ALLOSTERIC COMMUNICATION NETWORKS AND ENZYMATIC REGULATION AS MEDIATED BY CONFORMATIONAL DYNAMICS IN THE CYCLOPHILIN FAMILY OF PEPTIDYL-PROLYL ISOMERASES

by

MICHAEL JOSEPH HOLLIDAY

B.A., Dartmouth College, 2005

M.S., Thayer School of Engineering, Dartmouth College, 2007

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Molecular Biology Program

2015
Proteins exhibit conformational flexibility over a wide range of timescales, ranging from picoseconds to hours. These motions are not simply random fluctuations, but are intimately involved in a range of critical protein functions including substrate binding and release, enzymatic catalysis, and allosteric communication. Critical to our increasing understanding of these dynamics and their relation to function have been nuclear magnetic resonance (NMR) and molecular dynamics (MD) based studies, which are able to inform on motions over a wide timescale at atomic resolution. This manuscript outlines a number of related studies in which NMR and MD, along with additional biophysical approaches, have been utilized to study conformational dynamics in the cyclophilin family of peptidyl-prolyl isomerases. We have determined the NMR solution structure of the first characterized thermophilic cyclophilin (GeoCyp) and found that it catalyzes isomerization at a remarkably high rate at low temperatures and utilizes a conserved catalytic mechanism as compared to human Cyclophilin A (CypA). We additionally identified a novel mechanism by which GeoCyp increases substrate affinity via a dynamic ‘clamp’ over the substrate binding pocket. Using lineshape analysis, we have determined the rate constants defining the catalytic cycles of multiple human cyclophilins, including CypA, Cyclophilin B, and Cyclophilin C, as well as GeoCyp. This study revealed a conserved isomerization rate among the human cyclophilins with variable substrate affinities mediated by differing on- and off-rates. Structural and dynamic comparison suggests that variable substrate affinity is primarily mediated by dynamic differences among the human cyclophilins. In CypA, we identified, via
a combination of NMR dynamic experiments and analysis of MD ensembles, distinct networks of allostERIC communication mediated among non-coherently dynamic residues. Through generation of active site distal mutations, we have altered the conformational sampling in the CypA active site independent of changes to the ground state of the substrate interface. These dynamic changes alter CypA binding affinity and catalytic turnover towards multiple substrates, likely via shifting the inherent conformational selection mechanism of substrate engagement, indicating a direct link between conformational fluctuations in CypA and its enzymatic function.

The form and content of this abstract are approved. I recommend its publication.

Approved: Elan Z. Eisenmesser
I dedicate this work to my parents, Mark and Mary Holliday.
ACKNOWLEDGEMENTS

I would like to thank my parents, Mark and Mary Holliday, for their constant support throughout my life and for instilling the love of learning in me from as early as I can remember. I would like to thank my brother Andy and sister Caroline. I would like to thank Jess, my best friend and lifelong partner, for her endless support and love. I thank Lily, for helping me to keep it all in perspective.

I would like to thank my advisor, Elan Eisenmesser for his guidance over the years, and particularly for his constant excitement and enthusiasm for scientific exploration. I would like to thank previous mentors, Simon Shepherd, for helping to shape the way that I think about scientific questions, and Carrie McCurdy, for giving me the chance to dive full-on into biological research. I would like to thank current and former members of the Eisenmesser Lab for creating an enjoyable and engaging lab environment, and particularly, Aga Kendrick, Jasmina Redzic, and Thomas Chi for great friendships throughout the years. I would like to thank my committee for their suggestions and guidance. I would like to thank the students, faculty, and administrators of the Department of Biochemistry and Molecular Genetics, The Molecular Biology Program, and the Program in Structural Biology and Biochemistry, for providing such a valuable and gratifying research and training environment.
A gatekeeper residue that regulates substrate binding .................. 39
An active site electric field is conserved between CypA and GeoCyp .......... 41
GeoCyp and CypA exhibit comparable temperature dependent activities .... 42
Dynamics are similar over multiple timescales between CypA and GeoCyp,
with variability in temperature dependence and magnitude .................. 43
GeoCyp weakly self associates in the millisecond timescale .................. 45
Summary ........................................................................................................ 45

III. DETERMINATION OF THE FULL CATALYTIC CYCLE AMONG MULTIPLE CYCLOPHILIN FAMILY MEMBERS .................................................. 63
Introduction ................................................................................................... 63
Results .......................................................................................................... 64
Functional characterization of multiple members of the cyclophilin family .... 64
Structural and dynamic comparison of multiple human cyclophilins ........... 66
Elucidation of the complete catalytic cycle for multiple cyclophilins utilizing
lineshape analysis ....................................................................................... 68
Summary ........................................................................................................ 74

IV. LIMITATIONS ON THE APPLICATION OF CPMG-RD IN REVERSIBLE CATALYTIC SYSTEMS ................................................................. 82
Introduction ................................................................................................... 82
Results .......................................................................................................... 83
Modeling CPMG-RD in reversible systems .................................................. 83
Experimental measurement of non-catalyzed enzyme-perspective
cis↔trans interconversion ........................................................................... 87
Segmental dynamics in CypA in the bound form ........................................ 88
Summary ........................................................................................................ 88

V. MAPPING NETWORKS OF NON-COHERENT DYNAMIC ALLOSTERIC COUPLING IN CYCLOPHILIN A .................................................. 97
Introduction ................................................................................................... 97
Results .......................................................................................................... 99
LIST OF TABLES

1. Structural statistics for the GeoCyp RASREC Rosetta structures ............................. 49
2. GSFGPDLRAGD model peptide chemical shift assignments .................................... 50
3. Binding affinities and catalytic efficiencies for CypA, GeoCyp, and mutants .......... 50
4. Apparent binding affinities and observed isomerization rates for multiple cyclophilins .............................................................................................................. 76
5. Microscopic rate constants determined for CypA using different values of \( R_{2\text{-Bound}} \) ..................................................................................................................... 76
6. Best-fit solutions of cyclophilin microscopic rate constants .................................... 76
7. Catalyzed and uncatalyzed exchange rates for multiple cyclophilins ..................... 90
8. Apparent dissociation constants and isomerization rates of cyclophilin mutants towards the FGP peptide ........................................................................ 114
9. Apparent dissociation constants as measured by ITC ............................................. 114
10. Apparent dissociation constants and isomerization rates of cyclophilin mutants towards the \( \text{HA}^{P2A} \) peptide ................................................................. 114
11. Best-fit solutions of microscopic rate constants for CypA and mutants .......... 115
LIST OF FIGURES

FIGURE

1. Peptidyl prolyl isomerization ................................................................. 29
2. Cyclophilin A structure bound to cyclosporine and model peptide .......... 29
3. NMR chemical exchange regimes .......................................................... 30
4. CPMG-RD refocusing ........................................................................... 31
5. CPMG-RD profiles ................................................................................. 31
6. ZZ-exchange example .......................................................................... 31
7. Classic models of allostery ................................................................. 32
8. The secondary structure elements of GeoCyp are similar to those of CypA 51
9. The structure of GeoCyp is similar to CypA, with some changes in substrate interface ................................................................. 52
10. GeoCyp is stable over a broad range of temperatures ......................... 53
11. CypA and GeoCyp binding interfaces ................................................. 54
12. Peptide bound MD ensembles of CypA and GeoCyp ......................... 55
13. Homologous mutations in the context of mutation to the catalytic arginine .... 56
14. Sequence alignment of all annotated cyclophilins from \textit{Bacillaceae} family members in the RefSeq database ................................................................. 57
15. GeoCyp exhibits a -Z electric field within the active site ..................... 58
16. Catalytic efficiency for CypA and GeoCyp from 0-45\degree C .................. 58
17. For GeoCyp, R_{ex} values, but not R_{20} or R_{1} values, exhibit a large concentration dependence ................................................................. 59
18. R_{1} relaxation rates for CypA and GeoCyp ........................................ 60
19. R_{20} relaxation rates for CypA and GeoCyp ........................................ 61
20. GeoCyp self-associates on the millisecond time scale .......................... 62
21. Functional characterization of multiple cyclophilins ......................... 77
22. Highly conserved structure and fast timescale dynamics among CypA, CypB, and CypC, and variable \mu s timescale exchange ................................. 78
23. Overlaid ribbon diagrams of CypA crystal structures ......................... 78
24. Testing self association in CypB and CypC ............................................................. 79
25. Divergent $\mu$-ms motions in CypA and CypC .......................................................... 80
26. Four state peptide-perspective minimal exchange model of cyclophilin catalysis ................................................................................................................... 80
27. Extraction and fitting lineshape data ........................................................................ 81
28. Minimal three state enzyme-perspective exchange model ........................................ 91
29. Validation of stochastic CPMG-RD model. ............................................................... 91
30. CPMG-RD can report on uncatalyzed exchange in reversible enzymatic systems ................................................................................................................... 92
31. $R_{ex}$ as a function of substrate concentration for uncatalyzed peptide interconversion ................................................................................................................... 92
32. Cyclophilin peptide saturation .................................................................................. 93
33. CPMG-RD simulations performed over a range of $k_{ex}$-catalyzed and $K_D$-app values .... 94
34. Comparing catalytic turnover of the model peptide and its thioamide variant by CypA ................................................................................................................... 95
35. Measured CPMG-RD data on $^2$H$^{15}$N CypA when it is saturated with catalyzable and uncatalyzable substrates ...................................................................................... 96
36. Identification of an active site distal residue coupled to active site dynamics ......... 116
37. Active site distal methyl CPMG data alone and with FGP peptide ......................... 116
38. Identification of putative allosteric pathways between the CypA active site and Val 29 in an MD ensemble .................................................................................... 117
39. Conformational transitions reported on by Ser 99 and Phe 113 .............................. 119
40. A second putative allosteric pathway between Val 29 and the active site .......... 120
41. Val 29 mutants are well folded and do not alter the global structure of CypA ....... 121
42. Experimental identification of allosteric communication pathways in CypA ........ 122
43. Additional allosteric coupling stemming from Val 29 .............................................. 123
44. Identification of a second active site distal location and characterization of its communication pathway to the active site ...................................................................... 124
45. Dynamic response to Val 6 mutants and double mutant ........................................ 125
46. Characterization of the HA and HA$^{P2A}$ peptide ............................................... 126
47. Peptide on-rates mirror CPMG-RD monitored shifts in dynamics............................. 127

48. $^{15}$N-HSQC spectra of AAG, AlkA, DMG2, and EndoV ............................................ 141
CHAPTER I
LITERATURE REVIEW AND THE CURRENT STATE OF THE FIELD

The cyclophilins

Cyclophilin biology

Cyclophilins are a near ubiquitously expressed protein family, found in all domains of life, and only known to be absent in a small number of obligately symbiotic bacteria and archaea.\(^1,2\) Among organisms with cyclophilins, they are highly abundant (0.1-0.3% of total protein in human tissues) and are often present as multiple isoforms (e.g., 17 isoforms exist in humans, 8 in \textit{S. cerevisiae}, and 2 in \textit{E. coli}).\(^3-5\) Most, although not all, cyclophilins catalyze the \textit{cis}$$\leftrightarrow$$\textit{trans} isomerization of peptidyl-prolyl bonds (see Figure 1), and are therefore also referred to as peptidyl-prolyl isomerases (PPIases). Cyclophilins are one of three major families of PPIases, along with the structurally unrelated FK506 binding proteins and the parvulins.\(^6\)

The structure of cyclophilins is highly conserved and consists of a flattened eight-sheet anti-parallel $\beta$-barrel capped by two 2-3 turn $\alpha$-helices and a short 3-10 helix adjacent to the active site, which lies parallel to the $\beta$-sheet along one side of the barrel (Figure 2). Amino acid sequence is fairly well conserved within the cyclophilin family, with a representative 34-36% homology between the two \textit{E. coli} cyclophilins and human Cyclophilin A (CypA). Within the protein active site, certain residues are highly or absolutely conserved, including a catalytic arginine, which is absolutely conserved among known catalytically active cyclophilins.\(^7\) As some cyclophilins have evolved to fulfill niches that do not require isomerization activity, the catalytic arginine has not been absolutely conserved among these non-catalytic cyclophilins.\(^8\)
The prototypical and first characterized cyclophilin, CypA, encoded by the \textit{PPIA} gene, was initially identified independently by two groups for its peptidyl prolyl isomerase activity and as being the binding partner of the fungal cyclic undecapeptide and immunosuppressant, cyclosporine A (CsA).\textsuperscript{9,10} Soon thereafter, it was recognized that these two functions were mediated by the same protein.\textsuperscript{11} The immunosuppressant activity of CsA, however, has been shown to be independent of the PPIase activity of CypA; upon binding, the CypA-CsA interface binds to and blocks catalytic activity of the phosphatase Calcineurin, which dephosphorylates the nuclear factor for activation of T-cells (NF-AT), a critical step in T-cell activation.\textsuperscript{12,13} CsA has also been shown to bind to the majority of other human cyclophilins,\textsuperscript{5} although in mice the immunosuppressant activity of CsA has been shown to be specifically dependent on CypA.\textsuperscript{14}

Cyclophilins are involved in a wide range of biological functions. Perhaps the most well understood role of the proteins is as a chaperone/foldase. While peptide bonds consisting of X-nonPro amino acids, where X is any residue and nonPro is any non-proline residue, overwhelmingly adopt a \textit{trans} conformation (~99.9\% of bonds in both unstructured peptides and in protein structures), X-Pro peptide bonds adopt the \textit{cis} conformation with significant frequency (~10-30\% in unstructured peptides and ~6.5\% of bonds in protein structures).\textsuperscript{15} As a given protein structure will generally only contain a single isomer at each proline position and the un-catalyzed isomerization rate is on the order of 10s-100s of seconds (shown in Chapter 2 to be >>1 s for the model peptide utilized here), the \textit{cis}$\leftrightarrow$\textit{trans} interconversion of prolyl-peptide bonds represents the limiting step in the folding of many proteins, particularly when the correct isomer is required for multiple prolines.\textsuperscript{16,17} Cyclophilins, along with other peptidyl-prolyl isomerases, act to accelerate these interconversions by a factor of $\sim 10^5$, circumventing this bottleneck to protein folding.\textsuperscript{1,9,17,18} Cyclophilins have been shown to specifically promote proper folding of several human proteins, including transferrin,\textsuperscript{19} collagen,\textsuperscript{20} and the $\alpha$7 neuronal nicotinic and type 3
serotonin homo-oligomeric receptors. Additionally, the conserved hydrophobic active site allows cyclophilins to act as chaperones by shielding partially folded intermediates from misfolding and preventing misfolded proteins from aggregating.

More recently, multiple additional biological roles of CypA have been identified in humans, independent of its chaperone activity. CypA regulates signal transduction through the tyrosine kinase Itk, with a CypA mediated proline conformational switch inhibiting Itk activity and subsequent downstream signaling events. Additional signal transduction pathways are proposed to rely on conformational switching by CypA, including via interactions with Calcineurin and NF-κB. CypA has also been implicated in mediating intra-cellular trafficking of numerous protein targets, including apoptosis inducing factor, heterogeneous nuclear ribonucleoprotein A2, and zinc finger protein 1.

While CypA is predominantly cytoplasmically localized, several other human cyclophilins are localized to specific cellular compartments with specific, although generally less well understood, functions. Human cyclophilins Cyclophilin B (CypB) and Cyclophilin C (CypC) both contain target sequences that localize them to the endoplasmic reticulum (ER), with CypB suppressing ER stress-related apoptosis. Cyclophilin D (PPIF) interacts with and regulates calcium sensitivity of the mitochondrial permeability transition pore. Cyclophilin H (PPIH) and PPI-like protein 1 (PPIL1) have both been shown to stably interact the spliceosome. Cyclophilin E regulates gene expression through interactions with the mixed lineage leukemia histone methyltransferase. The mechanisms by which cyclophilins carry out most of these functions remain poorly characterized, including whether PPIase activity is required.

In addition to the specific cellular localizations of the human cyclophilins, CypA, CypB, and CypC are secreted extracellularly, acting as cytokines, with a role in driving many cancerous and inflammatory conditions, including those described below. CypA’s major known extracellular binding partner is extracellular matrix metalloprotease inducer
EMMPRIN, also known as CD147); blocking EMMPRIN leads to a potent reduction in cell migration in multiple contexts.\textsuperscript{39,40} The interaction of CypA with EMMPRIN, however, is very weak \textit{in vitro}, with a dissociation constant of more than 1 mM, suggesting that other binding partners may play a role in mediating CypA extracellular activity.\textsuperscript{41} CypB has likewise been found to utilize CD147 as a receptor for its extracellular activity.\textsuperscript{42}

Among the many biological roles identified for cyclophilins, their ability to provide tolerance to a range of stresses, including high salinity, oxidative stress, osmotic stress, infection, cold, and heat, has been identified across many species.\textsuperscript{43-46} Multiple studies have also exogenously expressed cyclophilins from one species into another, often conferring a range of tolerances in the recipient species.\textsuperscript{47-49} For example, a cyclophilin derived from a salt-tolerant strain of rice, when expressed in \textit{E. coli} or \textit{S. cerevisiae}, provides a survival advantage not only under high salt conditions, but also when exposed to osmotic, oxidative, or heat stress.\textsuperscript{47} The specific mechanism by which cyclophilins provide these various stress appears to be multifaceted. For example, some studies have pointed to targeted functions such as RNA binding activity or hyperpolarization of the mitochondrial membrane as mechanisms by which specific cyclophilins provide survival advantages.\textsuperscript{46} However, the breadth of protection provided has led most studies to hypothesize that, in general, tolerance is mediated through protein chaperone activity as cyclophilins act to maintain protein homeostasis and promote proper protein folding.\textsuperscript{6}

Given the evolutionary conservation of the family and the range of biological functions of cyclophilins, knockouts of cyclophilins are surprisingly well tolerated in many organisms. In \textit{S. cerevisiae}, simultaneous knockout of all 8 cyclophilins results in viable yeast, with only mild growth inhibition and slight temperature sensitivity.\textsuperscript{50} Overexpression of CypA from \textit{S. cerevisiae} is, however, able to rescue knockout of the one otherwise essential PPIase, the parvulin Ess1, indicating some overlapping roles in the function of these two functionally similar yet structurally distinct proteins.\textsuperscript{51} Knockout of the most abundant
cyclophilin, CypA, in mice results in somewhat decreased viability, but with largely healthy adults with normal lifespans. PPIA knockout mice do exhibit an allergic disease phenotype, consistent with hyperactivity of Itk without CypA downregulation, and are resistant to immunosuppression by CsA. In D. melanogaster, knockout of the human CypB homologue, ninaA, prevents proper generation of two isoforms of the rhodopsin G-protein-coupled receptor in the photoreceptor cells, presumably through a chaperone or foldase activity, suggesting that, while cyclophilins can target substrates promiscuously, certain cyclophilins may have highly specialized roles. Interestingly, mutations to CypB in humans have been shown not to impact rhodopsin, but to lead to Osteogenesis Imperfecta, a heterogeneous bone disease caused by impaired collagen biosynthesis, indicating that the specific protein targets of cyclophilins are not necessary evolutionarily conserved.

**The role of cyclophilins in pathology**

In addition to the known biological roles of cyclophilins catalogued above, cyclophilins have been implicated in a wide range of pathological conditions, including roles in promoting viral infectivity, multiple inflammatory diseases, and the progression of multiple cancers.

Human immunodeficiency virus (HIV) specifically incorporates host CypA into its viral coat via binding to a loop region within p24 capsid protein including the critical proline. This incorporation has been shown to enhance both viral replication and infectivity via numerous proposed mechanisms, including stabilization of the capsid and transport to the nucleus as well as in coordinating proper uncoating of the capsid. CypA has also been implicated in masking or otherwise protecting the virus from host recognition by pathogen associated molecular pattern receptors. Interestingly, cloning of the p24 binding sequence into the simian immunodeficiency virus (SIV), which does not naturally incorporate CypA, does lead to viral incorporation of CypA; the incorporation, however, inhibits viral replication in SIV, indicating that other structural components of the p24 capsid protein, specifically a
presence or absence of a type II tight turn, dictate whether CypA enhances or inhibits replication. While no relevant polymorphisms within the CypA coding region have been identified in relation to HIV susceptibility or other clinical outcomes, one polymorphism in the regulatory region has been associated with increased susceptibility to HIV, suggesting that levels of CypA production may influence HIV infectivity in vivo. Additionally, serum levels of CypA are increased in HIV infected patients.

Human hepatitis C virus (HCV) likewise utilizes cyclophilins to promote infectivity and replication. While the specific mechanisms are less well understood in HCV as compared to HIV, HCV is thought to incorporate CypA via binding to numerous sites within the NS5A protein which promotes formation of a double membrane vesicle that can protect the virus during replication from host detection and restriction. CypB is likewise thought to play a role in HCV replication via positive regulation of the viral RNA polymerase. Numerous other viruses have been shown to utilize various human host cyclophilins in their life cycles, including hepatitis virus B, influenza, and human papilloma virus. Given the role of cyclophilins in promoting the life cycle of these numerous viruses, significant efforts have been made to develop non-immunosuppressant homologues of CsA, including several currently in clinical trials.

CypA has been found to be up-regulated in a wide range of cancers, including lung, pancreatic, hepatocellular carcinoma, colorectal, and squamous cell carcinoma. In addition to potential application as a biomarker, in many of these cases, the up-regulation appears to be functionally relevant in the cancer development and/or metastasis. CypA effects cancer promotion through multiple complimentary inter- and extracellular mechanisms, including driving cell proliferation, blocking apoptosis, protection from reactive oxygen species, and promoting cell migration and invasion. Additionally, CypA has been shown to promote chemotherapeutic resistance via up-regulation of drug metabolism and transport proteins and has been found to be down-regulated by a
number of treatments. CypA levels are regulated by both p53 and HIF-1α, and CypA may engage in a positive feedback mechanism whereby it stabilizes p53. Other human cyclophilins are also up-regulated in multiple cancers, including CypB, CypD, and Cyp40, where they have been found to block apoptotic cell death, and CypC, where it has been found to increase cell migration and invasion.

Cyclophilins have also been implicated in the progression of a wide range of inflammatory diseases. These include cardiovascular disease, where CypA has been shown to promote vascular remodeling, vascular smooth muscle cell proliferation, inflammation, and the recruitment of immune cells. In rheumatoid arthritis, macrophages release high concentrations of extracellular CypA, resulting in an increased release of inflammatory cytokines and extracellular matrix reconstruction, potentially through interactions with EMMPRIN. Similarly, CypA has been associated with increased inflammation and immune cell recruitment in Alzheimer’s disease, sepsis, asthma, and peritonitis.

**Methods to monitor protein dynamics**

Proteins exist as a conformationally diverse collection of structures, interconverting, locally and globally, over a large range of timescales, from picoseconds to hours. The dominant theories of protein binding posit that dynamic mobility is crucial for protein-ligand interactions to occur and, in studies of enzymes, alterations to dynamics have been correlated with changes to both substrate interaction and catalytic turnover. Recognition of neither the existence nor the importance of dynamics in proteins is new; indeed, since ‘lock-and-key’ models began facing scrutiny over half a century ago, descriptions of protein interactions have evolved toward our current understanding, a broad combination of conformational selection and induced fit models. Only recently, however, have advances in experimental techniques and instrumentation, particularly in biomolecular nuclear magnetic resonance (NMR), permitted direct observation and
quantitation of these dynamic motions over a range of biologically important timescales at atomic resolution.

**Biomolecular NMR spectroscopy**

NMR spectroscopy relies on the phenomenon by which atomic nuclei with non-zero spin exhibit both a magnetic moment and angular momentum. While the dominant natural hydrogen isotope ($^1$H) does exhibit a non-zero spin, biomolecular NMR experiments often require isotopic labeling of additional atoms within a protein, generally $^{15}$N and/or $^{13}$C which have spin 1/2, as the most abundant isotope of carbon ($^{12}$C) has is zero-spin and is therefore insensitive to magnetic manipulation and the most abundant isotope of nitrogen ($^{14}$N) has a spin of 1, which introduces quadrupole relaxation that is prohibitively rapid for use in biomolecular NMR. Introducing non-zero spin nuclei into a strong static magnetic field $B_0$ (14.1-21.1 T for experiments described herein) leads to a preferential alignment with the $B_0$, providing a net alignment of the bulk magnetization ($M$) of a sample along the magnetic field (+z). This bulk magnetization can then be manipulated by application of a short radio frequency (RF) pulse, generating a weaker magnetic field perpendicular to $B_0$ (i.e., in x or y), causing rotation of $M$ into the x-y (transverse) plane, where the magnetization then processes about $B_0$ with a characteristic Larmor frequency determined by $B_0$ and the atom type. Additional small changes in the procession frequency of individual atoms are introduced by changes to the local magnetic environment due to their localization in the context of the rest of the protein. The NMR signal is recorded by measuring the current induced by the nuclear procession of all nuclei in a sample after cessation of the RF pulses, a signal known as the free induction decay (FID). To deconvolute the FID, a Fourier transform is applied to the FID to transform it to the frequency domain, yielding peaks with a distinct frequency, also known as chemical shift ($\omega$), associated with each monitored atom in the protein. $M$ can be transferred between different NMR active nuclei within the protein,
allowing collection of correlated chemical shift information among homo- or heteronuclei and yielding multidimensional spectra upon Fourier transformation.

The evolution of \( \mathbf{M} \) in the absence of an applied RF pulse is described by the Bloch equations, written here in the rotating reference frame\(^{110} \).

Equation 1

\[
\frac{d\mathbf{M}}{dt} = \begin{bmatrix} -R_2 & -\Omega & 0 \\ \Omega & -R_2 & 0 \\ 0 & 0 & -R_1 \end{bmatrix} \mathbf{M} + R_1 \mathbf{M}_0 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}
\]

where \( \Omega \) is the frequency offset (i.e., \( \Omega = \omega - \omega_{\text{rotating frame}} \)) and \( \mathbf{M}_0 \) is the initial magnetization aligned with \( \mathbf{B}_0 \). \( R_1 \) (or alternatively, its inverse, \( T_1 \)) describes longitudinal or spin-lattice relaxation that reflects the stochastic return of magnetization to the equilibrium state, aligned with the background magnetic field. \( R_2 \) (or alternatively its inverse \( T_2 \)) describes transverse or spin-spin relaxation and reflects the loss of coherence of magnetization in the transverse plane.

Three fundamental properties of a signal can be monitored by NMR spectroscopy. The chemical shift (\( \omega \)) is reflective of the local magnetic environment of a given atom within the protein, and may provide information on local secondary structural elements of the protein. The signal intensity (\( I \)), given by the peak area, is reflective of the total number of atoms processing at a given frequency. The signal linewidth (\( \lambda \)), given by the peak width at half height which is related to transverse relaxation by \( R_2 = \lambda \pi \) (in Hz). By monitoring these signal properties in the context of various NMR experiments, dynamics over wide range of timescale can be monitored.

**NMR methodologies to monitor protein dynamics**

Fast motions (ps-ns) in proteins are generally described in terms of the rigidity of a given bond vector, which in turn is reflective of the structural constraint on that vector due to the local secondary structure within the protein. Fast timescale motions in proteins can be
monitored by measurement of the transverse ($R_2$) and longitudinal ($R_1$) relaxation rates described above, along with the heteronuclear-NOE (hnNOE), a measurement of through space magnetization transfer between two heteronuclei (i.e., transfer from a proton to an attached $^{15}$N-labeled nitrogen). $R_2$ and $R_1$ can be directly monitored by observing the decay of magnetization as a function of delay time while magnetization is stored in the transverse plane or longitudinal axis, respectively. hnNOEs are measured on $^{15}$N amides by comparing signal intensity with or without proton pre-saturation. $R_1$, $R_2$, and hnNOE each have distinct dependencies on the spectral density function $J(\omega)$, which describes the amplitudes of motion experienced at a given frequency. $J(\omega)$ is influenced by the global tumbling of a protein, which occurs on the ns timescale, as well as any localized bond motions, which generally occur on the ps timescale. $R_2$ is additionally impacted by slower motions in the µs-ms timescale.

