© Copyright 2020

Michelle Redhair

Ligand and Protein Dynamics of CYP3A4

Michelle Redhair

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

William M. Atkins, Chair

Kelly K. Lee

Abhinav Nath

Rheem A. Totah

Program Authorized to Offer Degree:

Medicinal Chemistry

University of Washington

Abstract

Ligand and Protein Dynamics of CYP3A4

Michelle Redhair

Chair of the Supervisory Committee:

Professor William M. Atkins

Department of Medicinal Chemistry

Cytochrome P450 3A4 (CYP3A4) dominates drug metabolism in humans. Examples of promiscuity (binding and catalytic) and allosterism in CYP3A4 are ubiquitous but poorly understood. This work is dedicated to improving our understanding of the linkage between ligand binding and protein conformational dynamics with theoretical considerations of their impact on CYP3A4 catalysis using biophysical and theoretical methodologies. Specifically, the results herein demonstrate the complexity (requiring more than 2-states) in the ligand binding interactions for drugs and small molecules. Chapter 2 details the ligand-induced effects on conformational dynamics of CYP3A4 via hydrogen-deuterium exchange mass spectrometry for the allosteric effector, midazolam (MDZ). The MDZ-induced effects on the structural dynamics of CYP3A4 are not confined to the active site and they include regions remote from the heme active center. The results suggest the location of the allosteric MDZ binding site. Chapter 3 details theoretical aspects of multi-state ligand binding mechanisms wherein full kinetic derivations are presented with emphasis on methods for differentiating them. The results delineate the kinetic and thermodynamic

signatures for various multi-state binding mechanisms through a global analytical approach. In addition, a novel experimental approach is presented wherein ligand dissociation is used to further distinguish ligand binding mechanisms when coupled with the more traditional ligand binding approach. In Chapter 4, the mechanism of ligand binding for three small molecules (1,2,3-triazole, imidazole, and cyanide) binding to CYP3A4 is characterized via UV/Vis stopped-flow spectroscopy using the analytical approach described in Chapter 3. Assignment of a ligand binding mechanism to kinetic data provides an indirect measure of the relationship between conformational dynamics and ligand binding mechanism. The results show that, even for small molecules, the binding mechanism is complex, requiring interrogation of 4-state binding mechanisms to adequately describe the data. The recovered ligand binding mechanism (4-state induced fit) is consistent with the observed heterogeneity reported by other equilibrium based spectroscopic methods. The functional implications for the ligand binding interactions will be presented.

TABLE OF CONTENTS

List of Figures.....x

List of Supporting Figures xii

List of Tables.....xiii

Chapter 1: Introduction to ligand binding interactions with CYP3A4

1.1 Cytochrome P450 enzymes..... 1

1.1.1 Function and scope of research..... 1

1.1.2 Catalytic cycle, uncoupling pathways, and the potential for multiple products..... 2

1.1.3 Spectral paradigm may not always predict metabolic fate 5

1.1.4 Multiple conformers of CYP3A4 have been observed by X-ray crystallography..... 6

1.2 Conformational dynamics: hydrogen-deuterium exchange mass spectrometry (HDX-MS) 9

1.2.1 Nanodiscs provide a native-like membrane environment for CYP3A4 structural studies 9

1.2.2 HDX-MS methodology..... 10

1.2.3 The ligand-induced effects on the conformational dynamics of CYP3A4 are globally distributed 13

1.3 Ligand binding kinetics: stopped-flow spectroscopy 14

1.3.1 Traditional kinetic approach 14

1.3.2 CYP3A4 binding kinetics review 15

1.4 Figures..... 17

Chapter 2: Dynamics and Location of the Allosteric Midazolam Site in Cytochrome P4503A4 in Lipid Nanodiscs

2.1 Introduction..... 25

2.2 Experimental Procedures 28
2.2.1 Protein expression and purification (CYP3A4, MSP1D1, and rCPR)..... 28
2.2.2 Self-assembly of nanodiscs 32
2.2.3 Equilibrium binding titrations 32
2.2.4 Reconstitution system and kinetic assays 33
2.2.5 LC-MS/MS method for MDZ metabolism 34
2.2.6 Hydrogen-deuterium exchange 34
2.2.7 Mass spectrometry and data analysis of hydrogen-deuterium exchange samples 35
2.2.8 System setup and conventional molecular dynamics 36
2.2.9 Gaussian accelerated molecular dynamics 37
2.3 Results and Discussion 38
2.3.1 MDZ displays complex equilibrium binding behavior to CYP3A4 nanodiscs 38
2.3.2 MDZ metabolism by CYP3A4 39
2.3.3 HDX-MS profile of CYP3A4 nanodiscs 40
2.3.4 MDZ binding induces changes in HDX-MS 40
2.3.5 GaMD simulations with MDZ in the catalytic and allosteric sites 42
2.3.6 Comparison of MDZ and KTZ effects on HDX 48
2.4 Conclusions 48
2.5 Figures 50
2.6 Supporting figures 58

Chapter 3: Functional Implications of Complex Ligand Binding Kinetics and Protein Conformational Dynamics: Consideration of Promiscuous Drug Metabolizing Enzymes

3.1 Introduction 64
3.1.1 Underlying goal 66
3.1.2 Nomenclature 68
3.1.3 Kinetic model overview 71
3.2 Model derivation 72

3.2.1 The general 2-state system: E1 ⇄ E2 72
3.2.2 The 2-state isomerization model: E ⇄ E* 78
3.2.3 The 2-state ligand binding model: E + L ⇄ EL 82
3.2.4 The general 3-state system: E1 ⇄ E2 ⇄ E3 88
3.2.5 Translating the general 3-state system expressions to the ligand binding models: induced fit model as an example case 94
3.3 Kinetic models in detail 99
3.3.1 Linkage between conformational change and ligand binding: IF and CS 100
3.3.2 Multiple ligand binding interactions: Sequential binding, substrate reorientation and persistent heterogeneity 101
3.4 Experimental Procedures 102
3.4.1 Kinetic simulations 102
3.4.2 Signal considerations 102
3.4.3 Limiting cases: rapid first step, rapid second step, and the stationary intermediate 103
3.4.4 Identical k_{obs} case: apparently equivalent IF and CS case 105
3.4.5 Identical k_{obs} case: apparently equivalent IF and SR case 106
3.4.6 Identical k_{obs} case: apparently equivalent CS and SR case 108
3.4.7 Dilution experiment 109
3.5 Results 112
3.5.1 Induced fit: limiting cases 115
3.5.2 Conformational selection: limiting cases 120
3.5.3 Sequential binding: limiting cases 124
3.5.4 Substrate reorientation: limiting cases 127
3.5.5 Persistent Heterogeneity 129
3.5.6 IF vs CS: equivalence in k_{obs} 130
3.5.7 IF vs SR: equivalence in k_{obs} 132
3.5.8 CS vs SR: equivalence in k_{obs} 133
3.5.9 Dilution 134
3.6 Discussion 137

3.6.1 The ‘kinetically silent’ ensemble of unbound states 137
3.6.2 Rearrangement rates may be ligand dependent 138
3.6.3 Binding mechanism may be ligand dependent 139
3.6.4 Kinetic and thermodynamic properties of binding interactions may be sensitive to experimental conditions 140
3.6.5 Functional implications for the different binding models: binding promiscuity and catalytic promiscuity 140
3.6.6 Summary decision tree 141
3.7 Figures 143

Chapter 4: Kinetic Resolution of Heterogenous Ligation States of Small Molecules Binding to CYP3A4

4.1 Introduction 169
4.1.1 Heterogeneity in the ligation state of ‘Type II’ compounds 169
4.1.2 Experimental approach and ligand selection 170
4.1.3 Improvements in kinetic analyses: a push for a global approach 171
4.2 Experimental Procedures 173
4.2.1 Materials 173
4.2.2 Protein expression and purification 173
4.2.3 Equilibrium binding titrations 173
4.2.4 Corrections to CN^- concentration 174
4.2.5 Stopped-flow kinetic: ligand binding (mixing experiments) 174
4.2.6 Stopped-flow kinetic: ligand dissociation (dissociation experiments) 175
4.2.7 Global fitting and statistical analysis of stopped-flow kinetic data 175
4.2.8 Singular value decomposition of kinetic data 177
4.3 Results 178
4.3.1 Equilibrium titrations 178
4.3.2 Stopped-flow kinetic measurements 178
4.3.3 123-TRZ binding mechanism 179
4.3.4 IMZ binding mechanism 180
4.3.5 CN^- binding mechanism 181

4.3.6 SVD analysis of stopped-flow kinetic measurements 181
4.4 Discussion 182
4.5 Figures 191
Appendix 200
References 229

LIST OF FIGURES

Figure 1.1. The canonical P450 catalytic cycle involving hydrophobic type I ligands. 17

Figure 1.2. CYP heme-ligand interactions. 18

Figure 1.3. CYP3A4 structure. 19

Figure 1.4. Comparison of 3A4 crystal structures. 20

Figure 1.5. CYP3A4 nanodiscs. 21

Figure 1.6. HDX-MS workflow for CYP3A4 nanodiscs. 22

Figure 1.7. Stopped-flow experimental setup. 23

Figure 1.8. Ligand binding kinetic schemes tested in this work. 24

Figure 2.1. MDZ structure, equilibrium titration into CYP3A4 nanodiscs, and product ratio vs MDZ concentration plot. 50

Figure 2.2. MDZ-induced changes in HDX changes. 51

Figure 2.3. Average RMSD values for CYP3A4 residues under different ligation states 52

Figure 2.4. Representative MDZ binding modes and time dependences of active site MDZ-CYP3A4 interactions. 53

Figure 2.5. G'-helix distance with B-B'-loop and membrane interactions in the simulation. 54

Figure 2.6. Distances between MDZ C1' and the heme iron atom in the simulation. 56

Figure 2.7. Comparison of HDX changes from MDZ and KTZ binding. 57

Figure 3.1. Two-state models 143

Figure 3.2. Multi-state binding models. 144

Figure 3.3. Induced Fit: test cases. 152

Figure 3.4. Conformational selection: test cases. 154

Figure 3.5. Sequential Binding: test cases. 156

Figure 3.6. Substrate Reorientation: test cases. 158

Figure 3.7. ‘Equivalent’ IF and CS cases. 161

Figure 3.8. ‘Equivalent’ IF and SR cases. 162

Figure 3.9. ‘Equivalent’ CS and SR cases. 163

Figure 3.10. Comparison of binding and dilution experimental results. 164

Figure 3.11. Different ligand selection criteria may yield different binding mechanisms for the same enzyme. 166

Figure 3.12. Potential functional significance of the different binding mechanisms. 167

Figure 3.13. Kinetic decision tree. 168

Figure 4.1. Ligand binding kinetic schemes tested in this study. 191

Figure 4.2. Difference spectra and resultant binding isotherms for 123-TRZ, IMZ and CN^- binding to CYP3A4. 192

Figure 4.3. IMZ binding and dissociation kinetics. 193

Figure 4.4. CN^- binding and dissociation kinetics. 194

Figure 4.5. Summary of ligand binding mechanisms for IMZ and CN^- binding to CYP3A4. 197

Figure 4.6. SVD analysis of IMZ binding data. 198

Figure 4.7. SVD analysis of CN^- binding data. 199

LIST OF SUPPORTING FIGURES

Figure S1. Initial position of the second MDZ (dark blue) for two-MDZ simulations. 58

Figure S2. Spectral data fit to Michaelis-Menten and Eadie-Hofstee models for MDZ binding, and catalytic turnover data fit to Michalis-Menten model. 59

Figure S3. Deuterium uptake profiles for all peptides. 61

Figure S4. Overlay of 5TE8 crystal structure with one MDZ bound and snapshots from the MD with two MDZ's bound. 62

Figure S5. Distances between the Arg-105 Cζ and the MDZ1 N5 (blue) and center of the MDZ1 fluorophenyl ring (orange). 63

LIST OF TABLES

Table 3.1. Kinetic and thermodynamic signatures of the 2-state systems. 145

Table 3.2. Kinetic and thermodynamic signatures of the IF binding model..... 146

Table 3.3. Kinetic and thermodynamic signatures of the CS binding model. 147

Table 3.4. Kinetic and thermodynamic signatures of the SB binding model. 148

Table 3.5. Kinetic and thermodynamic signatures of the SR binding model. 149

Table 3.6. Kinetic and thermodynamic signatures of the PH binding model. 150

Table 3.7. Simplified observed rate expressions for the limiting cases. 151

Table 4.1 Model parameters and statistics for IMZ binding to CYP3A4. 195

Table 4.2 Model parameters and statistics for CN^- binding to CYP3A4. 196

ACKNOWLEDGEMENTS

I would like to thank the following people:

Bill Atkins for his support and guidance over the years – without his patience, trust, and insight, this work would not have been possible. Thanks for all the laughs and for letting me "not let [you] get in the way."

My committee members: Kelly Lee, Abhi Nath, and Rheem Totah. I have learned so much from each of you over the years. Thank you for helping me believe in myself, especially during times when I found it difficult. Thank you, Nina Isoherranen, for your feedback as my GSR.

Our collaborator, John Hackett. Thank you for your help with MD simulations and initial MATLAB scripts.

Caleb Woods for introducing me to kinetics, spectroscopy, and so much more. Thank you for the long discussions about science and about life. I would be a lesser scientist without you.

Mike Guttman for his help with HDX. Thank you for always making time to answer my questions and for all your feedback over the years.

Members of the Atkins lab: You all made working in lab fun and memorable. I would especially like to thank Mike Dabrowski, Kip Conner, Nick Treuheit, Mavis Li, and Amanda Clouser for helpful scientific discussions.

Eric Evangelista, thank you for being my friend and for helping me figure out how Webster would define stopped-flow spectroscopy: stopped(not moving)-flow(moving) spectroscopy or "not moving-moving spectroscopy".

Jenny Chang, thank you for your friendship and all the memories over the years. Thank you for making me laugh and reminding me to enjoy the moment.

My cohort: Rob Pelletier and James Williams. Your bright personalities and funny stories made me forget we were all working in windowless basement labs. Thank you to Rob, for teaching me so much about CYPs, everything I ever wanted to know about mass specs and wonky metabolism, and for your scientific insight over the years. Thank you to James, for all the happy memories and for helping develop the concepts for so many web series – too many to list.

All the friends I've made in the department. Thank you all.

The departments of Medicinal Chemistry and Pharmaceutics.

My family for their support. Thanks to my mom, dad, and brother. Also thank you to the Yehs.

Last but not least, my husband, Andrew Yeh. Thank you for always being supportive, kind, and understanding, especially when this work seemed to take so much of my time.

DEDICATION

To my parents, for always believing in me.

Chapter 1

Introduction to ligand binding interactions with CYP3A4

Note: portions of this chapter have been taken from the published manuscript:

Redhair, M., Clouser, A. F., and Atkins, W. M. (2019) Hydrogen-deuterium exchange mass spectrometry of membrane proteins in lipid nanodiscs. Chem. Phys. Lipids. 220, 14–22.

1.1 Cytochrome P450 enzymes

1.1.1 Function and scope of research

Cytochrome P450s (CYPs) form a superfamily of heme-containing mono-oxygenases that oxidize a range of endogenous and exogenous compounds [1]. CYPs are involved in a number of cellular processes including drug metabolism, hormone synthesis, glucocorticoid synthesis and metabolism of other natural products. Among the hepatic CYPs, the CYP3A4 isoform, which is the subject of this thesis, dominates drug metabolism [2]. CYP3A4 displays remarkable promiscuity, presumably as a result of conformational plasticity [3,4]. The 'promiscuity' includes the ability to bind and metabolize structurally dissimilar compounds and the ability to perform a range of oxidative processes on a variety of distinct functional groups [4]. In addition, atypical or non-Michaelis-Menten kinetics are commonly observed with CYP3A4 [5-8]. Importantly, the observed promiscuity and atypical kinetics cannot be explained by single-site or single binding mode models. Characterization of the ligand binding mechanism may provide mechanistic insight into how CYP3A4 is able to achieve its broad catalytic profile.

Enzymes sample distinct conformations over a range of time scales. It is widely appreciated that ligand binding directly influences the conformational landscape and, therefore, the catalytic activity of various enzymes in ligand-dependent and enzyme-dependent ways [9]. For example, the binding mechanism may uniquely contribute to the fractional contribution of the catalytically

relevant enzyme-substrate complex(es) and/or the fraction of enzyme primed for substrate recognition or product release. Hence, the ligand binding mechanism may have functional significance.

The purpose of this dissertation is to improve our understand of ligand binding and its potential impact on CYP3A4 structural dynamics and the potential linkage to promiscuity (binding and catalytic) and allosterism in CYP3A4. Specifically, in Chapter 2, I present the midazolam-induced effects on the conformational dynamics of CYP3A4. I hypothesized that the allosteric effector, midazolam, will differently affect the conformational dynamics compared to previous analyses with the potent azole antifungal inhibitor, ketoconazole. In Chapter 3, I present methods to distinguish different mechanisms of ligand binding for multi-state binding mechanisms which have been suggested for CYP3A4. I present an analytical approach to distinguish the binding mechanisms. In addition, I present a novel experimental approach wherein the dissociation kinetics improve the confidence in discrimination of competing binding models when globally analyzed with complementary binding kinetics. Finally, in Chapter 4, I present the ligand binding mechanism for small azoles (1,2,3-triazole and imidazole) and a diatomic ligand (cyanide) using methods described in Chapter 3, with the rationale that the binding mechanism is an indirect reporter of the dynamics of CYP3A4. I hypothesized that the ligands will binding to CYP3A4 through a binding mechanism that contains an induced fit component, which envisions an ensemble of different bound states (or ligation modes).

1.1.2 Catalytic cycle, uncoupling pathways, and the potential for multiple products

CYPs are mixed function oxidases that require molecular oxygen and two reducing equivalents from pyridine nucleotide electron donors, supplied by their respective redox partner proteins, to oxidize organic compounds in the reaction below:

$$\mathrm{RH} + \mathrm{NAD(P)H} + \mathrm{O_2} + 2\mathrm{H^+} \rightarrow \mathrm{ROH} + \mathrm{NAD(P)^+} + \mathrm{H_2O} \quad (1.1)$$

The hydroxylation reaction shown in equation 1.1 is common, however, CYPs are capable of catalyzing a range of other oxidation reactions [10]. The CYP catalytic cycle is a multi-step process which typically involves oxidation by compound I (Figure 1.1). The heme in the resting state contains a hexacoordinate low-spin state with water at the sixth axial ligand site. The following steps occur in the canonical catalytic cycle of CYP3A4: **1**) substrate binds resulting in a spin-state change, **2**) CYP reductase (CPR) reduces the high-spin (HS) ferric P450, **3**) oxygen binds the ferrous P450, **4**) CPR or cytochrome b5 (Cyt b5) reduce the oxygen bound P450 to form the ferric peroxy anion species, **5**) the distal oxygen is protonated to form the highly reactive ferric hydroperoxo complex also known as compound 0, **6**) the distal oxygen is protonated a second time resulting in a heterolytic O-O bond cleavage to release water and form the perferryl-oxy species, also known as compound I, and **7**) oxygen insertion into substrate by compound I. Notably, a complex proton relay network is responsible for proton delivery and the canonical cycle relies on precise delivery to the distal oxygen atoms.

The catalytic cycle is not as finely tuned in CYP3A4 (and other CYPs) and deviations from the canonical catalytic cycle may include formation of uncoupling products (superoxide, hydrogen peroxide, or a second water; these are referred to as uncoupling reaction products because their formation consumes reducing equivalents and the catalytic cycle resets to the resting state without oxidation of the bound ligand) or oxidation of the substrate at multiple sites [11,12]. In fact, multiple metabolites are commonly observed in CYP3A4 with different degrees of uncoupling in

some cases. The uncoupling reactions result in the release of the following 'products': superoxide, hydrogen peroxide, or a second water molecule (shown in grey in Figure 1.1). The uncoupling reactions (excluding superoxide formation) are shown below:

$$RH + NAD(P)H + O_2 + 2H^+ \rightarrow RH + NAD(P)^+ + H_2O_2 \quad (1.2)$$

$$RH + 2NAD(P)H + O_2 + 4H^+ \rightarrow RH + NAD(P)^+ + 2H_2O \quad (1.3)$$

In the autoxidation pathway, superoxide is released and the enzyme returns to its resting state without completion of the cycle. In the peroxide pathway, the proximal oxygen (not the distal oxygen) is protonated resulting in hydrogen peroxide release and a return to the resting state without completion of the cycle. In the oxidase pathway, compound I is further reduced and two protonation steps result in the release of an additional water and a return to the resting state without oxidation of the substrate. Regarding multiple oxidation products of the substrate, multiple reactive intermediates (ferric peroxy anion and compound 0) are formed in the catalytic cycle that can perform different oxidative reactions on the substrate (e.g. deformylation, epoxidation, etc), Alternatively, it is possible that substrate may adopt multiple conformations (i.e. multiple ligation states) throughout the cycle resulting in oxidation at multiple regions and functional groups on the substrate (not shown in the catalytic cycle figure). The multiple substrate orientations may result from conformational dynamics in the protein, allosteric effects, or from binding of multiple substrates. Collectively, the observation of multiple metabolites and differences in the degree of uncoupling for different ligands highlights our lack of understanding of the relationship between dynamics (in CYP3A4 structure and in ligand orientation) and catalysis. In the next two sections, I explore the relationships between ligand binding interactions and CYP3A4 structure with CYP3A4 function.

1.1.3 Spectral paradigm may not always predict metabolic fate

A common strategy for design of CYP3A4 inhibitors, or the rationale for using existing drugs as inhibitors, is based on the spectral paradigm that classifies three types of drug-heme interactions with CYPs (Figure 1.2). The three binding interactions have distinct spectral properties [13,14]. Within this paradigm a 'Type I' compound displaces water from the sixth axial ligand site resulting in a spin-state transition from a hexacoordinate low-spin (LS) state to a catalytically competent pentacoordinate HS state. 'Type II' compounds replace the axial water via direct iron-nitrogen bond, thereby stabilizing the LS state. The third spectral class, 'Reverse Type I' complexes, have been proposed to occur via direct ligation to the heme-iron with other heteroatoms or by hydrogen bonding between the axial water on the heme-iron and heteroatoms on the compound. The linkage between spin-state and catalysis is well-established in the P450cam isoform of CYPs [15,16]. Specifically, the LS state of P450cam cannot be reduced by its natural redox partner protein. Since oxygen may only bind to the reduced state, stabilization of the LS state effectively halts the catalytic cycle. Thus, the LS → HS state transition has been regarded as a 'catalytic gating mechanism', leading to the expectation that 'Type I' compounds will behave as substrates while 'Type II' and 'Reverse Type I' compounds will behave as inhibitors.

However, recent work challenges the spectral paradigm and its implications for CYP catalysis. Paradoxically, faster metabolic rates for type II compounds compared to analogous type I compounds have been observed [17-19]. One possible explanation is that the 'catalytic gating mechanism' is not as finely tuned in CYPs. In fact, evidence for direct reduction of the type II complexes has been presented [20]. In another study, it was shown that the redox potential increased for a type II compound relative to the ligand-free state, which is a behavior previously associated with only type I compounds [21]. Another possibility is that differences in heme

coordination, and therefore heme-ligation dynamics, contributes to the likelihood of achieving different categories of binding modes (e.g. catalytically productive vs inhibitory). Incomplete heme-coordination of the azole moiety in type II compounds compared to the sterically unhindered azole standards [22] has been observed by UV/Vis spectroscopy. Moreover, heterogeneity in heme-ligand coordination (type II and reverse type I coordination) has been observed by continuous wave electron paramagnetic resonance spectroscopy (EPR) for various compounds that were previously thought to bind only by type II coordination, as assessed by UV/Vis spectroscopy [23].

Collectively, the spectroscopic data suggest additional complexity in the heme-ligand coordination equilibria and highlight their potential effects on catalysis. Importantly, the spectroscopic techniques (EPR and UV/Vis spectroscopy with equilibrated solutions of ligand-bound CYP3A4) only highlight the thermodynamic properties of the heme-ligand interaction. Specifically, the different ligation states cannot be resolved by UV/Vis spectroscopy, and thus, the spectra are an average of the ligation states. The EPR spectra may resolve the different ligation species, however, the spectra only report on the relative equilibrium amounts of each state (at physiologically irrelevant enzyme concentrations and temperature). Hence, the results highlight the need to clarify our understanding of the thermodynamic and kinetic properties of LS complexes. In this work, I determined the mechanism of ligand binding for three type II ligands (1,2,3-triazole, imidazole, and cyanide) binding to CYP3A4 via stopped-flow spectroscopy.

1.1.4 Multiple conformers of CYP3A4 have been observed by X-ray crystallography

CYP3A4 is a ~57 kDa enzyme located in the membrane of the endoplasmic reticulum of the hepatocytes. The ligand-free crystal structure of CYP3A4 is shown in Figure 1.3 (PDB:1TQN)

[24]. The structure is rich in α-helical content with four β-sheet domains. The enzyme has an N-terminal transmembrane helix (~5 turns) that serves to anchor and superficially embed the enzyme to the membrane. The active site contains an iron protoporphyrin IX prosthetic group, which is ligated by Cys442. The heme is adjacent to the I-helix, which spans across the interior of the protein. The F/G region, which is perpendicular to the I-helix, acts as a 'lid' to the active site on the distal side of the heme. At the membrane interface, a phenylalanine cluster at the F/F' loop forms a hydrophobic core for large non-polar ligands to bind.

CYP3A4 is known for its large active site, which is capable of accommodating a range of chemically dissimilar compounds in a variety of binding poses [25]. Several crystal structures of CYP3A4 with different ligands have been published with ligands such as progesterone, metyrapone, erythromycin, ketoconazole, ritonavir, bromocriptine, progesterone, midazolam, metformin, and a variety of inhibitors and ritonavir analogs [24,26-41]. The ligands occupy different physical space (active site and peripheral binding sites) and in some cases multiple ligands are found within the active site. An overlay of the existing crystal structures reveals the largest degree of variability among the crystal structures at the F/G region and the intervening helices in addition to the H- and beginning of the I-helix (Figure 1.4A). In addition, molecular dynamics simulations show a high degree of flexibility in the resides spanning the loop between the F-F' helices (residues 211-218), which may contribute to the broad substrate specificity in CYP3A4 [42].

In two recent analyses, 24 different CYP3A4 X-ray crystal structures published before 2015 were analyzed and four channels and three major conformations were identified [43,44]. The publications summarize similar previous analyses, however, here I report the most recent findings by Benkaidali and coworkers. The four channels (named 2a, 2f, S, and 2e) were identified at the

following locations: 2a) between F/G region and the BC-loop and the β1-sheets, with an apolar mouth at the membrane 2f) between the F/G region and the C-terminal loop, with openings at the membrane interface and cytosol, S) between the E-, F-, and I-helices and the C-terminal loop, with cytosolic openings, and 2e) through the BC-loop with cytosolic openings. The conformations include one closed conformation (which contains the ligand-free crystal structures and a few other ligand-bound complexes), and two open conformations, open1 and open2. The closed and open1 conformations contained channels 2a and 2f, while the open2 conformation contained all channels. Additionally, the conformations differed between the F- and I-helices, and the intervening loops and helices (Figure 1.4B). Specifically, residues 212-218 in the F-F' loop point inwards towards the active site heme in the closed conformation, while the R212 residue and the F-F' loops rotate outwards towards the membrane at varying degrees in the open conformations.

Collectively, the x-ray crystallography data provide direct evidence that CYP3A4 is able to access multiple conformations, which in turn may influence solvent and ligand accessibility. However, although X-ray crystallography provides highly detailed structural information of CYP3A4 under different ligation states, it only captures a snapshot of the conformations suitable for crystallography. Among the uncertainties regarding the conformations of CYP3A4 captured by X-ray crystallography is their functional relevance. Additionally, the X-ray crystallography data were collected on soluble versions (heavily truncated and lacking the N-terminal transmembrane helix) in the absence of a membrane, which may influence the distribution of crystallizable conformations. Speculatively, the number of accessible conformations in the ligand-bound and ligand-free states in solution (and/or in membrane) may differ from the conformations in the crystalline state and have different catalytic properties, thus highlighting the necessity to understand the ligand-induced effects on conformational dynamics. In this work, I measured the

ligand-induced effects on conformational dynamics of CYP3A4 nanodiscs by the allosteric effector, midazolam.

1.2 Conformational dynamics: hydrogen-deuterium exchange mass spectrometry (HDX-MS)

1.2.1 Nanodiscs provide a native-like membrane environment for CYP3A4 structural studies

CYP3A4 nanodiscs are used in this work to study the conformational dynamics of CYP3A4. Nanodiscs are self-assembled proteolipid particles with membrane scaffold proteins (MSP) that encapsulate a well-defined number of phospholipid molecules in a bilayer [45,46]. Two MSPs per nanodisc particle form helical belts around the periphery of the lipid bilayers. The size of the discs and number of encapsulated lipid molecules are dictated by the specific MSP, which may be chosen from a number of engineered variants of Apo lipoprotein A. Several reviews describe these MSPs in detail [45,47-49]. Interestingly, the Apo lipoprotein A from which MSPs are derived is involved in the formation of spherical LDL particles from nascent discoidal particles in vivo. It should be emphasized that each target protein-MSP combination requires optimization of lipid:MSP:protein ratios and reconstitution conditions, as reviewed previously [49,50]. This can require significant investment of time, but once a protocol is optimized generation of protein-nanodiscs is highly reproducible. Importantly, it is possible to identify conditions where one CYP3A4 molecule is contained in one nanodisc.

Despite the challenges, lipid nanodiscs provide a way to maintain monomeric, monodisperse preparations of purified membrane proteins in a lipid bilayer environment, without the need for detergents to prevent aggregation. Furthermore, they can, in principle, allow for the systematic study of the effects of lipid composition on membrane protein structure and dynamics.

Liposomes provide an alternative membrane mimic for HDX-MS of membrane proteins, with some potential advantages, but they can also be polydisperse and the higher ratios of lipid:protein can add to difficulties in chromatography required for peptide resolution.

1.2.2 HDX-MS methodology

The static conformations of a protein obtained by X-ray crystallography and electron-microscopy are rarely sufficient to capture all of the relevant mechanistic features for a given system. In solution, a protein can sample a vast conformational landscape and a firm understanding of a protein's conformational dynamics is critical to understanding its function. Despite advances in crystallography, it is still difficult to perform crystallographic analyses for many membrane proteins [51-53]. HDX-MS in combination with nanodisc technology is an alternative approach that allows one to characterize the spatial and temporal conformational dynamics of a protein in a native-like membrane context. While NMR is heavily utilized for a wide range of protein dynamics experiments, the protein/complex size limitations, isotope costs, and large sample requirement make this approach rather intractable for many challenging, large membrane protein/nanodisc systems. Many older, elegant NMR studies demonstrated the utility of mapping HDX of individual amides to probe structure of soluble proteins [54,55], but the requirement for high sample concentrations and the presence of lipids make HDX-NMR studies intractable for membrane proteins. In contrast, the exchange of protons for deuterons is easily monitored by mass spectrometry and can be localized to specific peptides within the primary structure, upon proteolytic digestion.

There is no theoretical protein size limitation for HDX-MS (though a higher degree of complexity in the chromatographic separation of peptides may occur as protein size increases,

which may be alleviated by ion mobility spectrometry), and typical experiments with several time points only require hundreds of picomoles to a few nanomoles of the target protein. Despite the utility of HDX as a tool to probe dynamics and structure, the technique is most powerful when combined with other methods. HDX is not capable of generating *de novo* structural models, nor can it unambiguously distinguish between changes in secondary structure vs. changes in solvation due to environmental changes. Thus, HDX should be viewed as a powerful, but complementary, probe of behavior.

Exposure of a protein to a D_2O buffer induces exchange of amide backbone hydrogens for deuterium. Note that the experiment can also be done in reverse, where a fully labeled protein is exposed to an H_2O buffer. In either case, deuterium exchange is typically measured at the peptide level by monitoring the shift in the mass envelope for each peptide over time. Regions with defined secondary structural elements, such as α-helices and β-sheets, are stabilized by hydrogen bonding networks involving the backbone amide groups, so they tend to exchange more slowly than unstructured regions. HDX will also report on sequestration of regions of a protein in a solvent-inaccessible hydrophobic core, as deuteration will require both exposure of the region to solvent and fluctuations in hydrogen bonding

The experimental workflow for a standard continuous labeling experiment involves multiple steps [56]. In the HDX analysis of soluble proteins, the target protein is diluted into D_2O buffer ($pD \approx pH + 0.4$, [57]) and incubated for varying times ranging from a few seconds to many hours. These time points are selected based on the inherent dynamics of the protein and the regions of interest to give an overall profile. The intrinsic rate of exchange is temperature and pH dependent, so keeping these conditions is consistent between experiments is critical [58,59]. After discrete times following introduction of D_2O, the exchange processes are quenched by rapid

acidification and the temperature is decreased (pH 2.5 and 0 °C). The quench conditions need to be maintained for the remainder of the experiment to minimize 'back exchange' (deuterated amides begin to exchange with the protonated quench and LC buffer). Samples are then digested typically with pepsin in solution or chromatographed through an immobilized pepsin column. Other proteolytic enzymes that function at low pH can be used for alternative peptide coverage. The resulting peptides are analyzed by LC–MS. Briefly, digested samples are separated on a reverse-phase column (typically C18) with an acetonitrile gradient at low pH (0.1% formic acid). Near-freezing conditions must be maintained during chromatography and digestion/separation time should be minimized to limit back-exchange.

The advantages of the lipid nanodiscs for providing a membrane mimic and for ensuring that the samples are monodisperse are bundled with the disadvantages of having significant lipid present, which can significantly decrease the sensitivity of most MS analyses. As a result, the successful combination of HDX-MS with nanodisc-captured proteins requires modifications to typical HDX-MS protocols used with soluble proteins (Figure 1.6). Specifically, the following modifications are made to the 'standard' experimental workflow: **1**) cholate is added immediately after the exchange process to facilitate nanodisc disassembly, **2**) immobilized pepsin beads are used to minimize the number of overlapping protein peptides, **3**) ZrO_2 resin is added in the last minute of the pepsin digestion for lipid removal and to increase reproducibility in chromatography and column life in the subsequent LC methods, **4**) filtration of the final sample is done prior to freezing in liquid nitrogen to remove the immobilized pepsin beads and ZrO_2 beads with the adsorbed lipid.

1.2.3 The ligand-induced effects on the conformational dynamics of CYP3A4 are globally distributed

Structural dynamics have been studied for various CYPs under different ligations states by HDX-MS. The studies include the following CYP isoforms: CYP2B4 [60], CYP3A4 [61,62], CYP19 [63], CYP46A1 [64,65], and BM3 [66]. Among the CYP3A4 studies, two ligands have been studied via HDX-MS which include: ketoconazole-bound CYP3A4 nanodiscs compared to ligand-free CYP3A4 nanodiscs [61] and more recently (published shortly after work in this thesis was published for the midazolam-induced effects with CYP3A4 nanodiscs), progesterone-bound CYP3A4 compared to ligand-free CYP3A4 in solution [62]. Notably, despite not being in a membrane system, the ligand-free dynamics of CYP3A4 observed by Ducharme and coworkers are consistent with those reported in our lab for detergent-solubilized CYP3A4 and CYP3A4 nanodiscs [61]. In addition, the ketoconazole and progesterone-induced changes in HDX were globally distributed across the CYP3A4 structure and include overlapping regions such as the F/G region, the E-helix, the I-helix, the K/β1 region. Finally, Ducharme and coworkers used bioconjugation in combination to HDX-MS to identify an allosteric binding site for progesterone at the F'-helix which is consistent with the work presented in Chapter 3 with midazolam. In this work, I present the midazolam-induced effects on CYP3A4 dynamics via HDX-MS. This work provides evidence that although different ligands may alter dynamics in similar regions of the CYP3A4 structure (which is consistent with the published X-ray crystals), the direction of change may be ligand dependent (e.g. less dynamic vs more dynamic relative to the ligand-free CYP3A4 state). Stated differently, this work strongly suggests ligand-dependent effects on the conformational landscape of CYP3A4.

1.3 Ligand binding kinetics: stopped-flow spectroscopy

1.3.1 Traditional kinetic approach

Traditional kinetic analyses rely on interpretation of the concentration dependent behavior of the observed rate of binding obtained by mixing unbound enzyme with varying concentrations of ligand (see mixing experiment in Figure 1.7). The number of equilibration steps in a given binding model is equal to the theoretical maximum number of observable relaxation rates. For example, binding data for a simple 2-state system, which has a single equilibration step (E + L ⇄ EL; where ligand, L, binds to enzyme, E, to form a complex, EL), will exhibit a single exponential process best described by equation 1.4, where A_1 is the amplitude and $k_{obs,1}$ is the observed rate constant. Binding kinetics for multi-step binding models proceed through a maximum of *n* relaxation rates for *n* equilibria, best described by equation 1.5, which has the same nomenclature as equation 1.4.

$$f(t) = A_1 exp\left(-k_{obs,1}t\right) + c \tag{1.4}$$

$$f(t) = A_1 exp\left(-k_{obs,1}t\right) + A_2 exp\left(-k_{obs,2}t\right) + c \tag{1.5}$$

Although the number of observed relaxation processes is closely linked to the number of equilibria in the binding model, there are several examples where multi-step binding processes result in single exponential kinetics. A few examples include cases where one step of the binding process is fast compared to others if there is a stationary or relatively ‘constant’ concentration of intermediate. Thus, the observation of multi-exponential binding kinetics requires careful examination of multi-step binding models equal to or greater than the number of observed exponential processes.

1.3.2 CYP3A4 binding kinetics review

Complex ligand binding kinetics are commonly observed in CYP3A4. Multiphasic binding kinetics have been observed for the following ligands binding to CYP3A4: α-naphthoflavone, bromocriptine, carbon monoxide, clotrimazole, cyanide, flavone, indinavir, itraconazole, ketoconazole, midazolam, morphiceptin, ritonavir, and testosterone [28,29,67-71]. Several different binding mechanisms have been proposed which include different components of the following mechanisms: sequential binding (multiple ligand binding), multi-site binding ('spectrally silent' peripheral encounter complex prior to 'spectrally active' binding near to the heme), conformational selection (ligand selects a binding competent enzyme conformer from a pre-existing ensemble of enzyme states), induced fit (ligand binds and induces a conformational change), persistent heterogeneity (multiple non-interconverting enzyme conformers with different binding properties), substrate reorientation (multiple ligand binding orientations that require ligand dissociation followed by bulk reorientation prior to rebinding) [28,29,68-70]. In addition, the influence ligand binding may have on downstream steps in the catalytic cycle (such as reduction kinetics, uncoupling, oxygen dissociation, etc.) has been tested kinetically [67,71]. Notably, discerning ligand binding mechanisms is surprisingly difficult because fits are often sensitive to initial guess and prone to multiple minima. Thus, an exhaustive interrogation of multiple ligand binding models has not been performed. However, recent analytic advances have reinvigorated the examination of ligand binding kinetic data. Guengerich and coworkers recently re-evaluated previously published kinetic binding data for bromocriptine, ketoconazole, midazolam, and testosterone [68,70] and concluded binding occurs through a conformational selection model [72]. However, the data were not adequately described across all ligand concentrations for all ligands, which suggests additional complexity in the binding mechanisms. In this work, I interrogate

binding kinetics for small molecules (1,2,3-triazole, imidazole, and cyanide) which includes an exhaustive interrogation of multiple kinetic mechanisms (Figure 1.8) for ligand binding and dissociation data (see mixing and dilution experiment in Figure 1.7).

1.4 Figures

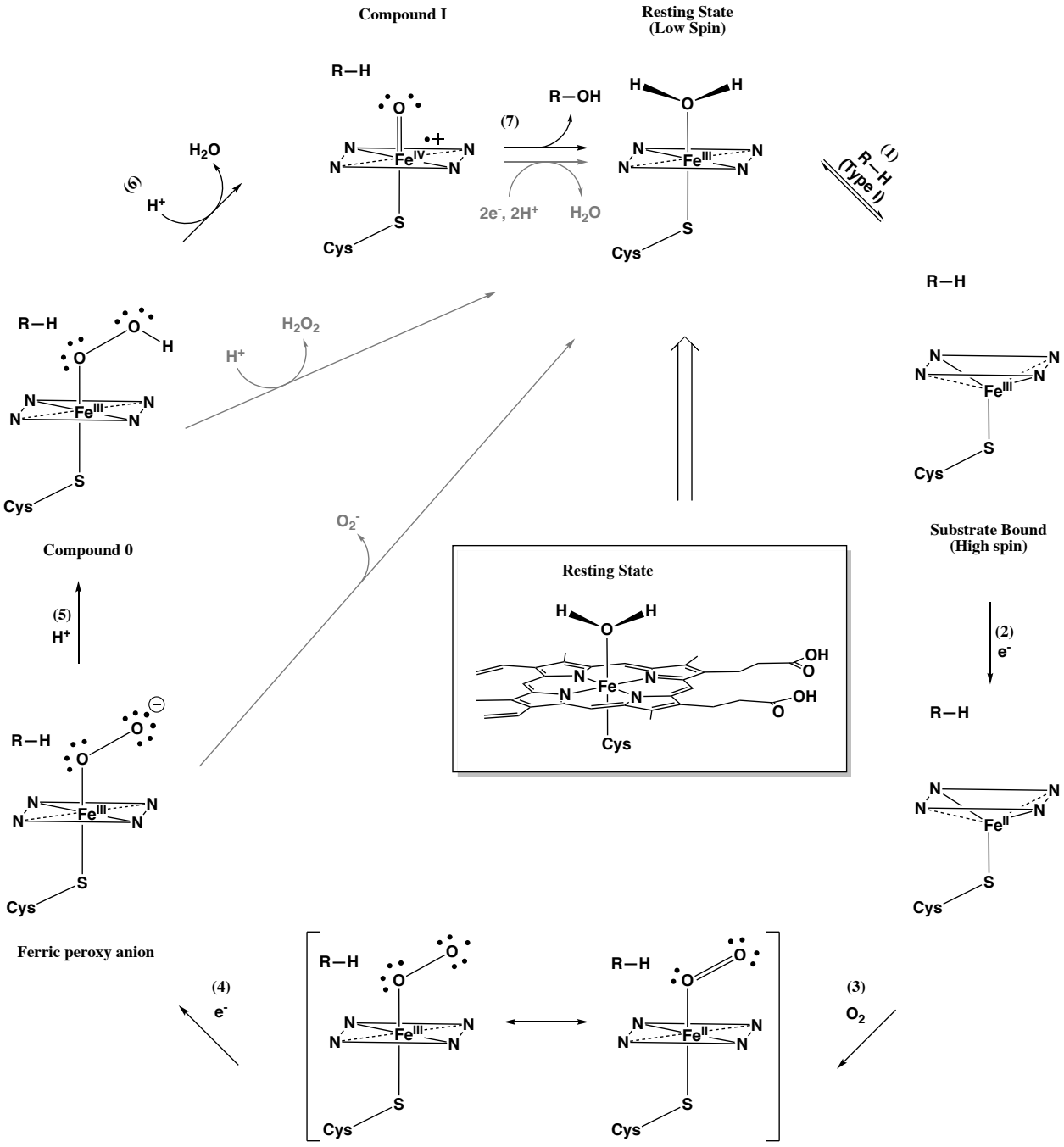


Figure 1.1. The canonical P450 catalytic cycle involving hydrophobic type I ligands. The heme in each species are shown in truncated form. The full structure of the resting state is shown for clarity. The uncoupling reaction pathways and products are shown in grey.

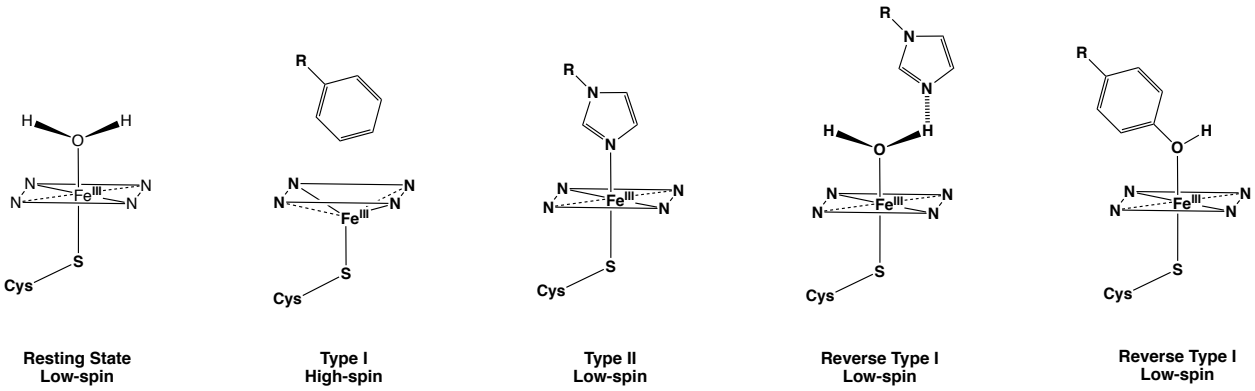


Figure 1.2. CYP heme-ligand interactions. Ligands are able to achieve different ligation modes with the active site heme which include a range of low-spin and high-spin interactions.

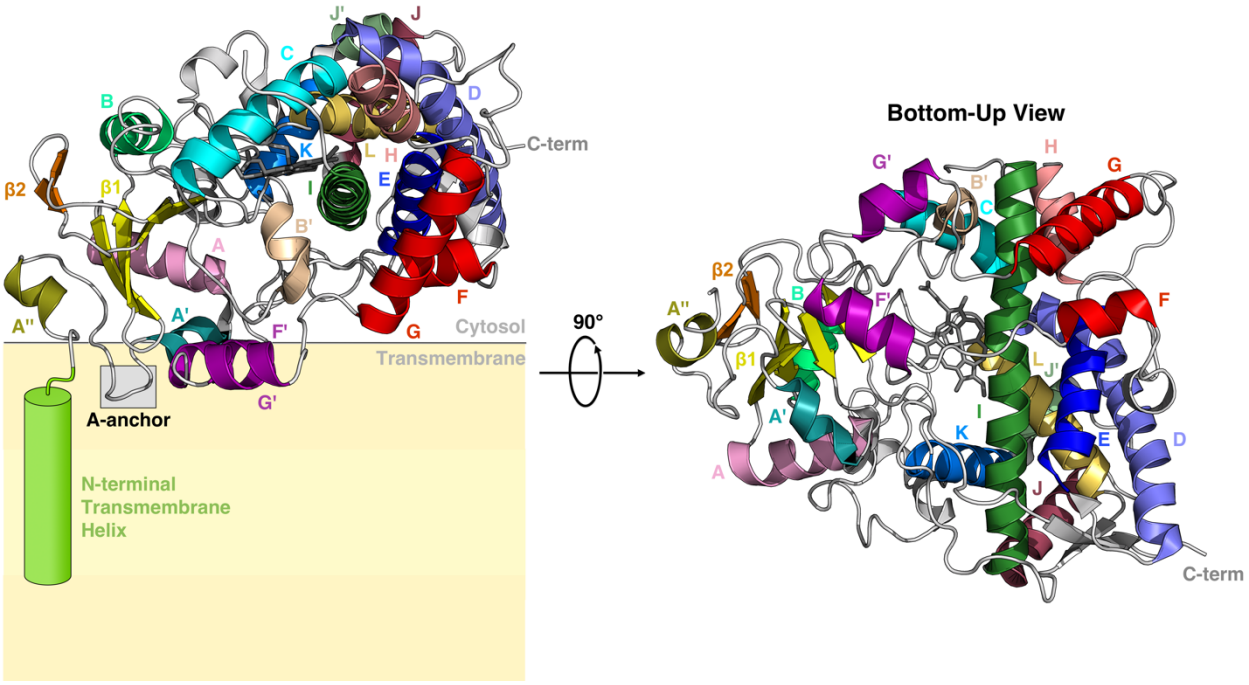


Figure 1.3. CYP3A4 structure. Cartoon representation of CYP3A4 ligand-free crystal structure (PDB: 1TQN) oriented on a membrane interface (yellow). This figure was modified from: [61].

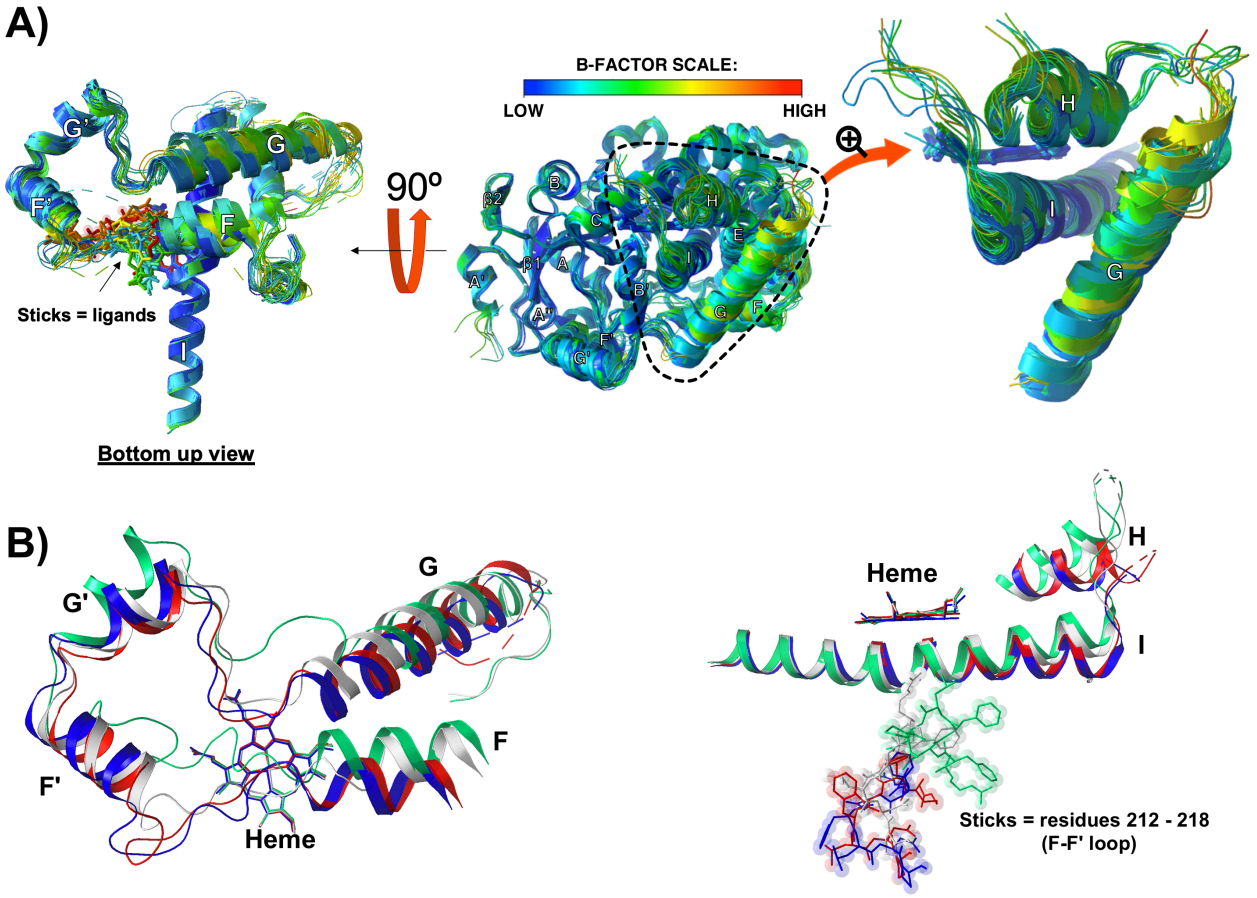


Figure 1.4. Comparison of 3A4 crystal structures. A) Overlay of 3A4 crystal structures (PDB: 1TQN, 1W0E, 1W0F, 2J0D, 2V0M, 3NXU, 3TJS, 3UA1, 4D6Z, 4D7D, 4D75, 4D78, 4I3Q, 4I4G, 4I4H, 4K9T, 4K9U, 4K9V, 4K9W, 4K9X, 4NY4, 4NZ2, 5A1P, 5A1R, 5G5J, 5TE8). **B**) Overlay of midazolam-bound CYP3A4 crystal structure with 3 representative 3A4 conformations identified in: [43]. The midazolam bound crystal structure is shown in green (PDB: 5TE8). The ligand free crystal structure is shown as a representative of the closed conformer in white (PDB: 1TQN). The ritonavir analog bound crystal structure is shown as a representative of the open 1 conformer in red: 4I4G. The ketoconazole bound crystal structure is shown as a representative of the open 2 conformer in blue

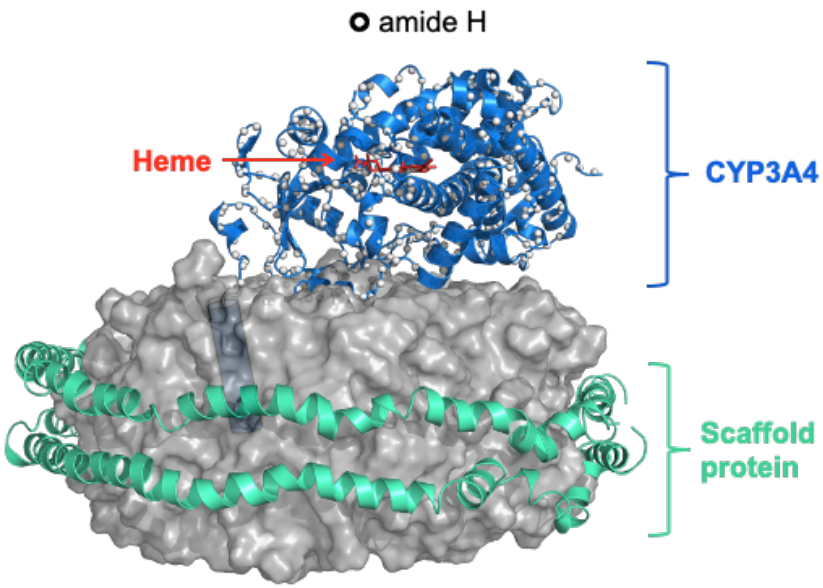


Figure 1.5. CYP3A4 nanodiscs. Cartoon representation of CYP3A4 nanodiscs.

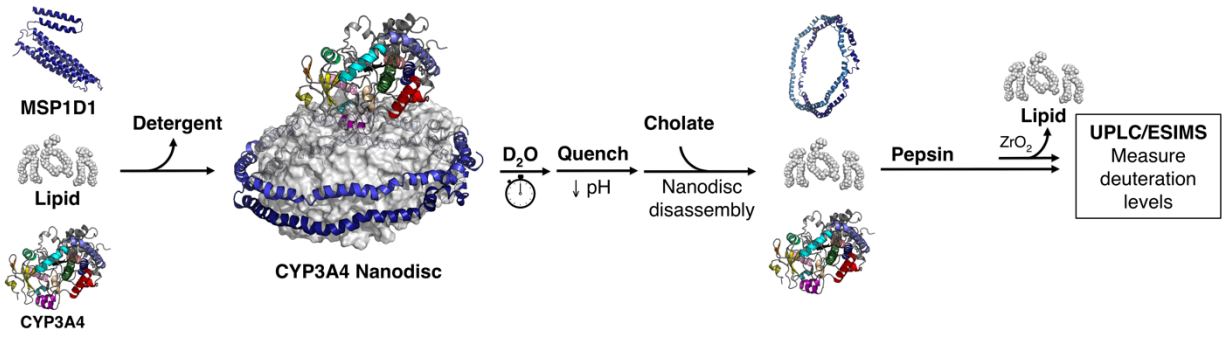
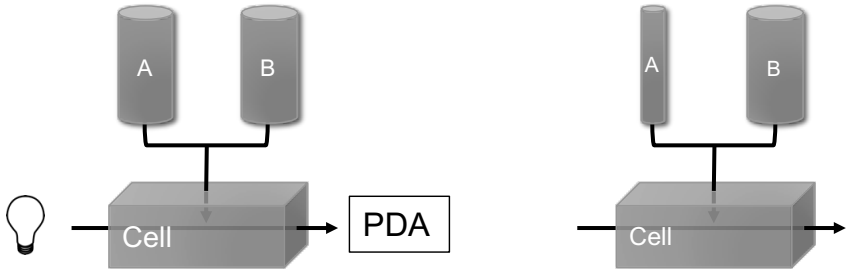


Figure 1.6. HDX-MS workflow for CYP3A4 nanodiscs. This figure was modified from: [61].



	Mixing	Dilution
Mixing Ratio:	1:1	1:1, 1:10
Start:	A) E B) L	A) E + L ⇌ EL B) Buffer
End:	E + L ⇌ EL	E + L ⇆ EL
Signal:	Formation of EL	Disappearance of EL
Information:	k_{on} and k_{off}	approach k_{off}

Figure 1.7. Stopped-flow experimental setup. The syringe contents and expected signal and kinetic information for the mixing and dilution experiments are shown for a simple 2-state process. 'E' represents the enzyme. 'L' represents the ligand. 'EL' represents the ligand-bound complex.

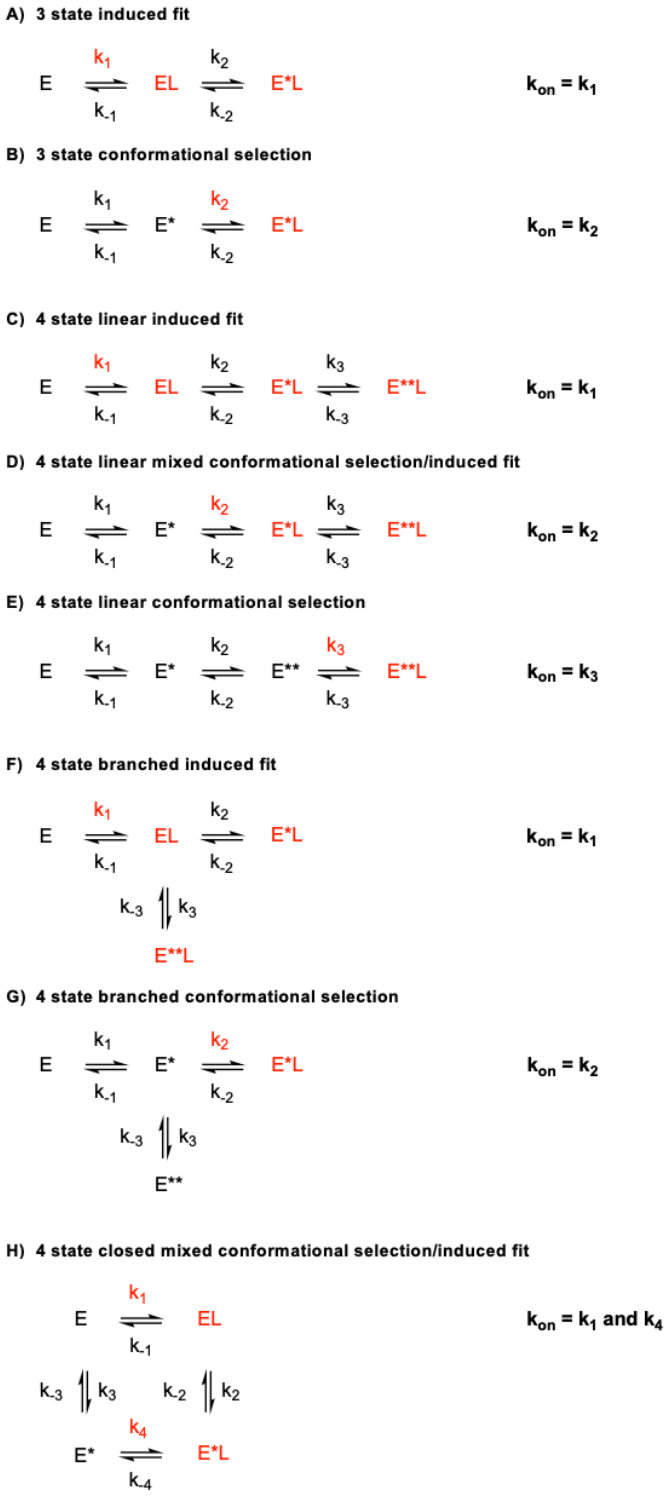


Figure 1.8. Ligand binding kinetic schemes tested in this work. The ligand binding step and ligand-bound enzyme states are highlighted in red text.

Chapter 2

Dynamics and Location of the Allosteric Midazolam Site in Cytochrome P4503A4 in Lipid Nanodiscs

Note: portions of this chapter have been published in this manuscript:

Redhair, M., Hackett, J. C., Pelletier, R. D., and Atkins, W. M. (2020) Dynamics and Location of the Allosteric Midazolam Site in Cytochrome P4503A4 in Lipid Nanodiscs. Biochemistry. 10.1021/acs.biochem.9b01001.

* I performed all experiments and data analysis except for the RMSD analysis and MD simulations, which were performed by John C. Hackett.

2.1 Introduction

Membrane-bound cytochrome P450 3A4 (CYP3A4) dominates drug metabolism in humans and displays remarkable substrate promiscuity, which is presumed to be a result of conformational plasticity. The plasticity of CYP3A4 has been documented via numerous crystal structures that reveal significant changes in active site structure and rearrangement of structural elements remote from the active site [27,36]. The ligand-dependent induced fit [73] is clearly operative with CYP3A4, wherein different protein conformations distinct from ligand-free CYP3A4 are observed when structurally distinct ligands bind. Despite the wealth of structural information, our understanding of CYP3A4-ligand interactions remains incomplete because the static models are based on the crystalline state in the absence of a membrane, which is likely to modulate the ligand-dependent conformational ensembles [74]. Furthermore, although elegant advances have been made in computational methods applied to CYPs [75,76], very few experimental approaches report directly on the dynamics of CYP3A4 in a membrane environment. As a result, our understanding of the detailed mechanisms by which drugs enter or leave the active

site is limited, and the time scales of protein dynamics or solvent reorientation in relevant access channels that link the membrane or bulk solvent to the active site are not established. By analogy with the structural plasticity indirectly suggested by CYP3A4 crystal structures, ligand-dependent protein and solvent dynamics are expected. It is possible, even likely, that different ligands bind and dissociate from the active site via different routes, but there are insufficient data to address this possibility.

Among the uncertainties regarding ligand-dependent CYP dynamics is the effect of allosteric ligands on the dynamics of CYP3A4. The allosteric behavior of CYPs, in general, is enigmatic because it parallels the substrate and catalytic promiscuity. In this regard and others, CYP3A4 allostery is distinct from traditional allostery wherein specific allosteric ligands are exploited to regulate the function of an active site. With CYP3A4, many ligands have substrate-dependent allosteric effects. The nature of such a promiscuous and substrate-dependent allosteric site is unknown. Several historical and recent studies suggest a set of overlapping allosteric sites near the Phe-cluster of CYP3A4 that provide for promiscuous allostery [77-80]. Hypothetically, the structural or dynamic properties of this allosteric site would be different from a ligand-specific allosteric site that regulates the function of substrate-specific enzymes.

Hydrogen–deuterium exchange mass spectrometry (HDX-MS) provides high-resolution information about the time-dependent exposure of peptide backbone amides of the folded protein state to solvent [58,59]. Structured regions involved in hydrogen bonding networks (helices or sheets) and regions buried in the protein exchange more slowly than unstructured regions (loops) and solvent-exposed areas. Although HDX-MS is straightforward for small cytosolic proteins, it remains challenging and less commonly used for membrane proteins [60,81], and there are very few examples of H/DX with membrane proteins in a membrane bilayer [61]. We previously have

initiated HDX-MS with CYP3A4 in nanodiscs and demonstrated that the lipid bilayer composition and structure could affect CYP3A4 dynamics [61]. Here we extend the previous work and monitor the HDX of CYP3A4 POPC nanodiscs in the presence of the classic CYP3A4 substrate midazolam (MDZ).

