha-ketoglutarate, or glutamate). Our results confirm that the α and β H-D exchange rates are dependent on the concentration of pyruvate, but also show that, in the absence of any added pyruvate, the exchange reaction proceeds at a much faster rate than Golichowski et al. were able to extrapolate based on their data. In addition, Golichowski et al. reported identical α and β exchange rates, while our data show the β exchange rate to be consistently slightly higher than that of α exchange (on a per-site basis, that is, the rate of β exchange is more than three times faster than α exchange as would be expected, simply because of multiplicity, i.e., there are 3 β_H for every α_H).

Although the experimental conditions used in this study attempted to replicate those reported in Golichowski et al., our experiments do not reproduce the rates they measured. One possible explanation for this could be a different D_2O:H_2O. The initial measured exchange rates depend, of course, on the degree to which the system is in disequilibrium; thus our experiments, which took place in a solution of approximately 75:25 D_2O:H_2O, may fail to replicate the Golichowski experiments, which reportedly took place in nominally 100% D_2O. However, it is doubtful that this was indeed quantitatively true; for instance, Golichowski did not document any method to deuterate the exchangeable sites on the alanine initially added to the reaction (i.e., alanine's three readily exchangeable hydrogen atoms, the amino and carboxyl hydrogen atoms). At the 1 M alanine concentration used in their experiments (and ours), there are 3 M exchangeable hydrogen atoms from alanine, leading to the addition of several percent hydrogen into the solution. It is also not documented in their study whether the ALT enzyme added in their experiments was in a solution of D_2O or H_2O. If it was in H_2O, adding this solution to D_2O would change the D_2O:H_2O. If the D_2O:H_2O of their experiments was indeed

higher than ours, the initial driving force for their reaction would have been higher, and, in that case, the rates they measured are expected to be higher than those we measured.

The measured $T_{\frac{1}{2}}$s in our study are longer than those of Golichowski et al. for experiments at 1 and 5 mM pyruvate, while the $T_{\frac{1}{2}}$ at our lowest experimental pyruvate concentration is much shorter than that predicted by extrapolation using a linear fit to their data, as plotted in Figure 5.3. Our data do not show a linear trend in reciprocal concentration space, while the data of Golichowski et al. do appear to follow such a trend. Golichowski et al. do not seem to have considered the possibility that alanine β HDX can occur without the addition of pyruvate. This is possible in part because the hydrogen exchange occurs at an intermediate stage of the alanine-pyruvate half-reaction, i.e., it is not necessary that alanine bind to the co-enzyme and leave as pyruvate in order for exchange to occur; instead, alanine can bind to the enzyme, proceed to an intermediate state in which HDX occurs, and then reverse direction and leave the way it came, as alanine. Also, in the absence of added pyruvate, a very small amount of pyruvate is still able to form. For every molecule of alanine that binds to the enzyme and leaves as pyruvate in the alanine $\rightarrow$ pyruvate half-reaction, that same number of enzyme + co-enzyme complexes are left in a configuration that will only accept a keto acid (in this case, pyruvate, as it is the only keto acid present; see Figure 4.4). Thus, by definition, the maximum possible pyruvate concentration under these circumstances cannot exceed the enzyme concentration, but will come to equilibrium at some concentration less than the concentration of the enzyme in solution. The exact relative proportions will depend on the binding affinities between pyruvate and enzyme, and alanine and enzyme. In our case, the concentration of alanine in solution is 1 M, while the enzyme concentration is lower by six orders of magnitude; we calculate the enzyme concentration to be roughly 2 μM, based on an approximate enzymatic mass of 100 kDa and a concentration of 0.35 mg of enzyme per 600 μL experimental volume. Thus we estimate the maximum pyruvate concentration in this experiment to be less than the ALT enzyme concentration of approximately 2 μM.

5.4.1.2 Water H/D Composition Dependence

Our second set of HDX experiments evaluated the kinetics of exchange over a wide range of D_2O:H_2O, and the corresponding equilibrium partitioning

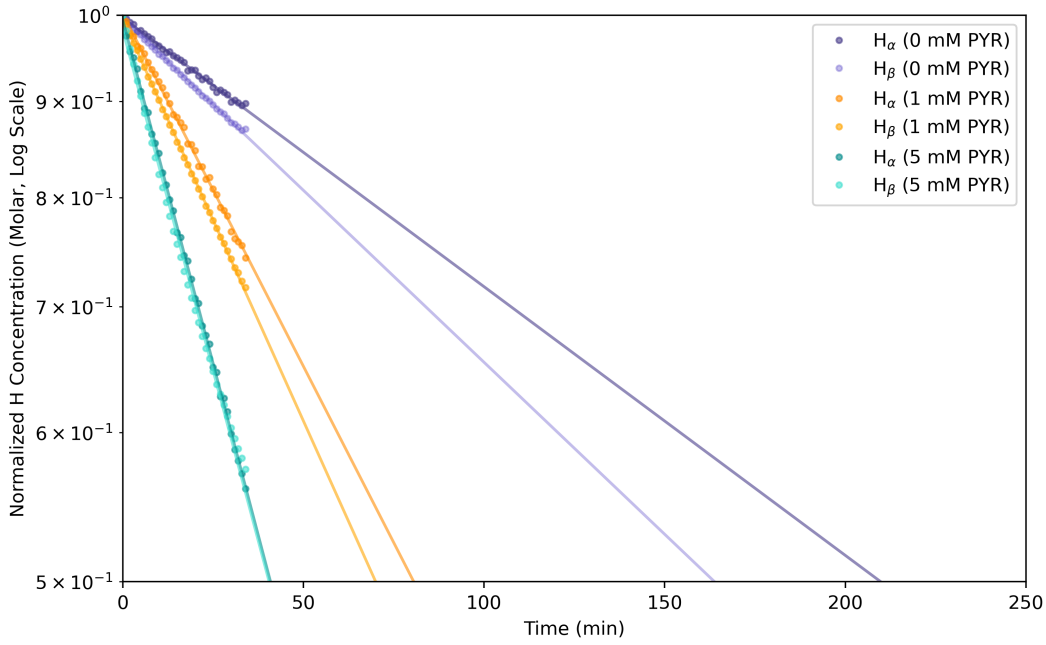

Figure 5.2: Comparing β and α protons loss rates in L-alanine as a function of pyruvate concentration, based on initial 35 minutes of each kinetics experiment. Here we normalize β_H signal (by using $\beta_H/3$) to start at a "concentration" of 1.0, for ease of comparison with α_H H signal loss rate. All reaction solutions were incubated at 303 K and contained 1.0 M L-alanine, 150 Units/ml of alanine transaminase, and varying pyruvate concentrations. These duplicate the experimental conditions of Golichowski et al., 1977. Exponential fits are shown as lines through these data, with equations for them given in the legend.

of deuterium between the α and β positions of alanine. Each experiment began with L-alanine, pyruvate, and pyridoxal phosphate (P5P) in water, with water composition ranging from ~18% D_2O (82% H_2O) to ~85% D_2O (15% H_2O). Our data show the β exchange rate constant to be consistently slightly higher than the α exchange rate constant. The overall shape of the time series $\beta_H/(3\alpha_H)$ (i.e., the ratio of β_H peak area divided by 3*α_H peak area, as depicted for this set of experiments in Figure 5.4) has notable features that invite explanation, namely the initial dip below 1 at early times, subsequent rise at intermediate times and the slow decay over long time scales. Loss of H in β is initially slightly faster than loss of H in α, resulting in an initial reduction in this ratio to a value ~0.95. As demonstrated by our selectively deuterium-labeled experiments discussed below, the H-loss rate of α is increased as β gains deuterium (i.e., $\beta CH_3 \rightarrow \beta CD_3$), while β H-loss rate decreases as α gains

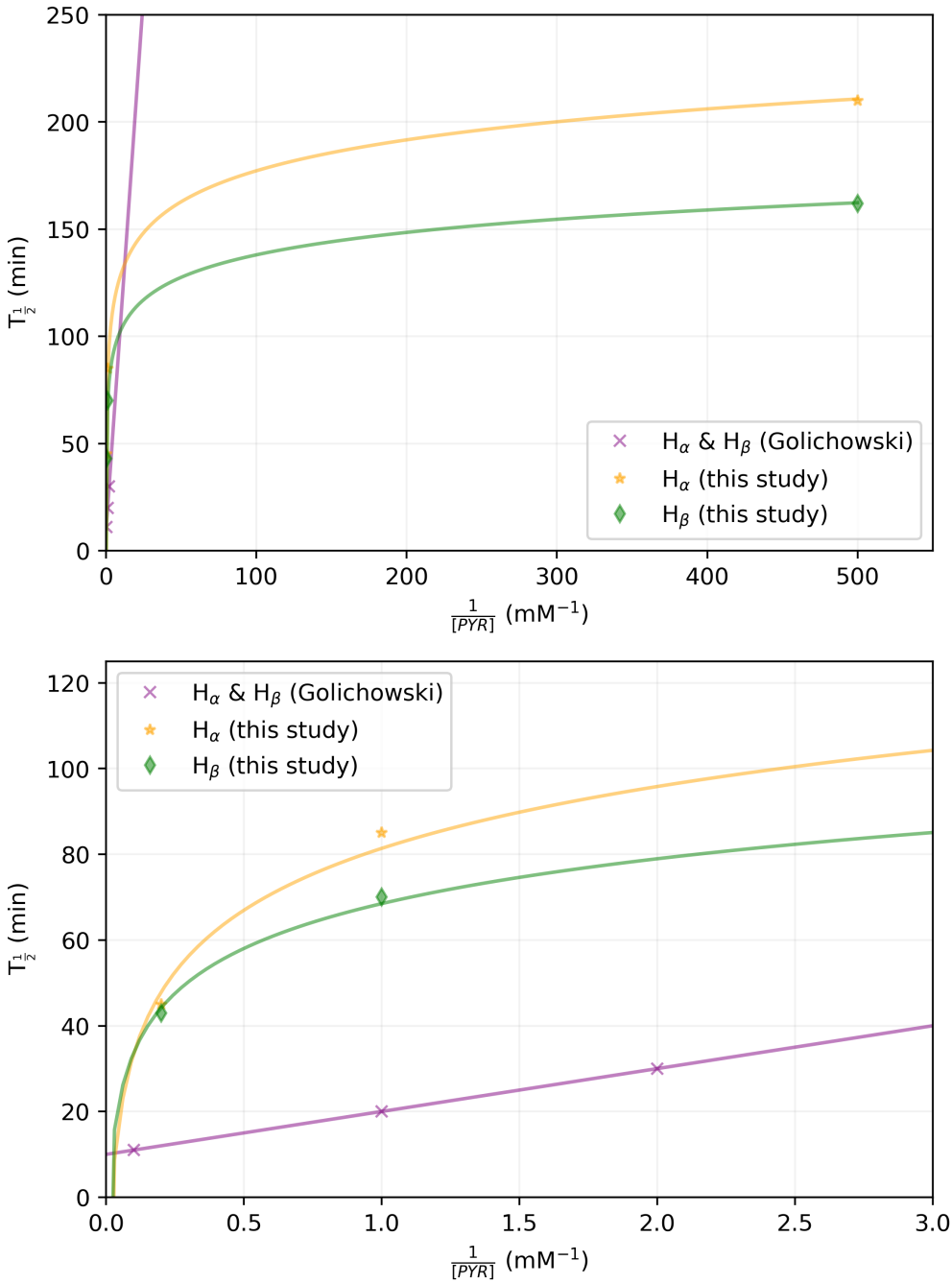

Figure 5.3: (a) Comparison of measured "half-life" for β and α proton exchange between Golichowski et al., 1977 (purple x's), and this study (α, yellow stars; β, green diamonds). (b) Magnified portion of (a) to give more detail in the region of higher pyruvate concentration. Note that we show two curves, one for α, and one for β, while Golichowski et al. were not able to differentiate between α and β exchange rates.

deuterium (i.e., as αCH $\rightarrow$ αCD). The combined effect of these H-loss rates, apparently influenced by the D/H ratio of the other site, is a crossover point at which α H-loss becomes faster than β H-loss, and $\int \beta_H/(3\int \alpha_H)$ evolves to values greater than 1. The higher the D_2O:H_2O, the greater this value overshoots the eventual equilibrium value.

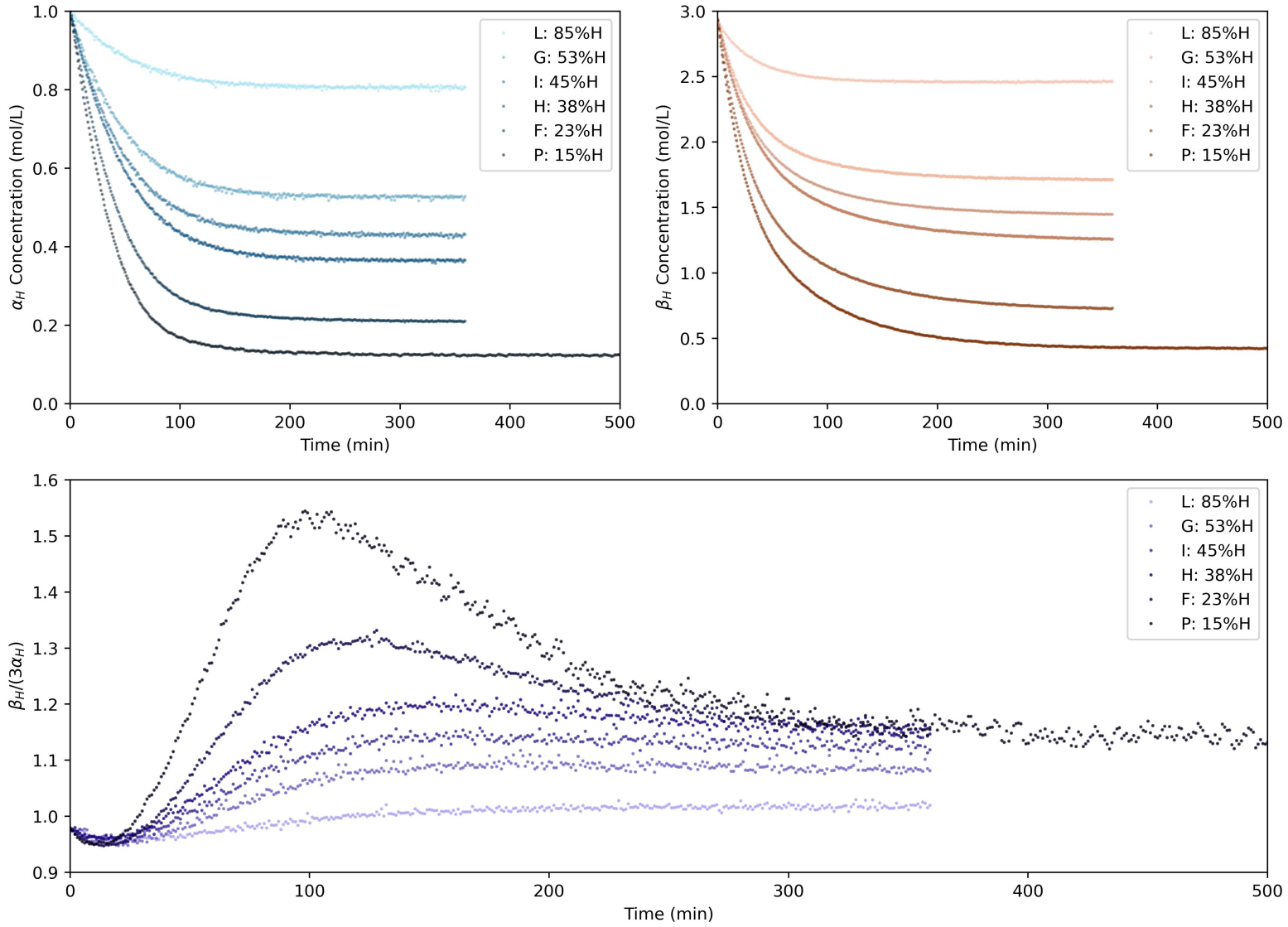

Figure 5.4: Top: Evolution of α_H and β_H concentrations (blue and brown points, respectively) from six ^{1}H NMR kinetics experiments "F", "G", "H", "I", "L", and "P", as described in Table 5.1. Bottom: Evolution of the ratio $\beta_H/(3\alpha_H)$ for this same set of experiments. The legend gives $\%H_2O$ of the solution at equilibrium. Initial alanine α and β composition for all experiments was $CHCH_3$, and the starting ratio $\beta_H/(3\alpha_H)$ was $\sim$0.997. Data are given in the Supplementary Materials.

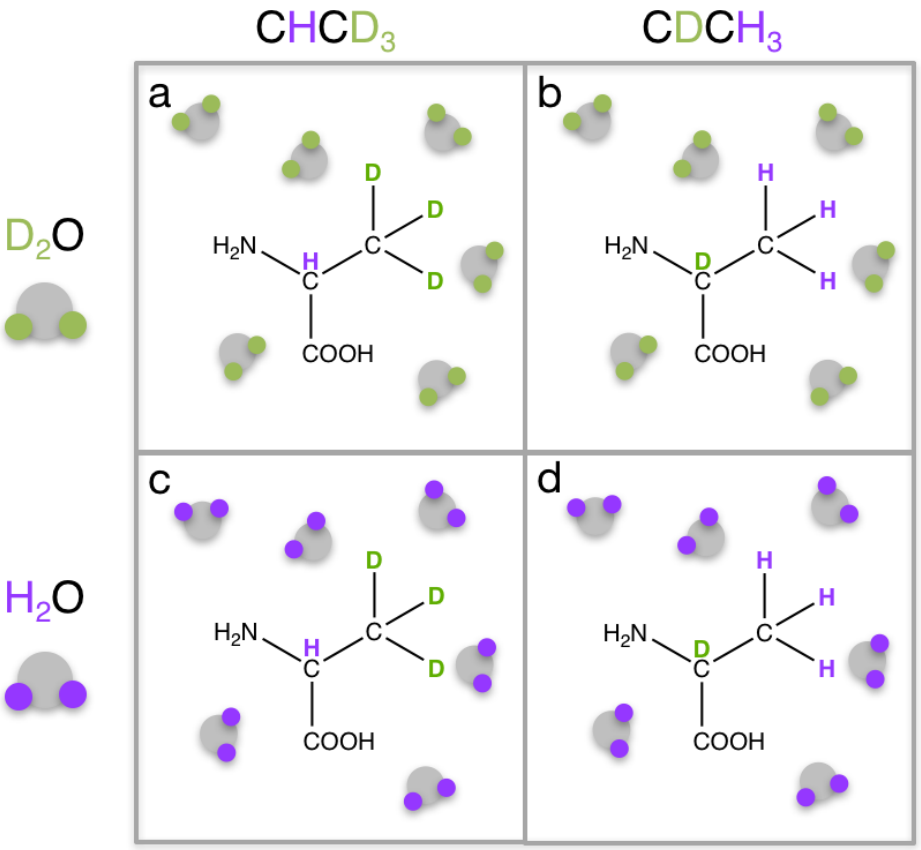

Figure 5.5: Approximate starting conditions of four kinetics experiments with selectively deuterium-labeled L-alanine: (a)"M", (b)"J", (c)"O", and (d)"N", also depicted in Figure 5.11.

5.4.1.3 Site-Specific D-Labeled Alanine Kinetics

Our third set of experiments was designed to evaluate the kinetics of exchange at one of the two (α or β) positions while the other position was already relatively close to equilibrium with water. We began with L-alanine labeled with 99% deuterium at either the α or β position. Each of these two deuterium-labeled alanine variants was reacted once in H_2O, and once in D_2O, resulting in a set of four experiments with approximate initial conditions as illustrated in Figure 5.5.

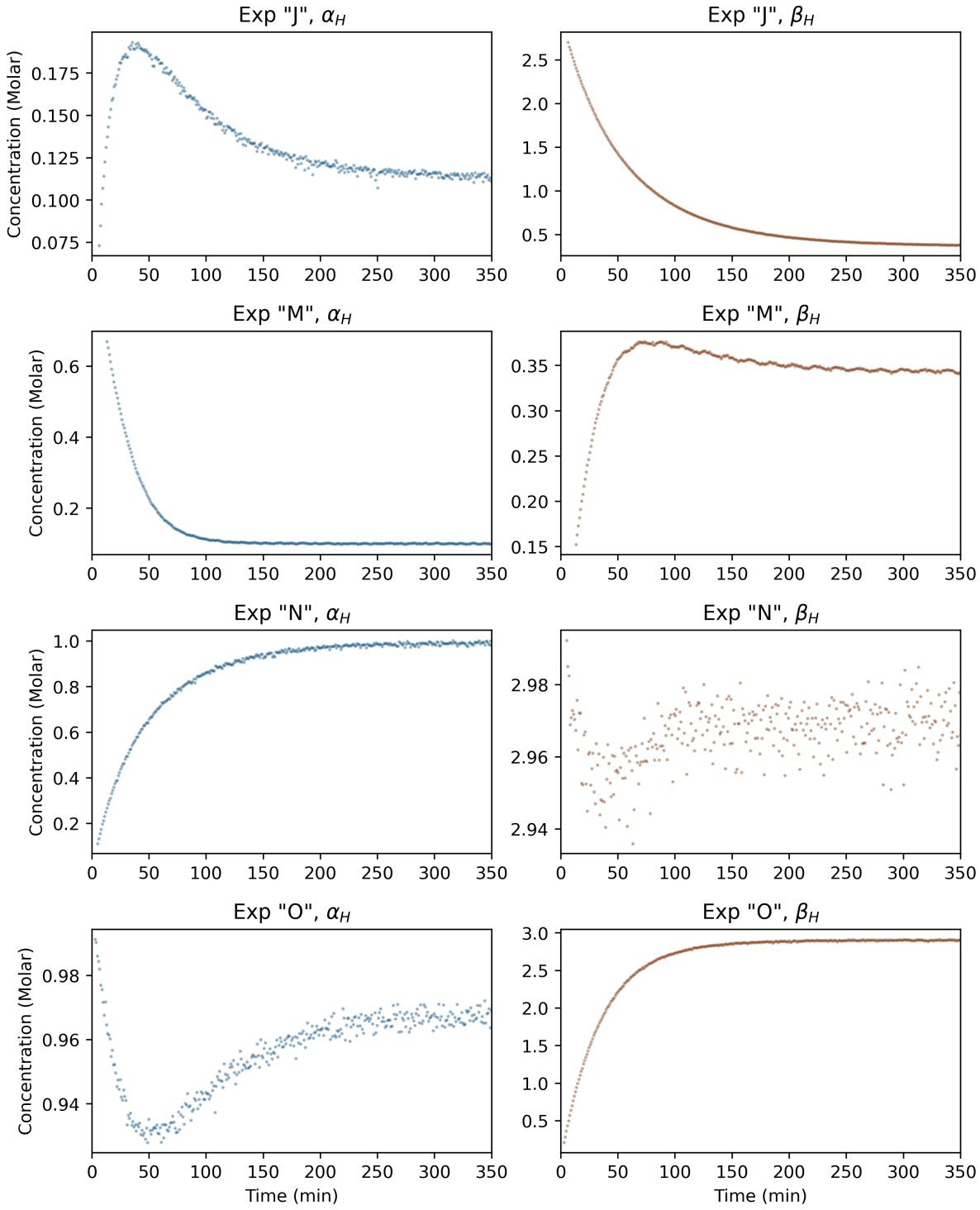

Figure 5.6: Evolution of α_H and β_H concentrations (blue and brown points, respectively) from four ^{1}H NMR kinetics experiments with selectively-deuterium-labeled L-alanine: "J", "N", "O", and "M", as depicted in Figure 5.5 and described in Table 5.1. Notes: (1) Vertical scales vary depending on α_H and β_H concentration ranges, particularly to highlight the transient overshoots in the species already close to equilibrium with solvent water, (2) The slight oscillation with ~20 minute period in "M" β_H originates from room temperature fluctuations due to A/C running at regular intervals during the experiment. Its effect on our modeling and conclusions is negligible.

These deuterium-labeled experiments are summarized thus:

- Experiment J: $CDCH_3$ in ∼85 % D_2O. The primary isotopic exchange reaction is the loss of H (gain of D) at the β site (Reaction 5.3 forward), there is a transient rise in the α signal as H leaks over from the β site, augmenting the contribution from ∼15 % H_2O in solution.
- Experiment N: $CDCH_3$ in 100 % H_2O. The primary isotopic exchange reaction is the loss of D (gain of H) α site (Reaction 5.1 backward), but there is a transient loss of β as D leaks over from the α site before being overwhelmed by H from the solution.
- Experiment M: $CHCD_3$ in ∼85 % D_2O. The primary isotopic exchange reaction is loss of H (gain of D) at the α site (Reaction 5.10 forward), but there is a transient rise of the β signal as H leaks over from the α site augmenting the contribution coming from the ∼15 % H_2O in solution.
- Experiment O: $CHCD_3$ in 100 % H_2O. The primary isotopic exchange reaction is the loss of D (gain of H) at the β site (Reaction 5.8 backward) but there is a transient loss of H from the α site as D leaks over from the β site before being overwhelmed by H from the solution.

We compared the initial rate (using the first 20 minutes of data collected) of α hydrogen loss in Figure 5.5a (Experiment "M", alanine with an initial composition of $^{12}CH^{12}CD_3$), to the analogous experiment with unlabeled alanine (Experiment "P") used in the second set of experiments (both experiments occur in 85:15 D_2O:H_2O). The initial composition of alanine is the only difference between these two experiments—in Figure 5.5b, the initial composition of alanine's α and β is $^{12}CH^{12}CD_3$, while in the unlabeled experiment, it is $^{12}CH^{12}CH_3$. Only the earliest portion of data collected is used, in part because this is the period during which the vast majority of α hydrogen loss occurs on alanine that has the original composition of either $^{12}CH^{12}CD_3$ (being converted to $^{12}CD^{12}CD_3$) or $^{12}CH^{12}CH_3$ (being converted to $^{12}CD^{12}CH_3$). As the reaction continues and more intermediate isotopic variants are formed, the HDX rates begin to deviate significantly from an exponential function. These data and fitted exponential functions are given in Figure 5.7a, plotted as the log of H concentration relative to the starting concentration for these two experiments (data are normalized to an initial value of 1). We observe a

substantial difference in α hydrogen loss rate between these two experiments, where α hydrogen loss from $^{12}CH^{12}CD_3$ is $\sim$1.4 times faster than α hydrogen loss from $^{12}CH^{12}CH_3$.

We also compared the rates of β hydrogen loss in two experiments, that is, Figure 5.5b (Experiment "J", alanine with an initial composition of $^{12}CD^{12}CH_3$), against the same analogous experiment with unlabeled alanine (Experiment "P"), as used in the α hydrogen loss comparison above. These data and fitted exponential functions are given in Figure 5.7b, where β hydrogen abundance is normalized to an initial value of 1.

We observe a difference in β hydrogen loss rate between these two experiments of similar magnitude, but it is in the opposite direction from that observed in α hydrogen loss, where β hydrogen loss from $^{12}CD^{12}CH_3$ is $\sim$1.4 times slower than β hydrogen loss from $^{12}CH^{12}CH_3$.

The site-specific deuterium labeling appears to have similar but opposite effects on the rate of hydrogen loss at the α and β carbon positions, that is, the rate of α hydrogen loss is *increased* by having deuterium in the β carbon site as opposed to hydrogen, while the rate of β hydrogen loss is *decreased* by having deuterium in the α carbon site as opposed to hydrogen. The faster α H loss observed in Experiment "M" relative to Experiment "P" is expected, if for no other reason than because the labeled $CHCD_3$ alanine has a nearby source of D. However, the slower β H loss observed in Experiment "J" relative to Experiment "P", however, must mean that the presence of D at the α site has an intrinsic retarding effect on the rate H loss from the β group, overwhelming any acceleration in H loss that might come from having a nearby D available at the α site. In other words, the rate constant for β exchange must faster for α=H than for α=D. But we cannot say from these data alone whether the rate constant for α exchange is really faster for β=D. We address this issue further below in the Modeling and Discussion section (5.5).

These experiments were inspired by somewhat similar experiments reported by Cooper where reaction rates were compared between starting compositions of variably deuterated alanine (Cooper, 1976). The distinction between Cooper's experiments and ours is that he measured the rate of transamination through the appearance of pyruvate, in addition to the rate of hydrogen exchange, while we have measured only the hydrogen exchange using the half-reaction alanine $\leftrightarrow$ pyruvate. Cooper found that the rate of transamination was faster

by a factor of 2.3 for alanine with α and β composition of $CHCH_3$, compared to the rate for alanine labeled with deuterium at the α site (i.e., $CDCH_3$). However, he found that labeling the β site with deuterium (i.e., $CHCD_3$), did not change the transamination rate. He found a similar factor of 2.4 difference in rate between β hydrogen loss from $CDCH_3$ and $CHCH_3$, with the α-deuterated variant resulting in the slower rate. This last result of his can be compared to our rate difference of 1.4 between β hydrogen loss from $CDCH_3$ and $CHCH_3$, which goes in the same direction. The reasons for the discrepancy in magnitude between our experiments and his are unknown. It is likely a matter of lower precision in his measurements given the technology available to him at the time. It may also have to do with the fact that Cooper was not running half-reactions but was running the full transamination ping-pong reaction, similar to what we did in Chapter 3 on carbon isotopes.

5.4.2 H/D Exchange Kinetics by ^{13}C NMR

Our final kinetics experiment, a single reaction observed using ^{13}C NMR, takes place in water with an isotopic composition close to 50:50 D_2O:H_2O, as described in Section 5.3.3.2. It begins with effectively all alanine present as $CHCH_3$, and ends with all eight isotopic species present. Figure 5.8 presents the relative abundances of each of the eight isotopic variants at two-minute intervals over the course of this reaction. As expected, the $^{12}CH^{13}CH_3$ species decays exponentially from its starting concentration, while the other seven species start at effectively zero abundance, each appearing to increase at a unique rate.

Note that this reaction is a closed system—total β carbon signal is conserved throughout the kinetics, i.e., loss of ^{13}C NMR signal from one of the eight species is always gained by another (see Figure 5.18 for the total integrated peak area of the β ^{13}C NMR region, which remains constant, despite some noise, throughout the experiment).

To our knowledge, the density and richness of the data presented here goes beyond any work previously published in isotope kinetics, and far surpasses any previous work on alanine. In general, isotopic kinetics studies are severely limited in time resolution, precision, or the number of isotopic species being tracked, and usually in all three. Here, the time resolution is fine relative to changes in isotopic composition, such that the quality of fits to the data

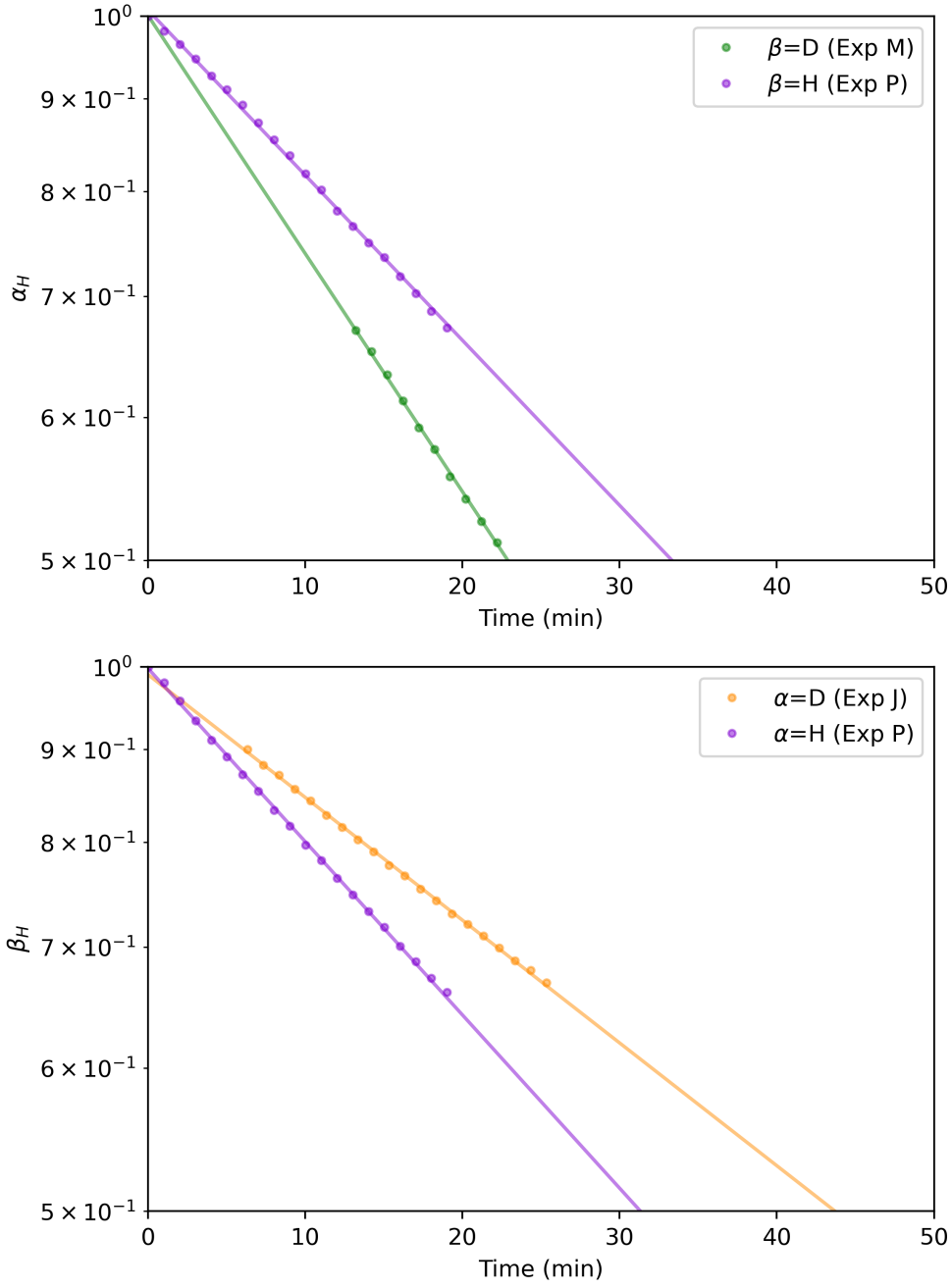

Figure 5.7: Comparing hydrogen loss rates for L-alanine, using the initial 20 minutes of data from kinetics experiments "P", "M", and "J" in ∼85:15 D_2O:H_2O. (a) α hydrogen loss rate in Experiments "M" ($^{12}CH^{12}CD_3$) and "P" ($^{12}CH^{12}CH_3$), denoted by purple circles (Experiment "P") and green circles (Experiment "M"). Lines are exponential fits to the data, y-axis is concentration in log scale, normalized to initial concentration. (b) β hydrogen loss rate in Experiments "J" ($^{12}CD^{12}CH_3$) and "P" ($^{12}CH^{12}CH_3$), denoted by purple circles (Experiment "P") and green circles (Experiment "J"). Lines are exponential fits to the data.

can easily be assessed. Measurement precision is qualitatively visible by the scatter of individual data points over the course of the reaction, and it is readily apparent that this degree of precision enables unprecedented tracking of each of the eight species throughout the course of the reaction.

This experiment proved to be the single most valuable kinetics experiment we conducted. The value of this experiment comes principally from the ability by ^{13}C NMR to distinguish and quantify through the course of reaction all eight isotopic variants of interest, which gives us the opportunity through modeling to determine rate constants for both intermolecular (between water and alanine) and intramolecular (within alanine) isotopic exchange, an impressive and exciting accomplishment we have not seen elsewhere for this reaction or for any other. Thus, in contrast to the ^{1}H kinetics experiments presented in the previous section, there are no previous studies with which we can compare our results, so we leave further discussion of this experiment, with added constraints from the ^{1}H kinetics experiments, to the modeling discussion below.

The design of this experiment was also key, as we chose to equilibrate unlabeled alanine, i.e., $CHCH_3$ with water having approximately 50:50 D_2O:H_2O composition. This final composition resulted in equilibrated amounts of the eight isotopic variants that vary by only a factor of $\sim$3. This factor of 3 is, of course, due to the multiplicity of three indistinguishable ways to configure four out of eight isotopic variants ($CHCH_2D$, $CDCH_2D$, $CHCHD_2$, and $CDCHD_2$), while the remaining four ($CHCH_3$, $CDCH_3$, $CHCD_3$ and $CDCD_3$) can exist in only one distinct configuration. That is, there is no way to construct this experiment to result in greater equality across all isotopic variants. The usefulness of this fact is that it enables us to acquire high-precision measurements on the equilibrated solution, as presented below, in Section 5.4.3, with up to ‰-level precision. An equilibrated solution with vastly different proportions of the eight isotopic species would result in greatly reduced measurement precision on the low abundance species, which would be reflected in the resulting precision on the calculated equilibrium constants that depend on the relative abundances of these species.