Interpretation of the relaxation measurements can take several forms. The most common is to use the Lipari-Szabo model-free approach, which assumes that internal motions are faster than and independent of the global tumbling of the protein, but otherwise assumes no structural information about the motions. The description of motions is defined by a global tumbling time ($\tau_m$), a site-specific correlation time ($\tau_e$), a site-specific squared order parameter ($S^2$), and any influence from µs-ms exchange ($R_{ex}$). Of these parameters, $S^2$ is most informative of fast timescale motions, and is generally interpreted in terms of the degree of restricted motion of the within a cone where a value of 1 is completely rigid and a value of 0 is completely unrestrained. Alternatively, $R_1$, $R_2$, and hnNOE can be interpreted directly across the sequence of a protein as reflective of regional internal motions within the protein.

Descriptions of slower motions (µs-s) are generally described as an exchange between distinct chemical environments (chemical exchange), described by the population
occupancies of the states \( P_A \) and \( P_B \), the rate of exchange \( k_{\text{ex}} = k_{A \rightarrow B} + k_{B \rightarrow A} \), and the chemical shift difference between the states \( \Delta \omega \). Depending upon the relationship between \( k_{\text{ex}} \) and \( \Delta \omega \), the exchange can fall into fast, intermediate, or slow exchange. In slow exchange, where \( k_{\text{ex}} \gg \Delta \omega \), distinct peaks are observable (assuming sufficiently high populations) for each chemical environment; in fast exchange, where \( k_{\text{ex}} \gg \Delta \omega \), a single peak is present at the population-averaged chemical shift position; intermediate exchange describes any case in between, \( k_{\text{ex}} \approx \Delta \omega \), wherein the peaks are linebroadened due to chemical exchange (see Figure 3 for representative spectra across the three regimes).

Slower timescale motions, from \( \sim 100\text{-}5000 \text{ s}^{-1} \) can be quantitatively monitored by the Carr-Purcell-Meiboom-Gill relaxation dispersion experiment (CPMG-RD), also known as \( R_2 \) relaxation dispersion\(^{112,113} \). These motions are on the timescale of sidechain reorientation, loop motions, and other structural rearrangements. The experiment consists of a constant relaxation time (\( \sim \)20-80 ms) during which magnetization is allowed to decay via \( R_2 \) relaxation. During this relaxation time, 180° refocusing pulses are applied in the transverse plane with varied frequency (\( \nu_{\text{cpmg}} \)) from \( \sim 50\text{-}1000 \text{ s}^{-1} \). Over that relaxation time, signal will decay according to the background \( R_{20} \) relaxation rate, with additional contribution from any chemical exchange on the \( \mu \text{s}-\text{ms} \) timescale (\( R_{\text{ex}} \)).

Equation 2

\[
R_2 = R_{20} + R_{\text{ex}}
\]

The application of the refocusing pulses leads to the retention of additional signal as a function of the refocusing frequency (see Figure 4), with increasing signal retained at higher refocusing frequencies. The signature of \( R_2 \) as a function of refocusing frequency is described for generalized two state exchange by the Carver-Richards equations\(^{114} \):
Equation 3

\[
R_2 = R_{20} + \frac{1}{2} \left[ k_{ex} - \frac{1}{2\tau} \cosh^{-1}(D_+ \cosh(\eta_+) - D_- \cosh(\eta_-)) \right]
\]

\[
D_\pm = \frac{1}{2} \left[ \pm 1 + \frac{\psi + 2\Delta\omega^2}{(\psi^2 + \xi^2)^{0.5}} \right]
\]

\[
\eta_\pm = \frac{2\tau}{\sqrt{2}} \left[ \pm \psi + (\psi^2 + \xi^2)^{0.5} \right]^{0.5}
\]

\[
\psi = (P_B k_{ex} - P_A k_{ex})^2 - \Delta\omega^2 + 4P_A P_B k_{ex}^2
\]

\[
\xi = 2\Delta\omega k_{ex}(P_B - P_A)
\]

where \(R_2\) is the observed \(R_2\) relaxation, \(R_{20}\) is the exchange independent component of \(R_2\) relaxation, \(\tau\) is the period of CPMG refocusing (\(\tau = 1/\nu_{cpmg}\)), \(k_{ex}\) is the exchange rate, defined as \(k_{ex}=k_{AB}+k_{BA}\), where \(k_{AB}\) and \(k_{BA}\) are the forward and reverse rate constants, respectively, \(\Delta\omega\) is the chemical shift difference between the two states, and \(P_A\) and \(P_B\) are the populations of the two states. Observed \(R_2\) relaxation is given by:

Equation 4

\[
R_2 = -\frac{1}{T} \log \left( \frac{l_T}{l_0} \right)
\]

where \(T\) is the constant time relaxation period, \(l_T\) is the peak intensity at a given refocusing time \(\tau\), and \(l_0\) is the peak intensity in the absence of the constant time relaxation period. Data are generally collected at two magnetic field strengths, which has been shown to be required to sufficiently constrain fits to the above equations.\(^{115}\) In theory, by fitting the data to the Carver-Richards equations, kinetic \((k_{ex})\), thermodynamic \((P_A)\) and structural \((\Delta\omega)\) information about the exchange process can be extracted, although this is only possible in the intermediate exchange regime (See Figure 5 for representative CPMG-RD \(R_{ex}\) profiles over a range of \(k_{ex}\) values).
CPMG-RD can be used to monitor exchange processes fully in the slow and fast exchange regimes, although less information can be determined about the exchange. Specifically, in the limit of slow exchange, $R_2$ is described by

Equation 5

$$R_2 = R_{20} + k_{AB} - k_{AB} \frac{\sin(\Delta \omega \tau)}{\Delta \omega \tau}$$

which contains information only on the ‘away’ rate (i.e., rate of exchange from A→B) and the chemical shift difference between the states. Alternatively, in the fast limit, $R_2$ is described by

Equation 6

$$R_2 = R_{20} + \frac{P_A P_B \Delta \omega^2}{k_{ex}} \left( 1 - \frac{2 \tanh(k_{ex} \tau)}{2k_{ex} \tau} \right)$$

for which the populations and $\Delta \omega$ are inherently coupled into a single parameter, precluding independent determination of either of them.

For much slower processes (~0.5-20 s$^{-1}$), fully in the slow exchange regime, and for which both states can be observed, ZZ-exchange spectroscopy (also known as exchange spectroscopy or EXSY) can be used to measure rates of exchange between states. In ZZ-exchange spectroscopy, the magnetization of each probe is labeled with its chemical shift and stored along the Z-axis, such that it is subject to $R_1$, and not $R_2$ relaxation. As $R_1$ is much smaller than $R_2$ for proteins, this allows for monitoring events occurring over timescales as long as multiple seconds. After a mixing time of varied length, the magnetization is transferred back to the transverse plane and recorded. Any atoms that have switched state during the mixing time will register as ‘cross-peaks’ as shown in Figure 6. The dependence of each of the four peaks is given by

Equation 7

$$R_2 = R_{20} + \frac{P_A P_B \Delta \omega^2}{k_{ex}} \left( 1 - \frac{2 \tanh(k_{ex} \tau)}{2k_{ex} \tau} \right)$$
Equation 7

\[
I_{A}(T) = I_{A}(0)(-\lambda_{2} - a_{11})e^{-\lambda_{1}T} + (\lambda_{1} - a_{11})e^{-\lambda_{2}T})/(\lambda_{1} - \lambda_{2}) \\
I_{B}(T) = I_{B}(0)(-\lambda_{2} - a_{22})e^{-\lambda_{1}T} + (\lambda_{1} - a_{22})e^{-\lambda_{2}T})/(\lambda_{1} - \lambda_{2}) \\
I_{AB}(T) = I_{A}(0)(a_{21}e^{-\lambda_{1}T} - a_{21}e^{-\lambda_{2}T})/(\lambda_{1} - \lambda_{2}) \\
I_{BA}(T) = I_{B}(0)(a_{12}e^{-\lambda_{1}T} - a_{12}e^{-\lambda_{2}T})/(\lambda_{1} - \lambda_{2}) \\
\lambda_{1,2} = 1/2[(a_{11} + a_{22}) \pm ((a_{11} - a_{22})^2 + 4k_{AB}k_{BA})^{0.5}] \\
a = \begin{bmatrix} R_{A} + k_{AB} & -k_{BA} \\ -k_{AB} & R_{B} + k_{BA} \end{bmatrix}
\]

where \(I_{A}\) and \(I_{B}\) are the peak intensities of peaks \(A\) and \(B\), \(I_{AB}\) and \(I_{BA}\) are exchange peak intensities, \(T\) is the mixing time, \(R\) is the longitudinal relaxation rate (\(R_{1}\)), \(k_{AB}\) and \(k_{BA}\) are the \(A\)\(\rightarrow\)B and \(B\)\(\rightarrow\)A exchange rates where \(k_{BA} = (P_{A}/P_{B})k_{AB}\) and \(P_{A}\) and \(P_{B}\) are the populations at equilibrium. By recording spectra while varying \(T\) over ~0.1-2 seconds, the exchange rate, \(k_{ex} = k_{AB} + k_{BA}\), and chemical species-specific \(R_{1}\) relaxation parameters can be determined.

The role of conformational exchange in regulating protein function

The role of dynamics in binding

The first qualitative descriptions of protein-substrate interactions followed the “lock-and-key” model by which a given protein is pre-formed to interact precisely with its partner.\(^\text{119}\) This model has been largely supplanted by two competing, if not directly contradictory, models of protein-substrate interactions, induced fit and conformational selection. The induced fit model posits that given binding partner will interact weakly with a protein, and in doing so, introduce conformational changes in the protein such that it then binds more tightly.\(^\text{120}\) Alternatively, the conformational selection model holds that a given protein is inherently sampling an ensemble of structures, dubbed the conformational landscape, with varying degrees of affinity for a given substrate, and that the binding event to a high affinity sub-population then shifts the overall population toward that high-affinity conformation.\(^\text{109,121}\) The two descriptions are analogous to the traditional competing models
of allostery, the Koshland-Némethy-Filmer (KNF) and the Monod–Wyman–Changeux (MWC) models, respectively, discussed further below and illustrated in Figure 7.

A particularly compelling landmark set of experiments in favor of conformational selection utilized low temperature (<200 K) flash photolysis to inhibit transitions between conformational states in myoglobin and demonstrated a variable affinity for O₂ among the various substates. Additional, the advent of high-resolution NMR and atomic resolution molecular dynamics simulations has demonstrated the sampling of high energy, bound-like states in native proteins, supporting the notion of conformational selection. The two models of binding are not mutually exclusive, however, and instead represent extremes on a continuum of mechanisms by which binding may occur. A more generalized and inclusive model, termed the ‘extended conformational selection’ model, suggests that conformational heterogeneity does exist within most free proteins, consistent with the conformational selection model, but that during the binding event, further conformational adjustment often occurs in the protein and/or binding partner, consistent with an induced fit. Analytical methods have been developed to theoretically distinguish the extent to which conformational selection or induced fit act in a given system, however, experimental limitations often limit the applicability of these approaches.

The role of dynamics in enzymatic catalysis

As described above, the role of inherent conformational fluctuations in substrate binding, via conformational selection, is well established. Likewise, numerous studies have established the role of inherent conformational fluctuations in the ligand-bound protein in regulating substrate release, as various substates of the bound complex also exhibit varied affinity for the substrate. For example, in dihydrofolate reductase (DHFR), a loop motion occurs in the ligand-bound form which confines dihydrofolate in the major form, trapping the substrate sufficiently long to permit catalysis, yet the loop transiently samples a low-population higher energy open form that permits substrate release. In this respect, the
role of inherent conformational fluctuations in regulating the complete catalytic cycle has been demonstrated.

Significant controversy, however, surrounds the idea that the inherent dynamics of a system influence the rate of catalysis subsequent to substrate binding. Broadly, enzymatic catalysis is thought to occur via an enzyme’s high affinity for the transition state of a reaction, stabilizing the transition state, and thereby lowering the activation energy required to transverse to product.\textsuperscript{132} The potential effect of conformational fluctuations in reaching the stabilized transition state can be dissected into two distinct impacts on catalysis: the conformational search of the enzyme-substrate system in reaching the optimal configuration to stabilize the transition state and a direct influence of dynamical motions on the chemical step of catalysis.

While the debate over the relevance of dynamics in catalysis has been waged by many parties, the controversy has been most notably argued between recent co-recipients of the Nobel Prize, Drs. Karplus and Warshel.\textsuperscript{133,134} Warshel has strongly argued against a role for conformational dynamics in influencing enzymatic catalysis, while Karplus has advocated its role. With regard to a direct influence on the chemical step, Warshel argues that the inherent friction present in a protein system prevents any ‘memory’ in the progression of the substrate over the activation energy barrier, such that any dynamic motions present prior to reaching the chemical step of catalysis are irrelevant in passing over the chemical barrier. He additionally argues that, in order to be considered as a dynamical influence on catalysis, motions must exhibit behavior deviating from a Boltzmann distribution, as otherwise any observed motions are simply a reflection of the energy landscape. Karplus largely agrees with the former point, indicating that energy dissipation is generally sufficiently fast as to prevent non-equilibrium behavior in a bound enzyme-substrate system. Oh the later point, however, Karplus argues that inherent conformational fluctuations in the system permit lowering of the quasithermodynamic free energy of
activation, providing evidence in the form of simulations in which a flexible enzyme proceeds through a lower activation barrier than a rigid one.\textsuperscript{135} The essence of the argument on the latter point may be largely semantic, as the motions of an enzyme necessarily follow Boltzmann behavior, and are dictated by (and reflective of) the energy landscape; nonetheless, understanding the particular motions present in an enzyme:substrate complex and the pathway by which the system reaches the activation barrier is important to understanding the catalytic mechanism of a given system.

\textit{The role of dynamics in mediating allosteric interactions}

While Pauling initially introduced the idea of intramolecular protein regulation to control oxygen binding in hemoglobin in 1935,\textsuperscript{136} the term allostery was not applied to the phenomenon until 1961 by Monod and Jacob.\textsuperscript{137} Subsequent to the determination of the structure of the prototypical allosteric protein, hemoglobin, in 1960,\textsuperscript{138} the long-range nature of the allosteric information transfer became apparent, as well as that the phenomenon was applicable not only to hemoglobin, but also to many additional proteins. Two phenomenological models arose to explain the nature of allostery (depicted in Figure 7). Both models concern symmetric connected ‘subunits’ that transition between distinct ‘tense’ and ‘relaxed’ states. The first, the ‘sequential’ or KNF model, which is similar to the model originally proposed by Pauling, posits that each subunit can individually transition from the tense state to the relaxed state upon ligand binding, and that the likelihood of ligand binding is influenced by the state of other subunits.\textsuperscript{120} The second, ‘symmetric’ or MWC model, assumes global transitions, such that all subunits adopt the tense or relaxed states simultaneously, with ligand affinity increased at each subunit in the relaxed state and the relaxed state stabilized by ligand binding.\textsuperscript{139} The KNF and MWC models essentially represent two-state versions of the induced fit and conformational selection models of binding discussed above. While the original formulations of these models concerned positive cooperativity in symmetric homooligomeric subunits, they have been expanded to
accommodate allosteric interactions within a single domain, and to incorporate negative cooperativity.\textsuperscript{140-142}

Inherent in the MWC model is the concept that proteins are constantly sampling a higher energy state, even in the absence of ligand.\textsuperscript{140,142} This concept has been expanded upon to encompass not only two-state transitions, but also systems for which a broad ensemble of structures is being sampled. This expanded view of allostery, referred to as ‘dynamic allostery,’ the ‘population-shift model,’ or the ‘ensemble nature of allostery,’ centers on the view that all proteins are dynamically transitioning to some extent between multiple states, and that any influence (e.g., binding event, post-translational modification, mutation) that preferentially stabilizes certain sub-populations within the ensemble will shift the population distribution with subsequent impact on protein function.\textsuperscript{143-145} In this regard, it has been suggested that, with perhaps the exception of fibrous proteins, all proteins are allosteric, with the predominant variability being in the degree to which known influences impact population sampling and the degree to which population shifts influence function.\textsuperscript{146} This model incorporates an important notion absent in the two-state model, namely that allosteric effects may be mediated absent significant changes to the average or ground state structure of a given protein. The homodimeric transcription factor CAP, which exhibits negative cooperativity between subunits for binding cyclic AMP (cAMP), presents an illustrative example of this effect. Binding of a single cAMP induces increased conformational flexibility in the other subunit without altering the ground-state structure, increasing the entropic penalty for a second cAMP binding, and thereby reducing affinity.\textsuperscript{147} Many other allosteric systems have been probed via CPMG-RD and networks of coherent or non-coherent residues have been identified that propagate dynamic allosteric information independent of ground state structural changes.\textsuperscript{148-151} Unlike the more rigid two-state model, this expanded, ensemble-perspective model also accommodates intrinsically disordered
proteins, which inherently sample an extremely broad conformational landscape that can nonetheless be allosterically modulated to influence subsequent binding events.\textsuperscript{152,153}

**Dynamic relationships between protein segments**

As described above, NMR provides a powerful approach to quantitatively measure motion with residue specific probes in proteins, particularly motions in the µs-ms timescale near many biologically relevant events via CPMG-RD. While the particular dispersion curves are readily obtainable for many systems, the mechanistic interpretation of the data is often challenging. Of particular interest is the establishment of the relationship between various sites for which dispersion is detected. Broadly, three ‘dynamic relationships’ may exist between any two probe sites.\textsuperscript{101} Two sites may be directly reporting on the same physical transition, for which we would expect identical exchange rates ($k_{\text{ex}}$) and population occupancies, as well as identical reactions to changing conditions (e.g., temperature, buffer, mutation). Alternatively, two probe sites may report on entirely independent physical phenomena, for which we may see similar or different $k_{\text{ex}}$ and population values, and similar or different responses to changing conditions (i.e., similar rates, populations, and responses to changing conditions are suggestive of reporting on the same physical phenomenon, but not definitive). Finally, two probe sites may report on distinct, yet coupled physical motions, for which the sites may report similar or different $k_{\text{ex}}$ and population values, but will experience correlated responses to altered conditions.

Systems for which largely coherent motions dominate have been most extensively studied in terms of relating protein dynamics and function. While experimental and technical improvements have opened the application of relaxation dispersion experiments to a wider range of applications, extensive studies of the role of µs-ms conformational dynamics have revolved around a thorough understanding of a small number of model systems that are readily amenable to study by NMR spectroscopic methods.
Kay and co-workers have been the dominant force in developing novel relaxation dispersion methods, including new labeling schemes and pulse sequences to access an increasing number of probe sites.\textsuperscript{154-158} In particular, the lab has utilized these techniques to access structural information about low-population states that are otherwise invisible to other structural methods.\textsuperscript{159-162} The most notable success of this approach was the determination of an atomic resolution model of a transient folding intermediate of a small FF domain with a millisecond lifetime and a population of 2-3%.\textsuperscript{163} To achieve this, relaxation dispersion experiments were carried out on five backbone probe sites (C\textsubscript{\alpha}, C\textsubscript{O}, H\textsubscript{\alpha}, H\textsubscript{N}, and N\textsubscript{H}); the data were fit to a single two-site exchange model, from which the low-population chemical shifts could be determined for each of the probe sites experiencing exchange. Using previously established empirical relationships between chemical shift values and secondary structure, an atomic resolution model of the low-population state could be reconstructed. This success highlights the power of relaxation dispersion approach, but likewise exposes limitations in that achieving an atomic resolution model was dependent on the global two-state nature of the conformational change and its application to a small (71 residue) protein.

Global two-state exchange does, however, appear to play a role in the catalytic cycles of many biologically relevant and NMR accessible systems. For instance, Loria and co-workers identified concerted conformational motions in an active site loop in ribonuclease A representing transitions between an open and closed state of the substrate binding pocket.\textsuperscript{164} Notably, the transition between the open and closed states exists in both the substrate-free and a substrate-mimic-bound form, with the substrate-mimic simply shifting the major state to the closed rather than open form, suggesting a role for the motion in substrate release. The rate constant of the coherent exchange remains similar in each case, and is likewise similar to the rate of substrate release, $k_{\text{off}}$, and to $k_{\text{cat}}$; combined, these data indicate that the transition to the open state is the limiting step in catalytic turnover in
ribonuclease A, such that the inherent conformational motions are directly responsible for regulation of the catalytic cycle.

Relaxation dispersion experiments have also been applied to systems with a surprisingly complex set of conformational rearrangements over a multi-step cycle. DHFR undergoes a five step catalytic cycle involving separate, sequential binding steps for the dihydrofolate substrate and NADPH co-factor, the catalytic step of proton transfer, and sequential release of the product and oxidized cofactor.\textsuperscript{131,165} Wright and co-workers have analyzed the conformational fluctuations in DHFR for each of the five steps in the cycle by utilizing a non-catalyzable mimic, allowing data collection on each of the five ground-state structures in the cycle. For each state of the cycle, they discovered that the enzyme samples a high-energy, low-population minor state; analysis of the chemical shift values of these minor states demonstrated that, in each case, the minor state sampled is similar to that of the next structure in the cycle. This study demonstrates the importance of conformational selection in both substrate binding and release, and also highlights the ability of ligand binding to impact both the ground state structure and the low-population minor state sampled. While the exchange rate and population occupancies of the minor states vary between ground state structures, in each case, the exchange parameters were sufficiently similar to fit to a global two-state exchange model. Wright and co-workers further identified a mutation in DHFR that abolishes detectable millisecond timescale dynamics within the active site independent of changes to the ground-state structure of the enzyme and observed a dramatic drop in $k_{\text{cat}}$.\textsuperscript{108} These results suggest that dynamics permit the conformational search for the optimized transition-state necessary for efficient catalysis in addition to regulating substrate binding and release.

In addition to the globally coherent systems described above, a smaller number of studies have investigated systems with more localized conformational sampling. Kay and co-workers investigated $\mu$s-ms motions in the enzyme complex of NAD(P)H:flavin
oxidoreductase bound to flavin adenine dinucleotide and identified three distinct regions of localized motion. One of the observed groups of residues appears to result from a localized unfolding event of one α-helix and is seemingly unconnected to the other groups. The other two groups, however, include residues involved in binding the proton acceptor (FAD) and the proton donor (Tyr 35), respectively, and are therefore in close physical proximity. No attempt was made to ascertain the degree to which motions in the second two groups are linked (i.e., if any allosteric interactions exist), so these two groups may exhibit fully independent localized motions or coupled, yet distinct motions.

McDonald et al. investigated conformational motions in the bacterial signaling protein CheY, which is known to exhibit allosteric regulation in which phosphorylation switches the protein to a predominantly active conformation. Utilizing CPMG-RD, they discovered that this conformational switch does not proceed through a coherent two state conformational switch. Rather, it appears that multiple residues or small groups of residues undergo localized two-state switching, yet are nonetheless coupled in that alterations to dynamics induced by Mg$^{2+}$ binding are propagated throughout the conformational flexible regions of the protein.

Kleckner et al. likewise investigated dynamic allostery in the trp-RNA binding attenuation protein via methyl 13C-CPMG-RD and identified a non-coherent network of coupled residues. The motions of these residues are quenched upon Trp binding and are proposed to coalesce, promoting subsequent binding to trp-RNA. This model represents a hybrid of localized MWC and KNF models in which individual segments each sample ‘active’ conformations, but the lack of coherence renders the entire binding site unlikely to simultaneously adopt a binding competent state. Trp binding, however, induces coherence among the segments, permitting the adoption of a fully binding competent site. A similar model is proposed by Xiao et al. for extracellular signal-regulated kinase 2 (ERK2). In the absence of phosphorylation, ERK2 exhibits largely non-coherent methyl dynamics; upon
phosphorylation, however, these motions appear to become largely coherent, such that a single global exchange phenomenon can explain the bulk of the observed dynamics.

These studies serve to highlight the diversity of mechanisms by which nature utilizes conformational flexibility in protein systems and the difficulty in formulating a coherent theory of the precise role of dynamics in protein systems.

**Studies of protein dynamics among homologues**

Given the appreciation for the role of conformational flexibility in regulating protein motions, many researchers have sought to determine the degree to which protein dynamics are conserved within a protein family, as structure tends to be.

Wolf-Watz et al. examined µs-ms dynamics in a mesophilic and thermophilic adenylate kinase (Adk) and identified a conserved regulatory mechanism by which a concerted lid opening is the rate-limiting step in catalytic turnover. This mechanism has been evolutionarily tuned and dramatically slows kinase activity in the thermophile at low temperatures, yet permits activity at high temperatures as the rate of lid opening increases. Further studies by the group have suggested that Adk provides a framework for continuous evolutionary tuning of loop dynamics to accommodate the specific needs of a given organism. Whittier et al. suggest a comparable mechanism among protein tyrosine phosphatases (PTPs), which exhibit a loop closing mechanism that has been evolutionarily tuned to accommodate the particular environment of its host.

Alternatively, Bhabha et al. found that, despite the highly conserved structure of DHFR between humans and *E. coli*, the inherent dynamics of the enzyme have been shifted dramatically in both nature and timescale, effecting adoption of a different catalytic cycle (i.e., ordering of folate and NADP(H) binding and release) in the two organisms. These changes are suggested to result from evolutionary adjustment to the levels of THF and NADP* in human vs. *E. coli* cells, but point to a fully altered functional mechanism as opposed to the continuous tuning of motions observed in Adk and PTPs.
Studies of dynamics in human Cyclophilin A

Experimental studies of dynamics in human Cyclophilin A

Along with systems described above, CypA has been an instrumental model system in understanding the role of inherent conformational fluctuations in regulating protein dynamics. The original studies originating from the Kern group (Eisenmesser et. al (2002)\textsuperscript{172} and Eisenmesser et. al (2005)\textsuperscript{106}) examined conformational dynamics in CypA alone and in the presence of high concentrations of a short tetra-peptide Suc-Ala-Phe-Pro-Phe-4-NA via CPMG-RD. The studies assumed that at sufficiently high peptide concentrations, dynamics measured on near-saturated CypA would be reporting solely on cis$\leftrightarrow$trans isomerization, with no effect from substrate binding and release. As demonstrated in Chapter 4 of this manuscript, the assumptions made in utilizing this weakly binding peptide are incorrect, and reported dynamics in the presence of the weakly binding tetra-peptide cannot be representative of on-enzyme isomerization. However, these studies did identify significant $\mu$s-ms conformational flexibility in CypA both alone and in the presence of substrate, in both cases localized predominantly to the active site and surrounding residues. This finding suggests that these motions are somehow linked to protein function. In the free protein, these studies identified two distinct regions of coherent motions, while identifying a globally coherent exchange in the bound state (although one that cannot be indicative of isomerization, as noted). Further, comparison of CPMG-RD profiles in multiple active site mutants suggested a ‘common dynamic network’ in which populations of all mobile resides in the protein are shifted proportionally in response to perturbation.

More recently, Schlegel et al.\textsuperscript{173} employed deuterated CypA, which improves signal quality by reducing the background $R_{20}$ relaxation rate,\textsuperscript{174} and data collection at multiple magnetic field strengths, which improves the accuracy of parameters derived from fitting the data,\textsuperscript{175} to measure $\mu$s-ms motions in free CypA over a range of temperatures from 0-25°C.
Because of the improvement in data collection, $R_{ex}$ was observed for nearly 50% of the residues in the protein, with nearly the entire active site exhibiting measureable exchange. Within the active site, they identified significant variability in the temperature dependency of different residues, indicating that the motions within CypA do not comprise two coherent regional motions as previously suggested. Rather, by individually fitting the residues within the active site to the Carver-Richards equations to extract exchange rates, they suggest that conformational fluctuations consist of a ‘dynamic continuum’ for which non-coherent rates of motion exist spanning nearly an order of magnitude across the active site. Furthermore, by generating single and double active site mutants of CypA, they demonstrate regional, not global, responses to mutational perturbation, indicating the existence of localized dynamic networks rather than the ‘common dynamic network’ previously suggested.