MDZ is metabolized to 1′-hydroxy MDZ (1′-OH MDZ) and 4-hydroxy MDZ (4-OH MDZ) by CYP3A4 (Figure 2.1A), and the product ratio exhibits homotropic and heterotropic allosteric effects *in vitro* [82], and *in vivo* [83], indicative of multiple ligand binding. The 1′-OH/4-OH MDZ ratio decreases with increasing MDZ concentrations. Mutagenesis work by Khan and co-workers supports the hypothesis that a discrete secondary binding site exists proximal to the catalytic binding site near the Phe-cluster [77]. Additionally, NMR work done by Roberts and co-workers indicates two MDZ molecules in or near the active site [82]. Furthermore, the equilibrium binding experiments described here indicate cooperative isotherms, consistent with multiple MDZ binding. As a result of the long-appreciated behavior of MDZ with CYP3A4, the current dogma is that CYP3A4 binds multiple MDZ molecules, as with other ligands, and here we aim to interrogate the structural dynamics of CYP3A4 when bound to MDZ in order to better characterize the allosteric sites.

Toward this end, we performed Gaussian accelerated molecular dynamics simulations (GaMD) for CYP3A4 in a POPC lipid bilayer under different ligation states to aid in our interpretation of the MDZ-induced changes in HDX. The HDX-MS and GaMD are reasonably cross-correlated, and they identify regions of CYP3A4 that have altered dynamics upon MDZ binding. Furthermore, the results suggest important differences between the structure of CYP3A4 in the crystalline state vs in a membrane, when MDZ is bound. The combination of several

structural and dynamic techniques appears necessary for a complete understanding of CYP3A4 interactions with ligands.

2.2 Experimental Procedures

2.2.1 Protein expression and purification (CYP3A4, MSP1D1, and rCPR)

Recombinant CYP3A4 was expressed in Escherichia coli C41(DE3) cells transformed with a pCWori+ vector encoding genes for ampicillin resistance and the Nf14 construct of CYP3A4 [84,85] preceding a C-terminal histidine tag, as previously described [86]. A single isolated colony was used to inoculate an overnight culture of luria bertani (LB) broth containing 100 μg/mL ampicillin. The overnight culture was used to inoculate terrific broth media containing 100 μg/mL ampicillin, 1 mM thiamine, and trace elements in a ratio of 1:100. After 4 hours growth at 37 ºC with shaking at 220 rpm, 0.5 mL 1M δ-ALA and 0.5 mL 1M IPTG were added to each flask to induce CYP3A4 expression and both temperature and speed of shaking were reduced to 27 ºC and 125 rpm. After 40 additional hours of growth, cells were harvested via centrifugation of each culture for 30 minutes at 4 ºC at 4,800 rpm. The resulting pellets were frozen at -80 ºC until further purification.

CYP3A4 growth cell pellets were thawed and gently resuspended in ice-cold CYP3A4 resuspension buffer (100 mM KPi, pH 7.4, 20% glycerol, 10 mM β-mercaptoethanol (BME), 3% Emulgen 911, 50 µM testosterone [TST], 20 mM imidazole (IMZ), supplemented with protease inhibitor cocktail for His-tagged proteins [EDTA-free], benzonase, and 2 mg/mL lysozyme) then stirred at 4 ºC for 1 hour before homogenizing several times through a hand-held homogenizer. A second round of stirring at 4 ºC for 4 hours followed by several more passes with the hand-held homogenizer greatly improved yield. The lysate was centrifuged for 1 hour at 4 ºC at 37,000 rpm

with a Beckman Ti45 rotor. The resulting pellet was discarded, and the supernatant was loaded onto a 30 mL Ni-NTA affinity column (Qiagen) equilibrated with wash buffer 1 (50 mM KPi, pH 7.4, 20% glycerol, 2 mM BME, 50 μM TST, 0.2% Anapoe $C_{12}E_{10}$, 300 mM KCl, 20 mM IMZ). The column was then washed with 250 mL wash buffer 1, then 500 mL wash buffer 2 (50 mM KPi, pH 7.4, 20% glycerol, 2 mM BME, 50 μM TST, 0.1% cholate, 300 mM KCl, 100 mM glycine, 20 mM IMZ), followed by a final 500 mL wash with wash buffer 3 (50 mM KPi, pH 7.4, 20% glycerol, 2 mM BME, 50 μM TST, 0.1% cholate, 40 mM IMZ). Protein was eluted with elution buffer (50 mM KPi, pH 7.4, 20% glycerol, 2 mM BME, 0.1% cholate and an IMZ gradient from 50 – 500 mM). P450 containing fractions were diluted 1:1 with an ice-cold solution of 20% glycerol and 0.1% cholate to reduce the KPi concentration before loading to a 20 mL hydroxyapatite (HA) column equilibrated with 10 mM KPi, pH 7.4, 20% glycerol, 2 mM BME, and 0.1% cholate. The column was then washed with 1000 mL HA wash buffer (25 mM KPi, 20% glycerol, 2 mM BME). Protein was eluted with 400 mM KPi, pH 7.4, 20% glycerol, and 2 mM BME. P450 containing fractions were dialyzed with a 10 kDa MWCO SnakeSkin dialysis tubing (Thermo) into 100 mM KPi, pH 7.4, 20% glycerol, 1 mM EDTA, and 0.5 mM TCEP. Protein was stored at -80 ºC till further use. Purity was >95% as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). CYP3A4 concentration was determined by "dithionite difference spectra," which is the difference between Fe^{3+}-H2O vs Fe^{2+}-CO, using an extinction coefficient of 99 mM^{-1} cm^{-1} [87,88]. The method is similar to those published by Omura and Sato (Fe^{2+} vs Fe^{2+}-CO) without the anaerobic conditions [89]. No P420 was detected from the difference absorbance spectrum.

Membrane scaffold protein, MSP1D1, was expressed and purified as previously described [61,90]. MSP1D1 was heterologously expressed in E. coli BL21-Gold (DE3) cells transformed

with a pET expression vector encoding genes for kanamycin resistance and MSP1D1 with a C-terminal histidine tag. An overnight culture grown in LB media containing 50 µg/mL kanamycin. The overnight culture was used to inoculate terrific broth media containing 100 µg/mL ampicillin and 1 mM thiamine in a ratio of 1:100. After 4 hours growth at 37 ºC with shaking at 220 rpm, 0.5 mL 1M IPTG was added to each flask to induce MSP1D1 expression. After 3 hours of growth, cells were harvested via centrifugation of each culture for 30 minutes at 4 ºC at 4,800 rpm. The resulting pellets were frozen at -80 ºC till purification.

MSP1D1 growth cell pellets were thawed and gently resuspended in ice-cold MSP1D1 resuspension buffer (20 mM KPi, pH 7.4, 1% Triton X-100, supplemented with protease inhibitor cocktail for His-tagged proteins (EDTA-free), benzonase, and 2 mg/mL lysozyme) and stirred at 4 ºC for 1 hour. Cells were lysed by french press and lysate was centrifuged for 45 minutes at 4 ºC at 22,000 rpm with a Beckman Ti42 rotor. The resulting pellet was discarded and the supernatant was loaded to a 20 mL Ni-NTA affinity column (Qiagen) equilibrated with wash buffer 1 (40 mM Tris, pH 8.0, 300 mM NaCl, 1% Triton X-100). The column was then washed with 250 mL wash buffer 1, then 100 mL wash buffer 2 (40 mM Tris, pH 8.0, 300 mM NaCl, 50 mM cholate, 20 mM IMZ), and finally with 100 mL washes buffer 3 (40 mM Tris, pH 8.0, 300 mM NaCl, 50 mM IMZ). Protein was eluted with 40 mM Tris, pH 7.4, 300 mM NaCl 400 mM IMZ. MSP1D1 containing fractions were dialyzed into 20 mM Tris, pH 7.4, 100 mM NaCl, 0.5 mM EDTA then filtered with a 0.2 µm filter. Protein concentration was spectrally determined using the calculated extinction coefficient for MSP1D1 absorbance at 280 nm (ε_{280} nm = 21 mM^{-1} cm^{-1}).

To aid in 3A4 nanodisc purification, the C-terminal histidine tag of MSP1D1 was cleaved by Pro-TEV Plus Protease (Promega Corporation) and further purified with a Ni-NTA column,

hereafter referred to as MSP1D1(-). Protein concentration was spectrally determined using the calculated extinction coefficient for MSP1D1(-) absorbance at 280 nm (ε_{280nm} = 18.2 mM^{-1} cm^{-1}).

Recombinant rat NADPH-cytochrome P450 (oxido)-reductase (rCPR) was expressed as previously described [86,91]. Briefly, cultures were grown as described for CYP3A4 except that LB media supplemented with 1 μg/mL riboflavin, 1 mM thiamine, and 100 μg/mL ampicillin was used. After 4 hr growth at 37 ºC with shaking at 220 rpm, expression was induced by the addition of 1 mM IPTG and cultures were further grown. The temperature and speed of shaking were reduced to 27 ºC and 125 rpm. After 16 additional hours of growth, cells were harvested via centrifugation of each culture for 30 minutes at 4 ºC at 4,800 rpm. The resulting pellets were frozen at -80 ºC till purification.

rCPR growth cell pellets were thawed and gently resuspended in ice-cold CPR resuspension buffer (75 mM Tris, pH 8.0, 0.25 M sucrose, 0.5 mM EDTA, 200 µM FMN, 0.02 mg/mL lysozyme, 0.5 µg/mL DNaseI, and 50 µM dithiothreitol [DTT], supplemented with protease inhibitor cocktail for His-tagged proteins [EDTA-free]) then stirred at 4 ºC for 1 hour before passage through a french press. The lysate was centrifuged for 1 hour at 4 ºC at 37,000 rpm with a Beckman Ti45 rotor. The resulting supernatant was discarded and the pellet resuspended with homogenization/wash buffer (50 mM Tris, pH 7.7, 0.1 mM EDTA, 50 µM DTT, 10% glycerol, and 0.15% Triton X-100) then passed through a hand-held homogenizer followed by centrifugation at 4 ºC at 37,000 rpm with a Beckman Ti45 rotor for 30 min. The resulting yellow/green supernatant was loaded to a 2',5'-ADP Sepharose column equilibrated with wash buffer. The column was washed with 20 column volumes of wash buffer. Protein was eluted with elution buffer (50 mM Tris, pH 7.7, 10% glycerol, 0.1 mM EDTA, 0.05 mM DTT and an NADP+ gradient from 0 – 6 mM). rCPR containing fractions were dialyzed with a 10 kDa MWCO

SnakeSkin dialysis tubing (Thermo) into 100 mM KPi, pH 7.5, 20% glycerol, and 0.1 mM EDTA. Protein was stored at -80 ºC till further use. Purity was >95% as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The rCPR concentration was determined spectrally, by measuring the difference in absorbance between 456 and 550 nm of the purified protein in solution of 0.01 mM $K_3Fe(CN)_6$ in 100 mM KPi using an extinction coefficient of 21.1 mM^{-1} cm^{-1} [92].

2.2.2 Self-assembly of nanodiscs

CYP3A4 nanodiscs were prepared as previously described [61,93]. A chloroform POPC stock solution was dried under a gentle stream of Nitrogen and remaining solvent was removed by vacuum desiccator overnight. Emulgen solubilized CYP3A4, MSP1D1(–), and POPC were mixed in a 0.1:1:65 molar ratio in disc forming buffer (DFB, 100 mM KPi, pH 7.4, 50 mM NaCl) with 20 mM sodium cholate in a bottle and placed on a Thermo Scientific Digital Bottle Roller for 1 hour at 4 ºC. Amberlite XAD-2 resin beads were added to the mixture at a final concentration 1 g/mL to remove detergent, and thus initiate nanodisc self-assembly to proceed on the bottle roller overnight at 4 ºC. The resulting solution was removed and beads were washed with 2 volumes DFB and beads removed from resulting solution. CYP3A4-nanodiscs were purified after Ni-NTA chromatography followed by size exclusion chromatography on a Superdex 200 10/300 GL column (GE Healthcare). Purified CYP3A4-nanodiscs were stored at -80 ºC till H/DX experiments.

2.2.3 Equilibrium binding titrations

Steady-state titrations were performed using a dual-beam OLIS/ Aminco DW2a spectrophotometer (OLIS, Bogart, GA) maintained at 25 °C. The sample cuvette contained CYP3A4 Nanodiscs at 0.73 μM in 50 mM HEPES pH 7.4, 50 mM NaCl. The reference cuvette contained exchange buffer. MDZ was added to both sample and reference cuvettes in 1 μL increments. Spectra were analyzed as the difference relative to ligand free spectra. The difference in absorbance between λ_{max} and λ_{min} were plotted as a function of ligand concentration and fit to the Hill equation using IGOR pro 6.35A (Wavemetrics, Lake Oswego, OR):

$$\Delta A = \frac{A_{max}[L]^n}{K_s^n + [L]^n} \quad (2.1)$$

where A_{max} is the maximum absorbance, n is the Hill coefficient, and K_s is the spectral binding affinity. It was not possible to estimate the amounts of the intermediate states of occupancy of CYP3A4 binding sites via sequential binding model (as described in [94]) without knowledge of spectral amplitudes for the separate binding events.

2.2.4 Reconstitution system and kinetic assays

Human CYP3A4 and rat CPR were reconstituted in a 1:1 ratio as described by Shaw et al. [95], except a 100× lipid stock containing 2 mg/mL POPC lipid was used in place of the lipid mixture. The final concentrations of the enzyme in each 100 μL incubation were 100 nM CYP3A4 and 100 nM rCPR. For each incubation, 80 μL of the lipid/protein reconstitution mixture was added to a 200 μL 96-well thin-walled PCR plate, followed by the addition of 10 μL of MDZ stock. All incubations contained 2% methanol. The PCR plate was placed in a hot water bath set to 37 °C and the temperature allowed to equilibrate for 5 min. The reaction was initiated by the addition of 10 μL of 5 mg/mL NADPH. After 5 min, the reaction was quenched by the addition of an equal

volume (100 µL) quench solution (1:1 mixture of ACN/MeOH containing 1′-OH MDZ-D4 as an internal standard). The plate was sealed and centrifuged for 10 min at 4 °C at 1200 rpm using a Sorvall Legend XTR tabletop centrifuge. The supernatant was transferred to a separate 96-well plate for LC-MS/MS analysis. A standard curve was created by adding known amounts of MDZ to identical incubation conditions without CYP3A4, rCPR, and NADPH.

2.2.5 LC-MS/MS method for MDZ metabolism

Analyses of 1′-OH MDZ and 4-OH MDZ activity were performed using an API 4000 QTrap (AB Sciex LLC, Framingham, MA) coupled with a Waters H Class UPLC System (Milford, MA). The metabolites were separated using a Waters Sunfire C18 column (50 × 2.1 mm, 5 µm particle size) with 0.1% formic acid in water and 0.1% formic acid in acetonitrile for mobile phases at 0.75 mL/min. The gradient used for the separation is the following: 0.0 – 0.8 min 100% A, 1.5 min 80% A, 3 min 60% A, 3.5 min 100% B, 3.6 – 6 min 100% A. The MRM transitions for this assay were 342 → 203 for 1′-OH MDZ, 342 → 297 for 4-OH MDZ, and 346 → 203 for 1′-OH MDZ-D4. 1′-OH MDZ-D4 was used as an internal standard for both 1′-OH MDZ and 4-OH MDZ. Integration and quantification were performed with AB Sciex Analyst 1.6 software using standard curves that ranged from 1 to 1000 nM for both metabolites.

2.2.6 Hydrogen-deuterium exchange

Deuterium exchange experiments were performed in duplicate with CYP3A4-nanodiscs as described previously [61,96]. A 15 µL aliquot (per incubation) of ligand-bound or ligand-free CYP3A4 nanodiscs was incubated at 25 °C for 5 min prior to dilution into 105 µL of ligand-containing or ligand-free D_2O buffer at 25 °C for 10 s, 30 s, 1.5 s, 4.5 s, and 13.5 min. The final

concentration during each exchange was 0.92 μM CYP3A4 nanodisc, 50 mM HEPES pD 7.4, 50 mM NaCl, 83% D_2O, 2% methanol, and MDZ at 0 μM, 6 μM, or 60 μM. The exchange was quenched by equal volume (120 μL) ice-cold 0.8% formic acid for a final pH of 2.5, followed by nanodisc disassembly initiated by the addition of a 4 mM ice-cold sodium cholate to nanodisc concentration. The sample was then rapidly transferred to a tube containing 200 μL of ice-cold immobilized pepsin beads (Pierce) and vortexed frequently for a total of 5 min. In the last minute of digest, 10 μL of 300 mg/mL ZrO_2 HybridSPE-Phospholipid beads (Supelco) in 0.8% formic acid was added to the digest to facilitate lipid removal [97]. After the digest was complete, the sample was filtered through a chilled 0.45 μm Spin-X microcentrifuge tube filter (Costar) to remove pepsin beads, precipitated sodium cholate, and ZrO_2 beads. The filtered sample was aliquoted and immediately flash-frozen in liquid nitrogen and stored at −80 °C until analysis. Undeuterated samples were prepared as described above without D_2O.

2.2.7 Mass spectrometry and data analysis of hydrogen-deuterium exchange samples

Text. Samples were thawed on ice over 5 min then injected with an ice-cold Hamilton syringe onto a Waters HDX Manager coupled to a Synapt G2-Si QTOF mass spectrometer, operating in positive ion mode with ion mobility enabled as described previously [61,98]. Peptides were trapped on a Vanguard BEH Shield RP18 1.7 μm trap column (2.1 × 5 mm; Waters), flowing 0.1% formic acid (FA) with 0.025% trifluoroacetic acid (TFA) at 200 μL/min for 3 min, after which the sample was resolved over a Hypersil 1.9 μm C18 column (1 × 50 mm; Thermo Scientific) using an acetonitrile (ACN) gradient of 8–50% solvent B at 40 μL/min for 8.5 min (solvent A 0.1% FA, 0.025% TFA, and 2% ACN in water; solvent B, 0.1% FA in ACN). Toward the end of the gradient, the sample flow was diverted to waste to prevent the eluting cholate from

entering the MS. Two rapid gradient steps (5–95% solvent B for 30 s) at the end of each sample run were added to minimize carryover between injections. The syringe, loop, and trap column were washed between injections with five subsequent injections of 10% FA, 30% trifluoroethanol, 80% methanol, 2:1 isopropyl alcohol/ACN, and 80% ACN [99]. Peptide identification was performed using PLGS Version 3 (Waters) by analyzing ion-mobility MSE data against a library containing sequences of CYP3A4, MSP1D1, and porcine pepsin using previously published criteria [61]. The relative deuterium uptake processed by DynamX Version 3 (Waters) and HX-Express v2 [100] is reported and not corrected for back-exchange.

2.2.8 System setup and conventional molecular dynamics

The same N-terminally modified version of the 2.05 Å resolution CYP3A4 structure (PDB ID: 1TQN) previously described [76] with the same sequence as that used in the HDX experiments was used as the starting point for all simulations. Protonation states of amino acids were assigned using Poisson–Boltzmann-based PropKa [101-104] calculations at pH 7.4. Curated CYP3A4 was then inserted into a 100 Å × 100 Å POPC membrane constructed with the Membrane Builder plugin of VMD [105]. POPC and water molecules overlapping and within 0.8 Å of the protein were removed. The system was solvated with a 10 Å layer of TIP3P water in the ± z directions. Potassium and chloride ions were added to an ionic strength of 0.15 M, thereby ensuring electrical neutrality. The final models included 253 POPC, 69 chloride ions, 56 potassium ions, and 19686 TIP3P water molecules.

All molecular dynamics procedures were performed with the NAMD 2.12 code [106] using the CHARM36 [105] force field for POPC, CHARMM27 [107] for the protein and heme, and parameters for MDZ were from CGenFF [108]. The ligand-free CYP3A4 system was minimized

for 10 000 steps, followed by equilibration of the lipids for 1 ns by freezing the positions of the lipid phosphate atoms as well the water, ions, and protein atoms. Melting of the lipid tails was followed by a 2 ns constant temperature and pressure (NPT) simulation with a harmonic restraint (5 kcal mol^{-1} $Å^{-2}$) applied to the protein Cα atoms. Bonds to hydrogen atoms were restrained with the SHAKE algorithm, and a 2 fs time step was used. The Langevin piston thermostat with a 0.5 ps^{-1} damping constant was used to maintain a constant temperature. The Nosé–Hoover Langevin piston method [109] was used with a target pressure of 1.01325 bar, oscillation period of 50 fs, decay period of 25 fs, and piston target temperature of 300 K. After completion of the preparative simulations, the complete CYP3A4 system was equilibrated in the NPT ensemble for an additional 100 ns. The end point of this simulation was used to generate both MDZ-containing systems. For the single-MDZ-containing system, MDZ was introduced only in the active site. In the two-MDZ system, the ligands were introduced into the active site as well as to a water-filled cavity bound by the β–β2 loop and the F′ helix (Figure S1). Both systems were minimized for 10,000 steps and subject to a two ns NPT simulation wherein the S119 OγH-MDZ N2 (2.6 Å) and Fe-MDZ-C1 (4.9 Å) distances of the active site MDZ were restrained with a force constant of 5 kcal mol^{-1} $Å^{-2}$. These restraints successfully reproduced the binding orientation assumed by MDZ in the crystal structure [36] and both served as starting points for an additional 100 ns unrestrained NPT equilibration.

2.2.9 Gaussian accelerated molecular dynamics

The three 100 ns unrestrained cMD simulations also served to collect the potential statistics (V_{max}, V_{min}, V_{avg}, and σ_V) necessary to calculate the GaMD boost potential Mass spectrometry and data analysis of hydrogen-deuterium exchange samples [110,111]. The endpoint of this simulation was used to spawn four each 0.5 μs GaMD simulations with independently-randomized velocities

at 300 K. GaMD simulations were performed in "dual-boost" mode meaning that boost potentials were added to both the dihedral and total system potential energies by setting the reference energy to lower bound, $E=V_{max}$. The average and standard deviation of the potential energy was calculated every 200 ps. The upper limit for the dihedral and total boost potential energies was set to 6 kcal mol^{-1} (~$10k_BT$). Trajectory frames were saved every 200 ps for further analysis. Only the latter 0.4 μs of each trajectory were subject to subsequent analyses.

2.3 Results and Discussion

2.3.1 MDZ displays complex equilibrium binding behavior to CYP3A4 nanodiscs

In order to assess the binding affinity of MDZ to CYP3A4 in nanodiscs, the CYP3A4 heme spin state was monitored at varying MDZ concentrations via UV/vis spectroscopy (Figure 2.1). The spectral titrations report on the ligation state of the ferric heme iron at the active site. In the resting state, the heme iron of CYP3A4 is predominantly in a hexacoordinate low-spin state with water in the sixth axial position. Displacement of the water upon binding of MDZ results in a transition to a pentacoordinate high-spin state, termed a type I spectral change. Figure 2.1B shows the characteristic type I difference spectra induced by MDZ and the resultant binding isotherm, consistent with previous reports. A fit to the Hill eq (equation 2.1) resulted in a K_S value of 14 ± 1.1 μM and a Hill coefficient of 1.9. For similar titrations in phosphate buffer and in the absence of a membrane, the data were fit to hyperbolic binding isotherms, and the recovered KD varied with the phosphate concentration [36,112].

For CYP3A4 in nanodiscs, our titration data do not fit well to hyperbolic binding, and significant cooperativity is apparent. A comparison of the data fit to a hyperbolic binding isotherm or Eadie–Hofstee plots is shown in Figure S2. Because the spectral titrations report only on the

displacement of the water at the heme iron, they are insufficient to determine the binding pose of MDZ and the stoichiometry. Importantly, unlike sigmoidal velocity vs [substrate] plots from catalytic turnover, sigmoidal binding isotherms at equilibrium demand multiple ligand binding [113]. Thus, in accordance with the existing dogma, we consider two MDZ binding sites for CYP3A4 in lipid nanodiscs. The arrows in Figure 2.1B indicate the concentrations of MDZ used in the H/DX experiments below. The turnover experiments are intended only to demonstrate the functional consequences of changing ratios of [CYP3A4 MDZ] vs. [CYP3A4 MDZ MDZ]. Obviously, the data indicate that the concentration dependence of the heme spin state in lipid nanodiscs is different from the concentration dependence of the product ratio in the turnover conditions, which include CPR and lack the membrane. It is important to note that the conformational status of CYP3A4 varies in nanodiscs vs detergent or lipid solution. In some solutions, CYP3A4 forms oligomers with functional and conformational differences compared to monomers, including differences in allosteric properties [67,114]. Thus, the quantitative and qualitative differences between the binding isotherms we observe in nanodiscs vs the isotherms observed by others in the absence of lipid bilayers are expected, and this precludes comparison of the catalytic data with the spectral binding data in terms of conformational changes that dictate regioselectivity.

2.3.2 MDZ metabolism by CYP3A4

In order to provide context for the multiple MDZ binding and to relate these studies to the previously appreciated behavior of CYP3A4, we confirmed the concentration-dependent switch in regioselectivity of MDZ hydroxylation. Determination of kinetic constants, K_M and V_{max}, using CYP3A4 and CPR followed the procedure of Shaw et al. [95] with the exception that only POPC

lipid was used in the reconstitution system to match better the HDX-MS experiments where POPC nanodiscs were used. Data for the rate of product formation fit well to the Michaelis–Menten equation (Figure S1C), and the change in product ratio vs concentration is shown in Figure 2.1C. The results are consistent with, but modestly different from, other reports with similar reconstitution systems containing recombinant enzymes [77,112] without the inhibition seen in human liver microsomes and supersomes [82,115,116]. The K_M values for 1′-OH MDZ and 4-OH MDZ formation were 17 and 389 μM, respectively. The Vmax values for 1′-OH MDZ and 4-OH MDZ formation were 3.5/min and 1.5/min, respectively. The different environments, with or without the redox partner or membrane, could contribute to differences in K_M values and K_S based on spectral titration. The choice of MDZ concentrations for the H/DX MS studies below was based on the spectral titrations in nanodiscs.

2.3.3 HDX-MS profile of CYP3A4 nanodiscs

CYP3A4 nanodiscs in different ligand-bound states (±MDZ) were analyzed by HDX-MS. After optimization of incubation and quench conditions, a sequence coverage of 88% with 70 unique peptides for CYP3A4 was achieved. Deuterium uptake profiles for the ligand-free state are consistent with previously published results [61]. More detailed analyses and comparison with the MDZ-bound states are provided below. Individual deuterium uptake plots for specific peptides are shown in Figure S3.

2.3.4 MDZ binding induces changes in HDX-MS

HDX was performed at two concentrations of MDZ, 6 μM and 60 μM. Although the HDX does not reveal changes due specifically to the first vs the second ligand, the binding data do

suggest that, at 60 μM MDZ, the overwhelming majority of CYP3A4 has two MDZ molecules bound. A general observation based on the results is that, for the majority of peptides, there is no effect of 6 μM MDZ, suggesting that neither the first nor second MDZ binding event causes dramatic changes in HDX for any peptides. At 60 μM MDZ, the MDZ-dependent changes in CYP3A4 dynamics are not limited to the active site but expand to distal regions of the enzyme.

Deuterium uptake plots for peptides with significant changes in HDX are shown in Figure 3.2. The differences are mapped onto the MDZ-bound CYP3A4 crystal structure (PDB ID: 5TE8) [36] where an increase or decrease in HDX relative to the ligand-free state is highlighted in red or blue, respectively. It should be emphasized that these changes are modest but reproducible across multiple CYP3A4 nanodisc preparations and separate experiments performed over the course of several years. The data shown are from same-day duplicate incubations that were injected onto the MS using optimized LC/MS methods such that all peptides were recovered from a single injection to allow for quantitative comparisons. The error for each time point is shown in the deuterium uptake plots and is usually smaller than the data points. Statistical significance was established by paired t tests for each time point across the different ligation states. Data for peptides 1, 3, and 5–8 yielded $p < 0.05$ for the comparison of 60 μM MDZ to 0 μM MDZ data. Data for peptide 4 approached $p = 0.07$ for the last two time points. There were no significant differences in the time points in peptide 2. Importantly, data for overlapping peptides, when available, displayed the same trends in differences, thus supporting the significance of the observed differences.

Regions affected by MDZ binding are globally distributed rather than confined to the active site. In our work, HDX indicates that the largest effects on dynamics upon multiple MDZ binding occur at the F- and G-helices and the intervening loops (Figure 3.2, peptides 3–6), as well as the loop connecting the K-helix and β1-sheet (peptide 7). There are modest effects at the BC loop as

well as on the backside of the protein at the β3/4-sheets (peptides 1 and 8, respectively), which become more protected. Although no difference in HDX was observed at the C-terminus of the B–C loop (peptide 2), we include the uptake for this peptide in Figure 3.2 because it contains Ser-119, which forms a hydrogen bond with MDZ in the crystal structure [36].

At the active site, peptides with residues in contact with the single MDZ in the crystal structure display changes in our HDX experiments at 60 μM MDZ. Noteworthy contacts in the crystal structure are present in peptides 2, 7, and 8 at the following residues: (A) Ser-119 from peptide 2 forms a hydrogen bond with N2 of MDZ, (B) Ile-369 and Ala-370 in peptide 7 contact the chlorophenyl ring of MDZ, and (C) Leu-482 in peptide 8 contacts the fluorophenyl group of MDZ. A decrease in HDX was observed for peptides 7 and 8, while no difference was observed for peptide 2. However, it must be emphasized that exchange at amino acid side chains in proteins is not quenched at a low pH, as an exchange at the backbone amide hydrogen bonds is. As a result, the exchange of deuterons at amino acid side chains is invisible to HDX, and only peptide amide hydrogen bonds are probed. Additionally, peptide 2 is a highly protected region in the ligand-free state, and any potential protective effect from MDZ might be masked in the time scale of the experiment. Notably, the region is also influenced by KTZ binding, which will be discussed in greater detail below.

At the membrane interface, the addition of 60 μM MDZ caused an increase in HDX at the F-, G-, and intervening helices (peptides 3–6), while a modest protective effect was observed at the opposite side of the membrane at the β3/4-sheets (peptide 8). Conformational changes at the membrane interface observed in the GaMD simulations will be discussed in greater detail below.

2.3.5 GaMD simulations with MDZ in the catalytic and allosteric sites

The CYP3A4-POPC membrane system used for simulations with MDZ was similar to the system described previously to simulate TST binding to CYP3A4 [76]. While available, features of the structure of the CYP3A4-MDZ crystal structure (PDB ID: 5TE8) are not consistent with the ability of CYP3A4 to bind multiple ligands. In the structure, the F–F′ loop collapses into the active site and packs against MDZ, precluding binding of a second ligand. This distinct conformational feature was attributed to the MDZ-induced fit; however, the observed collapse of a hydrophobic region into the active site is also likely due in part to the high ionic strength required to obtain the crystal. Binding studies at similarly high ionic strengths, supporting hydrophobic collapse, also occur in solution, yielding hyperbolic binding curves consistent with a single MDZ binding site. The K_D values of other ligands also shift to lower values with increasing ionic strength, demonstrating that the phenomenon is not unique to MDZ. We attribute the distinct conformational features of the 5TE8 structure, in part, to ionic strength-induced hydrophobic collapse rather than the ligand-induced fit. This proposal is further supported by the inability to crystallize the complex at a lower ionic strength due to MDZ's diffusion from the crystal, as noted by the authors. Had these structures been reported, we speculate that they would have revealed a more open active site. Therefore, we used the ligand-free 1TQN structure of CYP3A4 as it was obtained at an ionic strength comparable to the solution studies described herein and had a more open active site consistent with multiple ligand binding. The steric clash of the F–F′ loop and the allosteric MDZ is shown in Figure S4, where the local structure of the 5TE8 structure is overlaid with 3 snapshots from MD simulations with 2 MDZ starting from the 1TQN crystal structure.

For the MD studies here, 100 ns of conventional MD (cMD) simulations were performed without a ligand, with one MDZ oriented in the active site, and with MDZ oriented in the active site, as well as in a probable entrance channel (Figure S1). Prior to the cMD simulations, the MDZ

occupying the catalytic site was initially subject to a short simulation, where it was restrained to reproduce the orientation of MDZ observed in the crystal structure [36]. The end points of the MDZ-free simulation as well those with one and two MDZ served as the initial configurations for each independent Gaussian accelerated molecular dynamics (GaMD) simulations. GaMD affords sampling of rare events such as infrequent conformational changes or ligand binding events that occur on longer time scales not accessible by cMD. GaMD achieves more extensive conformational sampling by lowering the energetic barriers and thereby increasing the likelihood of escape from energetic basins [117].

Comparison of the backbone amide nitrogen root-mean-square deviations calculated for each group of GaMD trajectories reveals only a modest correlation between the MDZ-induced changes in RMSD (Figure 3.3) and the regions that exhibit a change in HDX in the mass spectrometry experiments (Figure 3.2). Apparently, the second MDZ affects backbone dynamics in several spatially distributed peptides.

In contrast, there is a correlation between the peptides affected in H/DX experiments and those contributing to the allosteric MDZ site in Figure 3.4. Of the very few peptides that show a change in H/DX (peptides 1, 3, 4, 5, 6, and 7), essentially all are contained within or directly adjacent to the binding site for the second MDZ. However, there are additional peaks or troughs in Figure 3.2B that are not in the binding site, some of which correspond to peptides for which we do not have coverage in the H/DX experiments. The effects of MDZ on the RMSD of these peptides are interesting inasmuch as they are varied with respect to the direction of change. For example, peptide 3 exhibits a large increase in RMSD, with either one or two MDZs present, and an increase in HDX at later times with two MDZs. In contrast, peptide 7 and the edge of peptide 1 exhibit an MDZ-dependent increase in RMSD, but they exhibit a decrease in HDX. Peptide 5

yields a decrease in RMSD but a pronounced increase in HDX. It is interesting that there are more "positive differences" in Figure 3.2B for the doubly bound state than negative differences, suggesting the enzyme is more dynamic in these peptides when two MDZ's are bound.

The differences in the direction of the effects of MDZ on HDX vs RMSD in some cases amplify the uncertainty about the factors that control H/D exchange rates in a membrane environment where a local charge may be a critical determinant and could oppositely affect backbone dynamics and H/D exchange [118]. An additional source of the distinction between the approaches is that the changes reflected in the MD simulation are sampling time regimes that are outside the scope of the HDX experiments [119] and vice versa. As shown by others [119,120], it is possible, even expected, that MD and HDX could exhibit apparently opposite effects of ligands because they sample such different time scales. For example, ligands could damp motion on fast time scales for some peptides but increase slow domain motions that increase HDX in the same peptides on longer time scales. Therefore, for the sake of locating the allosteric MDZ site, the most relevant comparison from MD and HDX is which peptides are affected, rather than the direction of the change. The overlap of peptides that are affected in the HDX at 60 μM MDZ and those affected in the GaMD upon the addition of the second MDZ suggests that the observed changes in HDX at 60 μM MDZ are due primarily to occupancy of the second, allosteric, site.

An essential point is that all of the CYP3A4 residues that make close contact with MDZ in the crystal structure also do so with overwhelming frequency in the simulations despite starting the simulations from a different structure (Figure 2.4A). However, when one MDZ is bound, it is highly mobile and rarely assumes the same binding mode that was observed in the crystal. This mobility is evident in Figure 2.4B that illustrates the distances between the MDZ nitrogens and Ser-119 as well as the observed sites of oxidation (1′ and 4) and the heme iron. The hallmark of

the CYP3A4-MDZ crystal structure is an apparent hydrogen bond between the N2 of MDZ and the Ser-119 OγH. Despite enforcing this hydrogen bond at the outset of the simulations to establish the crystallographic binding mode, this hydrogen bond is only transiently maintained, appearing in less than ~25% of the combined trajectories with an average distance of 5.6 ± 1.2 Å. The MDZ N5 is not oriented for hydrogen bonding to Ser-119 either, though it occasionally hydrogen bonds with the guanidinium side chain of Arg-105 (Figure S5). As evidenced by the sparsity of the S119 hydrogen bond, the absence of the collapsed F–F′ loop and the more open active site provides an expanded volume for the single MDZ to explore.

When a second MDZ is introduced into a possible entrance channel bound by the β1–β2 loop and F′ helix, it rapidly moves into a binding site adjacent to the first MDZ surrounded by residues of the B–C loop and the Phe-cluster (Figure 2.4C). Many of the residues in contact with the second MDZ such as Ile-120, Ile-301, Phe-241, and Phe-304 were also predicted to partially define an intermediate state on TST's binding path [76]. It was proposed that the intermediate site is also an allosteric site that could be occupied at high TST concentrations. The introduction of a second MDZ into the active site substantially attenuates the mobility of the first. However, more remarkable is the recovery of the MDZ N2-Ser119 OγH hydrogen bond observed in the crystal, evidenced by an average distance between the relevant oxygen and nitrogen atoms of 3.4 ± 0.8 Å (Figure 2.4D). Recovery of the S119 hydrogen bond observed in the crystal structure is readily explained by comparing the 5TEQ structure with snapshots from the MD simulations (Figure S4). The hydrophobic collapse of the F–F′ loop in the 5TEQ structure orients several residues, most notably Leu-216, into the same volume occupied by the allosteric MDZ. Despite their structural dissimilarity, the addition of a second MDZ enforces the same binding mode and protein

interactions and similarly limits the orientational heterogeneity of the first MDZ as filling this space with residues of the F–F′ loop.

The simulations reveal MDZ-dependent, correlated changes in interactions between several CYP3A4 regions, namely, regions including peptides 1, 4, and 5. The B–B′ loop, including peptide 1, undergoes distinct conformational changes in the presence of the second MDZ. This is at least partially attributable to direct interactions with the MDZ occupying the allosteric site where it primarily packs against the Val-111 side chain (Figure 2.4C) in more than 90% of the trajectory and transiently accepts an apparent hydrogen bond at the N5 position from the Gly-109 backbone amide.

What is more striking is that the MDZ occupation of the allosteric site stabilizes a bidentate hydrogen bond between Glu-234 on the G′-helix and the backbone amides of Val-111 and Gly-112. (Figure 2.5A,B) This interaction impacts the local fold of the B–B′ loop (Figure 2.5C–F) and induces changes in dynamics that propagate to peptides 4 and 5 to impact the interactions of these regions with the membrane. Figure 2.5G,H illustrates the cumulative probabilities of finding the mass-weighted centers of peptides 4 and 5 in the bilayer. For reference, these are superimposed on the average positions of the POPC membrane acylglycerols and phosphates. Remarkably, while the occupation of the catalytic site alone results in a small disengagement of peptide 4 from the membrane and small changes in the position of peptide 5, the occupation of the allosteric site by a second MDZ shifts both of these peptides deeper into the membrane, but they do not become immobilized.

The striking result that emerges from the GaMD is that the first MDZ bound at the active site is highly mobile and makes transient interactions with several active site residues. Upon addition of the second MDZ that binds in the allosteric site, the active site MDZ becomes

significantly less mobile. In contrast, the allosteric MDZ remains dynamic. The RMSD for the distance between the center of mass of each MDZ is shown in Figure 2.6. The allosteric MDZ is highly mobile, and its movement toward and away from the active is coupled to rearrangements in the Phe-cluster, whereas the active site MDZ remains relatively static. Thus, the GaMD results suggest that the ligand in allosteric sites remains dynamic.

2.3.6 Comparison of MDZ and KTZ effects on HDX

A comparison between the MDZ and previously documented KTZ-induced changes in HDX reveal that these ligands differently affect the dynamics of CYP3A4. The regions with the largest changes in HDX are mapped onto the ligand-free CYP3A4 crystal structure in Figure 2.7 (PDB ID: 1TQN). Peptides at the F- and G- helices that are exclusively altered by MDZ (increase in HDX) or KTZ (decrease in HDX) are highlighted in red or blue, respectively. Both ligands cause a decrease in uptake at the loop preceding the β1-sheets for both ligands, which is highlighted in purple. At the F′-helix, MDZ binding results in an increase in HDX, while the opposite effect is observed when KTZ is bound, highlighted in green. It is interesting that HDX is dramatically different at the F- and G-helices, given their potential role in ligand entry/exit channels.

2.4 Conclusions

The location and dynamic nature of the allosteric MDZ site are noteworthy. The GaMD and the HDX suggest that the Phe-cluster and its local environment comprise a dynamic site between the lipid bilayer and the active site. This site has been suggested to accommodate several allosteric ligands and to mediate allosteric effects [77-79]. The effects of the allosteric MDZ on the Phe-cluster are essentially opposite to the effects caused by KTZ in this region, based on H/DX.

Apparently, the promiscuity that is well appreciated for the active site of CYP3A4 is paralleled by the promiscuity of the allosteric site. Interestingly, allosteric sites used for regulation of enzymes in well-defined metabolic pathways are typically highly specific. The current results suggest a possible mechanistic basis for a promiscuous allosteric site, wherein ligand binding, either at the active site or the allosteric site, increases the local dynamics of the latter and generates a highly fluid binding site directly adjacent to the active site. The fact that an allosteric site can remain dynamic or fluid even when occupied by a ligand, as suggested by the HDX and GaMD, is consistent with a highly promiscuous allostery, which seems to be the case with CYP3A4. Alternatively, the changes in H/DX for peptides comprising the allosteric site could be due to their movement deeper into the membrane, which changes the local electrostatic environment and affects the H/DX, or a combination of both electrostatics and dynamics.

2.5 Figures

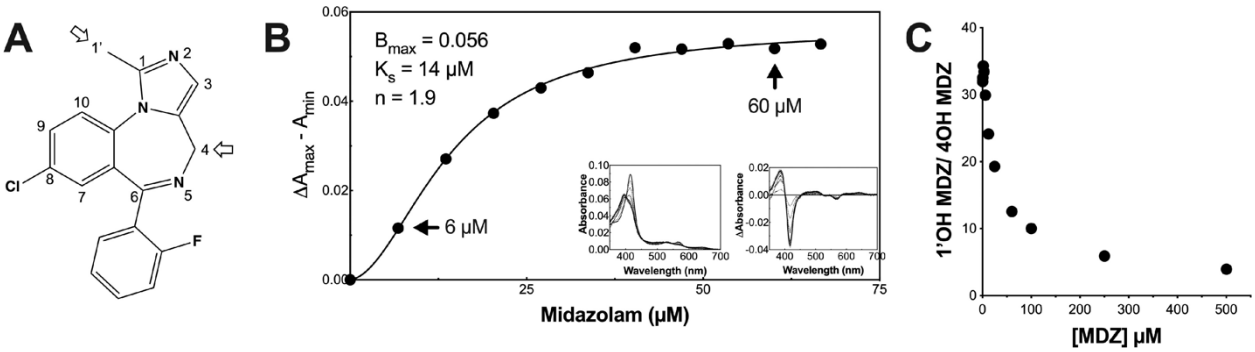


Figure 2.1. MDZ structure, equilibrium titration into CYP3A4 nanodiscs, and product ratio vs MDZ concentration plot. (A) MDZ structure with the indicated primary (C1′) and secondary (C4) hydroxylation sites. (B) Binding isotherm for MDZ with CYP3A4 nanodiscs, based on optical difference spectra as described in the methods section. The black line is the fit to the Hill equation, and black points are the optical titration data. The left inset and right inset are the absolute and difference spectra at varying concentrations of MDZ, respectively. The arrows indicate the two concentrations used in the HDX experiments, 6 or 60 μM MDZ. (C) 1′-OH MDZ/4OH MDZ product ratio vs MDZ concentration (μM).

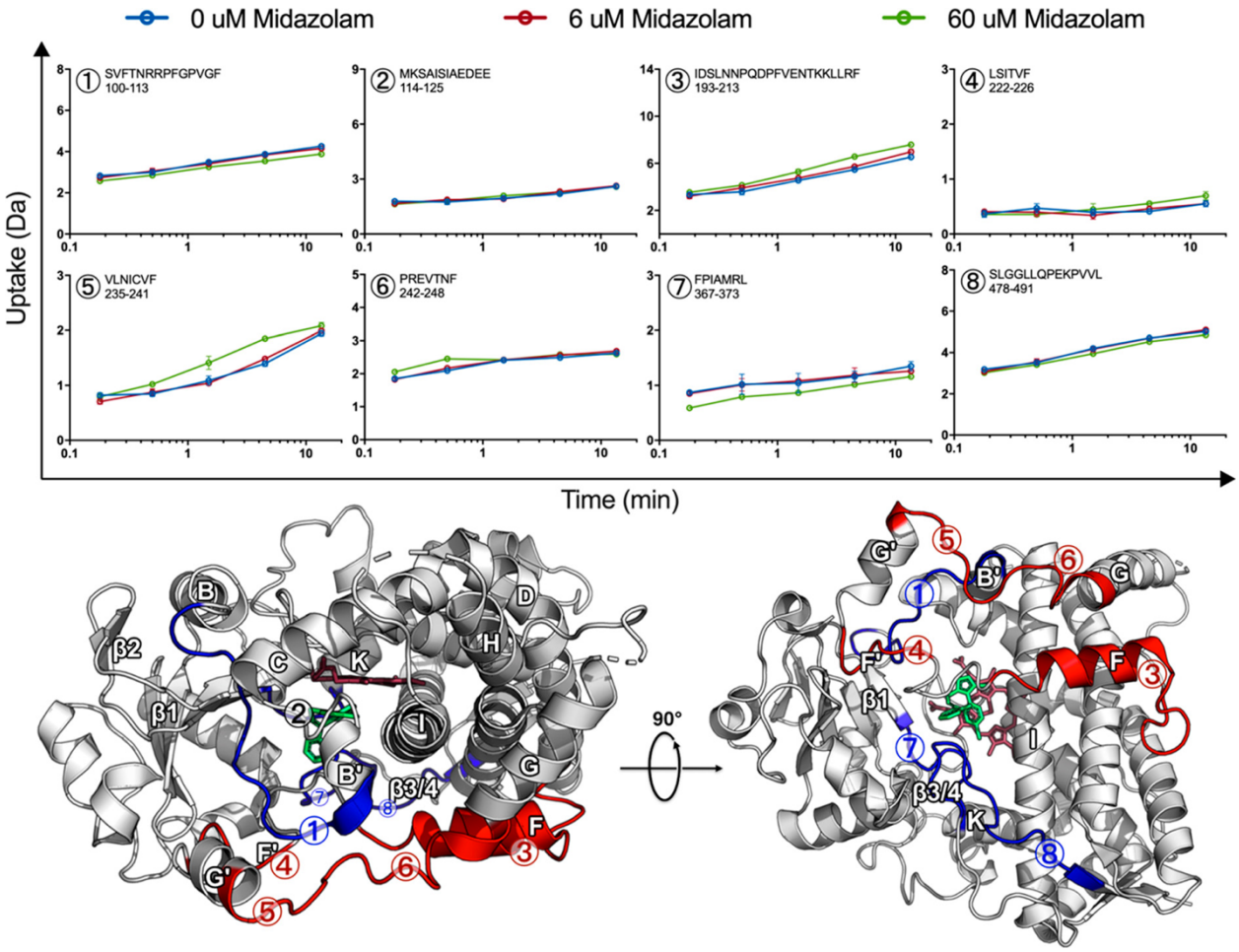


Figure 2.2. MDZ-induced changes in HDX changes. Uptake plots for selected peptides, 0 μM MDZ (blue), 6 μM MDZ (red), and 60 μM MDZ (green) with error bars representing the standard deviation for duplicate measurements. Differences are mapped onto the MDZ-bound CYP3A4 crystal structure (PDB ID: 5TE8), where the MDZ-induced increase and decrease in HDX for the 60 μM MDZ state are highlighted red and blue, respectively. The heme and MDZ are shown by stick representation in dark red and green, respectively.

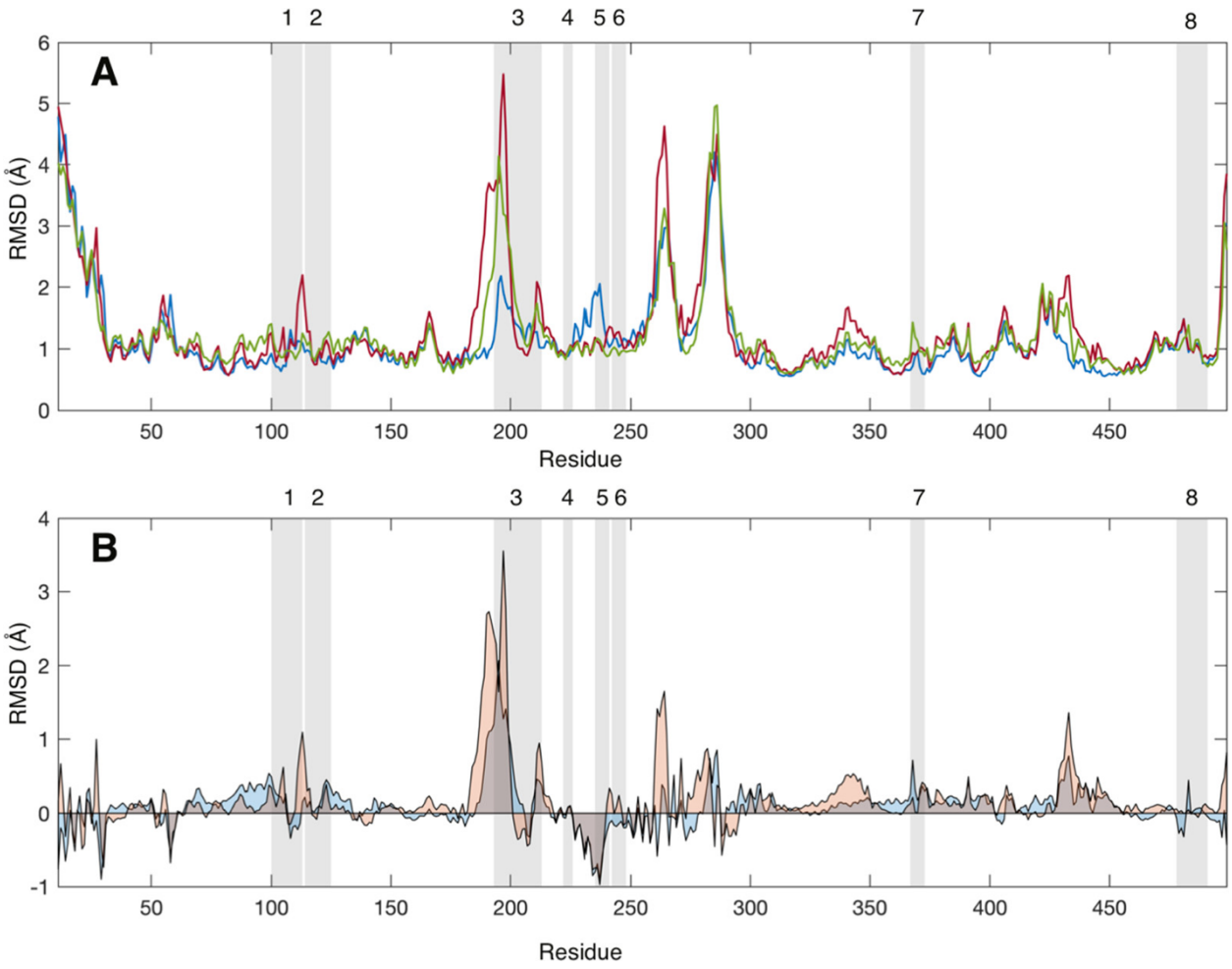


Figure 2.3. Average RMSD values for CYP3A4 residues under different ligation states. (**A**) Average root-mean-square deviation (RMSD, Å) values for each CYP3A4 residue without a ligand (blue), with one MDZ (red), and with two MDZ (green) bound in the active site. (**B**) Average RMSD values for each CYP3A4 with one MDZ (blue) or two MDZ (orange) less the average values for the ligand-free enzyme. RMSD values were calculated relative to the average structure over the entire trajectory. Gray-shaded regions highlight the peptides identified in the mass spectrometry experiments. RMSD values were calculated using the CA, CB C, O, N, HA, HA1, HA2, and HN atoms for all frames in the latter 0.4 μs of each simulation.

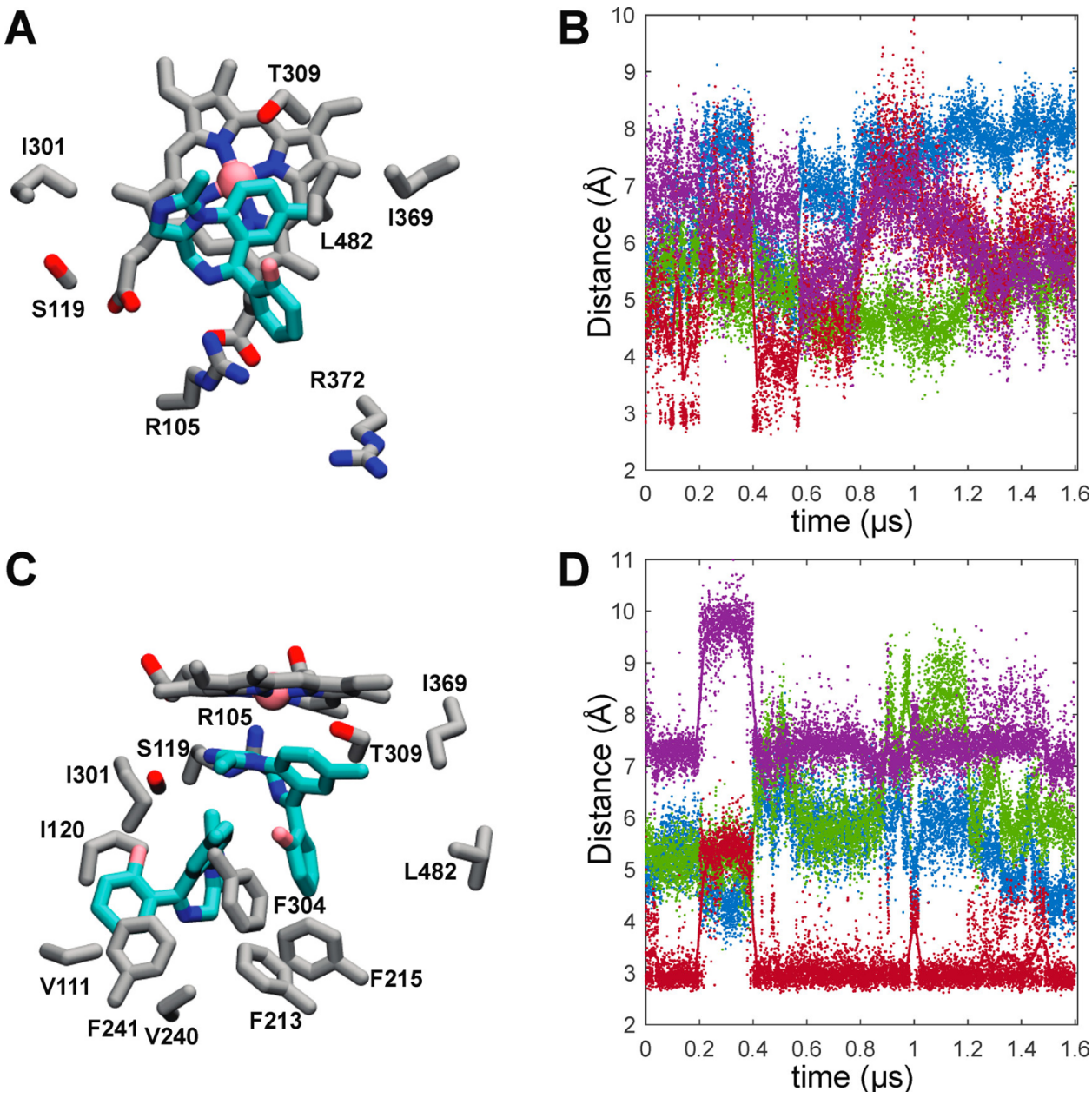


Figure 2.4. Representative MDZ binding modes and time dependences of active site MDZ-CYP3A4 interactions. (**A**) Representative MDZ binding mode from the single-MDZ system with side chains of residues that contact MDZ in 90% or more trajectory frames. (**B**) Time dependence of active site MDZ-CYP3A4 distances in the single-MDZ simulations. (**C**) Representative MDZ binding modes from the double-MDZ simulation with side chains of residues that contact each MDZ in 90% or more trajectory frames. (**D**) Time dependence of the active site MDZ-CYP3A4 distances with two MDZ bound. In plots B and D, MDZ N2-Ser119 Oγ (red), MDZ N5-Ser119 Oγ (purple), MDZ C4-Heme Fe (green), and the MDZ C1-Heme Fe (blue) distances are illustrated with solid points.

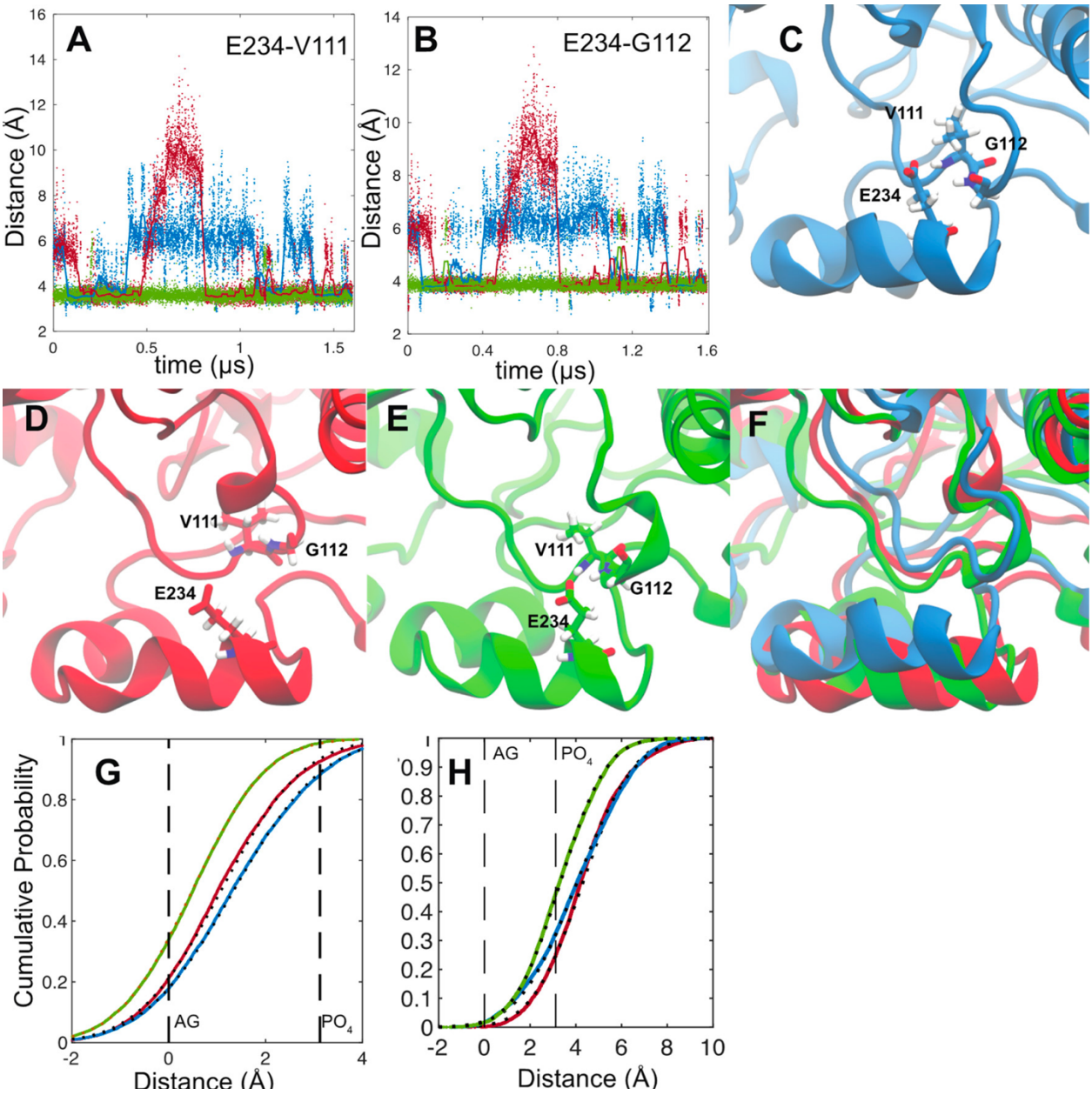


Figure 2.5. G'-helix distance with B-B'-loop and membrane interactions in the simulation. Distances between the Glu-234 carboxylate center of mass (calculated using Cδ and both O atoms) and the backbone amide nitrogen (**A**) Val-111 and (**B**) Gly-112. Distances from the ligand-free (blue), single-MDZ (red), and double-MDZ (green) simulations are illustrated as solid points. Solid lines represent a 30 ns moving average of the distances. Representative snapshots from (**C**) ligand-free, (**D**) single-MDZ, and (**E**) double-MDZ simulations. The three snapshots are superimposed in panel **F**. Cumulative probability plots of the distance of (**G**) peptide 4 (222–226) and (**H**) peptide 5 (235–241) centers of mass from the average position of the POPC acylglycerols (AGs) and phosphates (PO4). Distance data for from the ligand-free (blue), single-MDZ (red), and double-MDZ (green) simulations are illustrated with solid points. Fits to a normal distribution are illustrated with black-hashed lines. Peptide centers of mass were calculated using the weighted masses of the peptide backbone C, O, N, H, Cα, and Cβ atoms. AG centers of mass were calculated

using the C1, C2, C3, O11, O21, and O31 atoms. Phosphate centers of mass were calculated using the P, O11, O12, O13, and O14 atoms. AG and phosphate atom numbers are from the CHARMM 36 lipid topology.

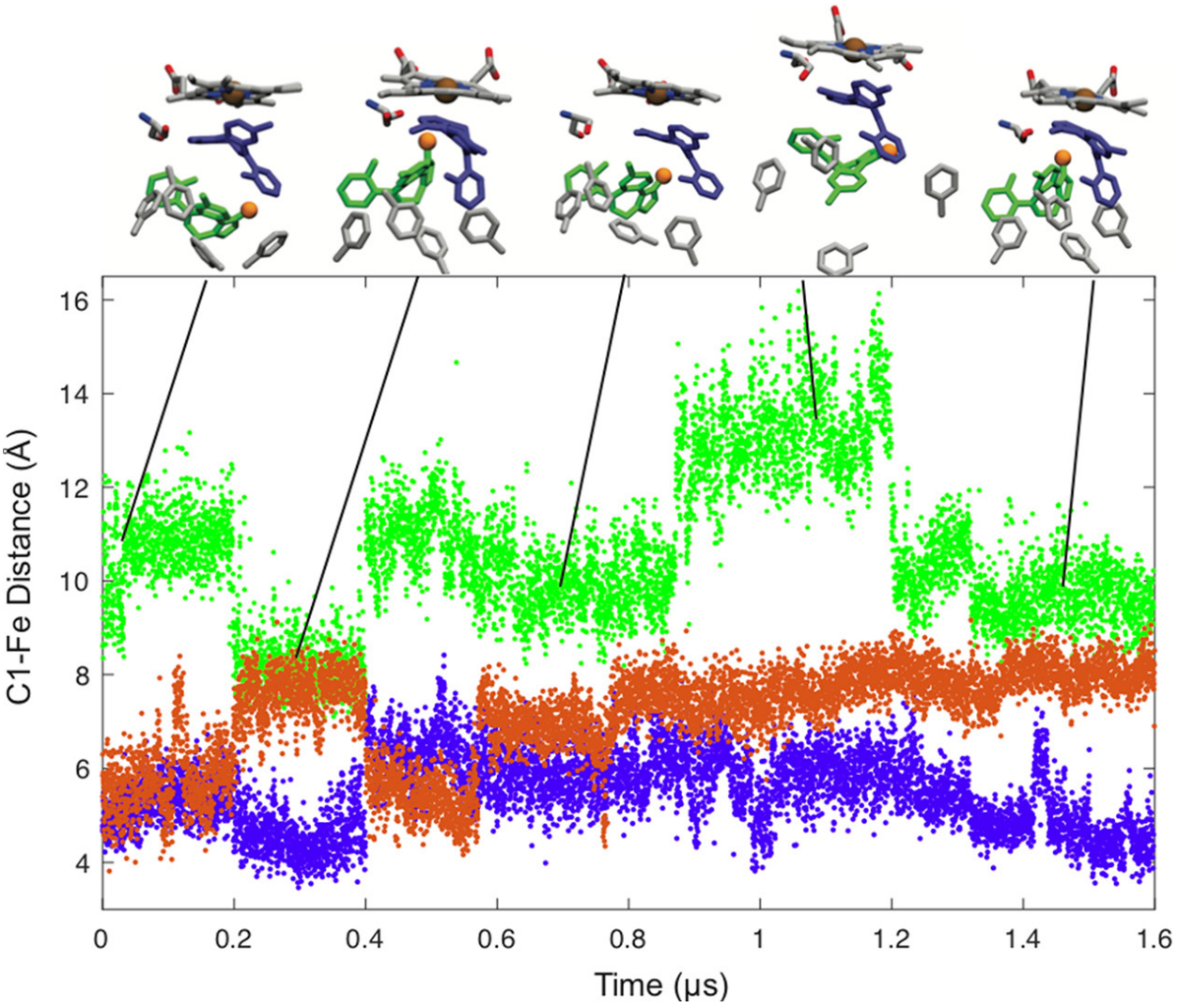


Figure 2.6. Distances between MDZ C1' and the heme iron atom in the simulation. Simulations with one MDZ (orange) and the simulation with an MDZ near the heme (purple) and in the allosteric site (green). Snapshots of the two-MDZ simulations at 0.08, 0.30, 0.66, 1.1, and 1.5 us are the inset. The positions of Ser119 and residues of the Phe-cluster (213, 215, 241, 304) are shown with the position of the C1 in the allosteric site MDZ highlighted with an orange sphere.