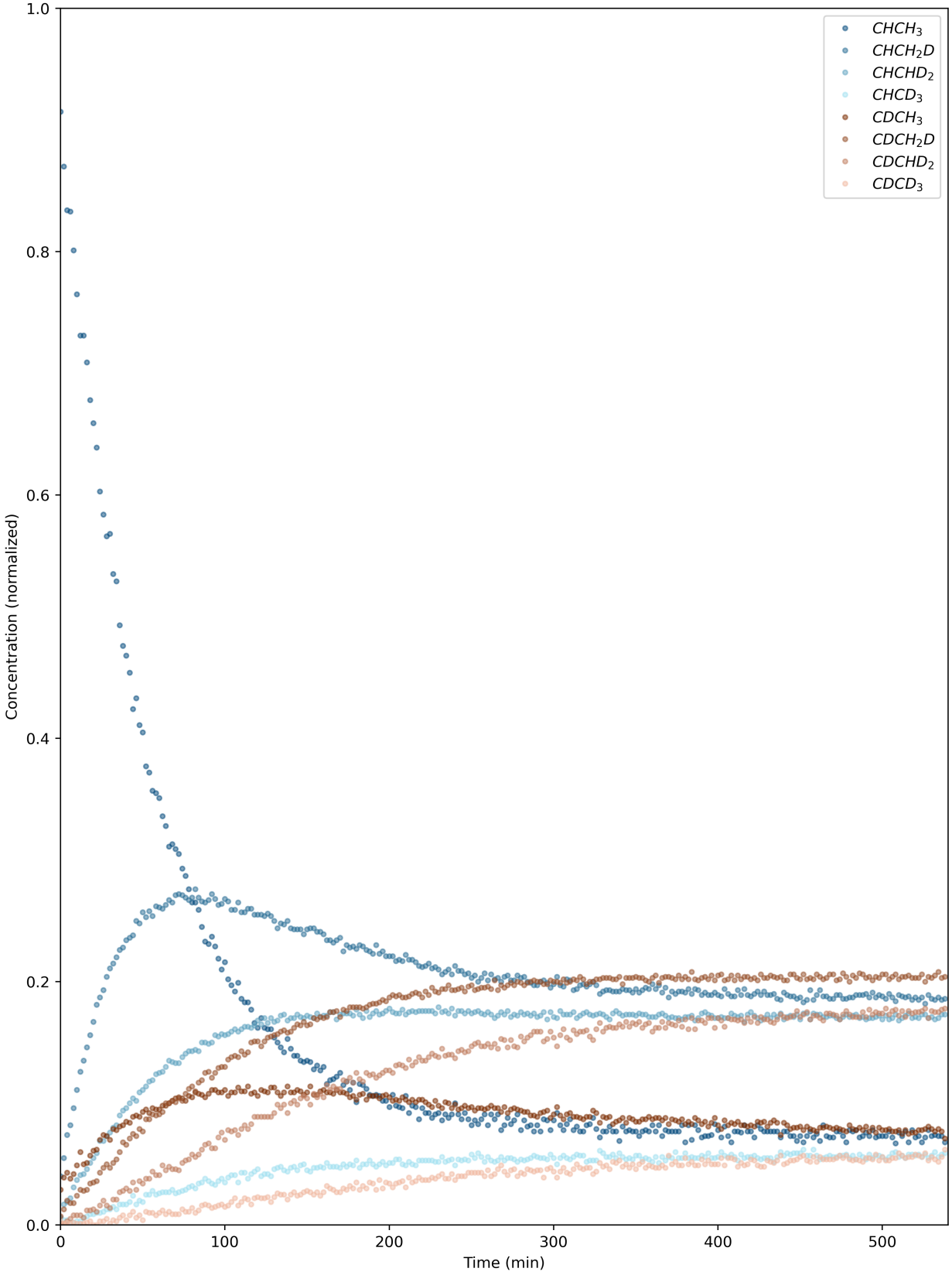

Figure 5.8: Time series of the eight isotopic species in Experiment "Q" using single-scan ^{13}C NMR acquisitions taken every two minutes for 540 minutes (9 hours) of reaction time. The observed reaction began with >99 % $^{12}CH^{13}CH_3$, and evolved towards the equilibrium distribution of eight isotopic variants reported in Tables 5.2 and 5.3. Isotopic species with H in the α site are shown in shades of blue (darkest = 4H, 0D $\rightarrow$ lightest = 1H, 3D) and those with D in the α site are shown in shades of brown (darkest = 3H, 1D $\rightarrow$ lightest = 0H, 4D).

5.4.3 H/D Isotopic Equilibrium Fractionation

Once the isotopic exchange reaction in Experiment "Q" (described in Section 5.4.2 above) reached equilibrium with water (over the following two days), we quantified the eight alanine isotopic variants at significantly higher precision than was used for the kinetics portion of Experiment "Q". Here we present results of our ^{13}C NMR equilibrium observations of the intermolecular hydrogen isotope fractionation between water and the two ALT enzyme-mediated exchangeable α- and β-carbon sites in alanine, as well as the intramolecular hydrogen isotope fractionation between the same α and β sites.

We used the fitted peak areas from the observations on equilibrated Experiment "Q" to calculate the relative abundances of each of the eight isotopic species, and calculated the ten equilibrium constants between alanine and solvent water using the equations given below. These reactions would proceed diagonally in Figure 5.9 and are given below, with solitary "D" and "H" representing these isotopes in solvent water:

$$\mathrm{CHCH_3 + D} \underset{k_{1_b}}{\overset{k_{1_f}}{\rightleftharpoons}} \mathrm{CDCH_3 + H} \tag{5.1}$$

$$\mathrm{CHCH_3 + D} \underset{k_{2_b}}{\overset{k_{2_f}}{\rightleftharpoons}} \mathrm{CHCH_2D + H} \tag{5.2}$$

$$\mathrm{CDCH_3 + D} \underset{k_{3_b}}{\overset{k_{3_f}}{\rightleftharpoons}} \mathrm{CDCH_2D + H} \tag{5.3}$$

$$\mathrm{CHCH_2D + D} \underset{k_{4_b}}{\overset{k_{4_f}}{\rightleftharpoons}} \mathrm{CDCH_2D + H} \tag{5.4}$$

$$\mathrm{CHCH_2D + D} \underset{k_{5_b}}{\overset{k_{5_f}}{\rightleftharpoons}} \mathrm{CHCHD_2 + H} \tag{5.5}$$

$$\mathrm{CDCH_2D + D} \underset{k_{6_b}}{\overset{k_{6_f}}{\rightleftharpoons}} \mathrm{CDCHD_2 + H} \tag{5.6}$$

$$\mathrm{CHCHD_2 + D} \underset{k_{7_b}}{\overset{k_{7_f}}{\rightleftharpoons}} \mathrm{CDCHD_2 + H} \tag{5.7}$$

$$\mathrm{CHCHD_2 + D} \underset{k_{8_b}}{\overset{k_{8_f}}{\rightleftharpoons}} \mathrm{CHCD_3 + H} \tag{5.8}$$

$$\mathrm{CDCHD_2 + D} \underset{k_{9_b}}{\overset{k_{9_f}}{\rightleftharpoons}} \mathrm{CDCD_3 + H} \tag{5.9}$$

$$\mathrm{CHCD_3 + D} \underset{k_{10_b}}{\overset{k_{10_f}}{\rightleftharpoons}} \mathrm{CDCD_3 + H} \tag{5.10}$$

Each of these reactions corresponds to an equilibrium constant that depends on the concentrations of the products and reactants:

$$K_1 = \frac{[\mathrm{CDCH_3}][\mathrm{H}]}{[\mathrm{CHCH_3}][\mathrm{D}]} \tag{5.11}$$

$$K_2 = \frac{[\mathrm{CHCH_2D}][\mathrm{H}]}{[\mathrm{CHCH_3}][\mathrm{D}]} \tag{5.12}$$

$$K_3 = \frac{[\mathrm{CDCH_2D}][\mathrm{H}]}{[\mathrm{CDCH_3}][\mathrm{D}]} \tag{5.13}$$

$$K_4 = \frac{[\mathrm{CDCH_2D}][\mathrm{H}]}{[\mathrm{CHCH_2D}][\mathrm{D}]} \tag{5.14}$$

$$K_5 = \frac{[\mathrm{CHCHD_2}][\mathrm{H}]}{[\mathrm{CHCH_2D}][\mathrm{D}]} \tag{5.15}$$

$$K_6 = \frac{[\mathrm{CDCHD_2}][\mathrm{H}]}{[\mathrm{CDCH_2D}][\mathrm{D}]} \tag{5.16}$$

$$K_7 = \frac{[\mathrm{CDCHD_2}][\mathrm{H}]}{[\mathrm{CHCHD_2}][\mathrm{D}]} \tag{5.17}$$

$$K_8 = \frac{[\mathrm{CHCD_3}][\mathrm{H}]}{[\mathrm{CHCHD_2}][\mathrm{D}]} \tag{5.18}$$

$$K_9 = \frac{[\mathrm{CDCD_3}][\mathrm{H}]}{[\mathrm{CDCHD_2}][\mathrm{D}]} \tag{5.19}$$

$$K_{10} = \frac{[\mathrm{CDCD_3}][\mathrm{H}]}{[\mathrm{CHCD_3}][\mathrm{D}]} \tag{5.20}$$

The H and D in these equations are associated with water molecules in the solvent, present as a mixture of H_2O, HDO, and D_2O. Since they are all readily exchangeable on time scales of interest to this study, we can take them as simply H or D, with initial populations given by [H]=2×[H_2O] and [D]=2×[D_2O]. In other words, we are not, at this time, creating a model with explicit H_2O, HDO, and D_2O species. The degree to which the constructed models are able to describe kinetics and equilibria across a range of initial H_2O:D_2O compositions will provide a check on the validity of assuming we do not have to concern ourselves with the specific molecular-level source of the D and H. The equilibrium constants K_1–K_{10} thus determined are given in Table 5.2. The large values of K_2 and K_3 and the small values of K_8 and K_9 simply result from multiplicity; i.e., there are three targets for the forward direction in Reactions 2 and 3, and only one for the backward direction, and three targets for the backward direction of Reactions 8 and 9, but only one for the forward direction.

The fractionations (in ‰) associated with these reactions are also given in parentheses in Table 5.2. These are just $1000ln(K_i)$ values, but also have the multiplicity effects removed (where indicated in the table) so that the intrinsic enthalphic vibrational effects are clear. For all HDX reactions taking place between the α site and water in the homogenized solvent pool (i.e., Reactions 1, 4, 7 and 10) the fractionations are positive, meaning that D prefers the α site in alanine over a solvated water molecule in the solvent pool. For HDX reactions taking place between the β site and water, the first two additions of a D atom on the β site (i.e., Reactions 2, 3, 5, and 6) have negative values, indicating that D would prefer to bind to a water molecule over the β site in alanine. However, if there are already two D atoms at the β site, the third D atom prefers to "clump" at the β site over a water molecule. These preferences hold regardless of whether the hydrogen atom bound to the α carbon is H (Reactions 2, 5, and 8) or D (Reactions 3, 6, and 9).

We also consider the possibility of "direct" exchange reactions between the α and β sites without the involvement of the homogenized solvent pool. Such reactions are useful for comparing with theoretical calculations because there is no solvent explicitly present in the reaction. Molecular modeling of the solvent is nontrivial, requiring sampling of many solvent structures to demonstrate convergence. Considering an isotopic swap within a single alanine molecule

suffers less from errors that may arise from modeling a reaction with completely different reactant and product environments. These reactions would proceed horizontally in Figure 5.9 and are as follows:

$$\mathrm{CHCH_2D} \underset{k_{\alpha\beta_{1_b}}}{\overset{k_{\alpha\beta_{1_f}}}{\rightleftharpoons}} \mathrm{CDCH_3} \tag{5.21}$$

$$\mathrm{CHCHD_2} \underset{k_{\alpha\beta_{2_b}}}{\overset{k_{\alpha\beta_{2_f}}}{\rightleftharpoons}} \mathrm{CDCH_2D} \tag{5.22}$$

$$\mathrm{CHCD_3} \underset{k_{\alpha\beta_{3_b}}}{\overset{k_{\alpha\beta_{3_f}}}{\rightleftharpoons}} \mathrm{CDCHD_2} \tag{5.23}$$

Each of these reactions corresponds to an equilibrium constant that depends on the concentrations of the products and reactants:

$$K_{\alpha\beta_1} = \frac{K_1}{K_2} = \frac{K_4}{K_3} = \frac{[\mathrm{CDCH_3}]}{[\mathrm{CHCH_2D}]} \tag{5.24}$$

$$K_{\alpha\beta_2} = \frac{K_4}{K_5} = \frac{K_7}{K_6} = \frac{[\mathrm{CDCH_2D}]}{[\mathrm{CHCHD_2}]} \tag{5.25}$$

$$K_{\alpha\beta_3} = \frac{K_7}{K_8} = \frac{K_{10}}{K_9} = \frac{[\mathrm{CDCHD_2}]}{[\mathrm{CHCD_3}]} \tag{5.26}$$

The two remaining isotopic species contain either all hydrogen or all deuterium. We can write these two species in a single reaction by describing the tendency for deuterium in the α carbon site to prefer clumping with an alanine molecule containing all deuterium versus all hydrogen at the β carbon site:

$$\mathrm{CHC\mathbf{D}_3 + C\mathbf{D}CH_3} \underset{k_{\alpha\beta_{4_b}}}{\overset{k_{\alpha\beta_{4_f}}}{\rightleftharpoons}} \mathrm{C\mathbf{D}C\mathbf{D}_3 + CHCH_3} \tag{5.27}$$

The corresponding equilibrium constant for this reaction is:

$$K_{(\alpha\beta)_4} = \frac{K_{10}}{K_1} = \frac{[\mathrm{CDCD_3}][\mathrm{CHCH_3}]}{[\mathrm{CHCD_3}][\mathrm{CDCH_3}]} \tag{5.28}$$

The measured and predicted (refer to Section 2.7) values for these equilibria indicating H/D site preferences between α and β carbon sites are given in Table 5.3. When multiplicity is accounted for, it is apparent that there is an enthalpic preference for D to sit at the α site in all cases, however the preference is significantly less for $K_{\alpha\beta_3}$ than for $K_{\alpha\beta_1}$ and $K_{\alpha\beta_2}$.

Positive values indicate deuterium preference for the α position. Recall from Section 2.7 Table 2.9 that we predicted, using density functional electronic structure calculations, a 169.9 ‰ site preference for deuterium in the α position in the case where all remaining exchangeable atoms are hydrogen and with ^{13}C in the beta position as done in these experiments. This value decreases by ∼16 ‰ per deuterium atom replacing hydrogen in the β position (i.e., the effect of deuterium clumping at the β position). We measure a site preference of ∼168 ‰ for deuterium in the α position for Reaction 5.24, in almost perfect agreement with the prediction of 170 ‰. In contrast, our measured site preference for clumped species given by Reactions 5.25 and 5.26 deviate significantly from the predicted values, where the presence of a single deuterium atom in the β position produces an even stronger deuterium site preference for the α position of ∼191 ‰, while the presence of two deuterium atoms in the β position weakens the site preference to ∼74‰. The preference for clumping of deuterium in both α and β positions within the same molecule (i.e., Reaction 5.26) is predicted to be negligible (∼0–1 ‰), and measured to be ∼-11 ‰, i.e., a slight anti-clumping preference extending beyond two bonds. The difference between theoretical predictions and experiment would be expected given all the issues involved in predicting H/D fractionations (non-Born-Oppenheimer effects, anharmonicity, barrier tunneling and recrossing, etc., Richet et al., 1977).

Table 5.2: Measured Equilibrium Constants, K_i, for Reactions 5.1–5.10, as depicted in Figure 5.9.

	K_i	$1000ln(K_i)$ (‰)
α (β=H or D)		
K_1	1.144±0.003	(134.9±2.9)
K_4	1.165±0.001	(152.4±1.2)
K_7	1.165±0.001	(152.7±1.4)
K_{10}	1.132±0.003	(123.9±2.8)
β (α=H)		
K_2	2.904±0.010	(-32.7±3.4)*
K_5	0.962±0.001	(-39.0±1.4)
K_8	0.361±0.002	(78.9±5.2)*
β (α=D)		
K_3	2.955±0.015	(-15.2±5.1)*
K_6	0.962±0.001	(-38.7±0.6)
K_9	0.350±0.001	(50.1±4.1)*

* Value has multiplicity effects removed by dividing observed K_2 and K_3 by 3 and multiplying observed K_8 and K_9 by 3.

Table 5.3: Measured and predicted site preference equilibrium constants, $K_{\alpha\beta_i}$, for intramolecular exchange Reactions 5.21, 5.22, and 5.23, as depicted in Figure 5.9.

	$K_{\alpha\beta_i}$	$1000ln(K_{\alpha\beta_i})$ (‰)	Predicted (‰)
$K_{\alpha\beta_1}$	1.182*	167.6±2.2*	170
$K_{\alpha\beta_2}$	1.211	191.4±0.3	154
$K_{\alpha\beta_3}$	1.077*	73.8±2.6*	139
$K_{\alpha\beta_4}$	0.990	-11.0	0

* Value has multiplicity effects removed by multiplying observed $K_{\alpha\beta_1}$ by 3 and dividing observed $K_{\alpha\beta_3}$ by 3

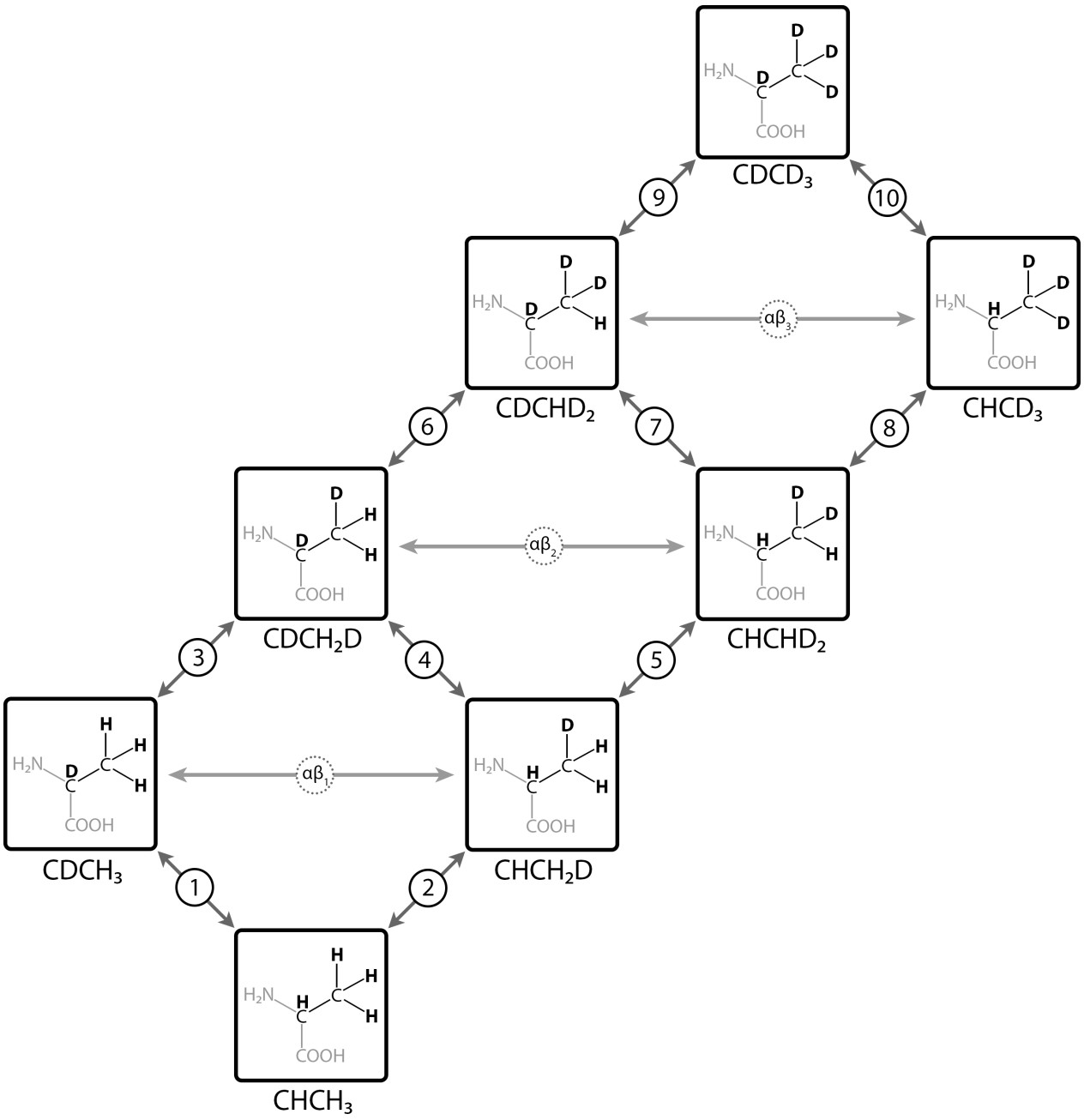

Figure 5.9: Graphical depiction of the system of ten intermolecular reactions (black diagonal arrows, circled) that convert the eight possible isotopic variants of alanine (i.e., the eight boxes) in our study from one to another by exchanging an H or D from either the α or β carbon position, and the three intramolecular reactions (gray horizontal arrows, circled) that swap an H and a D between the α and β carbon position. Forward reactions (up) remove a D from the solvent and emit an H to the solvent. Backward reactions (down) remove an H from the solvent and emit a D to the solvent. (Interaction with solvent for Reactions 1–10 is implicit in this diagram.)

5.5 Modeling and Discussion

It is readily apparent that the experimental data collected through the series of experiments presented above gives us an unprecedented look at this enzyme-mediated isotope exchange system, made possible through a combination of experimental capabilities that have advanced since earlier work was performed (see literature review in Chapter 4) and through experimental design choices. Perhaps less obvious is how to make sense of this dataset, and what useful information can be extracted from it, both with respect to this particular system and potentially to isotopic exchange systems more generally. To aid in this interpretive effort, we built a phenomenological kinetic model that allows determination of the reaction rates that convert each one of the eight isotopic variants into the others via H/D exchange with solvent water.

5.5.1 A Kinetic Model of Enzyme-Catalyzed Exchange Processes

Our model assumes that the eight isotopic alanine species populations follow a phenomenological rate law as given in the following ordinary differential equations, constructed using the isotope exchange Equations 5.1–5.10 and 5.21–5.23:

$$\begin{aligned}\frac{d[\mathrm{CHCH_3}]}{dt} &= +k_{1_b}[\mathrm{CDCH_3}][\mathrm{H}] - k_{1_f}[\mathrm{CHCH_3}][\mathrm{D}] \\ &+ k_{2_b}[\mathrm{CHCH_2D}][\mathrm{H}] - k_{2_f}[\mathrm{CHCH_3}][\mathrm{D}]\end{aligned} \tag{5.29}$$

$$\begin{aligned}\frac{d[\mathrm{CHCH_2D}]}{dt} &= +k_{2_f}[\mathrm{CHCH_3}][\mathrm{D}] - k_{2_b}[\mathrm{CHCH_2D}][\mathrm{H}] \\ &+ k_{4_b}[\mathrm{CDCH_2D}][\mathrm{H}] - k_{4_f}[\mathrm{CHCH_2D}][\mathrm{D}] \\ &+ k_{5_b}[\mathrm{CHCHD_2}][\mathrm{H}] - k_{5_f}[\mathrm{CHCH_2D}][\mathrm{D}] \\ &- k_{\alpha\beta_{1_f}}[\mathrm{CHCH_2D}] + k_{\alpha\beta_{1_b}}[\mathrm{CDCH_3}]\end{aligned} \tag{5.30}$$

$$\begin{aligned}\frac{d[\mathrm{CHCHD_2}]}{dt} &= +k_{5_f}[\mathrm{CHCH_2D}][\mathrm{D}] - k_{5_b}[\mathrm{CHCH_2D}][\mathrm{H}] \\ &+ k_{7_b}[\mathrm{CDCHD_2}][\mathrm{H}] - k_{7_f}[\mathrm{CHCHD_2}][\mathrm{D}] \\ &+ k_{8_b}[\mathrm{CHCD_3}][\mathrm{H}] - k_{8_f}[\mathrm{CHCHD_2}][\mathrm{D}] \\ &- k_{\alpha\beta_{2_f}}[\mathrm{CHCHD_2}] + k_{\alpha\beta_{2_b}}[\mathrm{CDCH_2D}]\end{aligned} \tag{5.31}$$

$$\begin{aligned}\frac{d[\mathrm{CHCD_3}]}{dt} &= +k_{10_b}[\mathrm{CDCD_3}][\mathrm{H}] - k_{10_f}[\mathrm{CHCD_3}][\mathrm{D}] \\ &+ k_{8_f}[\mathrm{CHCHD_2}][\mathrm{D}] - k_{8_b}[\mathrm{CHCD_3}][\mathrm{H}] \\ &- k_{\alpha\beta_{3_f}}[\mathrm{CHCD_3}] + k_{\alpha\beta_{3_b}}[\mathrm{CDCHD_2}]\end{aligned} \tag{5.32}$$

$$\begin{aligned}\frac{d[\mathrm{CDCH_3}]}{dt} &= +k_{1_f}[\mathrm{CHCH_3}][\mathrm{D}] - k_{1_b}[\mathrm{CDCH_3}][\mathrm{H}] \\ &+ k_{3_b}[\mathrm{CDCH_2D}][\mathrm{H}] - k_{3_f}[\mathrm{CDCH_3}][\mathrm{D}] \\ &+ k_{\alpha\beta_{1_f}}[\mathrm{CHCH_2D}] - k_{\alpha\beta_{1_b}}[\mathrm{CDCH_3}]\end{aligned} \tag{5.33}$$

$$\begin{aligned}\frac{d[\mathrm{CDCH_2D}]}{dt} &= +k_{3_f}[\mathrm{CDCH_3}][\mathrm{D}] - k_{3_b}[\mathrm{CDCH_2D}][\mathrm{H}] \\ &+ k_{4_f}[\mathrm{CHCH_2D}][\mathrm{D}] - k_{4_b}[\mathrm{CDCH_2D}][\mathrm{H}] \\ &+ k_{6_b}[\mathrm{CDCHD_2}][\mathrm{H}] - k_{6_f}[\mathrm{CDCH_2D}][\mathrm{D}] \\ &+ k_{\alpha\beta_{2_f}}[\mathrm{CHCHD_2}] - k_{\alpha\beta_{2_b}}[\mathrm{CDCH_2D}]\end{aligned} \tag{5.34}$$

$$\begin{aligned}\frac{d[\mathrm{CDCHD_2}]}{dt} &= +k_{6_f}[\mathrm{CDCH_2D}][\mathrm{D}] - k_{6_b}[\mathrm{CDCHD_2}][\mathrm{H}] \\ &+ k_{7_f}[\mathrm{CHCHD_2}][\mathrm{D}] - k_{7_b}[\mathrm{CDCHD_2}][\mathrm{H}] \\ &+ k_{9_b}[\mathrm{CDCD_3}][\mathrm{H}] - k_{9_f}[\mathrm{CDCHD_2}][\mathrm{D}] \\ &+ k_{\alpha\beta_{3_f}}[\mathrm{CHCD_3}] - k_{\alpha\beta_{3_b}}[\mathrm{CDCHD_2}]\end{aligned} \tag{5.35}$$

$$\begin{aligned}\frac{d[\mathrm{CDCD_3}]}{dt} &= +k_{10_f}[\mathrm{CHCD_3}][\mathrm{D}] - k_{10_b}[\mathrm{CDCD_3}][\mathrm{H}] \\ &+ k_{9_f}[\mathrm{CDCHD_2}][\mathrm{D}] - k_{9_b}[\mathrm{CDCD_3}][\mathrm{H}]\end{aligned} \tag{5.36}$$

$$
\begin{aligned}
&\frac{d[D_{Solution}]}{dt} = \\
&+ k_{1_b}[\mathrm{CDCH_3}][\mathrm{H}] - k_{1_f}[\mathrm{CHCH_3}][\mathrm{D}] \\
&+ k_{2_b}[\mathrm{CHCH_2D}][\mathrm{H}] - k_{2_f}[\mathrm{CHCH_3}][\mathrm{D}] \\
&+ k_{3_b}[\mathrm{CDCH_2D}][\mathrm{H}] - k_{3_f}[\mathrm{CDCH_3}][\mathrm{D}] \\
&+ k_{4_b}[\mathrm{CHCH_2D}][\mathrm{H}] - k_{4_f}[\mathrm{CDCH_2D}][\mathrm{D}] \\
&+ k_{5_b}[\mathrm{CHCHD_2}][\mathrm{H}] - k_{5_f}[\mathrm{CHCH_2D}][\mathrm{D}] \\
&+ k_{6_b}[\mathrm{CDCHD_2}][\mathrm{H}] - k_{6_f}[\mathrm{CDCH_2D}][\mathrm{D}] \\
&+ k_{7_b}[\mathrm{CHCHD_2}][\mathrm{H}] - k_{7_f}[\mathrm{CDCHD_2}][\mathrm{D}] \\
&+ k_{8_b}[\mathrm{CHCHD_2}][\mathrm{H}] - k_{8_f}[\mathrm{CHCD_3}][\mathrm{D}] \\
&+ k_{9_b}[\mathrm{CDCHD_2}][\mathrm{H}] - k_{9_f}[\mathrm{CDCD_3}][\mathrm{D}] \\
&+ k_{10_b}[\mathrm{CHCD_3}][\mathrm{H}] - k_{10_f}[\mathrm{CDCD_3}][\mathrm{D}]
\end{aligned}
\tag{5.37}
$$

$$
\frac{d[H_{Solution}]}{dt} = -\frac{d[D_{Solution}]}{dt} \tag{5.38}
$$

We wrote a code to integrate these kinetic equations, given the initial concentrations of each isotopic variant and the thirteen rate constants $\mathrm{k}_{i_{f,b}}$. To fit the parameters, this code was embedded into an optimization code (originally developed in Benjamin Hay's group at Pacific Northwest National Laboratory in the early 1990s) derived from the downhill simplex code "AMOEBA" in Press et al., 1986. The code from Hay's group is very easily configured for any optimization problem as it has system calls to a script that runs the model and calculates the error. In other words, one does not need to interface FORTRAN subroutines to run the model and calculate the error. While we originally wrote the code to integrate the differential equations 5.29-5.36 and to calculate the misfit in Python, this proved to be very slow, and so we translated the code into FORTRAN, which ran about three orders of magnitude faster. Both the Python and FORTRAN versions are given in the Supporting Information.

We configured the code to fit up to thirteen $k_{n_{f,b}}$ parameters. While there are twenty-three parameters in the model (thirteen rate constants and ten equilibrium constants), the corresponding $k_{n_{b,f}}$ parameters are calculated from the

equilibrium constants, determined by the relationship $K_n = k_{n_f} / k_{n_b}$. The code reads in an initial guess and then runs a specified number of fits from random perturbations, of a specified magnitude, about the initial guess. For each "simulation" of the reaction, we determine a "misfit" value, defined as the sum of the absolute value of the difference (experiment - model) for each data point for each of the eight species. The downhill simplex algorithm finds parameters that minimize the misfit, producing a "local" minimum. The details of the implementation fitting procedure are discussed further in Section 5.5.3. The merits (and deficiencies) of the downhill simplex method are discussed in Numerical Recipes (Press et al., 1986). This method creates a set of $n + 1$ points in an n-dimensional parameter space (a "simplex") and then specifies a series of moves on the vertex with the worst fit to the data, mainly projecting new guesses along a line through this vertex and the centroid of the n remaining vertices (all with better fits to the data). If no better fit can be found, the whole simplex shrinks towards the point having the best fit to the data. The process terminates when the values of the simplex differ by some specified amount, in our case 0.0001. This method is appropriate because it is particularly robust when presented with parameters whose initial values are highly uncertain, and when gradients of the parameters with respect to the misfit are expensive. In our case evaluation of the misfit involves running a simulation and cannot be expressed as an analytical function. Thus the gradient would have to be calculated numerically by evaluating $\frac{\partial \chi}{\partial p_i}$ where χ is the misfit and p_i is the ith parameter. Numerical gradients can introduce noise into the objective function to be optimized that can cause problems for higher-order optimization methods (see discussion in Numerical Recipes, Press et al., 1986).

For each solution we obtained (which locates a "local" minimum), we ran a Monte-Carlo-like procedure where each parameter was kicked with a random perturbation, scaled (by trial and error) to cause the first iteration of the downhill simplex method to have parameter sets that scored about a factor of three higher (i.e. worse) in the misfit parameter, and the downhill simplex method was allowed to proceed to reach an optimal parameter set. This was done over tens of thousands of iterations to search the parameter space as completely as reasonably possible. Through this procedure we attempt to find a "global" minimum.

Our kinetic model is shown in Figure 5.9. Each of the eight species are represented by a node in the diagram. The α=CH species are shown along the lower diagonal and the α=CD species on the upper diagonal. The four reactions converting α=CH to α=CD for β=CH_3, CH_2D, CHD_2, and CD_3 are given by the four northwest-southeast arrows. As we move upward (northeast) along each diagonal we go from β=CH_3 to β=CD_3, adding one D at each step. The scheme is arranged so that species with equivalent numbers of D are aligned horizontally (for example $CDCH_3$ and $CHCH_2D$ are level with each other). The horizontal reactions $K_{\alpha\beta_1}$-$K_{\alpha\beta_3}$ allow for "direct" intramolecular exchange of H/D between the α site and the β site without involving the homogenized solvent pool. Note that this homogenized solvent pool is implicit in Figure 5.9, but is explicitly present in the corresponding equations. These intramolecular reactions are unlikely to be elementary in the sense that they could involve intermediates that do not appear in the reactions. These intermediates could be bound to internal solvent water molecules, the co-enzyme, the enzyme itself, or a combination of all or a subset of these. The main point is that the D/H can be exchanged on the alanine without directly changing the bulk solution composition. While the ALT crystal structures determined by x-ray diffraction provided a major update to our understanding of the molecular mechanisms involved in the function of this enzyme, there are still many unknowns when it comes to the dynamics of this reaction. It is worth saying a bit more about what is presently known about the mechanism before getting further into the results of our model.

5.5.2 The Role of Lysine in Hydrogen-Deuterium Exchange

A lysine residue in ALT is now known to play a key role during transamination (Figure 4.4). During the reaction from alanine to pyruvate, this lysine's terminal amine ($:NH_2$) accepts the α hydrogen from alanine, resulting in a positive charge on the amine (NH_3^+). One of those three hydrogen atoms on the lysine terminal amine is then swapped onto the C4' position of the pyridoxal moiety, allowing the double bonds in the pyridoxal moiety to shift (forming a $-CH_2-$ where there was previously a $=CH-$). In the side-reaction where the double bond between the α carbon and the alanine nitrogen atom (forming a Schiff base) shifts to the bond between the alanine α and β carbons, the hydrogen atom lost from the β position is gained by a nearby base.

In his 1976 paper, nearly three decades before an enzyme crystal structure had

been determined for alanine transaminase (Zhou et al., 2004), Arthur Cooper entertained the idea that a lysine residue could, in theory, single-handedly mediate H/D exchange in both the α and β site. He believed it most probable that one or more bases must sit within the active site to assist in hydrogen removal in the α site, while β site hydrogen exchange occurred independently via a different basic residue in the enzyme active site. Nevertheless, he recognized the possibility that a single base could conceivably receive and donate hydrogen between both the α and β sites, and suggested that, if this were the case, it would have to be a lysine residue, for any other basic amino acid (e.g., histidine) would not be flexible enough to move between the various enzyme-substrate complexes. At the time of his writing, the mechanisms and structures of pyridoxal-5'-phosphate enzymes were largely unknown, and experimental work such as his was, in part, attempting to develop understanding of the molecular mechanisms at work in the active-sites of such pyridoxal-5'-phosphate-dependent enzymes. Only in the last couple decades has work on transaminase structures and mechanisms fully resolved Cooper's question about the presence of a single, versus multiple basic amino acid residues in the active site.