Expanding upon those studies utilizing the weakly binding tetra-peptide described above, Bosco et al.\textsuperscript{176} examined conformational dynamics of CypA in complex with the N-terminus of the HIV Capsid protein (CA\textsuperscript{N}). As described in Chapter 4, CypA’s binding affinity for CA\textsuperscript{N} is sufficiently tight that measurement of CPMG-RD on the complex can potentially monitor isomerization, although all measured motions in the complex are not \textit{a priori} indicative of isomerization. In the context of the complex, both CypA and CA\textsuperscript{N} exhibit similar rates of motion, and rates that differ significantly from those of either protein alone, supporting the notion that the measured dynamics are reporting on isomerization. They further analyze the impact of a number of active site mutants in terms of their ability to catalyze isomerization of the tetra-peptide and CA\textsuperscript{N}. Interestingly, several of the mutations decrease isomerization of the peptide by \textasciitilde1000 fold, while only minimally impacting CA\textsuperscript{N} isomerization. This dichotomy is a result of the rate-limiting step for CypA’s interaction with either substrate. With respect to the peptide, on-enzyme isomerization is the rate limiting step for CypA\textsuperscript{WT} and all mutants; however, with respect to CA\textsuperscript{N}, substrate release from CypA\textsuperscript{WT} is nearly two orders of magnitude slower than on-enzyme isomerization. Analysis
reveals that mutations that both reduce on-enzyme isomerization and increase off-rates (decrease substrate affinity) can result in a shift of the rate-limiting step to on-enzyme isomerization for CA\textsuperscript{N}. In these mutants, however, the more rapid substrate release results in comparable flux through the enzymatic cycle as in CypA\textsuperscript{WT}. This study highlights the importance of considering the full catalytic cycle in the analysis of CypA isomerization and in interpreting functional impacts of a given mutation.

In addition to CPMG-RD based studies of conformational sampling in CypA, the Alber\textsuperscript{177} and Fraser\textsuperscript{178} groups have utilized room temperature X-ray crystallography to identify minor state conformations of residues in CypA which are represented by low-contour electron density. They identified a series of residues in and adjacent to the active site which sample lowly occupied secondary states in the crystal and designed a single site mutation immediately adjacent to the active site (Ser 99 \rightarrow Thr) that shifts the sampling of the adjacent Phe 113 and Met 61 such that they now exist predominantly in what are the minor states in CypA\textsuperscript{WT}.\textsuperscript{177} This mutation leads to a drastic decrease in substrate isomerization with a relatively minor decrease in substrate affinity, suggesting that the mutation impacts isomerization to a much greater degree than affinity. They further suggest that isomerization is directly coupled to the sampling of the minor state by Phe113. Van den Bedem et al.\textsuperscript{178} proceeded to automate the identification of sampled minor states from X-ray crystal structures and have identified small clusters of coupled residues in CypA which are consistent with the degree of localization identified by Schlegel et. al,\textsuperscript{173} but lack the timescale information necessary to directly correlate these sampled states to the rates of motion measured by CPMG-RD.

**Computational studies of dynamics in Cyclophilin A**

In addition to the many experimental studies examining the role of conformational fluctuations in CypA function, a number of computational studies have provided additional insight as to the mechanism of CypA isomerization and the role of dynamics in mediating
catalysis. The peptidyl-prolyl isomerization process is an attractive one to approach via molecular dynamics (MD), due to the lack of bond breakage or formation. However, even with the $\sim 10^5$ acceleration of isomerization due to enzymatic catalysis, turnover occurs over millisecond timescales, while the conventional MD simulations only sample within the 10s of microseconds. Likewise, the conformational fluctuations measured throughout the CypA active site exist on the µs-ms timescale. Thus, a number of parallel approaches have been attempted to understand the mechanism of CypA isomerization.

Li and Cui\textsuperscript{179} investigated the role of the ‘catalytic arginine’ (Arg 55 in CypA), which is absolutely conserved among catalytic cyclophilins, using short hybrid quantum mechanical/molecular mechanics modeling of CypA in complex with a short peptide in cis, trans, and transition states. They found that the predominant role of the arginine was to stabilize the transition state through electrostatic interactions with the carbonyl oxygen of the isomerized proline, with additional stabilizing contributions the highly conserved Gln 63 and Asn 102. More recently, Camilloni et al.\textsuperscript{125} utilized NMR chemical shift constrained MD simulations in concert with density functional theory calculations to identify a critical and conserved electric field within the CypA active site that acts in an ‘electrostatic handle’ mechanism on the dipole of the prolyl carbonyl group to lower the energy barrier. They found that Arg 55 provides much of the electrostatic charge to generate this field, but that a weaker field remained even in the absence of charge on the Arg 55 guanidino group, consistent with persistence of much slower catalysis even with an Arg 55 $\rightarrow$ Ala mutation.

In addition to probing the electrostatic contributions to transition state stabilization, many studies have probed the role of dynamics in the CypA cycle. Camilloni et al. also identified significant conformational selection in CypA, such that the free protein, CypA samples states corresponding to both the cis-bound and trans-bound states of the enzyme.\textsuperscript{128} Agarwal et al.\textsuperscript{180} used quasi-harmonic analysis in concert with umbrella sampling of near-transition state complexes to identify networks of vibrational motion over picosecond
to millisecond timescales coupled to residues involved in hydrophilic and hydrophobic interactions with the substrate. In a subsequent study, Agarwal\textsuperscript{181} further analyzed the role of vibrational motion by computationally introducing kinetic energy to particular modes thought to be associated with catalysis, permitting transition state crossing in computationally accessible timescales. He was able to show that adding energy to particular modes of motion promoted turnover and are therefore relevant to catalysis, while adding energy to other modes had no impact. Ramanathan et al.\textsuperscript{182} expanded upon this work utilizing quasi-anharmonic analysis to identify a hierarchy of catalytically relevant sub-states along the trajectory toward the transition state.

Doshi et al.\textsuperscript{183} introduced an alternative approach to analyzing the role of conformational dynamics in CypA function by applying accelerated molecular dynamics,\textsuperscript{184} which introduces a bias potential preferentially to low energy states, promoting more rapid transition state crossing. This method, in combination with lowering the energy barrier in the isomerized prolyl bond, allows for unbiased isomerization to occur either in water or in the enzyme pocket. They found that conformational motions, in fact, restrict rotational motions of the substrate during enzymatic catalysis relative to a substrate alone in water, but that the inherent conformational sampling is nonetheless critical for the correct positioning of active site residues to stabilize the transition state. Further studies by the group have demonstrated that, as was shown by Camilloni et al.,\textsuperscript{125} the free enzyme undergoes conformational selection, sampling sub-states corresponding to the cis-bound and trans-bound conformers. Further, McGowan and Hamelberg\textsuperscript{185} show that inherent sampling present in CypA while bound to the cis or trans isomers of the substrate is restricted and stabilized by the transition state, and that the inherent motions of the enzyme complex are necessary to reach the stabilized transition state.
Figure 1. Peptidyl prolyl isomerization Isomerization reaction catalyzed by cyclophilins and other PPIases.

Figure 2. Cyclophilin A structure bound to cyclosporine and model peptide (a) Structure of CypA bound to CsA. (b) CypA with the peptide substrate GSFGPDLRAGD docked into the active site.
Figure 3. NMR chemical exchange regimes Simulated lineshapes for two residues undergoing two-state exchange over a range of $k_{ex}$ values from 0 (no exchange) to $10^6$ s$^{-1}$. In the slow limit ($k_{ex} \ll \Delta \omega$), the peaks are distinct. In the fast limit ($k_{ex} \gg \Delta \omega$), the peak appears at the population averaged position with an intensity corresponding to the sum of the magnetization. In the intermediate state ($k_{ex} \approx \Delta \omega$), both peaks are linebroadened and shifted toward the average. The minor state population was set to 20% of the major state, and $\Delta \omega$ was set to 300 Hz.)
Figure 4. CPMG-RD refocusing  Simulated lineshape from a probe with no chemical exchange (dotted lined) or experiencing chemical exchange on the µs-ms timescale (solid lines), with the application of a varied frequency of refocusing pulses.

Figure 5. CPMG-RD profiles  Simulated CPMG-RD profiles, calculated using a range of $k_{ex}$ values to demonstrate the profile shapes and field dependencies. Profiles were calculated at 21.1 T (red) and 12.1 (blue). For all simulations, $\Delta \omega$ was set to 2 ppm, and the minor state population was set to 10%.

Figure 6. ZZ-exchange example  Representative ZZ-exchange spectra collected with mixing times of 0-1.2 s. Peak intensities are plotted to the right as a function of time (dots), and the best fit solution to the ZZ-exchange equations is plotted as lines.
Figure 7. Classic models of allostery. KNF and MWC models of allostery as illustrated by a 2-subunit model protein complex. In the KNF model, substrate binding produces a conformational change in the bound subunit and induces a higher affinity for the substrate in the additional subunit(s). In the MWC model, the subunits transition coherently between states such that all subunits adopt identical conformations; substrate binding stabilizes the relaxed state, increasing substrate affinity in the unbound subunit(s). Squares represent subunits in the tense state, while circles represent subunits in the relaxed state. Filled shapes indicate a substrate-bound subunit, while open shapes indicate an unbound subunit.
CHAPTER II
A STRUCTURAL, FUNCTIONAL, AND DYNAMIC COMPARISON OF THE CYCLOPHILIN FROM THE THERMOPHILIC BACTERIUM *GEOBACILLUS KAUSTOPHILUS* WITH HUMAN CYCLOPHILIN A

Introduction

Previous studies of temperature extremophiles and mesophilic homologs have been utilized to explore the interplay between protein dynamics, stability, and enzymatic function. Broadly, these studies have found that a trade-off exists between stability and function; while conformational dynamics are required for enzymatic function, increased dynamics require less rigidity, and are therefore also destabilizing.\(^{168,189}\) Thus, evolution has tuned the dynamics of proteins to optimize this balance for the environment in which a given organism lives. Psychrophilic (cold-adapted) enzymes function well at low temperatures, but quickly destabilize and unfold at moderate temperatures, while mesophilic enzymes are too stable at low temperatures to function efficiently and are destabilized at high temperatures, but function ideally at moderate temperatures.\(^{190,191}\) Likewise, thermophilic enzymes at low or moderate temperatures generally function inefficiently due to hyper-stabilization, but become sufficiently dynamic at higher temperatures to function effectively.\(^{192}\)

While CypA has been extensively utilized as a model system with which to study the links between protein dynamics and function, including multiple studies described herein, no extremophilic cyclophilin has been previously characterized. Although numerous biological roles have now been discovered for cyclophilins, their function was initially ascribed to their chaperone-like activity in protein folding where they catalyze the often rate-limiting step of

---

proline isomerization. This chaperone-like activity may be especially important under extreme conditions where cyclophilins are often over expressed, such as in tumors for human cyclophilins, but also in organisms that thrive under extreme conditions, such as thermophilic bacteria. Thus, we have characterized the structure, enzymatic function, and dynamics of the sole cyclophilin (GeoCyp) encoded in the genome of the thermophilic bacterium *Geobacillus kaustophilus*, including a structural, functional, and dynamic comparison to human CypA. These studies reveal that, despite the vast evolutionary time separating the two cyclophilins and the drastically different environments for which the two proteins are adapted, their *in vitro* binding and catalytic functions are remarkably similar, and that they likewise exhibit similar dynamic profiles over a relatively broad range. These findings are not in line with previous findings of other thermophile-mesophile enzyme pairs, hinting that cyclophilins’ specific roles within the cell may have subjected the proteins to different evolutionary pressures than other protein families previously studied.

**Results**

*Chemical shift determination and secondary structure prediction of GeoCyp*

As no thermophilic homologue of human Cyclophilin A had been previously characterized, we identified the sole cyclophilin (GeoCyp) in the genome of the thermophilic bacterium *Geobacillus kaustophilus*, which had been classified in a survey of deep water vents and sequenced by Takami et al. Using purification via nickel affinity and gel filtration chromatography, we generated and purified $^{13}$C$^{15}$N-isotopically labeled GeoCyp for use in determining the NMR chemical shift assignments and solution structure of the protein. We determined near complete backbone and sidechain chemical shift assignments of GeoCyp at 25°C (BMRB accession number 18628). Specifically, assignments included 100% of amide protons, 99% of amide nitrogens, 99% of C$\alpha$’s and C$\beta$’s, and 91% of sidechain protons.
Prior to structural determination, we utilized the chemical shift assignments to determine the probable secondary structure of GeoCyp and perform a comparative analysis to humans CypA.\textsuperscript{196} Specifically, the $C_{\alpha}$ chemical shift differences relative to a random coil are shown using the Random Coil Index (RCI) along with the RCI predicted secondary structure (Figure 8).\textsuperscript{197} For CypA, both the previously published $C_{\alpha}$ and $C_{\beta}$ chemical shifts were used\textsuperscript{196} and for GeoCyp all nuclear chemical shifts determined here were used. While the majority of predicted secondary structure elements are similar between the two proteins, the predicted $\beta$-strand of GeoCyp comprising residues 68–71 and the predicted $\alpha$-helix comprising residues 126–129 are somewhat shifted in sequence from their CypA counterparts. This may indicate that the bioinformatically estimated sequence alignment does not necessarily represent the structural alignment.

It is important to recognize that the chemical shift predictions of secondary structural elements for CypA are somewhat different than the actual structures determined by both NMR and X-ray crystallography (see PDB accession numbers 1OCA, 1RMH, 3CYS) and are, therefore, likely significantly different for GeoCyp as well.\textsuperscript{186,196,198} The underlying reason for such discrepancies in secondary structure predictions solely using chemical shifts may be that there is a significantly higher representation of aromatic residues in both cyclophilins relative to other proteins (i.e., His, Trp, Phe, and Tyr). These include 20 aromatic residues of the 146 total for GeoCyp (14%) and 22 aromatic residues of the 165 total for CypA (13%). Specifically, for the previously determined structures of CypA, observed $\beta$-strands comprise residues 5–12, 15–24, 53–57, 61–64, 97–100, 112–115, 127–134, 156–163 and $\alpha$-helices comprise residues 30–41, 120–122 and 136–145, which all differ by at least several residues from predictions based on chemical shifts alone (in Figure 8). In fact, for CypA the critical 3$_{10}$ helix of residues 120–122 (sequence, EWL) and a relatively long $\beta$-strand of residues 127–134 (sequence VVFGKVKE) are not even identified by RCI. The converse is
also true for CypA where two chemical shift predicted \(\beta\)-strands are found within relatively flexible loops as indicated by relaxation rates.\(^{172,199}\) For GeoCyp, several \(\beta\)-strands likely exist that are not identified by chemical shift predictions alone. For example, one likely \(\beta\)-strand is within the active site and includes the highly conserved catalytic Arg 47 in GeoCyp (Arg 55 in CypA), which is responsible for destabilization of the prolyl-peptide bond of substrates. For GeoCyp this \(\beta\)-strand includes residues 45–49 (sequence FHRVI) that are highly conserved relative to CypA residues 53–57 (sequence FHRII) and is therefore likely present despite not being predicted by the RCI.

**Solution structure of GeoCyp and comparison to CypA**

In order to determine a solution structure of GeoCyp, \(^{15}\)N-edited and \(^{13}\)C-edited NOESY experiments were collected, allowing assignment of 1300 inter-residue NOEs. Using the CS-ROSETTA implementation of the ROSETTA structure prediction platform, we determined the NMR solution structure of GeoCyp with an average RMSD among the 20 lowest energy structures of 0.72 Å (see Figure 9 and Table 1 for full statistics). GeoCyp adopts a typical cyclophilin fold, consisting of 8 antiparallel \(\beta\)-strands arranged in a \(\beta\)-barrel, with 2-3 turn \(\alpha\)-helices capping either end of the barrel, and an additional short 3\(_{10}\) helix aligned parallel to the \(\beta\)-strands within the catalytic active site. Despite only having 41% identity, the \(\beta\)-barrel structures of and GeoCyp and CypA are nearly superimposable, as are the 3\(_{10}\) helix and the first of the \(\alpha\)-helices (backbone RMSD of \(\alpha\)-helices and \(\beta\)-sheets is 0.95 Å vs. a global backbone RMSD of 1.55 Å). Figure 9b shows an alignment of CypA with the lowest energy GeoCyp structure determined. GeoCyp contains several significantly shortened loops relative to CypA. Residues 12-15, 43-45, and 77-79 in CypA are absent in GeoCyp (Figure 9b, green); additionally, a 12 residue span (Figure 9b, red) comprising the final turn of the second \(\alpha\)-helix (\(\alpha2\)) and a loop between the helix and final \(\beta\)-sheet (\(\beta8\)) is absent in GeoCyp, which instead contains a two residue stretch directly linking the truncated
α-helix to the β-sheet. This α2-β8 loop comprises CypA residues 143-154 that are replaced by GeoCyp residues 132-133.

Previous studies have shown that thermophilic proteins tend to be shorter in sequence length and contain deletions in solvent exposed loops relative to their mesophilic homologues, suggesting that the overall shorter length of GeoCyp, including the significantly shortened α2-β8 loop, and perhaps other loop deletions, may play roles in enhancing the stability of GeoCyp. While the α2-β8 loop is a region of structural diversity among cyclophilins, the extent of the deletion in GeoCyp is only present in two other cyclophilin-like domains whose structures have been determined, both from other thermophilic bacteria, *Archaeoglobus fulgidus* and *Thermotoga maritime*. While these other two thermophilic proteins do adopt typical cyclophilin folds, they retain very low homology to catalytically active cyclophilins, including lacking the absolutely conserved catalytic arginine (Arg 55 in CypA and Arg 47 in GeoCyp), indicating that they are not functional isomerases. Alternatively, the deletions to residues 12-15 and 43-45 have been previously identified in other cyclophilins, including other human homologues.

While GeoCyp retains most of the canonical cyclophilin active site residues, there are two exceptions: His 110 in GeoCyp, which is typically a tryptophan, though is present as a histidine in several catalytically active cyclophilins, and Val 53, which is typically a methionine, but occasionally replaced by other hydrophobic residues. Notably, in the unbound state for which we have determined the structure, the GeoCyp ‘gatekeeper’ residues adopt an occluded conformation, with Arg 92 and Thr 64 blocking the hydrophobic groove into which substrate residues N-terminal of the isomerized proline generally fit (Figure 9c). While some cyclophilins with occluded active sites are still competent to bind substrates, bulky and occluding gatekeeper residues have been associated with a reduction or ablation of substrate binding. GeoCyp retains the hydrophobic active site characteristic
of cyclophilins, and a similar charge distribution as CypA, with a number of basic residues flanking the active site pockets (Figure 9c).

**GeoCyp is thermostabilized as compared to CypA**

To test the integrity of the GeoCyp structure with temperature, $^{15}$N-HSQC spectra were collected over a wide range of temperatures from 0°C to 40°C (Figure 10b). As indicated by the well-dispersed spectra, GeoCyp remains well folded over this temperature range. GeoCyp maintains its structure, with no significant alteration to its $^{15}$N-HSQC spectrum, for at least 1 week at 40 °C, in contrast to CypA that forms a visible precipitate within hours under the same conditions.

To further probe the stability of GeoCyp over a wider temperature range, both CypA and GeoCyp were subjected to circular dichroism (CD) thermal denaturation. While CypA exhibits a melting temperature ($T_m$) of 51.3°C, and is fully denatured by 55°C, the only measured alteration to GeoCyp’s secondary structure occurs with a substantially higher transition midpoint of 67.9°C (Figure 10c). Interestingly, even after this transition, GeoCyp maintains ~70% of the ellipticity of the fully folded form, suggesting that significant secondary structure remains present up to at least 95°C. These studies confirm that GeoCyp is thermostable.

**Differential conformational sampling between GeoCyp and CypA when bound to a peptide substrate**

We have previously reported a peptide substrate for cyclophilins, GSFGPDLRAGD, a slightly modified version of one identified by Piotukh et al. in a phage display screen, in which the G-P peptide bond is readily isomerized by both CypA and GeoCyp (see Table 2 for peptide NMR assignments). The peptide is representative of a large class of putative biological cyclophilin targets, and amino acid substitutions have shown that residues Phe 3 and Gly 4 are critical in mediating the cyclophilin-substrate binding interaction. To investigate the conformational landscape of GeoCyp during turnover of this
peptide, we docked the peptide to the solution structure using inter-Nuclear Overhauser Effect (NOE) distance restraints and generated an ensemble of chemical shift guided MD structures in the bound cis and bound trans forms, following a previously described method applied to CypA. The peptide binds in a similar mode to CypA, consistent with chemical shift perturbations observed upon addition of the substrate, which map to the canonical active site in both proteins (Figure 11). As shown in Figure 12a/b, in both CypA and GeoCyp the substrate proline localized to the hydrophobic S1 pocket, while Phe 3 of the peptide samples into and out of the S2 pocket in the peptide-bound ensembles. While the structure of the CypA S2 pocket remains relatively static throughout the ensemble, the loops creating the S2 pocket are highly dynamic in GeoCyp, permitting access to the pocket despite the occluded nature of the pocket in the unbound form. This mobility is reflected in the significantly higher root mean square deviation (RMSD) values of these loops in GeoCyp when compared to CypA (Figure 12c). In particular, Arg 92 in GeoCyp is highly mobile, acting both to clamp Phe 3 in place when in the pocket and forming \( \pi \)-cation interactions with the Phe 3 aromatic ring when out of the pocket (Figure 12a).

A gatekeeper residue that regulates substrate binding

Given the apparent role of S2 pocket adjacent residues in regulating substrate binding in the substrate bound GeoCyp ensembles, we sought to experimentally examine the role of Arg 92 (Ala 103 in CypA), as the bulkiest ‘gatekeeper’ residue present and because of the apparent \( \pi \)-cation interactions between the side-chain guanidinium group and the substrate phenol ring when Phe 3 shifts out of the S2 pocket. Homologous swap mutations were thus generated in each protein (CypA\textsuperscript{A103R} and GeoCyp\textsuperscript{R92A}). NMR titrations with the peptide substrate revealed that the mutations strengthened and weakened binding in CypA and GeoCyp, respectively, such that, in each case, the presence of arginine led to tighter binding (Table 3), consistent with clamping of the substrate and/or \( \pi \)-cation
interactions as identified in ensembles of GeoCyp. In addition to binding, catalytic efficiency was measured for both wild type and mutant proteins. Due to the reversible nature of the proline isomerization process, catalysis is measured via a ZZ-exchange NMR experiment wherein the substrate is $^{15}$N-labeled and a low, catalytic concentration (20 µM) of protein is added. The time-dependent appearance of cross-peaks indicates turnover and can be used to determine an effective isomerization rate ($k_{iso}$), which is a measurement of catalytic efficiency as previously described. This assay is not a measure of the catalytic rate on the enzyme that is not readily measurable, but instead is a measure of the isomerization rate at the particular catalytic concentration of the enzyme used. Nonetheless, ZZ-exchange is an effective way to compare the catalytic rates of cyclophilin-mediated isomerization. As shown in Table 3, CypA$^{A103R}$ and GeoCyp$^{R92A}$ catalyzed less and more efficiently than the wild-type enzymes, respectively, indicating that the increase in substrate affinity brought about by the presence of arginine in this position corresponds to a reduction in catalytic turnover.

To further probe this trade-off between substrate binding and catalytic turnover in cyclophilins, we sought to examine the impact of altering affinity in the context of a second mutation that is known to significantly reduce the background binding affinity. We therefore generated the R92A and A103R swap mutations in the context of mutation to the catalytic arginine (Arg 55 in CypA and Arg 47 in GeoCyp). In addition to significantly reducing binding affinity, mutation of the catalytic arginine alone also nearly, but not entirely, ablates catalysis (see cross-peaks in Figure 13a, only in the presence of enzyme). For these double mutants (CypA$^{R55A/A103R}$ and GeoCyp$^{R47A/R92A}$), both proteins still bind more tightly with an arginine than alanine at site GeoCyp-92/CypA-103, as evidenced by a larger change in the chemical shift of $^{15}$N-labeled peptide peaks upon addition of protein (Figure 13a). As shown in Figure 13b, however, catalytic turnover follows an opposite trend when the catalytic arginine mutation is present, with CypA$^{R55A/A103R}$ catalyzing more efficiently than CypA$^{R55A}$ and
GeoCyp<sup>R47A</sup> catalyzing more efficiently than GeoCyp<sup>R47A/R92A</sup>. Because significant line broadening occurs due to protein binding at the high concentrations needed to observe turnover in the context of the mutation to the catalytic arginine, accurate quantitation of $k_{iso}$ cannot be made. However, the ratio of cis peak intensity to cross-peak intensity is used as a proxy for rate of turnover, where a larger slope corresponds to faster turnover (Figure 13b). Collectively, these data provide a rationale for the trade-off between binding affinity and substrate turnover in cyclophilins, wherein binding must be sufficiently tight to engage the substrate, yet sufficiently weak to allow for substrate release after turnover.

Our hypothesis of evolutionary tuning of the S2 loop is further bolstered by comparison of cyclophilin protein sequences across the Bacillaceae family. As shown in Figure 14, among Bacillaceae cyclophilins annotated in the RefSeq database, only in the closely related thermophilic genera *Geobacillus* and *Anoxybacillus* is the homologous site occupied by an arginine. The site contains alanine, or occasionally threonine or serine, throughout the other 71 members of the family, as in human CypA. Combined with our functional mutagenesis data, these results indicate an evolutionarily tuned functional role for this S2 site in regulating substrate interactions. The one other thermophilic Bacillaceae cyclophilin annotated, from *Caldalkalibacillus thermarum*, does contain an alanine at this homologous site, indicating that acquisition of this arginine adaptation is not universal among all thermophiles.

**An active site electric field is conserved between CypA and GeoCyp**

As described above, we utilized chemical shift guided molecular dynamics methods for to generate an ensemble of GeoCyp structures bound to the cis or trans peptide isoforms. Recently, we demonstrated the existence of an electric field within the CypA active site in the -Z direction (defined as normal to the pyrrolidine ring of the isomerized proline) that acts, in a so-called ‘electrostatic handle’ mechanism, to facilitate isomerization. Specifically, this electric field exists in both the cis-bound and trans-bound states and
functions to reduce the energy of the $\omega=90^\circ$ transition state barrier by about 30 kJ/mol. To determine whether this catalytic mechanism is also conserved, we analyzed the ensembles of peptide-bound GeoCyp described above. As shown in Figure 15, GeoCyp exhibits a similar -Z electric field within the active site in both the cis and trans conformations, with a mean value of $\sim$30 MV/cm, compared to a mean value of $\sim$45 MV/cm previously identified in CypA. Relative to CypA, this weaker field in GeoCyp, along with the tighter binding affinity identified above, explains the reduction in catalytic efficiency (Table 3). These results suggest that GeoCyp and CypA utilize an evolutionarily conserved mechanism of isomerization that is fine-tuned with respect to binding via the S2 binding pocket.