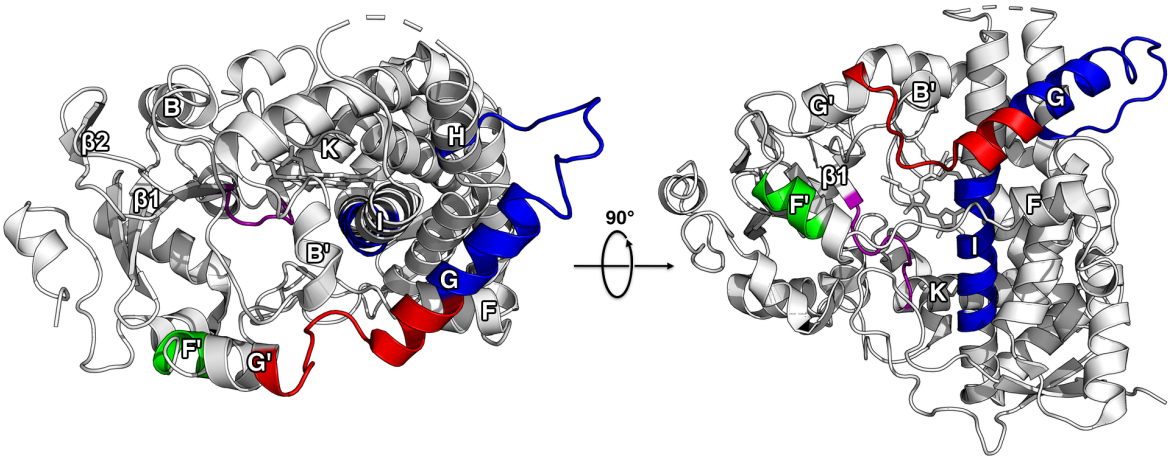


Figure 2.7. Comparison of HDX changes from MDZ and KTZ binding. The changes are mapped onto the ligand-free crystal structure of CYP3A4 (PDB ID: 1TQN). Regions where KTZ binding induces a decrease in HDX relative to ligand-free CYP3A4 are highlighted in blue, green, and purple. Regions where MDZ binding induces an increase in HDX relative to ligand-free CYP3A4 are highlighted in red and green, while regions where a decrease in HDX occurs are highlighted in purple.

2.6 Supporting figures

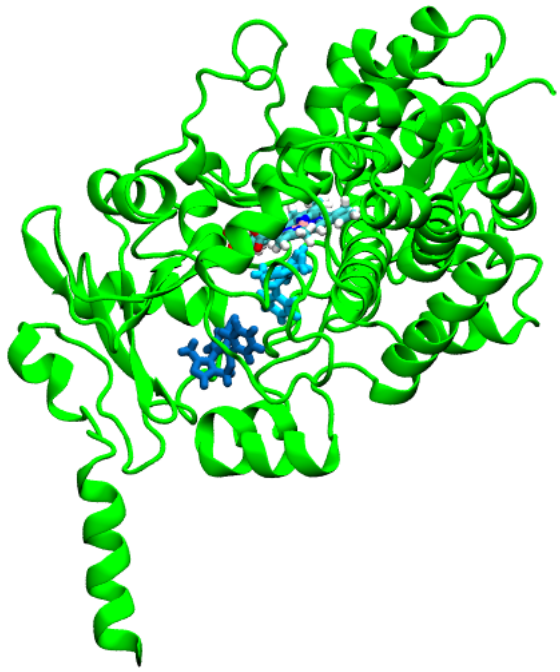


Figure S1. Initial position of the second MDZ (dark blue) for two-MDZ simulations.

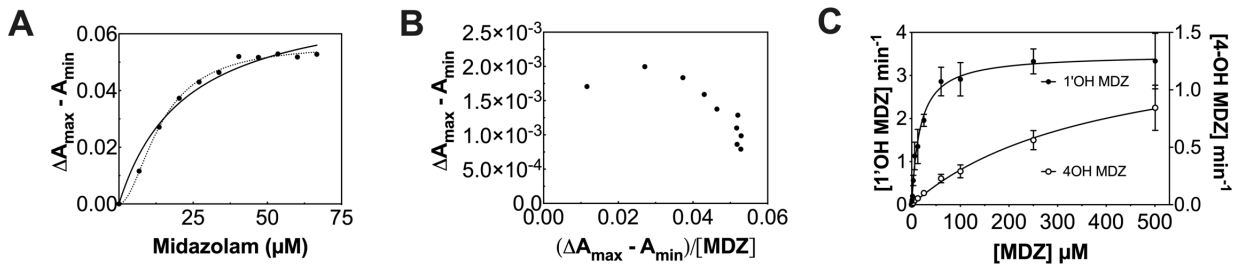
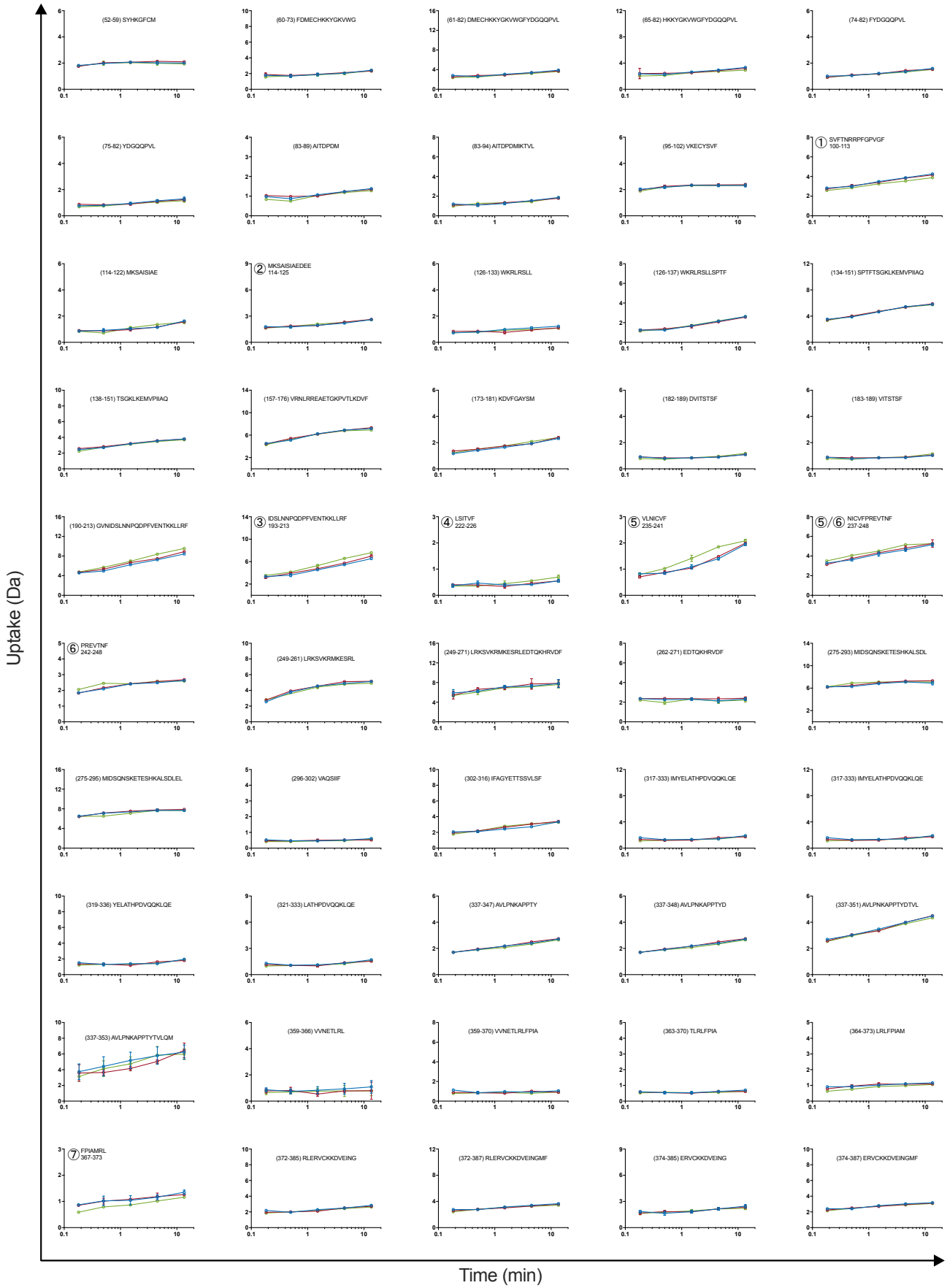


Figure S2. Spectral data fit to Michaelis-Menten and Eadie-Hofstee models for MDZ binding, and catalytic turnover data fit to Michalis-Menten model. (**A**) The dashed line shows the fit of the binding data to the hill equation. The black line shows the fit to the Michaelis-Menten equation. (**B**) Eadie-Hofstee plot for MDZ binding to CYP3A4 ND. (**C**) Midazolam hydroxylation kinetics. Product formation (nmol product / nmole CYP3A4 / min, or min^{-1}) are plotted vs MDZ concentration (μM). The closed and open circles represent 1'OH MDZ and 4OH-MDZ product formation, respectively.


(52-59) SYHKGFCM
(60-73) FDMECHKKYGKVWG
(61-82) DMECHKKYGKVWGFYDGQQPVL
(65-82) HKKYGKVWGFYDGQQPVL
(74-82) FYDGQQPVL
(75-82) YDGQQPVL
(83-89) AITDPDM
(83-94) AITDPDMIKTVL
(95-102) VKECYSVF
① SVFTNRRPFGPVGF 100-113
(114-122) MKSAISIAE
② MKSAISIAEDEE 114-125
(126-133) WKRLRSLL
(126-137) WKRLRSLLSPTF
(134-151) SPTFTSGKLKEMVPIIAQ
(138-151) TSGKLKEMVPIIAQ
(157-176) VRNLRREAETGKPVTLKDVF
(173-181) KDVFGAYSM
(182-189) DVITSTSF
(183-189) VITSTSF
(190-213) GVNIDSLNNPQDPFVENTKKLLRF
③ IDSLNNPQDPFVENTKKLLRF 193-213
④ LSITVF 222-226
⑤ VLNICVF 235-241
⑤/⑥ NICVFPREVTNF 237-248
⑥ PREVTNF 242-248
(249-261) LRKSVKRMKESRL
(249-271) LRKSVKRMKESRLEDTQKHRVDF
(262-271) EDTQKHRVDF
(275-293) MIDSQNSKETESHKALSDL
(275-295) MIDSQNSKETESHKALSDLEL
(296-302) VAQSIIF
(302-316) IFAGYETTSSVLSF
(317-333) IMYELATHPDVQQKLQE
(317-333) IMYELATHPDVQQKLQE
(319-336) YELATHPDVQQKLQE
(321-333) LATHPDVQQKLQE
(337-347) AVLPNKAPPTY
(337-348) AVLPNKAPPTYD
(337-351) AVLPNKAPPTYDTVL
(337-353) AVLPNKAPPTYTVLQM
(359-366) VVNETLRL
(359-370) VVNETLRLFPIA
(363-370) TLRLFPIA
(364-373) LRLFPIAM
⑦ FPIAMRL 367-373
(372-385) RLERVCKKDVEING
(372-387) RLERVCKKDVEINGMF
(374-385) ERVCKKDVEING
(374-387) ERVCKKDVEINGMF
Uptake (Da)
Time (min)

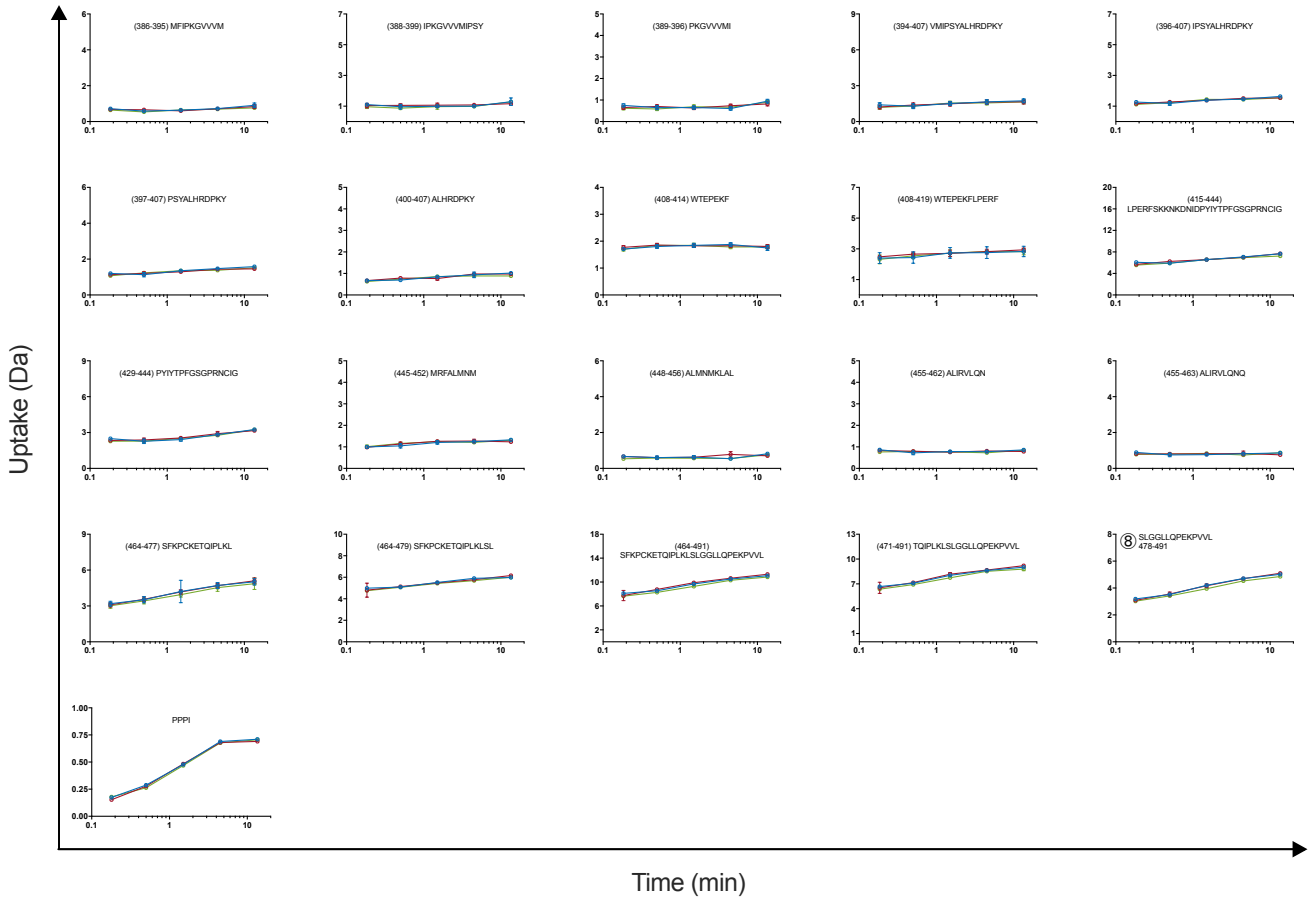


Figure S3. Deuterium uptake profiles for all peptides. Deuterium uptake (Da) vs Time (min) plots for each peptide. Exchange for the 0 μM, 6 μM and 60 μM MDZ samples is shown in blue, red, and green open circles, respectively. The error bars represent the standard deviation from duplicate experiments.

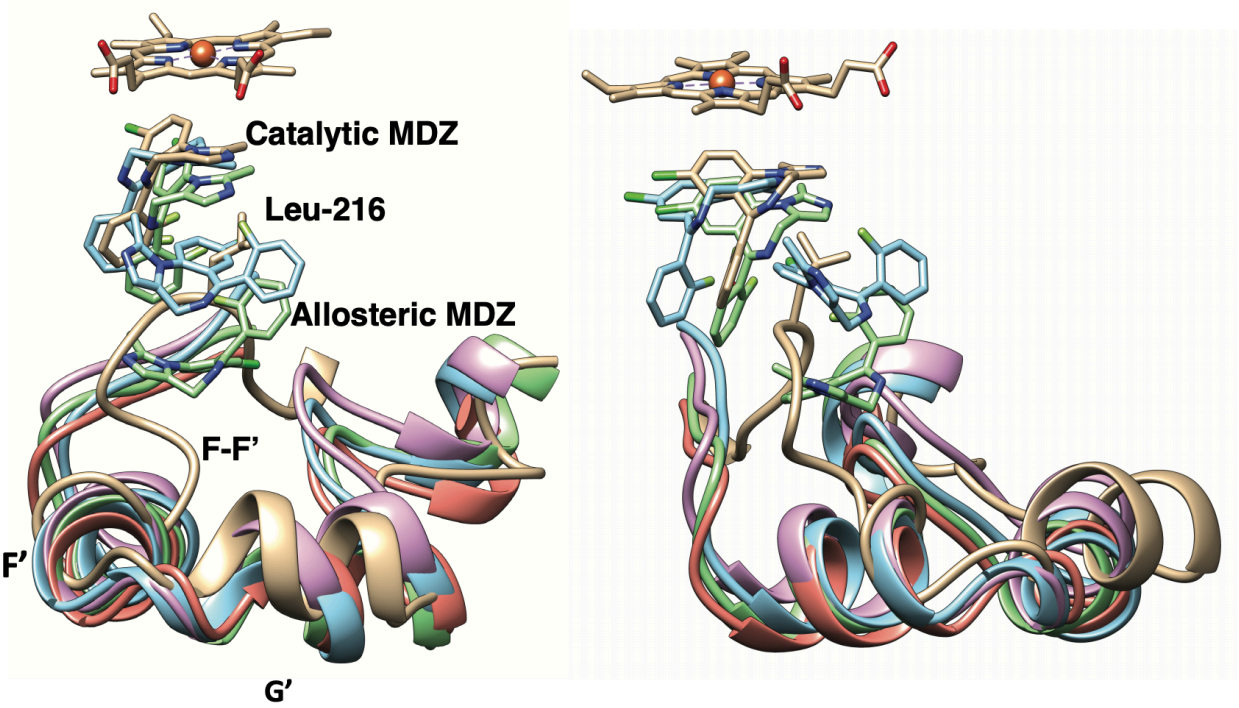


Figure S4. Overlay of 5TE8 crystal structure with one MDZ bound and snapshots from the MD with two MDZ's bound. Two views depicting comparisons of the F-G regions in the 5TE8 structure (tan) with the endpoints of the four GaMD simulations (blue, green, pink and orange) with two MDZ molecules bound. Representative MDZ positions form two of the four GaMD snapshots are depicted for clarity. The F-F' loop of the 5TE8 structure, specifically in the vicinity of Leu-216, occupies the same space as the predicted MDZ allosteric site. Binding of the allosteric MDZ is prohibited in the 5TE8 structure due to steric clash with the F-F' loop, but allowed upon addition of two MDZ's to the ligand-free structure.

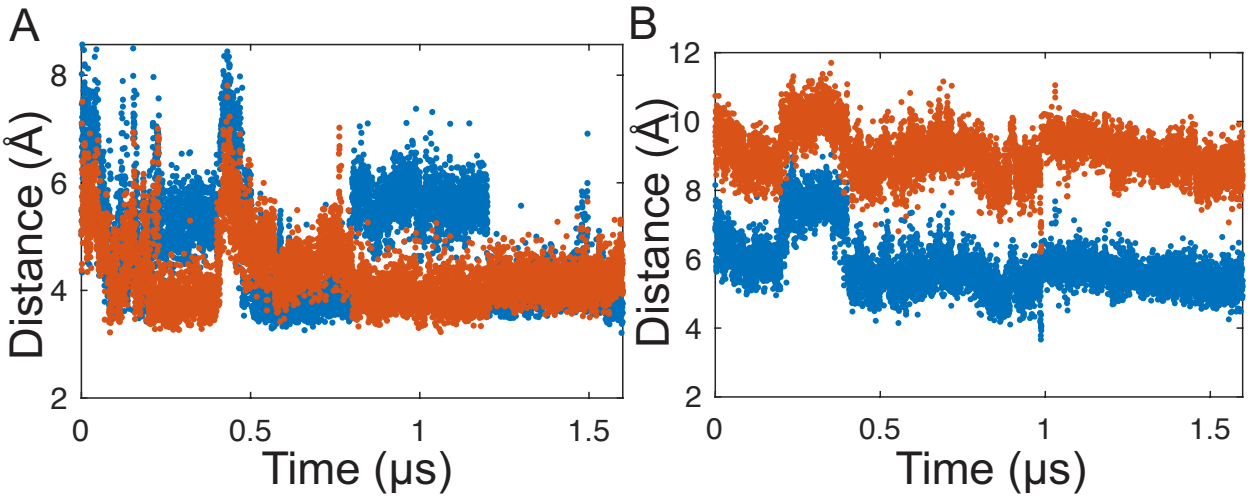


Figure S5. Distances between the Arg-105 Cζ and the MDZ1 N5 (blue) and center of the MDZ1 fluorophenyl ring (orange).

Chapter 3

Functional Implications of Complex Ligand Binding Kinetics and Protein Conformational Dynamics: Consideration of Promiscuous Drug Metabolizing Enzymes

3.1 Introduction

The interplay between ligand binding and conformational dynamics is of considerable interest to enzymologists. In principle, characterization of the thermodynamic and kinetic properties of an enzyme-ligand interaction, namely the mechanism of ligand binding, may provide details about the relative amounts of the states that comprise the ligand-free and the ligand-bound ensemble (and also the kinetics). Moreover, the ligand binding mechanism has functional significance, particularly when the members of the ligand-free ensemble have different binding properties or when members of the ligand-bound ensemble have different catalytic properties. Finally, the differential conformational dynamics in the ligand free vs ligand bound states provides mechanisms for achieving substrate specificity vs. promiscuity [3,4].

Drug metabolizing enzymes such as Cytochrome P450s (CYPs), which are the subject of this thesis, display remarkable differences in specificity vs promiscuity for closely related homologs [3,4,27]. In addition, their propensity for allosterism and their ability to form multiple products from a single substrate are well-documented [5-7,121,122]. Simultaneous binding of multiple ligands, multiple ligand binding modes, and/or CYP conformational dynamics are potential contributing factors to the observed atypical kinetics [8]. As a result of these considerations, characterization of the linkage between ligand binding and conformational change have practical and theoretical consequences.

In practice, assignment of a ligand binding model based on kinetic data is seldom straightforward. As a result, oversimplified models are often used, and they do not capture the complexity of ligand and protein dynamics. These oversimplified models include the case wherein ligand binds to a single conformation that is retained in the bound state. This two-state minimal model can easily be described by a single exponential process. However, most binding data require multiexponential models to obtain good fits, which requires a minimum of three enzyme states. Moreover, the application of more sophisticated models to ligand binding data, for CYPs in particular, is rare. Because drug metabolizing enzymes differ from traditional substrate specific enzymes, and are much more complex, there is a particular need to develop more sophisticated ligand models and methods for interrogating them.

In order to develop tools to understand the conformational dynamics of enzymes at a more realistic level of detail than previously considered, an in-depth analysis of several multi-state ligand binding models is presented in this chapter. Specifically, five multi-state ligand binding models were selected because each potentially results in double exponential binding kinetics (which are commonly observed in CYP ligand binding data) and because some of the models have been used to describe the observed atypical kinetics in CYPs. The five ligand models include: induced fit, conformational selection, sequential binding, substrate reorientation, and persistent heterogeneity.

Previous analyses have nicely described the analytical expressions for the relationships between the rates of change in protein bound species upon addition of ligand as the system relaxes to a new equilibrium for some of the models presented here [123-126]. In addition, more recent analyses have clarified analytical aspects of experimental data to distinguish the induced fit and conformational selection models [127-130]. Despite these insightful analyses, the functional

implications of different kinetic schemes have not been considered exhaustively in all cases. In particular, these analyses have been limited to ligand specific proteins and substrate specific enzymes, but many are highly ligand promiscuous and their function depends on their ability to interact with a wide range of substrates or ligands. Here, conclusions from the previous analyses are re-iterated for easy reference in addition to some new insights that have not been fully appreciated with regard to the kinetic analyses and methodological approach. Specifically, amplitude analysis, which is often ignored, in addition to the more commonly used observed rate analysis is detailed here. In addition, a new experimental approach which entails diluting bound enzyme-ligand complex is described. Briefly, the experiment takes advantage of the fact that the observed rates of change are independent of the direction of change of the relaxation process, thus widening the experimental concentration range to lower ligand concentration ranges to increase the likelihood of detecting observable kinetics when adequate characterization of kinetic behavior is not possible at higher ligand concentrations. Finally, a discussion of the ligand binding models and their potential implications for drug metabolism in CYPs is presented.

3.1.1 Underlying goal

The goal of this chapter is to provide an intuitive and methodological framework to aid in definitive assignment of a ligand binding model to kinetic data where double-exponential kinetics are observed. A double exponential process may be described by the expression below:

$$f(t) = A_0 + A_1 e^{-k_{obs,1}t} + A_2 e^{-k_{obs,2}t} \quad (3.1)$$

where the change is a function of time, t, A_n are the pre-exponential factors or amplitudes that describe the magnitude of change, $k_{obs,n}$ are the observed rates of change, and the first term, A_0, is an offset that may represent either the initial or final signal, depending on the experimental setup.

Traditionally, experimental data wherein a fixed amount of enzyme is mixed with varying amounts of ligand are fit to equations such as equation 3.1. Assignment of a kinetic model is usually selected on the basis of the behavior of the ligand concentration dependent behavior of the k_{obs}, often with no regard to the amplitude contributions. Unfortunately, this approach is not applicable to all situations and there exist several cases where multiple ligand binding models adequately describe the k_{obs} behavior [131]. Herein, a more rigorous analytical approach is presented where the model specific expressions will be examined. The expressions take the form of equation 3.1 but contain unique and model specific information about the observed rates *and* amplitudes.

The observation of multi-exponential kinetics requires careful examination of multi-state binding models. The theoretical maximum number of observable exponential processes for a given model is equal to the number of enzyme states minus one. Emphasis is placed on the number of *possible* observable processes. To illustrate this point, cases where multi-state binding models yield kinetics with less than the theoretical maximum number of exponential processes will be presented. In addition, the magnitude of change (i.e. the pre-exponential factors), which is an often-overlooked parameter in traditional kinetic analyses, will be emphasized in the forthcoming sections. Specifically, the utility of amplitude analysis will be presented for cases that are otherwise indistinguishable on the basis of k_{obs} behavior alone. Finally, experimental considerations will be presented for cases where k_{obs} and amplitude kinetic signatures are not apparent in the observable ligand concentration range. Specifically, a dilution or dissociation approach will be presented, which serves as a means of signal amplification for otherwise unobservable (due to low signal to noise) low ligand concentration ranges in the comparable mixing or binding experimental approach, thus widening the range for kinetic analyses. Here,

various three-state binding models will be presented with emphasis on their unique kinetic and thermodynamic signatures with the goal of differentiating the models.

The focus of the following introductory sections is to provide a quantitative and intuitive understanding of enzyme kinetics. The expressions describing the equilibrium amounts of each enzyme species in each 2-state model and the observed rates of change have been discussed extensively and are easily solved for by algebraic methods. However, the derivation for the more complex 3-state systems are not as straightforward by similar methods. Thus, techniques of linear algebra are used to solve the system of differential equations. The techniques are not new; in fact, the processes may be reviewed in any introductory linear algebra textbook. In addition, several others have detailed similar derivations, however, the goal of this introductory section is to present a more detailed and explicit presentation of the theory for experimental considerations. To that end, full solutions will be presented in a stepwise fashion, in an effort to aid the mathematical novice. With mathematical platforms such as MATLAB, Mathematica, and Maple, computing the full solutions for the system of differential equations is simple. The process is detailed in the following sections.

3.1.2 Nomenclature

Two classes of enzyme systems will be discussed in this chapter: 2- and 3-state mechanisms or models, wherein the number of states refers to the number of distinct enzyme states or species in each model. While the 3-state systems are the focus of this chapter, it is worthwhile to discuss the simpler 2-state systems as different combinations of the 2-state systems constitute the 3-state systems. Thus, the 2-state systems (Figure 3.1) are presented before the more complex 3-state systems (Figure 3.2). A general system will first be described (Figure 3.1A and Figure

3.2A) for the two classes of enzyme systems. The general cases provide no physical descriptions for the identity of the enzyme states or for the individual rate constants involved in the transitions between the enzyme states. As will become apparent in the remaining sections, the general case expressions may be easily converted to the more descriptive 2- and 3-state models, thereby preventing lengthy redundancy in the introductory sections. Regardless, all 2-state models will be presented in the forthcoming introductory sections because each 3-state system contains different combinations of the 2-state binding or isomerization steps.

The general case nomenclature is universal. In the general cases, the different enzyme states are numbered in the order of the transitions (i.e. from left to right as seen in Figures 3.1 and 3.2). For example, the general 2-state model (E1 ⇄ E2, Figure 3.1A) describes the transition between two enzyme states, E1 and E2. The general 3-state model (E1 ⇄ E2 ⇄ E3, Figure 3.2A) describes the transitions between three enzyme states E1, E2, and E3, wherein the transition from E1 to E3 proceeds through an intermediate state, E2, and vice versa. The individual rate constants are shown above the associated arrows in Figures 3.1A and 3.2A. The different rate constants, k_{if}, contain subscripts that inform the reader of the initial, i, and final, f, enzyme states for the transition. For example, k_{12} refers to the rate constant describing the transition from state E1 to E2, while k_{23} refers to the transition from state E2 to E3. Importantly, the general case nomenclature is sometimes used to lessen confusion, particularly when comparing different binding models which may have different physical descriptions for each enzyme state and kinetic process.

The ligand binding models (Figures 3.1B – 3.1C and Figures 3.2B – 3.2F) that will be discussed are simply descriptive variants of the analogous general case wherein the general case terminology has been replaced by their more descriptive counterparts. Individual enzyme states in the binding models differ in terms of the enzyme conformation, ligation state, and/or binding

orientation, and are renamed accordingly. In addition, the rate constants for each model have been renamed to reflect the type of transition, relating to either a ligand binding or an isomerization event. The analogous enzyme states and rate constants are presented directly below the related general case. For example, in the 2-state isomerization model (Figure 3.1B), the general case enzyme states, E1 and E2, have been replaced by the E and E* states, respectively. Likewise, the general case rate constants, k_{12} and k_{21}, have been replaced by the forward (represented by the right facing arrow) and the reverse isomerization rates (left facing arrow), k_f and k_r, respectively. Similar substitutions may be applied to the remaining models and examples are presented in the sections that follow. However, one model is an exception: the persistent heterogeneity model in Figure 3.2F describes two unlinked 2-state binding models, whose general case is the 2-state model in Figure 3.1A.

The physical differences in the enzyme states have different textual representations. An asterisk directly following the enzyme state (e.g. E* or E*L) represents a conformationally distinct enzyme state compared to the asterisk free counterpart (e.g. E or EL). An asterisk directly following the ligand (e.g. EL*) represents a distinct binding orientation compared to the asterisk free counterpart (e.g. EL). A bound complex is represented by an L directly following the enzyme state where the number of L correlates to the number of bound ligand molecules (e.g. EL and E*L represent singly ligated complexes while the ELL complex represents a doubly ligated complex). When multiple ligand binding steps are involved, the related on- and off-rates for each binding event are differentiated by a 1 or 2 in the subscript (e.g. k_{on1} represents the left most binding step in the model while k_{on2} represents the second or right most binding step). Visual representations of each model are provided in Figures 3.1 and 3.2 for each model to lessen confusion.

3.1.3 Kinetic model overview

Two physical scenarios of the 2-state system will be presented: conformational isomerization and ligand binding. The conformational isomerization model envisions a flexible enzyme that is capable of sampling distinct enzyme conformations, E and E* (Figure 3.1B). The isomerization from the E to the E* enzyme state is governed by the forward and reverse isomerization rates, k_f and k_r, respectively. The binding model envisions ligand, L, reversibly binding to enzyme, E, to form a complex, EL (Figure 3.1C). The ligand binding event resulting in the transition from the E to EL enzyme states is governed by the ligand on- and off-rate rates, k_{on} and k_{off}, respectively. The 2-state binding models will be discussed in detail in the upcoming sections to serve as background to better understand the behavior of the more complicated 3-state binding models.

Five binding models will be discussed in this chapter and they contain different combinations of the two 2-state scenarios (Figure 2B – 2F). In addition, each model has the potential of displaying double-exponential or biphasic re-equilibration kinetics. Among the 3-state binding models, two models contain both a ligand binding and isomerization event: induced fit (IF) and conformational selection (CS). The IF and CS binding models differ in the order of the binding and isomerization events. The remaining 3-state binding models contain two ligand binding steps: sequential binding (SB) and substrate reorientation (SR). The SB and SR binding models differ in their binding stoichiometry. Finally, a 4-state model is also presented, which contains two unlinked binding events, wherein two non-interconverting and distinct enzyme states bind to ligand: persistent heterogeneity (PH). Notably, the PH model contains two unlinked 2-state binding models (i.e. the general 2-state case is applicable to PH model).

3.2 Model derivation

The next few sections are written in a linear fashion. Model specific concepts such as binding affinity and experimental considerations are presented where applicable so as not to overwhelm the reader in the general case descriptions.

3.2.1 The general 2-state system: E1 ⇄ E2

Before attempting to explain the binding kinetics for the more complex multi-step binding models, it is best to start with a description of the simplest binding model, which is that of a two-state system (Figure 3.1A). The system is said to be two-state because there are two enzyme states, E1 and E2, which are in equilibrium. The E1 enzyme state may be reversibly converted to the E2 enzyme state, and vice versa. The two rate constants, k_{12} and k_{21}, represent the forward and reverse reactions, respectively. Perturbation of the equilibrium by an external factor (e.g. change in pH, temperature, pressure, addition of ligand etc.) results in a re-adjustment of the system to a new equilibrium based on the new conditions. Below, the kinetic and thermodynamic relationships for the general 2-state model will be presented.

The rate of change for each enzyme state under any set of conditions are defined by the differential equations below:

$$\frac{d[E1]}{dt} = -k_{12}[E1] + k_{21}[E2] \tag{3.2}$$

$$\frac{d[E2]}{dt} = +k_{12}[E1] - k_{21}[E2] \tag{3.3}$$

An equilibrium is established when the forward and reverse reactions are equal. In addition, the concentrations of both enzyme species are constant at equilibrium (i.e. d[E1]/dt = d[E2]/dt = 0).

Thus, the equilibrium constant, K_{eq}, may be derived by setting either differential equation equal to zero. The K_{eq}, which is a thermodynamic constant, is defined by the expression below:

$$K_{eq} = \frac{k_{12}}{k_{21}} = \frac{[E2_{eq}]}{[E1_{eq}]} \tag{3.4}$$

The K_{eq}, may be expressed as a ratio of the kinetic constants, k_{12} and k_{21}, or as a ratio of the equilibrium concentrations of both enzyme states, $[E1_{eq}]$ and $[E2_{eq}]$. Hence there is a relationship between the kinetic and thermodynamic properties of the system. The K_{eq} may be directly measured provided reliable measurements of the equilibrium enzyme concentrations are possible. However, quantification of the individual rate constants requires measurement of the re-equilibration kinetics. Hence, full characterization of the system requires both thermodynamic and kinetic assessment.

The system is defined in matrix notation below:

$$\begin{bmatrix} d[E1]/dt \\ d[E2]/dt \end{bmatrix} = \begin{bmatrix} -k_{12} & k_{21} \\ +k_{12} & -k_{21} \end{bmatrix} \begin{bmatrix} [E1] \\ [E2] \end{bmatrix} \tag{3.5}$$

and in shortened form:

$$\mathbf{\frac{dE}{dt} = KE} \tag{3.6}$$

where **dE/dt** is the 2x1 column vector of the time derivatives, **K** is the 2x2 coefficient matrix, and **E** the 2x1 column vector of enzyme concentrations in equation 3.5. The solution of the system may be found by integrating equation 3.6:

$$\int_{\mathbf{E_0}}^{\mathbf{E}} \mathbf{\frac{dE}{E}} = \int_0^t \mathbf{K dt} \tag{3.7}$$

$$ln\left(\mathbf{\frac{E}{E_0}}\right) = \mathbf{K}t \tag{3.8}$$

$$\mathbf{E} = \mathbf{E_0} e^{\mathbf{K}t} \tag{3.9}$$

where $\mathbf{E_0}$ is a 2x1 column vector of initial enzyme concentrations ([$E1_0$, $E2_0$]), **E** is the same 2x1 column vector in equation 3.5 ([E1, E2]), and **K** is the same 2x2 matrix from equation 3.5. The solution of equation 3.7 normally requires solving for the eigenvalues and eigenvectors of matrix **K**. Because **K** is a diagonizable matrix, the solution is given below:

$$\mathbf{E} = \mathbf{V}\begin{bmatrix} e^{\lambda_0 t} & 0 \\ 0 & e^{\lambda_1 t} \end{bmatrix}\mathbf{V^{-1}E_0} \tag{3.10}$$

where **V** is a 2x2 matrix whose columns are the eigenvectors of matrix **K,** λ_n are eigenvalues of the matrix **K**, and $\mathbf{V^{-1}}$ is the inverse of matrix **V**. Stated differently, perturbation of the equilibria results in a relaxation process at a rate described by the eigenvalue expressions. The magnitude of change is related to the eigenvectors.

Here, the eigenvalues are solved for. The eigenvalues, λ, satisfy the characteristic equation of matrix **K** shown below:

$$\det(\mathbf{K} - \lambda\mathbf{I}) = 0 \tag{3.11}$$

where the determinant of (K- λ**I**) is set to 0, and **I** is the identity matrix with the same dimensions as matrix **K**. The equation yields the characteristic polynomial of matrix **K**, whose roots are equal to the eigenvalues, as shown below:

$$\det\left(\begin{bmatrix} -k_{12} - \lambda & k_{21} \\ +k_{12} & -k_{21} - \lambda \end{bmatrix}\right) = 0 \tag{3.12}$$

$$(-k_{12} - \lambda)(-k_{21} - \lambda) - k_{12}k_{21} = 0 \tag{3.13}$$

$$\lambda^2 + (k_{12} + k_{21})\lambda = 0 \tag{3.14}$$

$$\lambda_0 = 0, \quad \lambda_1 = -k_{12} - k_{21} = -k_{obs} \tag{3.15}$$

The roots, λ_n, are the eigenvalues of the matrix **K** [132]. Perturbation of the system will result in single exponential kinetics (one non-zero eigenvalue) with an observed rate equal to k_{obs}. Intuitively, this means that depletion of E1 to form E2 occurs at the same rate, which reflects the mass conservation in the system. In this particular system, the observed rate is constant and contains the forward and reverse rates constants additively. This is because as the system approaches equilibrium, the forward and reverse processes are competing for the equilibration.

At this point, the binding kinetics may be described by the single exponential equation shown below:

$$f(t) \ = A_0 + A_1 e^{-k_{obs}t} \tag{3.16}$$

That is, one need not solve for the pre-exponential factors to recover the kinetic parameters. In fact, most often equation 3.16 is used with 'placeholder' amplitudes that have no physical linkage to a model. For simple systems, this is okay. However, as models become more complicated, it becomes necessary to understand the nature of the pre-exponential factors. To solve for the amplitude factors, or the so-called 'normal modes' of the reaction coordinate, one must obtain solutions for the eigenvectors matrix **K**. Methods to obtain the eigenvectors have been presented by many [123-125,133-136]. The process can become quite cumbersome, particularly when larger systems are involved (such as the 3x3 systems in this chapter). Therefore, an alternative analytical approach is described herein, which expresses the amplitude factors in terms of the eigenvalues.

Sylvester's theorem provides a rapid way to solve functions of a matrix (such as $e^{\mathbf{K}t}$) in terms of the eigenvalues (which are easy to solve for) without the requirement of solving for the eigenvectors (which are less easy to solve for) [137]. The exponential function in equation 3.9 is rewritten using Sylvester's theorem below:

$$e^{\mathbf{K}t} = \sum_{i=0}^{n} \mathbf{Q}_i e^{\lambda_i t} \tag{3.17}$$

where n is the number of enzyme states (n=2 here) and **Q** are the Frobenius covariants of matrix **K**. The Frobenius covariants are defined as follows:

$$\mathbf{Q}_i = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{(\mathbf{K} - \lambda_j \boldsymbol{I})}{(\lambda_i - \lambda_j)} \tag{3.18}$$

where **K** is the 2x2 coefficient matrix, **I** is the identity matrix with the same dimensions as **K**, and λ are the eigenvalues of the matrix **K**, for n = 2 states. The covariants for the general 2-state case are written below:

$$\mathbf{Q}_0 = \frac{(\mathbf{K} - \lambda_1 \boldsymbol{I})}{-\lambda_1} = \begin{bmatrix} \dfrac{k_{21}}{k_{12}+k_{21}} & \dfrac{k_{21}}{k_{12}+k_{21}} \\ \dfrac{k_{12}}{k_{12}+k_{21}} & \dfrac{k_{12}}{k_{12}+k_{21}} \end{bmatrix} \tag{3.19}$$

$$\mathbf{Q}_1 = \frac{(\mathbf{K})}{\lambda_1} = \begin{bmatrix} \dfrac{k_{12}}{k_{12}+k_{21}} & \dfrac{k_{21}}{-k_{12}-k_{21}} \\ \dfrac{k_{12}}{-k_{12}-k_{21}} & \dfrac{k_{21}}{k_{12}+k_{21}} \end{bmatrix} \tag{3.20}$$

where $\mathbf{Q_i}$ have the same dimensions as **K**. Thus, the solution in equation 3.9 is further expanded below:

$$\mathbf{E} = \mathbf{Q_0 E_0} e^{\lambda_0 t} + \mathbf{Q_1 E_0} e^{\lambda_1 t} = \mathbf{Q_0 E_0} + \mathbf{Q_1 E_0} e^{\lambda_1 t} = \mathbf{Q_0 E_0} + \mathbf{Q_1 E_0} e^{-k_{obs} t} \tag{3.21}$$

where $\mathbf{Q_i}$ are defined by equations 3.19 – 3.20. Multiplication of matrices $\mathbf{Q_i}$ (2x2) and $\mathbf{E_0}$ (2x1) results in a 2x1 column vector, $\mathbf{A_i}$. The final result is two equations (one for each enzyme state, E1 and E2) of the form 3.16 with the defined amplitudes in the rows of $\mathbf{A_i}$ below:

$$\mathbf{A_0} = \mathbf{Q_0 E_0} = \begin{bmatrix} \frac{k_{21}}{k_{12}+k_{21}} & \frac{k_{21}}{k_{12}+k_{21}} \\ \frac{k_{12}}{k_{12}+k_{21}} & \frac{k_{12}}{k_{12}+k_{21}} \end{bmatrix} \begin{bmatrix} [E1_0] \\ [E2_0] \end{bmatrix} = \begin{bmatrix} \frac{k_{21}[E_{total}]}{k_{12}+k_{21}} \\ \frac{k_{12}[E_{total}]}{k_{12}+k_{21}} \end{bmatrix} \tag{3.22}$$

where $[E_{total}] = [E1_0] + [E2_0]$

$$\mathbf{A_1} = \mathbf{Q_1 E_0} = \begin{bmatrix} \frac{k_{12}}{k_{12}+k_{21}} & \frac{-k_{21}}{k_{12}+k_{21}} \\ \frac{-k_{12}}{k_{12}+k_{21}} & \frac{k_{21}}{k_{12}+k_{21}} \end{bmatrix} \begin{bmatrix} [E1_0] \\ [E2_0] \end{bmatrix} = \begin{bmatrix} \frac{+[E1_0]k_{12} - [E2_0]k_{21}}{k_{12}+k_{21}} \\ \frac{-[E1_0]k_{12} + [E2_0]k_{21}}{k_{12}+k_{21}} \end{bmatrix} \tag{3.23}$$

The $\mathbf{A_i}$ column vectors contain the pre-exponential factors or amplitudes of equation 3.21 (and 3.16). Examination of equation 3.21 with the k_{obs} (defined in equation 3.15) and the amplitudes (defined by equations 3.22 – 3.23) reveal that the first set of terms ($\mathbf{Q_0 E_0} = \mathbf{A_0}$) are constant while the second set of terms ($\mathbf{Q_1 E_0} e^{-kobs*t} = \mathbf{A_1} e^{-kobs*t}$) range from A_1 (at t = 0) to 0 (at t = ∞). That is, the first step of terms describes the final equilibrium amounts of the corresponding enzyme state (expressed as a fraction of $[E_{total}]$; i.e. $\mathbf{A_0} = [\mathbf{E_{eq}}]$, where $[\mathbf{E_{eq}}]$ is a column vector defining the equilibrium concentrations, $\mathbf{E_{eq}} = [E1_{eq}, E2_{eq}]$) while the second set of terms describe the temporal change over time. The amplitude in the second set of terms, $\mathbf{A_1}$, is a column vector defining the magnitude of change where $\mathbf{A_1} = [A_{E1}, A_{E2}]$).

The full solutions for the 2-state system in Figure 3.1A are written below:

$$\mathbf{E(t)} = \mathbf{A_0} + \mathbf{A_1} e^{\lambda_1 t} = \left[\mathbf{E_{eq}}\right] + \mathbf{A_1} e^{\lambda_1 t} \tag{3.24}$$

$$[E1](t) = [E1_{eq}] + (A_{E1}) e^{-k_{obs} t} \tag{3.25}$$

where $[E1_{eq}] = \frac{k_{21}}{k_{12}+k_{21}} [E_{total}]$

$$A_{E1} = \frac{[E1_0]k_{12} - [E2_0]k_{21}}{k_{12}+k_{21}}$$

$$k_{obs} = k_{12} + k_{21}$$

$$[E2](t) = [E2_{eq}] + (A_{E2})e^{-k_{obs}t} \tag{3.26}$$

where $[E_{eq}] = \frac{k_{12}}{k_{12}+k_{21}}[E_{total}]$

$$A_{E2} = \frac{-[E1_0]k_{12} + [E2_0]k_{21}}{k_{12} + k_{21}}$$

$$k_{obs} = k_{12} + k_{21}$$

where the transient enzyme concentrations, [**E**], are a function of time, t, which vary exponentially with rate, $k_{obs,}$ and amplitudes, A_i. The solutions contain thermodynamic (equilibrium concentration) and kinetic (observed rate of change) information. The first term describes the final equilibrium amounts of each species ($[E1_{eq}]$ and $[E2_{eq}]$) expressed as a ratio of total enzyme ($[E_{total}]$ = [E1] + [E2]). The second set of terms describe the temporal change in the respective enzyme states with magnitude, A_i, and rate, k_{obs}. Three observations may be pointed out here that reflect the reversibility and conservation of mass in the system: **1**) the k_{obs} expressions are identical for each enzyme state, **2**) the sum of the equilibrium amplitudes is equal to the total enzyme concentration, $[E1_{eq}] + [E2_{eq}] = [E_{total}]$, and **3**) the sum of the pre-exponential factors equal zero, $A_{E1} + A_{E2} = 0$. That is, the formation of E2 resulting from the depletion of E1 happens at rate, k_{obs}, with equal and opposite magnitude of state E1. The last statement is true for the remaining 2-state models.

3.2.2 The 2-state isomerization model: E ⇄ E*

Enzymes sample distinct conformations over a range of time scales and it is widely appreciated that ligand binding directly influences the conformational landscape. Two of the ligand binding models discussed in this chapter involve an isomerization step (induced fit and conformational selection in Figures 3.2B and 3.2C, respectively). Therefore, it is useful to discuss

the simplest model containing an isomerization step to aid in understanding the more complex ligand binding models which include both ligand binding and isomerization steps. In this system, the enzyme, E, can reversibly isomerize to populate the E* state (Figure 3.1B). The forward and reverse rates for the isomerization or conformational change are defined as k_f and k_r, respectively. Both rate constants have units of time^{-1}.

The differential equations that describe the model are shown below:

$$\frac{d[E]}{dt} = -k_f[E] + k_r[E^*] \tag{3.27}$$

$$\frac{d[E^*]}{dt} = +k_f[E] - k_r[E^*] \tag{3.28}$$

The differential equations are re-written in matrix form below:

$$\begin{bmatrix} d[E]/dt \\ d[E^*]/dt \end{bmatrix} = \begin{bmatrix} -k_f & k_r \\ +k_f & -k_r \end{bmatrix} \begin{bmatrix} [E] \\ [E^*] \end{bmatrix} \tag{3.29}$$

Comparison of equations 3.27 – 3.29 reveal that the solutions derived for the general 2-state scenario may be converted to the current 2-state isomerization model with the following substitutions: $k_f = k_{12}$, $k_r = k_{21}$, and E = E1, E* = E2 for all associated enzyme state constants (e.g. $A_E = A_{E1}$, $A_{E^*} = A_{E2}$, $E_0 = E1_0$, $E^*_0 = E2_0$, etc.). That is, one may arrive at the expressions described for the general case for a model of interest, simply by substituting the general case expression enzyme states and rate constants with those of the model of interest. Hence, K_{eq} is defined by the expression below:

$$K_{eq} = \frac{k_f}{k_r} = \frac{[E^*_{eq}]}{[E_{eq}]} \tag{3.30}$$

The same relationship between the kinetic and thermodynamic constants for the isomerization model is established compared to the general case. As in the general case, the equilibrium constant may be directly measured when reliable measurements of the equilibrium concentrations of each

enzyme state are possible. In practice, it may be difficult to distinguish the two enzyme states, thus rendering direct measurement of K_{eq} impossible. However, perturbation of the system by addition of ligand may provide an indirect measure of the isomerization rate constants, which will be discussed in later sections (see the conformational selection model, Figure 3.2C). Regardless, kinetic experiments are required to quantify the individual rate constants (k_f and k_r).

The eigenvalues of the rate matrix in equation 3.29, are obtained via the characteristic equation defined in equation 3.11 in a similar fashion as for the general case. The resulting eigenvalues are calculated as follows:

$$det\left(\begin{bmatrix} -k_f - \lambda & k_r \\ +k_f & -k_r - \lambda \end{bmatrix}\right) = 0 \tag{3.31}$$

$$\lambda_0 = 0, \quad \lambda_1 = -k_f - k_r \; = \; -k_{obs} \tag{3.32}$$

Because the system is reversible, k_{obs} is identical for each enzyme state. Perturbation of the system by an external factor, such as a change in temperature or pH results in a re-equilibration process with kinetics governed by the microscopic rate constants, k_f and k_r. For this system, the re-equilibration process proceeds at the same rate, regardless of the initial enzyme conditions ($k_{obs} = k_f + k_r$; i.e. the k_{obs} expression does not contain any terms for the enzyme concentrations). Since k_{obs} is equal to the sum of both rate constants, quantification of the individual rate constants requires thermodynamic (K_{eq}) and kinetic (k_{obs}) measurements. Given measurements of K_{eq} and k_{obs}, one may arrive at the individual rate constant values using the expressions below:

$$K_{eq} = \frac{k_f}{k_r} = \frac{k_f}{k_{obs} - k_f} \tag{3.33}$$

Isolation of k_f from equation 3.33 yields:

$$k_f = \frac{K_{eq} k_{obs}}{1 + K_{eq}} \tag{3.34}$$

Substitution of equation 3.34 into the k_{obs} expression (equation 3.32) yields k_r:

$$k_r = k_{obs} - k_f = k_{obs} - \frac{K_{eq} k_{obs}}{1 + K_{eq}} \tag{3.35}$$

For completeness, the full solutions for the 2-state isomerization system are given below:

$$\mathbf{E(t) = A_0 + A_1 e^{\lambda_1 t} = [E_{eq}] + A_1 e^{\lambda_1 t}} \tag{3.36}$$

$$[E](t) = [E_{eq}] + (A_E)e^{-k_{obs}t} \tag{3.37}$$

where $[E_{eq}] = \frac{k_r}{k_f + k_r}[E_{total}]$

$$A_E = \frac{[E_0]k_f - [E^*{}_0]k_r}{k_f + k_r}$$

$$k_{obs} = k_f + k_r$$

$$[E^*](t) = [E^*{}_{eq}] + (A_{E^*})e^{-k_{obs}t} \tag{3.38}$$

where $[E^*{}_{eq}] = \frac{k_f}{k_f + k_r}[E_{total}]$

$$A_{E^*} = \frac{-[E_0]k_f + [E^*{}_0]k_r}{k_f + k_r}$$

$$k_{obs} = k_f + k_r$$

To clarify, the solutions were obtained by making the following substitutions into the general case equations 3.23 – 3.26: $k_f = k_{12}$, $k_r = k_{21}$, and E = E1, E* = E2 for all associated enzyme state constants (e.g. $A_E = A_{E1}$, $A_{E^*} = A_{E2}$, $E_0 = E1_0$, $E^*_0 = E2_0$, etc.). Notably, the solutions presented here are equivalent to an EL ⇄ E*L isomerization model with the following general case substitutions: EL = E1, E*L = E2. These details become important for descriptions of the limiting cases for the induced fit (Figure 3.2B) and conformational selection models (Figure 3.2C), which both contain isomerization events involving the bound (EL ⇄ E*L) or unbound (E ⇄ E*) enzyme

states, respectively. This is reinforced in section 3.2.5 for the induced fit mechanism. The derivation process is repeated for the 2-state ligand binding model in the next section.

3.2.3 The 2-state ligand binding model: E + L ⇄ EL

The 2-state ligand binding model is the simplest ligand binding model (Figure 3.1C). In most cases, binding is a reversible bimolecular process wherein ligand, L, binds to enzyme, E, to form a complex, EL. The intrinsic rate constants, k_{on} and k_{off}, represent the ligand on- and off-rates, respectively, which have units of concentration^{-1} time^{-1} and time^{-1}, respectively. For ligand binding processes, it is useful to define the equilibrium dissociation constant, K_D, instead of the equilibrium constant, K_{eq}. Like K_{eq}, the equilibrium dissociation constant or binding affinity, K_D, is a thermodynamic constant comprised of a ratio of kinetic constants, given by the equation below:

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[E][L]}{[EL]} \tag{3.39}$$

The dissociation constant has units of concentration and is equal to the concentration at which half of all binding sites are occupied ([EL] = [E] when ligand is in excess of enzyme or [EL] = [L] for the reverse condition). The 'strength' of the binding interaction is inversely proportional to the K_D. Importantly, different ligands may have identical binding affinities but very different kinetic properties. Thus, complete characterization of a binding interaction requires kinetic and thermodynamic assessment.

Traditional kinetic analyses rely on interpretation of the concentration dependent behavior of the observed rate of binding, k_{obs}, obtained by mixing varying amounts of ligand with a fixed amount of enzyme. Traditional thermodynamic analyses parallel the kinetic experiments but only the final equilibrium amounts of the bound complex, EL, or the disappearance in the unbound

complex, E, are measured when ligand concentration is varied (usually around the K_D). Here, the kinetic and thermodynamic relationships for the 2-state binding model are presented.

The system can be described by the differential equations below:

$$\frac{d[E]}{dt} = -k_{on}[E][L] + k_{off}[EL] \tag{3.40}$$

$$\frac{d[L]}{dt} = -k_{on}[E][L] + k_{off}[EL] \approx 0 \tag{3.41}$$

$$\frac{d[EL]}{dt} = +k_{on}[E][L] - k_{off}[EL] \tag{3.42}$$

In this work, ligand is in excess of total enzyme concentration, thus the change in ligand concentration is negligible. The system is said to be under pseudo-first order conditions provided [L] is at least 10x greater than [E] [138]. In addition, this allows the following substitutions into the general case expression: $k_{on}[L] = k_{12}$. The following substitutions for the 2-state binding model will be made to the general case expressions: $k_{on}[L] = k_{12}$, and $k_{off} = k_{21}$, and E = E1, EL = E2 for all associated enzyme state constants (e.g. $A_E = A_{E1}$, $A_{EL} = A_{E2}$, $E_0 = E1_0$, $EL_0 = E2_0$, etc.)

Here, a description of the relevant parameters from the derivation process of the two-state binding system is presented. Similarly, with the general two-state system in the forgoing example, the solutions take the form of equation 3.24. As will be described below, the change in k_{obs}, is independent of the initial amounts of each enzyme species and only depends on the final ligand concentration. In contrast, the pre-exponential factors, or amplitudes, are sensitive to the initial amounts of each enzyme species. These are concepts that are useful for experimental considerations, which will be described in detail in a later section of this chapter.

The differential equations are shown below in matrix form:

$$\begin{bmatrix} d[E]/dt \\ d[EL]/dt \end{bmatrix} = \begin{bmatrix} -k_{on}[L] & k_{off} \\ +k_{on}[L] & -k_{off} \end{bmatrix} \begin{bmatrix} [E] \\ [EL] \end{bmatrix} \tag{3.43}$$

The eigenvalues are shown below:

$$\lambda_0 \; = \; 0 \quad \lambda_1 = -k_{on}[L] - k_{off} \; = -k_{obs} \tag{3.44}$$

Because the system is reversible, k_{obs} is identical for each enzyme state. For this reason, equation 3.16 ($A_0 + A_1 e^{\lambda_1 t}$) with eigenvalue substitutions from equation 3.44 is often sufficient to recover the kinetic (k_{on} and k_{off}) and thermodynamic parameters ($K_D = k_{off}/k_{on}$) for a given binding interaction. More specifically, linear regression of a plot of k_{obs} vs [L] will yield a slope equal to k_{on} and y-intercept equal to k_{off}, which may be used to calculate K_D. Therefore, it is possible to fully characterize the kinetic and thermodynamic properties of a binding interaction from a kinetic experiment (spanning multiple ligand concentrations). This is in contrast to the 2-state isomerization model which requires both kinetic and thermodynamic measurements to characterize the K_{eq} and rate constants (note: this is because k_{obs} is concentration independent in the isomerization case, $k_{obs} = k_f + k_r$). Experimental considerations are saved for the end of this section.

The full solutions for the 2-state binding mechanism are shown below:

$$\mathbf{E(L,t) = A_0 + A_1 e^{\lambda_1 t} = \left[E_{eq}\right] + A_1 e^{\lambda_1 t}} \tag{3.45}$$

$$[E](L,t) = [E_{eq}] \; + (A_E)e^{-k_{obs}t} \tag{3.46}$$

$$\text{where } [E_{eq}] \; = \frac{k_{off}}{k_{on}[L]+k_{off}}[E_{total}] = \frac{K_D}{[L]+K_D}[E_{total}]$$

$$A_E = \frac{[E_0]k_{on}[L] - [EL_0]k_{off}}{k_{on}[L] + k_{off}} = \frac{[E_0][L] - [EL_0]K_D}{[L] + K_D}$$

$$k_{obs} = k_{on}[L] + k_{off}$$

$$[EL](L,t) = [EL_{eq}] \; + (A_{EL})e^{-k_{obs}t} \tag{3.47}$$

$$\text{where } [EL_{eq}] \; = \frac{k_{on}[L]}{k_{on}[L]+k_{off}}[E_{total}] = \frac{[L]}{[L]+K_D}[E_{total}]$$

$$A_{EL} = -\frac{[E_0]k_{on}[L] - [EL_0]k_{off}}{k_{on}[L] + k_{off}} = -\frac{[E_0][L] - [EL_0]K_D}{[L] + K_D}$$

$$k_{obs} = k_{on}[L] + k_{off}$$

The initial amounts of each enzyme state, $[E_0]$ and $[EL_0]$, in equations 3.46 and 3.47 represent the equilibrated initial enzyme concentrations. The initial enzyme concentrations depend on the initial ligand concentration, $[L_0]$, given by the equations below:

$$[E_0] = \left[E_{eq,0}\right] = \frac{K_D}{K_D + [L_0]}[E_{total}] \tag{3.48}$$

$$[EL_0] = \left[EL_{eq,0}\right] = \frac{[L_0]}{K_D + [L_0]}[E_{total}] \tag{3.49}$$

For the ligand binding model, the changes in enzyme concentration are now a function time, t, *and* ligand concentration, [L]. As in the previous examples, the full solutions have thermodynamic and kinetic information. Again, the first term describes the final equilibrium amounts of each species, $[E_{eq}]$ and $[EL_{eq}]$, which vary hyperbolically with respect to K_D. The second set of terms describe the temporal change in the respective enzyme state with magnitude, A_i, and rate, k_{obs}. Consistent with the isomerization model, the change in k_{obs}, is independent of the initial amounts of each enzyme species ($[E_0]$ and $[EL_0]$). However, in this case, the k_{obs} depends on the final ligand concentration, [L]. In contrast, the amplitudes, are sensitive to the initial amounts of each enzyme species. Because this is a two-state system, the amplitudes, A_E and A_{EL}, for each enzyme state are equal and opposite which is a consequence of the mass conservation in the system. These concepts are considered in detail in later sections but briefly expanded on below.