In the last two decades, several research groups have experimentally determined, using x-ray diffraction crystallography, the crystal structures of four ALT isoenzymes from radically different organisms: Pyrococcus Furiosus, Homo sapiens, Hordeum vulgare, and E. coli (i.e., an archea, a mammal, a plant, and a bacterium, respectively). These structures are published and publicly available on the Protein Data Bank (PDB): Pyrococcus furiosus (407 aa residues, Zhou et al., 2004), Homo sapiens ALT 2 (523 aa residues, Wisniewska et al., 2009), E. coli (405 aa residues, Peña-Soler et al., 2014), Hordeum vulgare (barley, 482 aa residues, Duff et al., 2012). Despite their significant variation in genetic encoding, as well as in sequence and number of amino acids, all of these isoenzymes share a very well-conserved active site structure and set of key amino acid residues. It is clear now from these structures that each active site contains only one basic amino acid residue, a lysine, which plays a direct role in stabilizing the transamination transition state. Thus the base receiving the hydrogen atom lost from the β position is indeed the same lysine residue that facilitates removal of the α hydrogen in transamination, which becomes NH_3^+ in this side reaction when it gains a β hydrogen (Figure 4.4). The approximate distance between the basic moiety on lysine and the

PLP carbon atom to which it temporarily binds during enzyme function can be determined from the ALT crystal structures and available tools on the Protein Data Bank (PDB) website. As we show in Figure 5.10, these distances are 1.60 Å (Homo sapiens ALT 2), 1.46 Å (E. coli), and 3.33 Å (Hordeum vulgare). The resolutions of these crystal structures (2.30 Å, 2.11 Å, and 2.71 Å, respectively) are of similar magnitude to these atomic distances, so these values are likely less precise than their significant figures suggest; nevertheless, they provide strong evidence of a tightly conserved mechanism, depending on a lysine residue from the enzyme interacting with the coenzyme and substrates. Unfortunately there does not yet exist a published crystal structure for Homo sapiens ALT 1, which is distinct from Homo sapiens ALT 2, with 496 amino acid residues instead of the 523 residues of ALT 2. In amino acid sequence, porcine ALT 1 and 2 are very similar to human ALT 1 and 2, respectively. There also does not yet exist a published crystal structure for either porcine ALT 1 or 2 (the ALT used in this study), but, given the well-conserved structure seen across the three radically diverse organisms with known crystal structures, it is safe to assume that the active site structures of Homo sapiens ALT 2 and porcine ALT 1 and 2 do not deviate significantly from these.

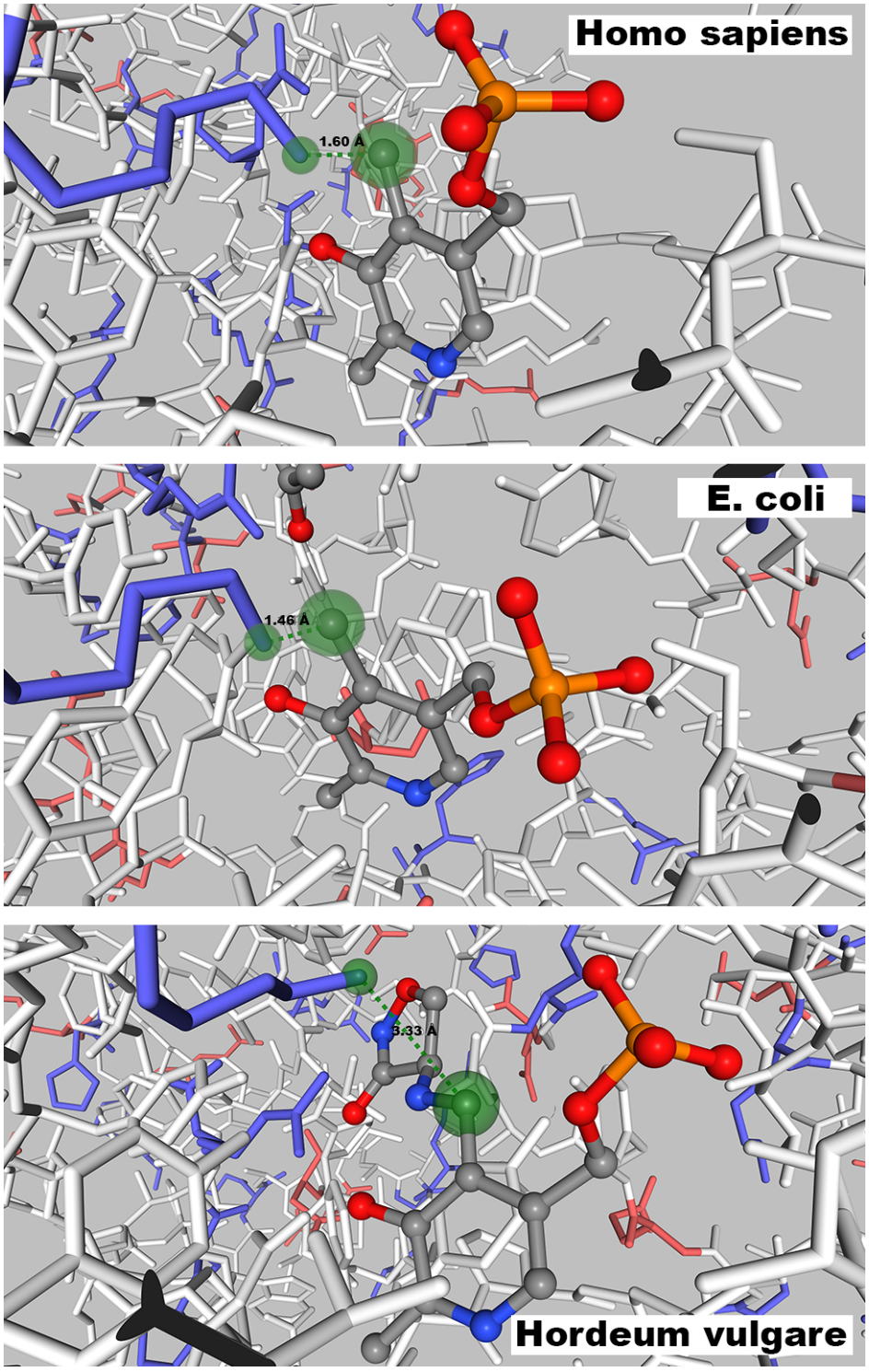

Figure 5.10: Views of the ALT enzyme active sites of Homo sapiens ALT 2, E. coli, and Hordeum vulgare (barley) showing the ball-and-stick PLP coenzyme (in the case of Horeum vulgare, PLP bound to cycloserine) with phosphate group in the center, and the lysine residue (purple) on left. Here basic amino acid residues are shown in purple, acidic residues are red, and the remaining residues are white. Structures are available on the Protein Data Bank. Distances between the basic moiety on lysine and the carbon atom on PLP to which it temporarily binds are determined from the crystal structure to be 1.60, 1.46, and 3.33 Å, respectively.

Strangely, Cooper came to the conclusion that a lysine residue was not the base receiving the hydrogen atom lost from the β position, based on his experimental results. Cooper proposed a test to distinguish between his two scenarios: if a single base was responsible for exchange at both sites, this should result in a "scrambling" (Cooper's terminology) of the hydrogen between the two sites; if two bases were responsible, such scrambling, he suggested, would not occur. Cooper believed that, for a reaction in which the α and β sites begin with known, contrasting hydrogen isotope compositions (e.g., $CDCH_3$), as they progress toward equilibrium with D_2O, this scrambling should be observable using the analytical tools available at that time (in his case, a Bruker WH 90 MHz NMR spectrometer). He tried this experiment with $CDCH_3$ in D_2O, but was not able to see any evidence of scrambling, and thus ruled out the single-base hypothesis. It is unclear whether his reported limit of detection of 5% incorporation at the α position of hydrogen derived from the β position was accurate. As can be seen in Figure 5.6 (Experiment "J", α_H) presented in Section 5.4.1.3 above, we ran an experiment analogous to his and observed temporary incorporation of hydrogen at the α position derived from the β position that is much greater than 5% relative to the final equilibrated incorporation, in apparent conflict with Cooper's results, but consistent with the now known enzyme structure.

In hindsight, Cooper's experiments may not have been able to provide a definitive answer; at the very least, the instruments and data processing methods he had available to him were not nearly as capable as those we have available to us today. Nevertheless, the experimental limitations of that day need not necessarily have hindered conceptual appreciation for the system at hand. An important question that Cooper did not discuss, but which naturally emerges when we consider the role of lysine in hydrogen exchange, is that of how the constrained water molecules inside the enzyme active site interact with the lysine residue, the co-enzyme, the enzyme-catalyzed transition state, the enzyme itself, and with the free water outside the enzyme. A related question is: What is the rate of exchange between this internal water and the exterior homogenized solvent pool? In other words, it is not simply the lysine residue that may play the role of mixing H and D between the α and β sites, but rather the tiny pool of water inside the enzyme active site, which may temporarily diverge in composition from the larger pool of well-mixed water outside the enzyme.

With the experimental tools now available to us, we are able to unambiguously observe the "scrambling", which we describe as "transient overshoots", between the α and β sites that Cooper sought. Communication between these two sites is most clearly realized in our four experiments beginning with selectively deuterated alanine (see Figures 5.5 and 5.6). In each of the four experiments of this type, "J", "M", "N", and "O", whichever alanine position began with a D/H composition most similar to the solvent water showed a temporary excursion away from final equilibrium, followed by an eventual return towards equilibrium (past its initial value which was slightly out of equilibrium with the final water composition). This initial trend towards greater disequilibrium indicates that the position closest to equilibrium with the solution is receiving hydrogen (in the case of the reactions in $\sim$85% D_2O) or deuterium (in the case of the reactions in $\sim$100% H_2O) from the position that is farther from equilibrium. The fact that the lysine residue is responsible for picking up hydrogen atoms from both sites means it can also donate the hydrogen atoms from one site to the other. Thus there will be some probability that a hydrogen atom taken from one position will be directly swapped onto the other. The alternative is that the hydrogen gets transferred to a water molecule inside the active site, which can either remain inside the active site to eventually undergo a further H/D exchange with lysine (and thus potentially exchange with another alanine), or go on to diffuse outside the enzyme and be lost to the free water molecules in solution outside of enzyme.

This raises several questions: (1) with what relative frequency does the lysine amine exchange its hydrogen atoms with the surrounding water, (2) what is the size of the pool of water within the enzyme active site with which the lysine can readily exchange, and (3) at what rate does the water inside the enzyme active site exchange with the external solution? These are questions about dynamical processes that cannot be answered with enzyme crystal structures alone, and which thus motivate both the experimental work presented above and the modeling work we present below. It is unknown to what extent a phenomenological model such as ours can conclusively answer these questions, but our unique data set gives us a rare opportunity to explore the possibilities and limitations of such a model.

5.5.3 Fitting Parameters for the Kinetic Model

The essential objectives of this section of our work are (1) to fit our experimental data presented in Section 5.4 with the kinetic model described by Equations 5.29–5.36, (2) to evaluate the extent to which our model might accurately represent the actual system, and, (3) using these fits, to look for patterns in the rate constants for this reaction network that may give insights into the factors governing the ALT-catalyzed hydrogen isotope exchange. The most basic null hypothesis is that all isotope exchange reactions occur with the same rate constant, independent of isotopic composition. A more reasonable one, however, is to allow for different rate constants for the chemically distinct α and β groups but to postulate that neither the α exchange rate constant nor the β exchange rate constant depends on the isotopic composition. Our strategy is to try out a number of models of increasing complexity and track the degree of misfit. If the system is well described by two rate constants for α and β then models that go beyond this should not give appreciably better fits to the data.

It is important to recognize that the approach we take here does not involve estimating rate constants by fitting exponential functions to early-stage reaction data, as in the previous work of, e.g., Cooper, 1976 (as we show in Figure 4.3) and Golichowski et al., 1977. Instead, we are globally fitting measured populations of the isotopic variants from the beginning to the end of each experiment. The movement of protons and deuterons from a bulk solvent into an enzyme constitutes an isotope-dependent transport problem in complex large-scale protein structure which we are ignoring in this formulation. It is not a straightforward task to figure out how the problem would simplify to some effective transport rate. Such transport effects could be influencing our rate constants in unknown ways that could require some type of detailed structure-kinetic coupling analysis such as recently presented (Kratochvil et al., 2023) or, in a geochemical context, the work by Bourg et al., 2012. On the other hand, it is likely that the better the fit achieved with the model chosen here, the more it indicates that one need not worry too much about such complex transport problems in the interpretation of our data. The issue of how mesoscale structural complexity can be folded into the rate constants does serve as an example of why these models cannot be considered elementary.

There are also, of course, considerations for influences from smaller scale pro-

cesses. Our model implicitly assumes that alanine enters into the enzyme as one of the isotopic variants, reacts, and exits as another isotopic variant that is related to the entering variant by one of the thirteen reactions identified in Figure 5.9. The turnover number for ALT is on the order of 1 to 100 s^{-1} (depending on experimental conditions and isoenzyme, as reported by, e.g., Duff et al., 2012, Escalera-Fanjul et al., 2017), so our NMR measurements average over on the order of 10^2-10^4 conversions per sampling interval. If some of these conversions were to result in a double substitution (for example $CHCH_3$ + 2D $\rightleftharpoons$ $CHCHD_2$ + 2H), or a simultaneous solvent and intramolecular exchange (for example $CDCH_3$ + D $\rightleftharpoons$ $CHCHD_2$ + H), the effectiveness of our model could be compromised.

As described at the beginning of Section 5.5, the unknown variables in this model include thirteen rate constants: k_{1_f}, k_{2_f}, k_{3_f}, k_{4_f}, k_{5_f}, k_{6_f}, k_{7_f}, k_{8_f}, k_{9_f}, k_{10_f}, $k_{\alpha\beta_{1_f}}$, $k_{\alpha\beta_{2_f}}$, $k_{\alpha\beta_{3_f}}$, and ten equilibrium constants: K_1, K_2, K_3, K_4, K_5, K_6, K_7, K_8, K_9, K_{10}. The relative size of the solvent pool with respect to the alanine concentration is fixed by the concentration of alanine dissolved in solution (1 Molar), and the concentration of H and/or D in water (2×∼55.5 Molar). Note that, as shown in Equations 5.24-5.26, $K_{\alpha\beta_1}$-$K_{\alpha\beta_3}$ are derivative and can be calculated through ratios of K_1-K_{10}. Naturally, a key question in such an endeavor is the uniqueness of the solution; we want to monitor the extent to which our model might be underconstrained, or overfitted, with respect to our experimental data.

To accomplish this, we begin by fitting the experimental data constraining k_{1_f}-$k_{\alpha\beta_{3_f}}$ to a single parameter representing a global rate constant (k_f). However, in this model, the corresponding k_{i_b} values are *not* restricted to a single value, but instead are fixed by the measured K_i through the relation k_{i_b}=k_{i_f}/K_i. In other words, while all the forward reactions are the same single parameter, each backward reaction constant in the single-parameter fit is unique, determined by the separate measurements of all ten unique equilibrium constants. This one-parameter model provides both a reference point for more complex models and also a well-defined starting guess for fitting such models.

To further evaluate our model, we also try this fitting procedure with three distinct subsets of our kinetics data (from experiments summarized in Table 5.1). The experiments included in these subsets, and the primary reactions receiving the highest flux in each are indicated in Figure 5.11. The subsets are

grouped as follows:

- Data Subset "Q": Data from Experiment "Q" only, which is our ^{13}C NMR kinetics experiment
- Data Subset "$QM_{\alpha}O_{\beta}$": Data from Experiments "Q", plus the α_H portion of data from Experiment "M", and the β_H portion of data from Experiment "O"
- Data Subset "QJMNO": All data for Experiments "Q", "J", "M", "N", and "O"

Fitting this 1-parameter model to the QJMNO dataset, we obtained a value of 2.3×10^{-4} $M^{-1}min^{-1}$ for the rate constant regardless of the initial guess, excluding initial guesses higher than 6×10^{-4} $M^{-1}min^{-1}$, where the model fails to converge at all, producing a nonphysical negative parameter during the optimization. Figure 5.13 shows the abundances of the isotopic species predicted by this 1-parameter model, plotted with the QJMNO dataset to which it was fitted.

Starting from this single-parameter model, we then explored a series of more complex models, to determine whether allowing for differences in the rate constants representing the thirteen isotope exchange reactions in Figure 5.9 would improve the level of agreement between model and experiment. This tests our null hypothesis that the system can be described terms of a rate constant for α exchange and a rate constant for β exchange. If so, we should see minimal improvements in allowing for variations beyond these models. In formulating and discussing these models it is useful to distinguish the intramolecular exchange reactions (with rate parameters $k_{\alpha\beta_{1f,b}}$-$k_{\alpha\beta_{3f,b}}$) from the solvent exchange reactions (with rate parameters $k_{1_{f,b}}$-$k_{10_{f,b}}$). This distinction can be intuited from a mechanistic point of view: we should expect the solvent-exchange rates and intramolecular exchange rates likely will occur on different time scales. Also, as just discussed, the SERPs are more likely to be influenced by transport effects associated with mesoscale structure. Notationally we denote these models as "NSMI", depending on the number N of the solvent exchange rate parameters (SERP) (N=1, 2, 3, 4, 10) and the number of intramolecular exchange rate parameters (IERP) (M=0, 1, or 2). In addition to adding IERPs to the 1P model just discussed, we progressively increased the number of SERPs from N=2 to

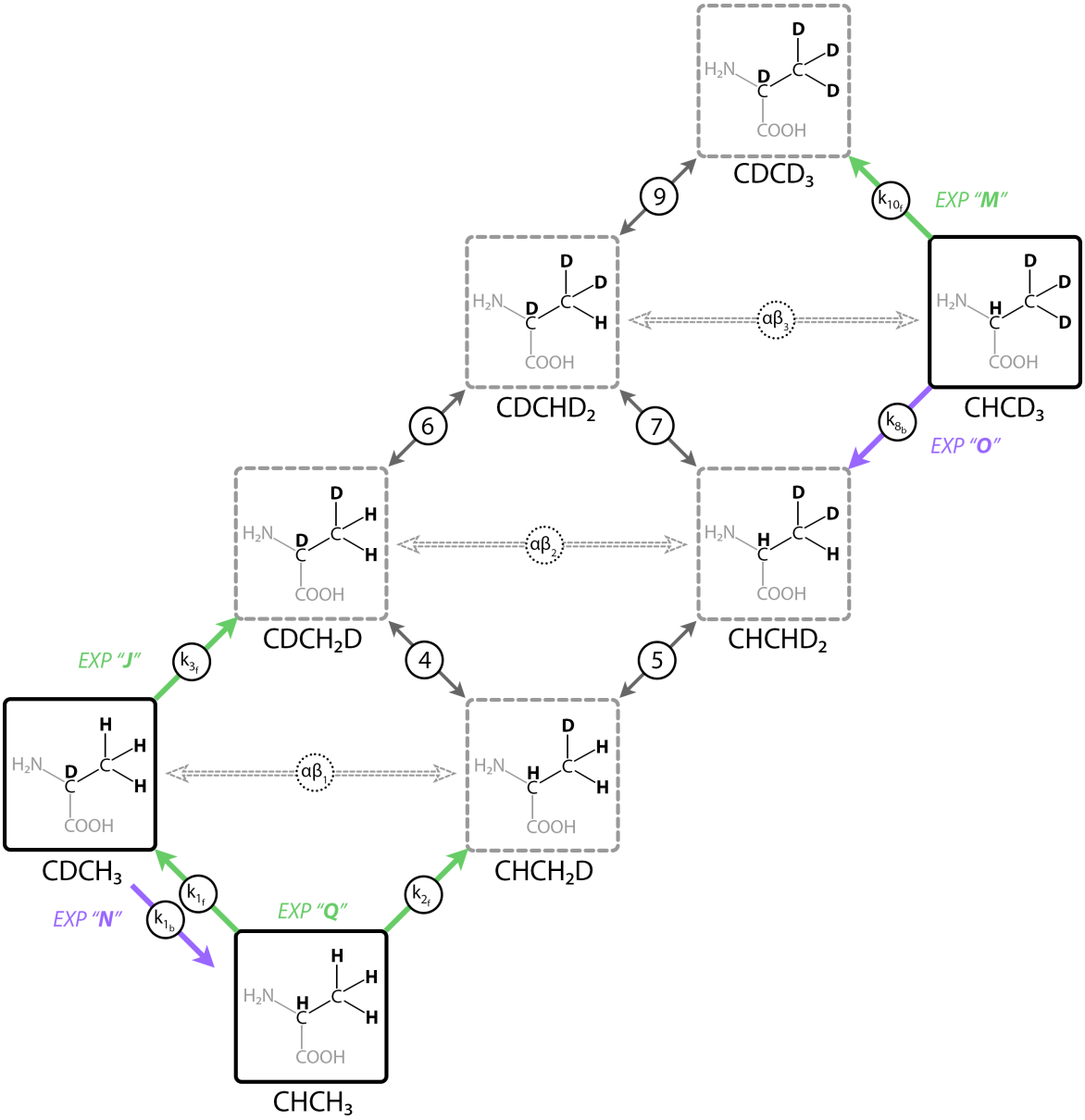

Figure 5.11: Graphical depiction of the system of ten intermolecular reactions that convert the eight possible isotopic variants of alanine in our study from one to another by exchanging an H or D from either the α or β carbon position, and three intramolecular reactions that swap an H and a D between the α and β carbon position, similar to Figure 5.9. Labeled green arrows (pointing upward) indicate reactions experiencing the highest flux for experiments "Q", "J" and "M". Labeled purple arrows (pointing downward) indicate reactions experiencing the highest flux for experiments "N" and "O". Solid black boxes surround isotopic variants that were used as initial pure starting points for these experiments. Dashed gray boxes surround remaining isotopic variants.

N=10, with and without IERPs. For models without any IERPs (M=0) we set $k_{\alpha\beta_{1_{f,b}}}$, $k_{\alpha\beta_{2_{f,b}}}$, and $k_{\alpha\beta_{3_{f,b}}}$ to zero. For models with a single IERP (M=1), we took two approaches: During the early fits we assigned $k_{\alpha\beta_{1_f}}$, $k_{\alpha\beta_{2_f}}$, and $k_{\alpha\beta_{3_f}}$ to the same single value. Later, after comparing our 8-species models with the 3-box model as described in Section 5.5.4, we realized it was better to assign the single parameter to $k_{\alpha\beta_{2_f}}$ and then assign $k_{\alpha\beta_{3_f}}=\frac{3}{2}k_{\alpha\beta_{2_f}}$ and $k_{\alpha\beta_{1_f}}=\frac{1}{2}k_{\alpha\beta_{2_f}}$ to better represent with the expected multiplicity effects. After realizing that the latter approach made more sense, this we re-ran 10,000 iterations of the downhill simplex parameter search for 10S1I, improving the misfit from 46.4 to 44.6. The multiplicity effects were still present in the first approach, but were folded entirely into $k_{\alpha\beta_{1_b}}$ and $k_{\alpha\beta_{3_b}}$. This is not necessarily incorrect, but it is probably not optimal to force all $k_{\alpha\beta_{i_f}}$ to be the same and correct for this only through the magnitude of the $k_{\alpha\beta_{1_b}}$ and $k_{\alpha\beta_{3_b}}$. For models with two IERPs (M=2), we set $k_{\alpha\beta_{1_f}}$ to the first parameter, $k_{\alpha\beta_{3_f}}$ to the second, and fix $k_{\alpha\beta_{2_f}}$ to $(k_{\alpha\beta_{1_f}}+k_{\alpha\beta_{3_f}})/2$. There is not a need to refit the (M=2) models as $k_{\alpha\beta_{1_f}}$ and $k_{\alpha\beta_{1_f}}$ are not constrained in any way. We also explored fitting all three IERPs, but with our given set of experiments, there was minimal improvement over setting $k_{\alpha\beta_{2_f}}$ to the average of $k_{\alpha\beta_{1_f}}$ and $k_{\alpha\beta_{3_f}}$) and, further the $k_{\alpha\beta_{2_f}}$ was very poorly defined through the fitting. Thus here we are really testing three different IERP models: (1) no IERPs; (2) single IERP (corrected for anticipated multiplicity); (3) Linear change in IERPs across the three reactions. Except for cases where the rate parameter is zero, the reverse reaction rate constant, as just discussed for the one-parameter model, is fixed by the measured equilibrium constant (K_i, both SERP and IERP reactions) from the ^{13}C NMR data using the 16-scan acquisitions (as presented in Section 5.4.3) with $k_{i_b}=k_{i_f}/K_i$.

We do not jump from here directly to the full thirteen-parameter model in the fitting process. Instead, we progressively and systematically increase model complexity so that we might track the response of the parameter values to this increasing complexity, thus providing a means to evaluate model robustness as well as the potential for overfitting. In particular, it is important to understand how the SERPs and IERPs affect each other as we add more parameters. To do this, we group parameter values using intuition regarding which rate constants are likely to be similar to one another.

NSMI	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
2S0I	k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}	k_{2_f}, k_{3_f}, k_{5_f}, k_{6_f}, k_{8_b}, k_{9_b}	—	—	—	—
3S0I	k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}	k_{2_f}, k_{5_f}, k_{8_b}	k_{3_f}, k_{6_f}, k_{9_b}	—	—	—
4S0I	k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}	k_{2_f}, k_{8_b}	k_{3_f}, k_{6_f}, k_{9_b}	k_{5_f}	—	—
2S1I	k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}	k_{2_f}, k_{3_f}, k_{5_f}, k_{6_f}, k_{8_b}, k_{9_b}	$k_{\alpha\beta_{1f}}$, $k_{\alpha\beta_{2f}}$, $k_{\alpha\beta_{3f}}$	—	—	—
3S1I	k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}	k_{2_f}, k_{5_f}, k_{8_b}	k_{3_f}, k_{6_f}, k_{9_b}	$k_{\alpha\beta_{1f}}$, $k_{\alpha\beta_{2f}}$, $k_{\alpha\beta_{3f}}$	—	—
4S1I	k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}	k_{2_f}, k_{8_b}	k_{3_f}, k_{6_f}, k_{9_b}	k_{5_f}	$k_{\alpha\beta_{1f}}$, $k_{\alpha\beta_{2f}}$, $k_{\alpha\beta_{3f}}$	—
2S2I	k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}	k_{2_f}, k_{3_f}, k_{5_f}, k_{6_f}, k_{8_b}, k_{9_b}	$k_{\alpha\beta_{1f}}$	$k_{\alpha\beta_{3f}}$	—	—
3S2I	k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}	k_{2_f}, k_{5_f}, k_{8_b}	k_{3_f}, k_{6_f}, k_{9_b}	$k_{\alpha\beta_{1f}}$	$k_{\alpha\beta_{3f}}$	—
4S2I	k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}	k_{2_f}, k_{8_b}	k_{3_f}, k_{6_f}, k_{9_b}	k_{5_f}	$k_{\alpha\beta_{1f}}$	$k_{\alpha\beta_{3f}}$

Table 5.4: Groupings of parameters for our simplified box models (NSMI-parameter, where N = 2, 3, or 4, and M = 0, 1, or 2).

For the models with fewer than ten solvent exchange rates, we grouped the parameters as given in Table 5.4. Note from the Table that we are assigning k_{8_b} and k_{9_b} rather than k_{8_f} and k_{9_f}. Given the sense of the multiplicity and remembering that $K_i = k_{i_f}/k_{i_b}$, it is not reasonable to force, for example, k_{8_f} (with $K_8 \approx 1/3$) to be equal to k_{2_f} (with $K_2 \approx 3$).

To give a few examples:

2S0I is a model with a rate for α exchange and a rate for β exchange that is independent of isotopic composition at the α site with no intramolecular exchange (this was effectively the model implicitly assumed in Cooper's work Cooper, 1976).

Model 3S0I has a single rate for α exchange with the possibility of *two* rates for β exchange, one with α=H (k_{2_f}, k_{5_f}, k_{8_b}) and one with α=D (k_{3_f}, k_{6_f}, k_{9_b}). Model 4S0I is similar but allows for different rate for k_{5_f}.

As just described we added either one or two IERPs. For example, 2S1I is a minimal model with a rate for α solvent exchange, a rate for β solvent exchange, and a single rate for intramolecular exchange.

The parameter groupings in Table 5.4 were informed by thousands of previous fits performed in a more exploratory and less systematic manner. This exploration suggested that: (1) the variation in the α exchange rates (k_{1_f}, k_{4_f}, k_{7_b}, k_{10_b}) was small relative to variation in β exchange rates (k_{2_f}, k_{3_f}, k_{5_f}, k_{6_f}, k_{8_b}, k_{9_b}); (2) there was a substantial difference in β exchange rates depending on whether α was H or D, (3) k_{5_f} was larger than k_{2_f} and k_{8_b}, which were approximately equal to one another; (4) the intramolecular exchange rate parameters were significantly larger than the solvent exchange rate parameters. Though much of this process was exploratory, it was nevertheless thorough. For example, in the 2S0I models we scanned every possible way $2^{10}/2$ (512) of assigning two parameters to ten variables, and, for 3S0I, every possible way $3^{10}/3!$ (9842) of assigning three parameters. Though the groupings above for the 2S0I and 3S0I were not necessarily the lowest-misfit groupings found (mainly because the full ten-parameter model turns out to be more complex than the symmetric lower-order models can recover), we are not blindly imposing preconceived groupings in our search.

We then searched for optimal parameters for each of these models. The starting guess for all SERPs were 2.0×10^{-4} $M^{-1}min^{-1}$ After discovering early on that

the IERPs were roughly a factor of ten larger than the SERPs, we set the initial parameters for IERPs to 20.0×10^{-4} $M^{-1}min^{-1}$ to achieve a reasonable fitting attempt success rate.

For each model we ran at least 10,000 unique fitting attempts, starting from random guesses with each of the initial values perturbed randomly (positively or negatively) by up to 40% of the initial value. Note that the initial values (before random perturbation) are the same for each of these fitting attempts. In other words, this is not a Monte-Carlo type of search where the initial guess is continually updated. Our approach is designed as a sort of reconnaissance operation to understand the space of possible solutions within the iterations of downhill simplex optimizations rather than to provide an additional level of optimization on top of the downhill simplex iterations. The results are shown for the 10S2I model in Figure 5.12, given as a plot of the value of each parameter versus the misfit. Note that the lowest-misfit obtained (39.8954) is subtracted from all values, so the distributions begin at zero. Overall, the parameters fitted to the QJMNO dataset are well constrained by the data. Although k_{4_f}, k_{7_f}, k_{5_f} and $k\alpha\beta_{3_f}$ have broad distribution bands at higher misfit levels, they converge to reasonably unambiguous values at low misfit levels. There is also a high density of points at approximately five above zero, indicating an alternate solution, which is associated with a set of lower values for k_{3_f} and k_{9_b}.

Additionally, it is apparent from Figure 5.12 that most of the parameter values would be much less well determined if our data were restricted to Experiment "Q" only, or $QM_\alpha O_\beta$. Experiment "Q" gives us tight constraints on k_{1_f} and k_{2_f} as these are associated with a high flux in that experiment, which starts from $CHCH_3$. From the opposite end of the isotopic concentration range, Experiment "M" provides some constraint on k_{10_f}, while Experiment "O" provides some constraint on k_{8_b}. We anticipate that kinetics experiments observing with ^{13}C NMR for the labeled systems (i.e., ^{13}C NMR experiments with the conditions of "J", "M", "N", and "O", which start with either $CDCH_3$ or $CHCD_3$), along with an additional ^{13}C NMR experiment starting with $CDCD_3$ in H_2O, would provide the specificity needed to tighten up the distributions for the other parameters to comparable levels.

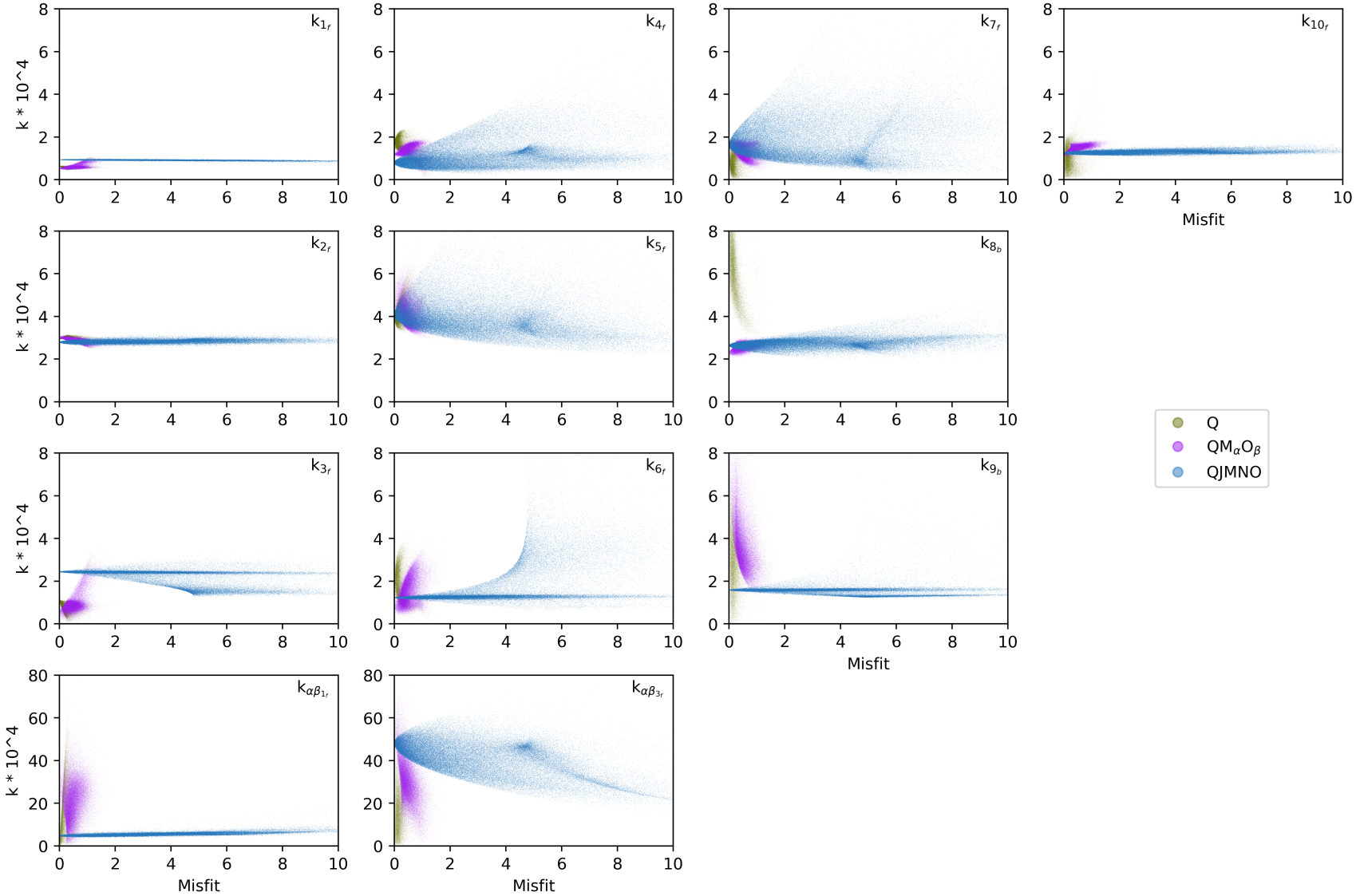

Figure 5.12: Parameter distributions for the 10S2I Model, comparing fits to "Q" (green), "$QM_\alpha O_\beta$" (purple), and "QJMNO" (blue) datasets. Misfits are shifted such that the x-axis is the misfit found by downhill simplex minus the lowest misfit value found for the group. Units of rate constants are $\times 10^{-4}$ $M^{-1}min^{-1}$

Although the fits that matter most are those to the full "QJMNO" dataset, we also look at models fitted to "Q" and to "$QM_\alpha O_\beta$", not only to check for evidence of overfitting in the full ten-parameter search, but also to examine the sensitivities of the parameters to selective omission of fitting data, in particular to look for inconsistencies between the labeled ^{1}H NMR ("J", "M", "N", "O") data sets and the ^{13}C NMR ("Q") data sets. We look for these in part because of the evidence presented above about the possibility of enzyme degradation between the ^{1}H NMR and ^{13}C NMR experiments. We choose $QM_\alpha O_\beta$ as one of the subsets because "M" and "O" provide constraints for Reactions 8 and 10 that "Q" alone cannot. It includes only the solution-driven species initially furthest from equilibrium, where Reactions 8 and 10 are most clearly involved, and does not directly incorporate any data resulting from the signals displaying "transient overshoots" discussed in Section 5.5.2 above.