**GeoCyp and CypA exhibit comparable temperature dependent activities**

Thermophilic enzymes are generally optimally functional at the source organism's optimal temperature, exhibiting a significant reduction in catalytic turnover at lower temperatures. As such, we measured the isomerization rate ($k_{iso}$) of both CypA and GeoCyp, using ZZ-exchange with $^{15}$N-labeled peptide substrate, over the range of temperature accessible by NMR spectroscopy. For the ZZ-exchange based catalytic experiment, $k_{iso}$ can only be accurately measured between about 0.5 and 30 s$^{-1}$. To remain within this range, two different enzyme concentrations were used: 20 $\mu$M for 0-20°C and 1 $\mu$M for 30-45°C, with 1 mM $^{15}$N-labeled peptide used in all cases. As shown in Figure 16, over the range of 0-37°C, GeoCyp consistently catalyzes isomerization at $\sim$70% the rate of CypA. Only at 45°C, above the physiologically optimal temperature for CypA and approaching its denaturation temperature of 51°C (Figure 10), does the increase in turnover with temperature slow for CypA, and continue for GeoCyp such that they catalyze turnover at a comparable rate. Unlike many previously studied thermophiles, therefore, GeoCyp does not exhibit a substantial impairment of function at lower temperatures relative to its mesophilic counterpart, CypA.
Dynamics are similar over multiple timescales between CypA and GeoCyp, with variability in temperature dependence and magnitude

Given the surprisingly high activity of GeoCyp at low temperatures, we decided to examine whether, like many other thermophilic proteins, GeoCyp is hyper-stabilized at low and moderate temperatures or, as might be predicted from our functional data, GeoCyp exhibits similar dynamics to CypA over these temperatures. NMR relaxation experiments were thus collected over a range of temperatures for both CypA and GeoCyp. As described in more detail below, by utilizing the Carr-Purcell-Meiboom-Gill Relaxation Dispersion (CPMG-RD) experiment, we found that GeoCyp exhibits weak self-association on the ms timescale, precluding direct analysis of these motions. To determine whether self-association would significantly impact measurement of faster timescale dynamics, we examined both fast timescale (ps-ns) dynamics via longitudinal ($R_1$) relaxation experiments and slower ($\mu$s) motions via $R_{20}$, the $R_{ex}$-independent component of transverse relaxation that was estimated via the CPMG-RD experiment collected with $\nu_{cpmg}$ of 1000 s$^{-1}$. As shown in Figure 20, we collected data at multiple concentrations and found minimal concentration dependent changes to either $R_{20}$ or $R_1$ values, indicating that the weak self-association predominantly impacts motions on the slower, millisecond timescale, and allowing comparison of CypA and GeoCyp dynamics over faster timescales such that we are able to determine to what degree dynamics are stabilized in GeoCyp. We attempted to determine order parameters ($S^2$) for both CypA and GeoCyp by measuring $R_1$, $R_2$, and $^1$H-$^{15}$N heteronuclear NOE relaxation values and applying Modelfree 4.15 via the FAST-Modelfree implementation. However, using either an isotropic or axially symmetric diffusion model, a large number of residues were unable to be fit by any set of model-free parameters for either CypA (34 of 144 residues unassigned) or GeoCyp (24 of 120 residues unassigned), indicating that the model-free formalism is unable to accurately describe the dynamics of these cyclophilins, perhaps due to the large number of residues exhibiting $\mu$s-
ms internal motions (nearly 50% in CypA\textsuperscript{173}). We have therefore directly compared R\textsubscript{1} and R\textsubscript{20} values between CypA and GeoCyp.

For CypA and GeoCyp, regions with elevated R\textsubscript{1} values localize well to homologous regions within the proteins (Figure 18), corresponding to the high degree of structural similarity (Figure 9).\textsuperscript{216} Namely, these regions include the two large active site loops (residues 60-80 and 100-110 in CypA and residues 57-76 and 89-99 in GeoCyp), as well as the β7-α2 loop (residues 135-137 in CypA and residues 124-126 in GeoCyp). For each protein, R\textsubscript{1} data were collected over a range of temperatures from 0-30°C. Fast timescale dynamics in CypA are largely unaffected by temperature over this range, aside from slight elevations in R\textsubscript{1} in the β7-α2 and α2-β8 loops. GeoCyp, however, exhibits significant temperature dependent changes in active site R\textsubscript{1} values over the same temperature range, including both large active site loops, suggesting that on the ps-ns timescale, GeoCyp may exhibit some of the low temperature dynamic dampening that has been identified in other thermophilic proteins relative to their mesophilic counterparts.\textsuperscript{168,189}

R\textsubscript{20} relaxation rates were determined for both GeoCyp and CypA over the same temperature ranges as for R\textsubscript{1} relaxation. As shown in Figure 19, localized variability in R\textsubscript{20} relaxation is largely temperature independent for both GeoCyp and CypA, indicating a lack of low temperature dynamic dampening of GeoCyp on the μs timescale. R\textsubscript{20} relaxation exhibits similar patterns throughout the two proteins, including the largest elevations mapping predominantly to loops within the active site, although with variability in the magnitudes. Notably, as with R\textsubscript{1} values, loop 89-99 in GeoCyp exhibits much higher R\textsubscript{20} values relative to the homologous loop in CypA, residues 100-110, consistent with the elevated flexibility identified in the molecular dynamics ensembles and required to access the occluded S2 binding pocket (Figure 9c, Figure 12c).
Interestingly, neither $R_1$ nor $R_{20}$ values were dramatically impacted upon addition of saturating concentrations of the peptide substrate in CypA or GeoCyp (Figure 18b and Figure 19b). The only significant changes seen are reductions in $R_{20}$ values in the active site loops of each protein, likely corresponding to a reduced mobility upon substrate binding.

**GeoCyp weakly self associates in the millisecond timescale**

The CPMG-RD experiment allows quantitative measurement of rates of motion in the slow $\mu$s to ms range (~100-5000 s$^{-1}$), which have previously been linked to enzymatic function in CypA.$^{172,217}$ As deuteration has been previously shown to drastically improve the quality of CPMG-RD experiments applied to CypA, we generated uniformly $^2$H$^{15}$N-labeled GeoCyp.$^{173}$ Previous studies have demonstrated that there is no dependence of CPMG-RD on protein concentration for CypA, indicating that no weak self-association is contributing to measured chemical exchange ($R_{ex}$). To test for self-association of GeoCyp, we measured $^{15}$N-CPMG-RD on GeoCyp over a concentration range of 0.5-2 mM and found that, unlike CypA, GeoCyp does exhibit a concentration dependent chemical exchange, indicating weak self-association (see Figure 20). $^{15}$N-CPMG-RD is a particularly sensitive means for monitoring weak self-association that is not readily observed in $^{15}$N-HSQC spectra due to relatively small chemical shift perturbations induced. However, the high solubility of GeoCyp allowed us to collect a $^{15}$N-HSQC at a concentration of 9 mM and pinpoint the chemical shift perturbations (see Figure 20). The majority of chemical shift changes between 0.5 mM and 9 mM mapped to the active site, suggesting that interactions between hydrophobic active site residues underlie this weak association. Thus, in the case of GeoCyp, this weak self-association precludes quantitative determination of self-association independent rates in the ms timescale.

**Summary**

In this study, we have combined both recently developed chemical shift-based methods and standard NMR relaxation experiments to perform a comprehensive
comparison of the structure, dynamics, and catalytic mechanism of a thermophilic/mesophilic pair of cyclophilins. GeoCyp adopts a typical cyclophilin fold, albeit with shortened loops commonly found in thermophilic proteins. Strikingly, despite the >20°C difference in optimal growth temperature of humans and G. kaustophilus and the 17°C increase in thermal melt transition of GeoCyp over CypA, catalytic activity is minimally reduced (~30%) for GeoCyp compared to CypA (Figure 16). This reduction in catalysis appears to be mediated predominantly through higher binding affinity, with an additional, independent contribution from a somewhat reduced active site electric field. These finding contrast with multiple other studies of thermophile/mesophile pairs, for which thermophilic protein activities are significantly impaired at lower temperatures, and become efficient only as the optimal organismal temperature is approached. Given cyclophilins’ roles in responding to cellular stress, including to heat and cold stress, perhaps maintenance of significant catalytic activity across a range of temperatures allows GeoCyp to respond efficiently to these stresses, especially given the temperature extremes that may be experienced in and around a deep-sea vent. Further investigation of other thermophilic cyclophilins may reveal whether this feature is unique to GeoCyp or common across thermophilic cyclophilins.

Studies here illustrate how cyclophilins have dynamically evolved to fine-tune mechanism. Specifically, structural analysis of GeoCyp in the free and substrate bound forms has revealed a loop in the ‘gatekeeper’ region of the protein that occludes the S2 binding pocket in the free enzyme, but is nonetheless sufficiently mobile to permit binding to the substrate. The homologous mutational analysis of Arg 92 in GeoCyp (Ala 103 in CypA) highlights the evolutionary trade-off that exists between substrate affinity and rates of turnover. A previous study\textsuperscript{16} examining a different cyclophilin substrate demonstrated that rates of substrate release are comparable to rates of enzyme-bound isomerization;
therefore, off-rates and isomerization rates both significantly impact \( k_{iso} \), such that a reduced off-rate, brought about by an arginine at residue GeoCyp-92/CypA-103, likewise reduces \( k_{iso} \). However, under conditions of significantly weakened binding brought about by mutation of Arg 47 in GeoCyp (Arg 55 in CypA), the rapid rate of substrate release becomes a limiting factor in isomerization, such that the increased binding affinity now increases \( k_{iso} \). These data appear to be consistent with the optimum temperatures at which the mesophilic/thermophilic cyclophilins function; CypA binding affinity for the peptide substrate is highly temperature dependent, indicating a significant entropic cost to binding. Thus, at the lower temperatures under which CypA operates, Ala 103 maintains sufficiently weak binding to allow efficient catalysis. In contrast, as the entropic cost of binding increases with the higher temperatures under which GeoCyp exists, the tighter binding that is mediated by loop clamping by Arg 92 functions to increase catalytic efficiency.

Numerous studies have focused on the interplay between protein dynamics, stability, and enzymatic function among thermophiles. In addition to maintaining stability at high temperatures, thermophilic proteins face the related challenge of retaining sufficient dynamic mobility to function efficiently. Given this balance, many studies have promoted the ‘corresponding state hypothesis’ wherein evolution has tuned the dynamics of proteins to render all members of a given family equally dynamic at their source organisms’ optimal temperatures, rendering them hyper-stabilized at lower temperatures and unstable at higher temperatures.\(^{190-192,208,209,211}\) Other studies, however, have refuted this notion, demonstrating cases for which thermophilic proteins exhibit comparable or elevated dynamics relative to a mesophile.\(^{212,218-220}\) The contradictions in these studies may reflect variability in the mechanisms of stabilization between different protein families, but are also likely influenced by the timescale of dynamics that are observed, which can vary substantially depending on the technique utilized.\(^{218}\) Given the unusually high activity of GeoCyp at low temperatures relative to CypA (Figure 16) when compared to other thermophile/mesophile pairs, analyses
of GeoCyp dynamics are unable to contribute directly to the debate over the corresponding state hypothesis; without substantial impairment of function at lower temperatures, the hypothesis makes no prediction as to how motions should be affected. However, the temperature-dependent increases in deviations of $R_1$ relaxation observed in GeoCyp but not CypA (Figure 18) suggest that comparable fast (ps-ns) timescale mobility exists in both proteins at lower temperatures, and that these motions increase more dramatically in GeoCyp with temperature. These findings are in line with several computational and experimental studies of other thermophiles, which found more fast timescale mobility at multiple temperatures when compared to mesophiles.\textsuperscript{218-221} Among these proteins, it appears that a more dynamic folded form of the protein contributes to a reduced entropic folding penalty, especially as temperature increases, such that more flexibility may actually stabilize the folded form.

In conclusion, while some dynamic variability exists between CypA and GeoCyp which may contribute to GeoCyp stability (elevated fast-timescale dynamics) and function (magnitudes of loop motions around the S2 pocket) the analyses presented here of CypA and GeoCyp largely highlight the high degree of structural, dynamic, and mechanistic conservation in the cyclophilin family.
Table 1. Structural statistics for the GeoCyp RASREC Rosetta structures

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of residues</td>
<td>146</td>
</tr>
<tr>
<td>NOE-based distance restraints</td>
<td></td>
</tr>
<tr>
<td>NOE distance restraints (violations ≥0.5 Å)</td>
<td>1687 (146±10)</td>
</tr>
<tr>
<td>Intra-residue</td>
<td>387</td>
</tr>
<tr>
<td>Inter-residue</td>
<td>1300</td>
</tr>
<tr>
<td>Sequential (</td>
<td>i – j</td>
</tr>
<tr>
<td>Medium-range (</td>
<td>i – j</td>
</tr>
<tr>
<td>Long-range (</td>
<td>i – j</td>
</tr>
<tr>
<td>Other restraints</td>
<td></td>
</tr>
<tr>
<td>ϕ+ψ dihedral-angle restraints (violations ≥5°)</td>
<td>224 (29±3.6)</td>
</tr>
<tr>
<td>Average RMSD to the average structure</td>
<td></td>
</tr>
<tr>
<td>Backbone (Å)</td>
<td>0.72 ± 0.14</td>
</tr>
<tr>
<td>Heavy atom (Å)</td>
<td>1.2 ± 0.14</td>
</tr>
<tr>
<td>Ramachandran plot summary</td>
<td></td>
</tr>
<tr>
<td>Most favored regions</td>
<td>84.5%</td>
</tr>
<tr>
<td>Allowed regions</td>
<td>14.8%</td>
</tr>
<tr>
<td>Generally allowed regions</td>
<td>0.6%</td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0.1%</td>
</tr>
<tr>
<td>Deviations from idealize geometry</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.011</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Statistics are given for the twenty best scored structures

*Torsion-angle restraints were derived from TALOS+
*RMSDs calculated using iCing server
*Analysis performed using PROCHECK
Table 2. GSFGPDLRAGD model peptide chemical shift assignments

<table>
<thead>
<tr>
<th>Residue</th>
<th>$\delta H^N$</th>
<th>$\delta N$</th>
<th>$\delta C\alpha$</th>
<th>$\delta C\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly 1</td>
<td></td>
<td>43.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser 2</td>
<td></td>
<td>58.45</td>
<td>63.82</td>
<td></td>
</tr>
<tr>
<td>Phe 3\text{trans}</td>
<td>8.40</td>
<td>121.49</td>
<td>39.84</td>
<td></td>
</tr>
<tr>
<td>Phe 3\text{cis}</td>
<td>8.38</td>
<td>121.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly 4\text{trans}</td>
<td>8.15</td>
<td>110.67</td>
<td>44.61</td>
<td></td>
</tr>
<tr>
<td>Gly 4\text{cis}</td>
<td>8.19</td>
<td>110.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro 5\text{trans}</td>
<td></td>
<td>63.26</td>
<td>32.16</td>
<td></td>
</tr>
<tr>
<td>Pro 5\text{cis}</td>
<td></td>
<td>34.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp 6\text{trans}</td>
<td>8.41</td>
<td>120.09</td>
<td>54.22</td>
<td>40.94</td>
</tr>
<tr>
<td>Asp 6\text{cis}</td>
<td>8.59</td>
<td>121.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu 7\text{trans}</td>
<td>8.18</td>
<td>123.42</td>
<td>55.33</td>
<td>42.13</td>
</tr>
<tr>
<td>Leu 7\text{cis}</td>
<td>8.35</td>
<td>123.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg 8\text{trans}</td>
<td>8.31</td>
<td>121.96</td>
<td>55.82</td>
<td>30.75</td>
</tr>
<tr>
<td>Arg 8\text{cis}</td>
<td>8.37</td>
<td>122.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala 9</td>
<td>8.31</td>
<td>125.69</td>
<td>52.64</td>
<td>19.25</td>
</tr>
<tr>
<td>Gly 10</td>
<td>8.28</td>
<td>109.05</td>
<td>45.17</td>
<td></td>
</tr>
<tr>
<td>Asp 11</td>
<td>7.88</td>
<td>126.07</td>
<td>55.87</td>
<td>42.16</td>
</tr>
</tbody>
</table>

Table 3. Binding affinities and catalytic efficiencies for CypA, GeoCyp, and mutants

<table>
<thead>
<tr>
<th>Protein</th>
<th>$K_D$ (µM)</th>
<th>$K_{iso}$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CypA$^{WT}$</td>
<td>76 ± 3$^a$</td>
<td>10.0 ± 0.5</td>
</tr>
<tr>
<td>CypA$^{A103R}$</td>
<td>42 ± 3</td>
<td>7.6 ± 0.3</td>
</tr>
<tr>
<td>GeoCyp$^{WT}$</td>
<td>39 ± 2</td>
<td>4.7 ± 0.2</td>
</tr>
<tr>
<td>GeoCyp$^{R92A}$</td>
<td>118 ± 2</td>
<td>7.1 ± 0.3</td>
</tr>
</tbody>
</table>

$^a$Errors are fit errors from a single experiment
Figure 8. The secondary structure elements of GeoCyp are similar to those of CypA. Comparisons between GeoCyp (red) and CypA (black) are shown using a sequence alignment (Needleman-Wunsch algorithm). RCI values based on the chemical shifts are shown above as β-strands (arrows) and α-helices (barrels) along with the Cα shift deviations from random coil.
Figure 9. The structure of GeoCyp is similar to CypA, with some changes in substrate interface. (a) Solution structure of GeoCyp. The 20 lowest energy structures are shown, with β-sheets colored blue and helices colored red. (b) Superimposed structures of CypA (from previously generated molecular dynamics ensemble, in white) and GeoCyp (lowest energy structure, in blue). Three short loop deletions from CypA are colored green, while the large α2-β8 loop deletion is in red. (c) Surface representations of CypA and GeoCyp free structures. Structures are colored according to surface charge, with blue indicating basic charge and red indicating acidic charge. S1 and S2 binding pockets are indicated, including the occluded S2 pocket of GeoCyp.
Figure 10. GeoCyp is stable over a broad range of temperatures (a) $^{15}$N-HSQC spectra of $^{15}$N-labeled GeoCyp show that the enzyme is fully folded over a broad range of temperatures (0°C (navy), 10°C (sky blue), 20°C (green), 30°C (orange), and 40°C (red)). For clarity, only 0°C, 20°C and 40°C plots are displayed for the full spectra. The inset shows several residues exhibiting chemical shift changes, which includes the catalytic Arg 47. (b) CD thermal denaturation curves for CypA (red) and GeoCyp (blue) measured at 222 nm. Solid lines are best-fit curves to the data, with midpoints represented by vertical dashed lines. (c) CD trace for CypA (red) and GeoCyp (blue) in the native state (10°C) and of GeoCyp (purple) after partial unfolding transition (95°C).
Figure 11. CypA and GeoCyp binding interfaces The peptide substrate was titrated into $^{15}$N-labeled CypA or GeoCyp. Residues for which the normalized chemical shift changes ($\sqrt{(5\Delta\delta H)^2 + (\Delta\delta N)^2}$) are greater than 0.25 ppm are colored in red.
Figure 12. Peptide bound MD ensembles of CypA and GeoCyp (a) Representative structures, with the peptide in the \textit{trans} conformation, from the molecular dynamics ensemble of GeoCyp bound to the model peptide. Within the ensemble, Phe 3 of the peptide samples the S2 pocket (top structure) and also exits the S2 pocket (bottom structure). Residues Arg 92 (blue) and Asp 66 (red) serve to clamp Phe 3 of the peptide into the pocket, or partially occlude the pocket when Phe 3 is not in it. (b) Representative structures, with the peptide in the \textit{trans} conformation, from the previously generated molecular dynamics ensemble of CypA bound to the model peptide. Phe 3 of the peptide also samples in (top structure) and out (bottom structure) of the S2 pocket within the CypA ensemble, but the S2 pocket structure remains fully open, and no clamping occurs. Residues Ala 103 and Gly 75, homologous to Arg 92 and Asp 66 in GeoCyp, are depicted in green. (c) Root mean square fluctuation (RMSF) values for the backbone residues of CypA and GeoCyp in chemical shift guided molecular dynamics simulations of the proteins bound to the peptide substrate in the \textit{cis} (red) and \textit{trans} (blue) conformations. GeoCyp residue numbers are shifted to match CypA, such that homologous residues are in line with one another.
**Figure 13. Homologous mutations in the context of mutation to the catalytic arginine**

ZZ-exchange spectra for the peptide substrate alone (black), with 100 µM of protein with mutation to the ‘catalytic’ residue CypA$^{R55A}$ or GeoCyp$^{R47A}$ (red), and with 100 µM of protein with mutation to both the ‘catalytic’ residue and swap mutant CypA$^{R55A/A103R}$ or GeoCyp$^{R47A/R92A}$ (blue). All spectra were collected at 30°C, with a mixing time of 0.768 s. For both CypA and GeoCyp, arginine in the swap mutation position leads to tighter binding, as evidenced by larger chemical shift changes of the Leu 6 cis peak and cross peak. **(b)** In the context of an alanine mutation to the catalytic arginine, tighter binding corresponds to faster turnover. Because of significant line broadening due to protein binding at the high concentrations needed to observe turnover in the context of mutation to the catalytic arginine, accurate determinations of $k_{iso}$ cannot be made, so the ratio of cis peak intensity to cross peak intensity is used as a proxy for rate of turnover, where a larger slope corresponds to faster turnover.
Figure 14. Sequence alignment of all annotated cyclophilins from *Bacillaceae* family members in the RefSeq database. *Percent identity to GeoCyp, calculated for entire cyclophilin sequence.* *Thermophilic.*
Figure 15. GeoCyp exhibits a -Z electric field within the active site Population distribution of the Z electrostatic fields in GeoCyp in the *cis* (black) and *trans* (gray) peptide conformations, where Z is defined as normal to the pyrrolidine ring of Pro 5 in the peptide substrate.

Figure 16. Catalytic efficiency for CypA and GeoCyp from 0-45°C 1 mM $^{15}$N-labeled peptide and 20 µM (0-20°C) or 1 µM (30-45°C) unlabeled protein were used. Error bars represent fit errors within a single experiment.
Figure 17. For GeoCyp, $R_{\text{ex}}$ values, but not $R_{20}$ or $R_{1}$ values, exhibit a large concentration dependence. (a) $R_{\text{ex}}$ values for GeoCyp at 0.5 mM (blue), 1 mM (red), and 2 mM (black), illustrating the global concentration dependence of exchange induced relaxation. (b) $R_{20}$ values for GeoCyp at 0.5 mM (blue), 1 mM (red), and 2 mM (black), illustrating the relatively small impact of self-association on faster (µs) timescale motions, measured via $R_{20}$. (c) $R_{1}$ values for GeoCyp at 0.5 mM (black) and two replicates at 0.25 mM (red and blue), illustrating the error in $R_{1}$ measurements and the lack of significant concentration dependence in the measurements. For $R_{20}$ and $R_{1}$, baseline adjustments were made to account for changes in baseline relaxation due to viscosity changes with protein concentration.
Figure 18. R\textsubscript{1} relaxation rates for CypA and GeoCyp (a) R\textsubscript{1} relaxation rates for CypA and GeoCyp at 0°C (black), 10°C (blue), 20°C (green), and 30°C (red). Baseline values were subtracted off of each data set to normalize values between temperatures, facilitating comparison of dynamic variations. (b) R\textsubscript{1} relaxation rates for CypA and GeoCyp at 10°C alone (blue) and in the presence of saturating concentrations (4mM) of the peptide substrate (red). For all plots, dots represent individual measurements and lines depict a five residue moving average. GeoCyp residue numbers are shifted to match CypA, such that homologous residues are in line with one another.
Figure 19. \( R_{20} \) relaxation rates for CypA and GeoCyp (a) \( R_{20} \) relaxation rates for CypA and GeoCyp at 0°C (black), 10°C (blue), 20°C (green), and 30°C (red). Baseline values were subtracted off of each data set to normalize values between temperatures, facilitating comparison of dynamic variations. (b) \( R_{20} \) relaxation rates for CypA and GeoCyp at 10°C alone (blue) and in the presence of saturating concentrations (4mM) of the peptide substrate (red). For all plots, dots represent individual measurements and the lines depict a five residue moving average. GeoCyp residue numbers are shifted to match CypA, such that homologous residues are in line with one another.
Figure 20. GeoCyp self-associates on the millisecond time scale (a) CPMG-RD data were collected at 21.1 T with concentrations of 0.5 mM (black dots), 1 mM (red dots), or 2 mM (blue dots) $^2$H$^{15}$N-labeled protein. Least-squares fits to the Carver-Richards equations of the 21.1 T data alone are represented as dotted lines. Baseline $R_{20}$ values are subtracted from each data set so that $R_{ex}$ values can be compared directly. For GeoCyp, but not CypA, $R_{ex}$ exhibits a clear concentration dependence, indicating a weak self-association on the millisecond time scale and precluding analysis of these data to extract exchange parameters. (b) For GeoCyp, no apparent chemical shift changes were observed by collecting $^{15}$N-HSQC spectra at 0.5 mM (black) and 1 mM (red) of unbound protein, indicating that collecting full CPMG-RD experiments are necessary to rule out the potential impact of weak self-association. (c) For GeoCyp, $^{15}$N-HSQC spectra were collected on $^{15}$N-labeled protein at concentrations of 0.5 mM (black) and 9 mM (red). Only by collecting data at this high protein concentration are significant chemical shift changes apparent in the $^{15}$N-HSQC spectra. CypA data were collected at 10 °C, GeoCyp data were collected at were collected at 30 °C, due to higher $R_{ex}$ values at this temperature, facilitating comparison of exchange at multiple concentrations.
CHAPTER III
DETERMINATION OF THE FULL CATALYTIC CYCLE AMONG MULTIPLE
CYCLOPHILIN FAMILY MEMBERS²

Introduction

With the exception of proline, the N-terminal peptide bonds of the other 19 common amino acids exist almost exclusively in the trans population in both unstructured peptides and in the context of proteins. However, X-Pro peptide bonds in free peptides, where X is any other amino acid, adopt the cis conformation ~10-30% of the time, depending predominantly on the identity of X. In the context of a folded protein, X-Pro bonds adopt the cis conformation ~3-10% of the time, and are generally locked into a single conformation in the context of a given protein structure.²²³ The inherent isomerization of the peptidyl-prolyl bond occurs with a rate constant on the order of $10^{-2}$ s⁻¹, while cyclophilins and other PPIases increase the rate of isomerization by ~5 orders of magnitude, facilitating proper protein folding and other isomer specific outcomes.⁹,¹⁷,¹⁸

Despite the diversity in cellular localization and biological roles of cyclophilins, few studies have directly compared enzymatic function across multiple members of the family or the degree to which the enzymatic cycles are conserved among them. Multiple human cyclophilins have been previously compared with respect to their binding affinity for the cyclic peptide inhibitor cyclosporine A (CsA) and qualitatively compared with respect to their catalytic activity toward a weakly binding model tetrapeptide substrate.⁵ However, we sought here to characterize the full enzymatic cycle among multiple cyclophilins as they catalyze a biologically representative peptide substrate. Because prolyl cis↔trans interconversion is a reversible process and both isoforms are significantly populated at equilibrium, direct

determination of the microscopic rate constants via measurement of substrate depletion or product formation is not possible. Measurement of the unidirectional cis-to-trans interconversion of isomerases can be achieved through a chymotrypsin-coupled assay, although this approach has significant limitations that have been previously outlined, including severe restrictions on the substrate, a low signal-to-noise ratio, and protease degradation of the enzyme. Alternatively, NMR lineshape analysis provides a powerful means of determining microscopic rate constants on arbitrary substrates at equilibrium. NMR lineshapes are exquisitely sensitive to exchange processes that occur on the timescale of chemical shift differences (~10’s-1000’s s\(^{-1}\)); when lineshapes are measured at multiple concentrations of an enzyme and substrate and combined with additional functional constraints, the microscopic rate constants defining the catalytic cycle can be determined. Specific to PPIases, lineshape analysis has been previously employed to characterize the microscopic rate constants defining CypA catalysis of a non-native modified tetrapeptide and for another structurally unrelated PPIase, Pin1, to a biologically relevant peptide substrate. Thus, here we have functionally characterized and applied lineshape analysis to three human cyclophilins, CypA, CypB, and CypC, as well as to the distantly related thermophilic cyclophilin GeoCyp.

Results

Functional characterization of multiple members of the cyclophilin family

While the enzymatic activities of multiple human cyclophilins have been qualitatively compared toward a short tetrapeptide, we sought here to quantitatively compare both binding affinities and catalytic activity toward a biologically representative substrate that binds tighter to multiple members of the family than previously used model peptide substrates. Specifically, we utilized a previously characterized and solution resonance-assigned peptide substrate, GSFGPDLRAGD, initially derived from a phage display screen by Piotukh et al., which was shown to be representative of multiple putative
cyclophilin targets.\textsuperscript{33} We have previously demonstrated that CypA and GeoCyp both bind to and catalyze isomerization of this peptide (Table 3).\textsuperscript{188} Here, we have also characterized the activities of two additional human cyclophilins, CypB and CypC, toward this substrate. The chemical shifts of CypA,\textsuperscript{196} CypB,\textsuperscript{226} and GeoCyp\textsuperscript{187} have been previously published, but not those for CypC. We therefore also determined the backbone chemical shift assignments for CypC (BMRB entry 25341).

The GSFGPDLRAGD peptide substrate exists with Pro 5 in both the cis and trans conformations, but titration of the substrate into any of the cyclophilins studied here results in a single set of bound peaks (Figure 21a), indicating that the enzyme-substrate interaction remains in the fast exchange regime once near-saturating concentrations are reached. Titration of the substrate allows for the determination of an apparent dissociation constant ($K_{D\text{-app}}$), which is influenced by binding to both the cis and trans isoforms. As shown in Table 4, all of the cyclophilins studied here bind to the peptide with affinities in the micromolar range, with an ~2.5-fold range in binding affinity among the human cyclophilins.