The relationship between the k_{obs}, the amplitudes, and the initial conditions are important for experimental considerations. As mentioned above, the k_{obs} are identical for all enzyme states, E and EL, and only depend on the final ligand concentration (i.e. k_{obs} is dependent on the final ligand concentration, [L], and not $[L_0]$ as shown in equations 3.46 and 3.47). However, the

amplitudes are unique to each enzyme state and sensitive to the initial enzyme amounts (i.e. the amplitude expressions contain terms for the initial conditions, and A_E in equation 3.46 ≠ A_{EL} in equation 3.47). Thus, perturbation of an equilibrated solution containing enzyme and ligand by additional ligand (e.g. equilibrium titration, sequential mixing experiment, etc. wherein $[L_0] \neq 0$ and $[L] > [L_0]$) will re-equilibrate at the same rate ($k_{obs} = k_{on}[L] + k_{off}$) as a parallel experiment wherein unbound enzyme is mixed with ligand at [L] (i.e. under the following initial conditions: $[E_0] = [E_{total}]$, $[L_0] = 0$, and $[EL_0] = 0$). Obviously, the sequential titration of ligand into a solution of ligand-bound enzyme will yield a lower change in fraction bound ($\Delta[EL] = [EL_{eq}] - [EL_0]$) compared to the latter experiment ($\Delta[EL] = [EL_{eq}] - 0$). This also holds for dilution experiments wherein an equilibrated solution of enzyme and ligand are diluted, which effectively lowers [L] and shifts the equilibrium towards the unbound enzyme state (i.e. under the following initial conditions: $[E_0] \neq [E_{total}]$, $[EL_0] \neq 0$, $[L_0] \neq 0$ and $[L] < [L_0]$), again at the same rate as a parallel binding experiment with [L]. Stated differently, k_{obs} is identical for binding, dilution, or sequential mixing experiments at [L], regardless of the magnitude or direction of change. Notably, the change in fraction bound may be enhanced in a dilution experiment as compared to the traditional binding experiment. This will be discussed in detail in later sections. For now, excluding dilution scenarios, examination of equations 3.46 and 3.47 and/or an intuitive understanding of ligand binding leads one to the conclusion that the largest change in each enzyme species concentration occurs when ligand is mixed with unbound enzyme. This scenario is discussed below.

As mentioned before, a typical kinetic experiment involves mixing unbound enzyme, [E], with varying amounts of ligand, [L] and measuring the formation of the bound complex [EL]. Thus, equations 3.46 and 3.47 are rewritten below. Here the expressions describe the changes in [E] and [EL] for the given initial conditions: $[E_0] = [E_{total}]$, $[EL_0] = 0$, $[L_0] = 0$.

$$[E](L,t) = \frac{K_D[E_{total}]}{K_D + [L]} + \frac{[L][E_{total}]}{K_D + [L]} e^{-k_{obs}t} \tag{3.50}$$

$$[EL](L,t) = \frac{[L][E_{total}]}{K_D + [L]} - \frac{[L][E_{total}]}{K_D + [L]} e^{-k_{obs}t} \tag{3.51}$$

Although the kinetic and thermodynamic parameters may be obtained from a well-planned kinetic experiment, a few experimental considerations must be discussed. If a detectible signal that is proportional to either enzyme state can be measured (be it the unbound or bound enzyme state, E or EL), a suitable ligand concentration range fits the following criteria: **1)** ligand concentration must be high enough to generate a detectible change in fraction bound (i.e. magnitude of signal, which is proportional to the pre-exponential factor in equations 3.50 and 3.51), **2)** ligand concentration should be selected such that the k_{obs}, which has a linear dependence with respect to [L], is within the limits of detection (i.e. frequency of measurement allows characterization of the rate of change). Regarding the first point, fraction bound varies hyperbolically with respect to the K_D. Therefore, it useful to measure the binding affinity via equilibrium titration prior to selecting the ligand concentration range. That is, though not necessary, one typically ends up performing both kinetic and thermodynamic experiments to fully characterize the kinetic and thermodynamic properties of a ligand binding event. Regarding both points, kinetic experiments at saturating ligand concentrations ([L] >> K_D) produce the largest change in fraction bound (i.e. signal), however, k_{obs}, which increases linearly with ligand concentration, may approach the limits of detection. Similarly, ligand concentrations well below the K_D will approach the lower limit of k_{obs} (k_{obs} approaches k_{off} when [L] approaches 0), however, the change in fraction bound may be lower than the level of detection (fraction bound approaches 0 as [L] approaches 0). Hence, there is a trade-off between signal intensity and kinetic detection. Moreover, for ligands with fast k_{off} values, it becomes increasingly difficult to measure the binding kinetics when k_{off} is near the limits of

detection. Lower ligand concentration limitations are addressed in the dilution experimental sections.

To summarize, an ideal experiment spans a range of ligand concentrations such that a fit of the k_{obs} values plotted against [L] are able to yield the k_{on} and k_{off} values from fits to the k_{obs} expression ($k_{obs} = k_{on}[L]+k_{off}$). While equation 3.16 ($A_0 + A_1 e^{\lambda_1 t}$) is often a good starting point when analyzing the data, there is no physical linkage between the generic exponential expression and a kinetic model. More specifically, data analysis with no regard to the amplitude contributions may have detrimental consequences when distinguishing models, particularly when more complex kinetics (multi-exponential) are involved. This complexity is highlighted in the next few sections.

3.2.4 The general 3-state system: E1 ⇄ E2 ⇄ E3

Here, the general 3-state case is described (Figure 3.2A) where three enzyme states, E1, E2, and E3 are in equilibrium. The kinetics for their interconversions are governed by the microscopic rate constants: k_{12}, k_{21}, k_{23}, and k_{32}. Two observable rates, $k_{obs,1}$ and $k_{obs,2}$, are expected for the three-state systems and a general solution takes the form of equation 3.1 ($A_0 + A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t}$). The derivation of the kinetic and thermodynamic relationships for the general 3-state model are presented in a similar fashion as for the 2-state model.

The differential equations for the general case are shown below in matrix form:

$$\begin{bmatrix} d[E1]/dt \\ d[E2]/dt \\ d[E3]/dt \end{bmatrix} = \begin{bmatrix} -k_{12} & +k_{21} & 0 \\ +k_{12} & -k_{21} - k_{23} & +k_{32} \\ 0 & +k_{23} & -k_{32} \end{bmatrix} \begin{bmatrix} [E1] \\ [E2] \\ [E3] \end{bmatrix} \tag{3.52}$$

and in shortened form:

$$\mathbf{\frac{dE}{dt} = KE} \tag{3.53}$$

where **dE/dt** is the 3x1 column vector of the time derivatives, **K** is the 3x3 coefficient matrix, and **E** the 3x1 column vector of enzyme concentrations in equation 3.52. The condensed description of the system (equation 3.53) is identical to the 2-state system (equation 3.9) but with different matrix dimensions defining the system. Hence, the derivation process for the 3-state system (or infinitely larger linear systems) follows the same order of operations presented for the general 2-state case. Thus, the eigenvalues may be obtained via the same methods employed for the 2x2 system described by equation 3.11 (with 3x3 matrix dimensions here). The determinant of (K- λ**I**), whose roots are equal to the eigenvalues of matrix **K** are presented below:

$$det\left(\begin{bmatrix} -k_{12}-\lambda & +k_{21} & 0 \\ +k_{12} & -k_{21}-k_{23}-\lambda & +k_{32} \\ 0 & +k_{23} & -k_{32}-\lambda \end{bmatrix}\right) = 0 \tag{3.54}$$

$$\begin{aligned}(-k_{12}-\lambda)\big((-k_{21}-k_{23}-\lambda)(-k_{32}-\lambda)-(k_{32})(k_{23})\big) \\ -(k_{21})\big((k_{12})(-k_{32}-\lambda)-(k_{32})(0)\big) \\ +(0)\big((k_{12})(k_{23})-(-k_{21}-k_{23}-\lambda)(0)\big)=0\end{aligned} \tag{3.55}$$

$$\lambda\big(\lambda^2+(k_{12}+k_{21}+k_{23}+k_{32})\lambda+(k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32})\big)=0 \tag{3.56}$$

The roots of equation 3.56 contain one zero eigenvalue, and two non-zero eigenvalues. The roots may be solved by completing the square of the second-degree polynomial in equation 3.56 as follows:

$$\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{3.57}$$

where $a = 1,\ b = k_{12}+k_{21}+k_{23}+k_{32}, and\ c = k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32}$

The eigenvalues are explicitly defined below:

$$k_{obs,0} = 0 \tag{3.58}$$

$$k_{obs,1} = \frac{(k_{12} + k_{21} + k_{23} + k_{32}) - \sqrt{(k_{12} + k_{21} + k_{23} + k_{32})^2 - 4(k_{12}k_{23} + k_{12}k_{32} + k_{21}k_{32})}}{2} \quad (3.59)$$

$$k_{obs,2} = \frac{(k_{12} + k_{21} + k_{23} + k_{32}) + \sqrt{(k_{12} + k_{21} + k_{23} + k_{32})^2 - 4(k_{12}k_{23} + k_{12}k_{32} + k_{21}k_{32})}}{2} \quad (3.60)$$

Among the observed rates, $k_{obs,1} < k_{obs,2}$ so $k_{obs,1}$ and $k_{obs,2}$ are hereafter referred to as k_{slow} and k_{fast}, respectively. The associated amplitudes, A_1 and A_2 are hereafter referred to as A_{slow} and A_{fast}, respectively.

The Frobenius covariants, defined by equation 3.18, are shown for the 3x3 system below:

$$\mathbf{Q}_0 = \frac{(\mathbf{K} - \lambda_1 \boldsymbol{I})(\mathbf{K} - \lambda_2 \boldsymbol{I})}{\lambda_1 \lambda_2} \quad (3.61)$$

$$\mathbf{Q}_0 = \begin{bmatrix} \frac{k_{21}k_{32}}{\lambda_1\lambda_2} & \frac{k_{21}k_{32}}{\lambda_1\lambda_2} & \frac{k_{21}k_{32}}{\lambda_1\lambda_2} \\ \frac{k_{12}k_{32}}{\lambda_1\lambda_2} & \frac{k_{12}k_{32}}{\lambda_1\lambda_2} & \frac{k_{12}k_{32}}{\lambda_1\lambda_2} \\ \frac{k_{21}k_{23}}{\lambda_1\lambda_2} & \frac{k_{21}k_{23}}{\lambda_1\lambda_2} & \frac{k_{21}k_{23}}{\lambda_1\lambda_2} \end{bmatrix}$$

where $\lambda_1\lambda_2 = k_{12}k_{23} + k_{12}k_{32} + k_{21}k_{32}$

$$\mathbf{Q}_1 = \frac{\mathbf{K}(\mathbf{K} - \lambda_2 \boldsymbol{I})}{\lambda_1(\lambda_1 - \lambda_2)} \quad (3.62)$$

$$\mathbf{Q}_1 = \begin{bmatrix} \frac{-k_{12}(k_{23} + k_{32} + \lambda_1)}{\lambda_1(\lambda_1 - \lambda_2)} & \frac{k_{21}(k_{32} + \lambda_1)}{\lambda_1(\lambda_1 - \lambda_2)} & \frac{k_{21}k_{32}}{\lambda_1(\lambda_1 - \lambda_2)} \\ \frac{k_{12}(k_{32} + \lambda_1)}{\lambda_1(\lambda_1 - \lambda_2)} & \frac{-k_{21}(k_{32} + \lambda_1) - k_{23}(k_{12} + \lambda_1)}{\lambda_1(\lambda_1 - \lambda_2)} & \frac{k_{32}(k_{12} + \lambda_1)}{\lambda_1(\lambda_1 - \lambda_2)} \\ \frac{k_{12}k_{23}}{\lambda_1(\lambda_1 - \lambda_2)} & \frac{k_{23}(k_{12} + \lambda_1)}{\lambda_1(\lambda_1 - \lambda_2)} & \frac{-k_{32}(k_{12} + k_{21} + \lambda_1)}{\lambda_1(\lambda_1 - \lambda_2)} \end{bmatrix}$$

where $\lambda_1(\lambda_1 - \lambda_2) = -(k_{12} + k_{21} + k_{23} + k_{32})\lambda_1 - 2(k_{12}k_{23} + k_{12}k_{32} + k_{21}k_{32})$

$$\mathbf{Q}_2 = \frac{\mathbf{K}(\mathbf{K} - \lambda_1 \boldsymbol{I})}{\lambda_2(\lambda_2 - \lambda_1)} \quad (3.63)$$

$$\mathbf{Q}_2 = \begin{bmatrix} \frac{-k_{12}(k_{23}+k_{32}+\lambda_2)}{\lambda_2(\lambda_2-\lambda_1)} & \frac{k_{21}(k_{32}+\lambda_2)}{\lambda_2(\lambda_2-\lambda_1)} & \frac{k_{21}k_{32}}{\lambda_2(\lambda_2-\lambda_1)} \\ \frac{k_{12}(k_{32}+\lambda_2)}{\lambda_2(\lambda_2-\lambda_1)} & \frac{-k_{21}(k_{32}+\lambda_1)-k_{23}(k_{12}+\lambda_2)}{\lambda_2(\lambda_2-\lambda_1)} & \frac{k_{32}(k_{12}+\lambda_2)}{\lambda_2(\lambda_2-\lambda_1)} \\ \frac{k_{12}k_{23}}{\lambda_2(\lambda_2-\lambda_1)} & \frac{k_{23}(k_{12}+\lambda_2)}{\lambda_2(\lambda_2-\lambda_1)} & \frac{-k_{32}(k_{12}+k_{21}+\lambda_2)}{\lambda_2(\lambda_2-\lambda_1)} \end{bmatrix}$$

where $\lambda_2(\lambda_2-\lambda_1) = -(k_{12}+k_{21}+k_{23}+k_{32})\lambda_2 - 2(k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32})$

Note that the numerators of the $\mathbf{Q_1}$ and $\mathbf{Q_2}$ covariants were changed such that the eigenvalue in the numerator matched that of the $\mathbf{Q_n}$ subscript for consistency and to save space. For example, the values in a_{12} of the $\mathbf{Q_1}$ matrix (top middle cell) were changed from '$-k_{21}(k_{12}+k_{21}+k_{23}+\lambda_2)$' to the shorter and equivalent form, '$k_{21}(k_{23}+\lambda_1)$.' For consistency, similar substitutions were made throughout.

The amplitudes were calculated by multiplying the Frobenius covariants by a 3x1 column vector of the initial enzyme concentrations: $\mathbf{E_0} = [E1_0, E2_0, E3_0]$. The resulting amplitudes are shown below:

$$\mathbf{A_{eq}} = \begin{bmatrix} \frac{k_{21}k_{32}[E_{total}]}{k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32}} \\ \frac{k_{12}k_{32}[E_{total}]}{k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32}} \\ \frac{k_{12}k_{23}[E_{total}]}{k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32}} \end{bmatrix} \tag{3.64}$$

where $[E_{total}] = [E1_0] + [E2_0] + [E3_0]$

$$\mathbf{A_{slow}} = \begin{bmatrix} \frac{-k_{12}(k_{23}+k_{32}+\lambda_1)[E1_0] + k_{21}(k_{32}+\lambda_1)[E2_0] + k_{21}k_{32}[E3_0]}{-(k_{12}+k_{21}+k_{23}+k_{32})\lambda_1 - 2(k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32})} \\ \frac{k_{12}(k_{32}+\lambda_1)[E1_0] - \big(k_{21}(k_{32}+\lambda_1)+k_{23}(k_{12}+\lambda_1)\big)[E2_0] + k_{32}(k_{12}+\lambda_1)[E3_0]}{-(k_{12}+k_{21}+k_{23}+k_{32})\lambda_1 - 2(k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32})} \\ \frac{k_{12}k_{23}[E1_0] + k_{23}(k_{12}+\lambda_1)[E2_0] - k_{32}(k_{12}+k_{21}+\lambda_1)[E3_0]}{-(k_{12}+k_{21}+k_{23}+k_{32})\lambda_1 - 2(k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32})} \end{bmatrix} \tag{3.65}$$

$$\mathbf{A_{fast}} = \begin{bmatrix} \dfrac{-k_{12}(k_{23}+k_{32}+\lambda_2)[E1_0] + k_{21}(k_{32}+\lambda_2)[E2_0] + k_{21}k_{32}[E3_0]}{-(k_{12}+k_{21}+k_{23}+k_{32})\lambda_2 - 2(k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32})} \\ \dfrac{k_{12}(k_{32}+\lambda_2)[E1_0] - \big(k_{21}(k_{32}+\lambda_2) + k_{23}(k_{12}+\lambda_2)\big)[E2_0] + k_{32}(k_{12}+\lambda_2)[E3_0]}{-(k_{12}+k_{21}+k_{23}+k_{32})\lambda_2 - 2(k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32})} \\ \dfrac{k_{12}k_{23}[E1_0] + k_{23}(k_{12}+\lambda_2)[E2_0] - k_{32}(k_{12}+k_{21}+\lambda_2)[E3_0]}{-(k_{12}+k_{21}+k_{23}+k_{32})\lambda_2 - 2(k_{12}k_{23}+k_{12}k_{32}+k_{21}k_{32})} \end{bmatrix} \quad (3.66)$$

The amplitude column vectors contain the different enzyme state values in each row in descending order: $\mathbf{A_{eq}}$ = [$E1_{eq}$, $E2_{eq}$, $E3_{eq}$], $\mathbf{A_{slow}}$ = [$A_{slow,E1}$, $A_{slow,E2}$, $A_{slow,E3}$], and $\mathbf{A_{fast}}$ = [$A_{fast,E1}$, $A_{fast,E2}$, $A_{fast,E3}$].

With the eigenvalues and amplitudes defined by equations 3.59 – 3.60 and equations 3.64 – 3.66, respectively, the full solutions are written below:

$$\mathbf{E(t) = A_0 + A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t} = [E_{eq}] + A_{slow} e^{\lambda_{slow} t} + A_{fast} e^{\lambda_{fast} t}} \quad (3.67)$$

$$[E1](t) = [E1_{eq}] + (A_{slow,E1})e^{-k_{slow}t} + (A_{fast,E1})e^{-k_{fast}t} \quad (3.68)$$

$$[E2](t) = [E2_{eq}] + (A_{slow,E2})e^{-k_{slow}t} + (A_{fast,E2})e^{-k_{fast}t} \quad (3.69)$$

$$[E3](t) = [E3_{eq}] + (A_{slow,E3})e^{-k_{slow}t} + (A_{fast,E3})e^{-k_{fast}t} \quad (3.70)$$

Again, the solutions contain thermodynamic (equilibrium concentration) and kinetic (observed rate of change) information as was seen with the 2-state models. The first term describes the equilibrium concentration of each enzyme state. The remaining sets of terms describes the temporal change in the respective enzyme states, which are biphasic in nature (i.e. a slow and fast phase are expected). As with the 2-state model, the same three observations may be pointed out here that reflect the reversibility and conservation of mass in the system: **1**) the k_{obs} expressions are identical for each enzyme state, **2**) the sum of the equilibrium amplitudes is equal to the total enzyme concentration, $[E1_{eq}] + [E2_{eq}] + [E3_{eq}] = [E_{total}]$, and **3**) the sum of the pre-exponential factors equal zero, $A_{slow,E1} + A_{slow,E2} + A_{slow,E3} = 0$ and $A_{fast,E1} + A_{fast,E2} + A_{fast,E3} = 0$. An obvious distinction from the 2-state system is that the amplitudes are no longer equal and opposite from

each other. Instead, the intermediate state amplitudes are equal and opposite to the terminal enzyme states, $A_{slow,E2} = -A_{slow,E1} - A_{slow,E3}$ and $A_{fast,E2} = -A_{fast,E1} - A_{fast,E3}$. That is, formation of E2 resulting from the depletion of E1 and E3 occurs at the same rates, k_{slow} and k_{fast}, with equal and opposite magnitude of states E1 and E3. The last statement is true for the remaining 3-state models.

In this work, the amplitude contribution is used as a diagnostic tool for distinguishing the different models. In addition, the kinetic analyses assume that the change in signal are related to the change in bound enzyme and that multiple bound states equally contribute to the signal (i.e. equal extinction coefficients or quantum yields for either bound state). That is, for models where multiple ligand bound states are populated, the change in signal is assumed to be from the sum of the bound enzyme states (i.e. the bound states are indistinguishable). This is important because this dictates how the change in amplitude contributions are determined. For initial analyses, the relative signal for multiple bound states are assumed to be equal. In later examples, the relative extinction coefficients or quantum yields are varied. The amplitude contributions are described in detail below.

Due to the lengthy nature of the amplitude values, it is simpler to discuss the amplitudes in terms of the percentage contribution of each phase (slow and fast phase) for the bound enzyme states. The percentage contribution of both phases is given by the equations below:

$$\%\text{slow phase} = 100\frac{A_{slow,bound}}{A_{slow,bound} + A_{\text{fast,bound}}} \tag{3.71}$$

$$\%\text{fast phase} = 100\frac{A_{fast,bound}}{A_{slow,bound} + A_{\text{fast,bound}}} \tag{3.72}$$

Depending on the model, the terms $A_{slow,bound}$ and $A_{fast,bound}$ in equations 3.71 and 3.72 may represent an individual species or a sum. For example, in the induced fit model (Figure 3.1B), two species, EL and E*L, represent bound enzyme states correlating to E2 and E3 in equations 3.69

and 3.70 (i.e. in equations 3.69 and 3.70 for the induced fit case: [EL] = [E2], [E*L] = [E3]). Thus, $A_{slow,bound}$ and $A_{fast,bound}$ in equations 3.71 and 3.72 represent the sum of the corresponding A_{slow} and A_{fast} values from each bound state (e.g. for the induced fit case, A_{slow} and A_{fast} in equations 3.71 and 3.72 are defined as follows: $A_{slow} = A_{slow,EL} + A_{slow,E^*L}$ and $A_{fast} = A_{fast,EL} + A_{fast,E^*L}$). In contrast, the conformational selection model (Figure 3.1C) has one bound species, E*L, which correlates to E3 in the general model (Figure 3.1A and equation 3.70). Thus, for the conformational selection model, $A_{slow,bound}$ and $A_{fast,bound}$ represent one species where $A_{slow,bound} = A_{slow,E^*L}$ and $A_{fast,bound} = A_{fast,E^*L}$. The amplitude contribution is expanded in the next section. As in the 2-state cases, an example of the conversion of the general 3-state case to a 3-state binding model with definitions of the binding affinity are presented in the next section with the induced fit model (Figure 3.1B) as an example case.

3.2.5 Translating the general 3-state system expressions to the ligand binding models: induced fit model as an example case

The equations describing the general 3-state binding model in the previous section can be easily converted to represent the individual 3-state ligand binding models, shown in Figure 3.1, by substituting the rate constants in the general case to the corresponding rate constants in the desired model. For example, the differential equations for the induced fit (IF) model are shown below in matrix notation:

$$\begin{bmatrix} d[E]/dt \\ d[EL]/dt \\ d[E^*L]/dt \end{bmatrix} = \begin{bmatrix} -k_{on}[L] & k_{off} & 0 \\ k_{on}[L] & -k_{off} - k_f & k_r \\ 0 & k_f & -k_r \end{bmatrix} \begin{bmatrix} [E] \\ [EL] \\ [E^*L] \end{bmatrix} \tag{3.73}$$

which were obtained by the following substitutions to the general case (equation 3.52): = $k_{on}[L] = k_{12}$', $k_{off} = k_{21}$, $k_f = k_{23}$, $k_r = k_{32}$, E = E1, EL = E2, and E*L = E3. Thus, the k_{obs} expressions may

be rewritten from the general case with similar substitutions to the k_{obs} expressions (equation 3.59 and 3.60) as follows:

$$k_{slow} = \frac{\left(k_{on}[L] + k_{off} + k_f + k_r\right) - \sqrt{\left(k_{on}[L] + k_{off} + k_f + k_r\right)^2 - 4\left(k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r\right)}}{2} \tag{3.74}$$

$$k_{fast} = \frac{\left(k_{on}[L] + k_{off} + k_f + k_r\right) + \sqrt{\left(k_{on}[L] + k_{off} + k_f + k_r\right)^2 - 4\left(k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r\right)}}{2} \tag{3.75}$$

Likewise, the amplitude expressions (general case expressions 3.64 – 3.66) are defined below:

$$\mathbf{E_{eq}} = \begin{bmatrix} \dfrac{k_{off}k_r[E_{total}]}{k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r} \\ \dfrac{k_{on}[L]k_{off}[E_{total}]}{k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r} \\ \dfrac{k_{on}[L]k_f[E_{total}]}{k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r} \end{bmatrix} \tag{3.76}$$

where $[E_{total}] = [E1_0] + [E2_0] + [E3_0]$

$$\mathbf{A_{slow}} = \begin{bmatrix} \dfrac{-k_{on}[L]\left(k_f + k_r + \lambda_1\right)[E1_0] + k_{off}(k_r + \lambda_1)[E2_0] + k_{off}k_r[E3_0]}{-(k_{on}[L] + k_{off} + k_f + k_r)\lambda_1 - 2(k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r))} \\ \dfrac{k_{on}[L](k_r + \lambda_1)[E1_0] + \left(-k_{off}(k_r + \lambda_1) + k_f(k_{on}[L] + \lambda_1)\right)[E2_0] + k_r(k_{on}[L] + \lambda_1)[E3_0]}{-\left(k_{on}[L] + k_{off} + k_f + k_r\right)\lambda_1 - 2\left(k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r\right)} \\ \dfrac{k_{on}[L]k_f[E1_0] - k_f(k_{on}[L] + \lambda_1)[E2_0] - k_r\left(k_{on}[L] + k_{off} + \lambda_1\right)[E3_0]}{-\left(k_{on}[L] + k_{off} + k_f + k_r\right)\lambda_1 - 2\left(k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r\right)} \end{bmatrix} \tag{3.77}$$

$$\mathbf{A_{fast}} = \begin{bmatrix} \dfrac{-k_{on}[L]\left(k_f + k_r + \lambda_2\right)[E1_0] + k_{off}(k_r + \lambda_2)[E2_0] + k_{off}k_r[E3_0]}{-(k_{on}[L] + k_{off} + k_f + k_r)\lambda_2 - 2(k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r))} \\ \dfrac{k_{on}[L](k_r + \lambda_2)[E1_0] + \left(-k_{off}(k_r + \lambda_2) + k_f(k_{on}[L] + \lambda_1)\right)[E2_0] + k_r(k_{on}[L] + \lambda_2)[E3_0]}{-\left(k_{on}[L] + k_{off} + k_f + k_r\right)\lambda_2 - 2\left(k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r\right)} \\ \dfrac{k_{on}[L]k_f[E1_0] - k_f(k_{on}[L] + \lambda_1)[E2_0] - k_r\left(k_{on}[L] + k_{off} + \lambda_1\right)[E3_0]}{-\left(k_{on}[L] + k_{off} + k_f + k_r\right)\lambda_2 - 2\left(k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r\right)} \end{bmatrix} \tag{3.78}$$

The amplitude column vectors contain the different enzyme state values in each row in descending order: $\mathbf{A_{eq}}$ = [E_{eq}, EL_{eq}, E^*L_{eq}], $\mathbf{A_{slow}}$ = [$A_{slow,E}$, $A_{slow,EL}$, A_{slow,E^*L}], and $\mathbf{A_{fast}}$ = [$A_{fast,E}$, $A_{fast,EL}$, A_{fast,E^*L}]. The initial enzyme concentrations are defined below:

$$[E1_0] = \frac{k_{off}k_r[E_{total}]}{k_{on}[L_0]k_f + k_{on}[L_0]k_r + k_{off}k_r} \tag{3.79}$$

$$[E2_0] = \frac{k_{on}[L_0]k_{off}[E_{total}]}{k_{on}[L_0]k_f + k_{on}[L_0]k_r + k_{off}k_r} \tag{3.80}$$

$$[E3_0] = \frac{k_{on}[L_0]k_f[E_{total}]}{k_{on}[L_0]k_f + k_{on}[L_0]k_r + k_{off}k_r} \tag{3.81}$$

where initial ligand concentration is equal to L_0.

With the eigenvalues, amplitudes, and initial enzyme concentrations defined by equations 3.74 – 3.81, the full solutions are written below:

$$\mathbf{E(L,t)} = \mathbf{[E_{eq}]} + \mathbf{A_{slow}}e^{-k_{slow}t} + \mathbf{A_{fast}}e^{-k_{fast}t} \tag{3.82}$$

$$[E](L,t) = [E_{eq}] + A_{slow,E}e^{-(k_{slow})t} + A_{\mathrm{fast,E}}e^{-(k_{fast})t} \tag{3.83}$$

$$[EL](L,t) = [EL_{eq}] + A_{slow,EL}e^{-(k_{slow})t} + A_{\mathrm{fast,EL}}e^{-(k_{fast})t} \tag{3.84}$$

$$[E^*\mathrm{L}](L,t) = [E^*\mathrm{L}_{eq}] + A_{\mathrm{slow},E^*\mathrm{L}}e^{-(k_{\mathrm{slow}})t} + A_{\mathrm{fast},E^*\mathrm{L}}e^{-(k_{\mathrm{fast}})t} \tag{3.85}$$

The changes in each enzyme state are functions of ligand concentration and time. The individual amplitude and k_{obs} expressions for each kinetic model are summarized in Tables 3.1 – 3.6.

Finally, the equations for the percentage contribution of the slow and fast phases (general case expressions 3.71 – 3.72) are shown below:

$$\%slow\ phase\ = 100\frac{\varepsilon_{EL}A_{slow} + \varepsilon_{E^*L}A_{slow,E^*L}}{\varepsilon_{EL}(A_{slow,EL} + A_{fast,EL}) + \varepsilon_{E^*L}(A_{slow,E^*L} + A_{fast,E^*L})} \tag{3.86}$$

$$\%fast\ phase\ = 100\frac{\varepsilon_{EL}A_{slow} + \varepsilon_{E^*L}A_{slow,E^*L}}{\varepsilon_{EL}(A_{slow,EL} + A_{fast,EL}) + \varepsilon_{E^*L}(A_{slow,E^*L} + A_{fast,E^*L})} \tag{3.87}$$

For the induced fit model, the percentage contribution of each phase represents a sum of the amplitude values from the two bound states, EL and E*L, multiplied by a scalar, ε_i, for each enzyme state that represents the relative contribution to the bound state signal (i.e. extinction

coefficient or quantum yield). Unless otherwise stated, the scalars are assumed to be equal to 1 for each enzyme state.

So far, the binding affinity for the three-state systems have not been mentioned because the general case does not include a ligand binding step. It is true that binding affinity or K_D still takes the form of equation 3.39 ($K_D = k_{off}/k_{on}$). Moreover, for models with multiple binding steps (Figure 3.1D-F), there are two K_D values corresponding to each binding equilibrium ($K_{D1} = k_{off1}/k_{on1}$ and $K_{D2} = k_{off2}/k_{on2}$). However, for multi-state systems, it is also informative to express the apparent binding affinity, $K_{D,app}$, which describes the distribution of bound and unbound enzyme states when multiple forms of either species are expected. A general equation is shown below:

$$K_{D,app} = \frac{[E_{unbound}][L]}{[E_{bound}]} \tag{3.88}$$

Depending on the model, the numerator or the denominator may represent an individual or sum of enzyme states. Regardless, the $K_{D,app}$ is the equal to the ligand concentration at which half of all binding sites are occupied ($[E_{unbound}] = [E_{bound}]$):

$$[E_{unbound}] = [E_{unbound}] \tag{3.89}$$

Using the IF case as an example, equation 3.89 may be re-written as follows (with substitutions from the equilibrium concentration column vector defined by equation 3.76):

$$[E_{eq}] = [EL_{eq}] + [E^*L_{eq}] \tag{3.90}$$

$$\frac{k_{off}k_r[E_{total}]}{k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r} = \frac{k_{on}[L]k_r[E_{total}]}{k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r} + \frac{k_{on}[L]k_f[E_{total}]}{k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r}$$

which reduces to:

$$k_{off}k_r = k_{on}[L]k_r + k_{on}[L]k_f = [L]k_{on}(k_r + k_f) \tag{3.91}$$

Isolating the expression for [L] yields $K_{D,app}$:

$$[L] = K_{D,app,IF} = \frac{k_{off}k_r}{k_{on}(k_r + k_f)} = K_D\left(\frac{k_r}{k_r + k_f}\right) \tag{3.92}$$

This process can be repeated for all the remaining models. The $K_{D,app}$ expressions for each model are summarized in Tables 3.1 – 3.6.

From here, the equilibrium concentrations may be expressed in terms of the apparent binding affinity as shown below:

$$[E_{eq}] = \frac{k_{off}k_r[E_{total}]}{k_{off}k_r + [L]k_{on}(k_f + k_r)} = \frac{K_{D,app,IF}}{K_{D,app,IF} + [L]}E_{max} \tag{3.93}$$

where $E_{max} = 1 * [E_{total}]$

$$[EL_{eq}] = \frac{k_{on}[L]k_r[E_{total}]}{k_{off}k_r + [L]k_{on}(k_f + k_r)} = \frac{[L]}{K_{D,app,IF} + [L]}EL_{max} \tag{3.94}$$

where $EL_{max} = \frac{k_r}{k_f + k_r} * [E_{total}]$

$$[E^*L_{eq}] = \frac{k_{on}[L]k_f[E_{total}]}{k_{off}k_r + [L]k_{on}(k_f + k_r)} = \frac{[L]}{K_{D,app,IF} + [L]}E^*L_{max} \tag{3.95}$$

where $E^*L_{max} = \frac{k_f}{k_f + k_r} * [E_{total}]$

The equations that describe the equilibrium concentrations for each enzyme state parallel the expressions for the 2-state model (defined in equations 3.46 and 3.47). Importantly, the expressions for the remaining models take similar forms. This becomes important for the dilution experiment theory. For now, a few points may be made here: **1**) the change in each enzyme state varies hyperbolically with respect to the apparent binding affinity and **2**) the maximum amount of each species (E_{max}, EL_{max}, and E^*L_{max}) are expressed as a ratio of total enzyme. Expanding on the first point in the context of the induced fit model, this means that the buildup of the bound states, EL and E*L, are scaled by the expressions EL_{max} and E^*L_{max} and vary hyperbolically with respect to

$K_{D,app}$. This is best illustrated for the case where $k_f = k_r$. In this case, $EL_{max} = E^*L_{max} = 0.5$, thus, a plot of the change in EL and E*L vs [L] yields two identical curves. Expanding on the second point in the context of the induced fit model, the maximum amount of each bound state depends on the isomerization equilibrium. Notably, the EL_{max} and E^*L_{max} expressions match the equilibrium expressions E_{eq} and E^*_{eq}, respectively, from the 2-state isomerization model (equations 3.37 and 3.38). The same is true for the conformational selection model, however, the same maximum value expressions apply to the initial enzyme concentrations (not the bound state concentration) in the absence of ligand.

To summarize, the expressions that describe the observed rates, equations 3.74 and 3.75 (see also Table 3.2), contain no terms for any of the enzyme states. Thus, as mentioned for the 2-state binding model, the observed rates are only dependent on the final ligand concentration, regardless of the initial enzyme conditions or experimental conditions. In addition, the k_{obs} expressions are identical for each enzyme state and they contain a term for each microscopic rate constant. That is, the k_{obs} do not correspond to a single step, rather, they reflect the fact that two equilibria are coupled by an intermediate state and thus include terms for all rate constants. However, the amplitudes are dependent on the initial amounts of each enzyme state ($[E1_0]$, $[E2_0]$, $[E3_0]$). In addition, the amplitude value relating to the equilibrium amounts of each species ($[E1_{eq}]$, $[E2_{eq}]$, and $[E3_{eq}]$) may be used to calculate the $K_{D,app}$. Finally, the expressions for k_{obs}, the amplitudes, and $K_{D,app}$ are all unique to a given model. In the next section, the two limiting cases of ligand binding mechanisms involving conformational changes, which are the subject of this thesis, are presented.

3.3 Kinetic models in detail

3.3.1 Linkage between conformational change and ligand binding: IF and CS

Thermodynamic linkage between ligand binding and protein conformational change is a hallmark of biological systems. The kinetics of the enzyme-ligand interactions and the associated conformational changes can be a critical aspect of protein function and enzyme catalysis. Complex kinetic models for these linkages have been considered in detail for substrate specific enzymes and ligand specific proteins, and the limiting cases include IF (Figure 3.2B) and CS (Figure 3.2C) [73,127,128,131,139]. Both limiting cases envision a free-energy landscape with multiple minima representing different ligand-free, E and E*, and ligand-bound enzyme conformations, EL and E*L. The IF case occurs when ligand binding induces a conformational change to stabilize the ligand bound state. In contrast, CS occurs when ligand 'selects' and only binds to one pre-existing enzyme conformation, E*, from an ensemble of unbound states, E and E*. In relation to the conformational landscape, ligand binding in the IF model results in expansion of the landscape to include a new conformation, E*L, whereas ligand binding in the CS model results in a population shift in the ensemble of conformations, E and E*, to the ligand 'preferred' conformation, E*L. It is important to note that the limiting case models are not mutually exclusive and both models may occur for a single enzyme-ligand pair. Additionally, the binding mechanism can be context dependent wherein different ligands can bind to the same enzyme through different mechanisms. Hence, differences in the 'selection criteria'

Traditional kinetic approaches aimed at distinguishing the IF and CS models from stopped-flow kinetic measurements rely on interpretation of the direction of change of the ligand dependent k_{slow} behavior from the double exponential fits of data, which has a hyperbolic dependence with respect to ligand concentration. For decades, kineticists interpreted a hyperbolic increase in k_{slow} as IF and a hyperbolic decrease in $k_{obs,slow}$ as CS. Vogt and Di Cera recently demonstrated that the

distinction is only valid if the conformational rearrangement step (k_f an k_r in Figure 3.1C) is rate-limiting [139]. In their analysis, they showed that CS can also display a hyperbolic increase in k_{slow} when the conformational rearrangements are not rate-limiting. They conclude that a hyperbolic decrease in k_{slow} behavior is unambiguous proof of CS while a hyperbolic increase or even concentration independent k_{slow} behavior can result from either the IF or CS model.

To further complicate the matter, Chakraborty and Di Cera recently discussed cases for which two different scenarios of IF and CS can yield equivalent k_{fast}, k_{slow} and $K_{D,app}$ values despite differences in the intrinsic kinetic rate constants for either model [131]. Importantly, the resulting amplitude values are not equivalent for the two models, demonstrating the utility of a global analysis approach which takes into account the fractional contribution of the fast and slow phases to the overall binding signal. In this chapter, we discuss the examples presented by Chakraborty and Di Cera and include examples of cases where different behaviors and k_{slow} and k_{fast} can be expected.

3.3.2 Multiple ligand binding interactions: Sequential binding, substrate reorientation and persistent heterogeneity

Here I aim to provide kinetic approaches to aid in differentiating IF and CS and expand our analysis to include alternative, infrequently considered, models where multi-step mechanisms of ligand binding are involved. The remaining models differ from the IF and CS models because they do not include a reversible isomerization step. The models contain two binding steps and are depicted in Figure 3.1D-F. The sequential binding (SB) and substrate reorientation (SR) models are linear three-state systems. Similar to the IF and CS models, the full solutions to the respective systems take the form of equations 3.26 – 3.28 with observed rates listed in Table 3.1. The

persistent heterogeneity (PH) model contains two two-state binding systems wherein two different non-interconverting enzyme conformers, E and E*, bind to ligand to form their respective bound complexes, EL and E*L. The full solutions to either enzyme conformer system take the form of equations 3.17 – 3.18 with observed rates listed in Table 3.1. The models in Figure 3.1 are expected to display a theoretical maximum of two observable rates (i.e. the re-equilibration processes due to ligand binding or dissociation may present as double exponential processes). Notably, the SR model also displays hyperbolic behavior of k_{slow} similar to the IF and CS models. However, the behavior of both observed rates in the SB and PH models have linear dependence with respect to ligand. In this chapter, I discuss the amplitude contributions for each model and also detail the kinetic and thermodynamic signatures with the goal of distinguishing the different models.

3.4 Experimental Procedures

3.4.1 Kinetic simulations

Simulations of the systems described in Figure 3.2 were performed using MATLAB. The transient concentrations of each enzyme state were governed by the differential equations listed in the appendix. The initial amounts of each enzyme state were calculated using the equations listed in the appendix. For each case, the rate constants are indicated in the figure legends.

3.4.2 Signal considerations

Because most kinetic studies of small molecule ligand binding utilize stopped-flow absorbance or fluorescence the rates of change in these optical properties are considered as the experimental probe, although other experimental probes will yield the same analytical expressions

as long as a few assumptions are true. Unless otherwise stated, the extinction coefficients for the bound species are equivalent. When different extinction coefficients are involved, the amplitude contributions are defined in a similar manner as shown for the induced fit example in section 3.2.5 (equations 3.86 and 3.87).

3.4.3 Limiting cases: rapid first step, rapid second step, and the stationary intermediate

Three limiting cases for each model are presented, which depend on the relative magnitudes of the rate constants associated with each model. The three test cases include: rapid first step, rapid second step, and stationary intermediate. Regarding the IF and CS, the first two cases deal with the relationship between the binding and isomerization steps. Regarding the remaining models, the first two cases deal with the relationship between the two binding steps. The stationary intermediate case involves a rapidly interconverting intermediate enzyme state (E2) for the three-state linear models. Notably, the stationary intermediate case is not applicable to the PH model because it involves binding interactions with two non-interconverting enzyme states. Below, the expressions for k_{fast} and k_{slow} are simplified for each limiting case. The simplified expressions for each model are summarized in Table 3.7. The process to arrive at the simplified expressions are detailed below using the general case nomenclature for simplicity.

To simplify the k_{obs} expressions, it is useful to express the sum and the product of the relaxation rates (equations 3.59 and 3.60) as rearrangement of these equations aids in simplification of the k_{obs} expressions in some cases [126]. The sum and product of the relaxation rates from the general 3-state model solution below:

$$\sum k_n = k_{slow} + k_{fast} = k_{12} + k_{21} + k_{23} + k_{32} \tag{3.96}$$

$$\prod k_n = k_{slow}k_{fast} = k_{12}k_{23} + k_{12}k_{32} + k_{21}k_{32} \tag{3.97}$$

It is essential to note that the equations 3.96 and 3.97 are equal to the 'b' and 'c' values in the quadratic expression that describes k_{obs} (equation 3.57), which are defined for each model in Tables 3.2 – 3.6. As will be mentioned later, the sum or product of the k_{obs} expressions can help differentiate models based on the ligand dependent behavior (linear vs quadratic).

Case 1: rapid first step

First, the k_{obs} expressions are simplified for the case where the first step is much faster than the second step (k_{12}, k_{21} >> k_{23}, k_{32}). Given that k_{fast} >> k_{slow}, equation 3.96 simplifies as shown:

$$k_{fast} = k_{12} + k_{21} \tag{3.98}$$

Substitution of equation 3.98 into equation 3.97 yields the simplified k_{slow} equation below:

$$k_{slow} = \frac{k_{12}k_{23} + k_{32}(k_{12} + k_{21})}{k_{12} + k_{21}} = k_{23}\left(\frac{k_{12}}{k_{12} + k_{21}}\right) + k_{32} \tag{3.99}$$

Case 2: rapid second step

The second simplification involves the opposite case where the second step is much faster than the second step (k_{23}, k_{32} >> k_{12}, k_{21}). When k_{fast} >> k_{slow}, equation 3.96 simplifies as shown:

$$k_{fast} = k_{23} + k_{32} \tag{3.100}$$

Substitution of equation 3.100 into equation 3.97 yields the simplified k_{slow} equation below:

$$k_{slow} = \frac{k_{12}(k_{23} + k_{32}) + k_{21}k_{32}}{k_{23} + k_{32}} = k_{12} + k_{21}\left(\frac{k_{32}}{k_{23} + k_{32}}\right) \tag{3.101}$$

Case 3: stationary intermediate

The final simplification involves a stationary intermediate case where the rates associated with the depletion of the intermediate enzyme state are faster than the rates associated with its formation (k_{21}, k_{23} >> k_{12}, k_{32}). Again, assume that the k_{fast} >> k_{slow}. Thus, equation 3.96 simplifies as follows:

$$k_{fast} = k_{21} + k_{23} \tag{3.102}$$

Substitution of equation 3.102 into equation 3.97 yields the simplified k_{slow} equation below:

$$k_{slow} = \frac{k_{12}k_{23} + k_{12}k_{32} + k_{21}k_{32}}{k_{21} + k_{23}} \tag{3.103}$$

$$= k_{12}\left(\frac{k_{23}}{k_{21} + k_{23}}\right) + \frac{k_{12}k_{32}}{k_{21} + k_{23}} + k_{32}\left(\frac{k_{21}}{k_{21} + k_{23}}\right)$$

The second term approaches zero as k_{12} and k_{32} << k_{21} and k_{23}, therefore, equation 3.103 further simplifies to:

$$k_{slow} = k_{12}\left(\frac{k_{23}}{k_{21} + k_{23}}\right) + k_{32}\left(\frac{k_{21}}{k_{21} + k_{23}}\right) \tag{3.104}$$

The simplified expressions for each kinetic model are summarized in Table 3.7.

3.4.4 Identical k_{obs} case: apparently equivalent IF and CS case

Two cases will be presented wherein the IF and CS models produce identical k_{obs} behavior. Chakraborty and Di Cera identified conditions where the k_{obs} for the IF and CS are nearly identical. This occurs under the following conditions:

$$E \underset{k_{off,IF}}{\overset{k_{on}[L]}{\rightleftarrows}} EL \underset{k_{r,IF}}{\overset{k_{f,IF}}{\rightleftarrows}} E^*L \quad = \quad E \underset{k_x}{\overset{k_{f,IF} + k_{r,IF}}{\rightleftarrows}} E^* \underset{k_{off,IF} - k_x}{\overset{k_{on}[L]}{\rightleftarrows}} E^*L \tag{3.105}$$

where the k_{obs} expressions for the IF case on the left is equivalent to the CS case on the right. The k_x term is defined by the difference in the IF model limits by the expression below:

$$k_x = \lim_{L\to 0} k_{fast,IF} - \lim_{L\to\infty} k_{slow,IF}$$
$$k_x = \frac{k_{off} - k_f - k_r + \sqrt{(k_{off} + k_f + k_r)^2 - 4k_{off}k_r)}}{2} \quad (3.106)$$

where the limits for the IF model k_{obs} expressions are listed in Table 3.2. In this work, two cases presented by Chakraborty and Di Cera will be discussed with additional commentary about the amplitude behavior. The two cases are shown below:

Case 1:

$$E \underset{3.6\ s^{-1}}{\overset{2.6\ \mu M^{-1}s^{-1}[L]}{\rightleftarrows}} EL \underset{2.6\ s^{-1}}{\overset{12\ s^{-1}}{\rightleftarrows}} E^*L \quad = \quad E \underset{3.1\ s^{-1}}{\overset{14.6\ s^{-1}}{\rightleftarrows}} E^* \underset{0.50\ s^{-1}}{\overset{2.6\ \mu M^{-1}s^{-1}[L]}{\rightleftarrows}} E^*L \quad (3.107)$$

Case 2:

$$E \underset{7.5\ s^{-1}}{\overset{0.54\ mM^{-1}s^{-1}[L]}{\rightleftarrows}} EL \underset{0.36\ s^{-1}}{\overset{0.44\ s^{-1}}{\rightleftarrows}} E^*L \quad = \quad E \underset{7.2\ s^{-1}}{\overset{0.8\ s^{-1}}{\rightleftarrows}} E^* \underset{0.34\ s^{-1}}{\overset{0.54\ mM^{-1}s^{-1}[L]}{\rightleftarrows}} E^*L \quad (3.108)$$

3.4.5 Identical k_{obs} case: apparently equivalent IF and SR case

Two cases will be presented whethe IF and SR models produce identical k_{obs} behavior. This occurs under the following conditions:

$$E \underset{k_{off,IF}}{\overset{k_{on,IF}[L]}{\rightleftarrows}} EL \underset{k_{r,IF}}{\overset{k_{f,IF}}{\rightleftarrows}} E^*L \quad = \quad EL \underset{k_{y,21}[L]}{\overset{k_{y,12}}{\rightleftarrows}} E \underset{k_{y,32}}{\overset{k_{y,23}[L]}{\rightleftarrows}} EL^* \quad (3.109)$$

where the k_{obs} expressions for the IF case on the left is equivalent to the SR case on the right. The SR rate constants, k_y, are defined in terms of the IF model in the expressions below:

$$k_{y,12} = \lim_{L\to 0} k_{slow,IF} = k_{off,IF} - k_x \quad (3.110)$$

$$= \frac{k_{off,IF} + k_{f,IF} + k_{r,IF} - \sqrt{(k_{off,IF} + k_{f,IF} + k_{r,IF})^2 - 4k_{r,IF}k_{off,IF}}}{2}$$

$$k_{y,21} = \frac{k_{on,IF}\left(-k_{off,IF} + k_{f,IF} + k_{r,IF} + k_x\right)}{-k_{off,IF} + k_{f,IF} + k_{r,IF} + 2k_x} \tag{3.111}$$

$$= \frac{k_{on,IF}(-k_{off,IF} + k_{f,IF} + k_{r,IF} + \sqrt{(k_{off,IF} + k_{f,IF} + k_{r,IF})^2 - 4k_{r,IF}k_{off,IF}})}{2\sqrt{(k_{off,IF} + k_{f,IF} + k_{r,IF})^2 - 4k_{r,IF}k_{off,IF}}}$$

$$k_{y,23} = \frac{k_{on,IF}k_x}{k_{f,IF} + k_{r,IF} - k_{off,IF} + 2k_x} \tag{3.112}$$

$$= \frac{k_{on,IF}(+k_{off,IF} - k_{f,IF} - k_{r,IF} + \sqrt{(k_{off,IF} + k_{f,IF} + k_{r,IF})^2 - 4k_{r,IF}k_{off,IF}})}{2\sqrt{(k_{off,IF} + k_{f,IF} + k_{r,IF})^2 - 4k_{r,IF}k_{off,IF}}}$$

$$k_{y,32} = \lim_{L\to 0} k_{fast,IF} = k_{f,IF} + k_{r,IF} + k_x \tag{3.113}$$

$$= \frac{k_{off,IF} + k_{f,IF} + k_{r,IF} + \sqrt{(k_{off,IF} + k_{f,IF} + k_{r,IF})^2 - 4k_{r,IF}k_{off,IF})}}{2}$$

where the limits for the IF model k_{obs} expressions are listed in Table 3.2 and k_x is described by equation 3.106. Due to the symmetry in the SR system, the assignment of the slower of the two off-rate steps in the SR model is inconsequential. Here, k_{off1} (associated with the left equilibrium in equation 3.109) is set to be less than k_{off2}. The IF cases are identical to those presented by Chakraborty and Di Cera (i.e. the SR cases are equivalent to the Chakraborty and Di Cera cases presented in section 3.4.4). The two cases are shown below:

<u>Case 1:</u>

$$E \underset{3.6\,s^{-1}}{\overset{2.6\,\mu M^{-1}s^{-1}[L]}{\rightleftarrows}} EL \underset{2.6\,s^{-1}}{\overset{12\,s^{-1}}{\rightleftarrows}} E^*L \quad = \quad EL \underset{2.13\,\mu M^{-1}s^{-1}[L]}{\overset{0.53\,s^{-1}}{\rightleftarrows}} E \underset{17.7\,s^{-1}}{\overset{0.47\mu M^{-1}s^{-1}[L]}{\rightleftarrows}} EL^* \tag{3.114}$$

Case 2:

$$E \underset{7.5\ s^{-1}}{\overset{0.54\ mM^{-1}s^{-1}[L]}{\rightleftarrows}} EL \underset{0.36\ s^{-1}}{\overset{0.44\ s^{-1}}{\rightleftarrows}} E^*L \quad = \quad EL \underset{0.032\ s^{-1}\ mM^{-1}s^{-1}[L]}{\overset{0.34\ s^{-1}}{\rightleftarrows}} E \underset{8.0\ s^{-1}}{\overset{0.51\ mM^{-1}s^{-1}[L]}{\rightleftarrows}} EL^* \tag{3.115}$$

3.4.6 Identical k_{obs} case: apparently equivalent CS and SR case

One case will be presented where the SR and CS models produce identical k_{obs} behavior. This occurs under the following conditions:

$$EL \underset{k_{on1,SR}[L]}{\overset{k_{off1,SR}}{\rightleftarrows}} E \underset{k_{off2,SR}}{\overset{k_{on2,SR}[L]}{\rightleftarrows}} EL^* \quad = \quad E \underset{k_{z,21}[L]}{\overset{k_{z,12}}{\rightleftarrows}} E^* \underset{k_{z,32}}{\overset{k_{z,23}[L]}{\rightleftarrows}} E^*L \tag{3.116}$$

where the k_{obs} expressions for the SR case on the left is equivalent to the CS case on the right. The CS rate constants, k_z, are defined in terms of the SR model in the expressions below:

$$k_{z,12} = \lim_{L\to\infty} k_{slow,SR} \tag{3.117}$$

$$= \frac{k_{off1,SR}k_{on2,SR} + k_{on1,SR}k_{off2,SR}}{k_{on1,SR} + k_{on2,SR}}$$

$$k_{z,21} = k_{off2,SR} - \lim_{L\to\infty} k_{slow,SR} \tag{3.118}$$

$$= k_{off2,SR} - \frac{k_{off1,SR}k_{on2,SR} + k_{on1,SR}k_{off2,SR}}{k_{on1,SR} + k_{on2,SR}}$$

$$k_{z,23} = k_{on1,SR} + k_{on2,SR} \tag{3.119}$$

$$k_{z,32} = \lim_{L\to 0} k_{slow,SR} = k_{off1,SR} \tag{3.120}$$

where the limits for the SR model k_{obs} expressions are listed in Table 3.5. Due to the symmetry in the SR system, the assignment of the slower of the two off-rate steps in the SR model is inconsequential. Here, k_{off1} (associated with the left equilibrium in equation 3.116) is set to be less than k_{off2}. The CS cases are identical to those presented by Chakraborty and Di Cera (i.e. the SR

cases are equivalent to the Chakraborty and Di Cera cases presented in section 3.4.4 and 3.4.5). The two cases are shown below:

Case 1:

$$EL \underset{2.13\ \mu M^{-1}s^{-1}[L]}{\overset{0.53\ s^{-1}}{\rightleftarrows}} E \underset{17.7\ s^{-1}}{\overset{0.47\mu M^{-1}s^{-1}[L]}{\rightleftarrows}} EL^* \quad = \quad E \underset{3.1\ s^{-1}}{\overset{14.6\ s^{-1}}{\rightleftarrows}} E^* \underset{0.50\ s^{-1}}{\overset{2.6\ \mu M^{-1}s^{-1}[L]}{\rightleftarrows}} E^*L \tag{3.121}$$

Case 2:

$$EL \underset{0.032\ s^{-1}\ mM^{-1}s^{-1}[L]}{\overset{0.34\ s^{-1}}{\rightleftarrows}} E \underset{8.0\ s^{-1}}{\overset{0.51\ mM^{-1}s^{-1}[L]}{\rightleftarrows}} EL^* \quad = \quad E \underset{7.2\ s^{-1}}{\overset{0.8\ s^{-1}}{\rightleftarrows}} E^* \underset{0.34\ s^{-1}}{\overset{0.54\ mM^{-1}s^{-1}[L]}{\rightleftarrows}} E^*L \tag{3.122}$$

3.4.7 Dilution experiment

The dilution experiment serves as a means of signal amplification at lower ligand concentrations. Importantly, the K_D (for two state models) or K_{Dapp} for (some) of the three-state models adequately describes the change in behavior when more than one bound species is expected. It is intuitive to think about the extent of binding in terms of $K_{D,app}$ for the multi-state binding models. Compared to the two-state systems, the principles are similar when comparing K_D to K_{Dapp}. That is, the final equilibrium amounts of each species depend only on the final ligand concentration (provided that [L] is in excess of [E]). Thus, the change in fraction bound may be described by the following equation shown below:

$$Y = \frac{[L]}{[L]+K_D} = \frac{[L]/K_D}{L]/K_D+1} = \frac{M}{M+1} \tag{3.123}$$

$$Y = \frac{[L]}{[L]+K_{D,app}} = \frac{[L]/K_{D,app}}{[L]/K_{D,app}+1} = \frac{M}{M+1} \tag{3.124}$$

where Y represents the fraction bound and M is the ligand concentration at some multiple of K_D or $K_{D,app}$. Equation 3.123 applies to the simple two-state binding model in Figure 1C while equation 3.124 applies to the three-state binding models in Figure 2 (excluding the sequential binding model). To derive the optimal concentration for the dilution experiment, the change in fraction bound is shown below:

$$\Delta Y = [E_{bound,0}] - [E_{bound,eq}] \tag{3.125}$$

$$\Delta Y = \frac{M_0}{M_0 + 1} - \frac{M_{eq}}{M_{eq} + 1} \tag{3.126}$$

Where $[E_{bound,0}]$ and M_0 represent the initial bound enzyme concentration and the initial ligand concentration at some multiple of K_D or $K_{D,app}$, respectively. The $[E_{bound,eq}]$ and M_{eq} represent the final bound enzyme concentration and the final ligand concentration at sum multiple of K_D or $K_{D,app}$. For experiments performed in the stopped-flow instrument, the initial ligand concentration is equal to the final ligand concentration multiplied by the dilution faction, D (remember L is diluted, so $L_{initial} > L_{final}$). Thus equation 3.126 can be rewritten as follows:

$$\Delta Y = \frac{DM_{eq}}{DM_{eq} + 1} - \frac{M_{eq}}{M_{eq} + 1} = \frac{D[L_{final}]/K_D}{D[L_{final}]/K_D + 1} - \frac{[L_{final}]/K_D}{[L_{final}]/K_D + 1} \tag{3.127}$$

The change in fraction bound is a function of M, which increases with ligand concentration to a maximum change in fraction bound followed by a decrease as the ligand concentration (or multiple of K_D or $K_{D,app}$) increases. Thus, the maximum change in absorbance occurs when the derivative of equation 3.127 is equal to zero, shown below:

$$\frac{d(\Delta Y)}{dM} = 0 \tag{3.128}$$

$$\frac{\frac{D}{M^2}(D-1)-D^2+D}{\left(\frac{1}{M}+1+D+DM\right)^2}=0 \tag{3.129}$$

Solving for M results in the following:

$$M=\sqrt{\frac{1}{D}} \tag{3.130}$$

Convert M back to the final ligand concentration:

$$M=\sqrt{\frac{1}{D}}=\frac{[L_{final}]}{K_D} \quad or \quad \frac{[L_{final}]}{K_{D,app}} \tag{3.131}$$

$$[L_{final}]=K_D\sqrt{\frac{1}{D}} \quad or \quad K_{D,app}\sqrt{\frac{1}{D}} \tag{3.132}$$

$$[L_0]=[L_{final}]D=K_D\sqrt{D} \quad or \quad K_{D,app}\sqrt{D} \tag{3.133}$$

Thus, the concentration at which the maximum change in fraction occurs at a starting ligand concentration described by equation 3.133.

The ligand concentration range where the change in fraction bound for the dilution experiment exceeds the observed change from the dilution experiment is described below. This occurs for ligand concentrations below the concentration where the change in fraction bound for the dilution and binding experiments are equal, as shown below:

$$\Delta Y_{dilution} = \Delta Y_{binding} \tag{3.134}$$

$$\frac{D[L_{final}]/K_{D,app}}{D[L_{final}]/K_{D,app}+1}-\frac{[L_{final}]/K_{D,app}}{[L_{final}]/K_{D,app}+1}=\frac{[L_{final}]/K_{D,app}}{[L_{final}]/K_{D,app}+1} \tag{3.135}$$

Isolating expression 3.135 for [L_{final}] yields the following:

$$\left[L_{final}\right] = K_{D,app}\left(\frac{D-2}{D}\right) \tag{3.136}$$

$$[L_0] = \left[L_{final}\right]D = K_{D,app}(D-2) \tag{3.137}$$

Thus, the signal (observed change in fraction bound) for the dilution experiment exceeds the signal the binding experiment when the initial ligand concentration is less than the concentration defined by equation 3.137. Here a few summary points may be made regarding the dilution experiment: **1**) the maximum change in fraction bound increases with the dilution factor (see equation 3.127), **2**) the maximum change in fraction bound occurs at an initial ligand concentration defined by the K_{Dapp} and dilution factor (see equation 3.133), **3**) the ligand concentration range where the observed change in fraction bound exceeds the binding experiment increases with the dilution factor (see equation 3.137; i.e. the range is wider for larger dilution factors), and **4**) the observed change in fraction bound exceeds the binding experiment only for dilution factors greater than 2 (see equation 3.137).

3.5 Results

Simulations for different kinetic cases for each model are presented here. The first three sets of test cases highlight the kinetic and thermodynamic behavior of each model for three 'limiting cases': rapid first step, rapid second step, and stationary intermediate. The 'limiting cases' illustrate how widely the k_{obs} and amplitude behavior can vary depending on the relative magnitudes of the rate constants. The second set of examples include two cases where IF and CS have nearly identical k_{obs} behavior despite having different dissociation and isomerization rate constants. The equivalent k_{obs} examples illustrate how incorporation of amplitude analysis may help distinguish IF from CS. The last set of examples illustrates the dilution experiment as a

method to amplify the change in signal in an effort to characterize k_{obs} (and amplitude) behavior at ligand concentrations that are otherwise inaccessible (due to low change in fraction bound) in the traditional binding experiment. Each set of examples are presented in individual sections below. However, a few generalizations are made here. To avoid confusion, the following general comments are written using the general case nomenclature.

The first set of examples includes rapid equilibration of either the first or second step which results in apparent uncoupling. This is evident in the simplified k_{obs} expressions (equations 3.98 – 3.101 or Table 3.7) where the k_{fast} expression contains terms that only describe the faster equilibrium (i.e. the simplified k_{fast} expression resembles the analogous 2-state k_{obs} expression and does not contain rate constants from the slower equilibration step). However, because the slower step is coupled to the system by the intermediate enzyme state, the k_{slow} expression contains terms related to the slower equilibration step but with an equilibration factor that describes the relative buildup of the intermediate state in terms of the faster equilibrium (equations 3.99 and 3.101 or Table 3.7). More specifically, the rate involved with the depletion of the intermediate state is multiplied by the equilibration factor that resembles the 2-state equilibrium definitions for the intermediate state. For example, when the first step is faster, the 'forward' k_{23} rate is multiplied by the equilibration factor in equation 3.99 ($k_{12}/(k_{12}+k_{21})$), which defines the equilibrium amounts of E2 for a two-state equilibrium between E1 and E2 (E1 ⇄ E2; see $[E2_{eq}]$ in equation 3.26). When the second step is faster, the 'reverse' k_{21} rate is multiplied by the scalar in equation 3.101 ($k_{32}/(k_{23}+k_{32})$), which defines the equilibrium amount of E2 for a two-state equilibrium between E2 and E3 (E2 ⇄ E3; similar to $E1_{eq}$ in equation 3.26 with the following substitutions E2 = E1, E3 = E2, $k_{23} = k_{21}$, $k_{32} = k_{21}$). More succinctly, the rate associated with the depletion of the

intermediate (k_{21} or k_{23}) is multiplied by an equilibration factor that represents the 'effective' concentration of the intermediate state.

The stationary intermediate case, in contrast to the rapid equilibration examples, differs considerably among the models. With the exception of the SR model (which contains the unbound enzyme state in the intermediate state), formation of the intermediate state is negligible. That is, the intermediate enzyme state is rapidly formed then rapidly depleted to form either terminal enzyme states. To clarify, with the exception of the SR model, the change in E1 is approximately equal and opposite of the change in E3. Thus, the experimenter may believe they are observing the following 2-state system shown below:

$$E1 \underset{k_{31,app}}{\overset{k_{13,app}}{\rightleftarrows}} E3 \qquad (3.138)$$

$$where\ \ k_{13,app} = \frac{k_{12}k_{23}}{k_{21} + k_{23}} \quad and \quad k_{31,app} = \frac{k_{21}k_{32}}{k_{21} + k_{23}}$$

In addition, depending on the model, there are differences in the likelihood of observing multi-exponential kinetics. This is discussed in detail for each model in the sections that follow.

The next set of examples include two cases where the IF and CS models are indistinguishable on the basis of k_{obs} behavior. The k_{obs} equivalent cases include intrinsic rate constants on similar orders of magnitude, in contrast to the limiting cases. The two cases have been discussed extensively by Di Cera and coworkers in the context of k_{obs} behavior. Here, discussion of the two cases has been expanded to include amplitude analysis. Importantly, inclusion of amplitude analysis can afford unambiguous discrimination between the IF and CS models when $[E^*_0]$ (in the absence of ligand) << $[E_{total}]$. Notably, some examples exist where equivalent k_{obs} *and* amplitude behavior may be expected resulting from unique instances where the signal contribution (e.g. extinction coefficients or quantum yields) for the ensemble of unbound or bound

enzyme states are varied for the IF and CS models, respectively. However, it is important to note that the equivalent k_{obs} *and* amplitude cases are specific to only two instances of signal variation and all other possible scenarios are distinguishable provided $[E^*_0]$ (in the absence of ligand) << $[E_{total}]$.