Ideally, the subspace of fits to the full data set would lie within the subspace of "$QM_\alpha O_\beta$" fits and the subspace of "$QM_\alpha O_\beta$" fits, in turn, would lie within the least restrictive subspace of fits to only the "Q" data set. If it should happen that the more restricted set of fits do *not* lie within the subspaces of the least restricted fits, that could indicate self-inconsistency within the data. Given that we are merging here both ^{1}H and ^{13}C NMR data, it is important to note known and potential sources of inconsistency. Recall that Experiment "Q" was performed on alanine labeled with ^{13}C at the β carbon position, which distinguishes it from the ^{1}H experiments ("J", "M", "N", "O") we are combining with "Q" in our models, which are not labeled with ^{13}C. From our computational predictions, we expect the presence of ^{13}C at the β site to introduce a clumping effect, where D will show a higher preference for the β site by 5–10 ‰ in the presence of ^{13}C relative to ^{12}C. This means there is a systematic error in this range of 5–10 ‰ built in to our data set, for which we do not try to correct. Since it is expected to be small compared to the H/D effects we are focused on in this study, we do not expect this error to lead to an inability to extract useful results from our model presented below in this section. In addition, the ^{13}C NMR kinetics appear to be approximately a factor of two slower than the ^{1}H NMR kinetics. In order to merge these data sets for this modeling, we attempted to correct for this rate discrepancy, as described Supplementary Materials, Section 5.8.4. We have provisionally attributed this discrepancy to enzyme degradation between experiments, however, the issues accounting for this discrepancy may be more complex. The obvious way to

resolve this issue is to repeat the experiments with modifications, as detailed at the end of this Chapter. It is important to note here, however, that if we are correct about the origin of the discrepancy coming from enzyme degradation, the rate constants reported here would need to be multiplied by two. As explained in more detail in Section 5.8.4, we have *divided* the time in the model predictions for experiments J, M, N, and O by a factor of two. In other words, we are speeding up the "model" time relative to what is used to represent Experiment "Q". We performed the fitting in a "Q"-centric framework due to the central role of the ^{13}C NMR measurements in motivating this work, even though, in retrospect, the ^{1}H NMR measurements may be the ones done on the un-degraded enzyme. We do the same for experiments "F", "G", "H", "I", "L". and "P" when comparing them to model predictions below; these experiments are not used to fit any parameters, so this has no influence on the model construction. This issue is discussed in more detail in the Appendix in Section 5.8.4.

Plots showing parameter distributions for all of the models are provided in the Supplemental Attachments. Some of the parameter distributions reveal a tendency for discontinuous areas of high density reflecting multiple minima in parameter space, particularly when fitting the reduced 2SMI, 3SMI, and 4SMI parameter sets to the restricted "QM$_\alpha$O$_\beta$" data when M=1 or 2. We interpret this as a split between a locally stable solution with small IERPs and another solution set with much faster IERPs and significantly lower degree of misfit. In some cases, for example, 3S0I "QM$_\alpha$O$_\beta$", there are effectively continuous linear subspaces of solutions where one of the parameters (e.g. k_{3_f} in the case of 3S0I) is locally stable at higher misfit values, but extrapolates to solutions that are in good agreement with "Q" and "QJMNO" parameter values at low values of misfit. In some cases the subspaces appear parabolic or bilinear rather than linear (as in 4S0I). We do not believe these subspaces result from the convergence tolerance, that is, specifying a tighter cutoff of, e.g., 10^{-5} rather than 10^{-4} between the simplex points does not appreciably focus the scatter in the parameters. We believe the parameter scatter really does reflect a distributed range of locally stable solutions.

The consistency check, in terms of the "QJMNO" fits falling within those of the less restrictive "Q" and "QM$_\alpha$O$_\beta$" fits, gives mixed results. The most significant discrepancies are in k_{3_f} and (to a lesser extent) k_{1_f}, $k_{\alpha\beta_{3_f}}$ and k_{4_f}.

Each parameter is reasonably well defined in the "Q" and "$QM_{\alpha}O_{\beta}$" sets but, for k_{3_f}, about a factor of two smaller than in the estimate taken from the full data set. We doubt that further exploration with our model would lead to a resolution to these discrepancies, as they likely arise from minor incompatibilities within our data set; to resolve these, the kinetics experiments should be repeated with modifications and additions, as discussed below in the recommendations for future work.

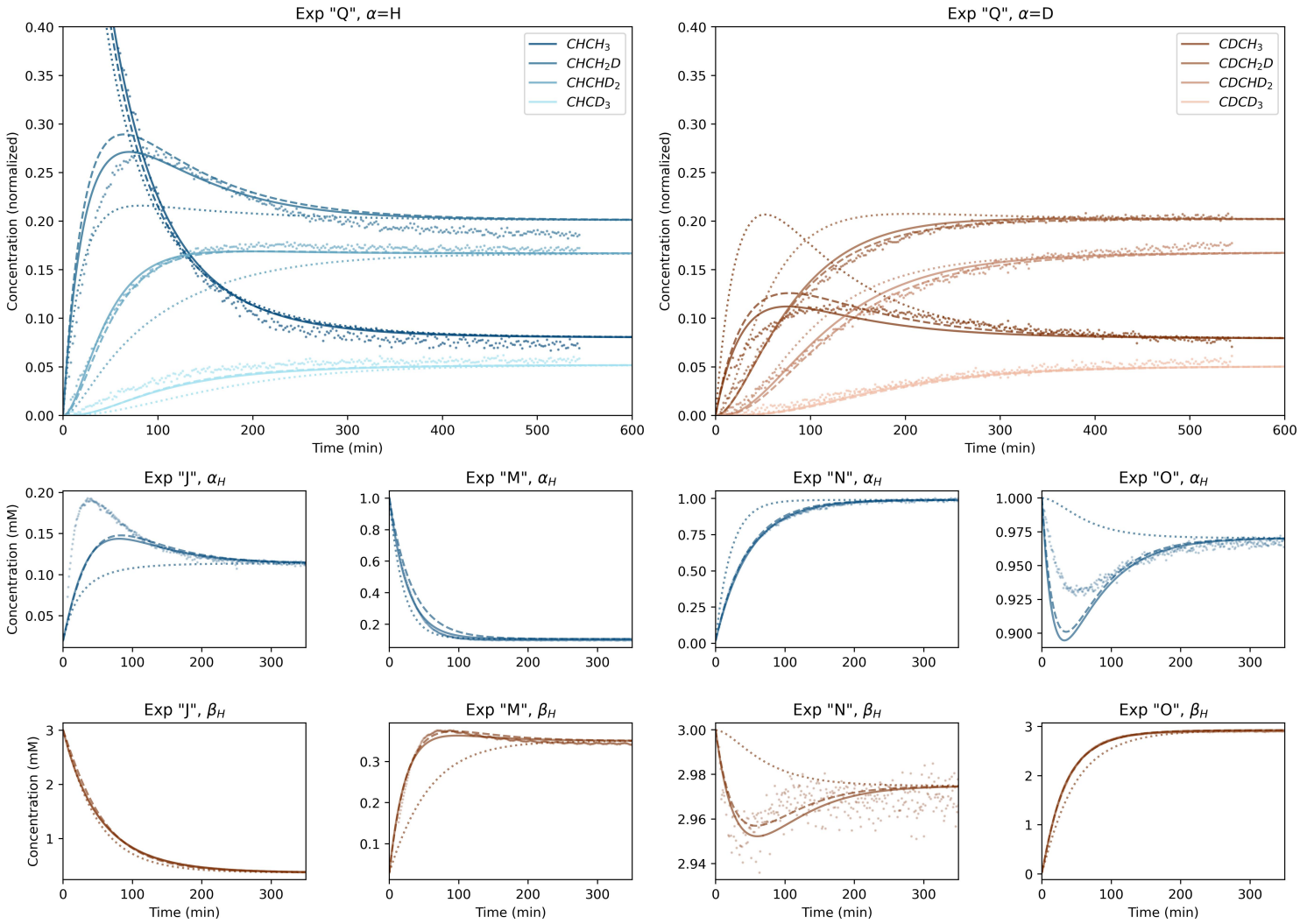

Figure 5.13: Time evolution of the 1S0I (dotted lines), 3S2I (dashed lines), and 10S2I (solid lines) models fitted to the experimental data for experiments "Q", "J", "M", "N", and "O" (colored points). Experiment "Q" is split into α=H (left) and α=D (right) species. For experiments "J", "M", "N", and "O", the α_H integrated species are shown on the top row and the β_H integrated species are shown on the bottom row. The experimental data displayed here is equivalent to that shown in Figures 5.6 and 5.8.

The general features of the 10S2I model (misfit=39.8954) are consistent with the simpler 3S2I model (misfit=53.8371), that is:

- An α SERP (k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}) close to 1×10^{-4} $M^{-1}min^{-1}$ (k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f})
- A fast (α=H) β SERP (k_{2_f}, k_{5_f}, k_{8_b}) close to 3×10^{-4} $M^{-1}min^{-1}$
- A slower (α=D) β SERP (k_{3_f}, k_{6_f}, k_{9_b}) close to 2×10^{-4} $M^{-1}min^{-1}$
- A $k_{\alpha\beta_{1_f}}$ IERP of around 5×10^{-4} $M^{-1}min^{-1}$ for $CDCH_3 \rightleftharpoons CHCH_2D$
- A $k_{\alpha\beta_{3_f}}$ IERP around 40–50×10^{-4} $M^{-1}min^{-1}$ for $CDCHD_2 \rightleftharpoons CHCD_3$

Examining the parameter distributions, the improved misfit for 10S2I over 3S2I may result from allowing for an increase in α exchange rate in going from k_{1_f} to k_{10_f}, along with the possibility of a faster rate for k_{5_f} and a slower rate for k_{6_f}. However, given the parameter uncertainties, specific aspects of the improvements in going from 3S2I to 10S2I are somewhat speculative. For example, it is not clear whether the increase in α exchange rates from k_{1_f} to k_{4_f} to k_{7_f} to k_{10_f} is gradual or non-monotonic, with k_{7_f} being greater than k_{10_f}. Further, because the 4S2I model has minimal improvement over 3S2I, it appears that the faster k_{5_f} needs to be coupled to greater flexibility in the other parameters.

A surprising aspect of the model is the very large value of the fitted $k_{\alpha\beta_{3_f}}/k_{\alpha\beta_{1_f}}$, with values of $\frac{k_{\alpha\beta_{3f}}}{k_{\alpha\beta_{1f}}}$ near 10. We would expect this ratio to be closer to 3, given the sense of the multiplicity. One might be led to question whether this value is meaningful; it is difficult to conceive of a rationalization for rate constants representing reactions for isotopic exchange alone resulting in values differing by such magnitude. Given the consistency of the IERP rate constants across the range of 2S2I-10S2I in Table 5.8, it seems unlikely that high ratio is a result of overfitting. We believe this may be an indication of multiple mechanisms involved in what we are representing as a pseudo-elementary reaction in our model. It is difficult attribute it to transport effects as IERPs are unlikely to be affected by transport, unless these effects are localized to multiple exchange pools involving, for example, lysine interacting with near-solvent and far-solvent pools. In any case, the mechanistic implications of this increase are likely more complex than can be accounted in the model we use here. This

is another area where molecular-level simulation models could provide some insight.

A similar observation can be made for the relative values of k_{2_f}, k_{5_f}, and k_{8_b}. When we allow these values to vary independently in the 4S and 10S models, the best fits have k_{5_f} close to 4×10^{-4} $M^{-1}min^{-1}$ and k_{2_f} and $k_{8_b} \approx 2.6$ to 2.8×10^{-4} $M^{-1}min^{-1}$. This is a more surprising variation when one realizes that, due to multiplicity effects, k_{2_f} and k_{8_b} would be expected to be about 1.5 times *larger* than k_{5_f}. Understanding the underlying molecular mechanisms that give rise to this behavior will likely require atomic-level simulations along the lines of, e.g., Dama et al., 2013, Jang et al., 2004, Pomès et al., 1996, and Cerqueira et al., 2011, but applied specifically to the ALT enzyme reaction.

The behavior of the 10S0I model is peculiar in that the parameter fitting procedure discovered physically implausible solutions with a very large k_{9_b} and a lower misfit ($\approx$60) than the physically plausible solution (misfit$\approx$70). This lower misfit model achieves a better fit (lower misfit) to the transient overshoots in Experiments J_α and M_β, with the trade-off being a worse fit (higher misfit) on Experiment "Q". We did not anticipate that a model without IERP parameters could have the capacity to recover transient overshoots. This appears to happen through artificial kinetic "traffic jams" in the model that arise due to the large contrasts in the magnitudes of the rate constants (a combination of very large or very small values). Note, however, that our fitting procedure did not discover any such solutions that are able to reproduce *all* the transient overshoots (i.e. J_α, M_β, N_β, and O_α), and it is likely that a model with only SERPs cannot accomplish this.

Interestingly, the lowest misfit of all models we discovered comes from a single 10S3I solution with misfit = 39.03. This was a single solution found only once in the search procedure. It is similar to the lowest-misfit 10S0I solution in that it has a large k_{9_b}, and a small $k_{\alpha\beta_{3_f}}$. The next-lowest misfit 10S3I solution (misfit = 39.76) more closely resembles the best 10S2I fits. As apparent in Table 5.10, a major reason for the success of the misfit=39.03 solution is that it is able to simultaneously represent both the J_α and O_α transients (albeit at the expense of a worse fit for J_β and M_β). Other solutions have trouble with this as, generally, getting a good value for the J_α transient overshoot coincides with an O_α transient "undershoot" that is too pronounced. It is likely that both the lowest-misfit 10S0I and 10S3I fits are spurious solutions

resulting from dataset inconsistencies that could be resolved via further work, as we suggest at the end of this section. The trade-offs between J_α and O_α may suggest an inconsistency in those datasets, but could also result from the limitations of the model. Although we reject the lowest-misfit models for 10S0I and 10S3I, it is encouraging that the searching procedure is doing its job of locating surprising alternative fitting approaches, even if they are nonphysical. Finding only a single anomalous solution for the 10S3I model also indicates that the 10,000 iterations of separate downhill simplex fits is not overkill.

In Figure 5.13 we present the time evolution of three distinct model simulations (1S0I, 3S2I, and 10S2I) for the experiments to which they were fitted (the QJMNO dataset) using parameters chosen from the lowest-misfit parameter set. Our choice of presenting the parameters of the lowest misfit solution may risk obscuring the diversity of nearby solutions having slightly higher total misfit. One could argue that it would be better to choose an "average" value in the middle of the distribution at slightly higher misfits (see Figure 5.12). For example, in 10S2I, the lowest-misfit model has k_{7_f}=1.77×10^{-4} M^{-1}min^{-1}, but this happens to be one value in a broad distribution covering ∼1–2×10^{-4} M^{-1}min^{-1}, without much increase in misfit. Thus there is potential trouble in using a single number to represent the goodness of fit, for it may be that some of these solutions have desirable attributes that we miss by just choosing a single one based on the scalar measure of total misfit. However, going beyond the total misfit criterion would require developing more sophisticated measures which should await a more complete data set. In the work presented here, we pick the model with the lowest misfit and use the parameter distributions as a visual indicator of parameter uncertainty. Given the non-Gaussian, irregular nature of the parameter distributions, this qualitative measure of uncertainty is the best we can do. The complexity inherent in these distributions is evident in Figure 5.12.

The sensitivity of our results to somewhat arbitrarily choosing the lowest-misfit solution can be addressed by plotting a range of solutions instead of just the lowest-error solution. We show this in Figure 5.19, which randomly samples solutions having misfit values amongst the lowest 10 %. Given the visible lack of variability in the trajectories of these models, we likely do not miss much by considering only the solution with the lowest error.

For the model with the lowest misfit, we plot the results of the time evolution

of the eight isotopic variants in Experiment "Q" and the integrated α_H and β_H species of Experiments "J", "M", "N" and "O", respectively, in Figure 5.13. While the one-parameter fit (i.e., 1S0I) is visually surprisingly good, remember that, by construction, the one-parameter model recovers the equilibrium distributions as the k_{n_b} rate constants will be different for each species through the relation $k_{n_b} = k_{n_f}/K_n$. The 10S2I model is a major improvement under conditions far from equilibrium. The most noticeable deficiency in the 10S2I model is the overshoot for $CHCH_2D$ in Experiment "Q", which may be related to a systematic discrepancy between the equilibrium defined by the single-scan runs and the more accurate determination from the multiscan runs at the end of the experiment. Similar plots showing the species evolution as a function of time are given in the Supplemental Attachments.

The experimental findings presented here invite molecular-level calculations (using a combination of molecular dynamics and first-principles electronic structure calculations) about the magnitudes of isotopic composition of neighboring atoms in the vicinity of an isotope exchange reaction. A possible approach would be to try to identify a reaction pathway for α and β exchange on alanine interacting with a lysine residue. If a transition state structure could be located for both α and β exchange reactions, the vibrational energy change on H/D isotopic substitutions could be calculated. First, We would expect to see modest increase in α exchange rates as D is progressively substituted in at the β site. A stronger effect would be predicted for β exchange. Substituting D for H at the should decrease the rate by about $\frac{2}{3}$. Given the uncertainties in transition state theory formulations it would be difficult to get quantitative estimates, but it should be possible to see whether substitution of D for H at the β site increases α exchange, and whether substitution of D for H at the α site decreases the rate of α exchange.

In the modeling work detailed above, we did not attempt to fit to ^{1}H NMR experiments FGHILP. Instead, we decided to use our models fitted to the QJMNO dataset to predict the course of these experiments, as an independent check on the validity of our models. This decision was based on the fact that the FGHILP experiments are more or less redundant to Experiment "Q", in the sense that they start from the same $CHCH_3$ isotopic variant, and the only variable changing across these experiments is the initial water composition, and thus in theory they should not provide useful additional constraints for the rate

constants in our model unless the rate constants are somehow dependent on solvent composition. Figure 5.14 presents these fits for 1S0I, 3S2I, and 10S2I models, as fitted to the QJMNO dataset. First note that the predictions for the ^{1}H β signal are systematically worse than for the α signal. Experimentally we found that the initial $\frac{\beta}{\alpha}$ ratio was close to 2.93 to 1 for Experiments "F", "G", "H", "I", "L", and "P", and we opted to normalize the signal such that the initial value for α was set to unity, and therefore to let the initial value of β be close to 2.93. In the model, the initial ratio $\frac{\beta}{\alpha}$ is 3, hence the misfit for the β signal suffers. Since we are not actually fitting these data, we have not attempted to correct the β signal. This means, however, that there is no physical significance to the fact that the β contribution to the misfit is worse than the α contribution; it is simply a result of our arbitrarily choosing to normalize the α to unity. Although we break down the contributions in terms of α and β, only the total misfit is significant. Further, as mentioned above, it is important to understand that the time for the model system has been divided by two when compared to the experimental data for all ^{1}H NMR data. As expected, the 3S2I fit is a dramatic improvement on the 1S0I fit, particularly on the α_H species. The 10S2I fit is a further improvement over the 3S2I fit, though less dramatic. These improvements can most easily be seen in the bottom panel of Figure 5.14, which gives the ratio of the data in the top two panels, i.e., $\beta_H/(3\alpha_H)$. In this format it is readily apparent that the 1S0I fit fails to capture any of the initial dip, and prematurely rises to a much higher ratio than the data actually do. Model 10S2I is our only model that predicts both the initial dip and subsequent rise of $\beta_H/(3\alpha_H)$ in these experiments, most clearly realized in Experiment "P" (15%H). Given the generally excellent ability of this model to predict Experiments "F", "G", "H", "I", "L", "P", it appears that the rate constants we have determined are independent of solution composition. We can take this agreement to indicate that the models do not suffer much due to not specifically accounting for H_2O-HDO-D_2O. Plots of the predictions of each of the models for Experiments "F", "G", "H", "I", "L", "P" are given in the Supplemental Attachments.

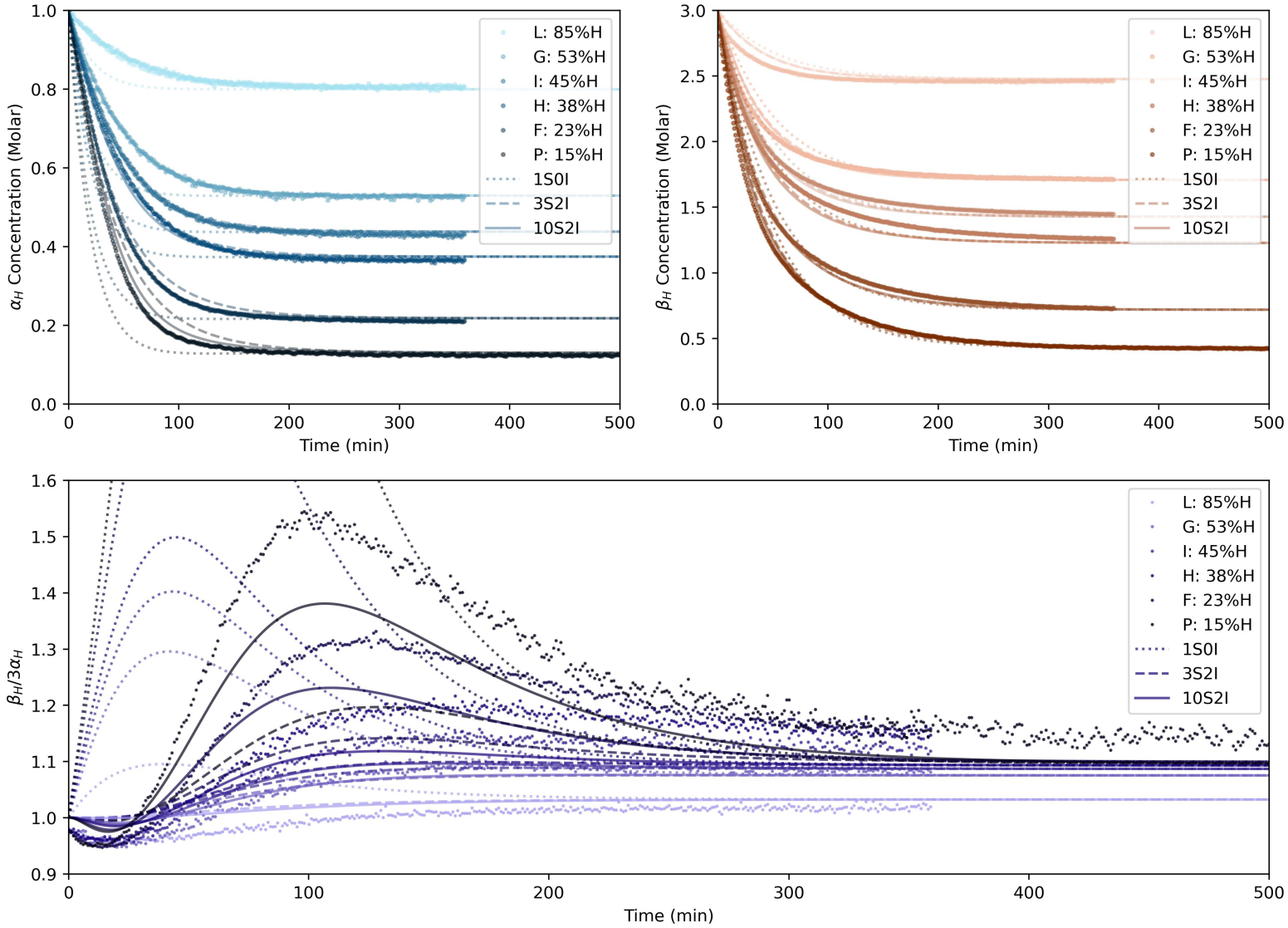

Figure 5.14: Time evolution of the 1S0I (dotted lines), 3S2I (dashed lines), and 10S2I (solid lines) models in a range of D_2O:H_2O (as displayed by experiments "F", "G", "H", "I", "L", and "P"), fitted to the QJMNO experimental dataset. Data from experiments "F", "G", "H", "I", "L", and "P", as described in Table 5.1 presented in Figure 5.4.

Figure 5.15 shows various model predictions for the data on the initial α-β H-loss experiments presented in Figure 5.5 from the labeled alanine experiments described in Section 5.4.1.3. An interesting feature of those experiments was that having D at the β site seemed to increase the rate of H-loss at the α site, whereas having D at the α site seemed to slow down the rate of H-loss at the β group. It is instructive to look at these data in light of some of the models presented above. For example, as shown in Figure 5.15 the 2S0I model with a single rate constant for α exchange and a single rate constant for β exchange cannot reproduce either the Experiment "M"–Experiment "P" comparison or the Experiment "J"–Experiment "P" comparison. This is obvious because in the 2S0I model, the two rate constants are independent. Model 3S0I can account qualitatively for the slow rate of β exchange observed in Experiment "J" (with D at α) relative to Experiment "P" (with H at α) but cannot account for the relative rates of α exchange between Experiment "M" (labeled with D in the β group) and Experiment "P" (with H at the β group). This is because the 3S0I model has a faster rate for k_{2_f}, k_{5_f}, and k_{8_b} (α=H), and a slower rate for k_{3_f}, k_{6_f}, k_{9_b} (α=D). But just as for model 2S0I, 3S0I cannot account for the observed comparison between Experiment "M" and Experiment "P" because, given the way we have grouped the parameters in Table 5.4, there is still only a single α exchange rate for k_{1_f}, k_{4_f}, k_{7_f} and k_{10_f}.

In contrast, the 2S1I model can qualitatively account for the increase in α H-loss caused by a deuterated CD_3 β site acting as a source of D. In other words, as mentioned above in the experimental section, the relative outcomes of Experiment "M" and Experiment "P" can be understood by including the possibility of intramolecular exchange. Further, the additional flexibility of the 2S2I model (with the large ratio of $k_{\alpha\beta_{3_f}}/k_{\alpha\beta_{1_f}} \approx 18$) is much better able to quantitatively capture this effect than 2S1I (with the smaller ratio of $k_{\alpha\beta_{3_f}}/k_{\alpha\beta_{1_f}} = 3$).

Figure 5.15 also shows that model 10S2I can account for the relative H-loss rates in Experiment "M" and "P" because there is a tendency for the α rate constants α exchange (k_{1_f}, k_{4_f}, k_{7_f} and k_{10_f}) to become faster with increasing D in the β group. It is also evident the slower rate of β exchange for α=D (in Experiment J) than for α=H (in Experiment "M") *cannot* be explained using IERPs; here the IERPs actually make the problem a little worse as they provide a source of D and thus accelerate the β H-loss rather than retard it.

As discussed in the experimental section, the relative outcomes of Experiment "J" and Experiment "P" cannot be reproduced (at least in the context of our model) without allowing for β exchange *rate constants* to be higher for α=H (k_{2_f}, k_{5_f}, k_{8_b}) than for α=D (k_{3_f}, k_{6_f}, k_{9_b}).

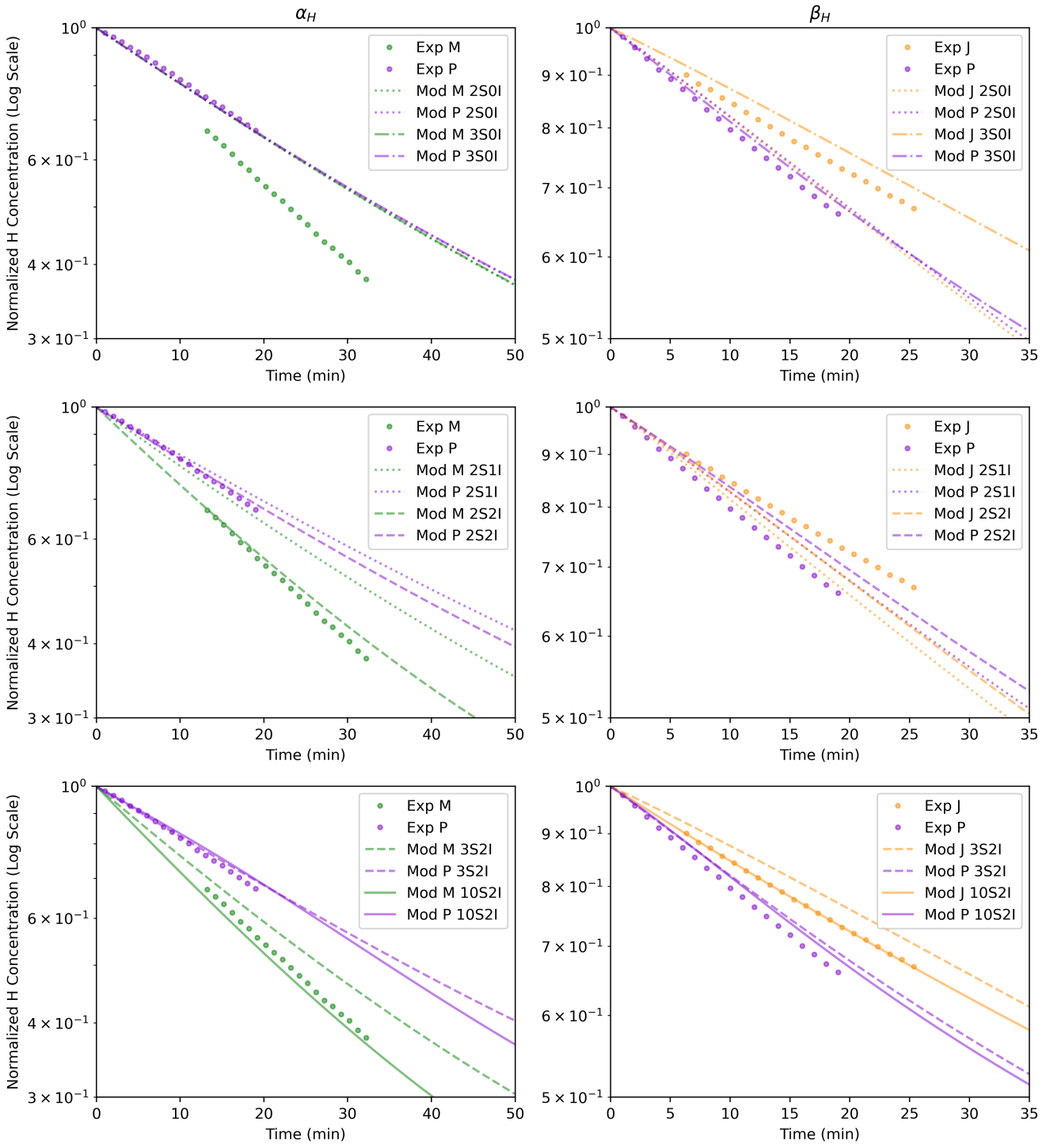

Figure 5.15: Comparing model predictions for hydrogen loss rates, using the initial 20 minutes of data from kinetics Experiments "P", "M", and "J" in ~85:15 D_2O:H_2O. Left: α hydrogen loss rate in Experiments "M" ($CHCD_3$, green circles) and "P" ($CHCH_3$, purple circles). Right: β hydrogen loss rate in Experiments "J" ($CDCH_3$, orange circles) and "P" ($CHCH_3$, purple circles). Lines are models as indicated, fitted to the QJMNO experimental dataset. Y-axis is concentration in log scale, normalized to initial concentration.

5.5.4 "3-Box" Simplified Model

In the early stages of this work, we created a simpler model fitted only to our data acquired using ^{1}H NMR (i.e., the FGHILP dataset). While we believe our later "8-species" model based on both ^{1}H NMR and ^{13}C NMR data far surpasses this earlier model, there is some value in presenting the earlier work here to show both the potential and limitations of working with observations that aggregate multiple distinct isotopic variants, i.e., in our case, ^{1}H NMR data. We also find that reducing the fully-resolved eight-isotopic-variant model into a model that differentiates only between the relative proportions of D and H at the α and β sites can give some guidance in constructing the "8-species" model, that is, both models benefit from the comparison.

This simplified model involves three "boxes", as shown in Figure 5.16: the α and β "reservoirs", whose equilibrium is mediated through a solvent reservoir, "s". This basic model has four parameters, $K_{\alpha\beta}$, k_{α,s_f}, k_{α,s_b}, and k_{β,s_f}, representing forward and backward exchange between both the α and β pools with the solvent. Figure 5.16 also depicts the subsets of isotopic variants that are aggregated in the counting of H abundance in the α "reservoir" ($CHCH_3$, $CHCH_2D$, $CHCHD_2$ and $CHCD_3$, each counted once) and β "reservoir" ($CHCHD_2$, $CDCHD_2$, $2\times CHCH_2D$, $2\times CDCH_2D$, $3\times CHCH_3$, and $3\times CDCH_3$). The system of ordinary differential equations defining the fluxes of H and D between the α and β sites and solution for this model are as follows:

$$\frac{d\alpha_H}{dt} = -k_{\alpha,s_f}[\alpha_H][s_D] + k_{\alpha,s_b}[\alpha_D][s_H] \tag{5.39}$$

$$\frac{d\alpha_D}{dt} = \frac{-d\alpha_H}{dt} \tag{5.40}$$

$$\frac{d\beta_H}{dt} = -k_{\beta,s_f}[\beta_H][s_D] + k_{\beta,s_b}[\beta_D][s_H] \tag{5.41}$$

$$\frac{d\beta_D}{dt} = \frac{-d\beta_H}{dt} \tag{5.42}$$

$$\frac{ds_H}{dt} = k_{\alpha,s_f}[\alpha_H][s_D] - k_{\alpha,s_b}[\alpha_D][s_H] + k_{\beta,s_f}[\beta_H][s_D] - k_{\beta,s_b}[\beta_D][s_H] \tag{5.43}$$

$$\frac{ds_D}{dt} = \frac{-ds_H}{dt} \tag{5.44}$$

The associated equilibrium constants for solvent exchange are:

$$K_1 = \frac{[\alpha_D][s_H]}{[\alpha_H][s_D]} \tag{5.45}$$

$$K_2 = \frac{[\beta_D][s_H]}{[\beta_H][s_D]} \tag{5.46}$$

In addition, in Section 2.7 we calculated a theoretical estimate for the direct exchange of H and D between the α and β sites:

$$\alpha_H + \beta_D \rightleftharpoons \alpha_D + \beta_H \tag{5.47}$$

with $K_{\alpha\beta}$:

$$K_{\alpha\beta} = \frac{[\alpha_D][\beta_H]}{[\alpha_H][\beta_D]} = \frac{K_1}{K_2} \tag{5.48}$$

Our first-principles calculations indicated a value of 1.191 for $K_{\alpha\beta}$. Fixing $K_{\alpha\beta}$ at 1.191 allowed us to reduce the number of fitted parameters from four to three, because $K_2 = \frac{K_1}{K_{\alpha\beta}} = \frac{k_{\beta,s_f}}{k_{\beta,s_b}}$. Our initial fits to a subset of the ^{1}H NMR data (experiments G, L, and P, representing the extremes of H-D water composition) gave, surprisingly, a slightly smaller rate constant for H/D exchange at the β site than for the α site. This disagreed with the work of Cooper, 1976 and Golichowski et al., 1977, both estimating the rate constant for β exchange was approximately three times the rate constant for α exchange.