Because the peptidyl-prolyl isomerization catalyzed by cyclophilins is a reversible reaction, the catalytic efficiency of each enzyme was determined by addition of a low, catalytic enzyme concentration (20\textmu{}M) to 1 mM $^{15}$N-labeled peptide substrate, followed by collection of a ZZ-exchange experiment as previously described.\textsuperscript{188} Multiple residues exhibit distinct cis and trans amide chemical shifts, including two residues (Asp 6 and Leu 7) for which the cis, trans, and cross-peaks are resolvable (one cross-peak for Asp 6 is overlapped with another residue peak). The time dependent disappearance of the cis and trans peaks and appearance of cross-peaks can be used to determine an isomerization rate ($k_{iso}$) between the two free isoforms under the given conditions.\textsuperscript{118} $k_{iso}$ does not report directly on the catalytic turnover on the enzyme, which will be addressed below, but is dependent on both substrate and enzyme concentrations and is influenced by cis and trans binding affinities as well as turnover on the enzyme. As shown in Table 4 and Figure 21b,
each of the cyclophilins studied here catalyzes isomerization of the substrate. Notably, the human cyclophilins catalyze isomerization with nearly identical $k_{iso}$ values, despite the much larger variability in substrate affinity, suggesting that the microscopic rate constants have been evolutionarily “fine-tuned” to maintain similar catalytic efficiencies.

**Structural and dynamic comparison of multiple human cyclophilins**

Previous studies have implicated both structural and dynamic factors in regulating cyclophilin function and in explaining functional differences within the family.\textsuperscript{5,188,217} Given the variable binding affinities of the three human cyclophilins for the model peptide and the more similar rates of turnover, we sought to determine whether the functional differences were dictated primarily through structure or dynamics. The structures of each of the three human cyclophilins have been previously determined by X-ray crystallography, although only for CypC in the presence of CsA. As such, we overlaid the structures of CypA, CypB, and CypC, each determined as bound to CsA. CypA is 64% and 54% identical to CypA for CypB and CypC, respectively. As shown in Figure 22a, the crystallized backbone conformations of the three proteins are nearly identical (backbone RMSDs of 1.40 Å and 1.06 Å with respect to CypA for CypB and CypC, respectively), with the only notable variability occurring in the orientation of the $\alpha$2-$\beta$8 loop. A comparison of the sidechains of the highly conserved active site and “gatekeeper” residues likewise reveals nearly identical conformations.\textsuperscript{5} While these structures were all determined in the presence of CsA, the X-ray structure determined for CypA alone reveals minimal perturbations relative to CypA’s structure when bound to CsA, suggesting that the protein ground state is not altered significantly upon CsA binding (Figure 23).

Given the high degree of structural conservation among these enzymes, we next measured dynamics over multiple timescales, as conformational flexibility may have a role in influencing function. As described in Chapter II, we have been unable to fit data collected on CypA to the model-free formalism, possibly because of the large number of residues
exhibiting µs-ms motions. We therefore determined $^{15}$N-longitudinal ($R_1$) relaxation rates for each of the three cyclophilins as reflective of fast timescale motions (Figure 22b). $R_1$ rates report on the rigidity of the backbone and are predominantly reflective of the secondary structure of the protein. Therefore, it is unsurprising that the pattern of $R_1$ rates is largely conserved among the three proteins, with the only major variability being localized to the highly dynamic loop comprising residues 65-73 (all residue numbering is given for CypA).

Slower (µs-ms) timescale motions have been previously reported as playing a role in the CypA catalytic cycle and can be measured quantitatively by CPMG-RD. However, we found that CypB weakly self associates on the slow µs-ms timescale, as demonstrated by a concentration dependence of exchange-induced relaxation (Figure 24) and as we have previously reported for GeoCyp (Figure 17). We have instead estimated the $R_{ex}$-independent component of $R_2$ relaxation ($R_{20}$) by measurement of CPMG-RD with the highest imparted $\nu_{cpmg}$ of 1000 s$^{-1}$. As shown in Figure 22c, the patterns of baseline deviations are largely similar among the three proteins. Notably, variability exists in the magnitude of $R_{20}$ deviations in both of the two large active site loops (residues 65-80 and 101-110 for CypA), with larger values for CypC and reduced values for CypB, relative to those of CypA. Structural variability in these regions has been previously suggested to impact substrate affinity and specificity and we have shown for GeoCyp that specific side chain differences and dynamic alterations in this region underlie its altered function relative to that of CypA.

CypA and CypC do not self-associate on the µs-ms timescale (Figure 24 and Eisenmesser et. al), so we compared µs-ms timescale motions between them by collecting CPMG-RD data on $^2$H$^{15}$N-labeled protein. Both enzymes exhibit mobility in and around the active site on this timescale; however, the localization and rates of motion are altered significantly (Figure 25a/b). Generally, inherent measured rates of motions in CypC are increased relative to CypA and localized predominantly to one side of the protein. In
combination with the highly conserved crystal structure and the variability in R20 rates, these data suggest that varied conformational sampling about a highly conserved ground state underlies the observed functional variability among the human cyclophilins studied here.

*Elucidation of the complete catalytic cycle for multiple cyclophilins utilizing lineshape analysis*

To more thoroughly determine the underlying mechanism of altered functionality among the cyclophilin family, we decided to characterize the enzymatic cycle of each of the human cyclophilins studied here, as well as the distantly related thermophilic cyclophilin, GeoCyp, using lineshape analysis. The minimal reaction model of cyclophilin isomerization is depicted in Figure 26 and consists of 8 microscopic rate constants. The uncatalyzed interconversion of the peptide substrate (described by kAD and kDA) occurs at a rate many orders of magnitude slower than that of the catalyzed reaction,188 and is therefore irrelevant on the NMR timescales utilized herein to experimentally determine the rate constants. We have therefore set this interconversion rate to zero for the remaining calculations. As demonstrated in previous studies utilizing lineshape analysis,16,224 the information content of the NMR lineshapes alone is generally insufficient to effectively constrain the parameters given in Figure 26; therefore, as in previous studies, we have included additional, independently determined constraints that are described below and allow for determination of the microscopic rate constants describing cyclophilin catalysis.

We collected high-resolution 15N-HSQC spectra for each of the cyclophilins with multiple 15N-labeled peptide (0.5-1 mM) and unlabeled enzyme (5-100 µM) concentrations. To minimize perturbations to the lineshape imposed by truncation of the free induction decay (FID) data or windowing, we collected the full FID by acquiring for 2.5 seconds after completion of the pulse sequence. This long acquisition time precludes application of decoupling during data application, preventing averaging of JNH and JHHC scalar coupling interactions and resulting in four peaks per residue per isoform. As shown in Figure 27a,
even with these coupling interactions, for residues Asp 6 and Leu 7, all cis and trans peaks are fully resolved and therefore have been utilized for lineshape analysis. Each of the sets of peaks was excised from the 2-D spectrum as shown in Figure 27a and averaged over the nitrogen dimension to eliminate effects of line broadening in the nitrogen dimension, yielding a series of 1-D traces in proton. Data were also normalized such that the total area under the 1-D trace for each cyclophilin/substrate concentration is set to be equal for all conditions for both measured data and calculated lineshapes, eliminating effects from variability in experiment-to-experiment signal intensity.

For a given set of microscopic rate constants, the equilibrium concentrations of states A-D (Figure 26) were determined by solving the set of coupled equations:

Equation 8

\[
\begin{align*}
[A][E_{\text{free}}]k_{AB} - [B]k_{BA} &= 0 \\
[B]k_{BC} - [C]k_{CB} &= 0 \\
[B]k_{CD} - [D][E_{\text{free}}]k_{DC} &= 0 \\
[B] + [C] + [E_{\text{free}}] &= [E_{\text{tot}}] \\
[A] + [B] + [C] + [D] &= [S_{\text{tot}}]
\end{align*}
\]

using the Levenberg–Marquardt algorithm (LMA) for solving non-linear least squares systems, where \([S_{\text{tot}}]\) is the total substrate concentration, \([E_{\text{tot}}]\) is the total enzyme concentration, and \([E_{\text{free}}]\) is the concentration of unbound enzyme. The evolution of magnetization \(M\) for a given residue is described by:

Equation 9

\[
\frac{dM}{dt} = KM
\]
where $M$ is the magnetization of states A-D:

$$M = \begin{bmatrix} M_A \\ M_B \\ M_C \\ M_D \end{bmatrix}$$

and $K$ is the evolution matrix:

$$K = \begin{bmatrix}
2\pi i(\omega_A + J) - R_{2A} k_{AB}[E_{free}] & k_{BA} & 0 & 0 \\
k_{AB}[E_{free}] & 2\pi i(\omega_B + J) - R_{2B} k_{BB} - k_{BC} & k_{CB} & 0 \\
0 & k_{BC} & 2\pi i(\omega_C + J) - R_{2C} k_{CB} - k_{CD} & k_{CD} \\
0 & 0 & k_{CD} & 2\pi i(\omega_D + J) - R_{2D} k_{DC}[E_{free}] 
\end{bmatrix}$$

where $\omega_{A-D}$ are the chemical shift of states A-D, $R_{2A-D}$ are the proton $R_2$ relaxation rates of states A-D, and $[E_{free}]$ is the concentration of unbound enzyme. $J$ is the scalar coupling constant, given by:

$$J = \pm J_{NH} \pm J_{HHa}$$

where $J_{NH}$ is the coupling between the amide proton and nitrogen, and $J_{HHa}$ is the coupling between the amide proton and the $C_\alpha$ proton. The free induction decay (FID) of the system is given by:

Equation 10

$$FID = V^{-1}M_0e^{\lambda t}V$$

where $V$ is the matrix of eigenvectors of $K$, $\lambda$ is the eigenvalues of $K$, and $M_0$ consists of the equilibrium concentrations of states A-D determined from equation 2:

$$M_0 = \begin{bmatrix} P_A \\ P_B \\ P_C \\ P_D \end{bmatrix}$$

FID, given in Equation 10, consists of four decaying exponential vectors describing each of states A-D. The frequency domain spectra were generated by taking the real component of the Fourier transform of the appropriate component (i.e., state A, free cis or state D, free trans) of the fid, which were then fit to lineshapes extracted from 2-D $^{15}$N-HSQC spectra as in Figure 27a. Spectra were generated for the free cis and free trans states of
each residue used for each concentration of protein and peptide. Fits were performed using LMA, allowing each of the 6 microscopic rate constants and the bound chemical shift values to vary, while fitting to the measured lineshapes (collected at 10°C on Varian 900 MHz spectrometer with a cryogenically cooled probe), the apparent $K_D$ ($K_{D-app}$), the effective isomerization rate ($k_{iso}$), and the chemical shift of each residue when saturated with each cyclophilin. Proton $R_2$ values of the bound form were directly measured on $^2$H$^{15}$N CypA in the free form and found to be $\sim40$ s$^{-1}$, variation of this value by $\pm10$ s$^{-1}$ had minimal effect on the solution for CypA, indicating an insensitivity of the solution to the particular bound $R_2$ value used (Table 5). $R_{2B}$ and $R_{2C}$ were therefore fixed at 40 s$^{-1}$ for all cyclophilins. At pH 6.5, at which all data were collected, the rate of amide exchange with solvent in an unstructured peptide is $\sim2$ s$^{-1}$, much lower than rates determined via lineshape analysis.$^{228}$ Additionally, because any exchange with water would be with a non-coherent proton, the net effect of solvent exchange would be a small increase in background $R_2$ (Equation 10, $K$ diagonal); because $R_2$ values are determined independently for each isoform of each residue, any variability in solvent accessibility would be accounted for by variability in these calculated $R_2$ values, precluding solvent exchange from impacting rate the rate constants determined here. $K_{D-app}$ was calculated by$^{224}$:

\[ K_{D-app} = \frac{1 + K_{eq}}{\frac{K_{eq}}{K_{D-trans}} + \frac{1}{K_{D-cis}}} \]

The catalytic efficiency ($k_{iso}$) given by a single set of microscopic rate constants was determined by generating artificial evolution matrices in which magnetization may only flow from cis $\rightarrow$ trans or trans $\rightarrow$ cis:
Equation 12

\[
\mathbf{M}_{\text{cis} \rightarrow \text{trans}} = \begin{bmatrix}
0 & k_{BA} & 0 & 0 \\
0 & -k_{BA} - k_{BC} & k_{CB} & 0 \\
0 & k_{BC} & -k_{CB} - k_{CD} & k_{DC}[E_{\text{free}}] \\
0 & 0 & k_{CD} & -k_{DC}[E_{\text{free}}]
\end{bmatrix}
\]

\[
\mathbf{M}_{\text{trans} \rightarrow \text{cis}} = \begin{bmatrix}
k_{AB}[E_{\text{free}}] & k_{BA} & 0 & 0 \\
-k_{AB}[E_{\text{free}}] & -k_{BA} - k_{BC} & k_{CB} & 0 \\
0 & k_{BC} & -k_{CB} - k_{CD} & 0 \\
0 & 0 & k_{CD} & 0
\end{bmatrix}
\]

Calculation of eigenvalues from these matrices yields two numbers of large magnitude that describe the rapid equilibration of the bound forms and a number of small magnitude, which describes the flux away from the cis or trans states (i.e., \(k_{\text{cis} \rightarrow \text{trans}}\) or \(k_{\text{trans} \rightarrow \text{cis}}\)). The catalytic efficiency is then given by:

Equation 13

\[
k_{\text{iso}} = k_{\text{cis} \rightarrow \text{trans}} + k_{\text{trans} \rightarrow \text{cis}}
\]

Data collected on the peptide alone were fit first, permitting determination of transverse relaxation, scalar coupling constants, and free peptide chemical shifts to be utilized for fits once enzyme was added. The free variables were then simultaneously fit to the measured lineshape data over the multiple enzyme and substrate concentrations as described above, allowing each of the microscopic rate constants to vary along with the chemical shifts of the bound cis and bound trans forms; additional restraints were imposed to fit the experimentally measured \(K_{D\text{-app}}\) and \(k_{\text{iso}}\) values described above (Table 4), as well as the averaged proton peak position in the bound form, which was determined by collection of \(^{15}\text{N}\)-HSQC spectra in the presence of saturating concentrations of each of the cyclophilins. To fully search the parameter space, 200 fits were performed for each cyclophilin while the initial conditions were randomly varied. The best 20 fits, as measured by the coefficient of determination, were then averaged as representative of the best-fit solution and are listed in Table 6.
To estimate the extent to which experimental variability may impact determined rate constants, a second, independent data set was also collected for CypA and fit using an identical routine. As shown in Table 6, the fitted rate constants are reproducible from a second data set, with errors between the data sets comparable to or lower than fitting uncertainties for most parameters in Table 2; we do observe somewhat larger experimental variability for both trans and cis on-rates, with differences of $0.6 \times 10^6$ and $0.4 \times 10^6$ s$^{-1}$M$^{-1}$ between the data sets for $k_{AB}$ and $k_{DC}$, respectively, suggesting that the true uncertainty in these values may be marginally higher than indicated from the uncertainty in the fitted parameters given in Table 6.

The catalytic cycles of the cyclophilins fall broadly into two categories: one comprising CypA and GeoCyp, for which on-enzyme interconversion is much faster than substrate release, and a second comprising CypB and CypC, for which on-enzyme interconversion occurs at a rate comparable to the rate of substrate release. For CypA and GeoCyp, the rate constants $k_{BC}$ and $k_{CB}$, describing cis↔trans interconversion on the enzyme, are poorly defined by the fitting routine. For these two enzymes, substrate release is the rate limiting factor in catalysis, allowing the substrate to stochastically equilibrate on the enzyme before release; under these conditions, the lineshape remains largely unaffected by increasing on-enzyme exchange, explaining the poor fits of these parameters. As has previously been reported for CypA catalyzing a tetrapeptide substrate, all of the cyclophilins maintain a much higher binding affinity for the cis isoform, leading to an overall slightly higher population of cis-bound enzyme than of the trans-bound form, even though the unbound substrate exists in ~87% trans.

A major result here is that, although the structures of these cyclophilins are largely similar, the specific microscopic rate constants vary, as do the measured dynamics on the slower µs-ms timescale, suggesting a relationship between these rate constants and motions. However, it remains difficult to extract a direct correlation between the microscopic
rates constants and rates of motions from the free enzymes considering that only substrate-
free CypA and CypC can be monitored by CPMG-RD because of the self-association of
CypB and GeoCyp. Moreover, our findings presented above also indicate that the motions
are highly localized within substrate-free CypA and CypC (Figure 25b/c), further clouding
such direct correlations. However, the microscopic rate constants determined here do
provide for direct assessment of the specific microscopic factors underlying observed
functional differences between these cyclophilins (Table 4). For example, in Chapter II, we
presented structural and mutagenesis data demonstrating that GeoCyp has a somewhat
occluded ground state active site but that upon substrate binding the increased binding
affinity and decreased catalytic efficiency of GeoCyp are mediated predominantly through a
substrate ‘clamp’ mechanism that holds the peptide in place in the active site. From our
lineshape analysis, it appears that the increased affinity of GeoCyp for the peptide as
compared to that of CypA is due predominantly to an increased affinity for the cis isoform
and consistent with our previous study, both the cis on-rate and off-rate are decreased
relative to those of CypA in accord with our proposed clamp mechanism. Via comparison of
CypA to the more weakly binding CypC and CypB, affinities for both the cis and trans
isoforms are decreased by similar percentages, effected by both decreased on-rates and
increased off-rates.

Summary

We have functionally characterized the interactions of multiple cyclophilins with a
biologically representative peptide substrate and found that the catalytic efficiency is well
conserved among human family members, while the binding affinities differ more
substantially. Because we have recently characterized a distantly related cyclophilin isoform,
GeoCyp, and found that its catalytic efficiency and apparent dissociation constant are lower
than its human counterparts, here we sought to determine the underlying differences
between catalysis among the cyclophilins. To this end, here we compared previously
determined structures, monitored dynamics on various timescales, and determined the complete catalytic cycle of three human paralogues (CypA, CypB, and CypC) and a thermophilic homologue, GeoCyp. The high degree of structural similarity among human cyclophilins along with the variability in µs-ms dynamics suggests that conformational flexibility plays a role in fine-tuning the catalytic cycle.

The application of lineshape analysis in combination with fitting functional data of binding and catalysis yields well-defined parameters for the full catalytic cycle of multiple cyclophilins. Particularly notable is the degree to which substrate on-rates vary between cyclophilins. The one previous study\textsuperscript{16} to utilize lineshape analysis as applied to CypA reported substrate on-rates of $18 \times 10^6$ and $27 \times 10^6$ s\textsuperscript{-1}M\textsuperscript{-1} for the trans and cis isoforms, respectively, near diffusion-limited, and many times the largest rates reported here. That study utilized a modified tetrapeptide, which may account for the difference in measured values; conformational selection has been suggested to play a major role in the cyclophilin binding mechanism, and the larger peptide utilized here may require a more restrictive, and thus less frequently sampled, conformation for binding. CypA has been shown to exhibit relatively low substrate specificity, implicating the high degree of active site conformational mobility in sampling a wide range of states competent to interact with diverse substrates.\textsuperscript{5,229} The variability in substrate on-rates between cyclophilins may, then, be reflective of subtly altered conformational landscapes whereby each cyclophilin has adapted its particular landscape to the suite of substrates with which it is likely to interact, such that we would expect variability in relative on-rates among the family if measured for an optimized substrate for a given cyclophilin. Indeed, variability in substrate binding energetics between cyclophilins has already been demonstrated and likely plays a role in modulating substrate release.\textsuperscript{5} Nonetheless, the full catalytic cycle appears to have been evolutionarily constrained, yielding nearly identical catalytic efficiencies for each of the human cyclophilins.
Table 4. Apparent binding affinities and observed isomerization rates for multiple cyclophilins

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$K_{D_{app}}$ (µM)</th>
<th>$k_{iso}$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CypA</td>
<td>76±3$^a$</td>
<td>10.0±0.5</td>
</tr>
<tr>
<td>CypB</td>
<td>171±5</td>
<td>8.4±0.4</td>
</tr>
<tr>
<td>CypC</td>
<td>146±4</td>
<td>9.0±0.5</td>
</tr>
<tr>
<td>GeoCyp</td>
<td>39±2</td>
<td>4.7±0.3</td>
</tr>
</tbody>
</table>

$^a$Errors in fits to a single experiment

Table 5. Microscopic rate constants determined for CypA using different values of $R_2$-Bound

<table>
<thead>
<tr>
<th>$R_2$-Bound</th>
<th>40 s$^{-1}$</th>
<th>30 s$^{-1}$</th>
<th>50 s$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{AB}$ (x10$^6$ s$^{-1}$M$^{-1}$)</td>
<td>3.2±0.2$^a$</td>
<td>3.2±0.3</td>
<td>3.2±0.2</td>
</tr>
<tr>
<td>$k_{BA}$ (s$^{-1}$)</td>
<td>550±160</td>
<td>530±190</td>
<td>590±203</td>
</tr>
<tr>
<td>$k_{BC}$ (s$^{-1}$)</td>
<td>1660±660</td>
<td>1810±990</td>
<td>1870±890</td>
</tr>
<tr>
<td>$k_{CB}$ (s$^{-1}$)</td>
<td>1070±230</td>
<td>1390±730</td>
<td>1120±350</td>
</tr>
<tr>
<td>$k_{CD}$ (s$^{-1}$)</td>
<td>170±40</td>
<td>200±90</td>
<td>170±60</td>
</tr>
<tr>
<td>$K_{DC}$ (x10$^6$ s$^{-1}$M$^{-1}$)</td>
<td>10.0±0.3</td>
<td>10.0±0.3</td>
<td>10.0±0.2</td>
</tr>
</tbody>
</table>

$^a$Errors determined among the 20 best fits to the data

Table 6. Best-fit solutions of cyclophilin microscopic rate constants

<table>
<thead>
<tr>
<th></th>
<th>CypA</th>
<th>CypA data set 2</th>
<th>CypB</th>
<th>CypC</th>
<th>GeoCyp</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{AB}$ (x10$^6$ s$^{-1}$M$^{-1}$)</td>
<td>3.2±0.2$^a$</td>
<td>3.8±0.4</td>
<td>2.1±0.2</td>
<td>4.2±0.8</td>
<td>5.2±1.8</td>
</tr>
<tr>
<td>$k_{BA}$ (s$^{-1}$)</td>
<td>550±160</td>
<td>670±240</td>
<td>960±430</td>
<td>1620±820</td>
<td>720±270</td>
</tr>
<tr>
<td>$k_{BC}$ (s$^{-1}$)</td>
<td>1660±660</td>
<td>1570±830</td>
<td>550±290</td>
<td>370±170</td>
<td>2410±1330</td>
</tr>
<tr>
<td>$k_{CB}$ (s$^{-1}$)</td>
<td>1070±230</td>
<td>1000±280</td>
<td>320±100</td>
<td>210±50</td>
<td>810±370</td>
</tr>
<tr>
<td>$k_{CD}$ (s$^{-1}$)</td>
<td>170±40</td>
<td>170±50</td>
<td>200±70</td>
<td>260±100</td>
<td>50±10</td>
</tr>
<tr>
<td>$K_{DC_{cis}}$ (µM)</td>
<td>17±4</td>
<td>18±4</td>
<td>40±14</td>
<td>32±11</td>
<td>7±1</td>
</tr>
<tr>
<td>$K_{DC_{trans}}$ (µM)</td>
<td>170±50</td>
<td>180±70</td>
<td>460±230</td>
<td>400±180</td>
<td>140±30</td>
</tr>
<tr>
<td>Bound cis/trans</td>
<td>1.6±0.8</td>
<td>1.7±1.1</td>
<td>2.1±1.5</td>
<td>2.1±1.4</td>
<td>3.1±1.0</td>
</tr>
<tr>
<td>$\omega_{trans-bound}$ D6 (ppm)</td>
<td>8.65±0.04</td>
<td>8.59±0.03</td>
<td>8.63±0.07</td>
<td>8.63±0.06</td>
<td>8.14±0.50</td>
</tr>
<tr>
<td>$\omega_{cis-bound}$ D6 (ppm)</td>
<td>8.43±0.03</td>
<td>8.47±0.04</td>
<td>9.76±0.96</td>
<td>8.20±1.05</td>
<td>8.93±0.04</td>
</tr>
<tr>
<td>$\omega_{trans-bound}$ L7 (ppm)</td>
<td>8.30±0.01</td>
<td>8.29±0.01</td>
<td>8.38±0.05</td>
<td>8.36±0.04</td>
<td>8.05±0.16</td>
</tr>
<tr>
<td>$\omega_{cis-bound}$ L7 (ppm)</td>
<td>8.26±0.01</td>
<td>8.27±0.01</td>
<td>8.25±0.02</td>
<td>8.23±0.05</td>
<td>8.36±0.07</td>
</tr>
</tbody>
</table>

$^a$Standard deviation determined from the 20 best fits to the data
Figure 21. Functional characterization of multiple cyclophilins (a) Titrations of the peptide substrate into $^{15}$N-labeled cyclophilins. $^{15}$N-HSQC spectra were collected on 0.5 mM cyclophilin with 0 mM (red), 0.1 mM (orange), 0.2 mM (green), 0.5 mM (cyan), 1 mM (blue), and 2 mM (violet) unlabeled peptide substrate. (b) ZZ-exchange spectra for residue Leu 7, collected with mixing times of 0 ms (black) and 144 ms (red) on 1 mM $^{15}$N-labeled peptide with the addition of 20 µM cyclophilin.
Figure 22. Highly conserved structure and fast timescale dynamics among CypA, CypB, and CypC, and variable µs timescale exchange (a) Overlaid ribbon diagrams of previously determined crystal structure of CypA (PDBID 2RMA, white230), CypB (PDBID 1CYN, blue231), and CypC (PDBID 2ESL, red, to be published). Sidechains of active site and gatekeeper residues are depicted as sticks. (b) $^{15}$N $R_1$ relaxation rates, collected at 900 MHz for $^{15}$N CypA, CypB, and CypC. (c) $^{15}$N $R_{20}$ relaxation rates, estimated by CPMG-RD with a $\nu_{cpmg}$ of 1000 s$^{-1}$ collected on $^2$H$^{15}$N CypA, CypB, and CypC at 900 MHz. Rates are shifted up by 15 s$^{-1}$ (CypB) and 30 s$^{-1}$ (CypA) to facilitate comparison. For both (b) and (c), residue numbers are those for CypA, dots represent individual values, and lines represent a 5 residue moving average, with lines plotted only for regions with three or more residues per window.

Figure 23. Overlaid ribbon diagrams of CypA crystal structures Structures determined alone (blue, PDBID 2CPL232) and bound to CsA (white PDBIDs 1CYN230). Conserved active site and gatekeeper residues are shown as sticks.
Figure 24. Testing self association in CypB and CypC For CypC and CypB, CPMG-RD data were collected at 21.1 T with concentrations of 0.5 mM (black dots) and 1 mM (red dots) $^2$H$^{15}$N-labeled protein. Least-squares fits to the Carver-Richards equations of the 21.1 T data alone are represented as dotted lines. Baseline $R_{20}$ values are subtracted from each data set so that $R_{ex}$ values can be compared directly. CypB, but not CypC, self-associates on the µs-ms timescale.
Figure 25. Divergent µs-ms motions in CypA and CypC (a) $k_{\text{ex}}$ values measured for CypA and CypC. Data were collected at 600 MHz and 900 MHz and fit individually for each residue to the Carver-Richards equations. Example data (dots) and fits (lines) are shown for homologous residues Cys 62 and Ile 96 at 600 MHz (dashed lines) and 900 MHz (solid lines). Residues are included for which $R_{\text{ex}}$ is greater than 0.5 s$^{-1}$ and the error in $k_{\text{ex}}$ is both less than 50% of $k_{\text{ex}}$ and less than 2000 s$^{-1}$.

(b) $k_{\text{ex}}$ distributions for CypA and CypC. Smoothed histogram showing µs-ms rate distribution for CypA (black) and CypC (red). For each $k_{\text{ex}}$ value, the number of residues is plotted exhibiting a measured rate of ±250 s$^{-1}$.