The final set of examples highlights a dilution or dissociation experimental setup wherein bound enzyme is diluted, thus shifting the equilibrium towards ligand dissociation from the initially bound enzyme complex. The experimental setup expands the experimental range to include lower ligand concentrations that are inaccessible in the traditional mixing or binding experimental setup due to low change in fraction bound. These experiments may be useful in cases where adequate curvature in the k_{slow} behavior for the IF, CS and SR models is not observed or for when k_{fast} is beyond the limits of detection at higher ligand concentration. Importantly, the k_{obs} behavior is unaltered by the experimental setup (binding vs dissociation). The amplitude behavior is discussed in detail in a section below.

3.5.1 Induced fit: limiting cases

The limiting cases for the IF model are shown in Figure 3.3 and the simplified k_{obs} expressions are presented in Table 3.7. In addition, the upper and lower limits for the k_{obs} expressions are presented in Table 3.2. A few general comments can be made upon examination of the limiting case behavior in Figure 3.3: **1**) the lower limit of k_{slow} and k_{fast} varies depending on the relative magnitudes of k_{off}, k_f and k_r, **2**) k_{slow} increases hyperbolically with an upper limit equal to $k_f + k_r$, **3**) the build-up of the ensemble of bound enzyme states (EL and E*L) varies hyperbolically with respect to K_{Dapp} but are scaled by different maximum value expressions which are related to the isomerization step, **4**) when both EL and E*L contribute equally to the observed

signal, the %fast phase always approaches 100% at high ligand concentrations, and **5**) the simplified k_{obs} expressions only apply to all ligand concentration ranges in the rapid binding step example. To clarify, when k_{on}*[L] is in the simplified k_{slow} expression (as in the rapid isomerization or the stationary intermediate cases), the k_{slow} eventually approaches k_{fast} and the simplified k_{obs} expressions no longer describe the behavior. Regarding the behavior of sum or product of the k_{slow} and k_{fast} expressions, two noteworthy features are worth mentioning (see Table 3.2): **6**) the sum of k_{slow} and k_{fast} varies linearly with increasing ligand concentrations, and **7**) the product of k_{slow} and k_{fast} varies linearly with increasing ligand concentrations.

Rapid step 1: rapid binding

The rapid binding step case represents a likely scenario for small molecules binding to enzyme wherein diffusion limited binding and dissociation kinetics are expected (Figure 3.3A). In this scenario, k_{slow} and %fast phase increases hyperbolically while the k_{fast} varies linearly with increasing ligand concentration, shown in the top plot of Figure 3A. This represents the only case where the simplified k_{obs} expressions adequately describe the kinetic behavior across all ligand concentration ranges. When $k_{off} >> k_f + k_r$, the lower limits of k_{slow} and k_{fast} approach k_r and k_{off}, respectively. In addition, the curvature in the k_{slow} behavior overlays with the %fast phase behavior. Notably, in this kinetic regime, the k_{slow} and %fast phase behavior varies hyperbolically with K_D (not $K_{D,app}$!). Importantly, $K_D > K_{D,app}$ for the IF binding mechanism. This feature is important for the comparison to the CS rapid binding case, which also varies with respect to K_D, however, $K_D < K_{D,app}$ for the CS binding mechanism.

The kinetic profiles of the transient concentrations of each enzyme state are shown in the middle four plots of the Figure 3A. Addition of ligand to unbound enzyme results in a rapid

'equilibration' of the fast process (relating to k_{fast}) from the rapid depletion of E and rapid formation of EL. What follows is a slow isomerization process and redistribution of the bound enzyme states with respect to $K_{Dapp,}$ as shown in the Eadie-hofstee plot (bottom plot of figure 3A). In the fast phase, the EL complex is rapidly formed hyperbolically with respect to K_D and approaches saturation ([EL] nears $[E_{total}]$) at low time points followed by slow conversion to the E*L state dictated by K_{Dapp}. If the observed signal results from an equal contribution of both EL and E*L (i.e. EL + E*L), then the %fast phase will vary hyperbolically with respect to K_D.

Rapid step 2: rapid isomerization

In some instances, rapid isomerization occurs and greatly exceeds the binding rate constants (Figure 3.3B). Notably, the simplified k_{obs} expressions only apply to low ligand concentration ranges. In addition, only the k_{slow} phase is observable in this ligand range. This is because the rapid isomerization process, which follows the slow formation of the [EL] complex, is limited by k_{slow} and therefore is kinetically 'invisible'. The experimenter observes single-exponential kinetics that behaves linearly with respect to ligand concentration with slope equal to k_{on} and intercept equal to k_{off} multiplied by an equilibration factor in terms of the isomerization step. However, as ligand concentration is increased, eventually k_{slow} (which contains k_{on}*[L] terms) approaches k_{fast} resulting in a window wherein both kinetic phases are observable. This window occurs because the upper and lower limits of k_{fast} and k_{slow}, respectively, are both equal to $k_f + k_r$, and thus k_{fast} and k_{slow} reach similar orders of magnitude, resulting in two observable kinetic phases. Notably, the width of the window and the ligand range depend on the k_{on}, highlighting the importance of data collection across a range of ligand concentrations. Unfortunately, because the lower limit of k_{fast} is equal to the upper limit of k_{slow}, the double-exponential window eventually

collapses, and the observed kinetics return to monophasic kinetics with slope equal to k_{on} and intercept equal to k_{off} multiplied by an equilibration factor in terms of the isomerization step. When a multi-state binding mechanism is expected from alternative experimental evidence, one may attempt to lower the isomerization kinetics by decreasing the experimental temperature or by addition of viscogen to decrease the gap in the intrinsic rate constants for the isomerization and binding events thus decreasing the likelihood of single-exponential kinetics. Unfortunately, this kinetic case is difficult to identify.

Stationary intermediate:

The stationary intermediate case presents a wide range wherein double-exponential kinetics are observable (Figure 3.3C). Like the rapid isomerization case, the simplified k_{obs} expressions only applies to the range wherein k_{slow} (which contains k_{on}*[L] terms) is orders of magnitude less than k_{fast}. Similarly, in this low ligand range, the kinetics are single-exponential and dominated by k_{slow}. As k_{slow} approaches k_{fast}, double exponential kinetics are observed, however, the window is wider than the rapid isomerization case because the lower limit of k_{fast} ($k_{off} + k_f$) and the upper limit of k_{slow} ($k_f + k_r$) are not equal. The midpoint of $\%A_{fast}$ occurs near K_D which is greater than $K_{D,app}$, but can vary depending on the relative magnitudes of the rate constants.

Notes on the limits of k_{slow} and k_{fast}:

Many descriptions of the increase in k_{slow} are rationalized by the ligand dependent buildup of the intermediate state, [EL], with formation of the intermediate by k_{on}*[L] as the 'driving' force for the increase in k_{slow}. However, no appreciable buildup of [EL] occurs in the stationary

intermediate case, yet k_{slow} still increases. Another counterintuitive point is that k_{slow} (and k_{fast}) behavior is unaltered by the direction of change (i.e. k_{obs} is identical for the binding and dilution experiment even though [EL] is formed or depleted with either experiment). So how can k_{slow} be explained if the ligand dependent depletion or formation (i.e. k_{on}*[L] description) of [EL] is not the sole determinant?

At low ligand concentrations, k_{on}*[L] approaches 0 and the kinetic system condenses to the following:

$$E \underset{k_{off}}{\longleftarrow} EL \underset{k_r}{\overset{k_f}{\rightleftarrows}} E^*L \tag{3.139}$$

where the equilibrium lies towards formation of E. Here, k_f is competing with k_{off} and k_r for formation of E. For any equilibration process, the kinetics are governed by the forward and reverse processes contributing to the favored product. Thus, k_f is always included in one of the limits with additive contribution of k_{off} and k_r which are scaled differently depending on the relative magnitudes of the three rate constants, explicitly stated by the equation in Table 3.2. However, when the isomerization event is relatively uncoupled from the dissociation event (i.e. when the rates are on different orders of magnitude), the following limits are expected: the lower limit of k_{fast} tends towards a sum k_f and the faster rate between k_{off} and k_r, while the lower limit of k_{slow} tends towards the slower rate between k_{off} and k_r scaled by an equilibration factor. For the <u>rapid binding case</u>, $k_{off} >> k_f + k_r$. In this case, the lower limit of k_{fast} approaches k_{off} (k_{fast} actually approaches $k_f + k_{off}$, but since $k_f << k_{off}$, $k_{fast} \sim k_{off}$). The lower limit of k_{slow} approaches k_r scaled by an equilibration factor, $k_{off}/(k_f + k_{off})$. For the <u>rapid isomerization case</u>, $k_{off} << k_f + k_r$. In this case, the lower limit of k_{fast} approaches $k_f + k_r$, and the lower limit of k_{slow} approaches k_{off} scaled by an equilibration factor related to effective [EL], $k_r/(k_f + k_r)$. For the <u>stationary intermediate case</u>, $k_{off} + k_f >> k_r$. In this case, the lower limit of k_{fast} approaches $k_{off} + k_f$, and the lower limit of k_{slow}

approaches k_r scaled by an equilibration factor, $k_{off}/(k_f + k_{off})$. In cases where the forward and reverse processes, are similar in magnitude, the limits of k_{fast} and k_{slow} become more complicated and include different contributions of the three rate constants: k_f, k_r, and k_{off}, as shown in Table 3.2 (see $\lim_{L\to 0} k_{obs1,2}$**)**.

At high ligand concentrations, k_{on}*[L] approaches infinity and the kinetic system condenses to the following:

$$E \xrightarrow{\infty} EL \underset{k_r}{\overset{k_f}{\rightleftarrows}} E^*L \tag{3.140}$$

where the binding process rapidly outcompetes the isomerization process. Thus, the upper limit of k_{fast} rate approaches infinity while the upper limit of k_{slow} approaches the sum of $k_f + k_r$, which both contribute to the formation of the desired products, EL and E*L. Examination of the lower and upper limits of k_{slow} reveals that k_{slow} increases for all conditions of the IF model. Notably, $\%A_{fast}$ always increases to 100% when the signal contribution from EL and E*L are equal.

3.5.2 Conformational selection: limiting cases

The limiting cases for the CS model are shown in Figure 3.4 and the simplified k_{obs} expressions are presented in Table 3.7. In addition, the upper and lower limits for the k_{obs} expressions are presented in Table 3.3. A few general comments can be made upon examination of the limiting case behavior in Figure 3.4: **1**) the lower limit of k_{slow} and k_{fast} varies depending on the relative magnitudes of k_{off} relative to the sum of k_f and k_r, where the lower limit of k_{slow} is equal to the lesser of k_{off} or k_f+k_r and the lower limit of k_{fast} is equal to the larger of k_{off} or k_f+k_r **2**) k_{slow} has an upper limit equal to k_f, **3**) a decrease in k_{slow} is unambiguous proof of CS, **4**) k_{slow} only decreases under the following conditions: a) $k_{off} > k_f + k_r$ or b) $k_{off} < k_f + k_r$ and $k_{off} < k_f$ **5**) the

depletion of the ensemble of unbound enzyme states (E and E*) varies hyperbolically with respect to K_{Dapp} but are scaled by different maximum value expressions related to the isomerization step, **6**) when E*L contributes to the observed signal, the %fast phase always approaches $\%E^*_0$ (in the absence of ligand) at high ligand concentrations and two kinetic phases are always observable provided $[E_{total}] >> [E^*_0] >> 0$, and **7**) the simplified k_{obs} expressions only apply to all ligand concentration ranges in the rapid binding step example. To clarify, when $k_{on}*[L]$ is in the simplified k_{slow} expression (as in the rapid isomerization or the stationary intermediate cases), the k_{slow} eventually approaches k_{fast} and the simplified k_{obs} expressions no longer describe the behavior. Regarding the behavior of sum or product of the k_{slow} and k_{fast} expressions, two noteworthy features are worth mentioning (see Table 3.3): **8**) the sum of k_{slow} and k_{fast} varies linearly with increasing ligand concentrations, and **9**) the product of k_{slow} and k_{fast} varies linearly with increasing ligand concentrations.

Rapid step 1: rapid isomerization

This case is similar to the rapid isomerization case of the induced fit scenario (Figure 3.4A). However, the upper limits of CS k_{slow} approach k_f (not $k_f + k_r$ as in the IF case). Similar to the IF case, the k_{slow} eventually increases to the upper limit k_f (which is now no longer orders of magnitude below k_{fast}) and the observation of two phases is possible. However, in contrast to the IF model, the two phases will be clearly distinguishable as ligand concentration continues to increase because the upper and lower limits of k_{slow} and k_{fast}, respectively, are not equal. In other words, k_{slow} and k_{fast} are resolvable at infinite ligand concentrations, in contrast to the IF binding mechanism wherein only a brief window of multi-exponential kinetics exists.

Rapid step 2: rapid binding

This is the 'textbook case' of CS (Figure 3.4B) wherein decreasing k_{slow} behavior is unambiguous proof of CS binding kinetics. There exist two conditions where CS may result in a hyperbolic decrease in k_{obs} behavior, one of which includes the given condition here where $k_{off} > k_f + k_r$. In this case, the lower limit of k_{slow} equals $k_f + k_r$ and the upper limit equals k_f. Like the IF case of rapid binding, the %fast phase perfectly overlaps with the k_{slow} behavior with a hyperbolic decrease with respect to K_D (again, not $K_{D,app}$!). However, in contrast to the IF model, $K_D < K_{Dapp}$ for the CS model. The decrease in observed k_{obs} with increasing ligand concentration may seem unintuitive, so this is discussed in detail below.

Stationary intermediate:

The stationary intermediate case presents a wide range wherein double-exponential kinetics are observable (Figure 3.4C). The behavior is similar to the stationary intermediate case presented for the IF model but with different upper and lower limits of the k_{obs} expressions. The midpoint of $\%A_{fast}$ occurs near K_D which is less than $K_{D,app}$, but can vary depending on the relative magnitudes of the rate constants. However, in contrast to the IF model, the fast phase may not be observable because the upper limit of $\%A_{fast}$, which approaches $\%E^*_0$ is very low in the stationary intermediate case because $k_r >> k_f$, so $\%E^*_0$ which equals $k_f/(k_f+k_r)$ tends towards 0. However, the hyperbolic behavior in the observed k_{slow} rate clues the experimenter in to the fact that multi-state kinetics are at play. Distinguishing CS from IF in this case is best done from global analysis of kinetic data across a range of ligand concentrations with simultaneous k_{obs} and amplitude analysis.

Notes on the limits of k_{slow} and k_{fast}:

In contrast to the IF model, whose lower limits may include a contribution from the k_{off}, k_f, and k_r rates, the CS k_{obs} lower limits contain discrete lower limits relating to the isomerization or dissociation process. At low ligand concentrations, k_{on}*[L] approaches 0 and the kinetic system condenses to the following:

$$E \underset{k_r}{\overset{k_f}{\rightleftarrows}} E^* \underset{k_{off}}{\longleftarrow} E^*L \tag{3.141}$$

where the equilibrium lies towards formation of E and E* governed by the isomerization rate constants. Here, k_{off} is perturbs the equilibrium between E and E*, established by k_f and k_r. For any equilibration process, the kinetics are governed by the forward and reverse processes contributing to the favored products. Thus, k_f is always included in one of the limits. Because only k_f and k_r contribute to formation of both E *and* E*, the lower limits are expected to be either k_{off} or the sum of $k_f + k_r$. The lower limit of k_{fast} tends towards the faster of k_{off} and the sum $k_f + k_r$, while the lower limit of k_{slow} tends towards slower of k_{off} and the sum $k_f + k_r$.

At high ligand concentrations, k_{on}*[L] approaches infinity and the kinetic system condenses to the following:

$$E \underset{k_r}{\overset{k_f}{\rightleftarrows}} E^* \overset{\infty}{\longrightarrow} E^*L \tag{3.142}$$

where the binding process rapidly outcompetes the isomerization process. Thus, the upper limit of k_{fast} rate approaches infinity while the upper limit of k_{slow} approaches the k_f, which is the only rate that contributes to the formation of the desired product, E*L. That is, at high ligand concentrations, the formation of E*L outcompetes the formation of E and k_f becomes rate limiting. Examination of the lower and upper limits of k_{slow} reveals that k_{slow} behavior can increase or decrease depending on the relationship between k_{off} and $k_f + k_r$ in addition to the relationship between k_{off} and k_f. Notably, k_{slow} decreases under the following conditions: **a**) $k_{off} > k_f + k_r$ or **b**) $k_{off} < k_f + k_r$ and k_{off}

< k_f. Case A is shown in Figure 3.4B and both k_{slow} and $\%A_{fast}$ decrease with ligand concentration. Case B, not shown, differs from Case A wherein k_{slow} decrease while $\%A_{fast}$ increases with ligand concentration, which helps differentiate the identity of the lower k_{slow} and k_{fast} for case A and case B.

Additional CS cases are worth noting here. The first occurs when when **c**) $k_{off} < k_f + k_r$ and $k_{off} = k_f$, then k_{slow} remains constant. The final related cases occur when $k_{off} = k_f + k_r$ and **d**) $k_f > k_f$, **e**) $k_f = k_f$, **f**) $k_f < k_f$. For cases D – F, the lower limits of $k_{fast} = k_{slow} = k_{off}$ or $k_f + k_r$, while the lower limit of $\%A_{slow}$ approaches 50%. Depending on the relationship between k_f and k_r, the $\%A_{fast}$ (which approaches $\%E^*_0$ in the absence of ligand, equal to $100*k_f/(k_f + k_r)$), can have different behaviors. More specifically, $\%A_{fast}$ either increases (when $k_f > k_r$), remains constant ($k_f = k_r$) or decreases (when $k_f < k_r$). Thus, the CS case can result in a variety of k_{obs} and $\%A_{fast}$ behaviors.

3.5.3 Sequential binding: limiting cases

The limiting cases for the SB model are shown in Figure 3.5 and the simplified k_{obs} expressions are presented in Table 3.7. In addition, the upper and lower limits for the k_{obs} expressions are presented in Table 3.4. A few general comments can be made upon examination of the limiting case behavior in Figure 3.5: **1**) the lower limit of k_{slow} and k_{fast} varies depending on the relative magnitudes of k_{off1} and k_{off2}, **2**) both k_{slow} and k_{slow} increase linearly with respect to either k_{on1} and k_{on2}, **3**) the formation of ELL results in the depletion of EL and the Eadie-Hofstee plots for the binding process are no longer linear as seen for the IF and CS models, Regarding the behavior of sum or product of the k_{slow} and k_{fast} expressions, two noteworthy features are worth mentioning (see Table 3.4): **5**) the sum of k_{slow} and k_{fast} varies linearly with increasing ligand concentrations, and **6**) the product of k_{slow} and k_{fast} takes the form of a second degree polynomial.

The sequential binding mechanism is distinguishable from the IF and CS mechanisms from all points mentioned above.

Rapid step 1: rapid binding step 1

This case represents the scenario wherein the first binding step is fast compared to the second ligand binding step (Figure 3.5A). Depending on the relative magnitudes of rate constants, the $\%A_{fast}$ behavior can increase or decrease. However, it is unlikely to confuse the binding mechanism with CS which can have similar $\% A_{fast}$ behavior, for the reasons mentioned above. In addition, for this particular case, a window wherein two observable rates are present followed by a collapse to single exponential kinetics. Again, it is unlikely to confuse the binding mechanism with IF which can have similar behavior, for the reasons mentioned above. In this particular case, one is most likely to measure k_{fast} which has a slope equal to k_{on2} at high ligand concentrations with intercept equal to k_{off2}. One may arrive at the other rate constants using the $K_{D,app}$ expression with k_{on2} and k_{off2} fixed to the values from the kinetic experiment in combination with fits to equilibrium binding data and/or global kinetic analysis with expressions describing k_{obs} and amplitude behavior.

Rapid step 2: rapid binding step 2

This case represents the scenario wherein the second binding step is fast compared to the first ligand binding step (Figure 3.5B). For this scenario, it is unlikely to observe multi-exponential kinetics and the experimenter will most likely only observe k_{slow} which has a slope equal to k_{on1} at high ligand concentrations with intercept equal to k_{off1}. As with the rapid step 1 case, one may similarly arrive at the k_{on2} and k_{off2} rate constants by simultaneously fitting kinetic with equilibrium

binding data and/or global kinetic analysis with expressions describing k_{obs} and amplitude behavior.

Stationary intermediate

The stationary intermediate case resembles single-exponential kinetics (Figure 3.5C). For this scenario, the experimenter will most likely only observe k_{slow} which has a slope equal to k_{on1} at high ligand concentrations with intercept equal to k_{off2}. As with the previous limiting cases, one may similarly arrive at the k_{on1} and k_{off1} rate constants by simultaneously fitting kinetic with equilibrium binding data and/or global kinetic analysis with expressions describing k_{obs} and amplitude behavior.

Notes on the limits of k_{slow} and k_{fast}:

The lower limits for the SB model contain discrete lower limits relating to the k_{off1} and k_{off2} rate constants. At low ligand concentrations, k_{on1}*[L] and k_{on2}*[L] approach 0 and the kinetic system condenses to the following:

$$E \underset{k_{off1}}{\longleftarrow} EL \underset{k_{off2}}{\longleftarrow} ELL \tag{3.143}$$

where the equilibrium lies towards formation of E. No rate is competing with the formation of E by k_{off1} and k_{off2}, so the lower limits are expected to be either of the two dissociation rate constants. The lower limit of k_{fast} tends towards the faster of k_{off1} and k_{off2}, while the lower limit of k_{slow} tends towards slower of k_{off1} and k_{off2}.

At high ligand concentrations, k_{on1}*[L] and k_{on2}*[L] approach infinity and the kinetic system condenses to the following:

$$E \overset{\infty}{\longrightarrow} EL \overset{\infty}{\longrightarrow} ELL \tag{3.144}$$

where the upper limits of k_{fast} and k_{slow} approach infinity.

3.5.4 Substrate reorientation: limiting cases

The limiting cases for the SR model are shown in Figure 3.6 and the simplified k_{obs} expressions are presented in Table 3.7. In addition, the upper and lower limits for the k_{obs} expressions are presented in Table 3.5. A few general comments can be made upon examination of the limiting case behavior in Figure 3.7: **1**) the lower limit of k_{slow} and k_{fast} varies depending on the relative magnitudes of k_{off1} and k_{off2}, **2**) k_{slow} increases hyperbolically with an upper limit equal to a ratio of all the rate constants described in Table 3.5, **3**) the build-up of the ensemble of bound enzyme states (EL and EL*) varies hyperbolically with respect to K_{Dapp} but are scaled by different maximum value expressions (see EL_{max} and E^*L_{max} in Table 3.5), **4**) when both EL and EL* contribute equally to the observed signal, the %fast phase always approaches 100% at high ligand concentrations, and **5**) due to the symmetry of the model, the limiting cases for rapid step 1 and rapid step 2 are mirrors of each other. Regarding the behavior of sum or product of the k_{slow} and k_{fast} expressions, two noteworthy features are worth mentioning (see Table 3.5): **6**) the sum of k_{slow} and k_{fast} varies linearly with increasing ligand concentrations, and **7**) the product of k_{slow} and k_{fast} varies linearly with increasing ligand concentrations. The model envisions two binding orientations of ligand wherein ligand must leave the enzyme active site and reorient externally prior to rebinding in order to achieve a different binding orientation. Often, this model can be ruled out for small molecules however, as the ligand becomes larger (MW > 500 g/mol), one should consider the substrate reorientation model.

Rapid step 1: rapid binding step 1

This case represents the scenario wherein the first binding step is fast compared to the second ligand binding step (Figure 3.6A). The behavior resembles the IF binding mechanism for the comparable scenario. Notably, as with the Di Cera examples wherein equivalent IF and CS k_{obs} behavior exists, so too do cases where IF and SR yield equivalent k_{obs} behavior. This is discussed in detail in a later section. However, for the majority of scenarios, the SR binding mechanism is distinguishable from the IF mechanism when global analysis of k_{obs} and amplitude behavior is applied to kinetic data.

Rapid step 2: rapid binding step 2

This case mirrors the rapid step 1 scenario due to the symmetry in the binding mechanism. To clarify, Figure 3.6B matches Figure 3.6A but EL and E*L are swapped. Due to the symmetry of the binding mechanism, the simplified expressions for k_{slow} and k_{fast} in Table 3.7 are similar.

Stationary intermediate

The stationary intermediate case resembles single-exponential kinetics across all ligand concentration ranges (Figure 3.6C). For this scenario, the experimenter will most likely only observe k_{fast} which has a slope equal to the sum of k_{on1}+k_{on2} and intercept equal to slower of k_{off1} or k_{off2}, which deviates from the simplified expression which predicts the intercept is equal to 0.

Notes on the limits of k_{slow} and k_{fast}:

The lower limits for the SR model contain discrete lower limits relating to the k_{off1} and k_{off2} rate constants. At low ligand concentrations, k_{on1}*[L] and k_{on2}*[L] approach 0 and the kinetic system condenses to the following:

$$EL \underset{}{\overset{k_{off1}}{\longrightarrow}} E \underset{k_{off2}}{\longleftarrow} E^*L \tag{3.145}$$

where the equilibrium lies towards formation of E. No rate is competing with the formation of E by k_{off1} and k_{off2}, so the lower limits are expected to be either of the two dissociation rate constants. The lower limit of k_{fast} tends towards the faster of k_{off1} and k_{off2}, while the lower limit of k_{slow} tends towards slower of k_{off1} and k_{off2}.

At high ligand concentrations, k_{on1}*[L] and k_{on2}*[L] approach infinity and the kinetic system condenses to the following:

$$EL \underset{\infty}{\longleftarrow} E \overset{\infty}{\longrightarrow} E^*L \tag{3.146}$$

where the upper limits of k_{fast} and k_{slow} approach infinity.

3.5.5 Persistent Heterogeneity

The limiting cases for the PH model are not shown in a figure because the model consists of two simple unlinked 2-state binding models. However, the k_{obs} and amplitude value expressions, are summarized in Table 3.6. In addition, simplifications of the k_{obs} expressions that were performed for the 3-state models were not required for the PH model, thus, the two k_{obs} expressions in Table 3.7 are identical to the k_{obs} expressions in Table 3.6. Two enzyme conformers in the PH model will be referred to as the E conformer (E and EL states) or the E* conformer (E* and E*L states). A few general comments can be made upon examination of the amplitude and k_{obs}

expressions: **1**) the amplitude and k_{obs} expressions are identical to those for the general 2-state binding system (i.e. PH consists of two uncoupled 2-state systems), **2**) the k_{obs} have linear dependence with respect to ligand concentration, **3**) the build-up or depletion of the ensemble of unbound, E or E*, and bound enzyme states, EL and EL*, vary hyperbolically with respect to K_{D1} or K_{D2} as shown in Table 3.6, **4**) the $K_{D,app}$, which is the concentration at which half of the total enzyme is bound is sensitive to the initial amounts of each enzyme conformer ($Enzyme_{total}$ = E + E* + EL + E*L), and **5**) when both EL and EL* contribute equally to the observed signal, the %phase contribution for each phase approaches the percentage of the respective enzyme conformer at saturating ligand concentrations. Regarding the behavior of sum or product of the k_{slow} and k_{fast} expressions, two noteworthy features are worth mentioning (see Table 3.6): **6**) the sum of k_{slow} and k_{fast} varies linearly with increasing ligand concentrations, and **7**) the product of k_{slow} and k_{fast} takes the form of a second-degree polynomial.

3.5.6 IF vs CS: equivalence in k_{obs}

The 'k_{obs} equivalent' cases for the IF and CS models are shown in Figure 3.7. Case 1 and 2 in section 3.4.4 are shown in the top and bottom panels of Figure 3.7, respectively. Because the k_{obs} plots are nearly identical, they are excluded from this figure. Notably, the IF and CS have similar $K_{D,app}$ (same thermodynamic endpoints in the kinetic curves), which also adds another layer of complexity when distinguishing the models by k_{obs} alone. Plots of the %fast phase contributions vs ligand concentration, bound species signal vs time at low (below and approaching saturating concentration) and high ligand concentrations (orders of magnitude above K_{Dapp}) are shown in the left, middle and right column respectively. Each row differs by the extinction coefficients that contribute to the signal, which are specified in the figure legend. In rows A and D, the extinction

coefficients for the bound species equally contribute to the signal. In rows and B and E, the extinction coefficients for the IF case remain equal to 1 while the extinction coefficients for the CS case are varied to yield a matching %amplitude plot compared to the IF model. In rows C and F, the extinction coefficients for the CS case remain equal to 1 while the extinction coefficients for the IF case are varied to yield a matching %amplitude plot compared to the CS model.

The signal for each extinction coefficient case are calculated as follows. First, Beer's law is assumed, thus signal is proportional to the enzyme state concentration multiplied by the extinction coefficient. For the IF case: signal = $[EL]\varepsilon_{EL} + [E^*L]\varepsilon_{E^*L}$. However, the only way to vary CS extinction coefficients to match the IF case requires using 'signal' from the depletion of unbound enzyme: signal = $([E_0]^*\varepsilon_E + [E^*_0]^*\varepsilon_{E^*})-([E]\varepsilon_E + [E^*]\varepsilon_{E^*})$ where $[E_0]=k_r/(k_f+k_r)$ and $[E^*_0]=k_f/(k_f+k_r)$. To clarify, the 'binding signal' is opposite the depletion of the unbound state, shifted so that at time zero 'bound signal' is equal to zero. Note that the signal is normalized to 1 (see the y-axis for the differences in the magnitude of signal).

The CS and IF models are difficult to distinguish when $\%E^*_0$ ($\%E^*_0 = 100^*k_f/(k_f+k_r)$), which is the upper limit of %fast phase for the CS case, is near to 100%, which is the upper limit for the IF case. This is illustrated by similar kinetic traces for Case 1 compared to Case 2 for the examples where the extinction coefficients for the IF and CS case are equal to 1 (rows A and D). However, under saturating conditions ([Ligand] >> $K_{D,app}$), it becomes easier to distinguish the models. This is because eventually the forward rate becomes limiting and the remaining portion of the signal (100% - $\%E^*_0$) is equal to k_f. That is, at infinite concentrations of ligand, there will always be a visible slow phase in the CS model (see right most plot in rows A and D). Thus, utilization of the amplitude and k_{obs} expressions for the data analysis allows one to distinguish the models with a higher degree of confidence.

In the remaining examples, the extinction coefficients are varied such that identical k_{obs} *and* amplitude plots may be expected for the CS and IF cases. Importantly, the only way that an equivalent amplitude plot may be observed for the CS case involves measurement of the depletion of the unbound species (i.e. the bound signal is equal to the negative value of the depletion of the unbound enzyme), thus making it somewhat easier to distinguish the two models (rows B and E). Notably, the CS case only matches the IF case when the terminal unbound enzyme state, E, has an extinction coefficient equal to 0. Thus, the 'bound signal' is a result of the depletion of E*. Elimination of either model requires knowledge of the species contribution to the signal (e.g. elimination of the CS model on the grounds of detectible signal from the E state). This is also true for the case where the extinction coefficients for the IF model are varied to match the CS %fast phase amplitude plots (rows C and F). Depending on the $\%E^*_0$, the extinction coefficients for the EL and E*L states may vary greatly. Again, knowledge of the relative contribution of the bound species for the IF case, simplifies the decision between models.

3.5.7 IF vs SR: equivalence in k_{obs}

The 'k_{obs} equivalent' cases for the IF and SR models are shown in Figure 3.8. Case 1 and 2 in section 3.4.5 are shown in the top and bottom panels of Figure 3.8, respectively. Both cases are equivalent to the cases presented in Figure 3.7 (except now the SR equivalent case is overlayed with the IF case). Because the k_{obs} plots are nearly identical, they are excluded from this figure. As with the equivalent IF and CS cases, the IF and SR have similar $K_{D,app}$ (same thermodynamic endpoints in the kinetic curves). Plots of the %fast phase contributions vs ligand concentration, bound species signal vs time and Eadie-Hofstee plots are shown in the left, middle and right column respectively. Because the %fast phase behavior has the same upper and lower limits

between both models, the %fast phase vs ligand concentration and bound species signal vs time plots are nearly perfectly overlayed. Examination of the Eadie-Hofstee plots reveals a potentially distinguishing feature: the build-up of the ensemble of bound enzyme states (EL and E*L for IF or EL and EL* for SR) varies hyperbolically with respect to K_{Dapp} but are scaled by different maximum value expressions which are summarized in Tables 3.2 and 3.5. Importantly, for the equivalent IF and SR cases, the maximum value of the bound species are never identical between models. Hence, it is possible that one may be able to distinguish the IF and SR models provided knowledge of the equilibrium amounts of the bound species.

The SR model represents an often overlooked but potentially equivalent model to interrogate when hyperbolic profiles for k_{slow} are observed. However, the ligand binding mechanism requires ligand dissociation followed by reorientation of the ligand (likely remote from the active site) before rebinding. If the ligand is particularly small such that reorientation within the active site may occur, then the ligand binding is less likely to occur via SR. Thus, assignment of the ligand binding mechanism requires the intuition of the experimenter to aid in distinguishing of the models.

3.5.8 CS vs SR: equivalence in k_{obs}

The 'k_{obs} equivalent' cases for the CS and SR models are shown in Figure 3.9. Case 1 and 2 in section 3.4.6 are shown in the top and bottom panels of Figure 3.9, respectively. Both cases are equivalent to the cases presented in Figure 3.7 (except now the SR equivalent case is overlayed with the CS case). Because the k_{obs} plots are nearly identical, they are excluded from this figure. As with the equivalent IF and CS cases, the CS and SR have similar $K_{D,app}$ (same thermodynamic endpoints in the kinetic curves). Plots of the %fast phase contributions vs ligand concentration,

bound species signal vs time at low (below and approaching saturating concentration) and high ligand concentrations (orders of magnitude above K_{Dapp}) are shown in the left, middle and right column respectively. Many of the same things can be said for the comparison of CS and SR that have been said for the k_{obs} equivalent cases for IF and CS. Specifically: **1**) CS becomes difficult to distinguish from SR when $\%E^*_0$ ($\%E^*_0 = 100*k_f/(k_f+k_r)$), which is the upper limit of %fast phase for the CS case, is near to 100%, which is the upper limit for the SR case, **2**) the CS equivalent %fast phase plots may only occur when the signal arises from the depletion of the ligand free enzyme states and the extinction coefficient for E is equal to 0 (as in Figures 3.7B and 3.7E for the comparison of the equivalent IF and CS cases), **3**) knowledge of the observable species simplifies the distinguishing between the CS and SR models. Finally, as stated for the IF vs SR comparison (section 3.5.7), the SR model represents an often overlooked but potentially equivalent model to interrogate when hyperbolic profiles for k_{slow} are observed.

3.5.9 Dilution

In this work, a dilution experiment is presented as a method to expand the experimental range to a lower ligand concentration range that is otherwise inaccessible through the traditional binding experiment (due to low signal). This is particularly useful, even necessary, for systems with very fast on rates (or even off rates, which typically represent the lower limit of k_{fast}). The dilution experiment takes advantage of the fact that the observed rates of change are unaltered by the direction of change (binding vs dilution). That is, the k_{obs} expressions do not contain terms for the enzyme concentration. The experimental setup includes diluting ligand-bound enzyme into ligand-free buffer, which results in a re-equilibration of the system towards the unbound complex ($[L_0] > [L_{final}]$ where initial ligand concentration is defined by: $[L_0] = [L_{final}]$*dilution factor) with

kinetics defined by [L_{final}]. The magnitude of the re-equilibration process is defined by the change in ligand concentration and the initial amounts of enzyme in each state (i.e. [$E1_0$], [$E2_0$], and [$E3_0$] which may represent different bound and unbound enzyme states). As will be demonstrated in this section, the amplitudes of change are greater for the dilution experiment (relative to the comparable binding experiment) when the initial ligand concentration is near $[L_0]=K_{D,app}\sqrt{\text{Dilution factor}}$, which defines the maximum change in fraction bound for the dilution experiment. Thus, the dilution experiment serves as a means of signal amplification at low ligand concentrations. This becomes useful for situations when the traditional binding experiment yield the following results: **1**) no observable curvature in the k_{slow}, and **2**) k_{fast} approaches the limits of detection. By lowering the experimental ligand concentration, the dilution experiment allows one to measure binding kinetics at near-zero ligand concentrations which: **1**) increases the likelihood to observe curvature in k_{slow} and **2**) increases the likelihood to capture the lower limits of k_{fast} (note: the lower limits of k_{fast} takes on different forms: k_{off}, $k_f + k_r$, or something more complex; see Tables 3.1 – 3.6 for the lower limits of k_{fast}).

The results from simulations for the comparison of the binding (black lines) and dilution (red lines, dilution factor = 10) experiments for the five different double-exponential ligand binding models are presented in Figure 3.10. Each row contains results for each binding model which include plots of the change in fraction bound, amplitude values, and %fast phase vs ligand concentration (scaled by $K_{D,app}$) in each column. The rate constants for each binding model are specified in the figure legend. Notably, the kinetic cases for the IF, CS, SB, and SR models match one of the limiting test cases in Figures 3.3 – 3.6 (specified in the Figure 3.10 legend).

The change in fraction bound for each binding model display similar behavior between the different models. Specifically, the change in fraction bound for the dilution experiment exceeds

the observed change for the binding experiment when the ligand concentration is below the intersection of both curves. In fact, identical plots are observed for the IF, CS, and SR models. Similar behavior for the SB and PH models is observed, however, the maximum value for the dilution experiment and the point of intersection between both experiments can vary depending on the relative binding affinities for the first and second binding event (i.e. K_{D1} and K_{D2}). In addition, the relative amounts of each enzyme state (E_{total} and E^*_{total}) may alter the maximum value and intercept for the PH model. For the dilution factor of 10, the dilution experiment yields a higher change in fraction bound for the dilution experiment when the initial ligand concentration is less than $8*K_{D,app}$. Thus, the dilution experiment will amplify the change in fraction bound for final ligand concentrations below $0.8*K_{D,app}$ ($[L_{final}] = [L_0]$/dilution factor; also recall that k_{obs} is dependent on $[L_{final}]$). To clarify, the dilution experiment amplifies the change in fraction bound for each model below the following final ligand concentrations: IF = 0.18 mM, CS = 0.72 mM, SB = 0.28 mM, SR = 0.25 mM, and PH = 0.51 mM. Examination of the k_{obs} and %fast phase plots for the relevant case in Figures 3.3 – 3.6 reveals that the dilution experiment range contains regions before either plot (k_{obs} or %fast phase) plateaus, particularly for the CS and SR models (Figures 3.4B and 3.6A). This is extremely useful for models where $K_D < K_{Dapp}$ (e.g. CS, SB, SR, PH) because the curvature in k_{slow} or %fast phase tends to vary around K_D (see detailed discussion for each kinetic case in sections 3.5.1 – 3.5.5).

The amplitude behavior and the resulting change in %fast phase plots highlight the relative changes in the pre-exponential factors for the fast and slow phases in the binding and dilution simulations. Moreover, the amplitude plots also further illustrate the differences in amplitude contribution for each model (different curvature, max values, etc). Strikingly, the %fast phase plots are identical for the binding and dilution experiments for the IF, CS, and SR models. However,

different behaviors may be expected for the SB and PH models depending on the relative KD values (for both models) and/or the relative amounts of each enzyme state (E_{total} and E^*_{total}, for the PH model). Importantly, global analysis of binding and dilution data takes the amplitude contribution into account and helps with model differentiation, particularly when data from both experiments are simultaneously fit.

3.6 Discussion

Different mechanisms of ligand binding were presented with the goal of providing a quantitative and qualitative understanding of limiting cases for each model. The solutions for k_{obs} and amplitude expressions are unique to each model. However, there are examples where the k_{obs} or the amplitudes are nearly identical between different ligand binding models. Importantly, the ligand binding mechanisms may be sufficiently resolved when the full solutions describing each model are applied to kinetic data. In addition, a new experimental approach, the dilution experiment, may increase the likelihood of distinguishing the different binding mechanisms. Here, the interpretation of the different ligand binding mechanisms in the context of a free-energy landscape of enzyme conformers are discussed in relation to functional properties such as binding promiscuity and catalytic promiscuity.

3.6.1 The ‘kinetically silent’ ensemble of unbound states

The binding mechanism for ligand binding involving an ensemble of unbound enzyme states may produce apparently 2-state binding kinetics (single-exponential process) for scenarios wherein an isomerization event has no bearing on the ligand ‘selection’ process (in addition to the rapid isomerization case in Figure 3.4A (when $k_r << k_f$) and the stationary intermediate case

discussed in Figure 3.3A). Consider the case where 3 different enzyme conformations constitute an ensemble of unbound enzyme states with 3 different ligand binding 'selection criteria' illustrated on the left, top, and right face of the enzyme (Figure 3.11A). The scenario with a 'kinetically silent' ensemble is illustrated in Figure 3.11B. In this example, the binding competent conformation persists in all enzyme states (binding site located on the top face of the different enzyme states, Figure 3.11A). The isomerization processes are irrelevant to the binding mechanism because ligand does not 'select' the other binding modes. Stated differently, while the isomerization processes may occur throughout the binding process, they remain 'invisible' to the kinetics provided the three different enzyme conformations have the same binding properties. Thus, the ligand binding kinetics would produce single-exponential kinetics and present as an apparently 2-state binding mechanism described by Figure 3.11B.

The 'kinetically silent' ensemble scenario offers a potential explanation (in addition to the stationary intermediate case in Figure 3.4C) for situations where one observes single exponential kinetics but may have expected multi-state or multi-exponential binding kinetics when structural evidence strongly suggests an ensemble of unbound enzyme conformations. To clarify, the lack of a CS component in the binding mechanism does not refute the possibility of a pre-existing ensemble of conformers as dictated by the CS binding mechanism. In this respect, ligand 'selects' all enzyme conformers and binds to 100% of the unbound enzyme conformations. Unfortunately for this scenario, recovery of the isomerization rate constants requires interrogation by different ligands (see next section) or alternative experimental approaches.

3.6.2 Rearrangement rates may be ligand dependent

An ensemble of unbound enzyme states may yield apparently different isomerization rates in the ligand binding mechanism for ligands with different 'selection criteria' for the binding competent state. Consider again, the ensemble of 3 enzyme states in the absence of ligand (Figure 3.11A). In this scenario, the binding competent conformers are the E2 and E3 conformers for the ligand 2 (purple ligand) and the E3 conformer for ligand 3 (green ligand), shown in Figures 3.11C and 3.11D, respectively. The binding kinetics may be expected to proceed via double and triple-exponential processes described by the apparent 3-state and 4-state binding mechanisms, respectively (Figure 3.11C and 3.11D). In either case, characterization of the binding mechanism for ligands 2 and 3 yield different rearrangement rates for the same enzyme, which may seem counterintuitive at first. Thus, caution is advised when defining the number of enzyme conformers in the absence of ligand. Importantly, the binding experiment for any individual ligand only reports on the kinetics and equilibria for the enzyme conformers involved in the ligand binding process for that ligand. To clarify, the kinetically determined binding mechanism defines the *minimum* number of ligand-free enzyme states.

3.6.3 Binding mechanism may be ligand dependent

The binding mechanism may be ligand dependent. The previous two sections outline cases where different ligands bind through different binding mechanisms despite interacting with the same ensemble of enzyme states (specifically yielding different CS mechanisms). Though not pictured, one might envision a variety of scenarios wherein different ligand types may induce different isomerization events (IF binding mechanism), multiple binding events (SB binding mechanism), or perhaps different binding orientations (SR binding mechanism). Additionally,

irreversible isomerization of enzyme to a new conformer may result in additional binding properties (PH binding mechanism).

3.6.4 Kinetic and thermodynamic properties of binding interactions may be sensitive to experimental conditions

The experimental conditions may contribute to mechanistic differences (mechanism or rate constants). For example, lipid content, ionic strength, pH, temperature, etc. should be considered when comparing kinetic and thermodynamic data because they may alter the conformational landscape for the accessible enzyme states. In addition, for enzymes that require additional cofactors for catalysis (e.g. Cytochrome P450 reductase or cytochrome b5 for CYPs), it is possible that conformational landscape may be further modified by their binding. However, this would require additional enzyme states, which are outside the scope of this work. The relevance of the different binding models are discussed in the next section.

3.6.5 Functional implications for the different binding models: binding promiscuity and catalytic promiscuity

Regardless of the rates at which binding takes place, the existence of multiple conformations in the absence of ligand can either be an advantage or a disadvantage. Exemplary binding scenarios include the CS and PH binding mechanisms (Figure 3.12). For a ‘traditional’ substrate specific enzyme with a defined substrate that binds to one ‘cognate’ member of a conformational ensemble, the other ‘noncognate’ conformations are a disadvantage. The noncognate conformations are a ‘cost’ to the enzyme because their presence lowers the overall affinity of the ligand for the enzyme, and the noncognate conformation results in the possibility of

inhibition by noncognate ligands that have affinity for these conformations. The presence of noncognate conformations in the ligand-free ensemble of substrate-specific enzymes might be considered as the 'cost' the enzyme must pay in order to have sufficient flexibility to exploit the other mechanistic extreme, induced fit, to achieve specificity for its cognate substrate.

On the other hand, the existence of multiple conformations in the ligand-free state could be a functional advantage for an enzyme that requires substrate promiscuity. For example, an enzyme may evolve to adopt a vast conformational landscape and therefore the ability to accommodate chemically diverse substrates wherein different ligand types selectively bind to different subsets of enzyme conformations. The gain in substrate promiscuity comes at the cost of the interconversion between conformations. In fact, the limiting step at high ligand concentrations is dependent on the forward rearrangement rate and the fraction of the binding competent state present before addition of the ligand.

The existence of multiple ligand-bound states could present a functional advantage or disadvantage for an enzyme that exhibits multi-functionality or catalytic promiscuity. Exemplary binding scenarios include the IF, SB, SR and PH binding mechanisms (Figure 3.12). One can envision a variety of scenarios where heterogeneity among ligand bound states can manifest in complex catalytic properties. The presence of multiple ligand-bound enzyme states can lead to evolutionary advantages wherein new protein function may be attained from different binding orientations or catalytic properties for the different bound states. Of course, this could also present a disadvantage as the adoption of multiple ligation modes potentially increases the likelihood of catalytically inefficient enzyme conformers (IF or PH) or binding poses (SB or SR).

3.6.6 Summary decision tree

A decision tree is provided to aid in early interpretation of kinetic data when double-exponential kinetics are observed. Each of the models discussed contain a linear k_{fast} component, so the best indicators for the different kinetic binding mechanisms include interpretations for the k_{slow} and % amplitude behavior. The decision tree is not exhaustive and it is best to consider the thermodynamic properties ($K_{D,app}$) in relation to the upper and lower limits of k_{fast} and k_{slow}, which are summarized in Tables 3.2 – 3.6. However, the best means of interrogating the different ligand binding models requires global analysis of kinetic binding (and likely dissociation) data with the full solutions to the different binding mechanisms (described in Tables 3.2 – 3.6).

3.7 Figures

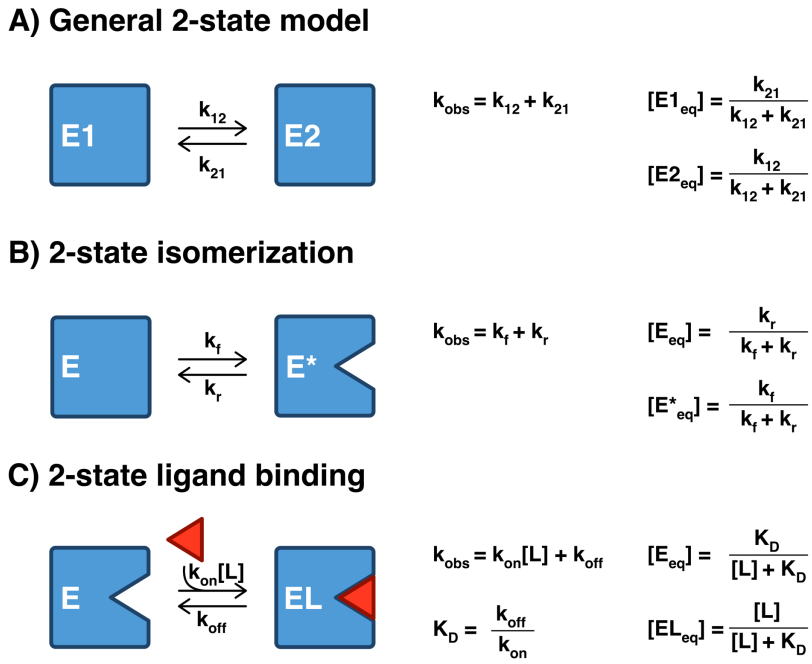


Figure 3.1. Two-state models. The general case represents the simplest two-state reversible model with no indication for the cause of differences between the two enzyme states, E1 and E2. More detailed variants of the general two-state model include a conformational isomerization or ligand binding event. **A)** The general 2-state model depicts the equilibrium between two interconvertible states, E1 and E2. The rate k_{12} describes the rate at which E1 is reversibly converted to E2. The rate k_{21} describes the reverse transition of E2 to E1. Perturbation of the system by an external factor such as pH or temperature results in a mono-exponential re-equilibration process with an observed rate, $k_{obs} = k_{12} + k_{21}$. The equilibrium amounts of each enzyme state, $[E1_{eq}]$ and $[E2_{eq}]$, are governed by the rate constants for the transition, k_{12} and k_{21}, as shown in the equations above. **B)** The 2-state isomerization model envisions a flexible enzyme that is capable of sampling distinct conformations, E and E*. The asterisk is the textual representation of differences in the enzyme conformers. The figure depicts equilibrium between a closed, E, to open, E*, enzyme state. The 2-state isomerization model parallels the general case and the expressions that describe the k_{obs} and equilibrium amounts of each enzyme state are similar to the general case expressions with the following substitutions: E = E1, E* = E2, $k_f = k_{12}$ and $k_r = k_{21}$. **C)** The 2-state binding model depicts differences in the ligation state of the enzyme, E and EL. Ligand, L (depicted as red triangles), binds to the unbound form of enzyme, E, to form a complex, EL. Under pseudo-first order conditions ([L] >> total enzyme), the observed rate has a linear dependence with respect to ligand concentration: $k_{obs} = k_{on}[L] + k_{off}$. The 2-state binding model parallels the general case and the expressions that describe the k_{obs} and equilibrium amounts of each enzyme state are similar to the general case expressions with the following substitutions: E = E1, EL = E2, $k_{on}[L] = k_{12}$ and $k_{off} = k_{21}$. In addition, the expressions describing $[E_{eq}]$ and $[EL_{eq}]$ were further simplified with the expression for binding affinity, $K_D = k_{off}/k_{on}$.

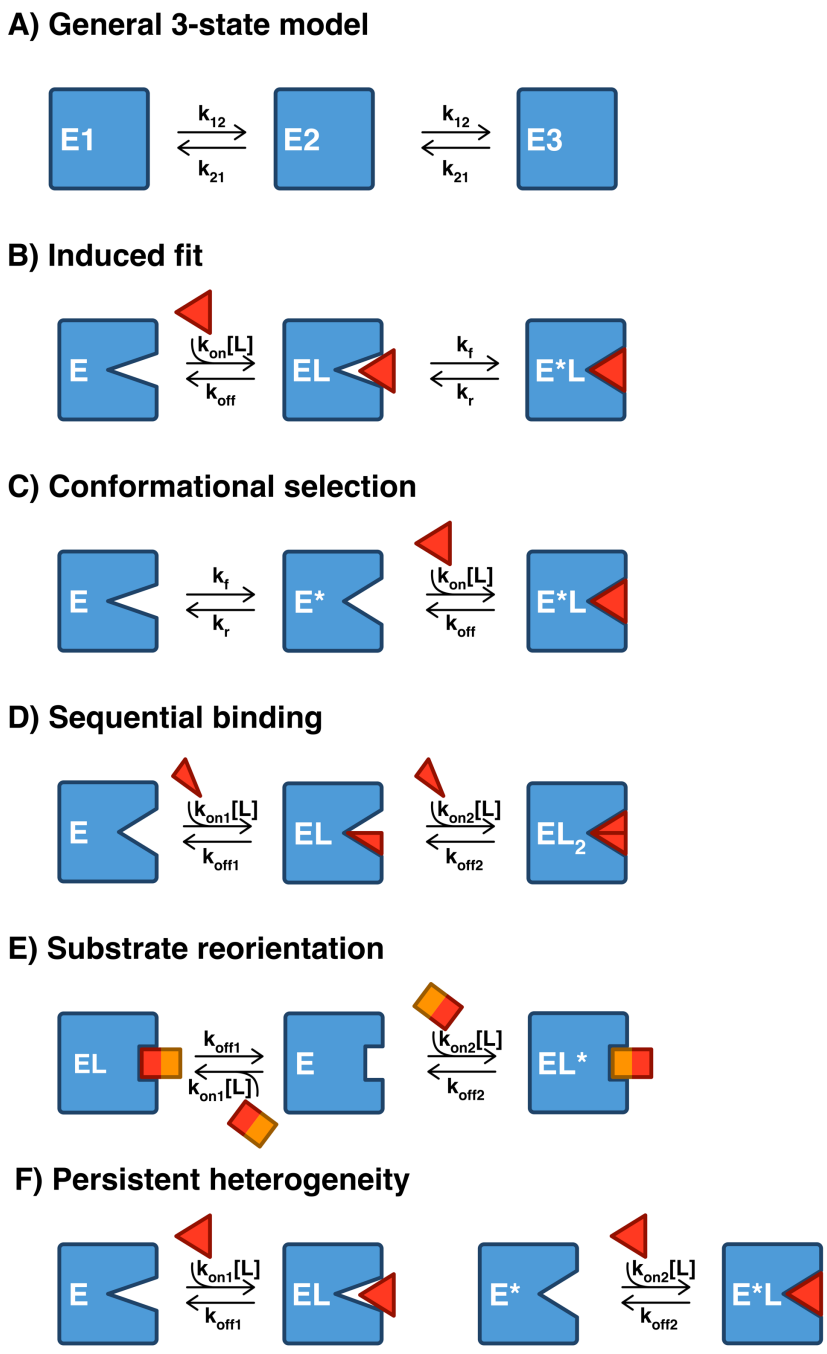


Figure 3.2. Multi-state binding models. The red shapes represent different ligands which have on- and off-rates equal to k_{on} or k_{off}, respectively. When multiple binding steps occur in a model, the rates associated with the first or second binding and dissociation events are noted in the subscripts (e.g. Model D: k_{on1} and k_{on2} represent on-rates for the first and second binding events, respectively). For models with isomerization processes, the forward and reverse rates of the transitions are equal to k_f and k_r, respectively.

Table 3.1. Kinetic and thermodynamic signatures of the 2-state systems. The k_{obs} have units of s^{-1} and the K_D have units of concentration. The PH model, which consists of two 2-state binding models, is summarized in Table 3.6.

2-state isomerization: E ⇄ E* (solution: [C] = [C_{eq}]+ $Ae^{-kobs*t}$)		
K_{eq}:	$\frac{k_f}{k_r}$	
[C_{eq}]:	$[E_{eq}] = \frac{k_r}{k_f + k_r}[E_{total}]$	$[E^*_{eq}] = \frac{k_f}{k_f + k_r}[E_{total}]$
A:	$A_E = \frac{[E_0]k_f - [E^*_0]k_r}{k_f + k_r}$	$A_{E^*} = \frac{-[E_0]k_f + [E^*_0]k_r}{k_f + k_r}$
k_{obs}:	$k_{obs} = k_f + k_r$	
2-state binding: E ⇄ EL (solution: [C] = [C_{eq}]+ $Ae^{-kobs*t}$)		
K_D:	$\frac{k_{off}}{k_{on}}$	
[C_{eq}]:	$[E_{eq}] = \frac{K_D}{[L] + K_D}[E_{total}]$	$[EL_{eq}] = \frac{[L]}{[L] + K_D}[E_{total}]$
A:	$A_E = \frac{[E_0]k_{on}[L] - [EL_0]k_{off}}{k_{on}[L] + k_{off}}$	$A_{EL} = \frac{-[E_0]k_{on}[L] + [EL_0]k_{off}}{k_{on}[L] + k_{off}}$
k_{obs}:	$k_{obs} = k_{on}[L] + k_{off}$	

Table 3.2. Kinetic and thermodynamic signatures of the IF binding model. The $K_{D,app}$ and K_D have units of concentration. The equilibrium concentrations, $[C_{eq}]$, are expressed as a ratio of $[E_{total}]$ where $[E_{total}]$= [E] + [EL] + [E*L]. The two observed rates, k_{slow} and k_{fast}, as well as their lower ([L] →0) and upper ([L] →∞) limits have units of s^{-1}. The coefficient matrix, **K**, is defined in equation 3.52 for the general case and can be applied to the IF model with the following substitutions: $k_{on}[L] = k_{12}$', $k_{off} = k_{21}$, $k_f = k_{23}$, $k_r = k_{32}$, E = E1, EL = E2, and E*L = E3. The matrix, **I**, represents the 3x3 identity matrix.

Induced fit: E ⇄ EL ⇄ E*L (solutions: $[C] = [C_{eq}] + A_{slow}e^{-k_{slow}t} + A_{fast}e^{-k_{fast}t}$)	
$K_{D,app}$:	$K_D\left(\frac{k_r}{k_r + k_f}\right) \qquad where\ K_D = \frac{k_{off}}{k_{on}}$
$[C_{eq}]$:	$[E_{eq}] = \frac{K_{D,app}[E_{max}]}{K_{D,app} + [L]} \qquad [EL_{eq}] = \frac{[L][EL_{max}]}{K_{D,app} + [L]} \qquad [E^*L_{eq}] = \frac{[L][E^*L_{max}]}{K_{D,app} + [L]}$ $where\ E_{max} = [E_{total}], \qquad [EL_{max}] = \frac{k_r}{k_f + k_r}[E_{total}], \qquad [E^*L_{max}] = \frac{k_f}{k_f + k_r}[E_{total}]$
$A_{1,2}$:	$A_{slow} = \frac{\boldsymbol{K}(\boldsymbol{K} - k_{fast}\boldsymbol{I})}{k_{slow}(k_{slow} - k_{fast})} \qquad A_{fast} = \frac{\boldsymbol{K}(\boldsymbol{K} - k_{slow}\boldsymbol{I})}{k_{fast}(k_{fast} - k_{slow})}$ $A_{E1,i} = \frac{-k_{on}[L](k_f + k_r - k_i)[E1_0] + k_{off}(k_r - k_i)[E2_0] + k_{off}k_r[E3_0]}{k_i(k_i - k_j)}$ $A_{E2,i} = -A_{E1,i} - A_{E3,i}$ $A_{E3,i} = \frac{k_{on}[L]k_f[E1_0] - k_f(k_{on}[L] - k_i)[E2_0] - k_r(k_{on}[L] + k_{off} - k_i)[E3_0]}{k_i(k_i - k_j)}$ $when\ i = slow, \quad j = fast \quad and\ vice\ versa$
$k_{obs1,2}$:	$k_{slow} = \frac{b - \sqrt{b^2 - 4ac}}{2} \qquad k_{fast} = \frac{b + \sqrt{a^2 - 4ac}}{2a}$ $where a = 1, b = k_{on}[L] + k_{off} + k_f + k_r, c = k_{on}[L]k_f + k_{on}[L]k_r + k_{off}k_r$ $and \sum k_{obs} = b, \quad and \prod k_{obs} = c$
$\lim_{L\to 0} k_{obs1,2}$:	$k_{slow} = \frac{b - \sqrt{b^2 - 4ac}}{2} \qquad k_{fast} = \frac{b + \sqrt{a^2 - 4ac}}{2a}$ $where\ a = 1, \qquad b = k_{off} + k_f + k_r, \qquad c = k_{off}k_r$
$\lim_{L\to \infty} k_{obs1,2}$:	$k_{slow} = k_f + k_r \qquad k_{fast} = k_{on}\infty + k_{off} = \infty$

Table 3.3. Kinetic and thermodynamic signatures of the CS binding model. The $K_{D,app}$ and K_D have units of concentration. The equilibrium concentrations, $[C_{eq}]$, are expressed as a ratio of $[E_{total}]$ where $[E_{total}] = [E] + [E^*] + [E^*L]$. The two observed rates, k_{slow} and k_{fast}, as well as their lower ($[L] \rightarrow 0$) and upper ($[L] \rightarrow \infty$) limits have units of s^{-1}. The coefficient matrix, **K**, is defined in equation 3.52 for the general case and can be applied to the CS model with the following substitutions: $k_f = k_{12}$, $k_r = k_{23}$, $k_{on}[L] = k_{23}'$, $k_{off} = k_{32}$, E = E1, E* = E2, and E*L = E3. The matrix, **I**, represents the 3x3 identity matrix.

Conformational selection: E ⇄ E* ⇄ EL (solutions: $[C] = [C_{eq}] + A_{slow}e^{-kslow*t} + A_{fast}e^{-kfast*t}$)
$K_{D,app}$: $K_D\left(\frac{k_f + k_r}{k_f}\right) \qquad where\ K_D = \frac{k_{off}}{k_{on}}$
$[C_{eq}]$: $[E_{eq}] = \frac{K_{D,app}[E_{max}]}{K_{D,app} + [L]} \qquad [E^*_{eq}] = \frac{K_{D,app}[E^*_{max}]}{K_{D,app} + [L]} \qquad [E^*L_{eq}] = \frac{[L][E^*L_{max}]}{K_{D,app} + [L]}$ $where\ E_{max} = \frac{k_r}{k_f + k_r}[E_{total}], \qquad [E^*_{max}] = \frac{k_f}{k_f + k_r}[E_{total}], \qquad [E^*L_{max}] = [E_{total}]$
$A_{1,2}$: $A_{slow} = \frac{\boldsymbol{K}(\boldsymbol{K} - k_{fast}\boldsymbol{I})}{k_{slow}(k_{slow} - k_{fast})} \qquad A_{fast} = \frac{\boldsymbol{K}(\boldsymbol{K} - k_{slow}\boldsymbol{I})}{k_{fast}(k_{fast} - k_{slow})}$ $A_{E1,i} = \frac{-k_f(k_{on}[L] + k_{off} - k_i)[E1_0] + k_r(k_{off} - k_i)[E2_0] + k_r k_{off}[E3_0]}{k_i(k_i - k_j)}$ $A_{E2,i} = -A_{E1,i} - A_{E3,i}$ $A_{E3,i} = \frac{k_f k_{on}[L][E1_0] + k_{on}[L](k_f - k_i)[E2_0] - k_{off}(k_f + k_r - k_i)[E3_0]}{k_i(k_i - k_j)}$ $when\ i = slow, \quad j = fast \quad and\ vice\ versa$
$k_{obs1,2}$: $k_{slow} = \frac{b - \sqrt{b^2 - 4ac}}{2} \qquad k_{fast} = \frac{b - \sqrt{a^2 - 4ac}}{2a}$ $where a = 1, b = k_f + k_r + k_{on}[L] + k_{off}, c = k_f k_{on}[L] + k_f k_{off} + k_r k_{off}$ $and \sum k_{obs} = b, \quad and \prod k_{obs} = c$
$\lim_{L \rightarrow 0} k_{obs1,2}$: $k_{slow} = k_f + k_r \quad and \quad k_{fast} = k_{off} \quad when\ k_{off} > k_f + k_r$ $k_{slow} = k_{off} \quad and \quad k_{fast} = k_f + k_r \quad when\ k_{off} < k_f + k_r$
$\lim_{L \rightarrow \infty} k_{obs1,2}$: $k_{slow} = k_f \qquad k_{fast} = k_{on}\infty + k_{off} + k_r = \infty$

Table 3.4. Kinetic and thermodynamic signatures of the SB binding model. The $K_{D,app}$ has units of concentration. The equilibrium concentrations, $[C_{eq}]$, are expressed as a ratio of $[E_{total}]$ where $[E_{total}] = [E] + [EL] + [ELL]$. The two observed rates, k_{slow} and k_{fast}, as well as their lower ($[L] \rightarrow 0$) and upper ($[L] \rightarrow \infty$) limits have units of s^{-1}. The coefficient matrix, **K**, is defined in equation 3.52 for the general case and can be applied to the SB model with the following substitutions: $k_{on1}[L] = k_{12}'$, $k_{off1} = k_{23}$, $k_{on2}[L] = k_{23}'$, $k_{off2} = k_{32}$, E = E1, EL = E2, and ELL = E3. The matrix, **I**, represents the 3x3 identity matrix.