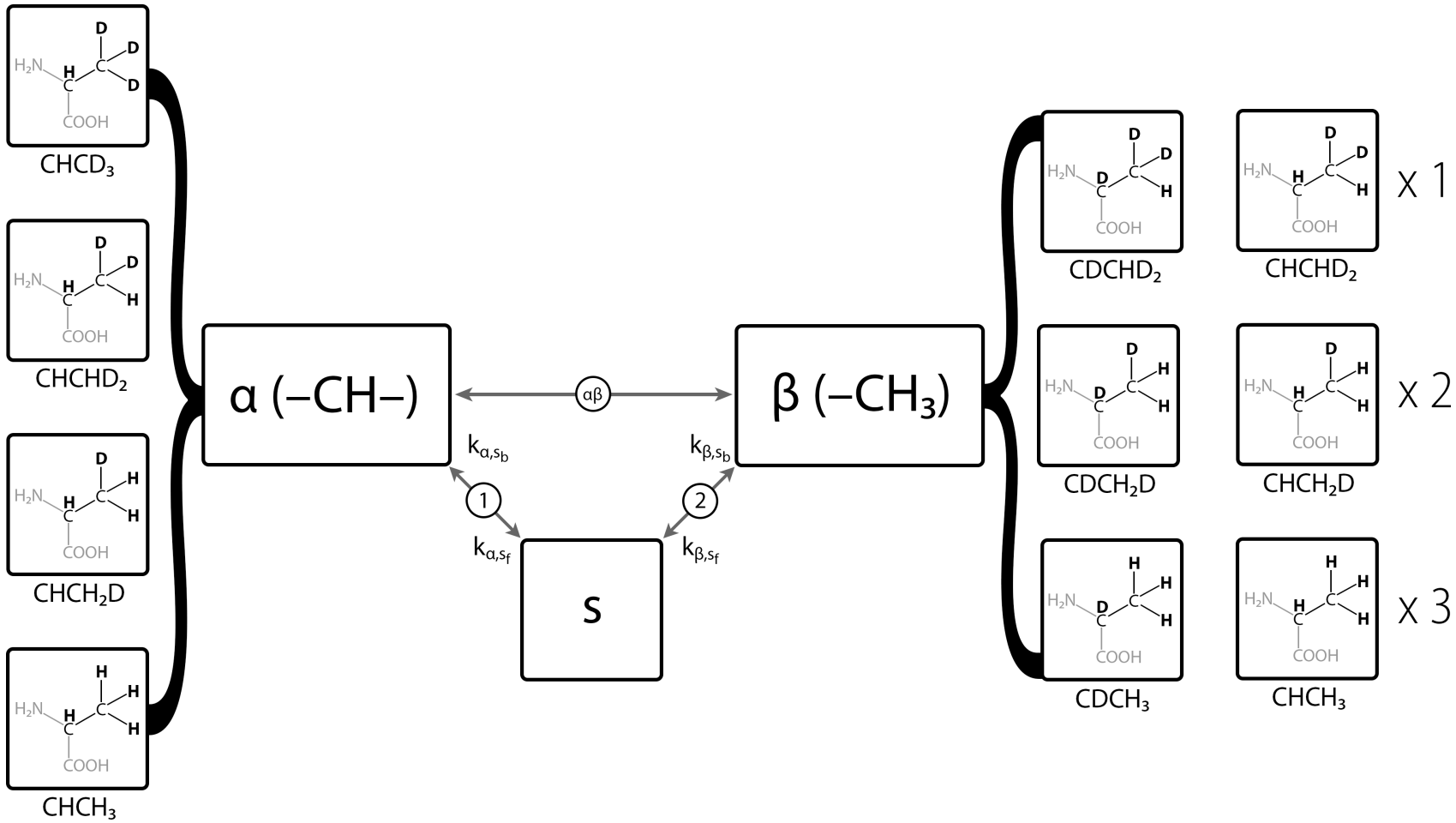

Figure 5.16: Graphical depiction of the "3-Box" model, in which we show the subsets of the eight isotopic variants making up the observed ^{1}H NMR signal. Note that those variants having H at the α site *and* the β site are "counted" in both the α "box" and the β "box", and variants having multiple H at the β site are "counted" according to this multiplicity. Corresponding equations are given in the text.

After working through the parameter fitting for the 8-species model (and greatly improving the parameter-fitting workflow and software) we revisited this earlier, more primitive model that was constructed before we had the ^{13}C-NMR data resolving all eight isotopic variants. Applying the improved software did not change the results. The improved software still produced a k_{α,s_f} that was close to, but a bit larger than, the k_{β,s_f}. We then modified the 8-species-based 2S0I model described in the main text to reproduce the 3-box model, but did not use any information from the ^{13}C-NMR experiments. As in the 2S0I model we forced all the α site exchange rates ($k_{1_f}, k_{4_f}, k_{7_f}, k_{10_f}$) to share an identical value, (k_{1_f}), and also forced all β-site exchange rates ($k_{2_f}, k_{3_f}, k_{5_f}, k_{6_f}, k_{8_b}, k_{9_b}$) to share an identical value, (k_{2_f}). Unlike the other reduced-parameter models MSNI (M < 10), which used the equilibrium constants, $K_1 - K_{10}$, measured in the ^{13}C-NMR experiments, we also fitted an equilibrium constant for K_1 (Equation 5.45), and enforced the K$_{\alpha\beta}$ (Equation 5.48) of 1.191. Thus these two model formulations (3-box and the restricted 2S0I 8-species model) appeared to us to be equivalent and were expected to yield the same result (subject, of course, to the randomness in the fitting process).

The fitted $k_{\alpha_s f}$ and k_{1_f} were very close (near 2.2×10^{-4} M^{-1}min^{-1}), but the fitted k_{β,s_f} and k_{2_f} were different, with k_{β,s_f} close to 1.9×10^{-4} M^{-1}min^{-1} and k_{2_f} close to 5.1×10^{-4} M^{-1}min^{-1}. The misfits for each of these models were essentially identical.

At this point it was necessary to more carefully formalize the equivalence between the 3-box model and the restricted 2S0I 8-species model to track down the source of the problem, since for two models that appeared to us to be equally valid (with nearly identical misfits obtained in the stochastic searching procedure) we were getting inconsistent results for the relative α and β exchange rates. After working through the algebra, it was realized that there was a problem in how we were dealing with the multiplicity of the forward and backward reactions for β exchange. For Reactions 2, 3, 8, and 9 we were simply dividing the forward rate constants k_{8_f} and k_{9_f} and backwards rate constants k_{2_b} and k_{3_b} by three to reflect the differences in free energy due to the multiplicity. This led to an inconsistency in matching the two models. We then tried a more symmetric approach, multiplying k_{2_f}, k_{3_f}, k_{8_b} and k_{9_b} by $\frac{3}{2}$ and k_{2_b}, k_{3_b}, k_{8_f} and k_{9_f} by $\frac{1}{2}$. When this is done, the "8-species" representation

and the "3-box" representation can be shown to be equivalent. The algebra is provided in Supplementary Materials Section 5.8.6. As can be seen in the last two equations, there is a factor of two reduction in the k_{β,s_f} when going between the "3-box" formulation and the restricted "8-species" formulation of the problem. Further, when the multiplicity was accounted for with the factors of $\frac{3}{2}$ and $\frac{1}{2}$ as just described, the fitted k_{2_f} changed from 5.1×10^{-4} $M^{-1}min^{-1}$ to 3.8×10^{-4} $M^{-1}min^{-1}$. Tables 5.5 and 5.12 show the near-perfect equivalence of these models once the proper symmetric multiplicity factors are included and the factor of two is applied to k_{β,s_f}.

The above discussion illustrates some of the potential problems that can arise when naively formulating "box" type models that bury unknown details within the boxes. In the context of this work, the complexities associated with multiplicity in the series of β-exchange reactions would be examples of such details. Here, these are knowable/measurable because we have the ^{13}C-NMR experimental data. In many cases, however, such complexities may not be knowable; this lack of knowledge is, of course, what would likely drive the generation of the "box"-type models in the first place. The exercise presented here shows that while the "3-box" model can describe the data just as well as the full "8-species" model, when fitted to an equivalent set of parameters, the conclusions drawn from the magnitudes of those parameters can be misleading. Even something as basic as the relative rates of the exchange between the α and β sites and solvent (k_{β,s_f} and k_{α,s_f}, respectively) can be obscured. In cases were we can only observe aggregations of isotopic variants, such as in the ^{1}H-NMR here, one must resist the temptation to go directly from the data to a model governed by the aggregated data. In such cases it is better to formulate the complete chemical model (in this case, one based on all eight isotopic variants) and then aggregate the complete model by grouping species and parameters as necessary, systematically moving upward in scale. If that step is bypassed, it is very easy to make errors analogous to the case of our initial misinterpretation of the relative rates of k_{β,s_f} and k_{α,s_f}. The algebra involved in performing the reduction from the "8-species" model to the "3-box" model is involved enough that issues like the factor of two reduction in the k_{β,s_f} exchange rate in the aggregated model are not obvious *a priori*.

Even if one cannot measure such indicators, it is important to work through the problem at the most fundamental level and then make approximations from

Parameter	8 Species	3-Box
$K_{\alpha\beta}$	**1.1910**	**1.1910**
k_{α,s_f}	2.1861544×10^{-4}	2.1861544×10^{-4}
k_{α,s_b}	1.9039785×10^{-4}	1.9039785×10^{-4}
k_{β,s_f}	3.8883611×10^{-4}	1.9441805×10^{-4}

Table 5.5: Parameters for the 8-species and 3-box models in Figure 5.16. $K_{\alpha\beta}$ is fixed by first principles calculations, and therefore given in bold.

there. In this work, for example, one can proceed by grouping parameters that cannot be distinguished experimentally, and then ensuring that the models are equivalent by working through the algebra. Further, the exercise of ensuring the consistency of an aggregated and fundamental model can give insights into how to interpret the fitted parameters of the fundamental model. Without going through this exercise it would have been easy to miss the factors of $\frac{3}{2}$ and $\frac{1}{2}$ that arise naturally in the equations. This is not an important issue in a completely unconstrained fit, of course, but it can matter when enforcing parameter groupings. For example, in the 3SNI fits, we set k_{2_f} and k_{8_b} to be the same as k_{5_f} and then enforced the multiplicity in Reactions 2 and 8 by setting k_{2_b} and k_{8_f} to be consistent with the measured equilibrium constants (2.904 for Reaction 2 and 0.361 for Reaction 8). It might have been more appropriate to set k_{2_f} and k_{8_b} to $\frac{3}{2}\mathrm{k}_{5_f}$. If we actually do this it turns out we get a worse fit, because we definitely find in the 10SNI fits, where the relative values of $\mathrm{k}_{2_f}/\mathrm{k}_{5_f}$ and $\mathrm{k}_{8_b}/\mathrm{k}_{5_f}$ are free to vary, that the best fits have roughly equal k_{2_f} and k_{8_b}, with both significantly less than k_{5_f}. Conversely, as discussed above, in the 10S1I model we significantly improved the fit by setting $\mathrm{k}_{\alpha\beta_{3_f}} = \frac{3}{2}\ \mathrm{k}_{\alpha\beta_{2_f}}$ and $\mathrm{k}_{\alpha\beta_{1_f}} = \frac{1}{2}\ \mathrm{k}_{\alpha\beta_{2_f}}$. This is simply because it appears that $\mathrm{k}_{\alpha\beta_{3_f}}$ is in fact faster than $\mathrm{k}_{\alpha\beta_{1_f}}$ as indicated in the unconstrained fits. That said, the formulation of the model with the correct factors does influence how the fitted parameters are interpreted. For example, the observation that k_{2_f} is slower than k_{5_f} is more of a surprise when its realized that it would be expected to be $\frac{3}{2}$ faster. The observation that $\frac{k_{\alpha\beta_{3_f}}}{k_{\alpha\beta_{1_f}}}$ is nearly 10 should be compared not with unity but to the factor of 3 that one would expect from the multiplicity. Of course, as a practical matter, it is important to keep in mind any differences between the fitted parameter and the actual value of the rate constant if the factors of $\frac{3}{2}$ and/or $\frac{1}{2}$ are enforced in the fitting process.

5.6 Implications

This fundamental study demonstrates that isotopic composition two bonds away from the exchanging site gives rise to measurable variations in the rate constant for exchange at said site. There are potential practical implications for this work as well, worth mentioning here. Given that ALT assays are routinely used to evaluate human liver function, and ALT is presently under investigation as to its role in cancer metabolism, it is plausible that further pursuit of answers to questions addressed in this chapter could be useful in developing new medical forensic tools. Several relevant questions come to mind: Could the pattern of measured rate constants be measurably sensitive to environmental factors, and such sensitivity be exploited as a kind of "fingerprint" for variations in enzyme tertiary structure and function *in vivo*? Given that there are many isozymes of this enzyme (two in humans, with many variations across other organisms), how similar are the rate constants across these isoforms? Further, even within one isoform, does the pattern of rate constants observed here vary with other factors? This is highly speculative, but certainly the ability to dynamically characterize H/D exchange and transport within an enzyme at this level of detail is a completely new capability, and, as such, has potential for unforeseen applications. In this sense our study gives a new kind of appreciation for the potential that isotomics has for broad application.

5.7 Conclusions

Here we used high-precision ^{1}H and ^{13}C NMR applied in novel ways to explore and characterize the details of hydrogen-deuterium exchange and site preference in alanine at both the α- and β-carbon positions, via an off-pathway reaction catalyzed by ALT. Our experiments yielded an unprecedentedly rich and informative dataset that tracks the kinetics of the exchange with high temporal resolution, following the abundances of eight isotopic variants of alanine involved in this reaction as they evolve through time.

By fitting a complete kinetic model of α- and β-hydrogen exchange (including rate constants for thirteen isotope exchange reactions) to the data from these experiments, we find evidence for two classes of rate constants, with one class governing exchange of H/D between solvent and alanine, and another governing intramolecular exchange within alanine. In general, intramolecular exchange proceeds approximately a factor of ten faster than solvent exchange, however we see variation in the solvent exchange rate constants by up to a fac-

tor of four, and by nearly a factor of ten in the rate constants for intramolecular exchange. Allowing for variations in rate constants over all thirteen isotope exchange reactions gives over a factor of three improvement in the misfit over a naive model assuming two exchange rates that are independent of isotopic variations elsewhere in the molecule (i.e., 2S0I in the notation discussed above, with one rate for exchange at the α site, and another for exchange at the β site). Given our findings, this "2S0I null hypothesis" model should probably be amended to include a reaction for intramolecular exchange between the α and β sites (i.e., a 2S1I model), which could have been postulated *a priori*. Even allowing for this modification, there is still a factor of 2.7 improvement in the misfit going from this restricted model to our fully fitted model. Our findings therefore disprove our null hypothesis; the observed behavior of the network of isotope exchange reactions in ALT catalysis is more complex than can be accounted for with a simple two-site model with intramolecular exchange, independent of the isotopic compositions of the two sites.

The H-D exchange reactions described here are decoupled from the chemical transamination reaction by leaving α-ketoglutarate (and thus also glutamic acid) out of the reaction. The ability to run the isotopic exchange reactions in absence of transamination, taking only one half of the "bi-bi ping-pong" reaction and thus having constant concentrations of chemical species throughout isotopic exchange makes the ALT reaction useful for exploring the fundamental issue of chemical/isotopic kinetic coupling, which has proven difficult when studied in other contexts, such as mineral precipitation reactions (e.g., Steefel et al., 2014, Druhan et al., 2013). In the ALT system we have the possibility in future experiments of understanding how the rate constants for H/D isotope exchange, measured here with the chemical reaction absent, would respond when the chemical reaction is proceeding at a finite rate.

While our model does not definitively resolve the relative importance of near and far solvent pools, the ALT active site lysine residue, and the PLP coenzyme in mediating exchange of H and D intermolecularly between alanine and water and intramolecularly between the α- and β-carbon atoms, it does give us confirmation that exchange is taking place over multiple time scales, as indicated by the "direct" exchange (intramolecular) rate constants, $k_{\alpha\beta_1}$, $k_{\alpha\beta_2}$, and $k_{\alpha\beta_3}$, which are consistently roughly an order of magnitude larger than the intermolecular rate constants, k_1-k_{10}.

References

Bigeleisen, Jacob and Maria Goeppert Mayer (1947). "Calculation of Equilibrium Constants for Isotopic Exchange Reactions". In: *The Journal of Chemical Physics* 15.5, pp. 261–267.

Bourg, Ian C and Carl I Steefel (2012). "Molecular dynamics simulations of water structure and diffusion in silica nanopores". In: *The Journal of Physical Chemistry C* 116.21, pp. 11556–11564. DOI: 10.1021/jp301299a.

Cerqueira, N M F S A, P A Fernandes, and M J Ramos (May 2011). "Computational Mechanistic Studies Addressed to the Transimination Reaction Present in All Pyridoxal 5'-Phosphate-Requiring Enzymes". English. In: *Journal of Chemical Theory and Computation* 7.5, pp. 1356–1368. ISSN: 1549-9618. DOI: 10.1021/ct1002219. URL: http://pubs.acs.org/doi/abs/10.1021/ct1002219.

Cooper, Arthur J L (1976). "Proton magnetic resonance studies of glutamate-alanine transaminase-catalyzed deuterium exchange. Evidence for proton conservation during prototropic transfer from the alpha carbon of L-alanine to the C4-position of pyridoxal 5'-phosphate." In: *Journal of Biological Chemistry* 251.4, pp. 1088–1096.

Dama, James F, Anton V Sinitskiy, Martin McCullagh, Jonathan Weare, Benoît Roux, Aaron R Dinner, and Gregory A Voth (2013). "The Theory of Ultra-Coarse-Graining. 1. General Principles". In: *Journal of Chemical Theory and Computation* 9.5, pp. 2466–2480.

DeRosa, G and Robert W Swick (1975). "Metabolic implications of the distribution of the alanine aminotransferase isoenzymes." In: *Journal of Biological Chemistry* 250.20, pp. 7961–7967.

Druhan, Jennifer L., Carl I. Steefel, Kenneth H. Williams, and Donald J. DePaolo (2013). "Calcium isotope fractionation in groundwater: Molecular scale processes influencing field scale behavior". In: *Geochimica et Cosmochimica Acta* 119, pp. 93–116. ISSN: 0016-7037. DOI: https://doi.org/10.1016/j.gca.2013.05.022. URL: https://www.sciencedirect.com/science/article/pii/S0016703713003098.

Duff, Stephen M G et al. (2012). "The Enzymology of alanine aminotransferase (AlaAT) isoforms from Hordeum vulgare and other organisms, and the HvAlaAT crystal structure". In: *Archives of Biochemistry and Biophysics* 528.1, pp. 90–101. DOI: https://doi.org/10.1016/j.abb.2012.06.006.

Eiler, John M (2007). "Clumped-isotope" geochemistry—The study of naturally-occurring, multiply-substituted isotopologues". In: *Earth and Planetary Science Letters* 262.3-4, pp. 309–327.

Escalera-Fanjul, Ximena, Carlos Campero-Basaldua, Maritrini Colón, James González, Dariel Márquez, and Alicia González (2017). "Evolutionary diversification of alanine transaminases in yeast: catabolic specialization and biosynthetic redundancy". In: *Frontiers in Microbiology* 8, p. 1150.

Golichowski, Alan, Richard C Harruff, and W Terry Jenkins (1977). "The effects of pH on the rates of isotope exchange catalyzed by alanine aminotransferase". In: *Archives of Biochemistry and Biophysics* 178.2, pp. 459–467.

Jang, Seung Soon, Valeria Molinero, Tahir Çagin, and William A Goddard (2004). "Nanophase-Segregation and Transport in Nafion 117 from Molecular Dynamics Simulations: Effect of Monomeric Sequence". In: *The Journal of Physical Chemistry B* 108.10, pp. 3149–3157.

Kratochvil, Huong T et al. (2023). "Transient water wires mediate selective proton transport in designed channel proteins". In: *Nature Chemistry*, pp. 1–10.

Peña-Soler, Esther, Francisco J. Fernandez, Miguel López-Estepa, Fernando Garces, Andrew J. Richardson, Juan F. Quintana, Kenneth E. Rudd, Miquel Coll, and M. Cristina Vega (July 2014). "Structural Analysis and Mutant Growth Properties Reveal Distinctive Enzymatic and Cellular Roles for the Three Major L-alanine Transaminases of Escherichia coli". In: *PloS ONE* 9.7, pp. 1–15. DOI: 10.1371/journal.pone.0102139. URL: https://doi.org/10.1371/journal.pone.0102139.

Pomès, R and B Roux (1996). "Structure and dynamics of a proton wire: a theoretical study of H+ translocation along the single-file water chain in the gramicidin A channel". In: *Biophysical Journal* 71.1, pp. 19–39.

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery (1986). *Numerical Recipes in FORTRAN 77, The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK.

Richet, P, Y Bottinga, and M Javoy (1977). "A review of hydrogen, carbon, nitrogen, oxygen, sulphur, and chlorine stable isotope fractionation among gaseous molecules". In: pp. 1–47.

Schauble, Edwin A, Prosenjit Ghosh, and John M. Eiler (2006). "Preferential formation of 13C–18O bonds in carbonate minerals, estimated using first-principles lattice dynamics". In: *Geochimica et Cosmochimica Acta* 70.10, pp. 2510–2529. ISSN: 0016-7037. DOI: 10.1016/j.gca.2006.02.011.

Steefel, Carl I., Jennifer L. Druhan, and Kate Maher (2014). "Modeling Coupled Chemical and Isotopic Equilibration Rates". In: *Procedia Earth and Planetary Science* 10, pp. 208–217. ISSN: 1878-5220. DOI: 10.1016/j.proeps.2014.08.022.

Urey, Harold C (1947). "The thermodynamic properties of isotopic substances". In: *Journal of the Chemical Society (Resumed)*, pp. 562–581.

Wang, Zhengrong, Edwin A. Schauble, and John M. Eiler (2004). "Equilibrium thermodynamics of multiply substituted isotopologues of molecular gases". In: *Geochimica et Cosmochimica Acta* 68.23, pp. 4779–4797. ISSN: 0016-7037. DOI: 10.1016/j.gca.2004.05.039.

Wisniewska, M. et al. (2009). *Human alanine aminotransferase 2 in complex with PLP*. DOI: https://doi.org/10.2210/pdb3IHJ/pdb.

Zhou, W et al. (2004). *Alanine aminotransferase from Pyrococcus furiosus Pfu-1397077-001*. DOI: https://doi.org/10.2210/pdb1XI9/pdb.

5.8 Supplementary Materials

5.8.1 Supplemental Attachments

EXPF_NMR.dat
EXPF.dat
EXPG_NMR.dat
EXPG.dat
EXPH_NMR.dat
EXPH.dat
EXPI_NMR.dat
EXPI.dat
EXPJ_NMR.dat
EXPJ.dat
EXPL_NMR.dat
EXPL.dat
EXPM_NMR.dat
EXPM.dat
EXPN_NMR.dat
EXPN.dat
EXPO_NMR.dat
EXPO.dat
EXPP_NMR.dat
EXPP.dat
EXPQ_NMR.dat
EXPQ.dat
Ch5_ExpQ-EQ-line-fit-abundances.dat

Ch5_ExpQ-EQ.py
FGHILP-data-only.py
QJMNO-data-only.py
GQ-scaling.py

5.8.2 Supplemental Tables

Species	Rep 1	Rep 2	Rep 3	Rep 4
$CHCH_3$	222.480	229.093	230.435	229.979
$CDCH_3$	236.019	245.950	246.097	247.078
$CHCH_2D$	609.488	620.531	623.125	620.487
$CDCH_2D$	661.000	674.958	679.125	676.416
$CHCHD_2$	546.114	558.130	558.975	559.434
$CDCHD_2$	593.846	607.529	610.596	607.115
$CHCD_3$	181.575	189.909	189.546	188.041
$CDCD_3$	193.191	200.959	200.518	197.409

Table 5.6: Line-fit abundances of four replicate measurements on alanine equilibrated with water from Experiment Q. These abundances are used to calculate the equilibrium constants given in Tables 5.2 and 5.3 as documented in attached Python script `Ch5_ExpQ-EQ.py`.

#	Species	Multiplicity	H:D	Mass (amu)
1	$H_3{}^{12}C-{}^{12}CH_3$	1	6:0	30.0470
2	$H_3{}^{12}C-{}^{13}CH_3$	1	6:0	31.0503
3	$H_3{}^{13}C-{}^{13}CH_3$	1	6:0	32.0537
4	$H_3{}^{12}C-{}^{12}CH_2D$, $DH_2{}^{12}C-{}^{12}CH_3$	6	5:1	31.0532
5	$H_3{}^{12}C-{}^{13}CH_2D$*	3	5:1	32.0566
6	$DH_2{}^{12}C-{}^{13}CH_3$*	3	5:1	32.0566
7	$H_3{}^{13}C-{}^{13}CH_2D$, $DH_2{}^{13}C-{}^{13}CH_3$	6	5:1	33.0599
8	$H_2D{}^{12}C-{}^{12}CH_2D$	9	4:2	32.0595
9	$H_2D{}^{12}C-{}^{13}CH_2D$	9	4:2	33.0629
10	$H_2D{}^{13}C-{}^{13}CH_2D$	9	4:2	34.0662
11	$H_3{}^{12}C-{}^{12}CHD_2$, $D_2H{}^{12}C-{}^{12}CH_3$	6	4:2	32.0595
12	$H_3{}^{12}C-{}^{13}CHD_2$*	3	4:2	33.0629
13	$D_2H{}^{12}C-{}^{13}CH_3$*	3	4:2	33.0629
14	$D_2H{}^{13}C-{}^{13}CH_3$	6	4:2	34.0662
15	$H_2D{}^{12}C-{}^{12}CHD_2$, $HD_2{}^{12}C-{}^{12}CH_2D$	18	3:3	33.0658
16	$H_2D{}^{12}C-{}^{13}CHD_2$*	9	3:3	34.0691
17	$HD_2{}^{12}C-{}^{13}CH_2D$*	9	3:3	34.0691
18	$H_2D{}^{13}C-{}^{13}CHD_2$, $HD_2{}^{13}C-{}^{13}CH_2D$	18	3:3	35.0725
19	$H_3{}^{12}C-{}^{12}CD_3$, $D_3{}^{12}C-{}^{12}CH_3$	2	3:3	33.0658
20	$H_3{}^{12}C-{}^{13}CD_3$*	1	3:3	34.0691
21	$D_3{}^{12}C-{}^{13}CH_3$*	1	3:3	34.0691
22	$H_3{}^{13}C-{}^{13}CD_3$, $D_3{}^{13}C-{}^{13}CH_3$	2	3:3	35.0725
23	$H_2D{}^{12}C-{}^{12}CD_3$, $D_3{}^{12}C-{}^{12}CH_2D$	6	2:4	34.0721
24	$H_2D{}^{12}C-{}^{13}CD_3$*	3	2:4	35.0754
25	$D_3{}^{12}C-{}^{13}CH_2D$*	3	2:4	35.0754
26	$H_2D{}^{13}C-{}^{13}CD_3$, $D_3{}^{13}C-{}^{13}CH_2D$	6	2:4	36.0788
27	$HD_2{}^{12}C-{}^{12}CHD_2$	9	2:4	34.0721
28	$HD_2{}^{12}C-{}^{13}CHD_2$	9	2:4	35.0754
29	$HD_2{}^{13}C-{}^{13}CHD_2$	9	2:4	36.0788
30	$HD_2{}^{12}C-{}^{12}CD_3$, $D_3{}^{12}C-{}^{12}CHD_2$	6	1:5	35.0783
31	$HD_2{}^{12}C-{}^{13}CD_3$*	3	1:5	36.0817
32	$D_3{}^{12}C-{}^{13}CHD_2$*	3	1:5	36.0817
33	$HD_2{}^{13}C-{}^{13}CD_3$, $D_3{}^{13}C-{}^{13}CHD_2$	6	1:5	37.0850
34	$D_3{}^{12}C-{}^{12}CD_3$	1	0:6	36.0846
35	$D_3{}^{12}C-{}^{13}CD_3$	1	0:6	37.0880
36	$D_3{}^{13}C-{}^{13}CD_3$	1	0:6	38.0913

Table 5.7: The 36 unique isotopic forms (isotopologues and isotopomers) of ethane (H_3C-CH_3), a possible additional molecular system for future further investigation. * denotes isotopomer pairs.

Model	k_{1_f}	k_{2_f}	k_{3_f}	k_{4_f}	k_{5_f}	k_{6_f}	k_{7_f}	k_{8_b}	k_{9_b}	k_{10_f}	$k_{\alpha\beta_{1_f}}$	$k_{\alpha\beta_{2_f}}$	$k_{\alpha\beta_{3_f}}$
1S0I	2.3016	2.3016	2.3016	2.3016	2.3016	2.3016	2.3016	2.3016	2.3016	2.3016	0.0000	0.0000	0.0000
1S1I	2.0099	2.0099	2.0099	2.0099	2.0099	2.0099	2.0099	2.0099	2.0099	2.0099	17.1854	17.1854	17.1854
1S2I	1.9883	1.9883	1.9883	1.9883	1.9883	1.9883	1.9883	1.9883	1.9883	1.9883	6.2085	55.1455	104.0825
2S0I	1.1135	2.8565	2.8565	1.1135	2.8565	2.8565	1.1135	2.8565	2.8565	1.1135	0.0000	0.0000	0.0000
2S1I	0.9743	2.7604	2.7604	0.9743	2.7604	2.7604	0.9743	2.7604	2.7604	0.9743	10.0625	10.0625	10.0625
2S2I	1.0290	2.5914	2.5914	1.0290	2.5914	2.5914	1.0290	2.5914	2.5914	1.0290	2.9427	28.2276	53.5125
3S0I	1.1106	3.1441	1.9284	1.1106	3.1441	1.9284	1.1106	3.1441	1.9284	1.1106	0.0000	0.0000	0.0000
3S1I	0.9918	3.0168	1.6544	0.9918	3.0168	1.6544	0.9918	3.0168	1.6544	0.9918	14.4630	14.4630	14.4630
3S2I	0.9934	2.9928	1.7564	0.9934	2.9928	1.7564	0.9934	2.9928	1.7564	0.9934	4.5117	22.2181	39.9245
4S0I	1.1092	2.9276	1.9289	1.1092	3.8348	1.9289	1.1092	2.9276	1.9289	1.1092	0.0000	0.0000	0.0000
4S1I	1.0094	2.7589	1.6394	1.0094	3.9020	1.6394	1.0094	2.7589	1.6394	1.0094	14.7464	14.7464	14.7464
4S2I	0.9914	2.7524	1.7523	0.9914	3.8960	1.7523	0.9914	2.7524	1.7523	0.9914	4.3925	22.9005	41.4084
$10S0I_A$	1.0243	2.7067	2.9927	5.8125	15.1224	0.5075	1.3727	2.2110	32.6628	1.8279	0.0000	0.0000	0.0000
$10S0I_B$	1.0513	2.7974	2.5718	0.6435	3.7569	1.5008	2.0746	3.0680	1.7909	1.7347	0.0000	0.0000	0.0000
10S1I	0.8186	2.8063	2.2597	1.1044	3.7412	1.2521	1.2345	2.7961	1.6233	1.3687	10.0176	20.0351	30.0527
10S2I	0.9296	2.7884	2.4394	0.7869	4.0492	1.2149	1.5867	2.6442	1.5808	1.2443	4.7184	26.1650	47.6116
$10S3I_A$	0.9405	2.7093	2.8817	0.3979	4.4451	0.4075	1.8572	2.7305	16.4969	1.7607	4.4436	62.3995	3.5657
$10S3I_B$	0.9317	2.7578	2.4301	0.7357	4.2045	1.1362	1.6052	2.6396	1.6030	1.2498	4.5894	38.4177	45.1290

Table 5.8: Parameters for 8-species models fitted to the "QJMNO" dataset ($\times 10^{-4}$ $M^{-1}min^{-1}$).

Model	$CHCH_3$	$CDCH_3$	$CHCH_2D$	$CDCH_2D$	$CHCHD_2$	$CDCHD_2$	$CHCD_3$	$CDCD_3$	Total Q	Total QJMNO
1S0I	4.7147	6.6535	5.0049	4.7689	5.9565	2.6534	3.2238	0.8498	33.8254	156.1131
1S1I	4.0134	5.8587	4.1074	3.6744	6.1063	1.8305	3.4849	1.2025	30.2781	141.8131
1S2I	4.3134	6.7906	4.6045	3.0675	6.0686	1.5255	3.1332	1.1498	30.6530	128.2469
2S0I	3.0244	2.6003	3.4641	1.3234	1.5908	2.0624	1.7754	0.7630	16.6039	126.1574
2S1I	2.7429	2.3807	4.1864	0.9563	1.4011	1.5532	1.8146	0.8056	15.8407	117.1725
2S2I	2.8562	2.2443	3.5022	1.0590	2.4486	1.5350	2.1813	0.8672	16.6937	92.7369
3S0I	3.9991	2.3650	3.8307	1.1626	1.1881	1.5346	1.6110	0.8959	16.5870	85.0908
3S1I	3.4578	2.0755	4.6050	0.8565	1.0716	1.3547	1.7583	1.2704	16.4498	68.0822
3S2I	3.3689	1.9955	4.3276	0.9276	1.3082	1.3316	1.8690	1.2074	16.3357	53.8371
4S0I	3.3774	2.4447	3.3760	0.9981	1.2016	1.5164	1.6234	0.8982	15.4359	84.2248
4S1I	2.9790	2.2113	3.3349	0.8096	1.0206	1.3199	1.7856	1.2576	14.7184	66.5873
4S2I	2.9081	2.0733	3.2503	0.8380	1.1011	1.3165	1.8713	1.2028	14.5615	52.3065
$10S0I_A$	2.7683	1.9629	6.0895	6.1682	1.8131	1.2178	1.8522	0.8826	22.7547	57.5441
$10S0I_B$	2.7506	2.3752	3.1900	0.8676	1.4201	2.1916	1.8252	0.7619	15.3823	66.6417
10S1I	2.5914	1.7489	3.2652	1.4922	1.1872	1.3587	1.9694	1.0614	14.6744	44.6027
10S2I	2.6059	1.9882	3.1738	1.3086	1.2449	1.6443	1.9784	1.0403	14.9846	39.8954
$10S3I_A$	2.5105	2.0374	2.9819	1.9304	1.4641	1.3390	1.8199	0.8252	14.9085	39.0331
$10S3I_B$	2.5741	1.9466	3.1220	1.4222	1.2771	1.5975	1.9967	1.0440	14.9802	39.7632

Table 5.9: Misfit values for Experiment "Q" for 8-species models fitted to "QJMNO" dataset.

Model	J_α	J_β	M_α	M_β	N_α	N_β	O_α	O_β	Transient	Major	Total JMNO	Total QJMNO
1S0I	10.0462	18.6738	5.7584	13.6568	25.6548	4.8667	7.0120	36.6190	35.5817	86.7060	122.2877	156.1131
1S1I	4.8010	15.0726	5.0879	11.3001	26.5105	6.3785	0.8375	41.5468	23.3171	88.2179	111.5350	141.8131
1S2I	2.6044	18.6939	9.0407	3.8196	24.1386	2.0949	7.4429	29.7589	15.9618	81.6321	97.5939	128.2469
2S0I	11.7581	43.5909	14.6318	11.2321	3.4515	5.1694	7.4327	12.2870	35.5923	73.9612	109.5535	126.1574
2S1I	6.7212	45.4736	15.6984	8.2580	6.8989	5.3872	1.4771	11.4174	21.8435	79.4883	101.3317	117.1725
2S2I	5.1335	40.0264	5.4317	2.4096	3.2880	2.5613	4.8292	12.3636	14.9336	61.1096	76.0432	92.7369
3S0I	12.1616	7.5024	14.7486	13.0501	3.3816	5.2085	7.3329	5.1181	37.7530	30.7508	68.5038	85.0908
3S1I	2.1297	4.3691	13.7010	7.4941	10.3409	7.0731	1.2780	5.2465	17.9749	33.6576	51.6325	68.0822
3S2I	4.8431	8.1981	8.3352	2.6748	3.3464	2.1537	2.8827	5.0673	12.5543	24.9471	37.5014	53.8371
4S0I	12.1656	7.4019	14.7966	13.5149	3.3502	5.2647	7.3440	4.9509	38.2892	30.4997	68.7889	84.2248
4S1I	2.0232	3.8597	13.0886	8.0350	10.9031	7.7565	1.3181	4.8848	19.1328	32.7361	51.8689	66.5873
4S2I	4.8662	8.0511	8.0923	2.6833	3.2304	2.2027	2.9609	5.6580	12.7132	25.0318	37.7450	52.3065
$10S0I_A$	4.0992	2.1498	3.0891	4.7728	2.8629	5.2563	6.5949	5.9645	20.7232	14.0663	34.7894	57.5441
$10S0I_B$	12.4864	0.5335	3.1904	14.3921	2.5814	5.2766	7.5447	5.2543	39.6998	11.5596	51.2594	66.6417
10S1I	4.1546	0.4994	3.6933	5.6918	2.4632	6.0771	2.4478	4.9010	18.3714	11.5570	29.9283	44.6027
10S2I	4.8970	0.4967	3.3686	2.7020	2.5927	2.3352	3.2769	5.2417	13.2111	11.6997	24.9108	39.8954
$10S3I_A$	2.7667	1.9335	3.0230	4.6256	2.5789	2.3194	1.4718	5.4056	11.1836	12.9410	24.1246	39.0331
$10S3I_B$	4.0776	0.5002	3.3034	2.7880	2.6017	2.3171	3.7105	5.4846	12.8933	11.8898	24.7831	39.7632

Table 5.10: Misfit values for Experiments "J", "M", "N", and "O" for 8-species models fitted to "QJMNO" dataset.