Figure 26. Four state peptide-perspective minimal exchange model of cyclophilin catalysis $k_{\text{AD}}$ and $k_{\text{DA}}$ are many orders of magnitude slower than the other exchange parameters.
Figure 27. Extraction and fitting lineshape data

(a) High resolution $^{15}$N-HQSC spectra were collected on $^{15}$N-labeled peptide with variable concentrations of unlabeled cyclophilin without nitrogen decoupling during signal acquisition. Shown here is the peptide in the absence of enzyme. Peaks corresponding to residues Asp 6 and Leu 7 with Pro 5 in the cis and trans conformations were extracted using the windows indicated and summed over nitrogen to yield 1-D proton spectra. (b) Lineshape fitting for CypA. Fitting was performed as described below, with data in black and best-fit lineshapes, determined simultaneously for all peptide and enzyme concentrations, in red. For the sake of clarity, 1D spectra are shown only for 1 mM $^{15}$N-labeled peptide with 5 µM, 10 µM, 20 µM, 50 µM, and 100 µM CypA (top to bottom for each residue). Because of the low intensity relative to that of trans peaks, cis peak intensities have been scaled up by a factor of five in bottom graphs for the purpose of visualization.
CHAPTER IV
LIMITATIONS ON THE APPLICATION OF CPMG-RD IN
REVERSIBLE CATALYTIC SYSTEMS

Introduction

The Carr-Purcell-Meiboom-Gill Relaxation Dispersion (CPMG-RD) experiment has been previously applied to study microscopic rate constants in reversible enzymatic systems. CPMG-RD permits quantitative measurement of chemical exchange phenomena in the ~100-5000 s\(^{-1}\) range. 2-D spectra are collected while a refocusing pulse is varied between ~50-1000 s\(^{-1}\) (\(\nu_{\text{cpmg}}\)), allowing, for each \(\nu_{\text{cpmg}}\), calculation of \(R_2\) relaxation rates, which comprise an intrinsic relaxation rate (\(R_{20}\)) and an exchange-induced relaxation rate (\(R_{\text{ex}}\)).\(^{104,213}\) The \(R_{\text{ex}}\) component can be fit to the Carver-Richards equations to extract kinetic (exchange rate, \(k_{\text{ex}}\)), thermodynamic (population occupancies), and structural (chemical shift difference, \(\Delta \omega\)) information about the exchange.\(^{114}\)

Studies of both CypA and the structurally unrelated PPIase Pin1 have measured chemical exchange via CPMG-RD on the enzyme in the presence of near-saturating concentrations of substrate and interpreted the motions as being reflective of on-enzyme turnover.\(^{172,176,225}\) These studies have been cited as evidence of a direct connection between the inherent conformational fluctuations in these enzymes and the rate of on-enzyme isomerization in these enzymes and utilized to identify those regions of the protein directly participating in the enzymatic cycle.\(^{106,183}\) While the findings of these studies are consistent with the broader consensus of the field in that the intrinsic dynamics of proteins are

generally functionally important, the particular findings of these studies as they relate to prolyl isomerization are dependent on those dynamics in the bound form reporting solely on the on-enzyme isomerization step catalyzed by a PPIase. Limited justification is provided in each case, however, to the assumption that binding and release have no impact on measured dynamics in these systems. Given the apparent importance of dynamics in regulating cyclophilin function (see chapters III and V), measurement of catalytically associated movements in CypA would be valuable in interpreting the role of dynamics in various aspects of the catalytic cycle, as would a means to determine how CPMG-RD measured dynamics are impacted by each aspect of the CypA catalytic cycle. Here, we identify limitations in the approach of measuring on-enzyme catalysis by CPMG-RD by providing both computational and experimental evidence indicating that CPMG-RD can be used only to measure on-enzyme catalysis in systems for which substrate release is much slower than on-enzyme interconversion.

Results

Modeling CPMG-RD in reversible systems

Multiple studies have measured exchange phenomena during catalytic turnover via measurement of CPMG-RD on enzymes in the presence of saturating concentrations of substrate, including for CypA and Pin1, and have concluded that these measured exchange phenomena may be representative of on-enzyme catalysis.\textsuperscript{106,168,172,176,225} To assess the validity of this approach for our current study, we have modeled the expected CPMG-RD data given by the minimal three-state exchange model from the enzyme perspective (Figure 28) by simulating the individual magnetizations of $10^5$ atoms as they proceed through the CPMG-RD pulse sequence for each $\nu_{cpmg}$ with each atom stochastically switching states as described by the microscopic rate constants determined by lineshape analysis and listed in Table 6.
For each model run, atoms were distributed among the three states depicted in Figure 28 as given by the equilibrium concentrations and begun with coherent magnetization. Magnetization was allowed to evolve over a 20 ms simulation time ($\tau_{\text{cpmg}}$), while stochastically transitioning between states as proscribed by the microscopic rate constants. *Trans*-bound and *cis*-bound chemical shifts of the simulated atom in the enzyme were arbitrarily set to 0.5 ppm and 1 ppm downfield of the free enzyme chemical shift, respectively, for all simulations. Refocusing pulses were applied with $\nu_{\text{cpmg}}$ of 50-1000 s$^{-1}$ by taking the complex conjugate of the magnetization at corresponding intervals. The absolute value of the sum of the magnetization of all atoms after the simulation time yields the peak intensity, and $R_{\text{ex}}$ values are calculated as for measured data (Equation 4) for which $I_0$ is set to the sum of magnetizations before evolution. We have validated this model by applying it to a simple two-state exchange simulation, for which it aligns with the expected results as described by the Carver-Richards equations over a range of input parameters (Figure 29). We have additionally simulated the CPMG-RD data, but blocked on-enzyme catalysis from occurring by setting $k_{BC}$ and $k_{CB}$ to zero.

As shown in Figure 30, even in the absence of on-enzyme catalytic turnover (blue lines), and at arbitrarily high substrate concentrations (see Figure 31), significant exchange-induced relaxation is expected for all four cyclophilins using the microscopic rate constants determined by lineshape analysis. The origin of this exchange in the absence of catalysis lies in interconversion, from the enzyme perspective, between the *cis*-bound form and *trans*-bound form via the free enzyme intermediate, by binding to different substrate molecules in different conformations (Figure 30b, blue arrows). The enzyme-perspective free$\leftrightarrow$*trans* (where $k_{\text{free-\text{trans}}}$ = $[A]k_{AB}+k_{BA}$) and free$\leftrightarrow$*cis* (where $k_{\text{free-\text{cis}}}$ = $k_{CD}+[D]k_{DC}$) interconversions each occur on timescales much faster than the CPMG-RD regime at high substrate concentrations, such as the peptide concentration of 6 mM utilized for all CPMG-RD
experiments here, and therefore do not directly influence $R_{ex}$ (see Figure 31 and Figure 32); however, the effective $\text{cis} \leftrightarrow \text{trans}$ interconversion via the free enzyme intermediate lies squarely in the CPMG regime for these parameters, as calculated by Equation 14 below. In other words, at high substrate concentrations, a residue with identical $\text{cis}$-bound and $\text{trans}$-bound chemical shifts and a different free chemical shift would register no chemical exchange via CPMG-RD, while a residue with different $\text{cis}$-bound and $\text{trans}$-bound chemical shifts would register chemical exchange even in the absence of isomerization. The apparent exchange rate ($k_{ex-app}$) that would be measured by any probe monitoring $\text{cis} \leftrightarrow \text{trans}$ exchange is given by:

Equation 14

\[ k_{ex-app} = k_{ex-catalyzed} + k_{ex-uncatalyzed} \]

with:

\[ k_{ex-catalyzed} = k_{bc} + k_{cb} \]

\[ k_{ex-uncatalyzed} = (k_{BA}k_{DC}[D] + k_{CD}k_{AB}[A])/(k_{DC}[D] + k_{AB}[A]) \]

where $[A]$ and $[D]$ are the equilibrium concentrations of unbound $\text{trans}$ and $\text{cis}$ peptide, respectively. This phenomenon is illustrated in Figure 30b by following the state occupancy of a single simulated atom over a 10 ms simulation. While virtually no time is spent in the free state, substrate release permits non-catalyzed exchange of the $\text{cis}$ and $\text{trans}$ isoforms on the enzyme (Figure 30b, $k_{ex-uncatalyzed}$: ‘*’) in addition to catalyzed exchange (Figure 30b, $k_{ex-catalyzed}$: ‘^’). Given that $k_{ex-app}$ is dependent all six rate constants (Equation 14), the conditions under which CPMG-RD measurements may be indicative of on-enzyme catalysis cannot be readily generalized to a given system without consideration of the full catalytic cycle. However, to estimate the range over which $k_{ex-uncatalyzed}$ may significantly impact CPMG-RD measurements in CypA, we simulated data over a range of $k_{ex-catalyzed}$ and $K_{D-app}$ values, keeping the remaining rate constants unchanged from those listed in Table 6 (Figure
In general, only in the limiting case for which substrate release (i.e., \( k_{BA} \) and \( k_{CD} \)) is much slower than on-enzyme catalysis can measured CPMG-RD data possibly be predominantly representative of on-enzyme catalysis. For all other cases involving reversible enzymatic catalysis, measured CPMG-RD data are instead influenced by both on-enzyme interconversion and interconversion via the free-enzyme intermediate. These two components cannot be readily parsed from the CPMG-RD data alone but can be separated by utilizing lineshape analysis as we have described above.

\[ k_{\text{ex-app}}, k_{\text{ex-catalyzed}}, \text{ and } k_{\text{ex-uncatalyzed}} \] values for each of the cyclophilins studied here, calculated from the microscopic rate constants in Table 6, are listed in Table 7. \( k_{\text{ex-app}} \) is minimally impacted by uncatalyzed interconversion for CypA or Geo, while it is significantly impacted for both CypC and CypB. To compare the measured CPMG-RD exchange rates to the \( k_{\text{ex-app}} \) values predicted by the microscopic rate constants, CPMG-RD experiments were conducted at multiple magnetic field strengths on \( ^2H^{15}N \)-labeled versions of each of the cyclophilins in the presence of saturating concentrations of the peptide substrate, and data were fit to the Carver-Richards equations. To ensure that interconversion with the free protein (i.e., free
\[ \text{trans} \leftrightarrow \text{cis} \] and free
\[ \text{trans} \leftrightarrow \text{cis} \]) would not directly impact measured exchange by CPMG-RD, data were collected at multiple substrate concentrations (Figure 32a). Additionally, the weak self-association of CypB and GeoCyp is blocked upon addition of saturating concentrations of substrate (Figure 32a), therefore allowing CPMG-RD data to be compared among all four cyclophilins. While all measured motions in the enzyme are not necessarily indicative of \( \text{trans} \leftrightarrow \text{cis} \) interconversion, measured motions detected at the conserved catalytic arginine (Arg 55 in CypA) are compared to the expected \( k_{\text{ex-app}} \) values in Table 7. With the exception of CypA, the expected \( k_{\text{ex-app}} \) values are similar to those experimentally observed at this residue. The overestimation of \( k_{\text{ex-app}} \) for CypA may be due to
the poorly defined $k_{BC}$ and $k_{CB}$ values that define $k_{\text{ex-catalyzed}}$ because of their minimal impact on lineshapes, Table 6.

_Eperimental measurement of non-catalyzed enzyme-perspective cis↔trans interconversion_

To experimentally demonstrate the existence of measurable CPMG-RD exchange under saturating conditions in the absence of catalysis, we utilized a substrate for which the G-P peptide bond in the peptide is replaced by a thioamide bond (TA-peptide). This substrate binds only ~50% weaker than the standard GSFGPDLRAGD peptide but is not isomerized by CypA.\textsuperscript{125} The TA-peptide does exhibit both _trans_ and _cis_ isoforms (~3% _cis_), but no exchange peaks can be detected indicating CypA mediated catalysis, even with elevated temperatures and very high enzyme concentration (500 µM) (Figure 34). We collected $^{15}$N-CPMG-RD data on $^2$N$^{15}$N-labeled CypA in the presence of saturating concentrations of either the peptide substrate or the uncatalyzable TA-peptide (see Figure 32b for validation of saturation by the TA-peptide). As shown in panels Figure 35a/b, chemical exchange can be measured throughout most of the active site for both substrates, albeit at a significantly slower rate for the modified peptide (~1500 s\(^{-1}\) for the peptide vs. ~300 s\(^{-1}\) for the TA-peptide). This lowered exchange rate is expected and indeed recapitulated in our simulation (Figure 30a), as the interconversion is only proceeding through the free-enzyme intermediate as opposed to a combined exchange through both the free-enzyme intermediate and on-enzyme exchange. To demonstrate that this exchange is reporting on binding and release of the substrate, we also collected $^{15}$N-CPMG-RD data on $^2$N$^{15}$N-labeled CypA in the presence of the inhibitor CsA, which binds with a dissociation constant of 7 nM.\textsuperscript{5} The exchange observed for CypA in the presence of both the catalyzed peptide and uncatalyzed TA-peptide that both exist in two unbound isoforms is quenched by CsA throughout most of the active site, as expected due to both the lack of _cis_ and _trans_ isoforms in CsA and the slow off-rate associated with binding at 7 nM affinity (Figure 35a/b).
**Segmental dynamics in CypA in the bound form**

Despite the limitations in utilizing CPMG-RD to directly measure on-enzyme catalytic turnover in cyclophilins due to the contribution of uncatalyzed exchange, CPMG-RD still allows us to identify the existence of localized dynamics in CypA in the presence of uncatalyzable binding partners. We mapped $k_{ex}$ values measured on $^{2}$H$^{15}$N-labeled CypA in the presence of saturating conditions of the peptide, the uncatalyzed TA-peptide, or CsA onto the structure of CypA. As described above and plotted in Figure 35a/b, for much of the active site, dynamics are completely quenched for the CypA:CsA complex and slowed dramatically in the CypA:TA-peptide complex as compared to those of the CypA:peptide complex. Within one region of the active site, however, comprising a loop of residues ~65-78 and adjacent residues ~108-110, significant exchange is observed for CypA:CsA and much faster motions (~2000 s$^{-1}$) are observed for CypA:TA-peptide (Figure 35b). The measured chemical exchanges for residues in this region are not fully independent of substrate binding, as $R_{ex}$ is altered significantly among the three binding partners (Figure 35c); nonetheless, these residues are not reporting directly on $trans$$\leftrightarrow$$cis$ interconversion and demonstrate the existence of localized, yet coupled, µs-ms conformational dynamics in CypA when bound to substrate.

**Summary**

We have identified significant limitations in applying CPMG-RD experiments to monitor on-enzyme catalytic activity in reversible enzymatic systems. Specifically, unless the enzyme-perspective substrate exchange via the free enzyme intermediate ($k_{ex$-uncatalyzed}, Figure 30) is much slower than on-enzyme exchange ($k_{ex$-catalyzed}$), the effective $cis$$\leftrightarrow$$trans$ exchange ($k_{ex-app}$) does not report solely on $k_{ex$-catalyzed} (Equation 14). Of previous studies analyzing substrate-saturated CypA dynamics, two utilized a weakly binding ($K_{D-app}$ of ~1 mM, $k_{off}$ of ~$10^{4}$ s$^{-1}$) tetrapeptide, while another used a domain of the HIV capsid
protein,\textsuperscript{176} which binds much more tightly ($K_{D\text{-app}}$ of 16 µM, $k_{\text{off}}$ of ~45 s\textsuperscript{-1}). In the former cases, the high value of $k_{\text{ex-uncatalyzed}}$ renders $k_{\text{ex-app}}$ too fast to be observed via CPMG-RD such that any observed dynamics by CPMG-RD cannot be indicative of cis↔trans interconversion; notably, a majority of the residues with measured exchange under substrate-bound conditions in these studies colocalize with dynamics identified in panels Figure 35a/b as not reporting directly on cis↔trans interconversion. Alternatively, because of the slow off-rates, dynamics measured during catalysis of the HIV capsid protein are potentially indicative of on-enzyme exchange as reported. A previous study that analyzed Pin1 dynamics during catalysis also utilized a weakly binding substrate ($K_{D\text{-app}}$ of ~0.8 mM), likewise precluding direct measurement of $k_{\text{ex-catalyzed}}$ by CPMG-RD.\textsuperscript{225} Thus, great care must be taken in interpreting CPMG-RD data collected on any multispecies systems with schemes more complex than a simple two-state binding model.
Table 7. Catalyzed and uncatalyzed exchange rates for multiple cyclophilins

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_{ex\text{-catalyzed}}$ (s$^{-1}$)</th>
<th>$k_{ex\text{-uncatalyzed}}$ (s$^{-1}$)</th>
<th>$k_{ex\text{-app}}$ (s$^{-1}$)</th>
<th>$k_{ex\text{-Arg}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CypA</td>
<td>2740±700$^a$</td>
<td>290±30</td>
<td>3030±700</td>
<td>1600±210$^b$</td>
</tr>
<tr>
<td>CypB</td>
<td>880±210</td>
<td>390±80</td>
<td>1270±280</td>
<td>980±330</td>
</tr>
<tr>
<td>CypC</td>
<td>580±130</td>
<td>560±120</td>
<td>1130±240</td>
<td>1080±420</td>
</tr>
<tr>
<td>GeoCyp</td>
<td>3200±1600</td>
<td>160±30</td>
<td>3380±1610</td>
<td>4010±2000</td>
</tr>
</tbody>
</table>

$^a$Standard deviation determined from the 20 best fits to the data

$^b$Errors in fitting the Carver-Richards equations
Figure 28. Minimal three state enzyme-perspective exchange model Model uses identical microscopic rate constants as depicted in Figure 26.

Figure 29. Validation of stochastic CPMG-RD model. CPMG-RD simulations were performed for a simple two-state exchange over a range of exchange values ($k_{ex}$ of 500, 1000, and 2000 s$^{-1}$), minor state populations ($P_B$ of 0.02, 0.10 and 0.20) and chemical shift differences ($\Delta\omega$ of 0.5, 1, and 2 ppm). Simulated data (black dots) from independent simulations are plotted over calculated values (red lined) given by the Carver-Richards equations for each set of parameters.
Figure 30. CPMG-RD can report on uncatalyzed exchange in reversible enzymatic systems (a) Using the rate constants determined from lineshape analysis (Table 2), CPMG-RD data were simulated from the cyclophilin perspective with (black) or without (blue) on-enzyme catalysis, with in silico concentrations of 1 mM cyclophilin and 10 mM peptide. For all cyclophilins, significant exchange persists even in the absence of catalysis. (b) State map of a single simulated atom over a 10 ms simulation, using the microscopic rate constants determined for CypC. Vertical lines represent transitions between states, while horizontal lines represent time spent in a single state. Transitions between Cyp:Pep\textsubscript{trans} and Cyp:Pep\textsubscript{cis} occur both through on-enzyme catalysis (\^) and through the free enzyme intermediate (*).

Figure 31. $R_{ex}$ as a function of substrate concentration for uncatalyzed peptide interconversion $R_{ex}$ values for $\nu_{cpmg}$ of 50 s$^{-1}$ were calculated for CPMG-RD data simulated for 1 mM CypA using the rate constants in Table 1 with $k_{BC}$ and $k_{CB}$ set to zero over a range of substrate concentrations. The chemical shifts of the trans-bound and cis-bound forms were arbitrarily set to 0.5 and 1 ppm downfield of the free chemical shift. CPMG-RD sensitive chemical exchange persists even at an arbitrarily high substrate concentration, as demonstrated by simulation at 1000 mM peptide.
Figure 32. Cyclophilin peptide saturation (a) For each cyclophilin used in this study, CPMG-RD data were collected on 0.5 mM enzyme in the presence of two concentrations (6 mM, black and 8 mM, red) of the model substrate to ensure that measured exchange was not monitoring interconversion between free and bound forms. For CypB and GeoCyp, which weakly self associate in the absence of substrate, data were also collected in the presence of 6 mM substrate with 0.25 mM enzyme (blue), confirming that the proteins no longer self associate in the presence of saturating substrate concentrations. (b) CPMG-RD data were collected on 0.5 mM $^2$H$^{15}$N CypA with 6 mM and 8 mM TA-FGP, confirming that free-bound interconversion is not being measured.
Figure 33. CPMG-RD simulations performed over a range of $k_{\text{ex-catalyzed}}$ and $K_{\text{D-app}}$ values. Beginning with the rate constants listed for CypA in Table 6, $k_{BC}$ and $k_{CB}$ were scaled via a constant value to alter $k_{\text{ex-catalyzed}}$, while $k_{BA}$ and $k_{CD}$ were scaled by a constant value to alter $K_{\text{D-app}}$. All data were simulated with substrate concentrations of 50 mM to ensure \textit{in silico} saturation of the enzyme, even at the highest $K_{\text{D-app}}$ value. The chemical shift difference between \textit{trans}-bound and \textit{cis}-bound was set to 0.5 ppm for all simulations.
Figure 34. Comparing catalytic turnover of the model peptide and its thioamide variant by CypA ZZ-exchange experiments were collected with a 0.5 s mixing time on each peptide alone (top), and in the presence of CypA (bottom). With only 20 µM CypA, the model substrate is robustly catalyzed, as indicated by the appearance of ‘cross-peaks’ between the cis and trans isomers. Even with a 25 fold increase in CypA concentration to 500 µM, however, the thioamide modified peptide exhibits no observable cross-peaks, indicating no detectable isomerization by CypA.
Figure 35. Measured CPMG-RD data on $^2$H$^{15}$N CypA when it is saturated with catalyzable and uncatalyzable substrates (a) $^{15}$N-CPMG-RD data collected at 900 MHz (dots) and best-fit curves (lines) using the Carver-Richards equations, fit to data collected at 600 MHz and 900 MHz for representative residues in $^2$H$^{15}$N CypA with saturating concentrations of the peptide substrate (black), the TA-peptide (blue), or CsA (red). (b) Measured exchange rates ($k_{\text{ex}}$) mapped onto the CypA structure in the presence of saturating concentrations of peptide, TA-peptide, or CsA. Data for each residue were individually fit. Residues are included for which $R_{\text{ex}}$ is greater than 0.5 s$^{-1}$ and the error in $k_{\text{ex}}$ is both less than 50% of $k_{\text{ex}}$ and less than 2000 s$^{-1}$. (c) CPMG-RD traces of representative residues in CypA that are not directly reporting on exchange between Cyp:Pep$_{\text{trans}}$ and Cyp:Pep$_{\text{cis}}$. 
CHAPTER V
MAPPING NETWORKS OF NON-COHERENT DYNAMIC
ALLOSTERIC COUPLING IN CYCLOPHILIN A

Introduction

Many proteins exhibit conformational fluctuations on the micro-millisecond timescale, which have been shown to play functional roles in the binding and release of binding partners as well as, for enzymes, in the on-enzyme conformational search for the transition state.\textsuperscript{108,131,164,167,217,233,234} Perhaps the most powerful development in probing these motions has been the Carr-Purcell-Meiboom-Gill relaxation dispersion (CPMG-RD) experiment, a nuclear magnetic resonance (NMR) based experiment that allows for atomic-resolution, quantitative measurement of low population states that are sampled with an exchange rate of \(\sim 100-5000 \text{ s}^{-1}\).\textsuperscript{112,113} The CPMG-RD experiment consists of a series of refocusing pulses at a variable frequency \(\nu_{\text{cpmg}}\), generally between \(\sim 50-1000 \text{ s}^{-1}\); linebroadening due to exchange induced relaxation (\(R_{\text{ex}}\)) on the micro-millisecond timescale is refocused by these pulses with a signature dependent upon the rate of exchange (\(k_{\text{ex}}\), where \(k_{\text{ex}}=k_{A\to B}+k_{B\to A}\) for two-site exchange between states A and B), the minor state population, and the chemical shift difference (\(\Delta \omega\)) between the states. To extract these parameters, the CPMG-RD signature is generally fit to the Carver-Richards equations, which describe generalized two-state exchange.\textsuperscript{114}

Out of studies utilizing CPMG-RD, many model systems have been extensively characterized and found to exhibit \(\mu\text{s-ms}\) dynamics representative of a single global exchange phenomenon, often a large hinge or loop motion or localized unfolding (i.e. most or all probes reporting exchange are indicative of a single structural transition).\textsuperscript{131,163,164,166,233,234} Additionally, a smaller number of studies have addressed systems for which motions are segmental in nature, generally consisting of side-chain
rearrangements or localized loop motions. Of these studies, a number of them have identified non-coherent motions in an enzyme that coalesce into global motion upon post-translational modification\textsuperscript{167} or ligand binding,\textsuperscript{151} perhaps suggesting this transition as a regulatory mechanism, while other systems appear to utilize non-coherent dynamics directly in function. A notable study in the latter group, by McDonald et al., demonstrated for the bacterial signaling protein CheY that μs-ms dynamics are segmental in nature, yet are still able to propagate an allosteric response to Mg+ binding.\textsuperscript{150} This study is representative of an increasing realization of the role of dynamics in propagating allosteric communication, in which a protein is viewed as an ensemble of interconverting structures for which binding or other modification is propagated via shifting the population sampling.\textsuperscript{143-145}

Cyclophilin A (CypA) is a highly dynamic protein on the μs-ms timescale that has been utilized previously in CPMG-RD based studies correlating its conformational fluctuations and enzymatic function. CypA is a peptidyl-prolyl isomerase, catalyzing the reversible \textit{cis}-\textit{trans} interconversion of prolyl-peptide bonds,\textsuperscript{11} and plays a role a range of biological functions, including as a foldase,\textsuperscript{1} as a protein chaperone,\textsuperscript{20} in signal transduction,\textsuperscript{22} and as an extracellular cytokine.\textsuperscript{38} CypA also functions in multiple pathogeneses, including a role in the promotion of viral infectivity of HIV, HCV, influenza, and multiple other viruses\textsuperscript{57} and in the development of multiple cancers and other inflammatory diseases.\textsuperscript{88,95}

Original studies of CypA dynamics identified two distinct groups of residues moving with different exchange rates in the protein alone and a single, global exchange when bound to substrate.\textsuperscript{106} More recently, utilizing deuteriation, data collection at multiple fields, and temperature variation, we have shown that CypA dynamics in the free protein are more prevalent and segmental than originally observed, comprising nearly 50% of the residues and a spanning range of nearly an order of magnitude in rates of motion.\textsuperscript{173} Additionally, we have recently shown that distinct segmental dynamics persist in CypA even in the presence
of saturating concentrations of substrate. Further, in CypA alone or when saturated with substrate, significant global communication occurs between non-coherent regions moving with varied exchange rates such that mutagenesis or substrate binding can impact measured motions across the protein at sites reporting on distinct structural transitions. The mechanisms by which this inter-segmental communication occurs in CypA and the corresponding functional ramifications have yet to be elucidated.

Here, we identify multiple active site distal residues in CypA that are nonetheless allosterically coupled to dynamics within the active site. Utilizing NMR chemical shift guided molecular dynamics (MD) ensembles of structures, we identify partially correlated conformational sampling which may propagate dynamic changes in the active site to these distal sites through putative non-coherent allosteric pathways. By generating multiple single site mutations at each of the distal sites and monitoring the corresponding global alterations to CPMG-RD signatures, we experimentally identify distinct allosteric pathways consistent with those identified in the MD ensembles. In addition to altering global conformational sampling, these active site distal mutants induce changes to CypA’s bulk catalytic function towards multiple substrates, definitively demonstrating a link between conformational dynamics and enzymatic function in CypA. Finally, we determine the specific impacts of these mutants on the microscopic rate constants that define the CypA catalytic cycle and demonstrate a role for altered conformational sampling in both substrate binding and release.