Sequential binding: E ⇄ EL ⇄ ELL (solutions: $[C] = [C_{eq}] + A_{slow}e^{-kslow*t} + A_{fast}e^{-kfast*t}$)	
$K_{D,app}$:	$\frac{-k_{on1}k_{off2} + \sqrt{(k_{on1}k_{off2})^2 + 4k_{on1}k_{off1}k_{on2}k_{off2}}}{2k_{on1}k_{on2}}$
$[C_{eq}]$:	$[E_{eq}] = \frac{k_{off1}k_{off2}[E_{total}]}{c}$ $[EL_{eq}] = \frac{k_{on1}[L]k_{off2}[E_{total}]}{c}$ $[ELL_{eq}] = \frac{k_{on1}k_{on2}[L]^2[E_{total}]}{c}$ $where\ c = k_{on1}k_{on2}[L]^2 + k_{on1}[L]k_{off2} + k_{off1}k_{off2}$
$A_{1,2}$:	$A_{slow} = \frac{\boldsymbol{K}(\boldsymbol{K} - k_{fast}\boldsymbol{I})}{k_{slow}(k_{slow} - k_{fast})}$ $A_{fast} = \frac{\boldsymbol{K}(\boldsymbol{K} - k_{slow}\boldsymbol{I})}{k_{fast}(k_{fast} - k_{slow})}$ $A_{E1,i} = \frac{-k_{on1}[L](k_{on2}[L] + k_{off2} - k_i)[E1_0] + k_{off1}(k_{off2} - k_i)[E2_0] + k_{off1}k_{off2}[E3_0]}{k_i(k_i - k_j)}$ $A_{E2,i} = -A_{E1,i} - A_{E3,i}$ $A_{E3,i} = \frac{k_{on1}k_{on2}[L]^2[E1_0] + k_{on2}[L](k_{on1}[L] - k_i)[E2_0] - k_{off2}(k_{on1}[L] + k_{off1} - k_i)[E3_0]}{k_i(k_i - k_j)}$ $when\ i = slow, \quad j = fast \quad and\ vice\ versa$
$k_{obs1,2}$:	$k_{slow} = \frac{b - \sqrt{b^2 - 4ac}}{2}$ $k_{fast} = \frac{b + \sqrt{a^2 - 4ac}}{2a}$ $where\ a = 1, b = k_{on1}[L] + k_{off1} + k_{on2}[L] + k_{off2}, \quad c = k_{on1}k_{on2}[L]^2 + k_{on1}[L]k_{off2} + k_{off1}k_{off2}$ $and \sum k_{obs} = b, \quad and \prod k_{obs} = c$
$\lim_{L \rightarrow 0} k_{obs1,2}$:	$k_{slow} = k_{off2}$ and $k_{fast} = k_{off1}$ $when\ k_{off1} > k_{off2}$ $k_{slow} = k_{off1}$ and $k_{fast} = k_{off2}$ $when\ k_{off2} > k_{off1}$
$\lim_{L \rightarrow \infty} k_{obs1,2}$:	$k_{slow} = \infty$ $k_{fast} = \infty$

Table 3.5. Kinetic and thermodynamic signatures of the SR binding model. The $K_{D,app}$ has units of concentration. The equilibrium concentrations, $[C_{eq}]$, are expressed as a ratio of $[E_{total}]$ where $[E_{total}] = [EL] + [E] + [EL^*]$. The two observed rates, k_{slow} and k_{fast}, as well as their lower ($[L] \rightarrow 0$) and upper ($[L] \rightarrow \infty$) limits have units of s^{-1}. The coefficient matrix, **K**, is defined in equation 3.52 for the general case and can be applied to the SR model with the following substitutions: $k_{off1} = k_{12}$, $k_{on1}[L] = k_{21}'$, $k_{on2}[L] = k_{23}'$, $k_{off2} = k_{32}$, EL = E1, E = E2, and EL* = E3. The matrix, **I**, represents the 3x3 identity matrix.

Substrate reorientation: EL ⇄ E ⇄ E*L (solutions: $[C] = [C_{eq}] + A_{slow}e^{-kslow*t} + A_{fast}e^{-kfast*t}$)	
$K_{D,app}$:	$\frac{k_{off1}k_{off2}}{k_{off1}k_{on2} + k_{on1}k_{off2}}$
$[C_{eq}]$:	$[EL_{eq}] = \frac{[L][E_{max}]}{K_{D,app} + [L]} \qquad [E_{eq}] = \frac{K_{D,app}[E^*{}_{max}]}{K_{D,app} + [L]} \qquad [E^*L_{eq}] = \frac{[L][E^*L_{max}]}{K_{D,app} + [L]}$ $where\ EL_{max} = \frac{K_{D,app}}{K_{D1}}[E_{total}], \qquad [E_{max}] = [E_{total}], \qquad [E^*L_{max}] = \frac{K_{D,app}}{K_{D2}}[E_{total}]$
$A_{1,2}$:	$A_{slow} = \frac{\boldsymbol{K}(\boldsymbol{K} - k_{fast}\boldsymbol{I})}{k_{slow}(k_{slow} - k_{fast})} \qquad A_{fast} = \frac{\boldsymbol{K}(\boldsymbol{K} - k_{slow}\boldsymbol{I})}{k_{fast}(k_{fast} - k_{slow})}$ $A_{E1,i} = \frac{-k_{off1}(k_{on2}[L] + k_{off2} - k_i)[E1_0] + k_{on1}[L](k_{off2} - k_i)[E2_0] + k_{on1}[L]k_{off2}[E3_0]}{k_i(k_i - k_j)}$ $A_{E2,i} = -A_{E1,i} - A_{E3,i}$ $A_{E3,i} = \frac{k_{off1}k_{on2}[L][E1_0] + k_{on2}[L](k_{off1} - k_i)[E2_0] - k_{off2}(k_{off1} + k_{on1}[L] - k_i)[E3_0]}{k_i(k_i - k_j)}$ $when\ i = slow, \qquad j = fast \quad and\ vice\ versa$
$k_{obs1,2}$:	$k_{slow} = \frac{b - \sqrt{b^2 - 4ac}}{2} \qquad k_{fast} = \frac{b + \sqrt{a^2 - 4ac}}{2a}$ $where\ a = 1, b = k_{off1} + k_{on1}[L] + k_{on2}[L] + k_{off2}, \ c = k_{off1}k_{on2}[L] + k_{off1}k_{off2} + k_{on1}[L]k_{off2}$ $and \sum k_{obs} = b, \quad and \prod k_{obs} = c$
$\lim_{L \rightarrow 0} k_{obs1,2}$:	$k_{slow} = k_{off2} \quad and \quad k_{fast} = k_{off1} \quad when \ k_{off1} > k_{off2}$ $k_{slow} = k_{off1} \quad and \quad k_{fast} = k_{off2} \quad when \ k_{off2} > k_{off1}$
$\lim_{L \rightarrow \infty} k_{obs1,2}$:	$k_{slow} = \frac{k_{off1}k_{on2} + k_{on1}k_{off2}}{k_{on1} + k_{on2}} \qquad k_{fast} = \infty$

Table 3.6. Kinetic and thermodynamic signatures of the PH binding model. The $K_{D,app}$, K_{D1}, and K_{D2} have units of concentration. The two observed rates, k_{obs1} and k_{obs2}, as well as their lower ([L] →0) and upper ([L] →∞) limits have units of s^{-1}.

Persistent Heterogeneity: E ⇄ EL and E* ⇄ E*L (solutions: [C] = [C_{eq}]+ $Ae^{-kobs*t}$)	
K_{Dapp}:	$\frac{-b+\sqrt{b^2-4ac}}{2a}$ $where\ a = E_{total} + E^*{}_{total},\quad b = E_{total}(K_{D2}-K_{D1}) + E^*{}_{total}(K_{D1}-K_{D2}),$ $c = -(E_{total} + E^*{}_{total})(K_{D1} * K_{D2}),\ E_{total} = E + EL,\ E^*{}_{total} = E^* + E^*L$ $K_{D1} = k_{off1}/k_{on1}\ , K_{D2} = k_{off2}/k_{on2}$
[C_{eq}]:	$[E_{eq}] = \frac{K_{D1}[E+EL]}{[L]+K_{D1}}$ $\quad$ $[E^*{}_{eq}] = \frac{K_{D,2}[E^*+E^*L]}{[L]+K_{D2}}$ $[EL_{eq}] = \frac{[L][E+EL]}{[L]+K_{D1}}$ $\quad$ $[E^*L_{eq}] = \frac{[L][E^*+E^*L]}{[L]+K_{D2}}$
A:	$A_E = \frac{[E_0]k_{on1}[L]-[EL_0]k_{off1}}{k_{on1}[L]+k_{off1}}$ $\quad$ $A_{E^*} = \frac{[E^*{}_0]k_{on2}[L]-[E^*L_0]k_{off2}}{k_{on2}[L]+k_{off2}}$ $A_{EL} = \frac{-[E_0]k_{on1}[L]+[EL_0]k_{off1}}{k_{on1}[L]+k_{off1}}$ $\quad$ $A_{E^*L} = \frac{-[E^*{}_0]k_{on2}[L]+[E^*L_0]k_{off2}}{k_{on2}[L]+k_{off2}}$
$k_{obs1,2}$:	$k_{obs1} = \mathrm{k_{on1}[L] + k_{off1}}$ $\quad$ $k_{obs2} = \mathrm{k_{on2}[L] + k_{off2}}$ $\sum k_{obs} = \mathrm{(k_{on1} + k_{on2})[L] + k_{off1} + k_{off2}}$ $\prod k_{obs} = \mathrm{k_{on1}k_{on2}[L]^2 + (k_{on1}k_{off2} + k_{on2}k_{off1})[L] + k_{off1}k_{off2}}$
$\lim_{L\to 0} k_{obs1,2}$:	$k_{obs1} = \mathrm{k_{off1}}$ $\quad$ $k_{obs2} = \mathrm{k_{off2}}$
$\lim_{L\to \infty} k_{obs1,2}$:	$k_{obs1} = \infty$ $\quad$ $k_{obs2} = \infty$

Table 3.7. Simplified observed rate expressions for the limiting cases. The rates for each case shown in the headers use the general case nomenclature: $E1 \underset{k_{21}}{\overset{k_{12}}{\rightleftarrows}} E2 \underset{k_{32}}{\overset{k_{23}}{\rightleftarrows}} E3$

	Rapid first step (k_{12}, k_{21} >> k_{23}, k_{32})	**Rapid second step** (k_{23}, k_{32} >> k_{12}, k_{21})	**Stationary Intermediate** (k_{21}, k_{23} >> k_{12}, k_{32})
IF: E ⇅ EL ⇅ EL	$\boldsymbol{k_{slow}} = k_f\left(\frac{[L]}{[L]+K_D}\right) + k_r$ $\boldsymbol{k_{fast}} = k_{on}[L] + k_{off}$	$\boldsymbol{k_{slow}} = k_{on}[L] + k_{off}\left(\frac{k_f}{k_f+k_r}\right)$ $\boldsymbol{k_{fast}} = k_f + k_r$	$\boldsymbol{k_{slow}} = k_{on}[L]\left(\frac{k_f}{k_{off}+k_f}\right) + k_r\left(\frac{k_{off}}{k_{off}+k_f}\right)$ $\boldsymbol{k_{fast}} = k_{off} + k_f$
CS: E ⇅ E* ⇅ EL	$\boldsymbol{k_{slow}} = k_{on}[L]\left(\frac{k_f}{k_f+k_r}\right) + k_{off}$ $\boldsymbol{k_{fast}} = k_f + k_r$	$\boldsymbol{k_{slow}} = k_f + k_r\left(\frac{K_D}{[L]+K_D}\right)$ $\boldsymbol{k_{fast}} = k_{on}[L] + k_{off}$	$\boldsymbol{k_{slow}} = k_f\left(\frac{k_{on}[L]}{k_r+k_{on}[L]}\right) + k_{off}\left(\frac{k_r}{k_r+k_{on}[L]}\right)$ $\boldsymbol{k_{fast}} = k_r + k_{on}[L]$
SB: E ⇅ EL ⇅ EL_2	$\boldsymbol{k_{slow}} = k_{on2}[L]\left(\frac{[L]}{[L]+K_{D1}}\right) + k_{off2}$ $\boldsymbol{k_{fast}} = k_{on1}[L] + k_{off1}$	$\boldsymbol{k_{slow}} = k_{on1}[L] + k_{off1}\left(\frac{K_{D2}}{[L]+K_{D2}}\right)$ $\boldsymbol{k_{fast}} = k_{on2}[L] + k_{off2}$	$\boldsymbol{k_{slow}} = k_{on1}[L]\left(\frac{k_{on2}[L]}{k_{on2}[L]+k_{off1}}\right) + k_{off2}\left(\frac{k_{off1}}{k_{on2}[L]+k_{off1}}\right)$ $\boldsymbol{k_{fast}} = k_{off1} + k_{on2}[L]$
SR: EL ⇅ E ⇅ EL*	$\boldsymbol{k_{slow}} = k_{on2}[L]\left(\frac{K_{D1}}{[L]+K_{D1}}\right) + k_{off2}$ $\boldsymbol{k_{fast}} = k_{off1} + k_{on1}[L]$	$\boldsymbol{k_{slow}} = k_{off1} + k_{on1}[L]\left(\frac{K_{D2}}{[L]+K_{D2}}\right)$ $\boldsymbol{k_{fast}} = k_{on2}[L] + k_{off2}$	$\boldsymbol{k_{slow}} = k_{off1}\left(\frac{k_{on2}}{k_{on1}+k_{on2}}\right) + k_{off2}\left(\frac{k_{on1}}{k_{on1}+k_{on2}}\right)$ $\boldsymbol{k_{fast}} = k_{on1}[L] + k_{on2}[L]$
PH: E ⇅ EL, E^* ⇅ E*L	$\boldsymbol{k_{slow}} = k_{on2}[L] + k_{off2}$ $\boldsymbol{k_{fast}} = k_{on1}[L] + k_{off1}$	$\boldsymbol{k_{slow}} = k_{on1}[L] + k_{off1}$ $\boldsymbol{k_{fast}} = k_{on2}[L] + k_{off2}$	Not applicable (the systems are not coupled by an intermediate).

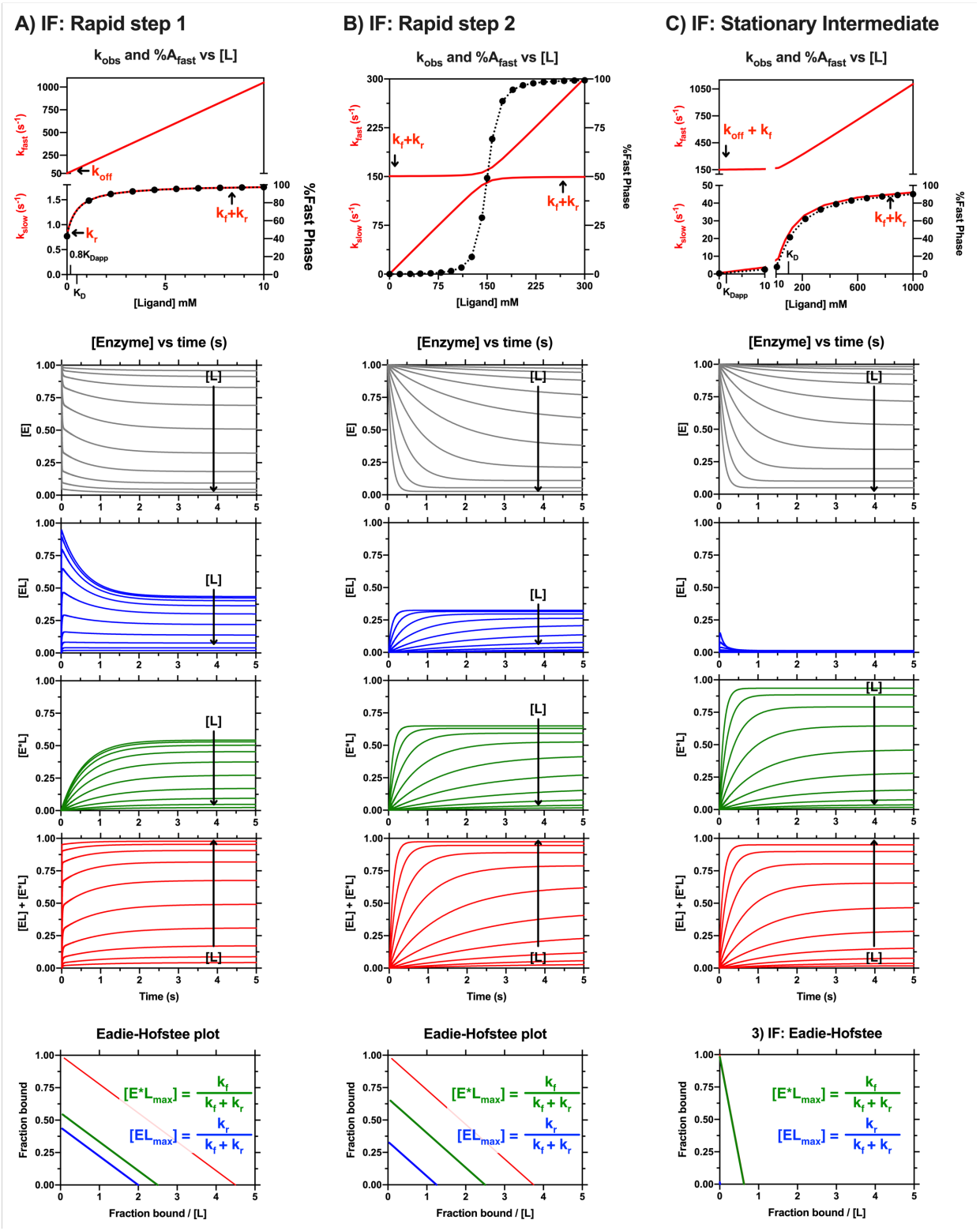


Figure 3.3. Induced Fit: test cases. Summary of limiting cases for the IF model. The simplified k_{obs} expressions for each case are summarized in Table 3.7. The top row contains plots of the k_{obs} vs [L] (red lines) and the %Fast phase vs [L] (black dots and dotted line). The middle four rows

contain plots of the transient concentrations of each enzyme species vs time: [E] = grey line, [EL] = blue line, [E*L] = green line, [EL] + [E*L] = red line. The bottom row contains the Eadie-Hofstee plots for each bound state: [EL] = blue line, [E*L] = green line, [EL] + [E*L] = red line. A) Rapid binding: k_{on} = 100 mM^{-1} s^{-1}, k_{off} = 50 s^{-1}, k_f = 1 s^{-1}, k_r = 0.8 s^{-1}, $K_{D,app}$ = 0.22 mM, K_D = 0.5 mM. This case was used for the dilution experiment so 0.8*K_{Dapp} and K_D are indicated on the x-axis. B) Rapid isomerization: k_{on} = 1 mM^{-1} s^{-1}, k_{off} = 0.8 s^{-1}, k_f = 100 s^{-1}, k_r = 50 s^{-1}, $K_{D,app}$ = 0.27 mM, K_D = 0.8 mM. The K_{Dapp} and K_D are not indicated on the x-axis because they are low compared to the concentration range. C) Stationary intermediate: k_{on} = 1 mM^{-1} s^{-1}, k_{off} = 100 s^{-1}, k_f = 50 s^{-1}, k_r = 0.8 s^{-1}, $K_{D,app}$ = 1.6 mM, K_D = 100 mM. The K_{Dapp} and K_D are indicated on the x-axis.

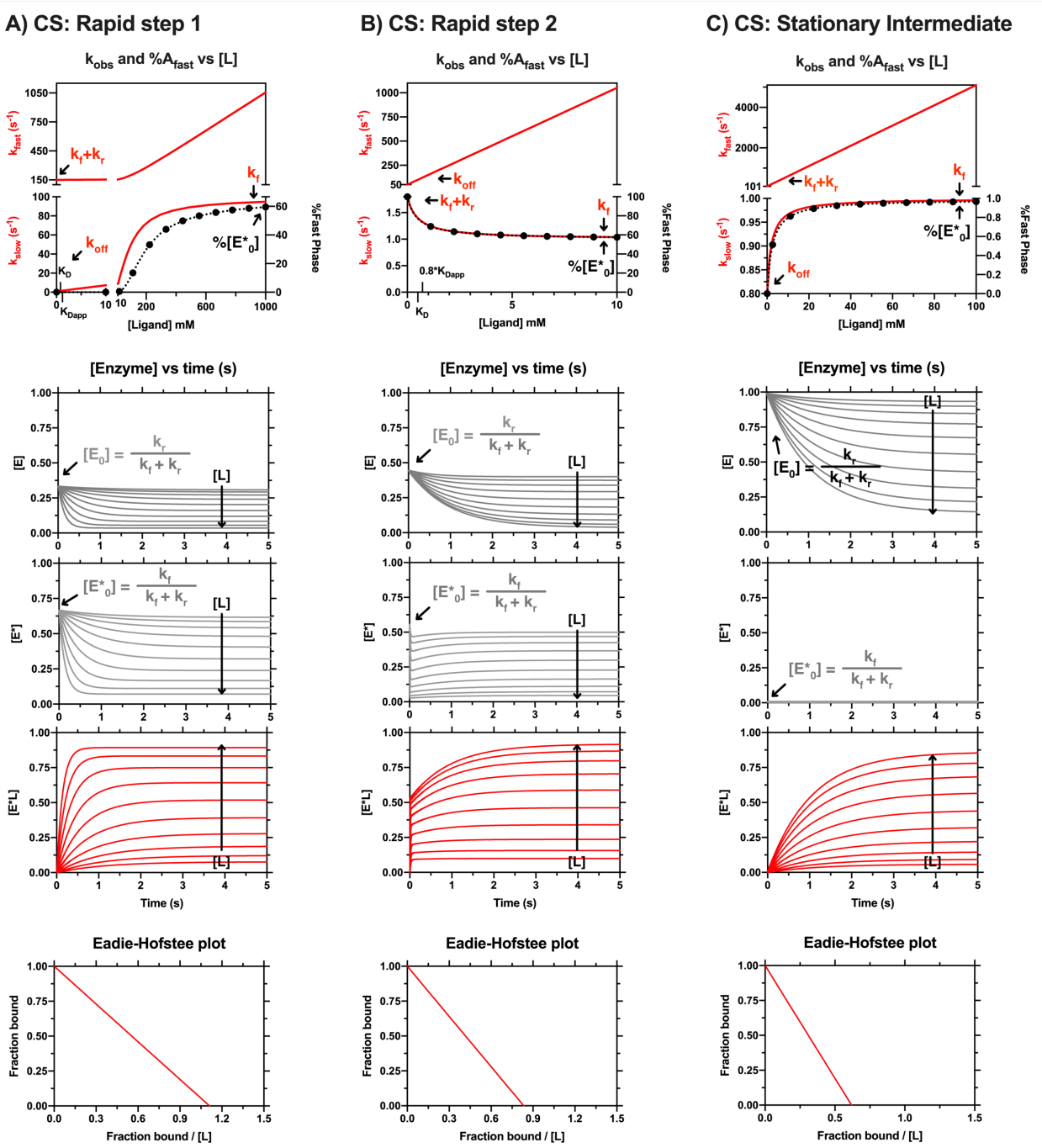


Figure 3.4. Conformational selection: test cases. Summary of limiting cases for the CS model. The simplified k_{obs} expressions for each case are summarized in Table 3.7. The top row contains plots of the k_{obs} vs [L] (red lines) and the %Fast phase vs [L] (black dots and dotted line). The middle three rows contain plots of the transient concentrations of each enzyme species vs time: [E] = grey line, [E*] = light grey, [E*L] = red line. The bottom row contains the Eadie-hofstee plots for the bound state: [E*L] = red line. A) Rapid isomerization: $k_f = 100\ s^{-1}$, $k_r = 50\ s^{-1}$, $k_{on} = 1\ mM^{-1}\ s^{-1}$, $k_{off} = 0.8\ s^{-1}$, $K_{D,app} = 1.2$ mM, $K_D = 0.8$ mM. The K_{Dapp} and K_D are indicated on the x-axis. B) Rapid binding: $k_f = 1\ s^{-1}$, $k_r = 0.8\ s^{-1}$, $k_{on} = 100\ mM^{-1}\ s^{-1}$, $k_{off} = 50\ s^{-1}$, $K_{D,app} = 0.9$ mM, $K_D = 0.5$ mM. This case was used for the dilution experiment so $0.8*K_{Dapp}$ and K_D are indicated on

the x-axis. C) Stationary intermediate: $k_f = 1\ s^{-1}$, $k_r = 100\ s^{-1}$, $k_{on} = 50\ mM^{-1}\ s^{-1}$, $k_{off} = 0.8\ s^{-1}$, $K_{D,app} = 1.62$ mM, $K_D = 0.016$ mM. The K_{Dapp} and K_D are not indicated on the x-axis because they are low compared to the concentration range.

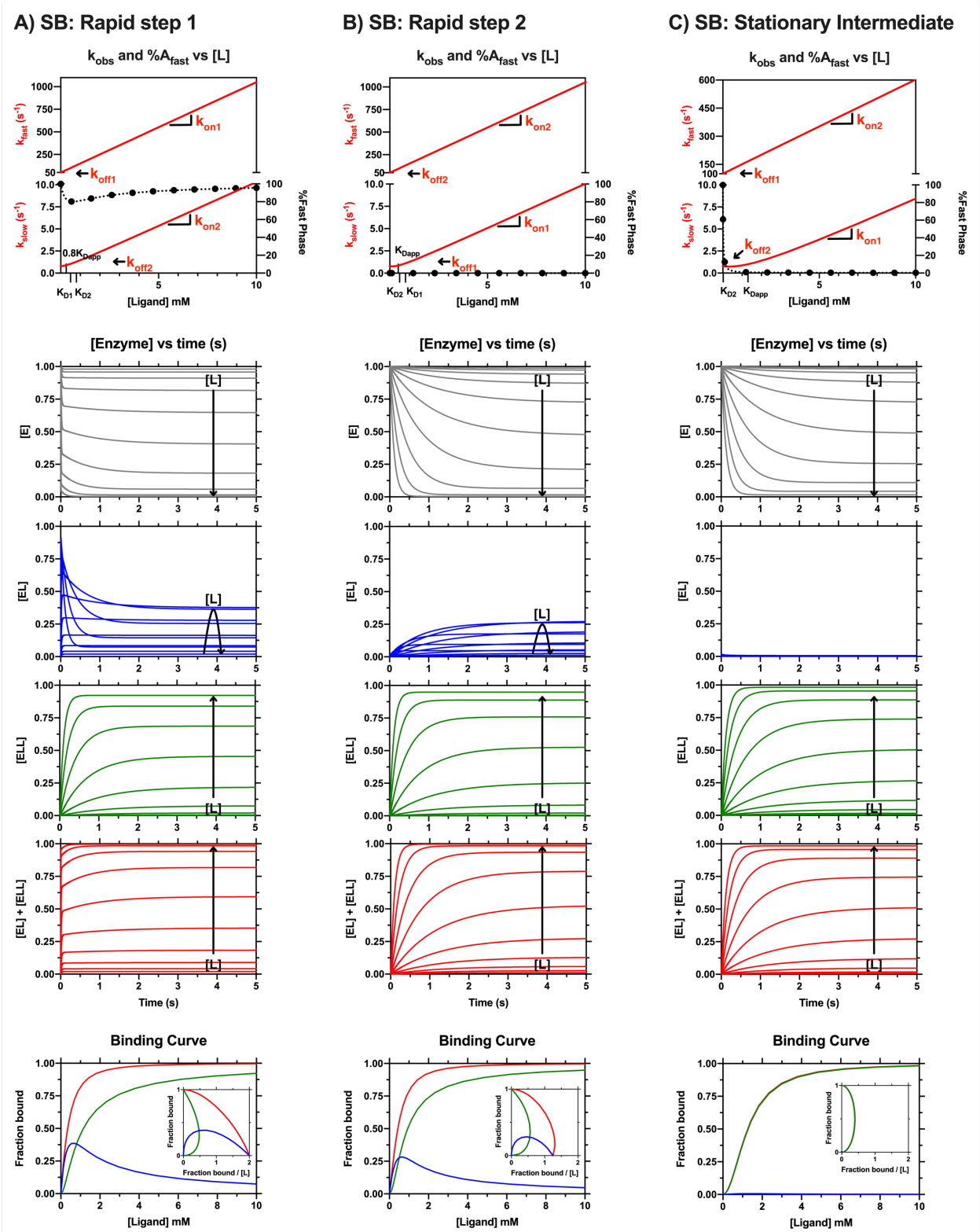


Figure 3.5. Sequential Binding: test cases. Summary of limiting cases for the SB model. The simplified k_{obs} expressions for each case are summarized in Table 3.7. The top row contains plots of the k_{obs} vs [L] (red lines) and the %Fast phase vs [L] (black dots and dotted line). The middle four rows contain plots of the transient concentrations of each enzyme species vs time: [E] = grey

line, [EL] = blue line, [ELL] = green line, [EL] + [ELL] = red line. The bottom row contains the binding curves with inset Eadie-hofstee plots for each bound state: [EL] = blue line, [ELL] = green line, [EL] + [ELL] = red line. A) Rapid binding step 1: k_{on1} = 100 mM^{-1} s^{-1}, k_{off1} = 50 s^{-1}, k_{on2} = 1 mM^{-1} s^{-1}, k_{off2} = 0.8 s^{-1}, $K_{D,app}$ = 0.35 mM, K_{D1} = 0.5 mM, K_{D2} = 0.8 mM. This case was used for the dilution experiment so 0.8*K_{Dapp}, K_{D1}, and K_{D2} are indicated on the x-axis. B) Rapid binding step 2: k_{on1} = 1 mM^{-1} s^{-1}, k_{off1} = 0.8 s^{-1}, k_{on2} = 100 mM^{-1} s^{-1}, k_{off2} = 50 s^{-1}, $K_{D,app}$ = 0.43 mM, K_{D1} = 0.8 mM, K_{D2} = 0.5 mM. The K_{Dapp}, K_{D1}, and K_{D2} are indicated on the x-axis. C) Stationary intermediate: k_{on1} = 1 mM^{-1} s^{-1}, k_{off1} = 100 s^{-1}, k_{on2} = 50 mM^{-1} s^{-1}, k_{off2} = 0.8 s^{-1}, $K_{D,app}$ = 1.3 mM, K_{D1} = 100 mM, K_{D2} = 0.016 mM. The K_{Dapp} and K_{D2} are indicated on the x-axis, however, K_{D1} is outside the axis range and not included.

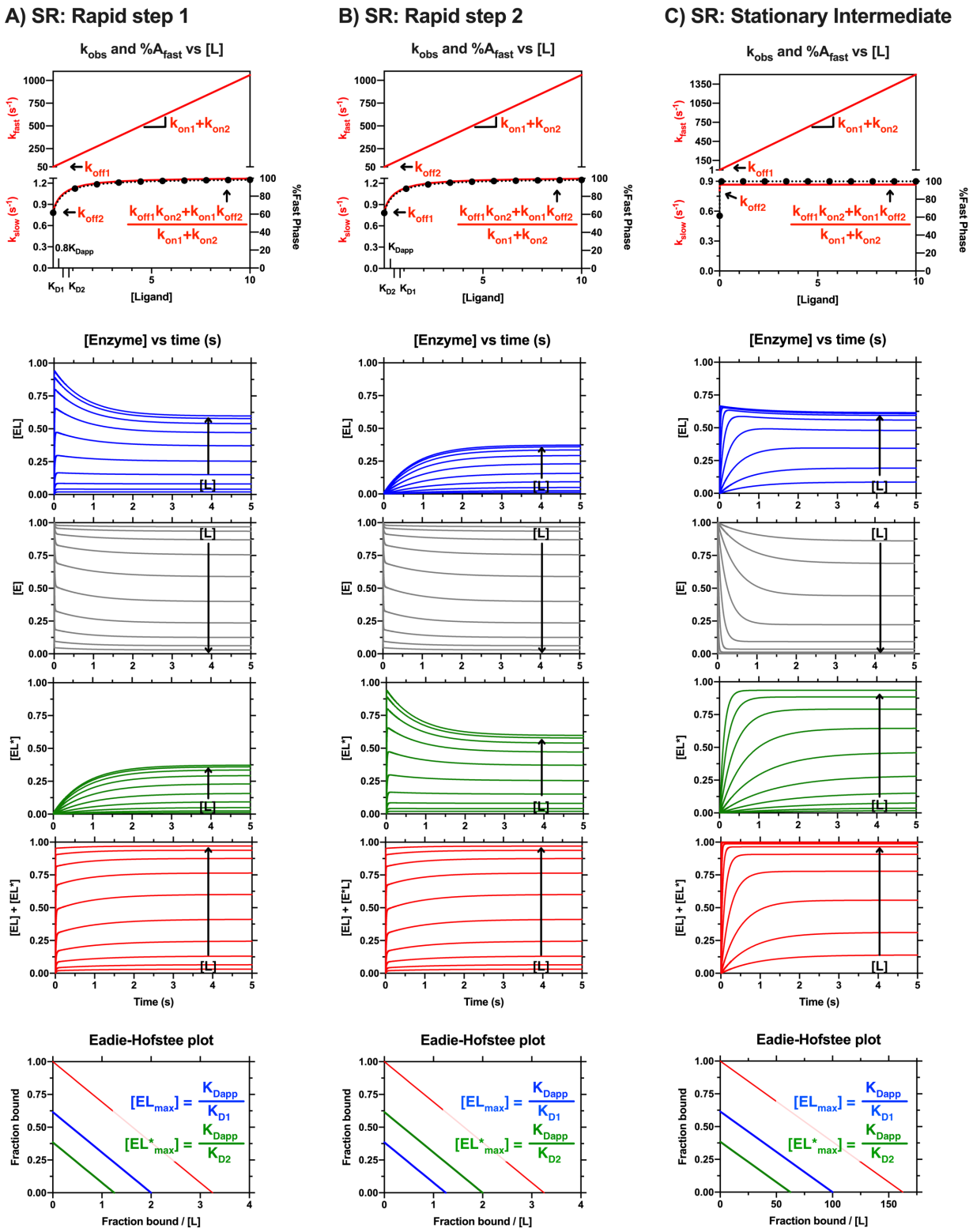
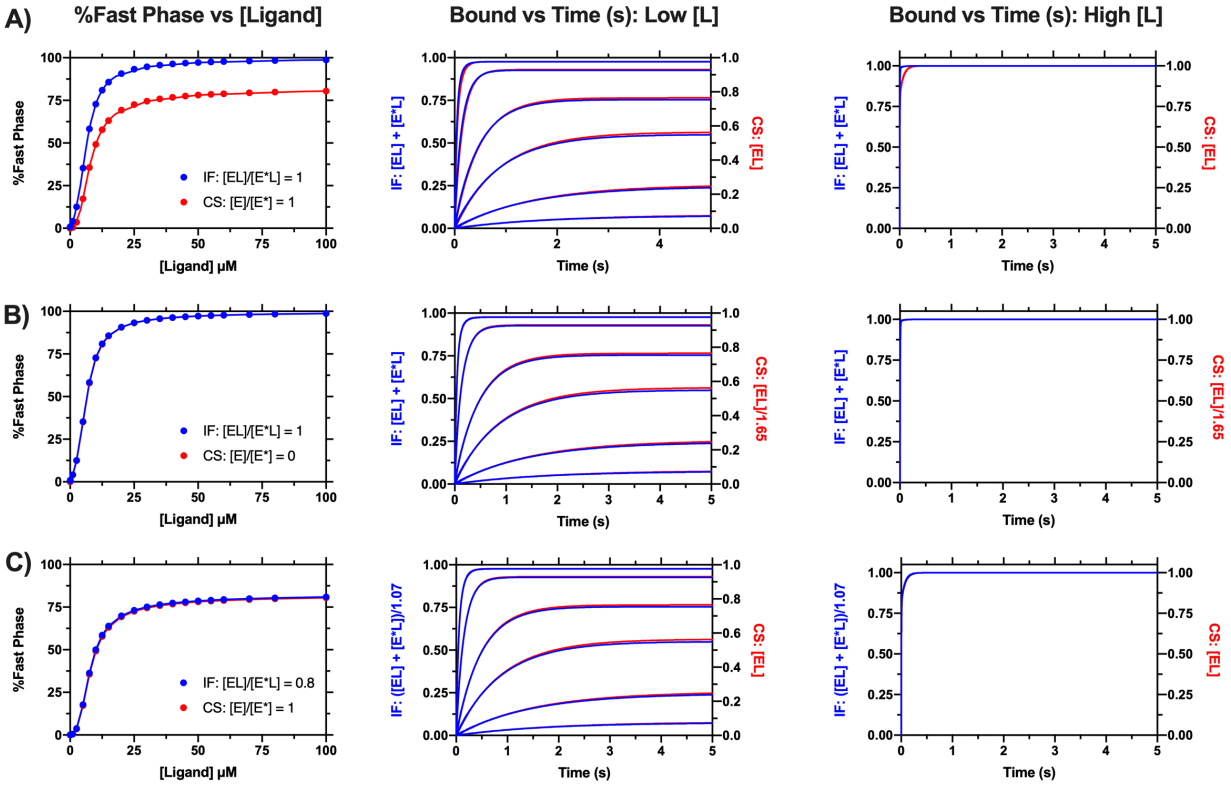


Figure 3.6. Substrate Reorientation: test cases. Summary of limiting cases for the SR model. The simplified k_{obs} expressions for each case are summarized in Table 3.7. The top row contains plots of the k_{obs} vs [L] (red lines) and the %Fast phase vs [L] (black dots and dotted line). The middle four rows contain plots of the transient concentrations of each enzyme species vs time:

[EL] = blue line, [E] = grey line, [EL*] = green line, [EL] + [EL*] = red line. The bottom row contains Eadie-hofstee plots for each bound state: [EL] = blue line, [EL*] = green line, [EL] + [EL*] = red line. A) Rapid binding step 1: k_{off1} = 50 s^{-1}, k_{on1} = 100 mM^{-1} s^{-1}, k_{on2} = 1 mM^{-1} s^{-1}, k_{off2} = 0.8 s^{-1}, $K_{D,app}$ = 0.31 mM, K_{D1} = 0.5 mM, K_{D2} = 0.8 mM. This case was used for the dilution experiment so 0.8*K_{Dapp}, K_{D1}, and K_{D2} are indicated on the x-axis B) Rapid binding step 2: k_{off1} = 0.8 s^{-1}, k_{on1} = 1 mM^{-1} s^{-1}, k_{on2} = 100 mM^{-1} s^{-1}, k_{off2} = 50 s^{-1}, $K_{D,app}$ = 0.31 mM, K_{D1} = 0.8 mM, K_{D2} = 0.5 mM. The K_{Dapp}, K_{D1}, and K_{D2} are indicated on the x-axis. C) Stationary intermediate: k_{off1} = 1 s^{-1}, k_{on1} = 100 mM^{-1} s^{-1}, k_{on2} = 50 mM^{-1} s^{-1}, k_{off2} = 0.8 s^{-1}, $K_{D,app}$ = 0.0062 mM, K_{D1} = 0.01 mM, K_{D2} = 0.016 mM. The K_{Dapp}, K_{D1}, and K_{D2} are not indicated on the x-axis because they are clustered closely and near zero.

IF vs CS: equivalent observed rate case 1



IF vs CS: equivalent observed rate case 2

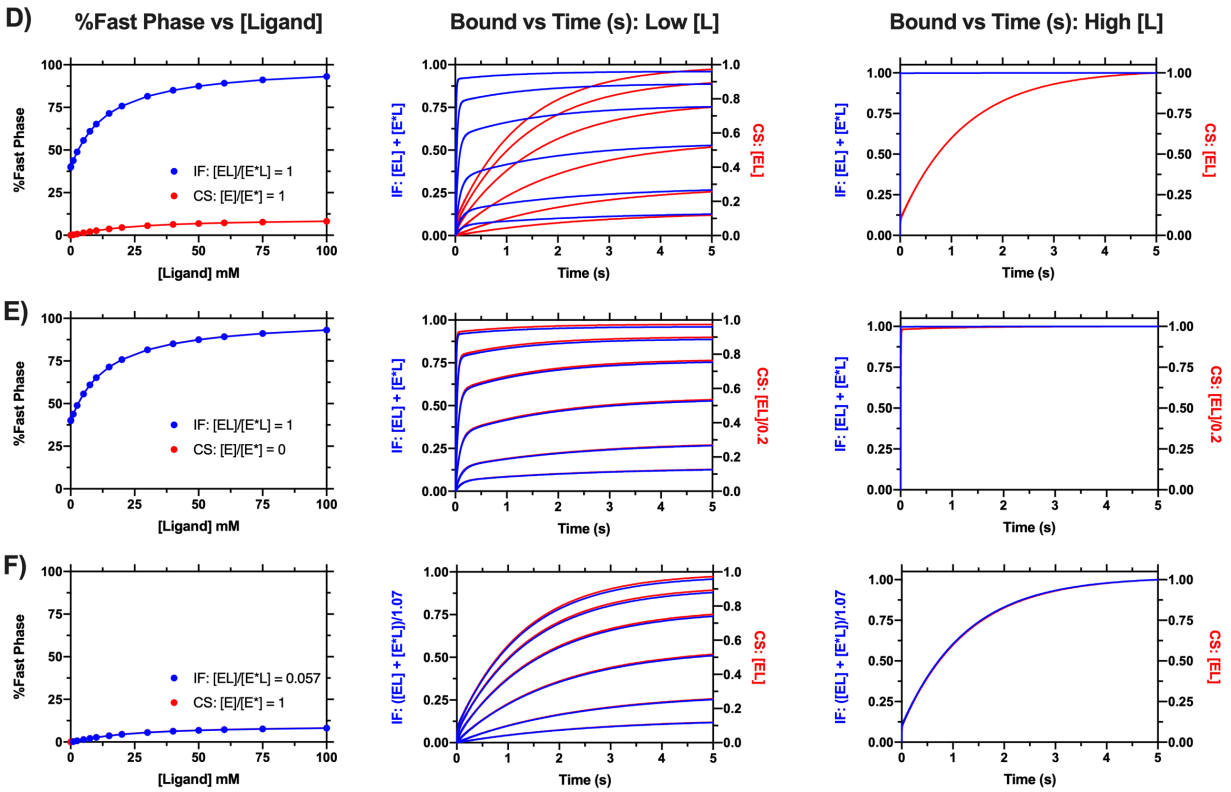


Figure 3.7. 'Equivalent' IF and CS cases. Case 1: IF: k_{on} = 2.6 μM^{-1} s^{-1}, k_{off} = 3.6 s^{-1}, k_f = 12 s^{-1}, k_r = 2.6 s^{-1}, K_D = 1.4 mM, K_D = 1.4 µM, $K_{D,app}$ = 0.25 µM, CS: k_f = 14.6 s^{-1}, k_r = 3.1 s^{-1}, k_{on} = 2.6 μM^{-1} s^{-1}, k_{off} = 0.50 s^{-1}, K_D = 0.19 µM, $K_{D,app}$ = 0.23 µM. **A**) IF model: ε_{EL} = 1, ε_{E^*L} = 1, CS model: ε_E = 1, ε_{E^*} = 1, $[E_0]$= 0.18, $[E^*_0]$ = 0.82, **B**) IF model: same as A, CS model: ε_E = 0, ε_{E^*} = 2, $[E_0]$= 0.18, $[E^*_0]$ = 0.82, **C**) IF model: ε_{EL} = 0.89, ε_{E^*L} = 1.11, CS model: same as A. **Case 2:** IF: k_{on} = 0.54 mM^{-1} s^{-1}, k_{off} = 7.54 s^{-1}, k_f = 0.44 s^{-1}, k_r = 0.36 s^{-1}, K_D = 14.0 mM, $K_{D,app}$ = 6.28 mM, CS: k_f = 0.8 s^{-1}, k_r = 7.2 s^{-1}, k_{on} = 0.54 mM^{-1} s^{-1}, k_{off} = 0.34 s^{-1}, K_D = 0.63 mM, $K_{D,app}$ = 6.30 mM. **D**) IF model: ε_{EL} = 1, ε_{E^*L} = 1, CS model: ε_E = 1, ε_{E^*} = 1, $[E_0]$= 0.9, $[E^*_0]$ = 0.1, **E**) IF model: same as D, CS model: ε_E = 0, ε_{E^*} = 2, $[E_0]$= 0.9, $[E^*_0]$ = 0.1, **F**) IF model: ε_{EL} = 0.11, ε_{E^*L} = 1.89, CS model: same as D.

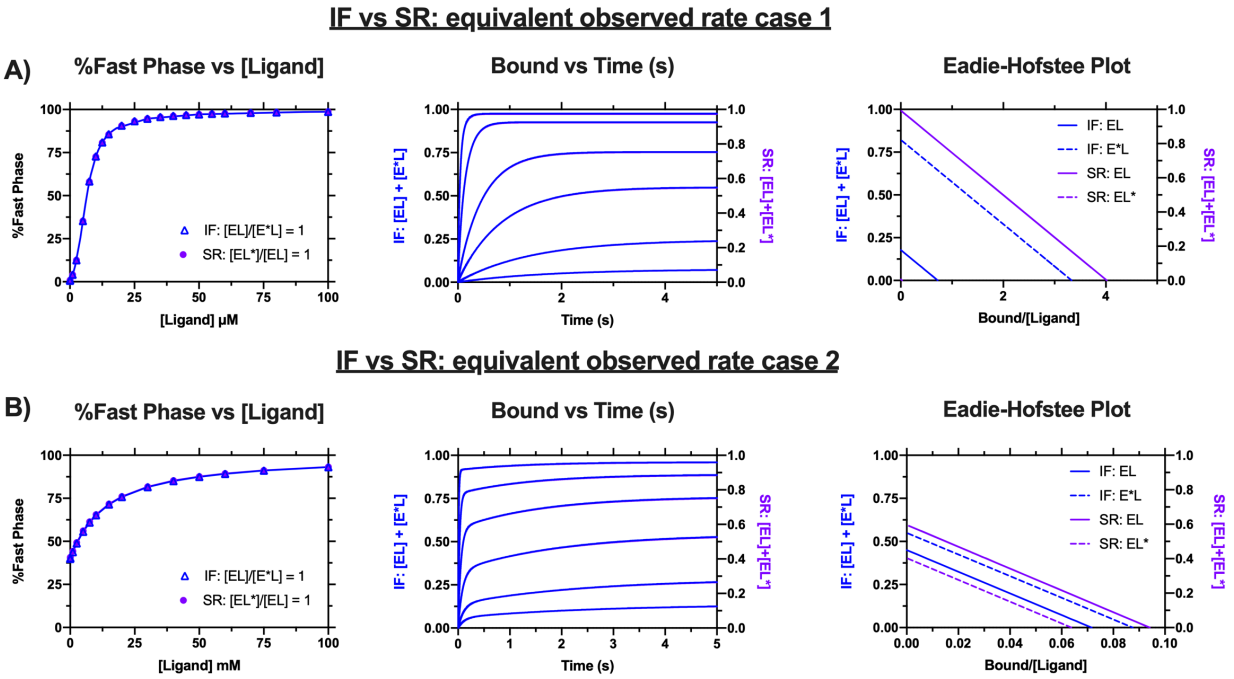


Figure 3.8. 'Equivalent' IF and SR cases. A) Case 1: IF: k_{on} = 2.6 µM^{-1} s^{-1}, k_{off} = 3.6 s^{-1}, k_f = 12 s^{-1}, k_r = 2.6 s^{-1}, K_D = 1.4 mM, K_D = 1.4 µM, $K_{D,app}$ = 0.25 µM, SR: k_{off1} = 0.53 s^{-1}, k_{on1} = 2.13 µM^{-1} s^{-1}, k_{on2} = 0.47 µM^{-1} s^{-1}, k_{off2} = 17.7 s^{-1}, K_{D1} = 0.25 µM, K_{D2} = 38 µM, $K_{D,app}$ = 0.25 µM. **A**) IF model: ε_{EL} = 1, ε_{E*L} = 1, SR model: ε_{EL*} = 1, ε_{EL} = 1, [E_0]= 1, **B) Case 2**: IF: k_{on} = 0.54 mM^{-1} s^{-1}, k_{off} = 7.54 s^{-1}, k_f = 0.44 s^{-1}, k_r = 0.36 s^{-1}, K_D = 14.0 mM, $K_{D,app}$ = 6.28 mM, SR: k_{off1} = 0.34 s^{-1}, k_{on1} = 0.032 µM^{-1} s^{-1}, k_{on2} = 0.51 µM^{-1} s^{-1}, k_{off2} = 8.0 s^{-1}, K_{D1} = 10.6 µM, K_{D2} = 15.7 µM, $K_{D,app}$ = 6.33 µM.

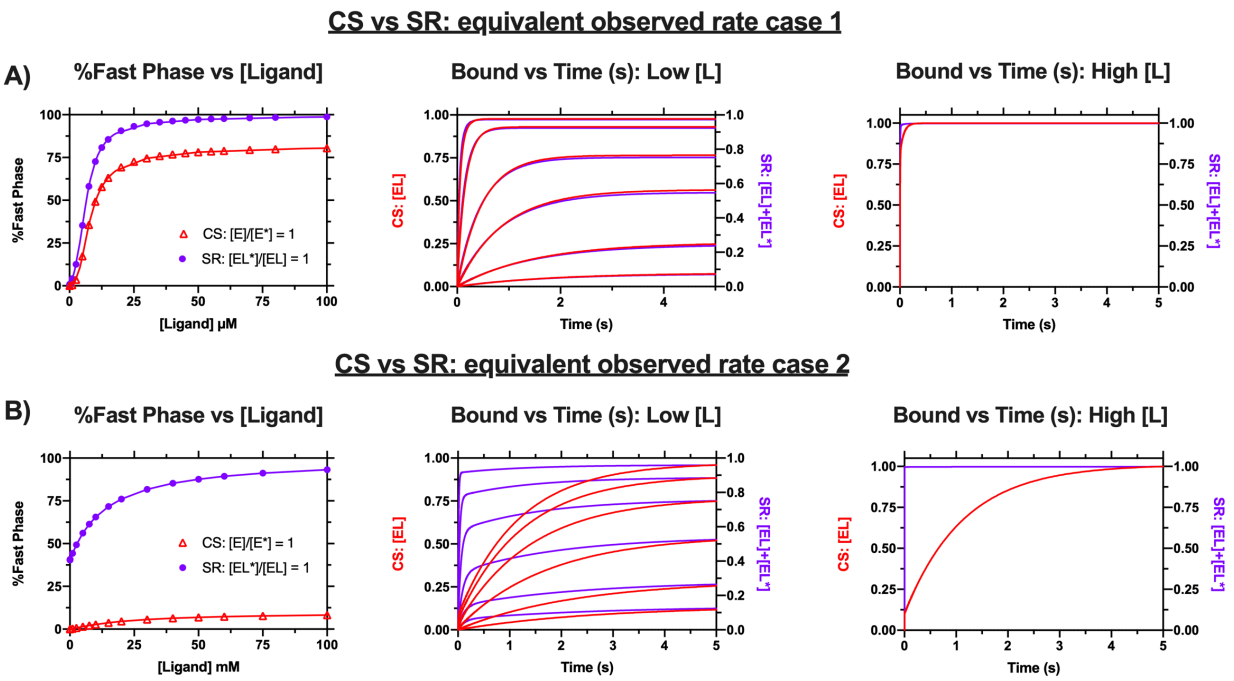


Figure 3.9. 'Equivalent' CS and SR cases. A) Case 1: CS: k_f = 14.6 s^{-1}, k_r = 3.1 s^{-1}, k_{on} = 2.6 μM^{-1} s^{-1}, k_{off} = 0.50 s^{-1}, K_D = 0.19 µM, $K_{D,app}$ = 0.23 µM, SR: k_{off1} = 0.53 s^{-1}, k_{on1} = 2.13 μM^{-1} s^{-1}, k_{on2} = 0.47 μM^{-1} s^{-1}, k_{off2} = 17.7 s^{-1}, K_{D1} = 0.25 µM, K_{D2} = 38 µM, $K_{D,app}$ = 0.25 µM. **A**) IF model: ε_{EL} = 1, ε_{E*L} = 1, SR model: ε_{EL*} = 1, ε_{EL} = 1, $[E_0]$= 1, **B) Case 2**: CS: k_f = 0.8 s^{-1}, k_r = 7.2 s^{-1}, k_{on} = 0.54 mM^{-1} s^{-1}, k_{off} = 0.34 s^{-1}, K_D = 0.63 mM, $K_{D,app}$ = 6.30 mM, SR: k_{off1} = 0.34 s^{-1}, k_{on1} = 0.032 μM^{-1} s^{-1}, k_{on2} = 0.51 μM^{-1} s^{-1}, k_{off2} = 8.0 s^{-1}, K_{D1} = 10.6 µM, K_{D2} = 15.7 µM, $K_{D,app}$ = 6.33 µM.

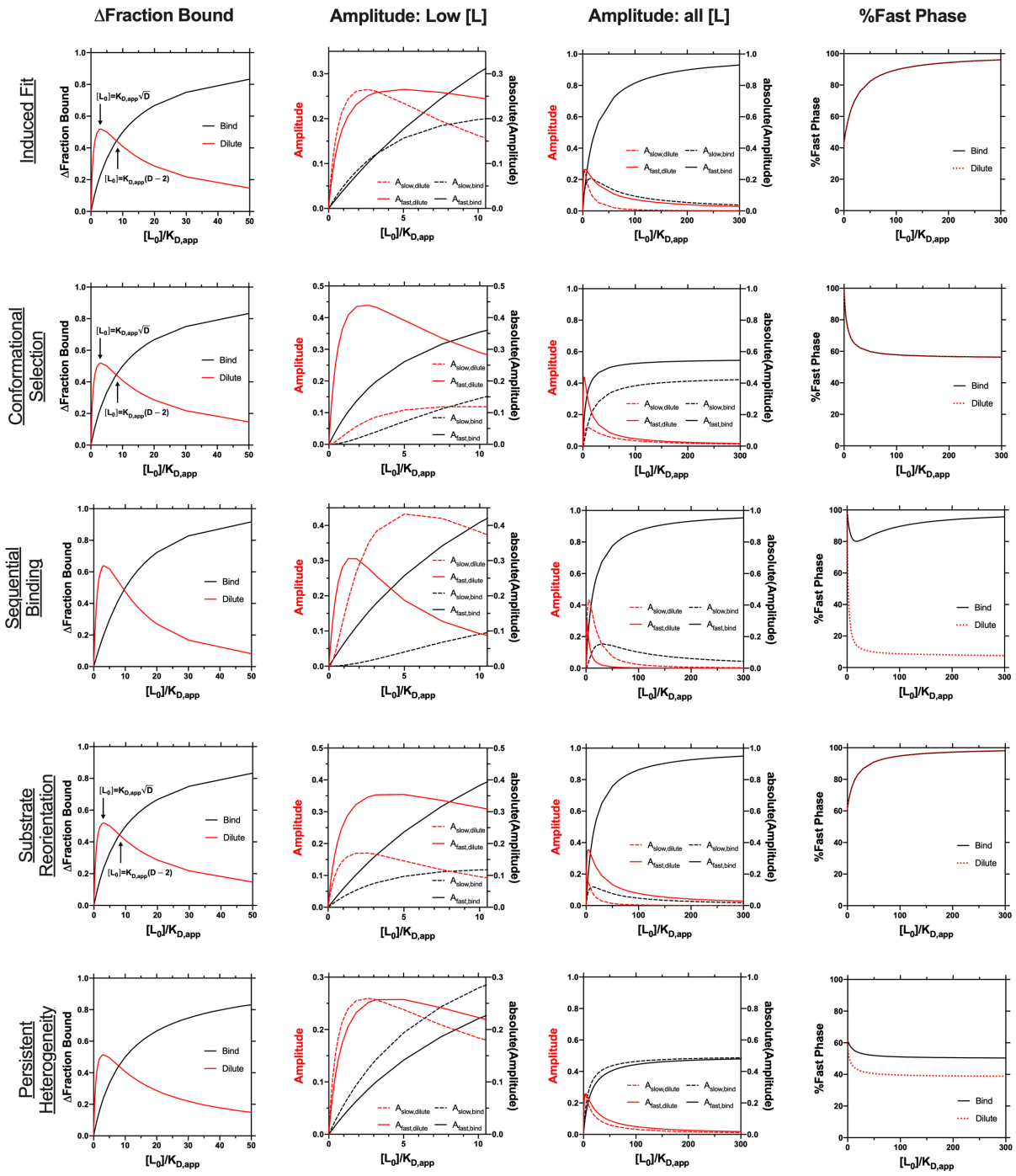


Figure 3.10. Comparison of binding and dilution experimental results. The dilution factor for all dilution simulations is equal to 10. Extinction coefficients for the bound species are equal to 1. **IF**: (Rapid binding example from Figure 3.3A) k_{on} = 100 mM^{-1} s^{-1}, k_{off} = 50 s^{-1}, k_f = 1 s^{-1}, k_r = 0.8 s^{-1}, $K_{D,app}$ = 0.22 mM, K_D = 0.5 mM, **CS**: (Rapid binding example from Figure 3.4B) k_f = 1 s^{-1}, k_r = 0.8 s^{-1}, k_{on} = 100 mM^{-1} s^{-1}, k_{off} = 50 s^{-1}, $K_{D,app}$ = 0.9 mM, K_D = 0.5 mM, **SB**: (Rapid step 1 example from Figure 3.5A) k_{on1} = 100 mM^{-1} s^{-1}, k_{off1} = 50 s^{-1}, k_{on2} = 1 mM^{-1} s^{-1}, k_{off2} = 0.8 s^{-1}, $K_{D,app}$ = 0.35 mM, K_{D1} = 0.5 mM, K_{D2} = 0.8 mM, **SR**: (Rapid binding example from Figure 3.6A)

k_{off1} = 50 s^{-1}, k_{on1} = 100 mM^{-1} s^{-1}, k_{on2} = 1 mM^{-1} s^{-1}, k_{off2} = 0.8 s^{-1}, $K_{D,app}$ = 0.31 mM, K_{D1} = 0.5 mM, K_{D2} = 0.8 mM, **PH**: k_{on1} = 100 mM^{-1} s^{-1}, k_{off1} = 50 s^{-1}, k_{on2} = 1 mM^{-1} s^{-1}, k_{off2} = 0.8 s^{-1}, $K_{D,app}$ = 0.63 mM, K_{D1} = 0.5 mM, K_{D2} = 0.8 mM, $[E_{total}] = [E^*_{total}] = E + EL = E^* + E^*L$.

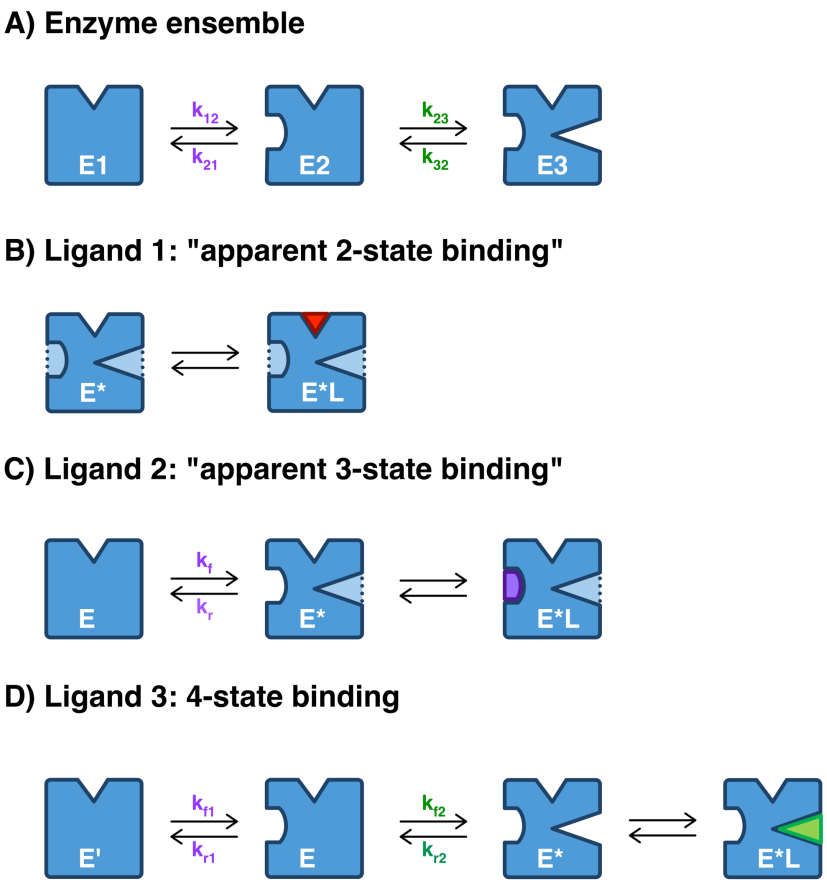
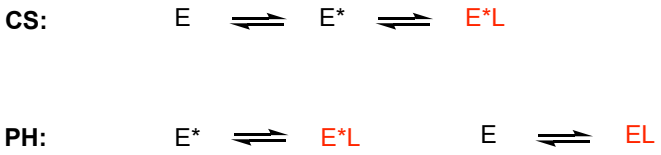


Figure 3.11. Different ligand selection criteria may yield different binding mechanisms for the same enzyme. Three different binding scenarios are presented with three different ligands (red ligand, purple ligand, and green ligand) that select different subsets of enzyme conformers (E1, E2, and E3). **A**) The ensemble of three unbound enzyme states are shown with isomerization rates describing their rates of interconversion in purple in green text. The ligand binding mechanisms are shown below for the red, purple, and green ligand, which possess different 'selection' criteria for the binding competent enzyme conformation. **B**) The ensemble is 'kinetically silent' when all enzyme conformers are 'binding competent.' Here, the ligand (red triangle) binds to all enzyme conformers in Figure 3.11A, which yields an apparently 2-state binding mechanism. The isomerization processes on the left and right face of the enzyme become 'kinetically silent' when they occur without altering the binding kinetics for the 'binding competent' conformation or binding site (top face of the enzyme). **C**) Ligand (purple curved rectangle) selects only E2 and E3 as the 'binding competent' conformers. In addition, the isomerization between E2 and E3 has no effect on ligand binding and becomes 'kinetically silent.' This results in an apparently 3-state CS binding mechanism which yields the k_{12} and k_{21} isomerization rates. **D**) Ligand (green triangle) selects only E3 as the 'binding competent' conformer. All isomerization processes become 'kinetically visible.' This results in a 4-state CS binding mechanism which yields the k_{12}, k_{21}, k_{23}, and k_{32} isomerization rates.