Model	Total, F-P	F_α	F_β	G_α	G_β	H_α	H_β	I_α	I_β	L_α	L_β	P_α	P_β
1S0I	200.4777	16.9859	21.2097	12.2117	15.1601	17.0337	21.7526	14.8021	17.8916	8.2977	17.9384	16.0549	21.1392
1S1I	190.9556	12.2808	23.1902	8.9226	18.3079	13.0889	22.3640	11.1007	20.1183	6.6341	19.2683	10.9880	24.6917
1S2I	198.9599	11.2136	25.6415	8.9641	19.6505	12.7181	23.7766	10.9135	21.3431	6.9830	20.7245	9.9669	27.0644
2S0I	145.7547	3.0791	26.6840	2.1562	10.5697	4.3444	29.4858	3.2041	19.1331	3.6482	12.4722	6.1847	24.7931
2S1I	144.2136	5.5295	24.8868	1.7101	10.6477	3.4513	27.5128	2.8181	18.1501	2.6163	13.3960	10.2623	23.2326
2S2I	135.4353	3.8128	22.7803	1.4476	11.6593	3.4826	24.7863	2.4709	17.3660	3.0621	14.8131	8.1054	21.6489
3S0I	109.9436	3.0004	16.3115	2.1067	8.3505	4.2743	24.4514	3.1456	15.2547	3.6109	11.4004	6.1204	11.9167
3S1I	104.9658	5.6754	12.6612	1.7190	8.6785	3.7367	20.2730	2.9330	13.3794	2.8357	12.6609	10.4734	9.9397
3S2I	104.5058	4.7732	13.5229	1.3947	8.8081	3.4164	20.3218	2.4550	13.4619	2.8679	12.6791	9.6436	11.1613
4S0I	111.2882	3.0125	16.4019	2.0924	8.7425	4.2570	24.3516	3.1309	15.3739	3.6075	12.3152	6.1551	11.8478
4S1I	106.4565	5.3748	12.7267	1.6787	9.3405	3.7068	20.5231	2.8802	13.8313	2.8468	13.6985	10.0568	9.7924
4S2I	106.1861	4.6765	13.5356	1.3773	9.3747	3.4100	20.5213	2.4237	13.8011	2.8441	13.7153	9.5640	10.9425
$10S0I_A$	125.5199	8.0053	14.9076	6.4518	7.5010	9.6501	24.6976	8.0673	14.3089	5.0982	12.2841	7.1478	7.4003
$10S0I_B$	109.3413	3.7499	16.1948	1.3900	9.8448	4.5318	25.9539	2.7969	17.0624	2.5595	12.5216	3.0373	9.6985
10S1I	96.1056	2.9186	11.8241	1.2925	9.0129	2.5868	21.4038	1.8637	14.1155	2.0549	13.1358	7.1050	8.7921
10S2I	94.6151	1.7416	11.7079	1.0857	8.9250	3.2416	22.3867	1.9520	14.5807	2.3056	12.8871	5.8950	7.9063
$10S3I_A$	104.6129	2.0317	15.9607	1.1296	8.9474	3.3371	23.9221	2.0047	15.1750	2.0859	13.1575	4.3692	12.4921
$10S3I_B$	94.7154	1.8376	11.4405	1.1265	8.9856	3.3770	21.9609	2.0594	14.3708	2.3460	13.0916	5.8527	8.2668

Table 5.11: Misfit values for Experiments "F", "G", "H", "I", "L", and "P" for 8-species models fitted to "QJMNO" dataset, as generated by the code in Supplemental Attachment `20240219-FORTRAN.tar`.

Misfits for Experiments FGHILP (used in fitting)													
Model	F_α	F_β	G_α	G_β	H_α	H_β	I_α	I_β	L_α	L_β	P_α	P_β	Total
8-Species	3.2418	15.6164	2.2429	8.7042	4.3729	16.7609	3.3351	10.8675	3.3695	14.0173	6.3575	21.0280	109.9141
3-Box	3.2418	15.6164	2.2429	8.7042	4.3729	16.7609	3.3351	10.8675	3.3695	14.0173	6.3575	21.0280	109.9141

Misfits for D-Labeled Experiments JMNO (not used in fitting)										
Model	J_α	J_β	M_α	M_β	N_α	N_β	O_α	O_β	Total	FGHILP+JMNO Total
8-Species	11.9878	26.6856	14.9872	12.7559	2.9986	5.1942	7.5534	21.5557	103.7184	213.63252496084010
3-Box	11.9878	26.6856	14.9872	12.7559	2.9986	5.1942	7.5534	21.5557	103.7184	213.63252493498513

Table 5.12: Misfit equivalence between 3-Box and 8-Species models, as generated by Supplemental Attachments `20240122-8SPEC-TO-3BOX.tar` and `20240122-3BOX.tar`. These models are equivalent to seven decimals.

5.8.3 Supplemental Figures

Here we provide several figures referenced, but not provided, in the main body of this chapter.

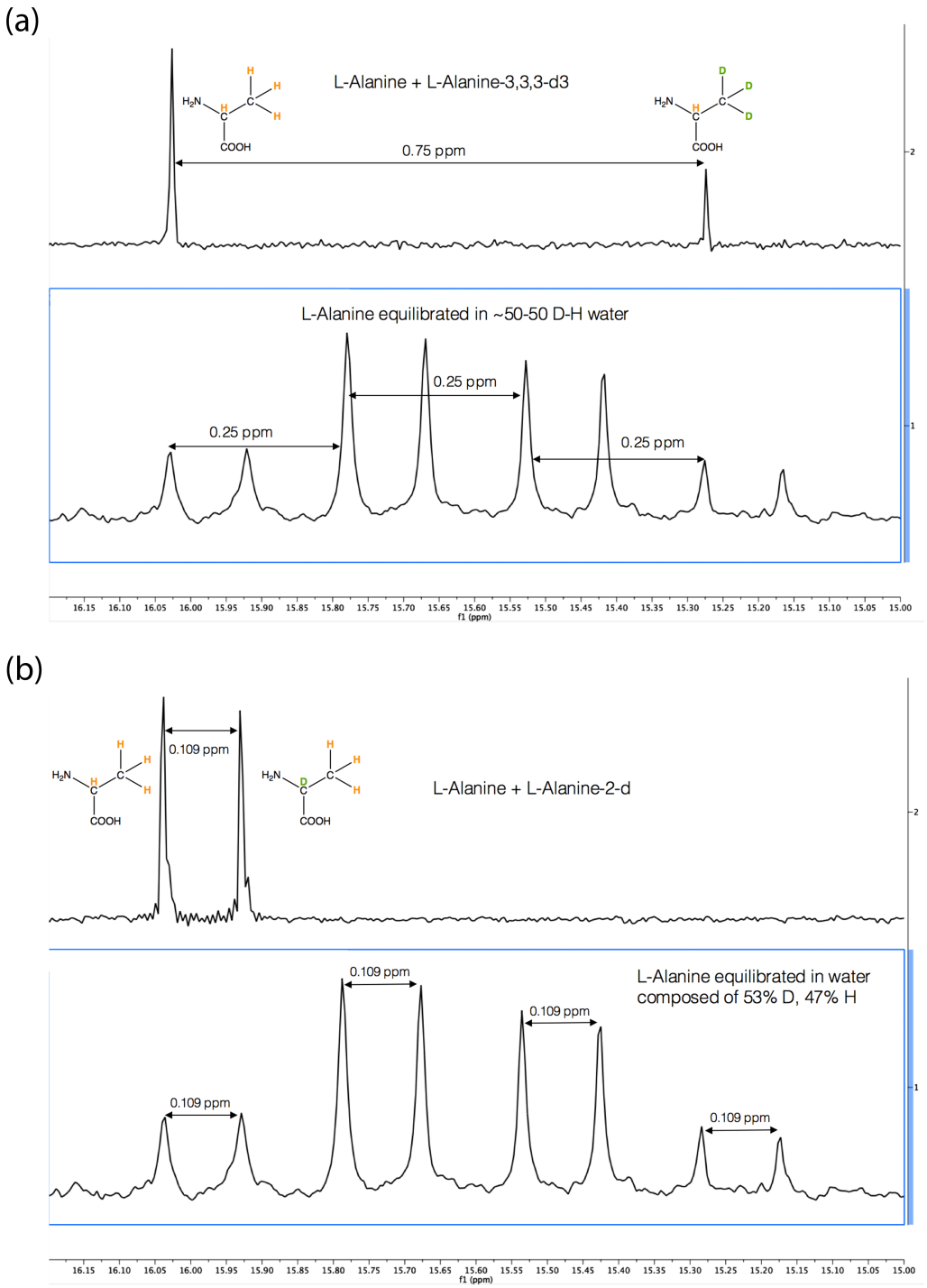

Figure 5.17: Verification by ^{13}C NMR of the identity of the isotopic species in our reactions, as described in Section 5.3.3.2. Distances between sets of peaks are measured in ppm. (a) Two vertically stacked spectra are shown, with alanine-$CHCH_3$ and alanine-$CHCD_3$ in the top spectrum, while the bottom spectrum contains all eight isotopic variants for comparison. For each additional ^2H on the β carbon, we observe a chemical shift of the β carbon signal for that species of 0.25 ppm, totaling 0.75 ppm shift between alanine-$CHCH_3$ and alanine-$CHCD_3$. (b) Two vertically stacked spectra are shown, with alanine-$CHCH_3$ and alanine-$CDCH_3$ in the top spectrum, while the bottom spectrum contains all eight isotopic variants for comparison. The two-bond shift that results from adding a ^2H at the α carbon position is 0.109 ppm.

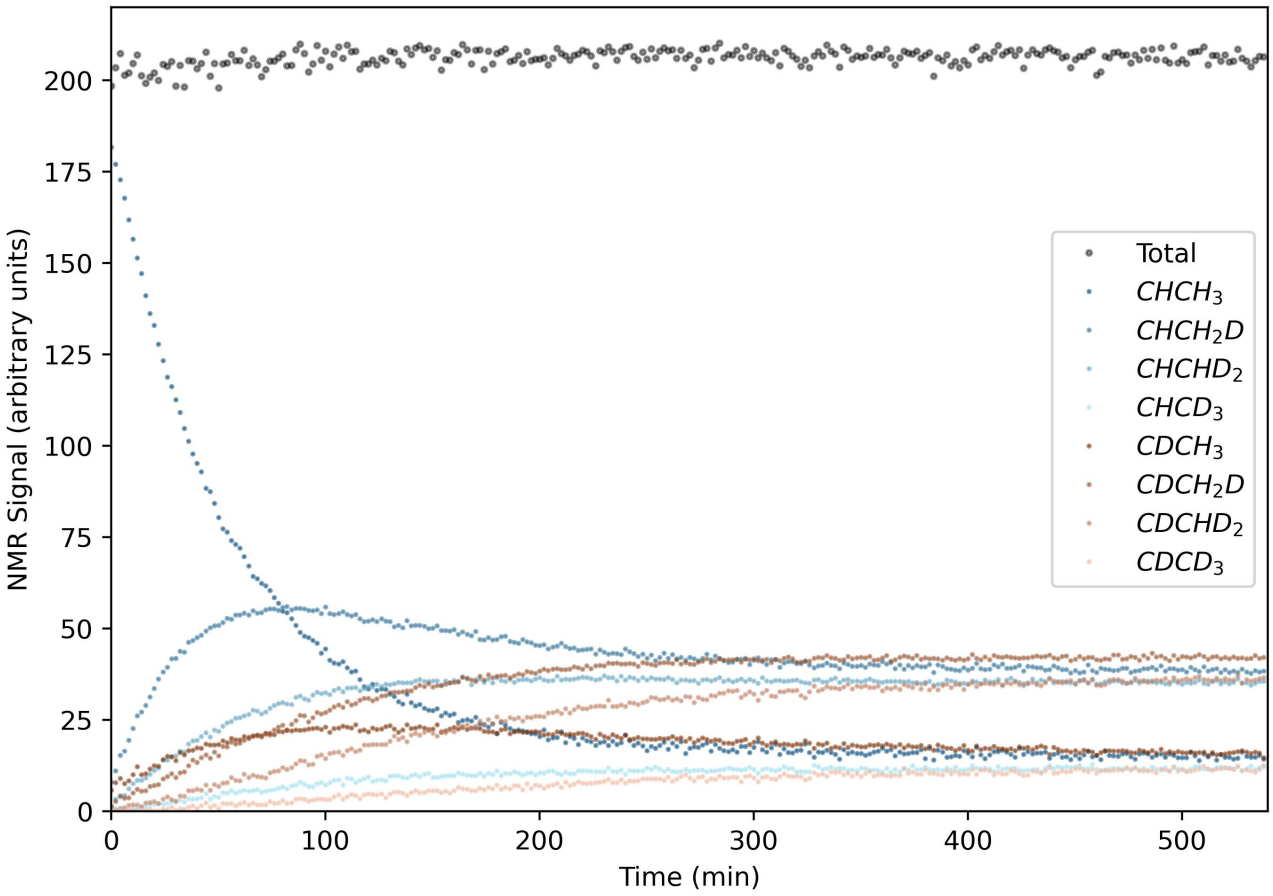

Figure 5.18: Time series of the eight isotopic variants in Experiment "Q" using single-scan ^{13}C NMR acquisitions taken every two minutes for 540 minutes (9 hours) of reaction time, as depicted in Figure 5.8, with added series giving the sum of ^{13}C NMR signal contribution from each of the eight isotopic variants measured in this experiment at each time step (black circles), to demonstrate that this sum is conserved (with some measurement noise) throughout this reaction.

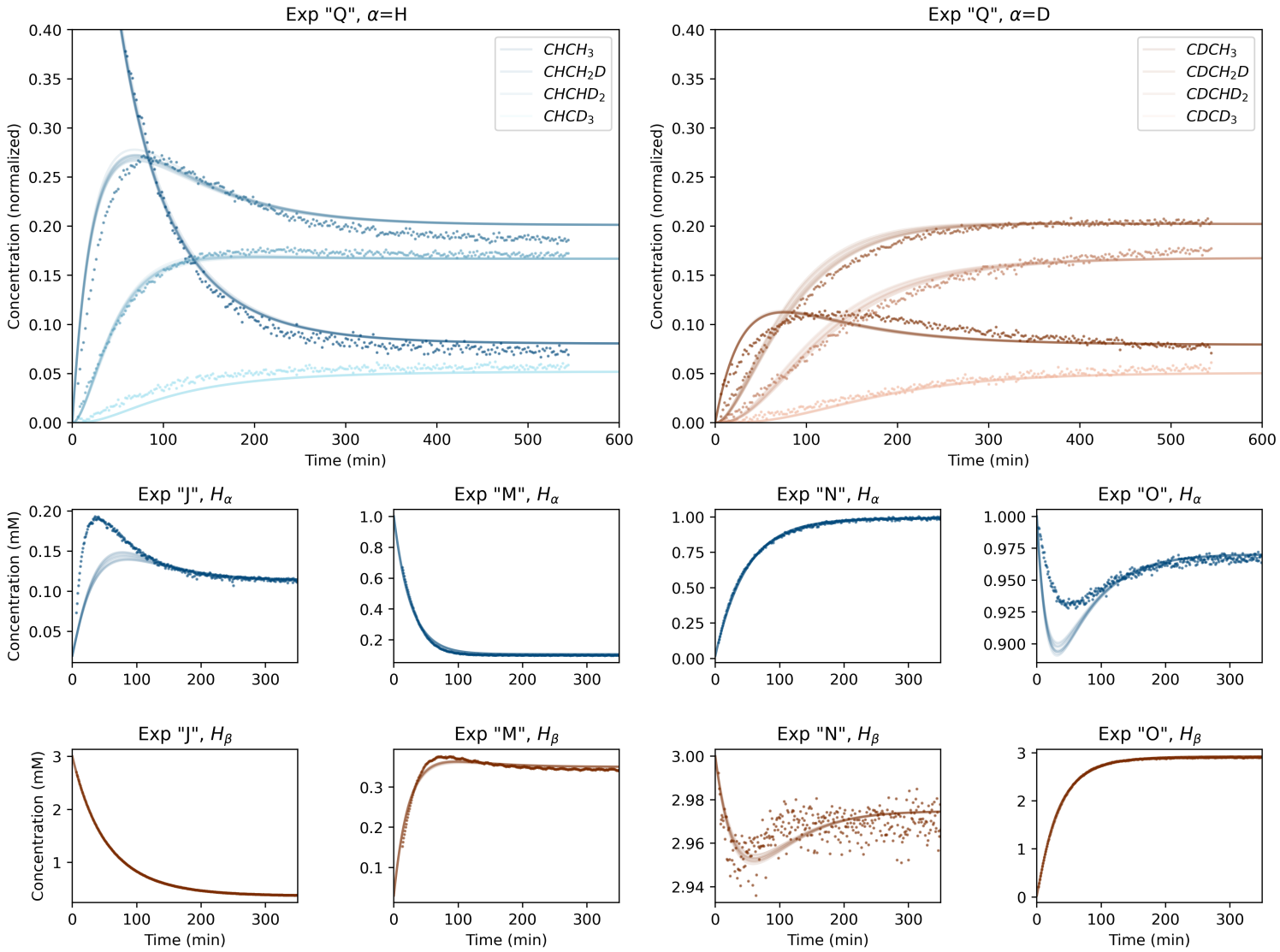

Figure 5.19: A random sampling of ten 10S2I model fits to the QJMNO experimental dataset in the range of misfit=39.89–40.28 (top ten percent of best fits), to illustrate the variability of these fits to experimental data. The variability of these fits is low, however in Experiment "Q", the fits to some isotopic species (e.g., $CHCH_2D$ and $CDCH_2D$) can be seen to vary more than others (e.g., $CHCH_3$ and $CDCH_3$).

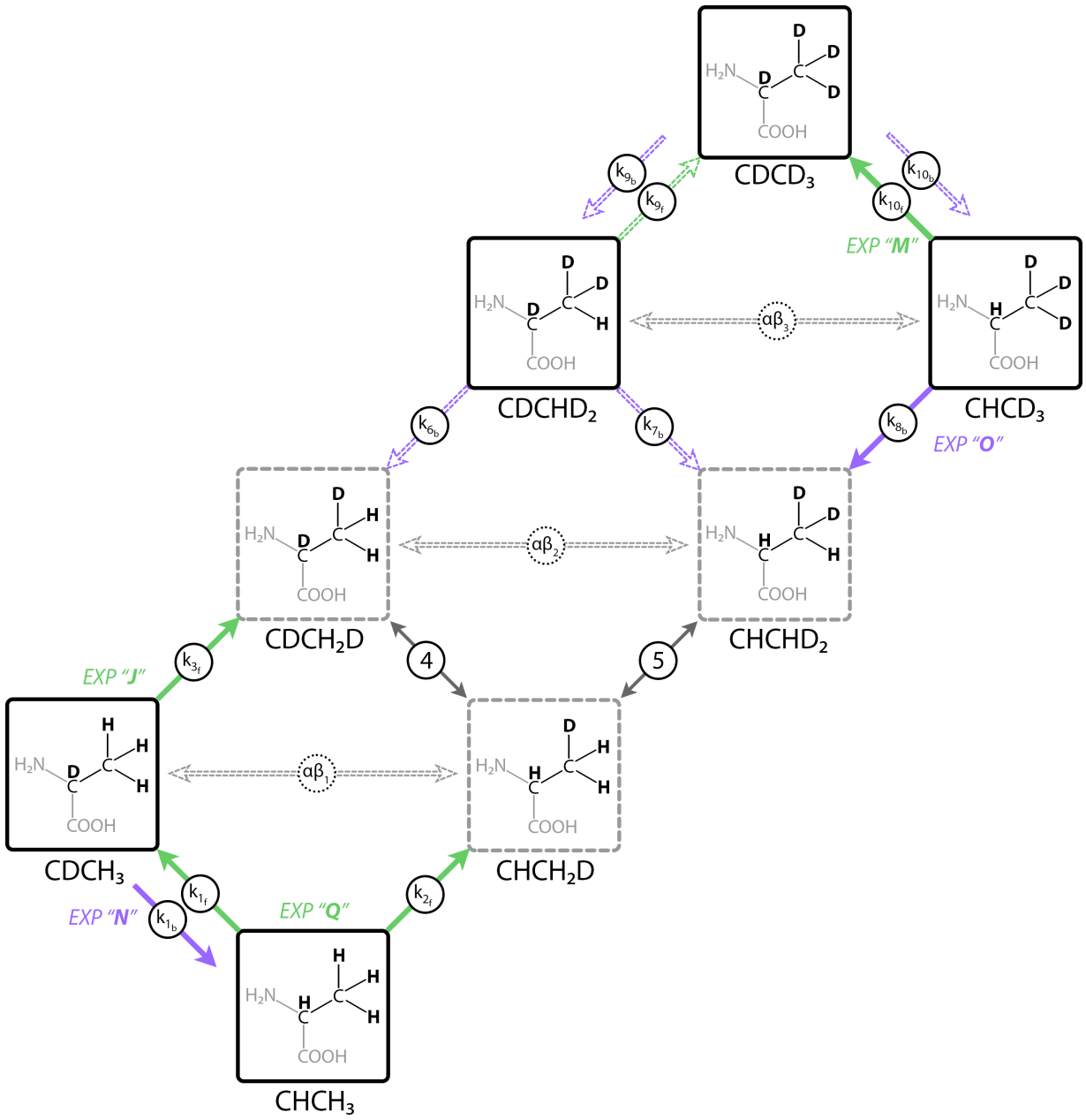

Figure 5.20: Graphical depiction of the system of ten reactions (circled) that convert the eight possible isotopic variants of alanine in our study from one to another (see also Figure 5.9). Forward reactions (up) remove a D from the solvent and emit an H to the solvent. Backward reactions (down) remove an H from the solvent and emit a D to the solvent. Colored arrows indicate kinetic rate constants most directly observed (and thus the direction in which each experiment proceeds) during the initial stages of reactions starting from pure isotopic variants, where purple (backward) reactions take place in water enriched with deuterium relative to the initial alanine isotopic variant, while green (forward) reactions take place in water depleted in deuterium relative to the initial alanine isotopic variant. Reactions explored via experiments presented in this work are labeled "J", "M", "N", "O", and "Q" as described in Table 5.1. Of these, "Q" is the only experiment performed with ^{13}C NMR. Isotopic variants available for purchase from Sigma Aldrich (SA) or Cambridge Isotope Laboratories (CIL), or able to be produced in our facilities, are outlined with solid black boxes, while those that are not available are outlined with dashed gray boxes.

5.8.4 Data Adjustments

5.8.4.1 Scaling

The proton NMR data from experiments in different D_2O:H_2O solution compositions (Experiments "F", "G", "H", "I", "L", and "P") as well as the D-labeled experiments beginning with either $CDCH_3$ (Experiments "J" and "N") or $CHCD_3$ (Experiments "O" and "M") can be compared with model predictions by summing the total H on the α and β sites in the model. In merging the ^{13}C NMR data (Experiment "Q", conducted 9/13/2019) with the proton NMR data (Experiments "F", "G", "H", "I", "L", "P" and "J", "M", "N", "O"), (conducted about two months earlier) we noted that the isotopic exchange rates seemed significantly faster for the proton NMR experiments. Experiment "G" (conducted 07/05/2019) has roughly the same solution composition as Experiment "Q" (see Table 5.1), so they can be directly compared without having to use a model. Figure 5.21 shows that the reaction in Experiment "G" proceeds roughly twice as fast as Experiment "Q". It is most likely that either there was some degradation of the enzyme over the time gap between these two experiments, or an issue with the addition of enzyme via pipette to initiate this reaction (e.g., an accidental bubble in the pipette tip would have reduced the amount of enzyme added). Based on Figure 5.21, we multiply the time scale by 2.0 when using any of the proton NMR experiments to constrain the model. An obvious question would be whether there was a systematic gradual degradation over the time period between early July and mid September, or whether there was a single event (for example, if the enzyme was accidentally allowed to sit at room temperature for a prolonged period). We find no indication of the later experiments being systematically slow, for example, Experiment "P" (conducted 8/15/2019), suggesting that the enzyme was not undergoing gradual degradation in the time between Experiment "P" and Experiment "Q". Since we performed no further kinetics experiments after Experiment "Q", we cannot distinguish between these two most plausible scenarios. Given this discrepancy, we only use the D-labeled experiments ("J", "M", "N", "O") in fitting the model, and then use the model thus produced to make predictions for Experiments "F", "G", "H", "I", "L", and "P". As we will see, the QJMNO-fitted model makes reasonably good predictions for the concentration dependence of the integrated α and β rates, so we do not expect any major issues using the "J", "M", "N", "O" experimental data to constrain the rate constants.

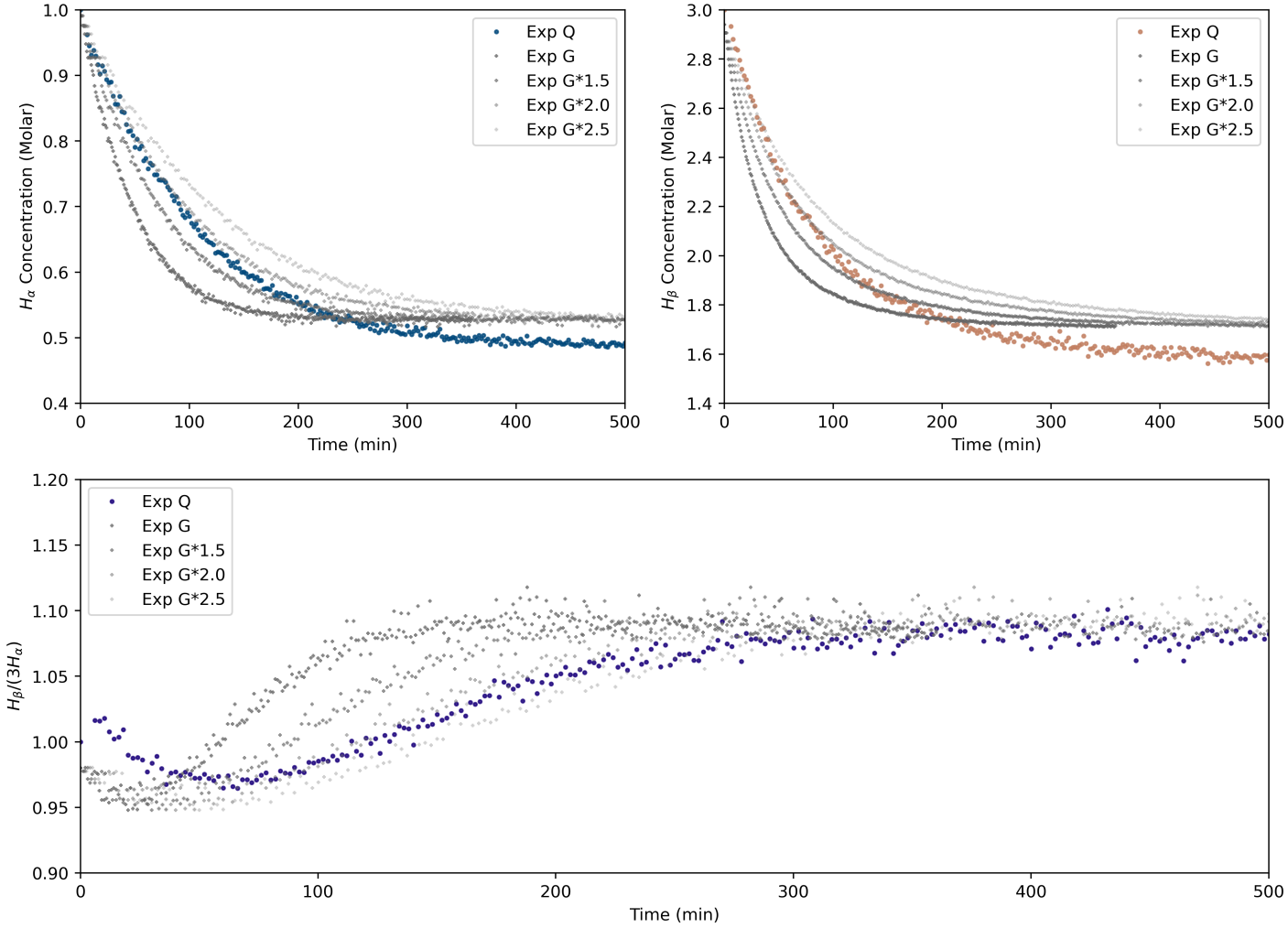

Figure 5.21: Depiction of method used to determine scaling Factor applied to Experiment "Q". Hydrogen isotope exchange reactions in Experiment "G" (acquired using ^{1}H NMR) and Experiment "Q" (acquired using ^{13}C NMR) took place in similar water D/H composition, and with the same initial $CHCH_3$ alanine hydrogen isotope composition. Several time scaling factors (1.5, 2.0, and 2.5) applied to Experiment G serve to illustrate that the reaction in Experiment "G" proceeds roughly twice as fast as Experiment "Q". We use this empirically derived factor of two to adjust the rate of Experiment "Q" so that is consistent with the earlier ^{1}H NMR experiments this entire dataset (from Experiments "Q", "J", "M", "N", and "O") can be incorporated into the same modeling effort.

5.8.4.2 T_0 Adjustments for Experiments Q, J, M, N, and O

As mentioned in Section 5.3.2, there is a delay on the order of several minutes between the addition of enzyme to the reaction mixture, marking the true beginning of the kinetics experiments, and the beginning of data collection. We did not anticipate the modeling work later conducted when performing these experiments, so this delay, while kept to a minimum with the equipment available, was not quantified with a stopwatch. This delay could in theory be eliminated with a specialized NMR tube, a "Sample Reaction System" by New Era, which we did not possess at the time. In order to combine this set of experiments so that they may provide independent constraints in our model, we attempt to account for this "missing" data, and correct for this time delay using two methods, depending on the nucleus observed for each experiment (^{13}C for Experiment "Q", ^{1}H for Experiments "J", "M", "N", and "O"). In the case of Experiment "Q", where the ^{13}C NMR observations provide the populations of each of the isotopic variants sum to a value that is constant (with some measurement noise, as depicted in Figure 5.18) over the course of the experiment, the most straightforward approach is to fit an exponential decay function to the early values of the concentration of the $CHCH_3$ variant and extrapolate it back to unity. This is possible because we know that the initial alanine was effectively entirely made up of the $CHCH_3$ variant, with all other variants initially at zero abundance. The result of this procedure is shown in Figure 5.22. The exponential function reaches unity at approximately -6.1 minutes. Thus, we estimate the time of the first data point collected to be 6.1 minutes after initial enzyme addition.

Correcting for the delay in the ^{1}H NMR Experiments "J", "M", "N", and "O" is slightly more involved, since these experiments do not yield conserved signal. In these cases we know the initial alanine composition has one site (either α or β) labeled with deuterium, so the ^{1}H NMR peak area of protons at that site is effectively zero. With this knowledge we can estimate the time delay by fitting a quadratic function to the initial 20–40 data points, and then use the fitted function to estimate when the signal from the labeled group goes to zero (refer to top panel of Figure 5.23 for the quadratic fit to Experiment "J" as an example; quadratic fits for Experiments "M", "N", and "O" are similar). We then shift the time by this amount and extrapolate the signal arising from the unlabeled site (either α or β) back to zero time (see bottom panel of Figure 5.23). Lastly, we normalize both the labeled and unlabeled

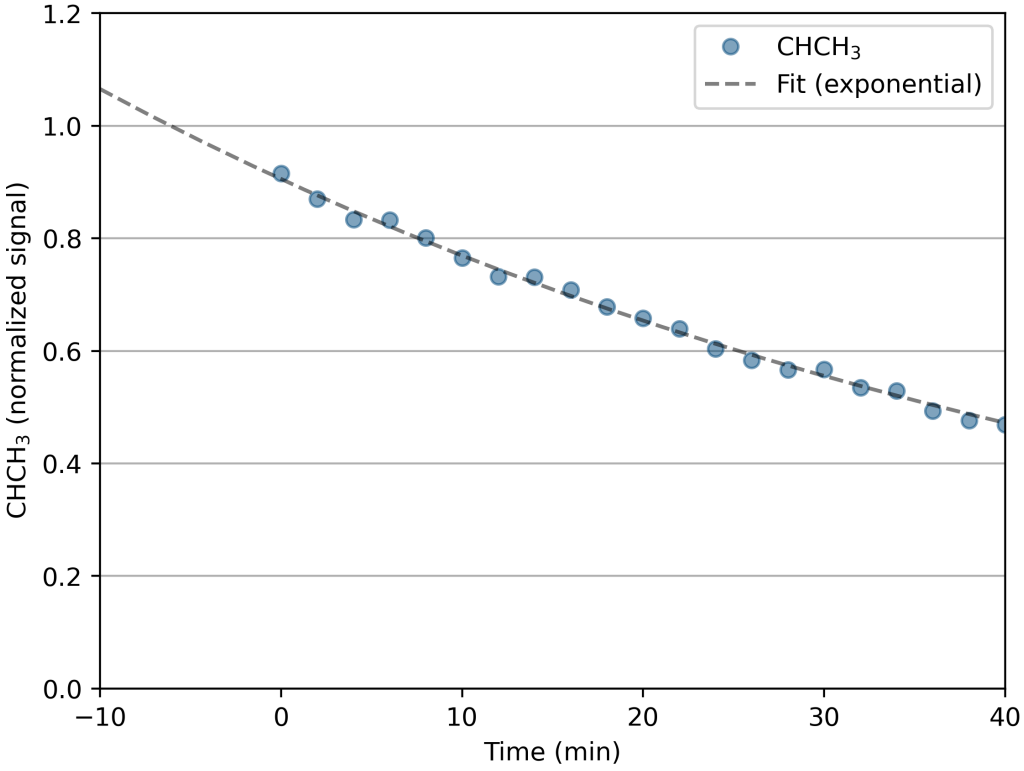

Figure 5.22: Fit of an exponential function to Experiment "Q" to determine delay between reaction initiation and start of analysis.

signal by the estimated zero-time NMR signal for the unlabeled component. Python scripts for processing the NMR signal for Experiments "Q", "J", "M", "N", and "O" are attached in Section 5.8.8.3.

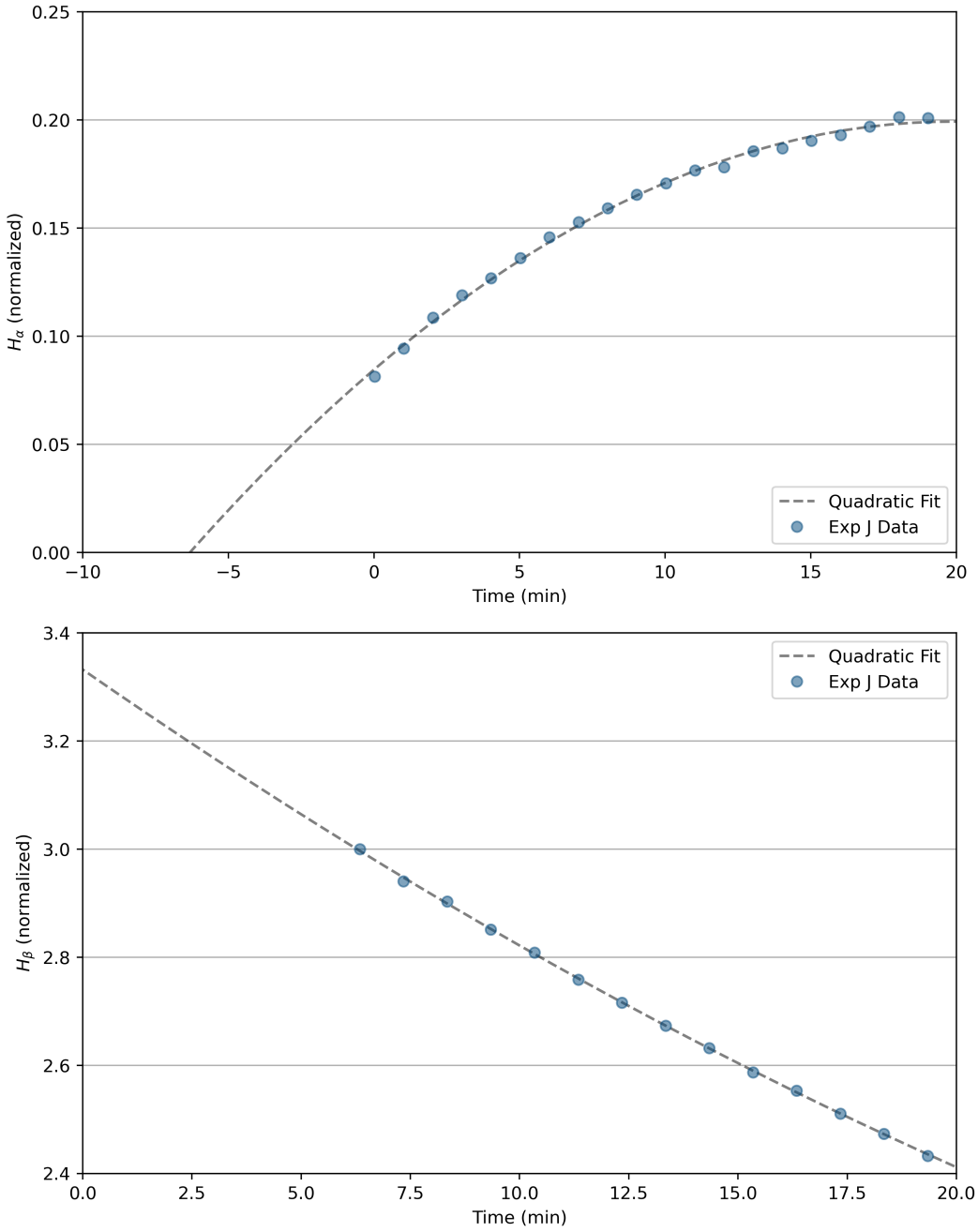

Figure 5.23: Fit of a quadratic function to Experiment "J" to determine delay between reaction initiation and start of analysis. Quadratic fits for Experiments "M", "N", and "O" are similar.