Results

Identification of a distal residue allosterically connected to active site dynamics

As has been previously shown, micro-millisecond conformational dynamics in CypA are highly localized, yet nonetheless able to intercommunicate. Specifically, within the active site, mutation of a single residue can alter monitored motions elsewhere in the active site at residues that experience different rates of motion and temperature dependencies.
through a currently unknown mechanism. We sought here to expand this approach and identify residues distal from the active site, yet linked allosterically to active site dynamics, thereby allowing perturbation of active site dynamics independent of changes to the substrate interface. While the majority of motions measured by CPMG-RD localize to the CypA active site, several distal residues also exhibit measurable relaxation dispersion. We monitored dynamics, via CPMG-RD, at these sites, in $^{2}\text{H}^{15}\text{N}$-labeled wild type CypA (CypA$^{\text{WT}}$) alone, as well as in the context of single-site mutations within the active site or in the presence of saturating concentrations of a well-characterized peptide substrate, GSFGPDLRAGD (referred to herein as FGP peptide). As previously described, measurement of CPMG-RD on labeled CypA in the presence of relatively weakly binding substrates, such as this FGP peptide, are not solely indicative of on-enzyme isomerization. Nonetheless, altered distal site motions due to active site mutation or substrate binding suggest an allosteric link to some aspect of the full catalytic cycle.

As shown in Figure 36a, residue Val 29 is $\sim$15 Å from the catalyzed substrate during isomerization. However, upon mutation of Phe 113 to alanine (CypA$^{F113A}$), mutation of Lys 82 to alanine (CypA$^{K82A}$), or addition of the FGP substrate, μs-ms dynamics measured via CPMG-RD on the amide of Val 29 are significantly altered (Figure 36b). Specifically, for each of these mutations or addition of substrate, $R_{\text{ex}}$ is increased by 2-10 fold. Notably, individual fits of the data to the Carver-Richards equations indicate that Phe 113 exhibits an exchange rate ($k_{\text{ex}}$) of 850±100 s$^{-1}$, while Lys 82 is moving with a $k_{\text{ex}}$ of 2650±100 s$^{-1}$, demonstrating a system of allosteric communication throughout much of the protein among residues moving non-coherently. Dispersion was also measured on selectively $^{13}\text{C}$-methyl-labeled CypA alone and in the presence of saturating substrate concentrations. While the data are of lower quality, a clear increase in dispersion can be identified in the methyl probes of Val 29 as well (Figure 37). Additionally, mutation to Val 29, discussed further below, leads to significant alterations in global exchange (Figure 42). Combined, these data
suggest a link between the distal residue Val 29 and conformational fluctuations within the active site alone and during interactions with a substrate.

**Identifying putative communication pathways in chemical shift guided MD ensembles.**

As demonstrated in Figure 36, regional motions in CypA intercommunicate throughout the protein. We sought here to identify a specific pathway or pathways by which this communication occurs and to determine the mechanism or mechanisms by which non-coherently mobile residues impact one another. While CPMG-RD provides a quantitative assessment of chemical shift changes at atomic resolution, the technique does not inform directly on the specific motions being monitored, nor provide a means to determine the mechanism by which regional intercommunication may occur. To address these issues, we utilized a previously calculated ensemble of CypA structures generated through successive in silico heating and cooling steps, guided by chemical shift restraints imparted by the CamShift method.\textsuperscript{125,235} While this method lacks explicit time-step information, and can therefore not directly inform on the timescale of motions, it provides a much broader sampling of conformational space than available in a typical MD simulation.\textsuperscript{205} In fact, this method of utilizing NMR-based restraints has been highly successful in probing the conformational changes that occur within enzymes on the micro-millisecond timescales.\textsuperscript{205,236}

Considering that we have identified an allosteric dynamic coupling of CypA Val 29 to the distal CypA active site through CPMG-RD approaches (Figure 36), we sought to address the underlying mechanism(s). To determine the means by which changes in the active site impart a change in the dynamics experienced by the amide group of Val 29, we first calculated the expected chemical shift of the Val 29 amide group over all structures in the ensemble to identify the specific structural change likely reported on by the Val 29 amide shift. We used a second chemical shift prediction method, SPARTA+,\textsuperscript{237} to back-calculate the expected chemical shift for each structure within the calculated ensemble. As shown in
Figure 38a (black bars), the chemical shift of Val 29 groups into two distinct chemical environments, centered at ~115 ppm and ~120 ppm. By correlating the chemical shift value to the rotameric angles of Val 29 and surrounding residues, we found that the chemical shift change measured at the Val 29 amide reports on the $\chi_1$ angle of Val 29 itself, which adopts two distinct values, corresponding to the two chemical shift values.

To identify putative pathways by which the active site impacts rotation at Val 29, we began by correlating the localization of Val 29 with other residues in the protein over the ensemble of structures. In order to directly compare atom localizations, we used principle component analysis (PCA) to map the three-dimensional position of each atom over the ensemble to a single variable, the first principle component (PC1), which captures the majority of a given atom’s variability in localization. This allowed us to calculate correlation coefficients between PC1 values and/or rotameric angles, which represent the co-occurrence in localization changes over the ensemble, including for weakly coupled residues. We calculated these values for the side chain of Val 29, identifying correlations to the $\chi_1$ angle or localization of $C_{\gamma_1}$ or $C_{\gamma_2}$ (Figure 38b). Three distinct peaks appear, one corresponding to residues nearby in sequence (27-32), as well as two corresponding to residues distant in sequence, but close in tertiary structure (85-87 and 129).

As shown in Figure 38c, the $\chi_1$ angle of Val 29 is correlated to the spatial localization of the benzyl ring of Phe 129. Notably, unlike Val 29 $\chi_1$, Phe 129 does not adopt two distinct states, but rather occupies a continuum of positions, subsets of which are compatible with a given Val 29 $\chi_1$ angle. Iterating this analysis, we find that Phe 129 localization also correlates with the $\chi_1$ angle of Leu 98 in a similar manner. The two Leu 98 $\chi_1$ states sampled in this simulation have previously been identified as major and minor states in X-ray electron density by Fraser et al.\textsuperscript{177} who also suggested a collective motion of Leu 98 with Ser 99 and Phe 113. For both Ser 99 and Phe 113, the previously identified minor states are likewise
sampled and the localization in these states appears to be correlated, but not collectively so. The previously identified minor state of 99 (dotted line circles in Figure 38c) is only sampled when both Leu 98 and Phe 113 are in the minor states identified by Fraser et al.; both Phe 113 and Leu 98, however, also sample these minor states independent of changes in Ser 99 $\chi_1$ (Figure 38c).

Experimentally, we can also identify this non-coherent coupling by comparing measured minor-state population occupancies of Leu 98 and Phe 113. Applying the same type of analysis as in Figure 38a, we find that the amide of Ser 99 likely reports on the rotameric state ($\chi_1$) of Leu 98 and that the amide of Phe 113 likely reports on the rotameric state ($\chi_1$) of Phe 113 (Figure 39). Utilizing a suite of mutants that alter minor state population occupancies, we used the Carver-Richards equation to calculate the minor state population occupancy for both Leu 98 and Phe 113 for each mutant. As shown in Figure 38d, the populations are positively correlated, but they are not identical, with Leu 98 exhibiting about twice the minor state population occupancy as Phe 113 for each of the mutants measured. Linking these partially correlated conformational states, we identify a pathway connecting Phe 113 in the active site to Val 29 (Figure 38e).

Following a similar iterative protocol for the other correlations identified in Figure 36b, we also identified a second putative communication pathway connecting Val 29 ultimately with loops comprising residues 103-109 and 65-75 within the active site (Figure 40). This pathway begins with a partial correlation between the Val 29 $\chi_1$ and loop motions in residues 85-86, which influence the rotameric state of Asp 85. Asp 85 is further correlated to motions in the active site loop comprising residues 103-109 which is likewise correlated to motions in a second active site loop comprising residues 65-75. The step-wise elucidation of the second pathway is outlined in Figure 40 and the residues involved in the pathway are mapped along with the residues involved in the first pathway onto the CypA structure in
Figure 38f. We have thus identified two putative pathways by which changes in active site conformational sampling may be propagated non-coherently to a distal region of CypA.

**Identification of allosteric pathways via CPMG-RD measurement on multiple single site mutations**

To experimentally probe the pathways identified above, we generated multiple mutations to Val 29. Only the well conserved mutations to alanine (CypA\textsuperscript{V29A}), threonine (CypA\textsuperscript{V29T}), and leucine (CypA\textsuperscript{V29L}) yielded stable, well-folded proteins. \(\text{C}_\alpha\), \(\text{C}_\beta\), and \(\text{N}_\text{H}\) chemical shifts were also determined for CypA\textsuperscript{V29L} and compared to WT (see \(^{15}\text{N}\)-HSQC spectra for all V29 mutants and \(\text{C}_\alpha\), \(\text{C}_\beta\), and \(\text{N}_\text{H}\) chemical shifts for CypA\textsuperscript{V29L} in Figure 41). The spectra exhibit minimal perturbations outside of the immediate vicinity of the mutation, indicating that no large-scale structural changes have been imposed by these mutations.

To determine whether conformational sampling had been perturbed globally in CypA by mutation to Val 29, we generated \(^2\text{H}\)\(^{15}\text{N}\)-labeled versions of each mutant and collected \(^{15}\text{N}\)-CPMG-RD experiments. As shown in Figure 42, \(\mu\text{s-}\text{ms}\) dynamics are altered for nearly every residue in the protein that exhibits exchange. Furthermore, the change in exchange falls into two distinct groupings: one for which the alanine mutation has no effect, but CypA\textsuperscript{V29T} and CypA\textsuperscript{V29L} exhibit comparable increases in \(R_{\text{ex}}\) over CypA\textsuperscript{WT} (Figure 42, Response 1) and a second group for which CypA\textsuperscript{V29A} and CypA\textsuperscript{V29T} exhibit comparable small increases in \(R_{\text{ex}}\) over CypA\textsuperscript{WT} and CypA\textsuperscript{V29L} induces a much larger increase over CypA\textsuperscript{WT} (Figure 42, Response 2). Within Response 2, rates of motion, determined from individual fits to the Carver-Richards equations, range from 700 s\(^{-1}\) to >3000 s\(^{-1}\); given the consistent pattern of \(R_{\text{ex}}\) alteration across the residues within the group, these data suggest that a shift in sampled minor state populations is propagated across residues moving in a non-coherent fashion.
By mapping residues exhibiting each response pattern onto the CypA structure, two clear clusters emerge, corresponding closely to the two pathways identified in Figure 38, providing experimental evidence in support of the proposed allosteric pathways. Two additional allosterically communicating regions are apparent in the experimental data: motions in the active site \(3_{10}\) helix comprising residues 119-124 that exhibit the pattern of Response 1 and the active site residues of \(\beta\)-sheets 3-4, composed of residues 54-62 that exhibit the pattern of Response 2. Further analysis of the ensembles reveals that localization of loop 70-75 is partially correlated with the localization of active site residues His 54 and Arg 55, providing a link between residues 54-62 and the previously identified Response 2 (Figure 43a). As no apparent link exists between the \(3_{10}\) helix and the localization of Val 29 in the MD ensemble generated for CypA\(^{WT}\), a chemical shift guided MD ensemble was generated for CypA\(^{V29L}\), following an identical protocol as for CypA\(^{WT}\). As shown in Figure 43b, the bulkier valine substitution at residue 29 sterically blocks the formation of a hydrogen bond between the backbone carbonyl of Arg 28 and the side chain amide of Asn 87, permitting more flexibility in the motions of residues Asn 87 and Phe 88, which appear to disrupt the hydrophobic packing around the \(3_{10}\) helix. The observed chemical shift changes in the helix appear to correspond to a partial unfolding, with broken hydrogen bonds between the backbones of 119 and 123 (Figure 43c). This mechanism is consistent with the observed responses to different mutations; alanine in place of valine would likely not interfere with the 28/87 hydrogen bonding, and therefore not influence the unfolding rate of the helix, while the polar threonine hydroxyl group is likely repelled from the hydrophobic valine binding pocket sufficiently to disrupt the 28/87 bond as does the bulky leucine.

**Integration of the influence of multiple active site distal mutations on CypA dynamics**

We have previously shown that combining multiple single site mutations within the active site of CypA leads to disparate responses in different regions of the protein, including
additive responses and responses in which the influence of one or the other dominates, consistent with the localized nature of CypA dynamics.\textsuperscript{173} We sought to replicate this effect utilizing residues distal from the active site in order to develop an understanding of how long range allosteric pathways may interact and may in turn affect function (probed below). Following the same approach as in Figure 36, we thus identified a second active site distal residue, Val 6, for which CPMG-RD measured dynamics on both the amide nitrogen and methyl carbons are altered by addition of substrate (see residue position in Figure 36, dynamics data in Figure 37 and Figure 44). Unlike Val 29, no observed changes appear in the dynamics of Val 6 upon mutation of the active site in the absence of substrate, suggesting that any allosteric linkage to the active site may exist only in the presence of substrate. As such, we performed a similar pathway analysis, beginning at Val 6, in an ensemble of structures previously generated with the FGP peptide bound to the active site; both \textit{cis} and \textit{trans} bound isoforms are included in the ensemble.\textsuperscript{125} As shown in Figure 44b, the chemical shift changes in the Val 6 amide appear to result from Val 6 itself sampling a minor state rotamer. In the peptide-bound ensemble, the rotameric state of Val 6 is correlated to the position of $\alpha$-helix 1 (residues 30-41), which is further coupled to the conformation of the active site loop comprising residues 65-75, apparently through interactions with the side chain of Phe 67 (Figure 44c/d). This active site loop interacts with the N-terminus of the peptide that rotates in the isomerization reaction;\textsuperscript{238} the conformational state of this active site loop is highly dependent on the peptide isomer bound in the ensembles, resulting in Val 6 only sampling its minor state in those structures bound to the \textit{cis} isoform, in which $\alpha$-helical position samples compatible conformations.

As with Val 29, only conservative mutations to Val 6 are tolerated, with alanine, isoleucine, serine, threonine, or leucine mutations yielding stable proteins. Of these, the serine (CypA\textsuperscript{V6S}), threonine (CypA\textsuperscript{V6T}), and leucine (CypA\textsuperscript{V6L}) were selected, based on their
functional influence on the enzyme (discussed further below and in Table 8), for $^2\text{H}^{15}\text{N}$-labeling followed by collection of CPMG-RD data on the free proteins. The mutations led to variable alterations to global dynamics, with no obviously distinct response signatures as with the Val 29 mutants, perhaps owing to the pathway passing through the localization of a full helix, which may enact multiple interaction outcomes. However, by mapping those residues that are affected by one or more of the mutations to Val 6 and those that are unaffected, an obvious pattern emerges in which the conformational sampling of the two active site loops (103-109 and 65-75) are impacted, as are β-sheets 3-4 (Figure 45a). These findings are consistent with the proposed allosteric pathway from Val 6 that impacts sampling in loop 65-75 as shown in Figure 44d.

A set of two mutations was selected from the panel of single-site mutations at Val 6 and Val 29 based on the functional outcomes of the mutants, discussed in the next section. We generated the $^2\text{H}^{15}\text{N}$-labeled double mutant CypA$^{V6T V29L}$ and measured dynamics by CPMG-RD. As we had previously shown for double mutants in the active site, we found localized signatures in response to the combined perturbations to the protein’s allosteric network. As shown in Figure 45b, the group of residues associated with the $3_{10}$ helix were not influenced by single mutation of Val 6, and subsequently not influenced by the V6T mutation in the context of the V29L mutant (i.e. the change in exchange induced in CypA$^{V29L}$ is identical to that in CypA$^{V6T V29L}$). A second group of residues (Figure 45c), located predominantly in the loops adjacent to $\alpha$-helix 1, experience increased exchange in the context of either CypA$^{V6T}$ or CypA$^{V29L}$ and further increased exchange in CypA$^{V6T V29L}$, corresponding to an additive effect. In the final group of residues (Figure 45d), comprising much of the active site, the V6T mutant partially or completely inhibits the effect of V29L, such that exchange is reduced in CypA$^{V6T V29L}$ relative to CypA$^{V29L}$. 
Active site distal residues influence substrate binding and catalysis

While correlative measures of enzyme dynamics and linebroadening methods that monitor substrate turnover have suggested a role for μs-ms dynamics in the catalytic cycle of CypA, here we show a direct connection by altering CypA motions through mutagenesis with concomitant changes to catalysis. To determine if the active site distal mutants generated above impact CypA function, we assayed CypA WT as well as each of the mutants for both binding and isomerization activity towards the FGP peptide substrate. Binding affinity for the peptide was measured via monitoring chemical shift changes to 15N-labeled proteins upon serial addition of unlabeled peptide substrate. Binding affinity is different for the cis and trans isoforms and they are in fast exchange in the bound form, so the apparent binding affinity (K_D-app) is a combination of the two. Because isomerization is a reversible reaction, catalysis was monitored by 15N-labeling the peptide substrate, adding catalytic concentrations (20 μM) of unlabeled enzyme, and monitoring isomerization directly on the substrate via ZZ-exchange as previously described. The effective isomerization rate (k_iso) is not a direct measure of on-enzyme catalysis, but is influenced by on-enzyme exchange as well as substrate binding and release. As listed in Table 8, mutation of these residues leads to a range of influences on substrate affinity, from minimal impact (for V29T, V6A, V6I, V6S, V6L), to an approximately 50% increase in affinity (for V29L). The mutants exhibit a range of effective isomerization rates as well, ranging from at 35% increase (for V29T) to a 35% decrease (for V29L) in turnover. The double mutant CypA V6TV29L exhibits affinity comparable to the more tightly binding of the two single mutants, CypA V29L, and a modestly reduced isomerization rate relative to CypA V29L.

We additionally measured the binding affinity of CypA WT and Val 29 mutants via isothermal titration calorimetry (ITC). When utilizing ITC, we systematically measure binding affinities (K_D-app-ITC) of about two fold weaker than by NMR titration, likely due the two-species nature of the substrate and the different physical parameters being measured by
NMR and ITC (i.e. the heat given off by the bulk combination of *cis* and *trans* binding as measured by ITC does not directly correspond to the fraction of bound CypA as measured by NMR). Nonetheless, the measured $K_{D-app-ITC}$ values indicate a similar change in affinity for the peptide as when measured by NMR titration, with CypA$^{V29T}$ exhibiting a minimal difference and CypA$^{V29A}$ and CypA$^{V29L}$ exhibiting increased affinity (Table 9).

To determine the degree to which the observed functional effects are substrate specific, we identified another peptide substrate with little sequence similarity to the FGP peptide. We discovered that CypA$^{WT}$ binds the HA tag sequence, YPTDVPDYA, originally identified in influenza hemagglutinin.\textsuperscript{239} Although previous work has shown that host CypA acts as a restriction factor for influenza infection through interactions with the viral M1 protein,\textsuperscript{240} our discovery here is the first to reveal other potential roles of CypA in influenza infection that may warrant further studies. An NMR titration of the peptide into $^{15}$N-labeled CypA$^{WT}$ indicated that the binding is in intermediate exchange for most residues, indicating tighter binding than to the FGP peptide and precluding accurate measurement of binding affinity by NMR titration (Figure 46a). Binding was thus measured by ITC and the $K_{D-app-ITC}$ for CypA$^{WT}$ was found to be 56 µM (see Figure 46c for a representative ITC trace), more than twice as tight as the FGP peptide utilized elsewhere in this study, which was optimized in a phage display screen. Given the well-established involvement of CypA in infection by multiple viruses, including influenza, this relatively tight binding affinity suggests a potential biological role for this interaction. Additionally, given this finding, an alternative affinity tag should be utilized in future biological experiments in which cyclophilins are involved to avoid potential complications. For the current study, the HA peptide was modified by mutation of Pro 2 to an Ala (HA$^{P2A}$), as ZZ-exchange experiments showed that CypA only catalyzed isomerization of Pro 6; the spectrum of HA is significantly complicated by the presence of up to four peaks (*cis* and *trans* for each Pro) for each residue; and the binding affinity is minimally impacted by the mutation (see Figure 46b for the HA$^{P2A}$ $^{15}$N-HSQC spectrum and
assignments, Table 10 for binding data). Binding affinity and isomerization rates of HA$^{P2A}$ were measured for CypA$^{WT}$, CypA$^{V29A}$, CypA$^{V29T}$, CypA$^{V29L}$, CypA$^{V6T}$ and CypA$^{V6TV29L}$. As shown in Table 10, $K_{D-app-ITC}$ and $k_{iso}$ are impacted in a generally similar manner by these mutations in HA as for the FGP peptide, indicating that these mutations influence enzyme function generally, and not in a manner limited to the FGP peptide. Some notable variation does exist between the substrates, most notably in the catalytic activity of CypA$^{V29T}$, which increases $k_{iso}$ for the FGP peptide, but decreases it for the HA peptide, and in the minimal impact of the V6T mutation on the binding affinity towards the HA peptide. This variability indicates that, while altering conformational sampling does influence enzymatic function in CypA towards multiple substrates, the specific outcomes are somewhat substrate specific, likely due to the differential interactions and dynamic pathways within each enzyme:substrate complex.

**Lineshape analysis suggests that altered conformational sampling regulates substrate binding**

The CypA catalytic cycle comprises, minimally, a four state reversible reaction mechanism, illustrated Figure 47a. Because the substrate exists in a pre-formed equilibrium, the individual rate constants cannot be readily determined by monitoring substrate depletion or product formation. Lineshape analysis has emerged as a powerful method to extract out the individual rate constants in reversible enzymatic reactions like prolyl-isomerization. NMR lineshapes are exquisitely sensitive to exchange rates in the µs-ms timescale. By collecting high resolution lineshapes of the $^{15}$N labeled peptide substrate at multiple concentrations of enzyme and fitting those data along with measured values of $K_{D-app}$ and $k_{iso}$, the individual rate constants can be extracted as we have previously described. Lineshape data were collected and microscopic rate constants were determined for CypA$^{WT}$, CypA$^{V29T}$, CypA$^{V29L}$, CypA$^{V29A}$, CypA$^{V6T}$, and CypA$^{V6TV29L}$ (Table 11). For each mutant, 200 fits were performed from randomly chosen initial conditions to ensure that a breadth of parameter space was
sampled; the 20 best fits, as determined by the highest coefficients of determination, were taken as representative of the solution for each mutant.

As shown in Table 11, the on-enzyme interconversion rates ($k_{BC}$ and $k_{CB}$) are poorly defined. Because the off-rates ($k_{BA}$ and $k_{CD}$) are much slower than the rates of on-enzyme interconversion, the lineshapes are minimally impacted by on-enzyme interconversion rates. Within these large uncertainties, conclusions about the impact of mutation on the on-enzyme interconversion are difficult to make. However, trends are apparent in the on and off rates of both the trans and cis isoforms of the substrate. Focusing specifically on the Val 29 mutants, both CypA$^{V29T}$ and CypA$^{V29L}$ exhibit a modest increase in the trans-peptide on-rate, and each of CypA$^{V29A}$, CypA$^{V29T}$, and CypA$^{V29L}$, exhibits an increase in cis on-rate. These increases in on-rates parallel the increase in conformational exchange observed via CPMG-RD for each of these mutants (i.e. CypA$^{V29L}$ exhibits overall greater exchange than CypA$^{V29T}$, which exhibits greater exchange than CypA$^{V29A}$, Figure 47), suggesting the inherent conformational sampling of CypA may be a limiting factor in substrate binding. CypA$^{V6T}$, however, exhibits elevated cis and trans on-rates relative to CypA$^{WT}$ with relatively modest alterations to the µs-ms conformational landscape, indicating that mutation at this site may bring about its functional effects through another mechanism, perhaps another timescale of motion or particular side-chain mobility that is minimally observable to the amide probes. Another notable functional outcome of mutagenesis is that the V29L mutation imparts a notable decrease in cis peptide off-rate ($k_{CD}$) alone or in the context of the V6T mutation, implicating the conformational fluctuations as relevant to substrate release in CypA as well as in substrate binding.

Summary

Cyclophilin A exhibits segmental motions on the µs-ms timescale. While motions on this timescale have been shown to influence protein function within a range of macromolecular systems, most well described systems on this timescale have been
described as globally coherent motions such that most or all probes are reporting on a single structural transition. The segmental nature of dynamics in CypA provides particular challenges as compared to globally coherent motions, namely that the specific localized structural changes reported on by CPMG-RD are difficult to identify and that, with multiple dynamic processes occurring simultaneously, the specific process being reported on by a given probe must be determined. Furthermore, a full description of the dynamic landscape of CypA must include mechanisms of intercommunication between these distinct segmental processes.

To address these challenges, we have developed an approach that combines the atomic level precision afforded by MD structural ensembles with experimental validation provided by the novel mutagenic approach of introducing multiple single site mutations at a distal residue found to be allosterically linked to the active site. The ensemble of structures utilized herein lacks an inherent timescale, and so to ensure that the ensemble does sample known minor states of the protein, we compared the states sampled to those identified by Fraser et al.'s powerful room temperature X-ray study of CypA. Their study identified low electron density contours corresponding to minor states of several active site and active site adjacent residues, including Leu 98, Ser 99, and Phe 113. As demonstrated in Figure 38, these states are sampled in the ensembles, demonstrating that the ensembles are conformationally diverse enough to encompass experimentally observed minor states. Further, we have expanded our understanding of these known minor states by using chemical shift-based approaches that monitor the solution behavior, as opposed to these previous studies that have observed a subset in the crystal. By doing so, we have demonstrated that the conformational states of the residues are indeed coupled (i.e. Ser 99 is only able to access its minor state when Phe 113 is also in its minor state), but not coherently so (i.e. Phe 113 can sample its minor state while Ser 99 remains in its major state). By further searching for atomic localization and rotameric states that are partially
conformationally coupled to others, we have identified multiple pathways of communication in CypA which allow non-coherent coupling through several mechanisms, including side-chain rotameric states coupled to the side-chain or backbone localization of adjacent residues, rotameric states of adjacent residues with partially correlated states, and partially coupled loop conformational states (Figure 38c, Figure 40, and Figure 43, respectively).