Binding promiscuity



Catalytic promiscuity

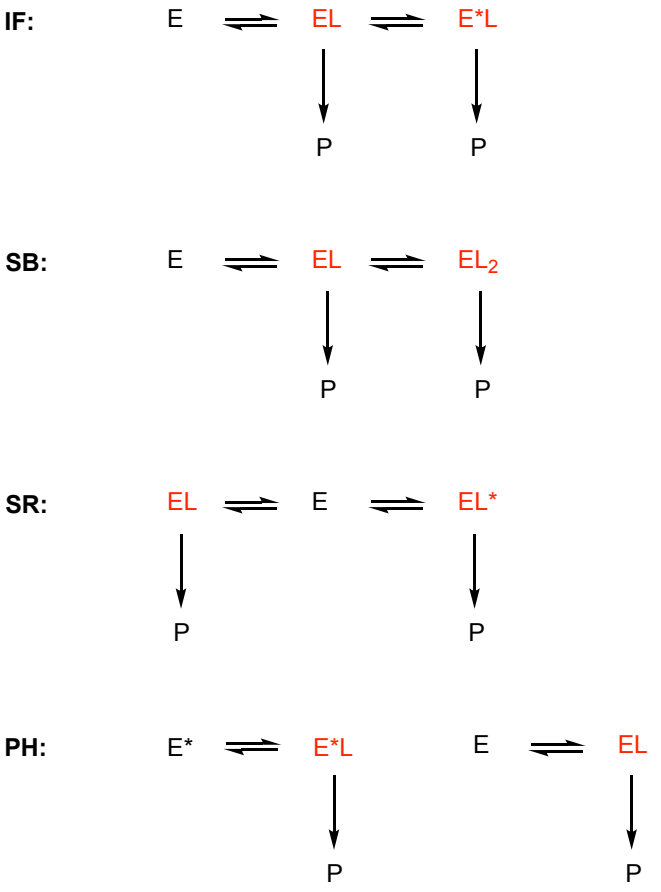


Figure 3.12. Potential functional significance of the different binding mechanisms. The different binding mechanisms are shown in the context of their potential role in binding and catalytic promiscuity. The CS and PH binding mechanisms potentially contribute to binding promiscuity when the different enzyme conformers have different binding preferences, particularly for different ligand types (e.g. polarity, size, planarity, etc.).The IF, SB, SR and PH binding mechanisms potentially contribute to catalytic promiscuity when the different bound states (highlighted in red text) have different catalytic properties (e.g. regiospecificity, catalytic efficiency, etc.).

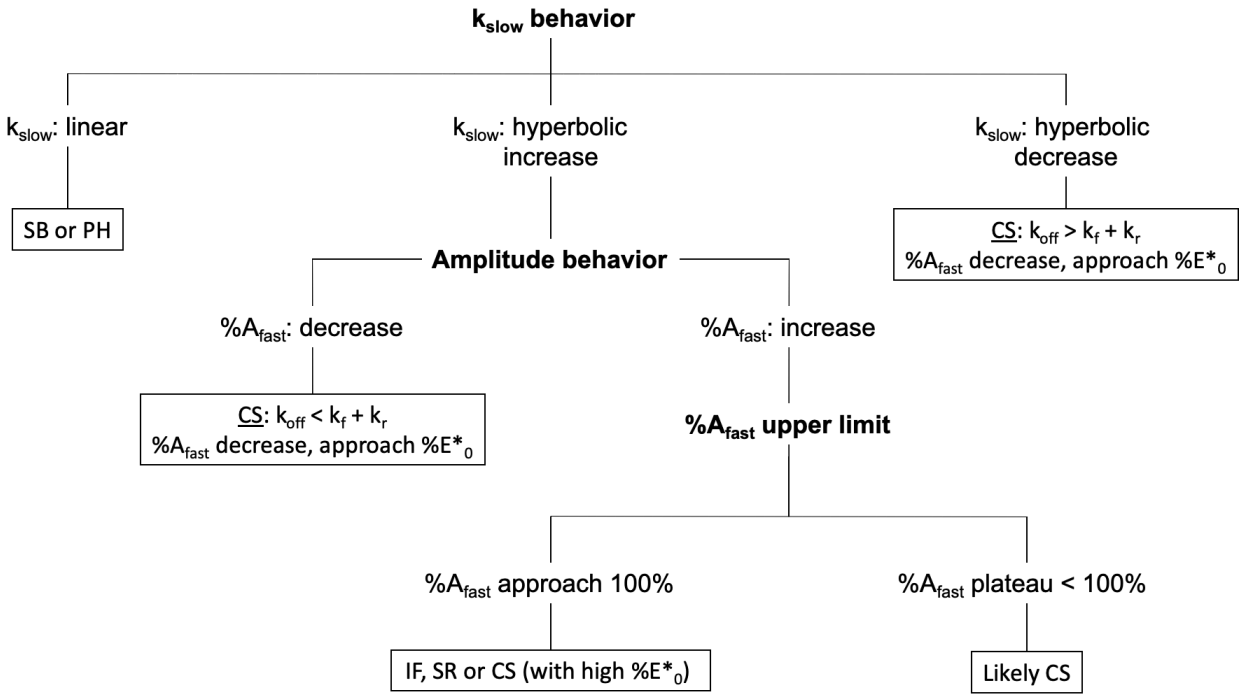


Figure 3.13. Kinetic decision tree.

Chapter 4

Kinetic Resolution of Heterogenous Ligation States of Small Molecules Binding to CYP3A4

4.1 Introduction

Cytochrome P450s (CYPs) are a pharmacologically important family of heme-containing mono-oxygenases that metabolize a range of endogenous and exogenous compounds. These CYPs metabolize, or are inhibited by, the majority of marketed drugs and this results in complex, and sometimes unpredictable, effects when multiple drugs are administered simultaneously. Among the hepatic CYPs, the CYP3A4 isoform is a major determinant of drug interactions. In fact, inhibition of CYP3A4 is an accepted strategy for modulating the clearance of other drugs known to be cleared by it [140-142]. A common strategy for the design of CYP3A4 inhibitors, or the rationale for using existing drugs as inhibitors, is based on the expectation that 'Type I' compounds will behave as substrates while 'Type II' and 'Reverse Type I' compounds will behave as inhibitors.

4.1.1 Heterogeneity in the ligation state of 'Type II' compounds

Recent observations have suggested that this design strategy for inhibition is oversimplified, as some imidazoles or triazoles generate a mixture of heme ligation states including multiple nitrogen-coordinated states, water-bridged states that retain the heme axial water hydrogen-bonded to the nitrogenous fragment, and heme with apparent 'bulk water' bound. These low-spin (LS) states are distinguishable by EPR parameters but are difficult to resolve by UV-visible spectroscopy [143]. Specifically, low temperature Hyperfine Sublevel correlations

Spectroscopy (HYSCORE) clearly demonstrates the ability of various nitrogen-containing compounds to hydrogen bond to the heme axial water. Mixtures of these water bridged species with nitrogen-coordinated species from a single compound have been deconvoluted by CW EPR and are distinguishable from water ligated 'resting' CYP [23]. Crystal structures further demonstrate the ability of nitrogenous compounds to hydrogen bond to the axial water rather than coordinate to the heme iron in the crystalline state [144]. However, the existence of such mixtures of ligation state in solution at near physiologic temperatures is not well documented.

Paradoxically, several nitrogenous compounds that yield these mixtures of heme states are efficiently metabolized by CYPs, in contrast to the simple canonical model that predicts inhibition for compounds that coordinate to the heme [17-20,145]. In contrast to P450cam, the nitrogen-coordinated species may be more efficiently reduced than the resting state for some CYPs [21]. Moreover, it is well established that CYP3A4 is able to form multiple products from the same compound, which is a property that implies multiple binding orientations within the active site [146]. As a result of these recent and long understood observations, it has become clear that effective design of CYP inhibitors requires a more complete characterization of the functional properties of the various LS complexes that are formed with nitrogenous ligands. In principle, the kinetics of exchange between different heme ligation states for a single drug could contribute to the overall catalytic efficiency or inhibitory properties.

4.1.2 Experimental approach and ligand selection

Here we employ kinetic methods to characterize the heterogeneous low spin complexes formed with model drug fragments to CYP3A4. In order to study the effects of local electronic structure on the kinetics of ligand binding, we aimed to minimize the effects of ligand interacting

with protein residues or backbone atoms. One simple approach is to use sterically unhindered chemically unsubstituted fragments that ligate to the heme. Imidazole (IMZ) and 1,2,3-triazole (123-TRZ) provide two such fragments with clearly established differences in the thermodynamics of their interaction with the heme. Specifically, it is well established that IMZ binds tightly whereas 123-TRZ has a much lower affinity (μM vs. mM) [143]. Because these fragments are sufficiently small to not interact with the active site protein residues when complexed to the heme iron, the spectrally measured kinetics are expected to be controlled primarily by the local electronic factors. We also include cyanide anion (CN^-), as a surrogate for small diatomic molecules known to bind the reduced heme-iron, such as oxygen or carbon monoxide. CN^- forms a unique 'Type II' binding spectra because it coordinates to the heme-iron via an electron rich carbon. Notably, binding kinetics for CN^- have been studied previously, but multiphasic kinetics were not analyzed in detail [71].

4.1.3 Improvements in kinetic analyses: a push for a global approach

There are many challenges to the determining of kinetic binding mechanisms, which are rooted in the methods of the analyses. One of these, involves the delicate balance between overparameterization and goodness of fit. Traditional analytical approaches rely on interpretation of the ligand-dependent behavior of the observed rates from multi-exponential fits of the data. The generic exponential equations have no physical linkage to a given kinetic model which may lead to erroneous assignment of a binding mechanism. A better approach is to fit the data to the characteristic eigenvalue expressions that describe the observed rates for a particular model. A second challenge arises from this approach, which involves the possibility of multiple minima. The eigenvalue expressions contain several terms and the final fits may be sensitive to the initial

parameter guesses (i.e. prone to local minima), particularly when data from a single set of experimental conditions (i.e. single ligand concentration) are locally fit. A global approach wherein multiple experimental conditions are simultaneously fit can help improve the fits. However, recent kinetic analyses reveal that different binding models can yield nearly equivalent observed rates for a range of experimental conditions, despite having different intrinsic rate constants for each model[131]. The binding kinetics are information-rich and the behavior of the recovered parameters from the multi-exponential fits of the data (the observed rates, k_{obs}, and the amplitudes) are unique to a given mechanism and set of initial conditions. Importantly, the models are more easily resolved when one considers the behavior of the amplitudes, which are often ignored in kinetic analyses. A final challenge emerges: the expressions that describe the amplitude behavior in tandem with the observed rate expressions may become prohibitively large for equation input to most analytical platforms.

Herein, a global approach is employed to analyze kinetic data wherein data are fit to full solutions for each tested model using MATLAB. The concept of global kinetic analysis is not new, however, the methodology in this work provides a framework for global kinetic analyses with a comprehensive list of 8 models to interrogate for cases when single-ligand binding is suspected. The approach can be easily adapted to include models with multiple binding sites, and I include the differential equations and the initial protein concentration equation expressions to do so (see Appendix). A potential reason that the amplitudes have been ignored in kinetic analyses is the possibility that different species in the binding mechanism differently contribute to the signal which could change the interpretation of the data. Herein, data are fit to the solutions of the system of differential equations that describe the tested kinetic mechanisms extinction coefficient parameters are included for each species in the model to account for the possibility that different

binding modes can differently contribute to the signal. To address the possibility of multiple minima in the optimization, a global approach is employed which simultaneously fits data from multiple experimental conditions (i.e. multiple ligand concentrations and different experimental approaches: binding vs dissociation). Additionally, in an effort to remove user input bias, a multi-start approach wherein multiple randomized initial guesses were used to search for a global minimum in the fits.

4.2 Experimental Procedures

4.2.1 Materials

1H-1,2,3-triazole, imidazole, and potassium cyanide were purchased from Sigma Chemical Co. (St. Louis, MO, USA) and used without further purification. All water used was Milli-Q quality from a Barnstead nanopure UV dispenser.

4.2.2 Protein expression and purification

Recombinant CYP3A4 was expressed and purified from *Escherichia coli* cells as described in Chapter 2.2.1. Purified protein was stored at -80 ºC in Buffer A (100 mM KPi pH 7.4, 20% glycerol, 0.5 mM TCEP, 1 mM EDTA) for further use.

4.2.3 Equilibrium binding titrations

Equilibrium binding titrations were conducted at 10.6 ºC with an Olis Modernized Aminco DW-2 (Olis, Inc., Bogart, GA) dual beam spectrophotometer equipped with a Julabo F30-C compact refrigerated circulator (Julabo USA, Inc., Allentown, PA) as described in [143]. Ligand

(IMZ, 123-TRZ or CN^-) and CYP3A4 stocks were prepared in Buffer A. Ligand was titrated into CYP3A4 (0.9 μM). Difference spectra were calculated as the difference in ligand bound CYP3A4 spectra relative to the ligand free CYP3A4 spectra. The difference in absorbance between λ_{max} and λ_{min}, ΔAbsorbance, was plotted as a function of ligand concentration, [L], and fit to the hyperbolic binding equation (equation 4.1) using IGOR pro 6.35A (Wavemetrics, Lake Oswego, OR):

$$\Delta\text{Absorbance} = \frac{A_{max}\,[L]}{K_D + [L]} \tag{4.1}$$

where A_{max} is the maximum change in absorbance and K_D is the ligand dissociation constant.

4.2.4 Corrections to CN^- concentration

Cyanide concentration in all figures and data analysis refer to the concentration of the cyanide anion species, [CN^-], at pH 7.4. The experiments are still considered to be under pseudo-first order conditions as there is a large pool of KCN in solution ($pK_{a(KCN)}$ = 11.0) as was done previously [147]. The final ligand cyanide anion concentration was calculated as follows:

$$[CN^-] = \left(\frac{K_{a(KCN)})}{[H]^+ + K_{a(KCN)}}\right)[KCN] \tag{4.2}$$

4.2.5 Stopped-flow kinetic: ligand binding (mixing experiments)

Ligand binding kinetic measurements were performed under pseudo-first order conditions ([Ligand] >> [CYP3A4]) at 10.6 ºC with an Applied Photophysics SX-20 stopped-flow instrument (Applied Photophysics, Leatherhead, U.K.) equipped with a photodiode array detector and

compact refrigerated circulator. Ligand free CYP3A4 (1.18 μM final concentration) was mixed with variable amounts of ligand in a 1:1 ratio in Buffer A. Difference spectra were calculated similarly to equilibrium titrations. The difference in absorbance between λ_{max} and λ_{min}, Δ Absorbance, was plotted as a function of time, t, for each ligand concentration.

4.2.6 Stopped-flow kinetic: ligand dissociation (dissociation experiments)

In general, when multiexponential kinetics are possible, a wide range of ligand concentrations is required to adequately characterize the concentration dependent behavior of the observed rate of change, k_{obs}, linked to each phase (typically assessed by k_{obs} vs [ligand] plots). In many cases, the observed rate of change for the fast phase, $k_{obs,fast}$, often has a linear relationship with respect to ligand concentration; the slope of which is often equal to the on-rate of ligand binding. Additional phases including the slowest rates of change, $k_{obs,slow}$, often have a hyperbolic dependence with respect to ligand concentration; the limits are often related to the rearrangements rates that govern the interconversion between enzyme states. Importantly, observation of the direction of change for $k_{obs,slow}$ with respect to ligand concentration may be indicative of the binding mechanism. In practice, the low ligand concentrations required to observe such curvature in the k_{obs} vs [ligand] plot are often impractical because the signal intensity diminishes as ligand concentration approaches zero. To circumvent such a problem, I performed a dilution experiment which 'amplifies' the signal intensity by measuring the dissociation of ligand from the bound species.

4.2.7 Global fitting and statistical analysis of stopped-flow kinetic data

Data were globally fit to eight different ligand binding models (Figure 4.1) in MATLAB with the goal of estimating the following parameters for each model: kinetic rate constants and relative extinction coefficients for each bound species. Total protein concentration was defined as the max absorbance change at saturating ligand concentrations (A_{max} in equation 1); this is valid for assumed linearity in the relationship between absorbance signal and concentration in accord with Beer's Law and when the kinetics are unaltered by ligand depletion in accord with the pseudo-first order experimental conditions. For models with a pre-existing ensemble of unbound enzyme conformations (models with a conformational selection component), the initial concentrations of each ligand-free species were calculated using the relevant rate constants involved in the interconversion of each species. For models with an ensemble of bound enzyme conformations (models with an induced fit component), parameters for the relative extinction coefficients of individual bound species are included to account for the possibility that different binding modes can differently contribute to the spectroscopic signal or even, in more extreme cases, be spectroscopically silent. The multi-start algorithm in the Global Optimization Toolbox was used to randomize initial guesses for the rate constants and extinction coefficients with lower bounds of 0 and upper bounds of 100 s^{-1} and 10^9 M^{-1} s^{-1} for the rate constants and bounds between 0 and 20 for the relative extinction coefficients of bound species [126]. The fits were constrained such that binding affinity was within 1-fold K_D obtained from the equilibrium titrations. The differential equations for each model (see appendix) were solved using the ode15s solver. The process was repeated for each initial guess to minimize the sum of residuals squared, RSS, using the *fmincon* optimizer.

The model that best described the data was selected using standard model selection criteria such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) under a

likelihood framework which ap[148,149]ply penalties to the increasing number of parameters in larger models . The log likelihood, LL, was calculated for each model fit using equation 4.3 where n is the number of data points, s is the standard deviation of the residuals, and RSS is the sum of residuals squared.

$$\mathrm{LL} = -\frac{\mathrm{n}}{2}\log(2\pi) - \mathrm{n} * \log(\mathrm{s}) - \frac{\mathrm{RSS}}{2\mathrm{s}^2} \tag{4.3}$$

The criteria were calculated using equations 4.4 and 4.5 below:

$$\mathrm{AIC} = -2 * \mathrm{LL} + 2 * \mathrm{p} \tag{4.4}$$

$$\mathrm{BIC} = -2 * \mathrm{LL} + \ln(\mathrm{n}) * \mathrm{p} \tag{4.5}$$

where LL is the log likelihood, n is the number of data points, and p is the number of parameters in the model. Due to the large sample size (thousands of data points), AIC and BIC were similar, so we only show AIC to conserve space. The lowest AIC or BIC value indicates the model with the best fit for a given data set.

4.2.8 Singular value decomposition of kinetic data

Singular value decomposition (SVD) is a useful matrix factorization technique that can be applied to spectral data to help identify the number of spectrally distinct species in a given dataset [150-153]. The absolute binding spectra from the kinetic experiments in this study were analyzed by SVD in MATLAB. The data matrix, A, was decomposed into the product of three matrices, U, S, and V^T:

$$\mathrm{A} = \mathrm{USV}^T \tag{4.6}$$

where A is an m x n data matrix (m = the number of wavelengths and n = the number of time points for a single stopped flow injection), U is an m x n orthogonal matrix in the wavelength domain containing spectral components, S is an n x n diagonal matrix containing the singular values and V^T is an n x n orthogonal matrix in the time domain containing the spectral amplitude vectors or concentration profiles.

4.3 Results

4.3.1 Equilibrium titrations

UV-visible difference spectra were performed in order to determine the equilibrium binding affinity of each ligand. Figure 4.2 shows the calculated difference spectra obtained from UV-Vis absorbance titrations and the resultant binding isotherms with CYP3A4 at 10.6 ºC in 100 mM KPi (pH 7.4, 20% glycerol, 0.5 mM TCEP and 1mM EDTA). Binding of all three ligands induced type II difference spectra which is consistent with previous reports [71,143]. The spectral K_D values for 123-TRZ, IMZ and CN^- binding to CYP3A4 are 13.4 mM, 201 μM, and 566 nM, respectively. CN^- binds to CYP3A4 with the tightest binding affinity. Between the azoles, IMZ binds more tightly than 123-TRZ.

4.3.2 Stopped-flow kinetic measurements

To understand the mechanism of ligand binding, kinetic data from mixing and dilution experiments spanning a range of ligand concentrations were conducted. The 1,2,3-TRZ binding kinetics were too fast to assign a model (data not shown). The results from the kinetic experiments from IMZ and CN^- shown in Figures 4.3 – 4.4. As is commonly observed for ligand binding

kinetics with CYP3A4 [72], the binding kinetics for all three ligands were multiexponential, thus necessitating interrogation of multistate binding mechanisms. The data were fit to a total of eight different ligand binding models (Figure 4.1) and analyses of those fits were used to identify the model that best described the data. It is important to acknowledge there are additional 3- or 4-state models and justify their omission from the pool of tested models below: On the basis of the linear Eadie-Hofstee plots for each ligand, which suggest a single heme binding site, all iterations of models containing multiple ligand binding sites were excluded from this study (example: $E + L \rightleftarrows EL \rightleftarrows EL_2$). Additionally, due to the small size of each ligand, it is conceivable that any small reorientations in binding mode occur within the active site. Thus, models containing heterogenous bound states that require active site egress and external reorientation prior to rebinding to achieve a different binding mode were excluded ($EL^* \rightleftarrows E + L \rightleftarrows EL$; prominent example: itraconazole [69]). Finally, triple exponential fits were sufficient to describe the kinetic data which strongly suggest a binding model containing at least three equilibration steps. Thus, in an effort to avoid overparameterization, models containing two or three equilibration steps were examined while models containing more than four equilibria were excluded, with the exception of the closed mixed conformational selection/induced fit model, as it has been extensively discussed in the literature.

4.3.3 123-TRZ binding mechanism

Binding and dilution kinetics were collected for 123-TRZ binding to CYP3A4. The binding kinetics yielded what appeared to be multi-exponential kinetics, however, the presence of a fast phase beyond the instrument limits of detection. Notably, a slow-phase was observed at concentrations up to 300 mM 123-TRZ, which potentially suggests a CS component (persistent slow phase with constant amplitude spanning large concentration range) or different extinction

coefficients for the bound species. The dilution experiment spanning ligand concentrations that were undetectable by the 'binding experiment' were performed. Multi-exponential kinetics were observed in the dilution experiment, however, the presence of a fast phase beyond the limits of detection made it impossible to interrogate a binding mechanism.

4.3.4 IMZ binding mechanism

The results of the global fitting of the IMZ data are shown in Figure 4.3. Goodness of fit was evaluated by RSS and LL and the models were compared by their AIC and BIC values, which are summarized in Table 4.1. The RSS values spanned 0.112 to 0.853 while the AIC values ranged from -7.70E+05 to -6.19E+05. Among the models tested, the purely CS models had the highest RSS and AIC values (Model B: 3-state CS, Model E: 4-state linear CS, Model G: 4-state branched CS) which indicate they do not describe the data well compared to the remaining models. Addition of an IF component to the models improved the fits (Model D: 4-state linear CS/IF and Model H: 4-state closed CS/IF). Both the purely induced fit 4-state models best described the data (Model C: 4-state linear IF and Model F: 4-state branched IF). Compared to the 3-state IF model, both the 4-state IF models yield better fits on the basis of RSS and AIC values supporting the inclusion of an additional state as compared to the 3-state model. Moreover, the additional parameters in the 4-state models outweigh the penalties associated with the AIC and BIC calculations. Between the 4-state IF models, the 4-state branched IF model best described the data with an RSS of 0.112 and an AIC of -7.70E+05. The recovered parameters from the 4-state branched induced fit model fit are: k_1 = 140 mM^{-1} s^{-1}, k_{-1} = 76.1 s^{-1}, k_2 = 3.93 s^{-1}, k_{-2} = 12.5 s^{-1}, k_3 = 0.110 s^{-1}, k_{-3} = 0.547 s^{-1}, $[EL]_{ext}$ = 0.870, $[E^*L]_{ext}$ = 1.38, $[E^{**}L]_{ext}$ = 1.33. The results from the 4-state branched induced fit model are summarized in Figure 4.5.

4.3.5 CN^- binding mechanism

The results of the global fitting of the CN^- data are shown in Figure 4.4. Goodness of fit was evaluated by RSS and LL and the models were compared by their AIC and BIC values, which are summarized in Table 4.2. The RSS values spanned 0.356-13.5 while the AIC values ranged from -4.51+05 to -2.69E+05. Similar to the IMZ data fits, of the models tested, the purely CS models were among the poorest fits based on RSS and AIC values (Model B: 3-state CS, Model E: 4-state linear CS, Model G: 4-state branched CS). Addition of an IF component improved the fit in the linear case but worsened the fit in the closed version of the model (Model D: 4-state linear CS/IF and Model H: 4-state closed CS/IF, respectively). Similar to the IMZ data fits, the 4-state branched induced fit model, Model F, best described the data with an RSS of 0.356 and an AIC of -4.51E+05. The recovered parameters from the 4-state branched induced fit model fit are: k_1 = 6.14 μM^{-1} s^{-1} (or 6140 mM^{-1} s^{-1}), k_{-1} = 8.39 s^{-1}, k_2 = 0.179 s^{-1}, k_{-2} = 1.46 s^{-1}, k_3 = 0.101 s^{-1}, k_{-3} = 0.0949 s^{-1}, $[EL]_{ext}$ = 0.544, $[E^*L]_{ext}$ = 3.09, $[E^{**}L]_{ext}$ = 1.19. The results from the 4-state branched induced fit model are summarized in Figure 4.5.

4.3.6 SVD analysis of stopped-flow kinetic measurements

Mechanistic analysis of the binding and dissociation kinetics suggested that an ensemble of ligand-bound states exists. Therefore, the kinetic data were analyzed by SVD to aid in potential identification of multiple distinct spectral components. The results from the SVD analysis for each ligand are shown in Figures 4.6 – 4.7. In a perfect experiment with noiseless data, the number of components with nonzero singular values reflect the number of species in the kinetic data. In practice, SVD analysis of experimental data will yield 'noise' components with near zero singular

values in addition to the desired spectral components. The components with the top four singular values are presented herein. The spectral and amplitude components do not present as 'random noise.'

4.4 Discussion

This chapter characterizes the mechanism of ligand binding for three small molecules (1,2,3-TRZ, IMZ, and CN^-) binding to CYP3A4 via a global kinetic analysis approach. All three ligands displayed multiphasic binding kinetics, however, the kinetics for 123-TRZ were too fast to be interrogated mechanistically. The analyses revealed that, even for small molecules, the ligand binding kinetics require complex multi-state binding models to adequately describe the data, which highlights the complexity in the binding interactions of low-spin complexes at the active site of CYP3A4. In general, the kinetics were best described by a 4-state induced fit binding model. These results provide evidence for an ensemble of bound states which may be attributed to heterogeneity in the ligation state of low-spin complexes at the active site heme. Quantitative estimates for the relative amounts of each bound species and estimates for their rates of interconversion were presented. In addition, a framework for global kinetic analyses with a comprehensive list of 8 models to test for cases was presented for cases when single-site ligand binding is expected.

The kinetics were best described by a 4-state induced fit binding model. Between the branched and linear versions of the model (Models C and F, respectively), the 4-state branched model best described the data with the lowest RSS, AIC, and BIC values for all models tested. A surprising result was that the fits for the CN^- data were inferior to the IMZ fits despite a significantly higher signal to noise ratio in the CN^- binding data. Additionally, the extinction coefficients were higher for each bound species compared to the IMZ fits, which were all close to

1 ($[ECN^-]_{ext}$ = 0.544, $[E^*CN^-]_{ext}$ = 3.09, $[E^{**}CN^-]_{ext}$ = 1.19 vs $[E\text{-}IMZ]_{ext}$ = 0.870, $[E^*IMZ]_{ext}$ = 1.38, $[E^{**}IMZ]_{ext}$ = 1.33). This suggests that perhaps a more complicated model is needed to describe the CN^- data. However, since a triple exponential was sufficient to describe the data, expansion of the analysis to include 5 state models (which require at least a quadruple exponential to describe the data) would be difficult without more data. Therefore, the implications from the current model fits are discussed below.

Before discussing the individual models in detail, mechanistic differences from a recent study must be addressed. The results in this work suggest that the ligands in this study readily bind to all CYP3A4 in solution (induced fit). This is counter to a recent study by Guengerich and coworkers wherein different drugs were found bind to CYP3A4 through a conformational selection model, which requires one enzyme conformer from an ensemble of states to not bind to ligand [72]. The methods in the study were similar to those presented here. While these experiments may provide evidence for a pre-existing ensemble of enzyme conformers, this is only possible for cases where the different conformations have distinct ligand binding properties. Stated more simply, the data are a reflection of what the ligand 'sees;' the spectroscopic signal reflects the ligand-induced change in the active site heme-iron spin-state and the kinetics may or may not be influenced by the conformational dynamics of the enzyme. The conformational dynamics may alter solvent accessibility, ligand binding channels or tunnels, auxiliary binding pockets, etc. Thus, the binding properties for each enzyme conformation in the ensemble may be ligand dependent. This is reflected in the study by Guengerich and coworkers wherein different rate constants for the rearrangement between unbound enzyme conformations (Figure 1B: k_1 and k_{-1}) were reported for different ligands with CYP3A4. In this work, it is possible that the ligands are sufficiently small such that ligand is able to access the active site-heme regardless of the relative dynamics of

possible CYP3A4 conformers. In the context of the CS/IF model comparison, perhaps the small molecules in this study bind to every member in the ensemble of CYP3A4 conformations in a similar manner. This becomes a case of parallel binding processes with similar kinetics, which can be condensed as described below. Consider the case where two unbound conformations of CYP3A4 exist in solution, E_1 and E_2, and bind to ligand through the same mechanism with equivalent rate constants as shown below.

Ensemble in the absence of ligand: $E_1 \rightleftarrows E_2$

Ligand binding for conformation 1: $E_1 \rightleftarrows E_1L \rightleftarrows E^*L \rightleftarrows E^{**}L$

Ligand binding for conformation 2: $E_2 \rightleftarrows E_2L \rightleftarrows E^*L \rightleftarrows E^{**}L$

Given that the rate constants are equivalent for ligand interactions with either conformation, the ligand binding model condenses to an induced fit model as shown below:

General ligand binding model: $E \rightleftarrows EL \rightleftarrows E^*L \rightleftarrows E^{**}L$

I would like to emphasize that 'induced fit' in experiments in this study, merely suggests that ligand is able to bind all enzyme conformations with similar binding properties. Note that a similar simplification can be made for the CS model provided E and E* (in Figure 4.1B) represent multiple conformations with similar binding properties, or lack thereof. Here it is important to make a clear distinction between IF and CS: IF is not definitive evidence against the existence of an ensemble of enzyme conformations. Whereas CS provides definitive evidence for a pre-existing ensemble of at least 2 distinct conformations, one of which ligand cannot bind to, however the relative amounts of each species (E and E*) can be ligand dependent depending on the selection criteria that the conformational dynamics influence. That is, different ligand types might interact with different subsets of enzyme conformations and thus one might recover apparently different rates for their interconversion (i.e. different rates for the $E \rightleftarrows E^*$ rearrangement as was seen by

Guengerich and coworkers). In this work, I do not refute the possibility of a pre-existing ensemble of enzyme conformations in the absence of ligand.

A summary of the recovered rate constants from the fits to the 4-state branched IF models are shown in the left panels of Figure 4.5. Comparison of the on- and off-rates for each ligand (Figure 4.5 step 1: $k_{on} = k_1$, $k_{off} = k_{-1}$), revealed that CN^-, the ligand with the highest affinity, had the fastest on-rate (CN^- k_{on} ~44x greater than IMZ k_{on}) and the slowest off-rate (CN^- k_{off} ~9x less than IMZ k_{off}). Notably, the results here for CN^- binding are consistent with measurements by Denisov and coworkers (T = 6 ºC, k_{on} = 40 μM^{-1} s-1 and k_{off} = 4.0 s^{-1}; concentration corrected for CN^- here from total CN reported in [71]). Moreover, the k_{on} and k_{off} for CN^- binding (T = 10.6 ºC, k_{on} = 6.14 μM^{-1} s-1 and k_{off} = 8.39 s^{-1}) are on the upper and lower end of reported rate constants for various ligands binding to CYP3A4, respectively [72].

Regarding the ensemble of ligand-bound states, the relative rates of their interconversion (Figure 4.5 steps 2 and 3) were more similar in magnitude compared to the binding step (Figure 4.5 step 1), which was much faster. Between the two rearrangement steps, the slowest steps (Figure 4.5 step 3) were comparable between the ligands (k ~0.1 – 0.5 s^{-1}). The model suggests that the ensemble of bound states have different equilibrium profiles for each ligand. The percentages next to each bound species represent their equilibrium amounts, which are visually represented by qualitative free energy diagrams in the right panel of Figure 4.5. The dominant species differed between the ligands (IMZ: EL = 66%, CN^-: E**L = 49%). The major species for IMZ was the initially formed ligand-bound complex, EL. The major species for CN^- was one of the branched rearrangement species, E**L, associated with the slow rearrangement step (step 3), however, the initially formed ligand-bound complex was nearly equal in amount (E**L = 49% vs EL = 46%). In the model, ligand rapidly binds and initially populates the EL state, followed by a relatively

slow rearrangement to form either the E*L or the E**L complex, hereafter referred to as 'branched species.' Between the 'branched species,' the 'kinetically favored' species, E*L, was preferred for IMZ (E*L = 21% vs E**L = 13%) while the 'thermodynamically favored species,' E**L, was overwhelmingly preferred for CN^- (E*L = 5% vs E**L = 49%). Speculatively, because the thermodynamic and kinetic profiles differed between the ligands, it is possible that the types of heme-ligand interactions differ between ligands (i.e. directly coordinated species, water bridged, etc.).

To explore the possibility that an ensemble of ligand-bound complexes was generated, as suggested from the model fits, the spectra were analyzed by SVD with the goal of identifying the number of spectral components in the kinetic experiments. The resulting spectral and concentration vectors from each component are shown in Figures 4.6 – 4.7. The magnitude of each singular value, S, is often an indication of the relative importance of each component. One may be tempted to include only the largest components from the analysis, however, omission of components with low singular values is not always recommended because short-lived intermediates, which may be significant to the mechanism, tend to have low singular values. Thus, the four components with the highest singular values that had non-zero spectral and concentration vectors were included. Comparison of the four spectral components reveal two components that appear to be inverted CYP spectra (components 1 and 3 which are colored blue and yellow, respectively) and two components that resemble difference spectra (components 2 and 4 which are colored red and purple, respectively). Comparison of the corresponding concentration vectors reveal that the first component, which has the highest singular value, appears to be relatively constant compared to the other components. This is because the first component is often the best approximation of all the data while the remaining components describe the differences over time. Matching simulations

of the kinetic data using the recovered parameters from the fits are included to compare the concentration vectors from the SVD analysis to the concentration profiles of individual species under the matching simulated conditions ([E_{total}] = 1 μM, [Ligand] matches the experimental concentration). The SVD are not a perfect match, which is commonly observed for CYPs. However, in all ligands, the concentration vector of component 4 resemble the E*L species. It is possible that E*L has slightly different spectral properties and is the reason that the SVD analysis was able to identify the component. Regardless, the observation of a time-dependent decrease in the concentration profile of a species is highly suggestive of an IF component in the binding mechanism. An interesting experiment for future work would be to perform rapid freeze-quench EPR experiments in an attempt to monitor the transient nature of each species.

A recent CW EPR study by Lockart and coworkers with several low-spin compounds and CYP3A4 revealed a mixture of low-spin complexes from flash-frozen samples of equilibrated ligand/protein solutions [23]. IMZ was among the ligands in the study and deconvolution of the spectra revealed 78.4% contribution from a directly coordinated low-spin species and 21.6% from an unknown ligation mode that could not be unambiguously assigned due to spectral overlap. The authors note that for the unknown ligation species, at least one complex contained water or was water bridged. Speculatively, it is possible that the E*L species (21%), which has faster associated rate constants of interconversion, may be related to the unknown ligation mode in the CW EPR study, and the remaining signal may be associated with directly ligated binding poses, EL and E**L (66% + 13% = 79%). If this is true, this suggests that direct coordination or ‘true type II’ interactions have longer residence time and are thermodynamically favored over a water-bridged complex. Another possibility is that the branched conformations, E*L and E**L, correlate to the unknown species (two different ligation modes: water-bridged and/or water containing mode =

21% + 13% = 34% ‘unknown’; 66% directly coordinated) and discrepancies in the final percentages can be explained by a temperature dependent shift in the ratios of each species (CW EPR experiments performed at room temperature vs kinetic experiments in this study performed at 10.6 ºC). Regardless, the results highlight different kinetic and thermodynamic properties for each member in the ensemble of bound complexes. If the final ratio of ligation modes for ‘type II ligands’ are ligand specific, this has potential functional implications in the context of drug metabolism for drugs that contain an IMZ moiety, if the different binding modes have different catalytic properties. This is consistent with the well-accepted that CYP3A4 is capable of forming multiple products for a given compound. This suggests that the substrate is able to access multiple binding orientations in the active site.

In the context of drug metabolism, the rate constants associated with step 3 were between 0.1 – 0.5 s^{-1} (6 min^{-1} – 30 min^{-1}) which are in the same time regime as k_{cat} for CYP3A4 and a majority of ligands. The ligands in this study were selected with the rationale that they are sufficiently small enough to not interact with any active site protein residues such that the kinetics for any possible rearrangements be governed by local electronic factors. That is, I hypothesized that the intrinsic rate constants for the rearrangement processes will be in the upper range compared to those expected from large drugs or drug-like molecules that have chemical handles capable of interacting with active site residues. If this is true, it is possible that type II compounds may form an ensemble of bound complexes whose rates of interconversion are comparable to their rates of metabolism and thus nullify rapid equilibrium assumptions. An interesting scenario arises when each of the different bound states (EL, E*L, and E**L), which binding with the same affinity, have different catalytic properties, be it differences in sites of metabolism, product formation rates, or even binding affinities for the required redox partner proteins. One possibility is that the different

bound complexes could potentially display non-Michaelis-Menten or 'atypical' kinetic profiles. Thus, detailed knowledge of the mechanism of ligand binding may help clarify some cases of multifunctionality observed in CYP3A4.

Related to the catalytic cycle, CN^- was selected in this study as a surrogate for oxygen (O_2). CN^- preferentially binds to the ferric form of CYPs whereas O_2 only binds to the ferrous form. Given the instability of the ferrous form of enzyme and the inherent complexities required to measure the kinetics of O_2 binding (reduction by redox partner proteins or by synthetic means, anaerobic conditions, difficulty regulating amounts of oxygen in solution, etc), I used CN^- as a surrogate diatomic ligand. The ferric-CN^- complex is considered as an O_2 mimic because it has been shown to reproduce structural changes observed in the I-helix of O_2-bound P450cam, detailed in crystallographic studies[154]. Moreover, CN^- is considered a better mimic than other diatomic ligands such as CO and NO because the negative charge in the distal atom (CN^- coordinates heme via an electron rich carbon) is able to participate in a hydrogen bonding network resulting from the reorientation of an Asp251 residue in the I-helix of P450cam. It is possible that the nature of the ensemble of bound states in our studies, here with CYP3A4, is a consequence of an induced hydrogen bonding network. Another potentially related possibility is fluctuations between bent and linear binding geometries, which have been observed by Resonance Raman spectroscopy in P450cam [155]. Unfortunately, comparisons of population distribution are not possible as the authors mention that the different species are sensitive to many factors that make derivation of their relative amounts difficult. Despite the body of evidence that has been presented for heterogeneity in CN^- ligation for P450cam, it is still possible that the results may not be fully recapitulated in the more complicated CYP3A4 isoform. The results here present interesting possibilities and more questions: Are similar reorientations possible for the ferrous-oxy complex

in CYP3A4? If an ensemble of ferrous-oxy complexes is generated, do different ligands or partner proteins influence the population distribution of the ensemble? Does this influence catalysis? Does this alter the propensity for uncoupling reactions?

The results presented here provide kinetic validation for the existence of equilibrating states within an ensemble of bound complexes for 123-TRZ, IMZ, and CN^- interacting with CYP3A4. Each ligand displayed multiphasic kinetics. Unfortunately, the binding and dissociation kinetics for 123-TRZ had components that were beyond the levels of detection for the instrument. However, IMZ and CN^- displayed kinetics that were able to be interrogated across multiple ligand concentrations and experimental setups. The data for IMZ and CN^- required complex 4-state binding models to describe the data, despite the small size of the compounds. The data for IMZ and CN^- were best described by a 4-state branched induced fit model. Together the kinetic results further reveal the complexity of the behavior of low-spin complexes at the active site of CYPs, provide quantitative estimates for the relative contributions of the states at equilibrium, and reveal the estimates for the rates of exchange between these states.

4.5 Figures

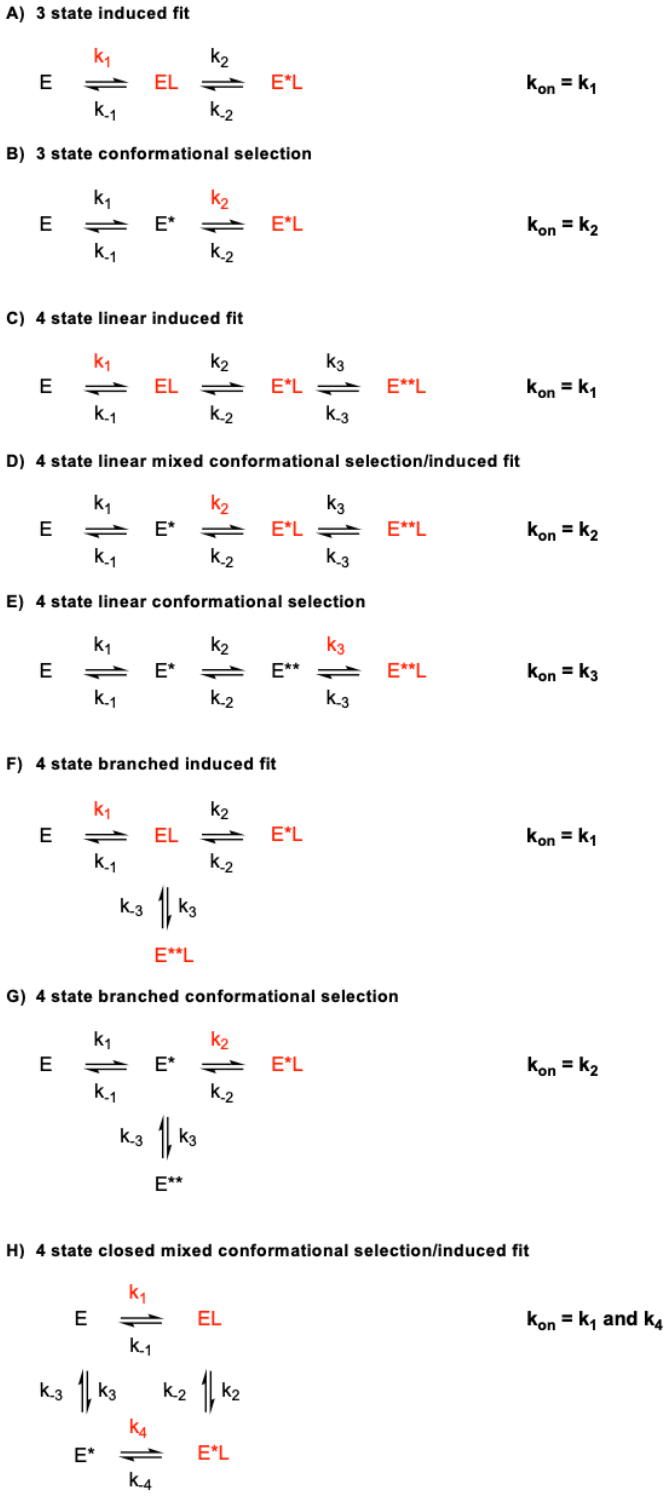


Figure 4.1. Ligand binding kinetic schemes tested in this study. The ligand binding step and ligand-bound enzyme states are highlighted in red text.

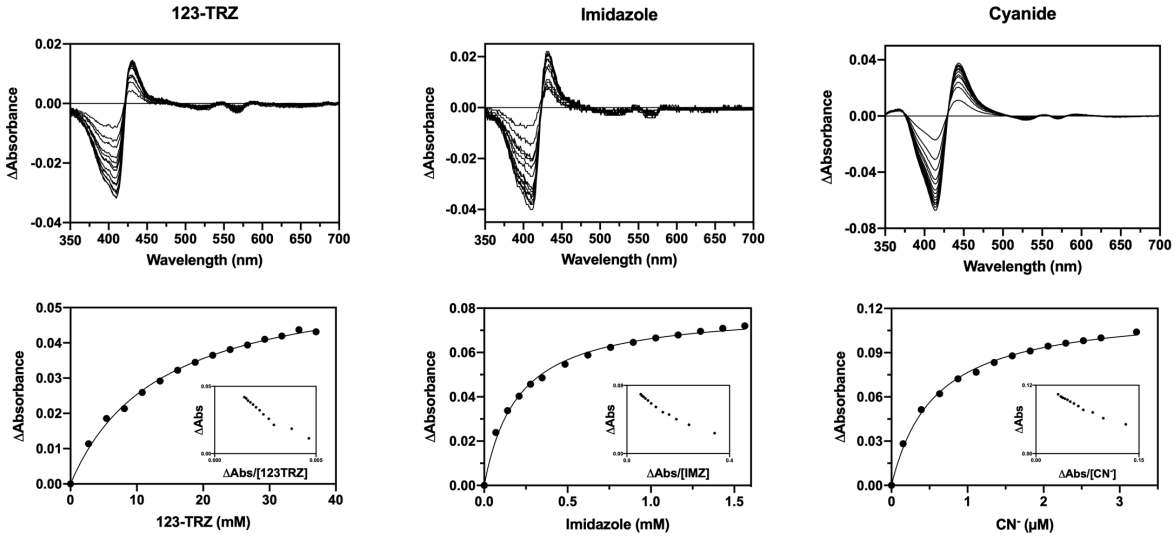


Figure 4.2. Difference spectra and resultant binding isotherms for 123-TRZ, IMZ and CN^- binding to CYP3A4. UV-Vis equilibrium titrations (10.6 ºC) of purified CYP3A4 (0.9 μM). The difference spectra calculated relative to the ligand-free spectra are shown in the top row. The resulting binding isotherm is shown on the bottom row (Change in absorbance vs Concentration) with inset Eadie-Hofstee plots (Change in absorbance vs change in absorbance/ligand concentration). The A_{max} and K_D for 123-TRZ was 0.0594 ± 0.00273 and 13.42 ± 1.5 mM. The A_{max} and K_D for IMZ was 0.0796 ± 0.0025 and 201 ± 23.3 μM. The A_{max} and K_D for CN was 0.120 ± 0.0035 and 566 ± 59.6 nM.

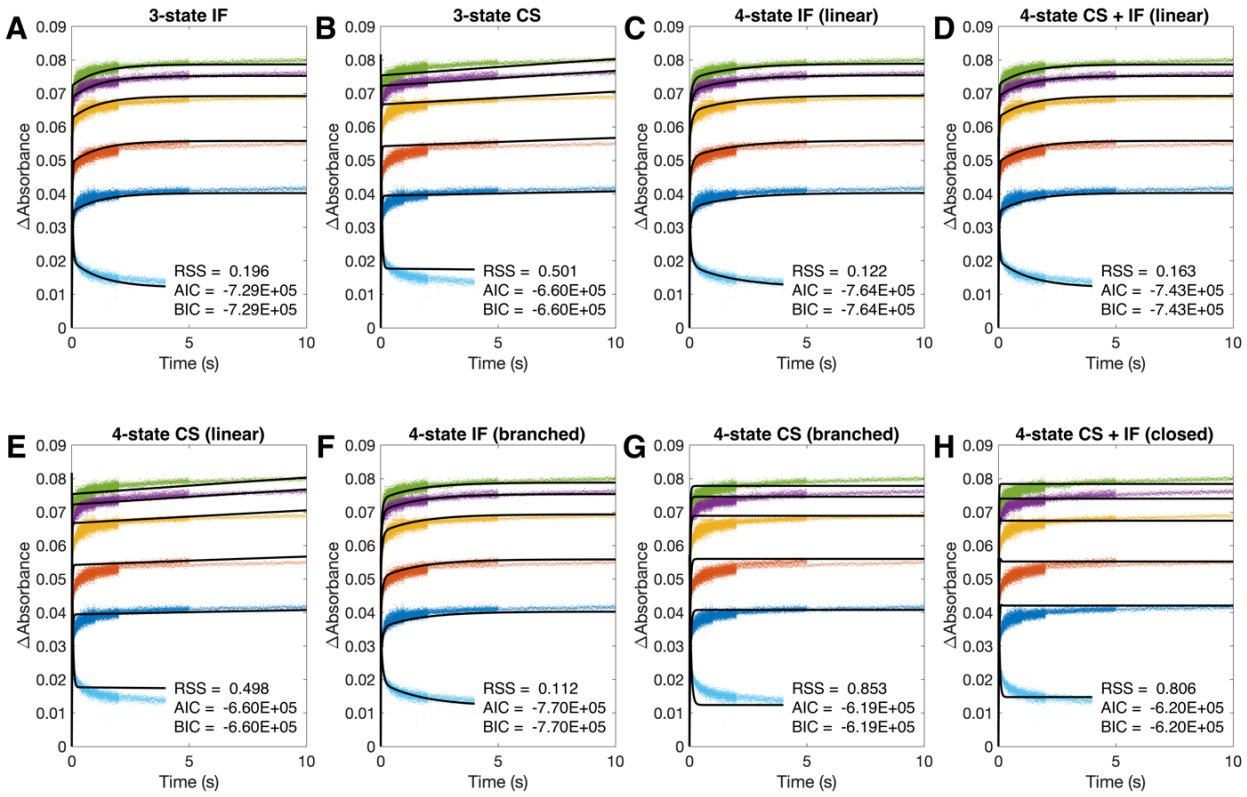


Figure 4.3. IMZ binding and dissociation kinetics. Plots of ΔA430-410 vs time (s) for imidazole binding and dissociating from CYP3A4. The model fits are depicted as black lines and the data from replicate experiments are depicted as colored dots. The recovered parameters for each model fit are shown in Table 2. Binding experiments: CYP3A4 was mixed with imidazole at 0.283 mM (blue), 0.566 mM (orange), 1.13 mM (yellow), 1.70 mM (purple) and 2.26 mM (green). Dissociation experiment: Data from an 11x dilution experiment are shown in light blue ($[IMZ]_{final}$ = 0.054 mM). A) 3-state induced fit model B) 3-state conformational selection model C) 4-state linear induced fit model D) 4-state mixed conformational selection/induced fit model E) 4-state linear conformational selection model F) 4-state branched induced fit model G) 4-state branched conformational selection model H) 4-state closed mixed conformational selection/induced fit model.

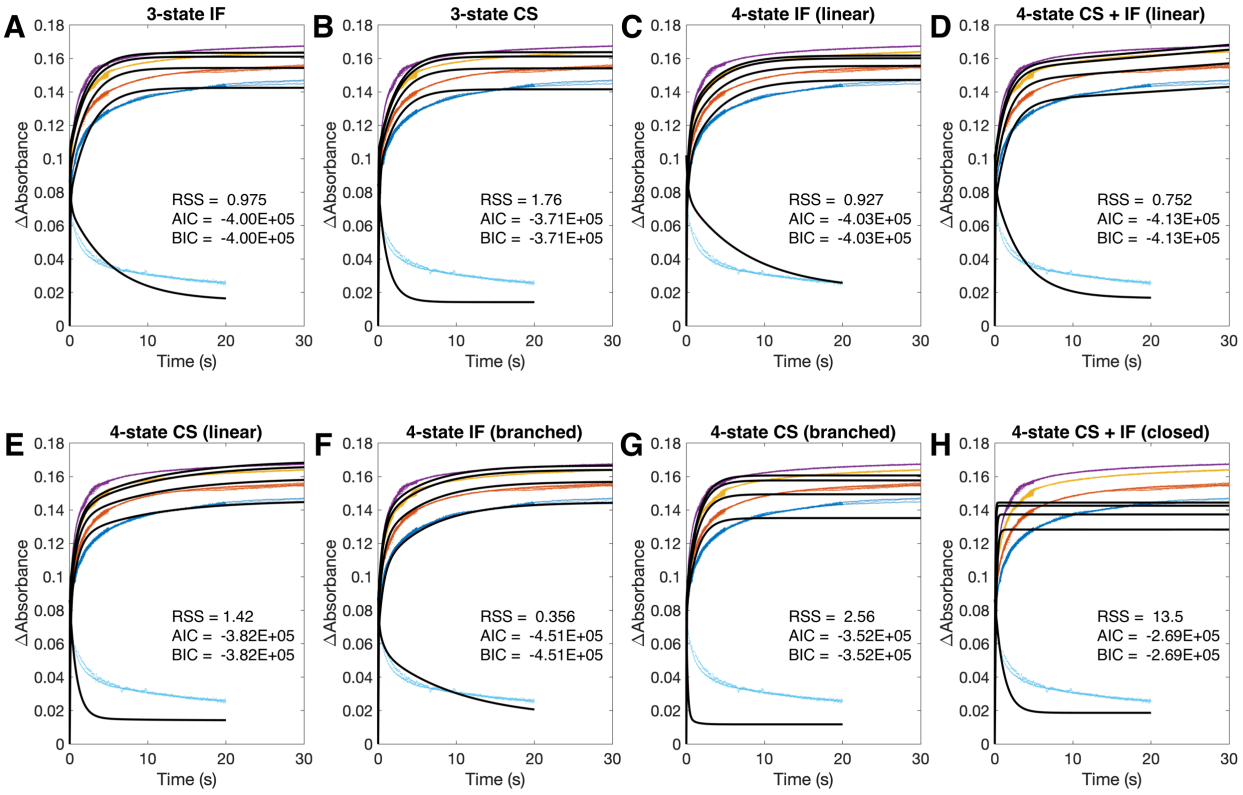


Figure 4.4. CN^- binding and dissociation kinetics. Plots of ΔA443-413 vs time (s) for imidazole binding and dissociating from CYP3A4. The model fits are depicted as black lines and the data from replicate experiments are depicted as colored dots. Binding experiments: CYP3A4 was mixed with cyanide at 3.30 µM (blue), 6.60 µM (orange), 13.2 µM (yellow), and 19.8 µM (purple). Dissociation experiment: Data from an 11x dilution experiment are shown in light blue ($[CN^-]_{final}$ = 0.0589 µM). A) 3-state induced fit model B) 3-state conformational selection model C) 4-state linear induced fit model D) 4-state mixed conformational selection/induced fit model E) 4-state linear conformational selection model F) 4-state branched induced fit model G) 4-state branched conformational selection model H) 4-state closed mixed conformational selection/induced fit model.

Table 4.1 Model parameters and statistics for IMZ binding to CYP3A4. The recovered rate constants, k_n, and relative extinction coefficients for the bound species, $[Species]_{ext}$, from individual model fits are shown in each column. The units for the first- and second-order rate constants have units of s^{-1} and mM^{-1} s^{-1}, respectively. For ease of comparison, the models are ranked from the best, 1, to the worst, 8, fit by the AIC and BIC values. The association rate, k_{on}, for each model is underlined. The model with the best fit, as determined by the lowest AIC and BIC values (n=73,000), is indicated in bold text. Each model is shown in Figure 4.1.

	3-state		**4-state**					
	A IF	**B** CS	**C** IF (linear)	**D** CS+IF (linear)	**E** CS (linear)	**F** IF (branch)	**G** CS (branch)	**H** CS+IF (closed)
k_1	<u>55.9</u>	0.00760	<u>101</u>	59.1	0.0253	**<u>140</u>**	69.3	<u>42.8</u>
k_{-1}	23.0	0.00779	45.0	24.1	90.4	**76.1**	41.6	48.3
k_2	0.115	<u>63.8</u>	0.894	<u>101</u>	38.9	**3.93**	<u>66.3</u>	19.2
k_{-2}	0.766	21.6	8.46	30.0	0.0112	**12.5**	13.4	23.7
k_3	-	-	0.693	0.120	<u>64.5</u>	**0.110**	1.52	42.6
k_{-3}	-	-	0.529	0.710	21.9	**0.547**	21.2	0.395
k_4	-	-	-	-	-	**-**	-	<u>57.7</u>
k_{-4}	-	-	-	-	-	**-**	-	1.59
$K_{D,app}$	0.358	0.686	0.358	0.357	0.688	**0.358**	0.337	0.287
$[EL]_{ext}$	0.971	2.00	0.884	0.98	2.00	**0.870**	1.02	0.697
$[E^*L]_{ext}$	1.46	-	2.18	1.36	-	**1.38**	-	5.08
$[E^{**}L]_{ext}$	-	-	1.28	-	-	**1.33**	-	-
RSS	0.196	0.501	0.122	0.163	0.498	**0.112**	0.853	0.806
LL ($x10^5$)	3.65	3.30	3.82	3.71	3.30	**3.85**	3.10	3.10
AIC ($x10^5$)	-7.29	-6.60	-7.64	-7.43	-6.60	**-7.70**	-6.19	-6.20
BIC ($x10^5$)	-7.29	-6.60	-7.64	-7.43	-6.60	**-7.70**	-6.19	-6.20
Rank	4	6	2	3	5	**1**	8	7

k_{on} = 140E05 M^{-1} s^{-1}, k_{off} = 76.1 s^{-1}, $K_D = k_{off}/k_{on}$ = 0.544 mM, $K_{D,app}$ = 0.358 mM

Table 4.2 Model parameters and statistics for CN^- binding to CYP3A4. The recovered rate constants, k_n, and relative extinction coefficients for the bound species, $[Species]_{ext}$, from individual model fits are shown in each column. The units for the first- and second-order rate constants have units of s^{-1} and mM^{-1} s^{-1}, respectively. For ease of comparison, the models are ranked from the best, 1, to the worst, 8, fit by the AIC and BIC values. The association rate, k_{on}, for each model is underlined. The model with the best fit, as determined by the lowest AIC and BIC values (n=50,000), is indicated in bold text. Each model is shown in Figure 4.1.

	3-state		**4-state**					
	A IF	**B** CS	**C** IF (linear)	**D** CS+IF (linear)	**E** CS (linear)	**F** IF (branch)	**G** CS (branch)	**H** CS+IF (closed)
k_1	<u>3.07</u>	0.416	<u>8.15</u>	0.00340	0.101	**<u>6.14</u>**	0.667	<u>1.32</u>
k_{-1}	4.84	0.240	8.20	0.00286	0.0464	**8.39**	0.574	1.04
k_2	0.298	<u>1.99</u>	0.207	<u>3.94</u>	1.02	**0.179**	<u>12.7</u>	6.13
k_{-2}	0.183	0.811	0.495	6.35	0.596	**1.46**	5.05	16.3
k_3	-	-	1.60	0.410	<u>2.89</u>	**0.101**	0.829	0.399
k_{-3}	-	-	0.611	0.287	1.05	**0.0949**	8.16	3.10
k_4	-	-	-	-	-	**-**	-	<u>6.09</u>
k_{-4}	-	-	-	-	-	**-**	-	0.100
$K_{D,app}$	0.601	0.641	0.400	1.22	0.674	**0.625**	0.777	0.474
$[EL]_{ext}$	0.674	0.986	0.490	1.18	1.02	**0.544**	0.974	2.91
$[E^*L]_{ext}$	1.17	-	4.17	2.12	-	**3.09**	-	15.6
$[E^{**}L]_{ext}$	-	-	0.167	-	-	**1.19**	-	-
RSS	0.975	1.76	0.927	0.752	1.42	**0.356**	2.56	13.5
LL ($x10^5$)	2.00	1.85	2.01	2.07	1.91	**2.25**	1.76	1.34
AIC ($x10^5$)	-4.00	-3.71	-4.03	-4.13	-3.82	**-4.51**	-3.52	-2.69
BIC ($x10^5$)	-4.00	-3.71	-4.03	-4.13	-3.82	**-4.51**	-3.52	-2.69
Rank	4	6	3	2	5	**1**	7	8

k_{on} = 6.14E06 M^{-1} s^{-1}, k_{off} = 8.39 s^{-1}, $K_D = k_{off}/k_{on}$ = 1.36 μM, $K_{D,app}$ = 0.625 μM

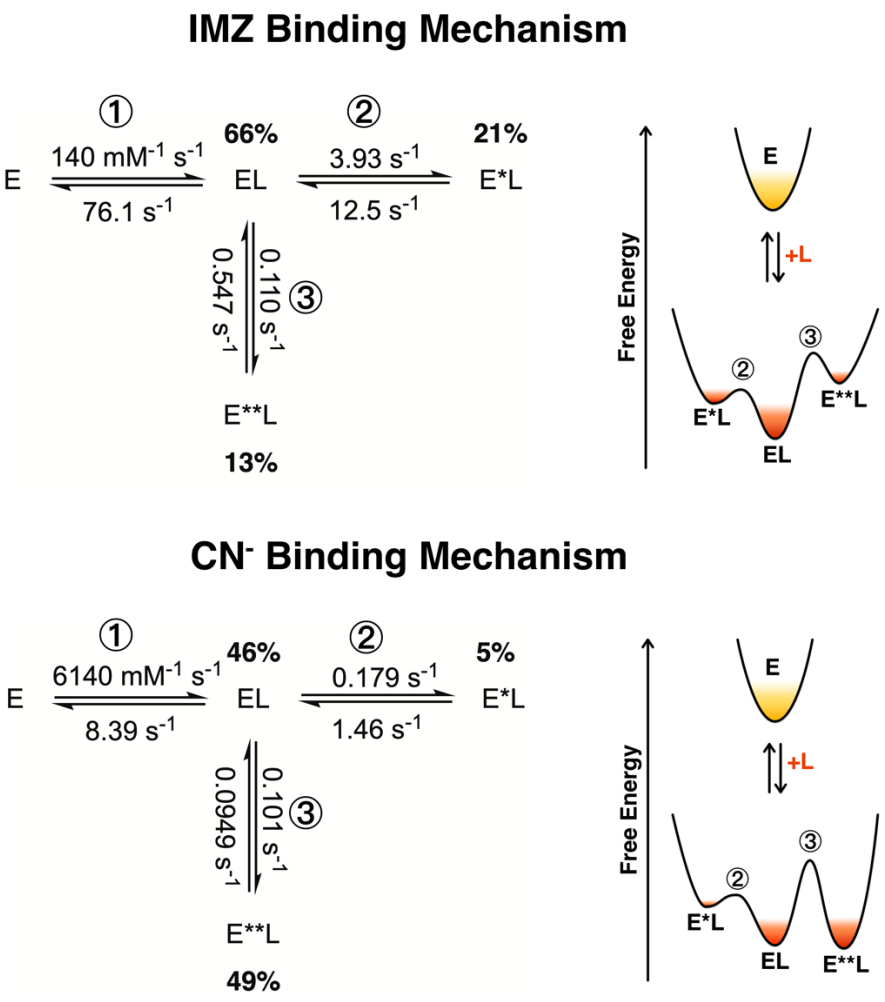


Figure 4.5. Summary of ligand binding mechanisms for IMZ and CN^- binding to CYP3A4. The recovered rate constants from the models that best described the data are shown. The percentages next to each species reflect their equilibrium amounts. The models are summarized in the free energy diagram, which detail the energy landscape profile for the unbound complex, and the three ligand-bound complexes.