5.8.5 Stochastic Distributions of Isotopic Variants at Equilibrium

While we have already presented the measured (and predicted, where possible) equilibrium constants for the isotopic exchange reactions involved in our experiments in Section 5.4.3, these results can also be presented in a format perhaps more familiar to geochemists studying clumped isotopes. A common basis for standardization in clumped isotope geochemistry is a state in which rare isotopes of known abundance are stochastically distributed among all possible isotopic variants of a molecule. For an isotopic variant of alanine (CH_3CHNH_2COOH), i, deviations from a random distribution are typically reported as Δ_i values:

$$\Delta_i = \frac{{}^iR}{{}^iR^*} - 1 \tag{5.49}$$

where ${}^iR = [i]/[{}^{12}CH_3{}^{12}CH{}^{14}NH_2{}^{12}C{}^{16}O_2H]$ (here the denominator would conventionally be the isotopic variant with no heavy isotope substitutions, i.e., the most common variant under conditions where isotopes are present at natural abundance) and ${}^iR^*$ denotes the value of ${}^iR$ predicted for the sample if the isotopes are stochastically distributed among all possible isotopic variants. A unique aspect of the experimental work presented in Section 5.4.3 is the presence of ^{13}C at the β carbon position for all eight quantified species, such that it is not reasonable to use the common definition for ${}^iR$. Here we are also concerned primarily with variants of alanine containing ^1H and/or ^2H (D) in its α- and β-carbon positions, i.e., the four "non-exchangeable" hydrogen atoms of alanine. In contrast with the hydrogen atoms at the carboxyl and amine sites, which spontaneously and rapidly (over timescales of microseconds to milliseconds) exchange with water when in solution, exchange of the hydrogen atoms at the α and β carbon sites requires an enzyme catalyst, and occurs at rates that are observable using the techniques applied in this study. For the remainder of this work, we do not consider the isotopic variation in the atoms making up the carboxyl and amine groups of alanine, for they are not directly involved with the hydrogen exchange reaction, and are not measured in our experiments.

Unlike typical studies in clumped isotope geochemistry, where isotopic abundances are close to natural terrestrial abundance (e.g., deuterium makes up only $\sim$0.015% of the total hydrogen), our experiments involve the incorporation of unnatural isotopic abundances (i.e., high abundances of heavy isotopes,

^{2}H and/or ^{13}C) to get insight into isotopic processes that would otherwise be difficult to probe. At natural abundance, the stochastic probability of having $CHCH_3$ will be 0.999, while that of $CDCD_3$ is $6 * 10^{-16}$ (here we are considering ^{1}H and ^{2}H isotopes only, i.e., ignoring variation in carbon isotopes). In the case of alanine, its naturally rare multiply substituted species range from difficult to impossible to measure with existing tools, and to our knowledge no prior observations exist for these isotopic species. In contrast, an experiment in which the hydrogen atoms in a species of interest are allowed to equilibrate with water that has a subequal concentration of deuterium and hydrogen results in a set of isotopic species that are close enough in relative abundance to be studied with existing tools. For example, equilibrating alanine's hydrogen using a water composition of 1:1 D_2O:H_2O results in a distribution of isotopic species that vary in relative stochastic abundance by only a factor of three (the four least abundant isotopic species that arise under these conditions each contribute ∼6.25% of the total, while the four most abundant each contribute ∼18.75%). Designing experiments in this way allowed us to measure the degree to which each of these eight isotopic variants diverge from a stochastic distribution upon equilibration, leading us to a deeper understanding of isotopic site preferences and clumping tendencies than would otherwise be possible.

In this context, where the most common isotopic variant is a species with multiple heavy isotope substitutions, and the quantities of all eight relevant isotopic variants are measured directly (rather than as ratios to a more abundant species, as is the case in isotope ratio mass spectrometry) a more straightforward way to calculate Δ_i is by using

$$\Delta_i = \frac{{}^{i}F}{{}^{i}F^{*}} - 1 \tag{5.50}$$

where ${}^{i}F = [i]/[Total]$, where $[Total]$ is the summed area of the signal arising from all eight isotopic species, and ${}^{i}F^{*}$ represents the expected percent of the total signal given a stochastic distribution. Using this definition, the isotopic variant containing the fewest heavy isotopes (in this case, $^{12}CH^{13}CH_3$) is no longer forced to have a Δ of 0 ‰, and the values for all eight species are given relative to their expected stochastic abundance. When calculated using this definition, the resulting Δ values shift by an amount approximately equal to the Δ value of the $^{12}C^{1}H^{13}C^{1}H_3$ species.

Stochastic distributions of the eight isotopic variants measured were determined based on measured D_2O concentration of the sample using ^2H NMR as described in the previous section. This value was measured to be 0.483:0.517 D_2O:H_2O. As a result, of course, the stochastic distributions that emerge from this water composition differ dramatically from those we would calculate in a natural abundance context. Because the ^{13}C abundances at the two carbon sites of interest are fixed (non-random, and non-exchangeable) in our experiments, we do not use ^{13}C abundance in our determination of stochastic distribution. Our computational work presented in Chapter 2 (Section 2.7) predicts a 5 ‰ ^{13}C-D clumping effect at the β carbon site, relatively minor compared to the H-D site preference of ∼170 ‰.

Our results indicate that we should not have any clumping across α and β, and that we should expect a progressive decrease of ∼16‰ for each D present in the β group.

The resulting fitted peak areas were used to calculate the relative abundances of each of the eight isotopic species and the Δ values for each species. These values are presented in Figure 5.24 and Table 5.13.

	Δ(‰)	STERR	MC error
$^{12}CH^{13}CH_3$	-36	1.7	4.8
$^{12}CD^{13}CH_3$	104	4.9	4.9
$^{12}CH^{13}CH_2D$	-67	1.7	1.5
$^{12}CD^{13}CH_2D$	87	1.0	1.4
$^{12}CH^{13}CHD_2$	-102	1.1	1.5
$^{12}CD^{13}CHD_2$	46	0.7	1.6
$^{12}CH^{13}CD_3$	-29	4.0	3.5
$^{12}CD^{13}CD_3$	100	4.4	3.7

Table 5.13: Values of Δ_i as shown in Figure 5.24. Standard errors are given for four replicate measurements. Monte Carlo simulated errors using 1E6 iterations, based on relative error using 1/SNR of peak intensities, are given for comparison and are in broad agreement with measurement error.

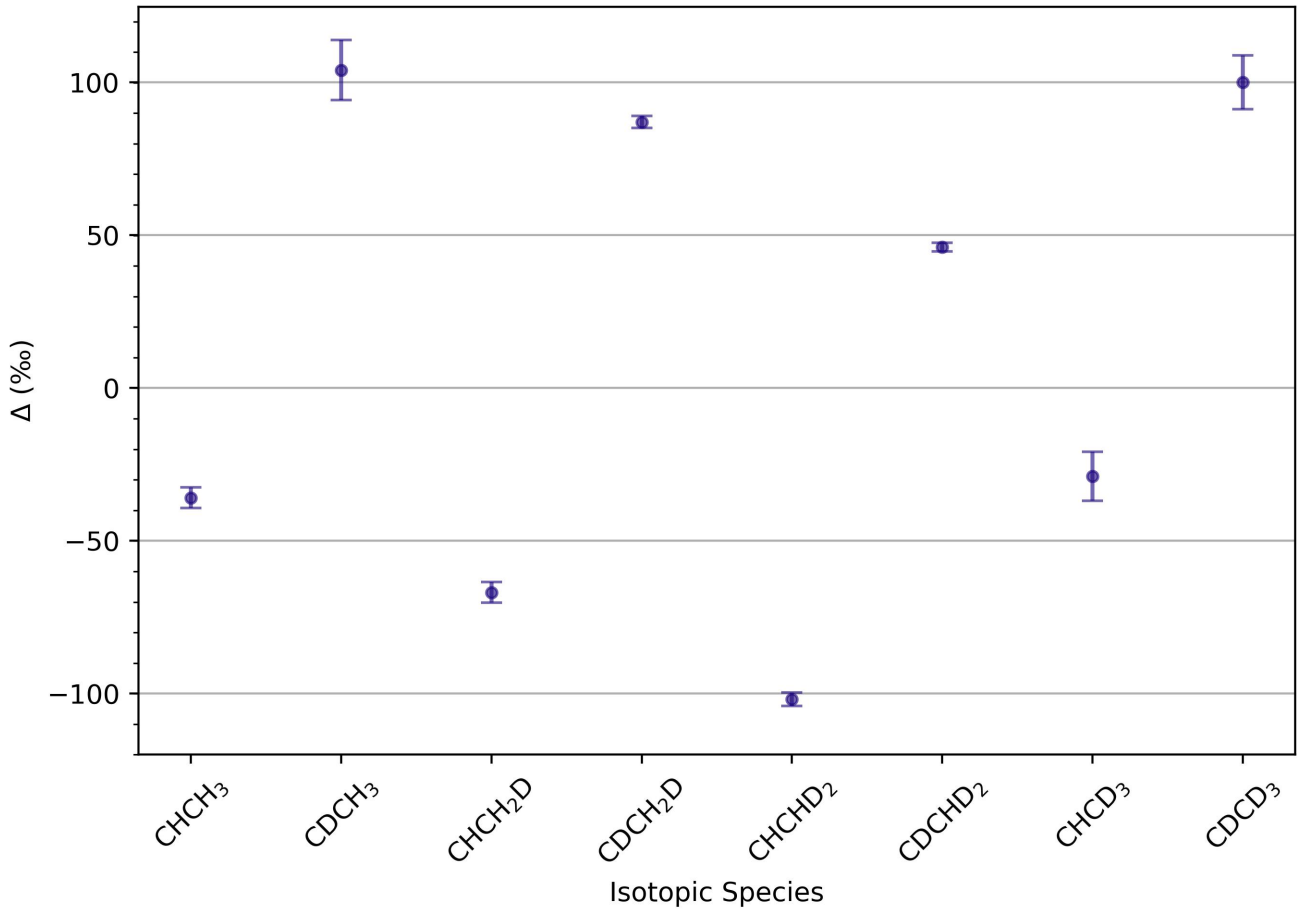

Figure 5.24: Abundance of eight isotopic species reported as deviations from a stochastic distribution, as Δ (‰). Here we define $\Delta_i = \frac{^iF}{^iF^*} - 1$ using $^iF = [i]/[Total]$, where $[Total]$ is the summed area of the signal arising from all eight isotopic species, and $^iF^*$ represents the expected percent of the total signal given a stochastic distribution. Error bars represent 2*sterr based on four replicate NMR measurements on a single sample.

The error bars for Δ in Figure 5.24 represent 2*standard error of four replicate measurements on a single equilibrated sample. Measurement reproducibility can also be evaluated based on the signal to noise ratio of each of the eight NMR signals corresponding to the eight isotopic variants of interest. We used the SNR tool built into MestReNova to calculate S/N, which ranges from $\sim$200 for the shortest peak (corresponding with $^{12}CH^{13}CH_3$) to $\sim$750 for the tallest ($^{12}CD^{13}CH_2D$), resulting in relative errors of 5.0‰ (1/200) and 1.3‰ (1/750), respectively. We used a Monte Carlo approach (code provided in Section 5.8.8.4) to model the errors in Δ that result from these relative errors in signal intensity. A limitation of using this approach is that S/N is a measure of the maximum intensity of a peak relative to the surrounding noise, and does not take peak width into account, thereby capturing only one aspect of the possible error. For example, two peaks may have the same maximum intensity, but different widths, and therefore different areas, corresponding to different abundances, but the relative error calculated for each will be the same. In the case of the data presented here, the equilibrium abundance of $^{12}CH^{13}CH_3$ is slightly greater than the abundance of $^{12}CD^{13}CD_3$, but the width of the $^{12}CD^{13}CD_3$ peak is a bit narrower than that of $^{12}CH^{13}CH_3$, such that the S/N reported is lower on the less abundant (but taller) $^{12}CD^{13}CD_3$ than on the more abundant (but shorter) $^{12}CH^{13}CH_3$. The calculation of relative error for these two species is different by $\sim$1–2‰, insignificant compared to the magnitude of the signals being measured.

5.8.6 Algebraic Transformation from "8-Species" Model to "3-Box" Model as Discussed in Section 5.5.4

Ten hydrogen-deuterium isotope exchange reactions:

$$\mathrm{CHCH_3} + \mathrm{D} \underset{k_{\alpha,s_b}}{\overset{k_{\alpha,s_f}}{\rightleftharpoons}} \mathrm{CDCH_3} + \mathrm{H} \tag{5.51}$$

$$\mathrm{CHCH_3} + \mathrm{D} \underset{\frac{1}{2}k_{\beta,s_b}}{\overset{\frac{3}{2}k_{\beta,s_f}}{\rightleftharpoons}} \mathrm{CHCH_2D} + \mathrm{H} \tag{5.52}$$

$$\mathrm{CDCH_3} + \mathrm{D} \underset{\frac{1}{2}k_{\beta,s_b}}{\overset{\frac{3}{2}k_{\beta,s_f}}{\rightleftharpoons}} \mathrm{CDCH_2D} + \mathrm{H} \tag{5.53}$$

$$\mathrm{CHCH_2D} + \mathrm{D} \underset{k_{\alpha,s_b}}{\overset{k_{\alpha,s_f}}{\rightleftharpoons}} \mathrm{CDCH_2D} + \mathrm{H} \tag{5.54}$$

$$\mathrm{CHCH_2D} + \mathrm{D} \underset{k_{\beta,s_b}}{\overset{k_{\beta,s_f}}{\rightleftharpoons}} \mathrm{CHCHD_2} + \mathrm{H} \tag{5.55}$$

$$\mathrm{CDCH_2D} + \mathrm{D} \underset{k_{\beta,s_b}}{\overset{k_{\beta,s_f}}{\rightleftharpoons}} \mathrm{CDCHD_2} + \mathrm{H} \tag{5.56}$$

$$\mathrm{CHCHD_2} + \mathrm{D} \underset{k_{\alpha,s_b}}{\overset{k_{\alpha,s_f}}{\rightleftharpoons}} \mathrm{CDCHD_2} + \mathrm{H} \tag{5.57}$$

$$\mathrm{CHCHD_2} + \mathrm{D} \underset{\frac{3}{2}k_{\beta,s_b}}{\overset{\frac{1}{2}k_{\beta,s_f}}{\rightleftharpoons}} \mathrm{CHCD_3} + \mathrm{H} \tag{5.58}$$

$$\mathrm{CDCHD_2} + \mathrm{D} \underset{\frac{3}{2}k_{\beta,s_b}}{\overset{\frac{1}{2}k_{\beta,s_f}}{\rightleftharpoons}} \mathrm{CDCD_3} + \mathrm{H} \tag{5.59}$$

$$\mathrm{CHCD_3} + \mathrm{D} \underset{k_{\alpha,s_b}}{\overset{k_{\alpha,s_f}}{\rightleftharpoons}} \mathrm{CDCD_3} + \mathrm{H} \tag{5.60}$$

Differential equations expressing change in species concentrations per time:

$$\frac{d[\mathrm{CHCH_3}]}{dt} = +k_{1_b}[\mathrm{CDCH_3}][\mathrm{H}] - k_{1_f}[\mathrm{CHCH_3}][\mathrm{D}] - \frac{3}{2}k_{2_f}[\mathrm{CHCH_3}][\mathrm{D}] + \frac{1}{2}k_{2_b}[\mathrm{CHCH_2D}][\mathrm{H}] \tag{5.61}$$

$$\frac{d[\mathrm{CDCH_3}]}{dt} = -k_{1_b}[\mathrm{CDCH_3}][\mathrm{H}] + k_{1_f}[\mathrm{CHCH_3}][\mathrm{D}] - \frac{3}{2}k_{2_f}[\mathrm{CDCH_3}][\mathrm{D}] + \frac{1}{2}k_{2_b}[\mathrm{CDCH_2D}][\mathrm{H}] \tag{5.62}$$

$$\begin{aligned}\frac{d[\mathrm{CHCH_2D}]}{dt} &= +k_{1_b}[\mathrm{CDCH_2D}][\mathrm{H}] - k_{1_f}[\mathrm{CHCH_2D}][\mathrm{D}] + \frac{3}{2}k_{2_f}[\mathrm{CHCH_3}][\mathrm{D}] \\ &\quad -\frac{1}{2}k_{2_b}[\mathrm{CHCH_2D}][\mathrm{H}] - k_{2_f}[\mathrm{CHCH_2D}][\mathrm{D}] + k_{2_b}[\mathrm{CHCHD_2}][\mathrm{H}]\end{aligned} \tag{5.63}$$

$$\begin{aligned}\frac{d[\mathrm{CDCH_2D}]}{dt} &= -k_{1_b}[\mathrm{CDCH_2D}][\mathrm{H}] + k_{1_f}[\mathrm{CHCH_2D}][\mathrm{D}] + \frac{3}{2}k_{2_f}[\mathrm{CDCH_3}][\mathrm{D}] \\ &\quad -\frac{1}{2}k_{2_b}[\mathrm{CDCH_2D}][\mathrm{H}] - k_{2_f}[\mathrm{CDCH_2D}][\mathrm{D}] + k_{2_b}[\mathrm{CDCHD_2}][\mathrm{H}]\end{aligned} \tag{5.64}$$

$$\begin{aligned}\frac{d[\mathrm{CHCHD_2}]}{dt} &= +k_{1_b}[\mathrm{CDCHD_2}][\mathrm{H}] - k_{1_f}[\mathrm{CHCHD_2}][\mathrm{D}] + \frac{3}{2}k_{2_b}[\mathrm{CHCD_3}][\mathrm{H}] \\ &\quad -\frac{1}{2}k_{2_f}[\mathrm{CHCHD_2}][\mathrm{D}] + k_{2_f}[\mathrm{CHCH_2D}][\mathrm{D}] - k_{2_b}[\mathrm{CHCHD_2}][\mathrm{H}]\end{aligned} \tag{5.65}$$

$$\begin{aligned}\frac{d[\mathrm{CDCHD_2}]}{dt} &= -k_{1_b}[\mathrm{CDCHD_2}][\mathrm{H}] + k_{1_f}[\mathrm{CHCHD_2}][\mathrm{D}] + \frac{3}{2}k_{2_b}[\mathrm{CDCD_3}][\mathrm{H}] \\ &\quad -\frac{1}{2}k_{2_f}[\mathrm{CDCHD_2}][\mathrm{D}] + k_{2_f}[\mathrm{CDCH_2D}][\mathrm{D}] - k_{2_b}[\mathrm{CDCHD_2}][\mathrm{H}]\end{aligned} \tag{5.66}$$

$$\frac{d[\mathrm{CHCD_3}]}{dt} = +k_{1_b}[\mathrm{CDCD_3}][\mathrm{H}] - k_{1_f}[\mathrm{CHCD_3}][\mathrm{D}] + \frac{1}{2}k_{2_f}[\mathrm{CHCHD_2}][\mathrm{D}] - \frac{3}{2}k_{2_b}[\mathrm{CHCD_3}][\mathrm{H}] \tag{5.67}$$

$$\frac{d[\mathrm{CDCD_3}]}{dt} = -k_{1_b}[\mathrm{CDCD_3}][\mathrm{H}] + k_{1_f}[\mathrm{CHCD_3}][\mathrm{D}] + \frac{1}{2}k_{2_f}[\mathrm{CDCHD_2}][\mathrm{D}] - \frac{3}{2}k_{2_b}[\mathrm{CDCD_3}][\mathrm{H}] \tag{5.68}$$

Defining β_H and β_D:

$$\beta_H = 3\Big([\mathrm{CHCH_3}] + [\mathrm{CDCH_3}]\Big) + 2\Big([\mathrm{CHCH_2D}] + [\mathrm{CDCH_2D}]\Big) + \Big([\mathrm{CHCHD_2}] + [\mathrm{CDCHD_2}]\Big) \tag{5.69}$$

$$\beta_D = 3\Big([\mathrm{CHCD_3}] + [\mathrm{CDCD_3}]\Big) + 2\Big([\mathrm{CHCHD_2}] + [\mathrm{CDCHD_2}]\Big) + \Big([\mathrm{CHCH_2D}] + [\mathrm{CDCH_2D}]\Big) \tag{5.70}$$

$$\frac{d\beta_H}{dt} = 3\Big(\frac{d[\mathrm{CHCH_3}]}{dt} + \frac{d[\mathrm{CDCH_3}]}{dt}\Big) + 2\Big(\frac{d[\mathrm{CHCH_2D}]}{dt} + \frac{d[\mathrm{CDCH_2D}]}{dt}\Big) + \Big(\frac{d[\mathrm{CHCHD_2}]}{dt} + \frac{d[\mathrm{CDCHD_2}]}{dt}\Big) \tag{5.71}$$

$$3\Big(\frac{d[\mathrm{CHCH_3}]}{dt} + \frac{d[\mathrm{CDCH_3}]}{dt}\Big) = -\frac{9}{2}k_{2_f}[\mathrm{CHCH_3}][\mathrm{D}] - \frac{9}{2}k_{2_f}[\mathrm{CDCH_3}][\mathrm{D}] + \frac{3}{2}k_{2_b}[\mathrm{CHCH_2D}][\mathrm{H}] + \frac{3}{2}k_{2_b}[\mathrm{CDCH_2D}][\mathrm{H}] \tag{5.72}$$

$$\begin{aligned}2\Big(\frac{d[\mathrm{CHCH_2D}]}{dt} + \frac{d[\mathrm{CDCH_2D}]}{dt}\Big) &= +3k_{2_f}[\mathrm{CHCH_3}][\mathrm{D}] - k_{2_b}[\mathrm{CHCH_2D}][\mathrm{H}] \\ &\quad +3k_{2_f}[\mathrm{CDCH_3}][\mathrm{D}] - k_{2_b}[\mathrm{CDCH_2D}][\mathrm{H}] - 2k_{2_f}[\mathrm{CHCH_2D}][\mathrm{D}] \\ &\quad +2k_{2_b}[\mathrm{CHCHD_2}][\mathrm{H}] - 2k_{2_f}[\mathrm{CDCH_2D}][\mathrm{D}] + 2k_{2_b}[\mathrm{CDCHD_2}][\mathrm{H}]\end{aligned} \tag{5.73}$$

$$\begin{aligned}\Big(\frac{d[\mathrm{CHCHD_2}]}{dt} + \frac{d[\mathrm{CDCHD_2}]}{dt}\Big) &= +\frac{3}{2}k_{2_b}[\mathrm{CHCD_3}][\mathrm{H}] - \frac{1}{2}k_{2_f}[\mathrm{CHCHD_2}][\mathrm{D}] \\ &\quad +k_{2_f}[\mathrm{CHCH_2D}][\mathrm{D}] - k_{2_b}[\mathrm{CHCHD_2}][\mathrm{H}] + \frac{3}{2}k_{2_b}[\mathrm{CDCD_3}][\mathrm{H}] \\ &\quad -\frac{1}{2}k_{2_f}[\mathrm{CDCHD_2}][\mathrm{D}] + k_{2_f}[\mathrm{CDCH_2D}][\mathrm{D}] - k_{2_b}[\mathrm{CDCHD_2}][\mathrm{H}]\end{aligned} \tag{5.74}$$

Collecting terms with $k_{2_f}[D]$:

$$k_{2_f}[D] \times \Big(-\frac{9}{2}[\mathrm{CHCH_3}] + 3[\mathrm{CHCH_3}] - \frac{9}{2}[\mathrm{CDCH_3}] + 3[\mathrm{CDCH_3}] \\ -2[\mathrm{CHCH_2D}] + [\mathrm{CHCH_2D}] - 2[\mathrm{CDCH_2D}] + [\mathrm{CDCH_2D}] - \frac{1}{2}[\mathrm{CHCHD_2}] - \frac{1}{2}[\mathrm{CDCHD_2}]\Big) \tag{5.75}$$

Simplifying:

$$k_{2_f}[D] \times \Big(-\frac{3}{2}[\mathrm{CHCH_3}] - \frac{3}{2}[\mathrm{CDCH_3}] - [\mathrm{CHCH_2D}] - [\mathrm{CDCH_2D}] - \frac{1}{2}[\mathrm{CHCHD_2}] - \frac{1}{2}[\mathrm{CDCHD_2}]\Big) \tag{5.76}$$

Collecting terms with $k_{2_b}[H]$:

$$k_{2_b}[\mathrm{H}] \times \Big(+\frac{3}{2}[\mathrm{CHCH_2D}] + \frac{3}{2}[\mathrm{CDCH_2D}] + 2[\mathrm{CHCHD_2}] - [\mathrm{CHCH_2D}] \\ +2[\mathrm{CDCHD_2}] - [\mathrm{CDCH_2D}] - [\mathrm{CHCHD_2}] + \frac{3}{2}[\mathrm{CHCD_3}] - [\mathrm{CDCHD_2}] + \frac{3}{2}[\mathrm{CDCD_3}]\Big) \tag{5.77}$$

Simplifying:

$$k_{2_b}[\mathrm{H}] \times \Big(+\frac{1}{2}[\mathrm{CHCH_2D}] + \frac{1}{2}[\mathrm{CDCH_2D}] + [\mathrm{CHCHD_2}] + [\mathrm{CDCHD_2}] + \frac{3}{2}[\mathrm{CHCD_3}] + \frac{3}{2}[\mathrm{CDCD_3}]\Big) \tag{5.78}$$

Defining $\mathrm{d}\beta_H/\mathrm{dt}$

$$\frac{d\beta_H}{dt} = k_{2_b}[\mathrm{H}] \times \Big(\frac{1}{2}[\mathrm{CHCH_2D}] + \frac{1}{2}[\mathrm{CDCH_2D}] + [\mathrm{CHCHD_2}] + [\mathrm{CDCHD_2}] + \frac{3}{2}[\mathrm{CHCD_3}] + \frac{3}{2}[\mathrm{CDCD_3}]\Big) \\ -k_{2_f}[D] \times \Big(\frac{3}{2}[\mathrm{CHCH_3}] + \frac{3}{2}[\mathrm{CDCH_3}] + [\mathrm{CHCH_2D}] + [\mathrm{CDCH_2D}] + \frac{1}{2}[\mathrm{CHCHD_2}] + \frac{1}{2}[\mathrm{CDCHD_2}]\Big) \tag{5.79}$$

$$\frac{d\beta_H}{dt} = \frac{k_{2_b}}{2}[\mathrm{H}] \times \Big([\mathrm{CHCH_2D}] + [\mathrm{CDCH_2D}] + 2[\mathrm{CHCHD_2}] + 2[\mathrm{CDCHD_2}] + 3[\mathrm{CHCD_3}] + 3[\mathrm{CDCD_3}]\Big) \\ -\frac{k_{2_f}}{2}[D] \times \Big(3[\mathrm{CHCH_3}] + 3[\mathrm{CDCH_3}] + 2[\mathrm{CHCH_2D}] + 2[\mathrm{CDCH_2D}] + [\mathrm{CHCHD_2}] + [\mathrm{CDCHD_2}]\Big) \tag{5.80}$$

$$\frac{d\beta_H}{dt} = \frac{k_{2_b}}{2}[\mathrm{H}] \times \beta_D - \frac{k_{2_f}}{2}[D] \times \beta_H \tag{5.81}$$

5.8.7 Recommendations for Future Study of This Isotopic Exchange System

We see ways in which the experimental component of our work could be carried further, at relatively low cost, beyond what we present here, potentially providing the information necessary to resolve some of the questions about this system that remain open. Thus we offer a set of recommendations for future experiments that build on the work presented here:

- In general, further HDX experiments with alanine should follow the methods of Experiment "Q" where possible, with ^{13}C NMR using simultaneous decoupling of ^{1}H and ^{2}H, and alanine labeled with ^{13}C at its β carbon position, thus providing distinction between all eight isotopic variants. This method stands in contrast to using ^{1}H (or ^{2}H) NMR, which does not provide the means to distinguish all eight variants, but instead aggregates the eight isotopic variants into two groups, distinguished only between α- and β-carbon-bound ^{1}H (or ^{2}H), and thus failing completely to observe the presence of $CDCD_3$ (or $CHCH_3$, if using ^{2}H NMR).

- An experimental setup in which the first several minutes of a reaction could be observed would be an improvement on our experiments, which did not capture the first several minutes of reaction time, as described in Section 5.3. The Sample Reaction System by New Era ($317) provides this option.

- A set of ^{13}C NMR kinetics experiments should be performed with initial conditions starting with pure single isotopic variants as can be acquired or synthesized, as presented in Supplementary Figure 5.20 and Supplementary Table 5.14 using experimental conditions delineated above. We expect this set of experiments will greatly reduce our uncertainties on several of the rate constants in our model. The current hourly rate for the 600 MHz instrument in our facility is $24.40. Eight experiments at 12 hrs each = 8*12*24.40 = $2342.40

- At the time of the experiments presented in this work, our Varian 600 MHz NMR was equipped with a broadband inverse (BBI) probe, which is optimized for ^{1}H observation and multinuclear, multidimensional experiments (e.g., HSQC, HMQC, HMBC). Replication of our experiments,

Table 5.14: Recommended future ^{13}C NMR kinetics experiments using isotopic variants available for purchase from Sigma-Aldrich (SA) or Cambridge Isotope Laboratories (CIL), or able to be produced in our facilities.

Initial Species	Water Composition	Rate Constant(s) Sampled	Product No.	Cost ($/gram)
$CH^{13}CH_3$	D_2O	k_{1_f}, k_{2_f}	489948 (SA)	$716
$CD^{13}CH_3$	H_2O	k_{1_b}	740055 (SA)	$1086
$CD^{13}CH_3$	D_2O	k_{3_f}	740055 (SA)	”
$CH^{13}CD_3$	H_2O	k_{8_b}	CDLM-3439-PK (CIL)	$2400
$CH^{13}CD_3$	D_2O	k_{10_f}	CDLM-3439-PK (CIL)	”
$CH^{13}CHD_2$	H_2O	k_{6_b}, k_{7_b}	710512 (SA)	$5100
$CH^{13}CHD_2$	D_2O	k_{9_f}	710512 (SA)	”
$CD^{13}CD_3$	H_2O	k_{9_b}, k_{10_b}	Not Available	–

in addition to execution of those recommended here, on a multinuclear broadband observe (BBO) probe should result in a ∼10x increase in carbon signal for otherwise identical experiments, thus resulting in an even higher precision data set.

- This same set of experiments may also be conducted on the full transamination reaction, i.e., initial conditions would include the presence of α-ketoglutarate. The experiments presented in this chapter involve only the ALA $\rightleftharpoons$ PYR half-reaction, in contrast to experiments presented in the previous chapter.

- Experiments performed in this study as well as those proposed above could be conducted with different ALT isoenzymes (e.g., mammalian cytosolic and mitochondrial ALT, plant ALT, bacterial ALT, and archaeal ALT). Combined with knowledge of protein sequence and structure, this may reveal any heterogeneity across ALT isoenzymes in terms of their rate-limiting transition states.

- Building molecular models has the potential to deepen our understanding beyond kinetic models. This would include models for finding transition states in the exchange reactions as well as molecular models of H/D transport in the enzyme environment.