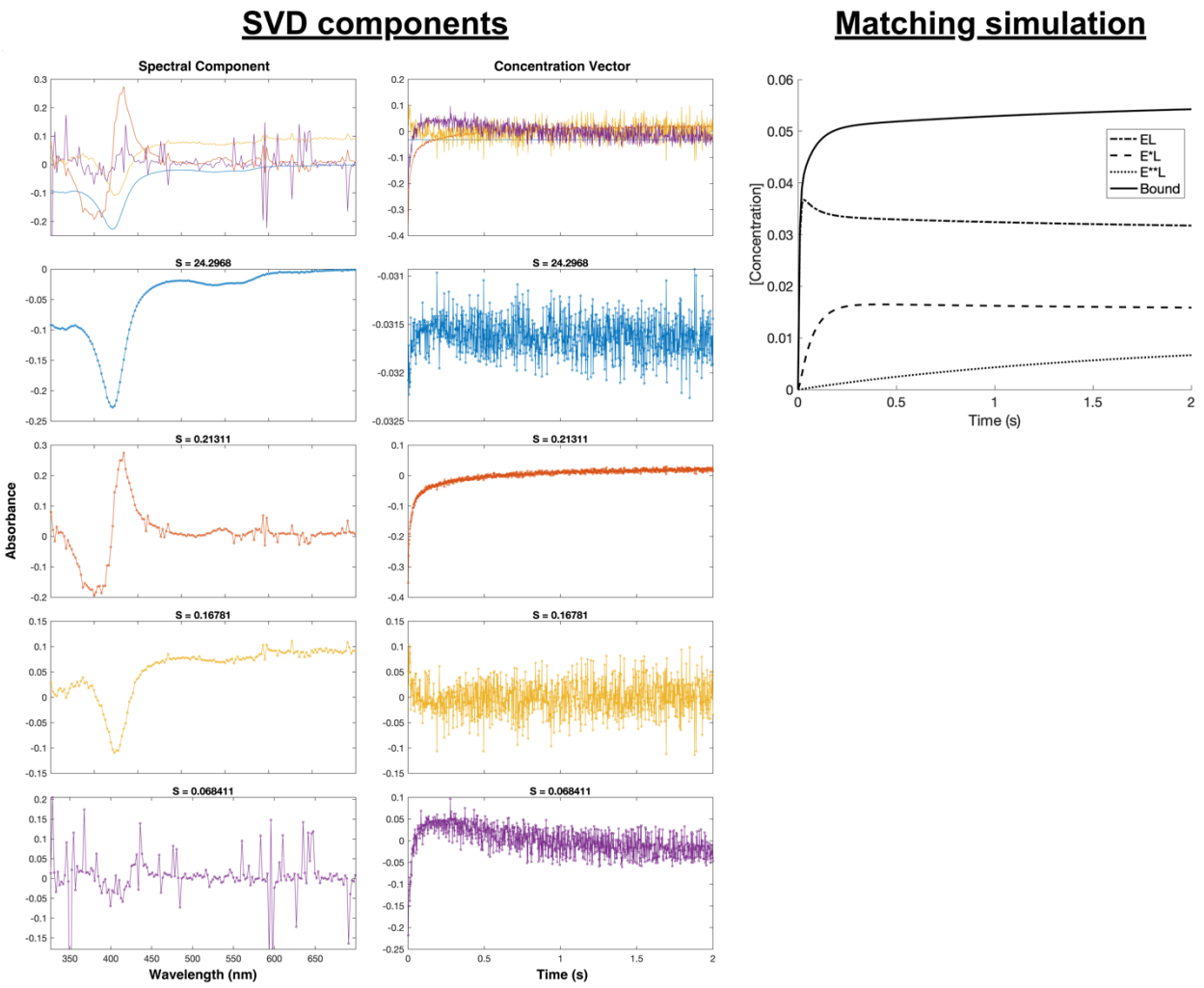


Figure 4.6. SVD analysis of IMZ binding data. Extracted spectral components and corresponding amplitude vectors obtained from SVD of IMZ binding data. Left) The top panel shows an overlay of each component. Individual components are shown in subsequent plots with the singular values indicated above each plot. Right) Matching simulated data using the recovered rate constants and extinction coefficients from the 4-state branched induced fit model in Table 4.1.

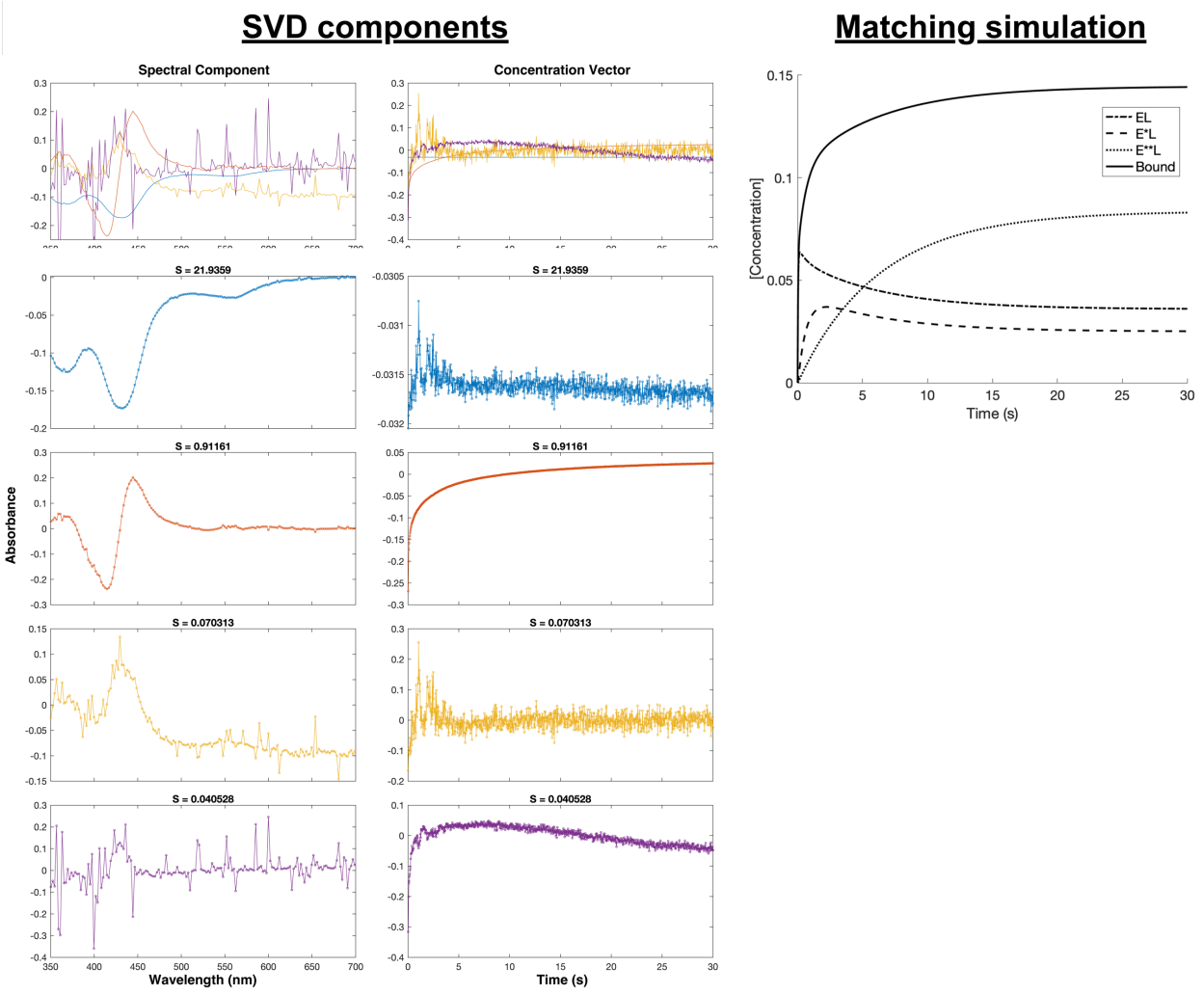


Figure 4.7. SVD analysis of CN^- binding data. Extracted spectral components and corresponding amplitude vectors obtained from SVD of CN^- binding data. Left) The top panel shows an overlay of each component. Individual components are shown in subsequent plots with the singular values indicated above each plot. Right) Matching simulated data using the recovered rate constants and extinction coefficients from the 4-state branched induced fit model in Table 4.2.

Appendix

The following pages list the differential equations, initial enzyme amounts, apparent binding affinities, and parameters for each model as input into MATLAB for fitting.

Differential equations:

At the beginning of each section, a generic scheme is shown which details the nomenclature for each enzyme. For example, in the generic 3-state mechanism, E1 corresponds to state "a," associated with the differential equation da/dt, initial protein concentration E1, and extinction coefficient, extA.

Initial protein concentrations:

The initial amounts of protein are useful for conformational selection cases where a pre-existing ensemble of unbound enzyme conformations exist. Initial protein amounts are also important for dilution experiments where enzyme is pre-incubated with ligand at an initial concentration, Linit, and diluted. For a binding experiment (mix unbound enzyme with ligand), Linit = 0.

Binding affinity:

The expressions for the apparent binding affinity, $K_{D,app}$, for each model are shown.

Parameters:

The kinetic parameters are listed as well as the corresponding extinction coefficients for each state. The extinction coefficient for the unbound states was set to zero (i.e. extA = 0 for the 3-state

induced fit model; extA = extB = 0 for the 3-state conformational selection model, etc). Where appropriate, the states with extinction coefficients set to zero are indicated for each model.

LIST OF MODELS

2 state – Generic 206

2 state – Isomerization 207

2 state – Binding 208

3 state – Generic 209

3 state – Induced Fit 210

3 state – Conformational Selection 211

3 state – Sequential Binding 212

3 state – Substrate Reorientation 213

4 state – Persistent heterogeneity 214

4 state – Linear Generic 215

4 state – Linear Induced Fit 216

4 state – Linear Mixed Conformational Selection/Induced Fit 217

4 state – Linear Conformational Selection 218

4 state – Linear Induced Fit with 2 Binding Sites (A) 219

4 state – Linear Induced Fit with 2 Binding Sites (B) 220

4 state – Linear Conformational Selection with 2 Binding Sites 221

4 state – Branched Generic 222

4 state – Branched Induced Fit 223

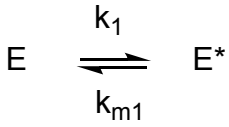
4 state – Branched Conformational Selection 224

4 state – Closed Generic 225

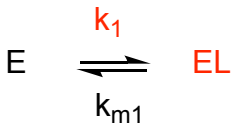
4 state – Closed Conformational Selection/Induced Fit 227

Summary of Models:

2 state isomerization (2 rates) – page 207



2 state binding (2 rates) – page 208

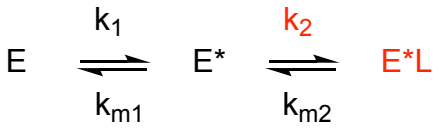


$k_{on} = k_1$

3 state induced fit (4 rates) – page 210

$k_{on} = k_1$

3 state conformational selection (4 rates) – page 211

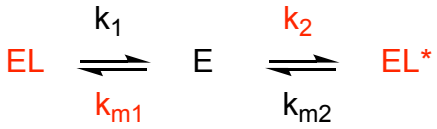


$k_{on} = k_2$

3 state sequential binding (4 rates) – page 212

$k_{on} = k_1$ and k_2

3 state substrate reorientation (4 rates) – page 213



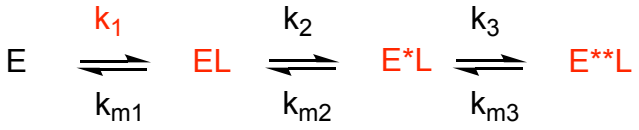
$k_{on} = k_{m1}$ and k_2

4 state persistent heterogeneity (4 rates) – page 214

$k_{on} = k_1$

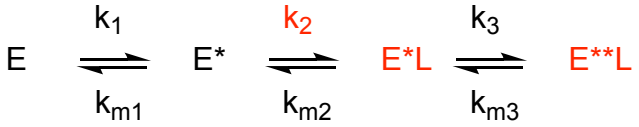
$k_{on} = k_2$

4 state linear induced fit (6 rates) – page 216



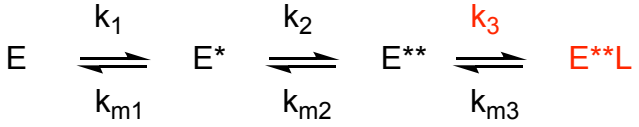
$k_{on} = k_1$

4 state linear mixed conformational selection/induced fit (6 rates) – page 217



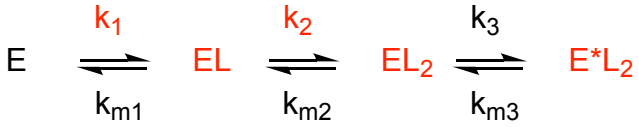
$k_{on} = k_2$

4 state linear conformational selection (6 rates) – page 218



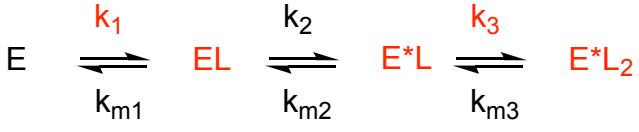
$k_{on} = k_3$

4 state induced fit with two binding sites (A) (6 rates) – page 219



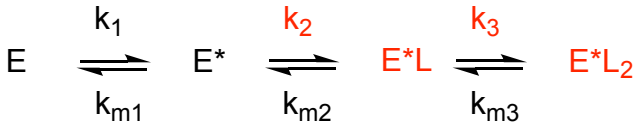
$k_{on} = k_1$ and k_2

4 state induced fit with two binding sites (B) (6 rates) – page 220



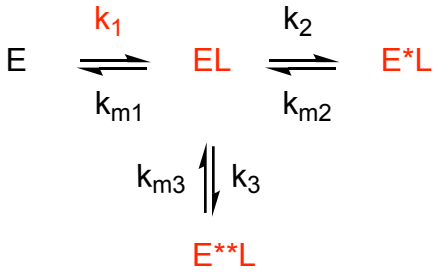
$k_{on} = k_1$ and k_3

4 state conformational selection with two binding sites – page 221



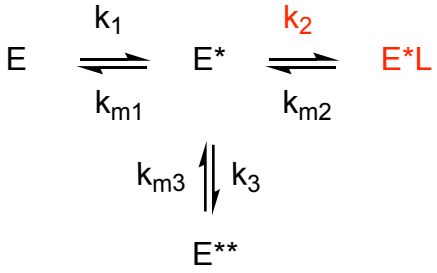
$k_{on} = k_2$ and k_3

4 state branched induced fit (6 rates) – page 223



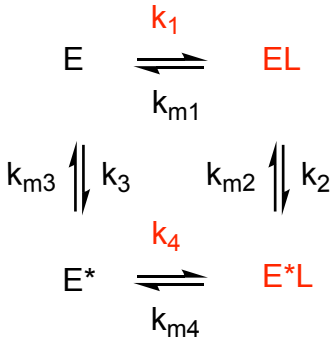
$k_{on} = k_1$

4 state branched conformational selection (6 rates) – page 224



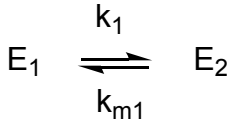
$k_{on} = k_2$

4 state closed mixed conformational selection/induced fit (8 rates) – page 227



$k_{on} = k_1$ and k_4

2 state – Generic




```
dE1/dt = dadt
dE2/dt = dbdt
```

Differential equations:

```
dadt=-k1*a+km1*b;
dbdt=k1*a-km1*b;
dYdt=[dadt;dbdt];
```

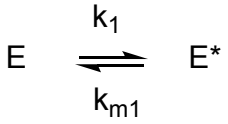
Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*(km1)/(km1+k1);
E2=Etotal*(k1)/(km1+k1);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
extA=K(3);
extB=K(4);
```

2 state – Isomerization



Rate constants:

k_f = k1
k_r = km1

Differential equations:

```
dadt=-k1*a+km1*b;
dbdt=k1*a-km1*b;
dYdt=[dadt;dbdt];
```

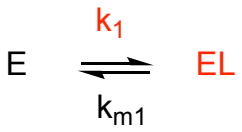
Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*(km1)/(km1+k1);
E2=Etotal*(k1)/(km1+k1);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
extA=K(3);
extB=K(4);
```

2 state – Binding



$k_{on} = k_1$

Rate constants:

k_{on} = k1
k_{off} = km1

Differential equations:

```
dadt=-k1*a*L+km1*b;
dbdt=k1*a*L-km1*b;
dYdt=[dadt;dbdt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*(km1)/(km1+k1*Linit);
E2=Etotal*(k1*Linit)/(km1+k1*Linit);
```

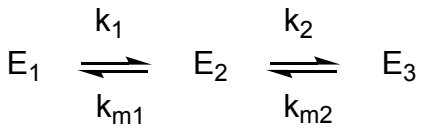
Binding affinity:

```
Kd=(km1)/(k1);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
extA=K(3);
extB=K(4);
```

3 state – Generic




```
dE1/dt = dadt
dE2/dt = dbdt
dE3/dt = dcdt
```

Differential equations:

```
dadt=-k1*a+km1*b;
dbdt=k1*a-km1*b-k2*b+km2*c;
dcdt=k2*b-km2*c;
dYdt=[dadt;dbdt;dcdt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*(km1*km2)/(km1*km2+k1*km2+k1*k2);
E2=Etotal*(k1*km2)/(km1*km2+k1*km2+k1*k2);
E3=Etotal*(k1*k2)/(km1*km2+k1*km2+k1*k2);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
extA=K(5)
extB=K(6);
extC=K(7);
```

3 state – Induced Fit

Binding step:

k_{on} = k1
k_{off} = km1

Differential equations:

```
dadt=-k1*L*a+km1*b;
dbdt=k1*L*a-km1*b-k2*b+km2*c;
dcdt=k2*b-km2*c;
dYdt=[dadt;dbdt;dcdt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*(km1*km2)/(km1*km2+k1*Linit*km2+k1*Linit*k2);
E2=Etotal*(k1*Linit*km2)/(km1*km2+k1*Linit*km2+k1*Linit*k2);
E3=Etotal*(k1*Linit*k2)/(km1*km2+k1*Linit*km2+k1*Linit*k2);
```

Binding affinity:

```
Kdapp=(km1*km2)/(k1*(k2+km2));
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
extA=K(5)=0;
extB=K(6);
extC=K(7);
```

3 state – Conformational Selection

Binding step:

k_{on} = k2
k_{off} = km2

Differential equations:

```
dadt=-k1*a+km1*b;
dbdt=k1*a-km1*b-k2*L*b+km2*c;
dcdt=k2*L*b-km2*c;
dYdt=[dadt;dbdt;dcdt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*(km1*km2)/(km1*km2+k1*km2+k1*k2*Linit);
E2=Etotal*(k1*km2)/(km1*km2+k1*km2+k1*k2*Linit);
E3=Etotal*(k1*k2*Linit)/(km1*km2+k1*km2+k1*k2*Linit);
```

Binding affinity:

```
Kdapp=(km2*(k1+km1))/(k1*k2);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
extA=K(5)=0;
extB=K(6)=0;
extC=K(7);
```

3 state – Sequential Binding

$k_{on} = k_1$ and k_2

Binding step:

$k_{on,1}$ = k1
$k_{off,1}$ = km1
$k_{on,2}$ = k2
$k_{off,2}$ = km2

Differential equations:

```
dadt=-k1*L*a+km1*b;
dbdt=k1*L*a-km1*b-k2*L*b+km2*c;
dcdt=k2*L*b-km2*c;
dYdt=[dadt;dbdt;dcdt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*(km1*km2)/(km1*km2+k1*Linit*km2+k1*Linit*k2*Linit);
E2=Etotal*(k1*Linit*km2)/(km1*km2+k1*Linit*km2+k1*Linit*k2*Linit);
E3=Etotal*(k1*Linit*k2*Linit)/(km1*km2+k1*Linit*km2+k1*Linit*k2*Linit)
;
```

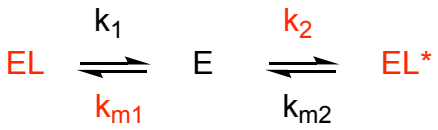
Binding affinity:

```
Kdapp=(-k1*km2+sqrt(k1^2*km2^2+4*k1*k2*km1*km2))/(2*k1*k2);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
extA=K(5)=0;
extB=K(6);
extC=K(7);
```

3 state – Substrate Reorientation



$k_{on} = k_{m1}$ and k_2

Rate constants:

$k_{off,1}$ = k1
$k_{on,1}$ = km1
$k_{on,2}$ = k2
$k_{off,2}$ = km2

Differential equations:

```
dadt=-k1*a+km1*L*b;
dbdt=k1*a-km1*L*b-k2*L*b+km2*c;
dcdt=k2*L*b-km2*c;
dYdt=[dadt;dbdt;dcdt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*(km1*Linit*km2)/(km1*Linit*km2+k1*km2+k1*k2*Linit);
E2=Etotal*(k1*km2)/(km1*Linit*km2+k1*km2+k1*k2*Linit);
E3=Etotal*(k1*k2*Linit)/(km1*Linit*km2+k1*km2+k1*k2*Linit);
```

Binding affinity:

```
Kdapp=k1*km2/(k1*k2+km1*km2);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
extA=K(5);
extB=K(6)=0;
extC=K(7);
```

4 state – Persistent heterogeneity

$k_{on} = k_1$

$k_{on} = k_2$

Rate constants:

k_{on1} = k1
k_{off1} = km1
k_{on2} = k2
k_{off2} = km2

Differential equations:

```
dadt=-k1*a*L+km1*b;
dbdt=k1*a*L-km1*b;
dcdt=-k2*c*L+km2*d;
dddt=k2*c*L-km2*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=(E1+E2)*(km1)/(km1+k1*Linit);
E2=(E1+E2)*(k1*Linit)/(km1+k1*Linit);
E3=(E3+E4)*(km2)/(km2+k2*Linit);
E4=(E3+E4)*(k2*Linit)/(km2+k2*Linit);
```

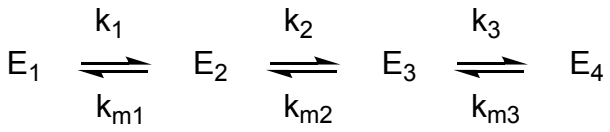
Binding affinity:

```
Kd1=(km1)/(k1);
Kd2=(km2)/(k2);
Kdapp=(-((E1+E2)*(Kd2-Kd1)+(E3+E4)*(Kd1-Kd2))+sqrt(((E1+E2)*(Kd2-
Kd1)+(E3+E4)*(Kd1-Kd2))^2-
4*Kd1*Kd2*(E1+E2+E3+E4)^2)/(2*(E1+E2+E3+E4));
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
extA=K(5);
extB=K(6);
extC=K(7);
extD=K(8);
```

4 state – Linear Generic




```
dE1/dt = dadt
dE2/dt = dbdt
dE3/dt = dcdt
dE4/dt = dddt
```

Differential equations:

```
dadt=-k1*a+km1*b;
dbdt=k1*a-km1*b-k2*b+km2*c;
dcdt=k2*b-km2*c-k3*c+km3*d;
dddt=k3*c-km3*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

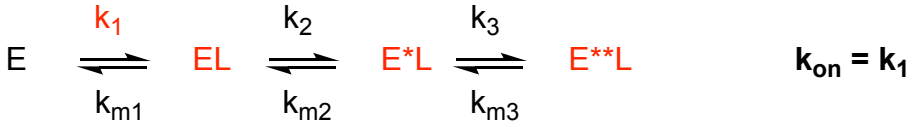
Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*km1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*k2*k3);
E2=Etotal*k1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*k2*k3);
E3=Etotal*k1*k2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*k2*k3);
E4=Etotal*k1*k2*k3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*k2*k3);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
extA=K(7);
extB=K(8);
extC=K(9);
extD=K(10);
```

4 state – Linear Induced Fit



Binding step:

k_{on} = k1
k_{off} = km1

Differential equations:

```
dadt=-k1*L*a+km1*b;
dbdt=k1*L*a-km1*b-k2*b+km2*c;
dcdt=k2*b-km2*c-k3*c+km3*d;
dddt=k3*c-km3*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*km1*km2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*km3+k1
*Linit*k2*k3);
E2=Etotal*k1*Linit*km2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*k
m3+k1*Linit*k2*k3);
E3=Etotal*k1*Linit*k2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*km
3+k1*Linit*k2*k3);
E4=Etotal*k1*Linit*k2*k3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*km3
+k1*Linit*k2*k3);
```

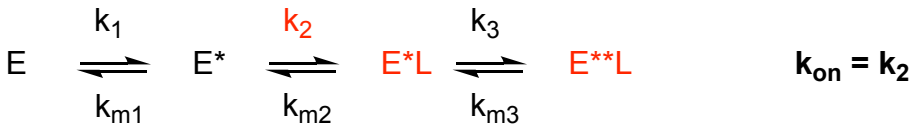
Binding affinity:

```
Kdapp=km1*km2*km3/(k1*(km2*km3+k2*km3+k2*k3));
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
extA=K(7)=0;
extB=K(8);
extC=K(9);
extD=K(10);
```

4 state – Linear Mixed Conformational Selection/Induced Fit



Binding step:

```
kon = k2
koff = km2
```

Differential equations:

```
dadt=-k1*a+km1*b;
dbdt=k1*a-km1*b-k2*L*b+km2*c;
dcdt=k2*L*b-km2*c-k3*c+km3*d;
dddt=k3*c-km3*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*km1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*k2*Li
nit*k3);
E2=Etotal*k1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*k2*Lin
it*k3);
E3=Etotal*k1*k2*Linit*km3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*k
2*Linit*k3);
E4=Etotal*k1*k2*Linit*k3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*k2
*Linit*k3);
```

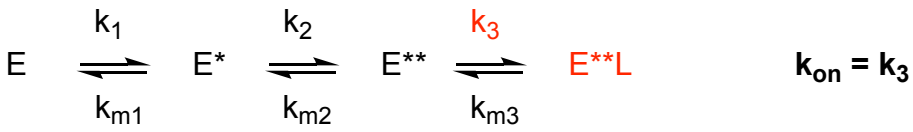
Binding affinity:

```
Kdapp=km2*km3*(km1+k1)/(k1*k2*(km3+k3));
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
extA=K(7)=0;
extB=K(8)=0;
extC=K(9);
extD=K(10);
```

4 state – Linear Conformational Selection



Binding step:

k_{on} = k3
k_{off} = km3

Differential equations:

```
dadt=-k1*a+km1*b;
dbdt=k1*a-km1*b-k2*b+km2*c;
dcdt=k2*b-km2*c-k3*L*c+km3*d;
dddt=k3*L*c-km3*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*km1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*k2*k3*Linit
);
E2=Etotal*k1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*k2*k3*Linit)
;
E3=Etotal*k1*k2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*k2*k3*Linit);
E4=Etotal*k1*k2*k3*Linit/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*k2*k3*Li
nit);
```

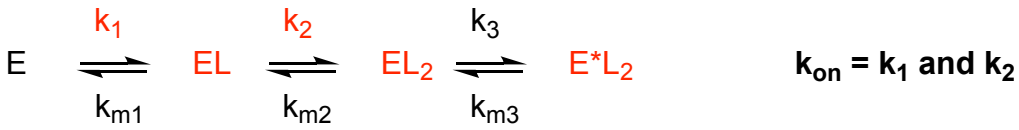
Binding Affinity:

```
Kdapp=km3*(km1*km2+k1*km2+k1*k2)/(k1*k2*k3);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
extA=K(7)=0;
extB=K(8)=0;
extC=K(9)=0;
extD=K(10);
```

4 state – Linear Induced Fit with 2 Binding Sites (A)



Binding step:

$k_{on,1}$ = k1
$k_{off,1}$ = km1
$k_{on,2}$ = k2
$k_{off,2}$ = km2

Differential equations:

```
dadt=-k1*L*a+km1*b;
dbdt=k1*L*a-km1*b-k2*L*b+km2*c;
dcdt=k2*L*b-km2*c-k3*c+km3*d;
dddt=k3*c-km3*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*km1*km2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*Linit*
km3+k1*Linit*k2*Linit*k3);
E2=Etotal*k1*Linit*km2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*L
init*km3+k1*Linit*k2*Linit*k3);
E3=Etotal*k1*Linit*k2*Linit*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit
*k2*Linit*km3+k1*Linit*k2*Linit*k3);
E4=Etotal*k1*Linit*k2*Linit*k3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*
k2*Linit*km3+k1*Linit*k2*Linit*k3);
```

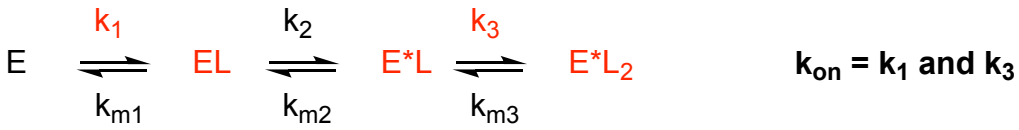
Binding Affinity:

```
Kdapp=(sqrt(k1*km2*km3*(k1*km2*km3+4*k2*k3*km1+4*k2*km1*km3))-
k1*km2*km3)/(2*k1*k2*(k3+km3))
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
extA=K(7)=0;
extB=K(8);
extC=K(9);
extD=K(10);
```

4 state – Linear Induced Fit with 2 Binding Sites (B)



Binding step:

$k_{on,1}$ = k1
$k_{off,1}$ = km1
$k_{on,2}$ = k3
$k_{off,2}$ = km3

Differential equations:

```
dadt=-k1*L*a+km1*b;
dbdt=k1*L*a-km1*b-k2*b+km2*c;
dcdt=k2*b-km2*c-k3*L*c+km3*d;
dddt=k3*L*c-km3*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*km1*km2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*km3+k1
*Linit*k2*k3*Linit);
E2=Etotal*k1*Linit*km2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*k
m3+k1*Linit*k2*k3*Linit);
E3=Etotal*k1*Linit*k2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*km
3+k1*Linit*k2*k3*Linit);
E4=Etotal*k1*Linit*k2*k3*Linit/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*
k2*km3+k1*Linit*k2*k3*Linit);
```

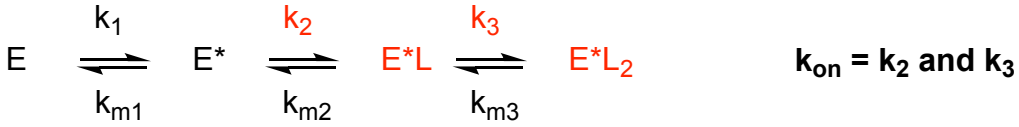
Binding Affinity:

```
Kdapp=(sqrt((k1*(k2+km2)^2*km3+4*km2*k2*k3*km1)*km3*k1)+(-k2-
km2)*km3*k1)/(2*k1*k2*k3)
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
extA=K(7)=0;
extB=K(8);
extC=K(9);
extD=K(10);
```

4 state – Linear Conformational Selection with 2 Binding Sites



$k_{on} = k_2$ and k_3

Binding step:

$k_{on,1}$ = k2*L
$k_{off,1}$ = km2
$k_{on,2}$ = k3*L
$k_{off,2}$ = km3

Differential equations:

```
dadt=-k1*a+km1*b;
dbdt=k1*a-km1*b-k2*L*b+km2*c;
dcdt=k2*L*b-km2*c-k3*L*c+km3*d;
dddt=k3*L*c-km3*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*km1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*k2*Li
nit*k3*Linit);
E2=Etotal*k1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*k2*Lin
it*k3*Linit);
E3=Etotal*k1*k2*Linit*km3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*k
2*Linit*k3*Linit);
E4=Etotal*k1*k2*Linit*k3*Linit/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3
+k1*k2*Linit*k3*Linit);
```

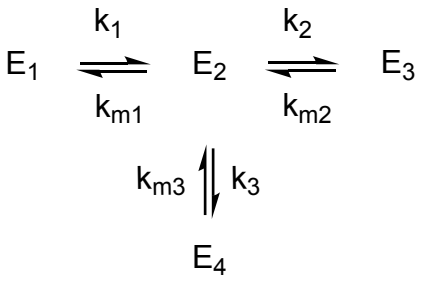
Binding Affinity:

```
Kdapp=(-
k1*k2*km3+sqrt((k1*k2*km3+4*km2*k3*(k1+km1))*k2*km3*k1))/(2*k1*k2*k3);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
extA=K(7)=0;
extB=K(8)=0;
extC=K(9);
extD=K(10);
```

4 state – Branched Generic




```
dE1/dt = dadt
dE2/dt = dbdt
dE3/dt = dcdt
dE4/dt = dddt
```

Differential equations:

```
dadt=-k1*a+km1*b;
dbdt=k1*a-km1*b-k2*b+km2*c-k3*b+km3*d;
dcdt=k2*b-km2*c;
dddt=k3*b-km3*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

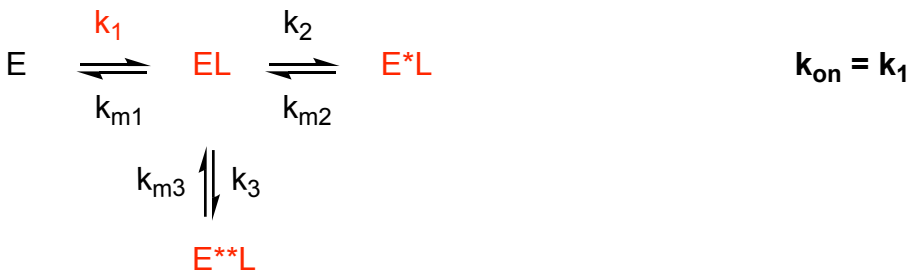
Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*km1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*km2*k3);
E2=Etotal*k1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*km2*k3);
E3=Etotal*k1*k2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*km2*k3);
E4=Etotal*k1*km2*k3/(km1*km2*km3+k1*km2*km3+k1*k2*km3+k1*km2*k3);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
extA=K(7);
extB=K(8);
extC=K(9);
extD=K(10);
```

4 state – Branched Induced Fit



$k_{on} = k_1$

Binding step:

k_{on} = k1
k_{off} = km1

Differential equations:

```
dadt=-k1*L*a+km1*b;
dbdt=k1*L*a-km1*b-k2*b+km2*c-k3*b+km3*d;
dcdt=k2*b-km2*c;
dddt=k3*b-km3*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*km1*km2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*km3+k1
*Linit*km2*k3);
E2=Etotal*k1*Linit*km2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*k
m3+k1*Linit*km2*k3);
E3=Etotal*k1*Linit*k2*km3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*km
3+k1*Linit*km2*k3);
E4=Etotal*k1*Linit*km2*k3/(km1*km2*km3+k1*Linit*km2*km3+k1*Linit*k2*km
3+k1*Linit*km2*k3);
```

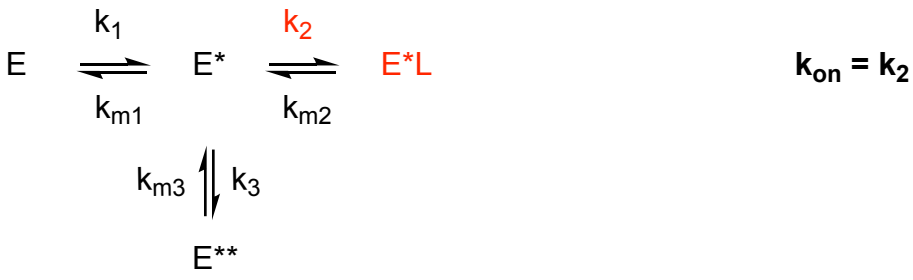
Binding affinity:

```
Kdapp=km1*km2*km3/(k1*(km2*km3+k2*km3+km2*k3));
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
extA=K(7)=0;
extB=K(8);
extC=K(9);
extD=K(10);
```

4 state – Branched Conformational Selection



$k_{on} = k_2$

Binding step:

k_{on} = k2
k_{off} = km2

Differential equations:

```
dadt=-k1*a+km1*b;
dbdt=k1*a-km1*b-k2*L*b+km2*c-k3*b+km3*d;
dcdt=k2*L*b-km2*c;
dddt=k3*b-km3*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*km1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*km2*k
3);
E2=Etotal*k1*km2*km3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*km2*k3
);
E3=Etotal*k1*k2*Linit*km3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*k
m2*k3);
E4=Etotal*k1*km2*k3/(km1*km2*km3+k1*km2*km3+k1*k2*Linit*km3+k1*km2*k3)
;
```

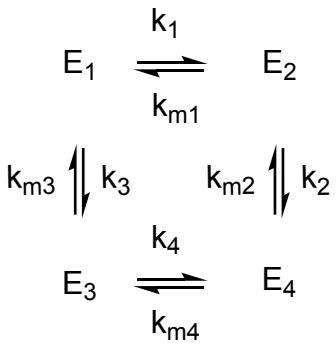
Binding affinity:

```
Kdapp=km2*(km1*km3+k1*km3+k1*k3)/(k1*k2*km3);
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
extA=K(7)=0;
extB=K(8)=0;
extC=K(9);
extD=K(10)=0;
```

4 state – Closed Generic




```
dE1/dt = dadt
dE2/dt = dbdt
dE3/dt = dcdt
dE4/dt = dddt
```

Differential equations:

```
dadt=-k1*a+km1*b-k3*a+km3*c;
dbdt=k1*a-km1*b-k2*b+km2*d;
dcdt=k3*a-km3*c-k4*c+km4*d;
dddt=k2*b-km2*d+k4*c-km4*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

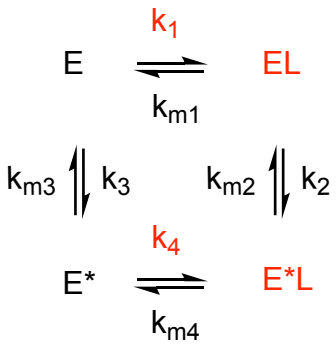
```
E1=Etotal*(k4*km1*km2+km3*((km2+km4)*km1+km4*k2))/(k4*k1*(km2+k2)+(k4*
km1*(k3+km2)+k1*(km2+km4+k2)*km3+k3*k4*km2+k2*(k1*km4+k3*k4))+(k3+km3)
*((km2+km4)*km1+km4*k2));
E2=Etotal*(k1*k4*km2+(k1*km3+k3*k4)*km2+k1*km3*km4)/(k4*k1*(km2+k2)+((
k1*km3+k4*(k3+km1))*km2+((k2+km4)*km3+km4*k2)*k1+k3*k4*(k2+km1))+((k2+
km1)*km4+km1*km2)*(k3+km3));
E3=Etotal*(k1*k2*km4+k3*((k2+km1)*km4+km1*km2))/(k4*k1*(km2+k2)+(k1*(k
2+km3)*km4+k4*(k2+km1+km2)*k3+k4*km1*km2+k1*km3*(km2+k2))+((k2+km1)*km
4+km1*km2)*(k3+km3));
E4=Etotal*(k1*k2*k4+(k1*km3+k3*k4)*k2+k3*k4*km1)/(k4*k1*(km2+k2)+((k3*
k4+k1*(km3+km4))*k2+((km1+km2)*k3+km1*km2)*k4+k1*km3*(km2+km4))+(k3+km
3)*((km2+km4)*km1+km4*k2));
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
k4=K(7);
km4=K(8);
extA=K(9);
extB=K(10);
```

```
extC=K(11);
extD=K(12);
```

4 state – Closed Conformational Selection/Induced Fit



k_{on} = k_1 and k_4

Binding step:

$k_{on,1}$ = k1
$k_{off,1}$ = km1
$k_{on,2}$ = k4
$k_{off,2}$ = km4

Differential equations:

```
dadt=-k1*L*a+km1*b-k3*a+km3*c;
dbdt=k1*L*a-km1*b-k2*b+km2*d;
dcdt=k3*a-km3*c-k4*L*c+km4*d;
dddt=k2*b-km2*d+k4*L*c-km4*d;
dYdt=[dadt;dbdt;dcdt;dddt];
```

Initial Protein Concentrations (equilibration concentration):

```
E1=Etotal*(Linit*k4*km1*km2+km3*((km2+km4)*km1+km4*k2))/(k4*k1*(km2+k2
)*Linit^2+(k4*km1*(k3+km2)+k1*(km2+km4+k2)*km3+k3*k4*km2+k2*(k1*km4+k3
*k4))*Linit+(k3+km3)*((km2+km4)*km1+km4*k2));
E2=Etotal*(Linit*k1*k4*km2+(k1*km3+k3*k4)*km2+k1*km3*km4)*Linit/(k4*k1
*(km2+k2)*Linit^2+((k1*km3+k4*(k3+km1))*km2+((k2+km4)*km3+km4*k2)*k1+k
3*k4*(k2+km1))*Linit+((k2+km1)*km4+km1*km2)*(k3+km3));
E3=Etotal*(Linit*k1*k2*km4+k3*((k2+km1)*km4+km1*km2))/(k4*k1*(km2+k2)*
Linit^2+(k1*(k2+km3)*km4+k4*(k2+km1+km2)*k3+k4*km1*km2+k1*km3*(km2+k2)
)*Linit+((k2+km1)*km4+km1*km2)*(k3+km3));
E4=Etotal*(Linit*k1*k2*k4+(k1*km3+k3*k4)*k2+k3*k4*km1)*Linit/(k4*k1*(k
m2+k2)*Linit^2+((k3*k4+k1*(km3+km4))*k2+((km1+km2)*k3+km1*km2)*k4+k1*k
m3*(km2+km4))*Linit+(k3+km3)*((km2+km4)*km1+km4*k2));
```

Binding affinity:

```
Kdapp=(-k4*(k2+km1+km2)*k3+k4*km1*km2+k1*(k2-km3)*km4-
k1*km3*(km2+k2)+sqrt((km3*km2+(km3-
km4)*k2+km4*km3)^2*k1^2+2*k4*(((k3+km1)*km3+2*k3*km1)*km2^2+(((2*k3+km
1+2*km4)*km3+(2*km1+km4)*k3+km1*km4)*k2+((km1+km4)*k3+km1*km4)*km3+2*k
3*km1*km4)*km2+(((k3+2*km4)*km3+k3*km4)*k2+k3*km3*km4)*(k2+km1))*k1+((
k3-km1)*km2+k3*(k2+km1))^2*k4^2))/(2*k4*k1*(km2+k2));
```

Parameters (rates and extinction coefficients):

```
k1=K(1);
km1=K(2);
k2=K(3);
km2=K(4);
k3=K(5);
km3=K(6);
k4=K(7);
km4=K(8);
extA=K(9)=0;
extB=K(10);
extC=K(11)=0;
extD=K(12);
```

References

1. P. R. Ortiz de Montellano, *Cytochrome P450* (Springer, 2015).
2. A. Saravanakumar, A. Sadighi, R. Ryu, and F. Akhlaghi, Clin Pharmacokinet **58**, 1281 (2019).
3. W. M. Atkins, J. Steroid Biochem. Mol. Biol. **151**, 3 (2015).
4. W. M. Atkins, Febs J. **287**, 1306 (2020).
5. M. Shou, Y. Lin, P. Lu, C. Tang, Q. Mei, D. Cui, W. Tang, J. S. Ngui, C. C. Lin, R. Singh, B. K. Wong, J. A. Yergey, J. H. Lin, P. G. Pearson, T. A. Baillie, A. D. Rodrigues, and T. H. Rushmore, Curr. Drug Metab. **2**, 17 (2001).
6. P. Hlavica and D. F. V. Lewis, Eur J Biochem **268**, 4817 (2001).
7. M. Shou, Meth. Enzymol. **357**, 261 (2002).
8. W. M. Atkins, Annu Rev Pharmacol Toxicol (2005).
9. S. J. Benkovic, G. G. Hammes, and S. Hammes-Schiffer, Biochemistry **47**, 3317 (2008).
10. F. P. Guengerich, Chem. Res. Toxicol. **14**, 611 (2001).
11. L. D. Gorsky, D. R. Koop, and M. J. Coon, J. Biol. Chem. **259**, 6812 (1984).
12. Y. V. Grinkova, I. G. Denisov, M. A. McLean, and S. G. Sligar, Biochemical and Biophysical Research Communications **430**, 1223 (2013).
13. J. B. Schenkman, S. G. Sligar, and D. L. Cinti, Pharmacology & Therapeutics **12**, 43 (1981).
14. K. P. Conner, C. M. Woods, and W. M. Atkins, Archives of Biochemistry and Biophysics **507**, 56 (2011).
15. R. Tsai, C. A. Yu, I. C. Gunsalus, J. Peisach, W. Blumberg, W. H. Orme-Johnson, and H. Beinert, Proc. Natl. Acad. Sci. U.S.a. **66**, 1157 (1970).
16. I. Schlichting, Science **287**, 1615 (2000).
17. C.-C. Peng, J. T. Pearson, D. A. Rock, C. A. Joswig-Jones, and J. P. Jones, Archives of Biochemistry and Biophysics **497**, 68 (2010).
18. U. P. Dahal, C. Joswig-Jones, and J. P. Jones, J. Med. Chem. **55**, 280 (2012).
19. C. C. Peng, W. Shi, J. D. Lutz, K. L. Kunze, J. O. Liu, W. L. Nelson, and N. Isoherranen, Drug Metabolism and Disposition **40**, 426 (2012).
20. J. Pearson, U. P. Dahal, D. Rock, C.-C. Peng, J. O. Schenk, C. Joswig-Jones, and J. P. Jones, Archives of Biochemistry and Biophysics **511**, 69 (2011).
21. M. N. Podgorski, J. S. Harbort, T. Coleman, J. E. Stok, J. A. Yorke, L.-L. Wong, J. B. Bruning, P. V. Bernhardt, J. J. De Voss, J. R. Harmer, and S. G. Bell, Biochemistry **59**, 1038 (2020).
22. C. W. Locuson, J. M. Hutzler, and T. S. Tracy, Drug Metab. Dispos. **35**, 614 (2007).
23. M. M. Lockart, C. A. Rodriguez, W. M. Atkins, and M. K. Bowman, Journal of Inorganic Biochemistry **183**, 157 (2018).
24. J. K. Yano, M. R. Wester, G. A. Schoch, K. J. Griffin, C. D. Stout, and E. F. Johnson, J. Biol. Chem. **279**, 38091 (2004).
25. S. C. Gay, A. G. Roberts, and J. R. Halpert, Future Med Chem **2**, 1451 (2010).
26. P. A. Williams, J. Cosme, D. M. Vinkovic, A. Ward, H. C. Angove, P. J. Day, C. Vonrhein, I. J. Tickle, and H. Jhoti, Science **305**, 683 (2004).
27. M. Ekroos and T. Sjögren, Proc. Natl. Acad. Sci. U.S.a. **103**, 13682 (2006).
28. I. F. Sevrioukova and T. L. Poulos, Proceedings of the National Academy of Sciences **107**, 18422 (2010).
29. I. F. Sevrioukova and T. L. Poulos, J. Biol. Chem. **287**, 3510 (2012).

30. I. F. Sevrioukova and T. L. Poulos, Archives of Biochemistry and Biophysics **520**, 108 (2012).
31. I. F. Sevrioukova and T. L. Poulos, J. Med. Chem. **56**, 3733 (2013).
32. I. F. Sevrioukova and T. L. Poulos, Biochemistry **52**, 4474 (2013).
33. G. Brändén, T. Sjögren, V. Schnecke, and Y. Xue, Drug Discovery Today **19**, 905 (2014).
34. I. F. Sevrioukova and T. L. Poulos, Biochemistry **54**, 4083 (2015).
35. P. Kaur, R. Chamberlin, T. L. Poulos, and I. F. Sevrioukova, J. Med. Chem. (2015).
36. I. F. Sevrioukova and T. L. Poulos, Proceedings of the National Academy of Sciences **114**, 486 (2017).
37. I. F. Sevrioukova, Biochemistry **56**, 3058 (2017).
38. Z. Guo, I. F. Sevrioukova, I. G. Denisov, X. Zhang, T.-L. Chiu, D. G. Thomas, E. A. Hanse, R. A. D. Cuellar, Y. V. Grinkova, V. W. Langenfeld, D. S. Swedien, J. D. Stamschror, J. Alvarez, F. Luna, A. Galván, Y. K. Bae, J. D. Wulfkuhle, R. I. Gallagher, E. F. Petricoin, B. Norris, C. M. Flory, R. J. Schumacher, M. G. O'Sullivan, Q. Cao, H. Chu, J. D. Lipscomb, W. M. Atkins, K. Gupta, A. Kelekar, I. A. Blair, J. H. Capdevila, J. R. Falck, S. G. Sligar, T. L. Poulos, G. I. Georg, E. Ambrose, and D. A. Potter, Cell Chem Biol **24**, 1259 (2017).
39. M.-H. Hsu, U. Savas, and E. F. Johnson, Mol. Pharmacol. **93**, 14 (2018).
40. E. R. Samuels and I. Sevrioukova, Mol. Pharm. **15**, 279 (2018).
41. E. R. Samuels and I. F. Sevrioukova, Bioorg. Med. Chem. **28**, 115349 (2020).
42. H. Park, S. Lee, and J. Suh, J. Am. Chem. Soc. **127**, 13634 (2005).
43. L. Benkaidali, F. André, G. Moroy, B. Tangour, F. Maurel, and M. Petitjean, Mol Inform **36**, (2017).
44. L. Benkaidali, F. André, G. Moroy, B. Tangour, F. Maurel, and M. Petitjean, Int J Mol Sci **20**, 987 (2019).
45. A. Nath, W. M. Atkins, and S. G. Sligar, Biochemistry **46**, 2059 (2007).
46. I. G. Denisov and S. G. Sligar, Chem. Rev. **117**, 4669 (2017).
47. Y. V. Grinkova, I. G. Denisov, and S. G. Sligar, Protein Eng. Des. Sel. **23**, 843 (2010).
48. F. Hagn, M. Etzkorn, T. Raschle, and G. Wagner, The Journal of Physical Chemistry (2013).
49. F. Hagn, M. L. Nasr, and G. Wagner, Nat Protoc **13**, 79 (2018).
50. M. J. Li, W. M. Atkins, and W. D. McClary, Curr Protoc Protein Sci **98**, e100 (2019).
51. E. P. Carpenter, K. Beis, A. D. Cameron, and S. Iwata, Curr. Opin. Struct. Biol. **18**, 581 (2008).
52. I. Moraes, G. Evans, J. Sanchez-Weatherby, S. Newstead, and P. D. S. Stewart, Biochim. Biophys. Acta **1838**, 78 (2014).
53. J. L. Parker and S. Newstead, Adv. Exp. Med. Biol. **922**, 61 (2016).
54. M. F. Jeng, S. W. Englander, G. A. Elöve, A. J. Wand, and H. Roder, Biochemistry **29**, 10433 (1990).
55. M. M. G. Krishna, L. Hoang, Y. Lin, and S. W. Englander, Methods **34**, 51 (2004).
56. J. R. Engen and T. E. Wales, Annu Rev Anal Chem (Palo Alto Calif) **8**, 127 (2015).
57. P. K. Glasoe and F. A. Long, J. Phys. Chem. **64**, 188 (1960).
58. S. W. Englander, J Am Soc Mass Spectrom **17**, 1481 (2006).
59. S. R. Marcsisin and J. R. Engen, Anal Bioanal Chem **397**, 967 (2010).
60. P. R. Wilderman, M. B. Shah, T. Liu, S. Li, S. Hsu, A. G. Roberts, D. R. Goodlett, Q. Zhang, V. L. Woods, C. D. Stout, and J. R. Halpert, J. Biol. Chem. **285**, 38602 (2010).

61. N. A. Treuheit, M. Redhair, H. Kwon, W. D. McClary, M. Guttman, J. P. Sumida, and W. M. Atkins, Biochemistry (2016).
62. J. Ducharme, C. J. Thibodeaux, and K. Auclair, bioRxiv 2020.06.09.142851 (2020).
63. L. Paço, F. Zarate-Perez, A. F. Clouser, W. M. Atkins, and J. C. Hackett, Biochemistry (2020).
64. W.-L. Liao, N. G. Dodder, N. Mast, I. A. Pikuleva, and I. V. Turko, Biochemistry **48**, 4150 (2009).
65. K. W. Anderson, N. Mast, J. W. Hudgens, J. B. Lin, I. V. Turko, and I. A. Pikuleva, J. Biol. Chem. **291**, 11876 (2016).
66. L. N. Jeffreys, K. J. Pacholarz, L. O. Johannissen, H. M. Girvan, P. E. Barran, M. W. Voice, and A. W. Munro, J. Biol. Chem. **295**, 7595 (2020).
67. D. R. Davydov, H. Fernando, B. J. Baas, S. G. Sligar, and J. R. Halpert, Biochemistry **44**, 13902 (2005).
68. E. M. Isin and F. P. Guengerich, J. Biol. Chem. **281**, 9127 (2006).
69. J. T. Pearson, J. J. Hill, J. Swank, N. Isoherranen, K. L. Kunze, and W. M. Atkins, Biochemistry **45**, 6341 (2006).
70. E. M. Isin and F. P. Guengerich, J. Biol. Chem. **282**, 6863 (2007).
71. I. G. Denisov, Y. V. Grinkova, M. A. McLean, and S. G. Sligar, J. Biol. Chem. **282**, 26865 (2007).
72. F. P. Guengerich, C. J. Wilkey, and T. T. N. Phan, J. Biol. Chem. **294**, 10928 (2019).
73. A. D. Vogt and E. Di Cera, Biochemistry **52**, 5723 (2013).
74. W. D. McClary, J. P. Sumida, M. Scian, L. Paço, and W. M. Atkins, Biochemistry **55**, 6258 (2016).
75. J. L. Baylon, I. L. Lenov, S. G. Sligar, and E. Tajkhorshid, J. Am. Chem. Soc. **135**, 8542 (2013).
76. J. C. Hackett, J. Biol. Chem. **293**, 4037 (2018).
77. K. K. Khan, Y. Q. He, T. L. Domanski, and J. R. Halpert, Mol. Pharmacol. **61**, 495 (2002).
78. D. R. Davydov, J. A. O. Rumfeldt, E. V. Sineva, H. Fernando, N. Y. Davydova, and J. R. Halpert, J. Biol. Chem. **287**, 6797 (2012).
79. I. G. Denisov, Y. V. Grinkova, J. L. Baylon, E. Tajkhorshid, and S. G. Sligar, Biochemistry **54**, 2227 (2015).
80. I. G. Denisov, J. L. Baylon, Y. V. Grinkova, E. Tajkhorshid, and S. G. Sligar, Biochemistry **57**, 805 (2018).
81. M. Redhair, A. F. Clouser, and W. M. Atkins, Chem. Phys. Lipids **220**, 14 (2019).
82. A. G. Roberts, J. Yang, J. R. Halpert, S. D. Nelson, K. T. Thummel, and W. M. Atkins, Biochemistry **50**, 10804 (2011).
83. J. Yang, W. M. Atkins, N. Isoherranen, M. F. Paine, and K. E. Thummel, Clinical Pharmacology & Therapeutics 1 (2009).
84. E. Gillam, Z. Y. Guo, and F. P. Guengerich, Archives of Biochemistry and … **312**, 59 (1994).
85. T. L. Domanski, J. Liu, G. R. Harlow, and J. R. Halpert, Archives of Biochemistry and Biophysics **350**, 223 (1998).
86. C. M. Woods, C. Fernandez, K. L. Kunze, and W. M. Atkins, Biochemistry **50**, 10041 (2011).
87. T. Matsubara, M. Koike, A. Touchi, Y. Tochino, and K. Sugeno, Anal. Biochem. **75**, 596 (1976).

88. A. A. Rutten, H. E. Falke, J. F. Catsburg, R. Topp, B. J. Blaauboer, I. van Holsteijn, L. Doorn, and F. X. van Leeuwen, Arch. Toxicol. **61**, 27 (1987).
89. T. Omura and R. Sato, J. Biol. Chem. **239**, 2370 (1964).
90. B. J. Baas, I. G. Denisov, and S. G. Sligar, Archives of Biochemistry and Biophysics **430**, 218 (2004).
91. A. L. Shen, T. D. Porter, T. E. Wilson, and C. B. Kasper, J. Biol. Chem. **264**, 7584 (1989).
92. J. S. French and M. J. Coon, Archives of Biochemistry and Biophysics **195**, 565 (1979).
93. A. Nath, Y. V. Grinkova, S. G. Sligar, and W. M. Atkins, J. Biol. Chem. **282**, 28309 (2007).
94. S. T. Lefurgy and T. S. Leyh, Anal. Biochem. **421**, 433 (2012).
95. P. M. Shaw, N. A. Hosea, D. V. Thompson, J. M. Lenius, and F. P. Guengerich, Archives of Biochemistry and Biophysics **348**, 107 (1997).
96. C. M. Hebling, C. R. Morgan, D. W. Stafford, J. W. Jorgenson, K. D. Rand, and J. R. Engen, Anal. Chem. **82**, 5415 (2010).
97. S. Adhikary, D. J. Deredge, A. Nagarajan, L. R. Forrest, P. L. Wintrode, and S. K. Singh, Proceedings of the National Academy of Sciences **114**, E1786 (2017).
98. M. J. Li, M. Guttman, and W. M. Atkins, J. Biol. Chem. **293**, 6297 (2018).
99. J. Fang, K. D. Rand, P. J. Beuning, and J. R. Engen, Int J Mass Spectrom **302**, 19 (2011).
100. D. D. Weis, J. R. Engen, and I. J. Kass, J Am Soc Mass Spectrom **17**, 1700 (2006).
101. T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker, Nucleic Acids Res. **35**, W522 (2007).
102. D. C. Bas, D. M. Rogers, and J. H. Jensen, Proteins **73**, 765 (2008).
103. G. Kieseritzky and E. W. Knapp, Journal of Computational Chemistry **29**, 2575 (2008).
104. M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen, J Chem Theory Comput **7**, 525 (2011).
105. J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, and R. W. Pastor, J. Phys. Chem. B **114**, 7830 (2010).
106. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, Journal of Computational Chemistry **26**, 1781 (2005).
107. A. D. MacKerell, M. Feig, and C. L. Brooks, Journal of Computational Chemistry **25**, 1400 (2004).
108. K. Vanommeslaeghe and A. D. MacKerell, J. Chem. Inf. Model. **52**, 3144 (2012).
109. S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, J Chem Phys **103**, 4613 (1995).
110. Y. Wang, C. B. Harrison, K. Schulten, and J. A. McCammon, Comput Sci Discov **4**, (2011).
111. Y. T. Pang, Y. Miao, Y. Wang, and J. A. McCammon, J Chem Theory Comput **13**, 9 (2017).
112. N. A. Hosea, G. P. Miller, and F. P. Guengerich, Biochemistry **39**, 5929 (2000).
113. H. Qian, Biophys. J. **95**, 10 (2008).
114. D. R. Davydov, B. J. Baas, S. G. Sligar, and J. R. Halpert, Biochemistry **46**, 7852 (2007).
115. T. Kronbach, D. Mathys, M. Umeno, F. J. Gonzalez, and U. A. Meyer, Mol. Pharmacol. **36**, 89 (1989).
116. J. A. Williams, B. J. Ring, V. E. Cantrell, D. R. Jones, J. Eckstein, K. Ruterbories, M. A. Hamman, S. D. Hall, and S. A. Wrighton, Drug Metab. Dispos. **30**, 883 (2002).
117. Y. Miao, V. A. Feher, and J. A. McCammon, J Chem Theory Comput **11**, 3584 (2015).
118. B. F. Shaw, H. Arthanari, M. Narovlyansky, A. Durazo, D. P. Frueh, M. P. Pollastri, A. Lee, B. Bilgiçer, S. P. Gygi, G. Wagner, and G. M. Whitesides, J. Am. Chem. Soc. **132**, 17411 (2010).

119. Y. Hamuro, J Am Soc Mass Spectrom **28**, 486 (2017).
120. D. Goswami, S. Devarakonda, M. J. Chalmers, B. D. Pascal, B. M. Spiegelman, and P. R. Griffin, J Am Soc Mass Spectrom **24**, 1584 (2013).
121. J. B. Houston and K. E. Kenworthy, Drug Metab. Dispos. **28**, 246 (2000).
122. K. E. Kenworthy, S. E. Clarke, J. Andrews, and J. B. Houston, Drug Metab. Dispos. **29**, 1644 (2001).
123. M. Eigen and L. DeMaeyer, *Relaxation Methods in "Technique of Organic Chemistry"* (1963).
124. G. G. Hammes and P. Schimmel, The Journal of Physical Chemistry **70**, (1966).
125. C. Bernasconi, *Relaxation Kinetics* (Academic Press: New York, 1976).
126. A. Fersht, *Enzyme Structure and Mechanism* (W.H. Freeman, 1985).
127. A. D. Vogt and E. Di Cera, Biochemistry **51**, 5894 (2012).
128. S. Gianni, J. Dogan, and P. Jemth, Biophys. Chem. **189**, 33 (2014).
129. F.-J. Meyer-Almes, Eur. Biophys. J. **45**, 245 (2016).
130. S. E. Faraj, R. C. Rossi, and M. R. Montes, Journal of Biological Education (2019).
131. P. Chakraborty and E. Di Cera, Biochemistry **56**, 2853 (2017).
132. G. W. Castellan, Berichte Der Bunsengesellschaft Für Physikalische Chemie **67**, 898 (1963).
133. K. Kustin, D. Shear, and D. Kleitman, J. Theor. Biol. **9**, 186 (1965).
134. D. Thusius, Biochimie **55**, 277 (1973).
135. D. Thusius, Biophys. Chem. **7**, 87 (1977).
136. W. Bujalowski, Chem. Rev. **106**, 556 (2006).
137. F. R. Gantmakher, *The Theory of Matrices* (American Mathematical Soc., 1959).
138. W. Stroberg and S. Schnell, Math Biosci **287**, 3 (2017).
139. A. D. Vogt, N. Pozzi, Z. Chen, and E. Di Cera, Biophys. Chem. **186**, 13 (2014).
140. A. Molla, M. Korneyeva, Q. Gao, S. Vasavanonda, P. J. Schipper, H. M. Mo, M. Markowitz, T. Chernyavskiy, P. Niu, N. Lyons, A. Hsu, G. R. Granneman, D. D. Ho, C. A. Boucher, J. M. Leonard, D. W. Norbeck, and D. J. Kempf, Nature Medicine **2**, 760 (1996).
141. D. J. Kempf, K. C. Marsh, G. Kumar, A. D. Rodrigues, J. F. Denissen, E. McDonald, M. J. Kukulka, A. Hsu, G. R. Granneman, P. A. Baroldi, E. Sun, D. Pizzuti, J. J. Plattner, D. W. Norbeck, and J. M. Leonard, Antimicrobial Agents and Chemotherapy **41**, 654 (1997).
142. R. K. Zeldin and R. A. Petruschke, J. Antimicrob. Chemother. **53**, 4 (2004).
143. K. P. Conner, P. Vennam, C. M. Woods, M. D. Krzyaniak, M. K. Bowman, and W. M. Atkins, Biochemistry **51**, 6441 (2012).
144. H. E. Seward, A. Roujeinikova, K. J. McLean, A. W. Munro, and D. Leys, J. Biol. Chem. **281**, 39437 (2006).
145. B. M. Rock, S. M. Hengel, D. A. Rock, L. C. Wienkers, and K. L. Kunze, Mol. Pharmacol. (2014).
146. W. Fitch, T. Tran, M. Young, L. Liu, and Y. Chen, Dml **3**, 191 (2009).
147. M. Oszajca, A. Franke, M. Brindell, G. Stochel, and R. van Eldik, Inorg Chem **50**, 3413 (2011).
148. H. Akaike, *Information Theory as an Extension of the Maximum Likelihood Principle in Petrov BN, Editor; & Csaki F., Editor.(Eds.) Second International Symposium on …* (1973).
149. G. Schwarz, *Estimating Thedimension of a Model. Annals of Statistics, 6 (2), 461–464* (1978).

150. E. R. Henry and J. Hofrichter, Meth. Enzymol. **210**, 129 (1992).
151. R. W. Hendler and R. I. Shrager, J. Biochem. Biophys. Methods **28**, 1 (1994).
152. E. R. Henry, Biophys. J. **72**, 652 (1997).
153. J. Rittle, J. M. Younker, and M. T. Green, Inorg Chem **49**, 3610 (2010).
154. R. Fedorov, D. K. Ghosh, and I. Schlichting, Archives of Biochemistry and Biophysics **409**, 25 (2003).
155. M. C. Simianu and J. R. Kincaid, J. Am. Chem. Soc. **117**, 4628 (1995).