5.8.8 Supplemental Scripts

5.8.8.1 Python Codes for Parameter Fitting

```
import numpy as np
from numpy import array
import pandas as pd
import matplotlib.pyplot as plt

pd.set_option('display.max_rows', None)

df1=pd.read_csv('kvals',names=['kf','kb'],sep="\s+")
df2=pd.read_csv('input',names=['input'])

kf=df1['kf']
kb=df1['kb']

input=df2['input']

trajectory=[]
trajectoryint=[]
solution=[]

t=0
nstep=input[0]
nstep=int(nstep)
dt=input[1]
tscal=input[2]
h2ofrac=input[3]
d2ofrac=input[4]
h0=input[5]
d0=input[6]
h1=input[7]
d1=input[8]
h2=input[9]
d2=input[10]
h3=input[11]
d3=input[12]


h2o=2.00*55.5*h2ofrac
d2o=2.00*55.5*d2ofrac

kab1=1000.0*np.log(d0/h1)
kab2=1000.0*np.log(d1/h2)
```

```
kab3=1000.0*np.log(d2/h3)
kab4=1000.0*np.log((d3*h0)/(h3*d0))

h2ofrac=(h2o/(h2o+d2o))
d2ofrac=(d2o/(h2o+d2o))

summh=4.*h0+3.*h1+2.*h2+1.*h3+3.*d0+2.*d1+1.*d2+0.*d0
summd=0.*h0+1.*h1+2.*h2+3.*h3+1.*d0+2.*d1+3.*d2+4.*d3

sumalphah=h0+h1+h2+h3
sumbetah=1.0*h2+2.0*h1+3.0*h0+1.0*d2+2.0*d1+3.0*d0

trajectory.append(array([t/tscal,h0,d0,h1,d1,h2,d2,h3,d3,\\
kab1,kab2,kab3,kab4,h2ofrac,d2ofrac]))
trajectoryint.append(array([t/tscal,sumalphah,sumbetah]))
solution.append(array([t/tscal,h2o,d2o,h2o+d2o,h2ofrac,d2ofrac
    ,\\
summh,summd,summh+summd]))

for i in range(1,nstep):

    t=t+dt

    flux1f=kf[0]*h0*d2o*dt
    flux1b=kb[0]*d0*h2o*dt
    flux1=flux1f-flux1b

    flux2f=kf[1]*h0*d2o*dt
    flux2b=kb[1]*h1*h2o*dt
    flux2=flux2f-flux2b

    flux3f=kf[2]*d0*d2o*dt
    flux3b=kb[2]*d1*h2o*dt
    flux3=flux3f-flux3b

    flux4f=kf[3]*h1*d2o*dt
    flux4b=kb[3]*d1*h2o*dt
    flux4=flux4f-flux4b

    flux5f=kf[4]*h1*d2o*dt
    flux5b=kb[4]*h2*h2o*dt
    flux5=flux5f-flux5b

```

```
84      flux6f=kf[5]*d1*d2o*dt
85      flux6b=kb[5]*d2*h2o*dt
86      flux6=flux6f-flux6b
87
88      flux7f=kf[6]*h2*d2o*dt
89      flux7b=kb[6]*d2*h2o*dt
90      flux7=flux7f-flux7b
91
92      flux8f=kf[7]*h2*d2o*dt
93      flux8b=kb[7]*h3*h2o*dt
94      flux8=flux8f-flux8b
95
96      flux9f=kf[8]*d2*d2o*dt
97      flux9b=kb[8]*d3*h2o*dt
98      flux9=flux9f-flux9b
99
100     flux10f=kf[9]*h3*d2o*dt
101     flux10b=kb[9]*d3*h2o*dt
102     flux10=flux10f-flux10b
103
104     flux11f=kf[10]*h1*dt
105     flux11b=kb[10]*d0*dt
106     flux11=flux11f-flux11b
107
108     flux12f=kf[11]*h2*dt
109     flux12b=kb[11]*d1*dt
110     flux12=flux12f-flux12b
111
112     flux13f=kf[12]*h3*dt
113     flux13b=kb[12]*d2*dt
114     flux13=flux13f-flux13b
115
116     dh2o=flux1+flux2+flux3+flux4+flux5+flux6+flux7+flux8+flux9+
    flux10
117     dh0=-flux1-flux2
118     dh1=flux2-flux4-flux5-flux11
119     dh2=flux5-flux7-flux8-flux12
120     dh3=flux8-flux10-flux13
121     dd0=flux1-flux3+flux11
122     dd1=flux3+flux4-flux6+flux12
123     dd2=flux6+flux7-flux9+flux13
124     dd3=flux9+flux10
125
```

```
126     h2o=h2o+dh2o
127     d2o=d2o-dh2o
128 
129     h0=h0+dh0
130     h1=h1+dh1
131     h2=h2+dh2
132     h3=h3+dh3
133 
134     d0=d0+dd0
135     d1=d1+dd1
136     d2=d2+dd2
137     d3=d3+dd3
138 
139     kab1=1000.0*np.log(d0/h1)
140     kab2=1000.0*np.log(d1/h2)
141     kab3=1000.0*np.log(d2/h3)
142     kab4=1000.0*np.log((d3*h0)/(h3*d0))
143 
144     h2ofrac=(h2o/(h2o+d2o))
145     d2ofrac=(d2o/(h2o+d2o))
146 
147     summh=4.*h0+3.*h1+2.*h2+1.*h3+3.*d0+2.*d1+1.*d2+0.*d0
148     summd=0.*h0+1.*h1+2.*h2+3.*h3+1.*d0+2.*d1+3.*d2+4.*d3
149 
150     sumalphah=h0+h1+h2+h3
151     sumbetah=1.0*h2+2.0*h1+3.0*h0+1.0*d2+2.0*d1+3.0*d0
152 
153     trajectory.append(array([t/tscal,h0,d0,h1,d1,h2,d2,h3,d3,\\
154     kab1,kab2,kab3,kab4,h2ofrac,d2ofrac]))
155     trajectoryint.append(array([t/tscal,sumalphah,sumbetah]))
156     solution.append(array([t/tscal,h2o,d2o,h2o+d2o,h2ofrac,
    d2ofrac,\\
157     summh,summd,summh+summd]))
158 
159 df3=pd.DataFrame(trajectory,columns=['time','H0','D0','H1','D1'
    ,'H2',\\
160 'D2','H3','D3','KAB1','KAB2',\\
161 'KAB3','KAB4','H2OFRAC','D2OFRAC'])
162 df4=pd.DataFrame(trajectoryint,columns=['time','ALPHAH','BETAH'
    ])
163 df5=pd.DataFrame(solution,columns=['time','H2O','D2O','
    SUMH2OD2O',\\
164 'H2OFRAC','D2OFRAC','TOTH','TOTD','TOTHD'])
```

```
165
166 c1=df3['time']
167 c2=df3['H0']
168 c3=df3['D0']
169 c4=df3['H1']
170 c5=df3['D1']
171 c6=df3['H2']
172 c7=df3['D2']
173 c8=df3['H3']
174 c9=df3['D3']
175 c10=df3['KAB1']
176 c11=df3['KAB2']
177 c12=df3['KAB3']
178 c13=df3['KAB4']
179 c14=df3['H2OFRAC']
180 c15=df3['D2OFRAC']
181
182 n=len(c1)
183
184 mod = open("mod.dat", "w")
185 modint = open("modint.dat", "w")
186 hdbal = open("hdbal.dat", "w")
187
188 i=0
189 while i < n:
190    print('{0:>10.3f} \
191 {1:>10.3f} \
192 {2:>10.3f} \
193 {3:>10.3f} \
194 {4:>10.3f} \
195 {5:>10.3f} \
196 {6:>10.3f} \
197 {7:>10.3f} \
198 {8:>10.3f} \
199 {9:>10.3f} \
200 {10:>10.3f} \
201 {11:>10.3f} \
202 {12:>10.3f} \
203 {13:>10.3f} \
204 {14:>10.3f}\
205 '.format(c1[i],\
206 c2[i],\
207 c3[i],\
```

```
c4[i],\
c5[i],\
c6[i],\
c7[i],\
c8[i],\
c9[i],\
c10[i],\
c11[i],\
c12[i],\
c13[i],\
c14[i],\
c15[i]),file=mod)
   i=i+1

c1=df4['time']
c2=df4['ALPHAH']
c3=df4['BETAH']

i=0
while i < n:
   print('{0:>10.3f} {1:>10.3f} {2:>10.3f}'.format(c1[i],\\
   c2[i], c3[i]),file=modint)
   i=i+1

c1=df5['time']
c2=df5['H2O']
c3=df5['D2O']
c4=df5['SUMH2OD2O']
c5=df5['H2OFRAC']
c6=df5['D2OFRAC']
c7=df5['TOTH']
c8=df5['TOTD']
c9=df5['TOTHD']


i=0
while i < n:
   print('{0:>10.3f} \
{1:>10.3f} \
{2:>10.3f} \
{3:>10.3f} \
{4:>10.3f} \
{5:>10.3f} \
```

```
{6:>10.3f} \
{7:>10.3f} \
{8:>10.3f}\
'.format(c1[i],\
c2[i],\
c3[i],\
c4[i],\
c5[i],\
c6[i],\
c7[i],\
c8[i],\
c9[i]),file=hdbal)
   i=i+1

mod.close()
modint.close()
hdbal.close()

exit()
```

Listing 5.1: box8.py

```
import numpy as np
import pandas as pd

df1=pd.read_csv('free_parameters',names=['FREE'])
free=df1['FREE']
#for 1-parameter model, need to modify for other models
k1f=free[0]
k2f=free[0]
k3f=free[0]
k4f=free[0]
k5f=free[0]
k6f=free[0]
k7f=free[0]
k8b=free[0]
k9b=free[0]
k10f=free[0]
k11f=0.0
k12f=0.0
k13f=0.0

K1=1.144
K2=2.904
K3=2.955
K4=1.165
K5=0.962
K6=0.962
K7=1.165
K8=0.361
K9=0.350
K10=1.132
K11=0.394153522
K12=1.210934903
K13=3.229704225

k1b=k1f/K1
k2b=k2f/K2
k3b=k3f/K3
k4b=k4f/K4
k5b=k5f/K5
k6b=k6f/K6
k7b=k7f/K7
k8f=K8*k8b
k9f=K9*k9b
```

```
44 k10b=k10f/K10
45 k11b=k11f/K11
46 k12b=k12f/K12
47 k13b=k13f/K13
48
49 kvals = open("kvals", "w")
50 print('{0:>20.10e} {1:>20.10e}'.format(k1f,k1b),file=kvals)
51 print('{0:>20.10e} {1:>20.10e}'.format(k2f,k2b),file=kvals)
52 print('{0:>20.10e} {1:>20.10e}'.format(k3f,k3b),file=kvals)
53 print('{0:>20.10e} {1:>20.10e}'.format(k4f,k4b),file=kvals)
54 print('{0:>20.10e} {1:>20.10e}'.format(k5f,k5b),file=kvals)
55 print('{0:>20.10e} {1:>20.10e}'.format(k6f,k6b),file=kvals)
56 print('{0:>20.10e} {1:>20.10e}'.format(k7f,k7b),file=kvals)
57 print('{0:>20.10e} {1:>20.10e}'.format(k8f,k8b),file=kvals)
58 print('{0:>20.10e} {1:>20.10e}'.format(k9f,k9b),file=kvals)
59 print('{0:>20.10e} {1:>20.10e}'.format(k10f,k10b),file=kvals)
60 print('{0:>20.10e} {1:>20.10e}'.format(k11f,k11b),file=kvals)
61 print('{0:>20.10e} {1:>20.10e}'.format(k12f,k12b),file=kvals)
62 print('{0:>20.10e} {1:>20.10e}'.format(k13f,k13b),file=kvals)
63 kvals.close()
```

Listing 5.2: `genk.py`

```
import numpy as np
from numpy import array
import pandas as pd
import matplotlib.pyplot as plt


pd.set_option('display.max_rows', None)

dfexperiment=pd.read_csv('EXPQ.dat',names=['TIME','CHCH3','
    CDCH3'\\
,'CHCH2D','CDCH2D','CHCHD2',\\
'CDCHD2','CHCD3', 'CDCD3'],sep="\s+")
dfmodel=pd.read_csv('modQ.dat',names=['TIME','CHCH3','CDCH3'\\
,'CHCH2D','CDCH2D','CHCHD2','CDCHD2',\\
'CHCD3','CDCD3',\
'KAB1','KAB2','KAB3','KAB4','H2OFRAC','D2OFRAC'],sep="\s+")

texp=dfexperiment['TIME']
tmod=dfmodel['TIME']

chch3exp=dfexperiment['CHCH3']
chch3mod=dfmodel['CHCH3']
cdch3exp=dfexperiment['CDCH3']
cdch3mod=dfmodel['CDCH3']

chch2dexp=dfexperiment['CHCH2D']
chch2dmod=dfmodel['CHCH2D']
cdch2dexp=dfexperiment['CDCH2D']
cdch2dmod=dfmodel['CDCH2D']

chchd2exp=dfexperiment['CHCHD2']
chchd2mod=dfmodel['CHCHD2']
cdchd2exp=dfexperiment['CDCHD2']
cdchd2mod=dfmodel['CDCHD2']

chcd3exp=dfexperiment['CHCD3']
chcd3mod=dfmodel['CHCD3']
cdcd3exp=dfexperiment['CDCD3']
cdcd3mod=dfmodel['CDCD3']

nexp=len(texp)
nmod=len(tmod)

```

```
43 dtm=tmod[1]-tmod[0]
44
45 tmodtotal = tmod[nmod-1]
46
47 i=0
48 misfit_chch3=0.0
49 misfit_cdch3=0.0
50 misfit_chch2d=0.0
51 misfit_cdch2d=0.0
52 misfit_chchd2=0.0
53 misfit_cdchd2=0.0
54 misfit_chcd3=0.0
55 misfit_cdcd3=0.0
56 misfitQ=0.
57
58 while i < nexp:
59       ti = texp[i]
60       ndt=int(ti/dtm)
61
62       chch3mod_interp=chch3mod[ndt]+((chch3mod[ndt+1]-chch3mod[
   ndt])\\
63       *((ti-tmod[ndt])/dtm))
64       misfit_chch3=misfit_chch3+np.absolute(chch3exp[i]-
   chch3mod_interp)
65
66       cdch3mod_interp=cdch3mod[ndt]+((cdch3mod[ndt+1]-cdch3mod[
   ndt])\\
67       *((ti-tmod[ndt])/dtm))
68       misfit_cdch3=misfit_cdch3+np.absolute(cdch3exp[i]-
   cdch3mod_interp)
69
70       chch2dmod_interp=chch2dmod[ndt]+((chch2dmod[ndt+1]-
   chch2dmod[ndt])\\
71       *((ti-tmod[ndt])/dtm))
72       misfit_chch2d=misfit_chch2d+np.absolute(chch2dexp[i]\\
73       -chch2dmod_interp)
74
75       cdch2dmod_interp=cdch2dmod[ndt]+((cdch2dmod[ndt+1]-
   cdch2dmod[ndt])\\
76       *((ti-tmod[ndt])/dtm))
77       misfit_cdch2d=misfit_cdch2d+np.absolute(cdch2dexp[i]\\
78       -cdch2dmod_interp)
79
```

```
80        chchd2mod_interp=chchd2mod[ndt]+((chchd2mod[ndt+1]-
    chchd2mod[ndt])\\
81        *((ti-tmod[ndt])/dtm))
82        misfit_chchd2=misfit_chchd2+np.absolute(chchd2exp[i]\\
83        -chchd2mod_interp)
84
85        cdchd2mod_interp=cdchd2mod[ndt]+((cdchd2mod[ndt+1]-
    cdchd2mod[ndt])\\
86        *((ti-tmod[ndt])/dtm))
87        misfit_cdchd2=misfit_cdchd2+np.absolute(cdchd2exp[i]\\
88        -cdchd2mod_interp)
89
90        chcd3mod_interp=chcd3mod[ndt]+((chcd3mod[ndt+1]-chcd3mod[
    ndt])\\
91        *((ti-tmod[ndt])/dtm))
92        misfit_chcd3=misfit_chcd3+np.absolute(chcd3exp[i]\\
93        -chcd3mod_interp)
94
95        cdcd3mod_interp=cdcd3mod[ndt]+((cdcd3mod[ndt+1]-cdcd3mod[
    ndt])\\
96        *((ti-tmod[ndt])/dtm))
97        misfit_cdcd3=misfit_cdcd3+np.absolute(cdcd3exp[i]\\
98        -cdcd3mod_interp)
99
100       i=i+1
101
102 misfitQ= misfit_chch3+misfit_cdch3+misfit_chch2d+misfit_cdch2d
    \\
103 +misfit_chchd2+misfit_cdchd2+misfit_chcd3+misfit_cdcd3
104
105 f = open("misfit", "r+")
106 misfitnow=float(f.read())
107 misfit=misfitQ+misfitnow
108 f.seek(0)
109 f.write(str(misfit))
110 f.close()
```

Listing 5.3: `getmisfitQ.py`

```
import numpy as np
from numpy import array
import pandas as pd
import matplotlib.pyplot as plt

pd.set_option('display.max_rows', None)

dfexperiment=pd.read_csv('EXPJ.dat',names=['TIME','ALPHAH','
    BETAH']\\
,sep="\s+")
dfmodel=pd.read_csv('modintJ.dat',names=['TIME','ALPHAH','BETAH
    ']\\
,sep="\s+")

texp=dfexperiment['TIME']
tmod=dfmodel['TIME']
alphahmod=dfmodel['ALPHAH']
alphahexp=dfexperiment['ALPHAH']

nexp=len(texp)
nmod=len(tmod)
dtm=tmod[1]-tmod[0]
tmodtotal = tmod[nmod-1]

i=0
misfitJa=0.0
while i < nexp:
      ti = texp[i]
      ndt=int(ti/dtm)
      alphahmod_interp=alphahmod[ndt]+((alphahmod[ndt+1]-
    alphahmod[ndt])\\
      *((ti-tmod[ndt])/dtm))
      misfitJa=misfitJa+np.absolute(alphahexp[i]-
    alphahmod_interp)
      i=i+1

f = open("misfit", "r+")
misfitnow=float(f.read())
misfit=misfitJa+misfitnow
f.seek(0)
f.write(str(misfit))
f.close()
```

Listing 5.4: `getmisfitJa.py`

5.8.8.2 *Python Code to Calculate Equilibrium Constants from Experiment "Q"*

```
import numpy as np
from numpy import array
import pandas as pd
import matplotlib.pyplot as plt

pd.set_option('display.max_rows', None)

df1=pd.read_csv('Ch5_ExpQ-EQ-line-fit-abundances.dat',names=['species','rep1','rep2','rep3','rep4'],sep="\s+")

D2O=0.483 # proportion of deuterium in solvent water
H2O=1-D2O # proportion of protium in solvent water

# calculate stochastic abundance expectations
CHCH3  = H2O**3*H2O
CDCH3  = H2O**3*D2O
CHCH2D = H2O**2*D2O*H2O*3
CDCH2D = H2O**2*D2O*D2O*3
CHCHD2 = H2O*D2O**2*H2O*3
CDCHD2 = H2O*D2O**2*D2O*3
CHCD3  = D2O**3*H2O
CDCD3  = D2O**3*D2O
#print(CHCH3,CDCH3,CHCH2D,CDCH2D,CHCHD2,CDCHD2,CHCD3,CDCD3)
sto = [CHCH3,CDCH3,CHCH2D,CDCH2D,CHCHD2,CDCHD2,CHCD3,CDCD3]
df1['sto']=sto
#print(sto)

# sum peak area signals for all eight species
Tot1=sum(df1['rep1'])
Tot2=sum(df1['rep2'])
Tot3=sum(df1['rep3'])
Tot4=sum(df1['rep4'])
#print(Tot1,Tot2,Tot3,Tot4)

# calculate relative abundance by dividing each species by the total peak area
df1['rep1rel']=df1['rep1']/Tot1
df1['rep2rel']=df1['rep2']/Tot2
df1['rep3rel']=df1['rep3']/Tot3
df1['rep4rel']=df1['rep4']/Tot4
#print(df1['rep1rel'])
```

```
40
41 rep1rel=df1['rep1rel']
42 rep2rel=df1['rep2rel']
43 rep3rel=df1['rep3rel']
44 rep4rel=df1['rep4rel']
45
46 relCHCH3  = (rep1rel[0],rep2rel[0],rep3rel[0],rep4rel[0])
47 relCDCH3  = (rep1rel[1],rep2rel[1],rep3rel[1],rep4rel[1])
48 relCHCH2D = (rep1rel[2],rep2rel[2],rep3rel[2],rep4rel[2])
49 relCDCH2D = (rep1rel[3],rep2rel[3],rep3rel[3],rep4rel[3])
50 relCHCHD2 = (rep1rel[4],rep2rel[4],rep3rel[4],rep4rel[4])
51 relCDCHD2 = (rep1rel[5],rep2rel[5],rep3rel[5],rep4rel[5])
52 relCHCD3  = (rep1rel[6],rep2rel[6],rep3rel[6],rep4rel[6])
53 relCDCD3  = (rep1rel[7],rep2rel[7],rep3rel[7],rep4rel[7])
54
55 # calculate the cap delta values relative to stochastic
      abundance
56 df1['delta1']=(df1['rep1rel']/sto-1)*1000
57 df1['delta2']=(df1['rep2rel']/sto-1)*1000
58 df1['delta3']=(df1['rep3rel']/sto-1)*1000
59 df1['delta4']=(df1['rep4rel']/sto-1)*1000
60 #print(df1['delta1'],df1['delta2'],df1['delta3'],df1['delta4'])
61
62 delta1=df1['delta1']
63 delta2=df1['delta2']
64 delta3=df1['delta3']
65 delta4=df1['delta4']
66
67 DelCHCH3  = (delta1[0],delta2[0],delta3[0],delta4[0])
68 DelCDCH3  = (delta1[1],delta2[1],delta3[1],delta4[1])
69 DelCHCH2D = (delta1[2],delta2[2],delta3[2],delta4[2])
70 DelCDCH2D = (delta1[3],delta2[3],delta3[3],delta4[3])
71 DelCHCHD2 = (delta1[4],delta2[4],delta3[4],delta4[4])
72 DelCDCHD2 = (delta1[5],delta2[5],delta3[5],delta4[5])
73 DelCHCD3  = (delta1[6],delta2[6],delta3[6],delta4[6])
74 DelCDCD3  = (delta1[7],delta2[7],delta3[7],delta4[7])
75
76 # calculate avg measured cap Delta for each of the eight
      species, along with the standard error
77 avg1=np.mean(DelCHCH3,dtype=np.float64)
78 avg2=np.mean(DelCDCH3,dtype=np.float64)
79 avg3=np.mean(DelCHCH2D,dtype=np.float64)
80 avg4=np.mean(DelCDCH2D,dtype=np.float64)
```

```
avg5=np.mean(DelCHCHD2,dtype=np.float64)
avg6=np.mean(DelCDCHD2,dtype=np.float64)
avg7=np.mean(DelCHCD3,dtype=np.float64)
avg8=np.mean(DelCDCD3,dtype=np.float64)

std1=np.std(DelCHCH3,dtype=np.float64,ddof=1)
std2=np.std(DelCDCH3,dtype=np.float64,ddof=1)
std3=np.std(DelCHCH2D,dtype=np.float64,ddof=1)
std4=np.std(DelCDCH2D,dtype=np.float64,ddof=1)
std5=np.std(DelCHCHD2,dtype=np.float64,ddof=1)
std6=np.std(DelCDCHD2,dtype=np.float64,ddof=1)
std7=np.std(DelCHCD3,dtype=np.float64,ddof=1)
std8=np.std(DelCDCD3,dtype=np.float64,ddof=1)

sterr1=std1/np.sqrt(4)
sterr2=std2/np.sqrt(4)
sterr3=std3/np.sqrt(4)
sterr4=std4/np.sqrt(4)
sterr5=std5/np.sqrt(4)
sterr6=std6/np.sqrt(4)
sterr7=std7/np.sqrt(4)
sterr8=std8/np.sqrt(4)

df1['Delta']=[avg1,avg2,avg3,avg4,avg5,avg6,avg7,avg8]
df1['STDev']=[std1,std2,std3,std4,std5,std6,std7,std8]
df1['STerr']=[sterr1,sterr2,sterr3,sterr4,sterr5,sterr6,sterr7,
    sterr8]
#print(df1['Delta'],df1['STDev'],df1['STerr'])

# calculate avg measured relative abundance for each of the
    eight species, along with the standard deviation and
    standard error
avgr1=np.mean(relCHCH3,dtype=np.float64)
avgr2=np.mean(relCDCH3,dtype=np.float64)
avgr3=np.mean(relCHCH2D,dtype=np.float64)
avgr4=np.mean(relCDCH2D,dtype=np.float64)
avgr5=np.mean(relCHCHD2,dtype=np.float64)
avgr6=np.mean(relCDCHD2,dtype=np.float64)
avgr7=np.mean(relCHCD3,dtype=np.float64)
avgr8=np.mean(relCDCD3,dtype=np.float64)

# calculate Ks
K1 = avgr2*H2O/(avgr1*D2O)
```

```
121 K2 = avgr3*H2O/(avgr1*D2O)
122 K3 = avgr4*H2O/(avgr2*D2O)
123 K4 = avgr4*H2O/(avgr3*D2O)
124 K5 = avgr5*H2O/(avgr3*D2O)
125 K6 = avgr6*H2O/(avgr4*D2O)
126 K7 = avgr6*H2O/(avgr5*D2O)
127 K8 = avgr7*H2O/(avgr5*D2O)
128 K9 = avgr8*H2O/(avgr6*D2O)
129 K10 = avgr8*H2O/(avgr7*D2O)
130
131 Kab1 = K1/K2
132 Kab2 = K4/K5
133 Kab3 = K7/K8
134 Kab4 = K10/K1
135
136 K1p=1000*np.log(K1)
137 K2p=1000*np.log(K2/3)
138 K3p=1000*np.log(K3/3)
139 K4p=1000*np.log(K4)
140 K5p=1000*np.log(K5)
141 K6p=1000*np.log(K6)
142 K7p=1000*np.log(K7)
143 K8p=1000*np.log(K8*3)
144 K9p=1000*np.log(K9*3)
145 K10p=1000*np.log(K10)
146 Kab1p=1000*np.log(Kab1*3)
147 Kab2p=1000*np.log(Kab2)
148 Kab3p=1000*np.log(Kab3/3)
149 Kab4p=1000*np.log(Kab4)
150
151 Ks={'Ks':[K1,K2,K3,K4,K5,K6,K7,K8,K9,K10,Kab1,Kab2,Kab3,Kab4],'
    permil':[K1p,K2p,K3p,K4p,K5p,K6p,K7p,K8p,K9p,K10p,Kab1p,
    Kab2p,Kab3p,Kab4p]}
152 df2=pd.DataFrame(Ks)
153 print(df2)
154
155 # plot it up
156 fig1 = plt.figure(figsize=(7,10))
157 ax1 = plt.subplot2grid((2, 1), (0, 0), colspan=1, rowspan=1)
158 ax2 = plt.subplot2grid((2, 1), (1, 0), colspan=1, rowspan=1)
159 ax1.plot(df1['species'],df1['rep1'],'o',color='#1a017c',alpha
    =0.4,markersize=5)
160 ax1.plot(df1['species'],df1['rep2'],'o',color='#1a017c',alpha
```

```
    =0.4,markersize=5)
ax1.plot(df1['species'],df1['rep3'],'o',color='#1a017c',alpha
    =0.4,markersize=5)
ax1.plot(df1['species'],df1['rep4'],'o',color='#1a017c',alpha
    =0.4,markersize=5)
ax1.set_ylim(0,700)
ax1.grid('on')
ax1.tick_params(axis='x', grid_color='none')
ax1.tick_params(axis='y')
ax1.xaxis.set_ticks(df1['species'])
ax1.xaxis.set_ticklabels(ax1.get_xticklabels(),rotation=45)
ax1.set_xlabel('Isotopic Species')
ax1.set_ylabel(r'Peak Area (arbitrary units)')

ax2.plot(df1['species'],df1['rep1rel'],'o',color='#1a017c',
    alpha=0.3,markersize=5,label='Replicate Measurements')
ax2.plot(df1['species'],df1['rep2rel'],'o',color='#1a017c',
    alpha=0.3,markersize=5)
ax2.plot(df1['species'],df1['rep3rel'],'o',color='#1a017c',
    alpha=0.3,markersize=5)
ax2.plot(df1['species'],df1['rep4rel'],'o',color='#1a017c',
    alpha=0.3,markersize=5)
ax2.plot(df1['species'],df1['sto'],'+',color='k',alpha=0.6,
    markersize=5,label='Stochastic Expectation')
ax2.set_ylim(0,0.25)
ax2.set_xlabel('Isotopic Species')
ax2.set_ylabel(r'Relative Abundance')
ax2.grid('on')
ax2.tick_params(axis='x', grid_color='none')
ax2.tick_params(axis='y')
ax2.xaxis.set_ticks(df1['species'])
ax2.xaxis.set_ticklabels(ax1.get_xticklabels(),rotation=45,
    horizontalalignment='right')
ax2.legend()

fig1.tight_layout()
fig1.savefig('Ch5_ExpQ-EQ-line-fit-abundances.png',dpi=400,
    transparent=False)
plt.show()

fig2 = plt.figure(figsize=(7,5))plt.errorbar(df1['species'],df1
    ['Delta'],yerr=df1['STerr']*2,fmt='o',color='#1a017c',alpha
    =0.6,markersize=4,capsize=4)
```

```
plt.grid('on')
plt.tick_params(axis='x', grid_color='none')
plt.tick_params(axis='y')
plt.xticks(rotation=45)
plt.yticks(np.arange(-120.1, 120.1, 100/10),minor=True)
plt.xlabel('Isotopic Species')
plt.ylabel(r'$\Delta$ ($^o/_{oo}$)')
fig2.tight_layout()
fig2.savefig('Ch5_Stochastic-Clumping.png',dpi=400,transparent=
    False)
plt.show()
```

Listing 5.5: Ch5_EXPQ-EQ.py

5.8.8.3 Python Codes for Data Adjustments

```
    import pandas as pd
import numpy as np
import array
from scipy.optimize import curve_fit
from matplotlib import pyplot as plt

df=pd.read_csv('EXPQ_NMR.dat',names=['TIME','CHCH3','CDCH3','
    CHCH2D','CDCH2D','CHCHD2','CDCHD2','CHCD3','CDCD3','TOTAL'
    ],sep="\s+")
df_select=df.head(35)
df_select2=df.head(17)

time=df['TIME'].to_numpy()
chch3=df['CHCH3'].to_numpy()
cdch3=df['CDCH3'].to_numpy()
chch2d=df['CHCH2D'].to_numpy()
cdch2d=df['CDCH2D'].to_numpy()
chchd2=df['CHCHD2'].to_numpy()
cdchd2=df['CDCHD2'].to_numpy()
chcd3=df['CHCD3'].to_numpy()
cdcd3=df['CDCD3'].to_numpy()
total=chch3+cdch3+chch2d+cdch2d+chchd2+cdchd2+chcd3+cdcd3
chch3=chch3/total
cdch3=cdch3/total
chch2d=chch2d/total
cdch2d=cdch2d/total
chchd2=chchd2/total
cdchd2=cdchd2/total
chcd3=chcd3/total
cdcd3=cdcd3/total
htime=time[0:35]
hchch3=chch3[0:35]
hcdch3=cdch3[0:35]
hchch2d=chch2d[0:35]

x=htime
y=hchch3

def test(x, a,b):
    return np.exp(-a * (x+b) )

param, param_cov = curve_fit(test, x, y)
```

```
41
42 a=param[0]
43 b=param[1]
44 time_shift=b
45
46 xx1 = np.linspace(-10, 100, num = 40)
47 f1= np.exp(-a * (xx1+b))
48
49 plt.plot(x, hchch3, 'o', color ='green', label =r"CHCH$_3$")
50 plt.plot(xx1, f1, '--', color ='blue', label ="Fit (exponential
    ")
51 #plt.axhline(y = 1.0, color = 'black', linestyle = '-')
52 plt.xlim(-10,30)
53 plt.legend()
54 plt.grid()
55 plt.xlabel('Time (min)')
56 plt.ylabel(r'CHCH$_3$ (normalized signal)')
57 plt.savefig('extrpltQ.jpg',dpi=400,transparent=False)
58 plt.savefig('extrpltQ.png',dpi=400,transparent=False)
59 plt.show()
60
61 time=time+time_shift
62 n=len(time)
63
64 f = open("EXPQ.dat", "w")
65 i=0
66 while i < n:
67    print('{0:>10.6f} {1:>10.6f} {2:>10.6f} {3:10.6f} {4:10.6f}
    {5:10.6f} {6:10.6f} {7:10.6f} {8:10.6f}'\
68 .format(time[i], chch3[i], cdch3[i], chch2d[i], cdch2d[i],
    chchd2[i], cdchd2[i], chcd3[i], cdcd3[i]),file=f)
69    i=i+1
70 f.close()
```

Listing 5.6: `ProcessQ.py`

```
import pandas as pd
import numpy as np
import array
from scipy.optimize import curve_fit
from matplotlib import pyplot as plt

dfnmr=pd.read_csv('EXPJ_NMR.dat',names=['time','alpha','beta'],
    sep="\s+")
time=dfnmr['time'].to_numpy()
alphanmr=dfnmr['alpha'].to_numpy()
betanmr=dfnmr['beta'].to_numpy()
normalize=betanmr[0]
alpha=alphanmr/normalize*3
beta=betanmr/normalize*3

#df_select=df.head(20)
#df_select2=df.head(17)

xa = time[0:20]
ya = alpha[0:20]

def test(x, a, b, c):
    return a * x**2 + b * x + c

param, param_cov = curve_fit(test, xa, ya)
a=param[0]
b=param[1]
c=param[2]

xx = np.linspace(-10, 20, num = 40)
yy = a * xx**2 + b * xx + c

root1=(-b-np.sqrt(b**2-4*a*c))/(2*a)
root2=(-b+np.sqrt(b**2-4*a*c))/(2*a)

tshift1=-root2

c=c-0.02

root1=(-b-np.sqrt(b**2-4*a*c))/(2*a)
root2=(-b+np.sqrt(b**2-4*a*c))/(2*a)

tshift2=-root2
```

```
43
44 xb=time[0:17]+tshift1
45 yb=beta[0:17]
46 param, param_cov = curve_fit(test, xb, yb)
47 a=param[0]
48 b=param[1]
49 c=param[2]
50
51 print(a,b,c)
52
53 xx2 = np.linspace(-10, 20, num = 40)
54 yy2 = a * xx2**2 + b * xx2 + c
55
56 fig1 = plt.figure(figsize=(8,10))
57 ax1 = plt.subplot2grid((2, 1), (0, 0), colspan=1, rowspan=1)
58 ax2 = plt.subplot2grid((2, 1), (1, 0), colspan=1, rowspan=1)
59
60 ax1.plot(xx, yy, '--', color ='k',alpha=0.5,label ="Quadratic
    Fit")
61 ax1.plot(xa, ya, 'o',color='#01497c',alpha=0.5,label ='Exp J
    Data')
62 ax1.set_xlim(-10,20)
63 ax1.set_ylim(0,0.25)
64 ax1.grid('on')
65 ax1.tick_params(axis='x', grid_color='none')
66 ax1.tick_params(axis='y')
67 ax1.legend(loc='lower right')
68 ax1.set_xlabel('Time (min)')
69 ax1.set_ylabel(r'$H_\alpha$ (normalized)')
70 ax2.plot(xx2, yy2, '--', color ='k',alpha=0.5,label ="Quadratic
     Fit")
71 ax2.plot(xb, yb, 'o',color='#01497c',alpha=0.5,label ='Exp J
    Data')
72 ax2.set_xlim(0,20)
73 ax2.set_ylim(2.4,3.4)
74 ax2.grid('on')
75 ax2.tick_params(axis='x', grid_color='none')
76 ax2.tick_params(axis='y')
77 ax2.legend(loc='upper right')
78 ax2.set_xlabel('Time (min)')
79 ax2.set_ylabel(r'$H_\beta$ (normalized)')
80 fig1.tight_layout()
81 plt.show()
```

```
fig1.savefig('Ch5_EXPJ-correction.png',dpi=400,transparent=
    False)

beta_0_0=c

x=time[0:17]+tshift2
y=beta[0:17]
param, param_cov = curve_fit(test, x, y)
a=param[0]
b=param[1]
c=param[2]

beta_02_0=c

t_0=time+tshift1
t_0=np.insert(t_0,0,0.0)
t_02=time+tshift2
t_02=np.insert(t_02,0,0.0)
alpha_0=alpha
alpha_0=np.insert(alpha_0,0,0.0)
alpha_02=alpha
alpha_02=np.insert(alpha_02,0,0.02)
beta_0=beta
beta_0=np.insert(beta_0,0,beta_0_0)
beta_02=beta
beta_02=np.insert(beta_02,0,beta_02_0)

beta_0=beta_0/beta_0_0*3.0
alpha_0=alpha_0/beta_0_0*3
beta_02=beta_02/beta_02_0*3
alpha_02=alpha_02/beta_02_0*3

n=len(t_0)

f = open("EXPJ.dat", "w")
i=0
while i < n:
    print('{0:>10.6f} {1:>10.6f} {2:>10.6f}'.format(t_0[i],
    alpha_0[i],beta_0[i]),file=f)
    i=i+1
f.close()

f = open("EXPJb.dat", "w")
```

```
123 i=0
124 while i < n:
125     print('{0:>10.6f} {1:>10.6f} {2:>10.6f}'.format(t_02[i],
        alpha_02[i],beta_02[i]),file=f)
126     i=i+1
127 f.close()
128
129 fig2 = plt.figure(figsize=(7,5))
130 plt.plot(t_0, alpha_0, ':', color ='#01497c',alpha=0.5,label ="
        extrap to 0",markersize=5)
131 plt.plot(t_0, beta_0, ':', color ='#7c2c01',alpha=0.5,label ="
        extrap to 0",markersize=5)
132 plt.plot(t_02, alpha_02, '-.', color ='#01497c',alpha=0.5,label
        ="extrap to 0.02",markersize=5)
133 plt.plot(t_02, beta_02, '-.', color ='#7c2c01',alpha=0.5,label
        ="extrap to 0.02",markersize=5)
134 plt.xlim(0,400)
135 plt.ylim(0,3)
136 plt.legend()
137 fig2.savefig('Ch5_EXPJ-correction-2.png',dpi=400,transparent=
        False)
138 plt.show()
```

Listing 5.7: `AdjustJ.py`

5.8.8.4 FORTRAN Code for Monte Carlo Error Simulation

```
      parameter(maxiter=1000000)
      real abund(8),err(8),sab(8),tmp(8),pab(8)
      real capdel(8,maxiter),acd(8),acddev(8)
      abund(1)=222.480
      abund(2)=236.019
      abund(3)=609.488
      abund(4)=661.000
      abund(5)=546.114
      abund(6)=593.846
      abund(7)=181.57475
      abund(8)=193.19091
      err(1)= 0.0054
      err(2)=0.0048
      err(3)=0.0017
      err(4)=0.0014
      err(5)=0.0018
      err(6)=0.0016
      err(7)=0.0038
      err(8)=0.0035
      sab(1)=7.1443
      sab(2)=6.6745
      sab(3)=20.0235
      sab(4)=18.7067
      sab(5)=18.7067
      sab(6)=17.4764
      sab(7)=5.8255
      sab(8)=5.4424

22    format(i6,f12.5)
      do i=1,maxiter
         do k=1,8
            tmp(k)=abund(k)
         enddo
         do k=1,8
         shift=4.0*err(k)*(rand()-0.5)
         tmp(k)=tmp(k)+shift*abund(k)
         enddo
         summ=0.0
         do k=1,8
            summ=summ+tmp(k)
         enddo
33    format(2f12.2)
```

```
34      format(2f10.0)
           do k=1,8
              pab(k)=(tmp(k)/summ)*100.0
           enddo
           do k=1,8
              capdel(k,i)=(pab(k)/sab(k)-1.0)*1000.
           enddo
        enddo
        do k=1,8
           acd(k)=0.0
        enddo
        do i=1,maxiter
           do k=1,8
           acd(k)=acd(k)+capdel(k,i)
           enddo
        enddo
        do k=1,8
           acd(k)=acd(k)/float(maxiter)
        enddo
        do k=1,8
           acddev(k)=0.0
        enddo
        do i=1,maxiter
           do k=1,8
           acddev(k)=acddev(k)+abs(capdel(k,i)-acd(k))
           enddo
        enddo
        do k=1,8
           acddev(k)=acddev(k)/float(maxiter)
        enddo
44      format(f12.1,f12.1)
        do i=1,8
           print 44,acd(i),acddev(i)
        enddo
        end
```

Listing 5.8: Ch5_EXPQ-EQ.py