To experimentally probe these coupled networks of residues, we identified multiple active site distal residues that are coupled to active site dynamics (Figure 36 and Figure 44), and generated multiple mutations at each site, allowing us to identify distinct response signatures corresponding to specific allosteric pathways. These mutations represent the first instances in which the conformational sampling of the CypA active site has been altered independent of changes to the ground state active site structure. Fraser et al. did shift active site sampling via an S99T mutation, but this mutation sterically forced Phe 113 fully into its minor state, necessarily altering the ground state of the active site. The functional changes associated with the Val 29 mutations generated herein (Table 8) definitively demonstrate a link between conformational sampling in CypA and progression through the enzymatic cycle. These mutations impact both the substrate binding and release steps, with the cis conformation of the peptide more strongly impacted. We have previously identified, in a thermophilic cyclophilin, a dynamic clamping mechanism for which motions in both the free protein and the protein:substrate complex are relevant for substrate binding. The findings here indicate a role for µs-ms dynamics in CypA as well, both in the conformational sampling involved in forming a complex with the substrate and in permitting substrate release.
Table 8. Apparent dissociation constants and isomerization rates of cyclophilin mutants towards the FGP peptide

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$K_{D-app}$ (µM)$^a$</th>
<th>$k_{iso}$ (s$^{-1}$)$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>76±3</td>
<td>10.4±0.3</td>
</tr>
<tr>
<td>V29A</td>
<td>57±2</td>
<td>8.9±0.2</td>
</tr>
<tr>
<td>V29T</td>
<td>73±3</td>
<td>12.3±1.1</td>
</tr>
<tr>
<td>V29L</td>
<td>38±2</td>
<td>6.9±0.3</td>
</tr>
<tr>
<td>V6A</td>
<td>82±5</td>
<td>10.1±0.2</td>
</tr>
<tr>
<td>V6I</td>
<td>86±4</td>
<td>10.2±0.1</td>
</tr>
<tr>
<td>V6S</td>
<td>65±3</td>
<td>8.9±0.2</td>
</tr>
<tr>
<td>V6T</td>
<td>48±2</td>
<td>9.1±0.1</td>
</tr>
<tr>
<td>V6L</td>
<td>69±3</td>
<td>9.4±0.1</td>
</tr>
<tr>
<td>V6TV29L</td>
<td>37±3</td>
<td>6.1±0.2</td>
</tr>
</tbody>
</table>

$^a$Errors in $K_{D-app}$ are in fits to a single experiment  
$^b$Errors in $k_{iso}$ are standard deviations of 2 or more independent experiments

Table 9. Apparent dissociation constants as measured by ITC

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$K_{D-app-ITC}$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>133±5$^a$</td>
</tr>
<tr>
<td>V29A</td>
<td>102±4</td>
</tr>
<tr>
<td>V29T</td>
<td>127±0.4</td>
</tr>
<tr>
<td>V29L</td>
<td>98±4</td>
</tr>
</tbody>
</table>

$^b$Errors are standard deviations of 2 or more independent experiments

Table 10. Apparent dissociation constants and isomerization rates of cyclophilin mutants towards the HA$^{PZ2A}$ peptide

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>$K_{D-app-ITC}$ (µM) $^a$</th>
<th>$k_{iso}$ (s$^{-1}$) $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>HA</td>
<td>56</td>
<td>-</td>
</tr>
<tr>
<td>WT</td>
<td>HA$^{PZ2A}$</td>
<td>58±2$^a$</td>
<td>4.2±0.1$^b$</td>
</tr>
<tr>
<td>V29A</td>
<td>HA$^{PZ2A}$</td>
<td>45</td>
<td>3.7</td>
</tr>
<tr>
<td>V29T</td>
<td>HA$^{PZ2A}$</td>
<td>50</td>
<td>3.7</td>
</tr>
<tr>
<td>V29L</td>
<td>HA$^{PZ2A}$</td>
<td>39</td>
<td>3.3</td>
</tr>
<tr>
<td>V6T</td>
<td>HA$^{PZ2A}$</td>
<td>54</td>
<td>3.9</td>
</tr>
<tr>
<td>V6TV29L</td>
<td>HA$^{PZ2A}$</td>
<td>35</td>
<td>3.3</td>
</tr>
</tbody>
</table>

$^a$Standard deviation from three independent experiments, all other ITC measurements were only conducted once  
$^b$Standard deviation from two independent experiments, all other isomerization measurements were only conducted once
Table 11. Best-fit solutions of microscopic rate constants for CypA and mutants

<table>
<thead>
<tr>
<th></th>
<th>CypA^{WT}</th>
<th>CypA^{V29A}</th>
<th>CypA^{V29T}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{AB}$ ($\times 10^6$ s^{-1}M^{-1})</td>
<td>3.8±0.6</td>
<td>4.0±0.7</td>
<td>5.1±1.0</td>
</tr>
<tr>
<td>$k_{BA}$ (s^{-1})</td>
<td>560±150</td>
<td>530±250</td>
<td>690±210</td>
</tr>
<tr>
<td>$k_{BC}$ (s^{-1})</td>
<td>1400±800</td>
<td>1400±600</td>
<td>1500±700</td>
</tr>
<tr>
<td>$k_{CB}$ (s^{-1})</td>
<td>1400±1600</td>
<td>1400±1500</td>
<td>1600±1000</td>
</tr>
<tr>
<td>$k_{CD}$ (s^{-1})</td>
<td>220±130</td>
<td>190±130</td>
<td>280±160</td>
</tr>
<tr>
<td>$k_{DC}$ ($\times 10^6$ s^{-1}M^{-1})</td>
<td>10.0±1.2</td>
<td>11.5±0.8</td>
<td>12.3±0.8</td>
</tr>
<tr>
<td>$K_{D-cis}$ (µM)</td>
<td>150±50</td>
<td>130±50</td>
<td>140±40</td>
</tr>
<tr>
<td>$K_{D-trans}$ (µM)</td>
<td>22±13</td>
<td>16±10</td>
<td>23±14</td>
</tr>
<tr>
<td>Bound cis/trans</td>
<td>1.1 +/- 0.7</td>
<td>2.1 +/- 1.0</td>
<td>1.1 +/- 0.7</td>
</tr>
</tbody>
</table>

$\omega$^{trans-bound} D6 (ppm) 8.48±0.06 8.53±0.15 8.50±0.09
$\omega$^{cis-bound} D6 (ppm) 8.62±0.17 8.86±0.87 8.55±0.16
$\omega$^{trans-bound} L7 (ppm) 8.25±0.01 8.24±0.02 8.25±0.01
$\omega$^{cis-bound} L7 (ppm) 8.27±0.01 8.30±0.05 8.27±0.01

<table>
<thead>
<tr>
<th></th>
<th>CypA^{V29L}</th>
<th>CypA^{V6T}</th>
<th>CypA^{V6T/V29L}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{AB}$ ($\times 10^6$ s^{-1}M^{-1})</td>
<td>4.8±0.7</td>
<td>5.4±1.2</td>
<td>4.2±1.1</td>
</tr>
<tr>
<td>$k_{BA}$ (s^{-1})</td>
<td>500±140</td>
<td>480±160</td>
<td>410±240</td>
</tr>
<tr>
<td>$k_{BC}$ (s^{-1})</td>
<td>2100±1200</td>
<td>1200±800</td>
<td>1600±1000</td>
</tr>
<tr>
<td>$k_{CB}$ (s^{-1})</td>
<td>1000±400</td>
<td>1200±600</td>
<td>1000±500</td>
</tr>
<tr>
<td>$k_{CD}$ (s^{-1})</td>
<td>100±20</td>
<td>190±80</td>
<td>110±90</td>
</tr>
<tr>
<td>$k_{DC}$ ($\times 10^6$ s^{-1}M^{-1})</td>
<td>12.8±0.6</td>
<td>13.7±1.0</td>
<td>11.7±0.9</td>
</tr>
<tr>
<td>$K_{D-cis}$ (µM)</td>
<td>100±30</td>
<td>90±30</td>
<td>90±30</td>
</tr>
<tr>
<td>$K_{D-trans}$ (µM)</td>
<td>8±2</td>
<td>14±7</td>
<td>9±6</td>
</tr>
<tr>
<td>Bound cis/trans</td>
<td>2.1 +/- 1.0</td>
<td>1.1 +/- 0.6</td>
<td>2.1 +/- 1.3</td>
</tr>
</tbody>
</table>

$\omega$^{trans-bound} D6 (ppm) 8.49±0.12 8.50±0.08 8.52±0.12
$\omega$^{cis-bound} D6 (ppm) 8.54±0.09 8.60±0.51 8.50±0.81
$\omega$^{trans-bound} L7 (ppm) 8.24±0.01 8.25±0.01 8.24±0.01
$\omega$^{cis-bound} L7 (ppm) 8.27±0.00 8.27±0.01 8.28±0.03

^Standard deviation determined from the 20 best fits to the data
Figure 36. Identification of an active site distal residue coupled to active site dynamics. (a) CypA structure bound to the peptide substrate, from a previously generated MD ensemble. Relevant residues are highlighted in green, the peptide is shown in black, with the isomerized proline in magenta. (b) CPMG-RD data (dots) collected at 900 MHz for on CypA<sub>WT</sub> alone (black, solid lines), CypA<sub>F113A</sub> alone (red), CypA<sub>K82A</sub> alone (blue) and CypA<sub>WT</sub> (black, dotted line) in the presence of a saturating concentration (6 mM) of the model substrate, as measured on the amide nitrogen of Val 29. Lines are determined from single residue linear least squares fitting to the Carver-Richards equation of data collected at 900 and 600 MHz.

Figure 37. Active site distal methyl CPMG data alone and with FGP peptide. <sup>13</sup>C-CPMG-RD data collected on selectively methyl labeled CypA<sub>WT</sub> alone (black) or in the presence of a saturating (6 mM) concentration of the model peptide (red) as measured for Val 6 and Val 29. Dots represent data while lines represent best-fit solutions to the Carver-Richards equations.
Figure 38. Identification of putative allosteric pathways between the CypA active site and Val 29 in an MD ensemble
Figure 38. Identification of putative allosteric pathways between the CypA active site and Val 29 in an MD ensemble (a) Top panel, Val 29 samples two rotameric states in the MD ensembles, plotted in blue and red. Bottom panels, the predicted chemical shift of the Val 29 amide nitrogen over all structured in the MD ensemble (black), structures corresponding to Val 29 $\chi_1$ state 1 (blue), and structures corresponding to Val 29 $\chi_1$ state 2 (red). (b) Maximum squared correlation coefficient for Val 29 $\chi_1$, $C_{\gamma 1}$, or $C_{\gamma 2}$ compared to the PC1 values for all other heavy atoms in CypA. Peaks correspond to residues that are partially coupled to the rotameric state of Val 29 and/or the localization of the $C_{\gamma 1}$ or $C_{\gamma 2}$. (c) For each pair of atoms analyzed, the rotameric angle or PC1 values are plotted against one another for all structures in the ensemble. Means are shown as solid lines. Structures are grouped, based on k-means clustering, to 2 or 3 clusters (red, blue, and green) based on the conformational state of the residue on the x-axis. Representative structures from each cluster are shown next to each correlation graph. (d) CPMG-RD data were simultaneously fit for CypA$^{WT}$ and the 6 mutants listed using both 600 MHz and 900MHz data, with a single $\Delta \omega$ value, but allowing $k_{ex}$ and $P_B$ to vary between mutants. Calculated $P_B$ for the amide of Phe 113 is plotted against $P_B$ for the amide of Ser 99 (which is representative of the rotameric state of Leu 98, see Figure 39). Error bars are representative of fit errors. (e) Putative allosteric pathway stemming from active site residue Phe 113 to Val 29. Rates were determined via single residue fitting to the Carver-Richards equations using 900 MHz and 600 MHz data for CypA$^{WT}$. (f) Two allosteric pathways (red and blue), identified here and in Figure 40, by which active site perturbations may propagate to Val 29.
Figure 39. Conformational transitions reported on by Ser 99 and Phe 113 Top panels, rotameric states of the $\chi_1$ angles of Leu 98 and Phe 113 throughout the ensemble of structures. Bottom panels, predicted chemical shifts of the amide nitrogens of Ser 99 and Phe 113 for all structures (black), for those corresponding to state 1 above (red), and for those corresponding to state 2 above (blue). States were determined via k-means clustering. Ser 99 and Phe 113 shifts are clustered based on the $\chi_1$ angle of Leu 98 and Phe 113, respectively.
Figure 40. A second putative allosteric pathway between Val 29 and the active site (a) The Val 29 $\chi_1$ angle is correlated to the localization of the Ser 32 sidechain and the backbone of residues 85 and 86. (b) The $\chi_1$ angle of Asp 85 is correlated to conformation of loop 103-109. (c) The two major active site loops are partially coupled such that loop 65-75 is much more tightly constrained when loop 103-109 is in the state colored blue than when in the state colored red.
Figure 41. Val 29 mutants are well folded and do not alter the global structure of CypA
Top panels, $^{15}$N-HSQC spectra collected on CypA$^{WT}$ (black), CypA$^{V29T}$ (green), CypA$^{V29A}$ (blue), and CypA$^{V29L}$ (red). Bottom panels, C$_\alpha$, C$_\beta$, and N$_H$ chemical shifts determined for CypA$^{WT}$ (black) and CypA$^{V29L}$ (red).
Figure 42. Experimental identification of allosteric communication pathways in CypA
CPMG-RD data collected at 900 MHz (dots) and best fit solutions to the Carver Richards
equations, solved for each residue individually using data collected at 600 and 900 MHz
(lines) for CypA$^{WT}$ (black), CypA$^{V29T}$ (green), CypA$^{V29A}$ (blue), and CypA$^{V29L}$ (red). Two
representative residues are shown for each of the two response patterns. Residues
following response pattern 1 are colored in magenta, those following response pattern 2 are
colored in cyan. Only residues for which data was available for CypA$^{WT}$ and all V29L
mutants are included.
Figure 43. Additional allosteric coupling stemming from Val 29 (a) The conformation of loop 65-75 is coupled to the $\chi_1$ rotameric state of His 54. His 54 samples three distinct states, but is restricted to the state at $\sim 50^\circ$ in one loop conformation (blue), and samples the other two angles ($\sim 200^\circ$ and $\sim 300^\circ$) when the loop samples an alternative conformation. The states of His 54 $\chi_1$ and Arg 55 $\chi_1$ are coupled such that the major state sampled by Arg 55 is $\sim 190^\circ$ when His 54 is in the state colored blue, but shifts to $\sim 300^\circ$ when His 54 samples its other rotameric states. (b) CypA$^{V29L}$ (red) occludes the hydrogen bond that forms in CypA$^{WT}$ between the side chain amine proton of Asn 87 and the carbonyl of Arg 28, disrupting the hydrophobic packing around the active site 3-10 helix. (c) The active site 3-10 helix samples a partially unfolded state (State 2), represented by the Thr 119 $\psi$ angle, which is reported upon by the amide chemical shift of Leu 122. Leu 122 is in slow exchange, which allows only for measurement of $k_s$, the rate of exchange to a minor state, which was determined to be $2.3 \pm 0.3$ s$^{-1}$ and $4.7 \pm 0.3$ s$^{-1}$ for CypA$^{WT}$ and CypA$^{V29L}$, respectively, by fitting to the slow exchange equation described by Tollinger et al.$^{116}$
Figure 44. Identification of a second active site distal location and characterization of its communication pathway to the active site (a) CPMG-RD data collected for CypA WT alone (solid line) or in the presence of a saturating concentration (6 mM) of the FGP peptide (dashed line) as measured at the amide nitrogen of Val 6. (b) Top, $\chi_1$ distribution of Val 6 in the ensemble of peptide-bound structures. Bottom, predicted chemical shift values for Val 6 N. In both panels, blue and red represent the two states of the $\chi_1$ angle, based on k-means clustering. (c) Pairs of atom PC1 or rotameric angle, plotted against one another for all structures in the peptide-bound ensemble. Red and blue dots represent structure for which the peptide is in the cis and trans conformation, respectively. (d) Representative structures illustrating the proposed allosteric pathway from an active site loop to Val 6. Red and blue structures are bound to the peptide in the cis and trans conformation, respectively.
Figure 45. Dynamic response to Val 6 mutants and double mutant (a) CPMG-RD data collected at 900 MHz (dots) and best-fit solutions to the Carver-Richards equations for CypA<sup>WT</sup> (black), CypA<sup>V6L</sup> (green), CypA<sup>V6T</sup> (blue), and CypA<sup>V6S</sup> (red). Representative residues are shown for which CPMG-RD measured dynamics are altered or not altered by mutation to Val 6. Residues for which dynamics are altered (cyan) or not altered (magenta) are indicated on the structure. (b-d) CPMG-RD data collected at 900 MHz (dots) and best fit solutions to the Carver Richards equations for CypA<sup>WT</sup> (black), CypA<sup>V29L</sup> (red), CypA<sup>V6T</sup> (blue), and CypA<sup>V6TV29L</sup> (green). Representative residues are shown for which the V6T mutation has no influence on the dynamics (b), for which the two mutants exhibit additive exchange (c), or for which the V6T mutation dampens or blocks the impact of the V29L mutation (d). Residues exhibiting each of the signatures are illustrated in structures.
**Figure 46. Characterization of the HA and HA$^{P2A}$ peptide**

(a) NMR titrations of the FGP peptide (top) and HA peptide (bottom) into 0.5 mM $^{15}$N CypA$^{WT}$, demonstrating fast exchange for the FGP peptide and intermediate exchange for the HA peptide. Peptide concentrations are 0 mM (red), 0.1 mM (orange), 0.2 mM (yellow), 0.5 mM (green), 1 mM (blue), 2 mM (purple). (b) $^{15}$N-HSQC spectrum of HA$^{P2A}$ with peaks assigned. All minor peaks correspond to the peptide in the cis conformation. (c) Representative raw ITC data (left), integrated heat values (right, squares), and fit to the data (right, line), with HA$^{P2A}$ titrated into CypA$^{WT}$.
**Figure 47. Peptide on-rates mirror CPMG-RD monitored shifts in dynamics** Top, on-rates for the cis and trans peptide isoforms as determined by lineshape analysis for CypA\textsuperscript{WT} and Val 29 mutants. Error bars represent standard errors, * = p-value < 0.05 as compared to CypA\textsuperscript{WT}. Bottom, additional examples of residues exhibiting Response 1 (left) and Response 2 (right) to Val 29 mutants, plotted as in Fig. 3.
DNA encoding the 146 amino acid GeoCyp was commercially purchased (Genewiz, South Planfield, NJ) and subsequently cloned into the pET15b plasmid encoding an N-terminal 6xHis-tag for expression in E. Coli strain BL21 (DE3). Cell pellets were lysed via sonication in nickel buffer (50 mM Na2HPO4, pH 7.5, 500 mM NaCl, 10 mM imidazole), applied to a 10-20 ml Ni-sepharose column, and eluted with nickel elution buffer (50 mM Na2HPO4, pH 7.5, 500 mM NaCl, 400 mM imidazole). Virtually all of the expressed GeoCyp was found to be soluble upon sonication. Elutions comprising 6xHis-tagged GeoCyp were then concentrated to approximately 2-3 ml, the thrombin tag was cleaved off with 3 units of thrombin incubated overnight at room temperature, and the protein was applied to a superose-75 sizing column equilibrated in NMR buffer (50 mM Na2HPO4, pH 6.5 with 1 mM DTT and 1mM EDTA). Elutions comprising GeoCyp were concentrated to approximately 0.5-1.5 mM with 95% H2O/5% D2O for subsequent NMR analysis.

CypA Purification

CypA was expressed and purified as previously published. Briefly, cells were lysed in 25 mM MES, pH 6.1, with 1 mM DTT, purified over an SP Sepharose exchange column, dialyzed into 50 mM Tris, 1 mM EDTA pH 6.8 and flowed through a Q Sepharose column, followed by size exclusion chromatography in NMR buffer.

CypB Purification

The cyclophilin domain of CypB (residues 39-216) was purified following a similar protocol to CypA. Briefly, for CypB, cells were lysed in lysis buffer (50 mM Tris, 50 mM NaCl, pH 7.0, with 1 mM DTT), purified over an SP Sepharose exchange column, dialyzed back
into the lysis buffer and flowed through a Q Sepharose column, followed by size exclusion chromatography in NMR buffer.

**CypC Purification**

The cyclophilin domain of CypC (residues 25-212) was purified following an identical protocol to GeoCyp. Briefly, cells were lysed and purified over a nickel affinity column, the 6xHis tag was removed with thrombin, followed by size exclusion chromatography in NMR buffer.

**Protein Deuteration**

To deuterate proteins, 50 ml of $^{2}\text{H}^{15}\text{N}$ or $^{2}\text{H}^{15}\text{N}^{13}\text{C}$ M9 were inoculated with 2 ml of cells spun down from a saturated LB growth. After 24 hours, 1 L of $^{2}\text{H}^{15}\text{N}$ or $^{2}\text{H}^{15}\text{N}^{13}\text{C}$ M9 was inoculated with 50 ml starter and allowed to grow overnight at room temperature, with temperature increased to 37 °C in the morning. Cells were induced at an OD600 of 0.5, and cells spun down after 4 hours. To allow for complete amide proton exchange, for all deuterated proteins, pellets were lysed in 5 M Guanidine HCl, 100 mM Tris, 100 mM EDTA, pH 7.5, then dialyzed for 24 hours into 1 M Arginine, 100 mM Tris, 100 mM NaCl, pH 7.5, then into the nickel equilibration buffer (GeoCyp, CypC) or SP equilibration buffer (CypA, CypB). From this point, all purifications proceeded as before.

**Peptide purification**

The GSFGPDLRAGD substrate (FGP peptide) was expressed in a pet30 vector with an N-terminal fusion to the small solubility tag GB1, a 6xHis tag, and a thrombin cleavage site. Cells were lysed in nickel buffer and purified over a nickel column. The elution was dialyzed back into nickel buffer, followed by cleavage of the thrombin tag overnight at RT. The protein was again passed over a Ni column to remove GB1, followed by reverse-phase HPLC. Pure peptide was identified in the HPLC elution by MALDI Mass Spectrometry, lyophilized, and resuspended in water. Peptide concentration was determined by amino acid analysis. Re-lyophilized peptide was dissolved in 50 mM PO₄, 1 mM EDTA, pH 6.5 and
brought to pH 6.5. The GSW peptide was purified following an identical protocol as the FGP peptide, but quantified by absorbance at an OD of 280 nm. The HA and HA\textsuperscript{P2A} peptides were purified following an identical protocol above, but with cleavage by the enzyme Factor Xa instead of thrombin. The TA-peptide substrate was synthetically generated by WuXi AppTec, further purified via reverse phase HPLC, and resuspended in 50 mM PO\textsubscript{4}, 1 mM EDTA, pH 6.5.

**NMR assignments**

**GeoCyp assignments**

\textsuperscript{1}H, \textsuperscript{13}C, and \textsuperscript{15}N resonances were assigned for GeoCyp at 25\textdegree C using a combination of 2-D and 3-D experiments collected on a Varian 600 MHz spectrometer acquired using a \textsuperscript{13}C\textsuperscript{15}N-labeled sample (for review and primary references see Sattler et al.\textsuperscript{241}). Backbone sequential assignments included a CBCA(CO)NH, HNCACB, HNCO, 3D \textsuperscript{15}N-edited and \textsuperscript{13}C-edited NOESY-HSQC experiments (both NOESY experiments employed a \(t_{\text{mix}} = 150\) ms). Aliphatic assignments were made using an HCCH-TOCSY, a TOCSY-HSQC (\(t_{\text{mix}} = 60\) ms), a 3-D HNHA and a 3D HNHB. Due to significant, specific interactions between DSS and GeoCyp, all spectra were referenced to the HDO resonance.

**CypC assignments**

Isotopically \textsuperscript{2}H\textsuperscript{13}C\textsuperscript{15}N-labeled CypC was purified, followed by collection of \textsuperscript{15}N-HSQC, HNCACB, HN(CO)CA, and HN(COCA)CB experiments at 25\textdegree C on an Agilent DD2 800 MHz spectrometer with a cryogenically cooled probe, allowing near-complete backbone sequential assignments to be determined (BMRB 25341).

**Peptide assignments**

The FGP and HA\textsuperscript{P2A} peptides were assigned (Table 2 and Figure 46) by collecting \textsuperscript{15}N-HSQC, \textsuperscript{13}C-HSQC, and HH-TOCSY experiments on \textsuperscript{13}C\textsuperscript{15}N-labeled peptide at 25\textdegree C using a Varian 600 MHz spectrometer, with assignments at other temperatures determined by following amide peak positions.
Assignments for cyclophilins bound to peptide: To determine amide peak assignments in the bound form for CypA (BMRB 25337), CypB (BMRB 25338), and GeoCyp (BMRB 25336) at 10°C, peaks were followed during titration of the model peptide. Additionally, HNCACB and CBCA(CO)NH experiments were collected on \(^{13}\text{C}\text{^{15}}\text{N}\)-labeled proteins, using a either a Varian 600 MHz spectrometer or an Agilent DD2 800 MHz spectrometer in the presence of saturating concentrations of the model peptide to resolve assignment of any ambiguous peaks. For CypC, titrations in both temperature and with the peptide substrate, along with a \(^{15}\text{N}\)-NOESY-HSQC experiment, allowed assignment of the amide resonances of CypC at 10°C in the bound form (BMRB 25339).

All data were processed using NMRPipe\(^{242}\) and analyzed using CCPNmr.\(^{243}\)

**NMR solution structure determination**

\(^{15}\text{N}\)-edited and \(^{13}\text{C}\)-edited NOESY experiments were collected on isotopically \(^{13}\text{C}\text{^{15}}\text{N}\)-labeled GeoCyp and used with previously determined chemical shift assignments\(^{187}\) to identify long-range interactions. Chemical shifts were analyzed using TALOS+\(^{244}\) and used to guide Rosetta\(^{245}\) fragment analysis. NOESY peak assignments were analyzed using CYANA 2.1,\(^{246}\) and converted to Rosetta constraint format using the CS-Rosetta Toolkit [www.csrosetta.org]. The RASREC-Rosetta algorithm\(^{247}\) was then used to calculate an ensemble of 20 lowest scoring structures. This was run using the Janus supercomputing cluster at the University of Colorado, employing Message Passing Interface (MPI) over 528 CPUs. Violation analysis of the resulting ensemble of structures was performed using PDBStat\(^{248}\) and PSVS analysis [psvs.nesg.org]. Electrostatic potentials were determined using the APBS web server\(^{249}\) [www.poissonboltzmann.org].

**Circular dichroism monitored thermal denaturation**

Samples were prepared at 38 µM CypA or 100 µM GeoCyp in 50 mM Na\(_2\)HPO\(_4\), pH 6.5. CD readings were measured from 30-70 °C for CypA or 30-70 °C for GeoCyp at 1
°C/min at 222 nm on a Jasco J-810 spectropolarimeter. To obtain mid-point values, data were least-squares fit to a variable slope sigmoidal curve.

**NMR titrations for measuring binding affinity**

Binding affinity was measured by NMR titration. $^{15}$N-HSQC spectra were collected on 500 µM $^{15}$N-labeled protein in the presence of 0, 0.1, 0.2, 0.5, 1, and 2 mM peptide substrate on a Varian 900 MHz spectrometer. For peaks with significant chemical shift changes upon titration, chemical shift changes were least squares fit individually to the steady state equilibrium binding equation below.

Equation 15

$$F([L]) = F_{\text{max}} - \frac{[P] + [L] + K_D}{\sqrt{([P] + [L] + K_D)^2 - 4[P][L]}} - F_{\text{max}}$$

Where $F([L])$ is the ligand dependent chemical shift change, $F_{\text{max}}$ is the chemical shift change upon full saturation, $[P]$ is the total protein concentration, $[L]$ is the total ligand concentration, and $K_D$ is the dissociation constant. All chemical shifts that could be fit well individually ($r^2 > 0.99$, indicating they are in the fast exchange regime needed to accurately calculate binding affinity) were then fit simultaneously, yielding a single dissociation constant determined for each protein.

**ZZ-exchange to measure isomerization rate**

Isomerization was measured via the ZZ-exchange NMR experiment. 1 mM isotopically $^{15}$N-labeled peptide substrate was mixed with 1 or 20 µM protein, depending on temperature used (20 µM for 0-20°C, 1µM for 30-45°C). Data were collected with mixing times of 0, 0.036, 0.072, 0.144, 0.24, 0.3, 0.36, 0.54, 0.72, 0.9, 1.08, and 1.2 s on a Varian 600 MHz spectrometer and fit, via linear least squared fitting, to the equations described by Farrow et al. The trans state of the peptide was found to comprise 89% of the total peptide by measuring peak intensities in the absence of enzyme. Cis, trans, and both exchange peaks of LEU 7 and cis, trans, and one exchange peak (the other is overlapped
with another residue) of ASP 6 were simultaneously fit to determine $k_{iso}$ values. Longitudinal relaxation was found to be nearly identical for both cis and trans conformations of both ASP 6 and LEU 7, so a single longitudinal relaxation parameter was used. For CypA$^{R55A}$, CypA$^{R55A/A103R}$, GeoCyp$^{R47A}$, and GeoCyp$^{R47A/R92A}$, 1 mM $^{15}$N-labeled peptide was mixed with 100 µM protein and data were collected at 30°C using the mixing times of 0, 0.192, 0.384, 0.576, 0.768, 1.014, 1.152, and 1.344 s.

NMR relaxation experiments.

$^{15}$N-CPMG-RD

$^{15}$N-TROSY CPMG-RD experiments were collected on 1 mM deuterated, isotopically $^{15}$N-labeled CypA, CypB, CypC, or GeoCyp on a Varian 900 MHz spectrometer with a cryogenically cooled probe and on a Varian 600 MHz spectrometer. Data were collected at 0°C, 10°C, 20°C, and 30°C for CypA, as indicated, using constant time relaxation periods of 50, 60, 80, and 90 ms. Data were collected at 0°C, 10°C, 20°C, and 30°C for GeoCyp, using constant time relaxation periods of 60, 70, 90, and 100 ms. Data were collected at 10°C for both CypB and CypC using constant relaxation periods of 30 and 40 ms, respectively. Data were also collected in the presence of substrate with 0.5 mM $^2$H$^{15}$N CypA, CypB, CypC, or GeoCyp and 6 mM peptide substrate, 6 mM TA-peptide substrate, or 0.7 mM CsA. The $R_2$ relaxation rate was calculated as in Equation 4$^{250}$ Exchange parameters were determined by least-squares fitting to the Carver-Richards equations, which describe generalized two state exchange (Equation 3).$^{114}$

$R_1$ relaxation

$R_1$ relaxation was measured on 0.5 mM $^{15}$N CypA, CypB, CypC, or GeoCyp, using mixing times of 10, 30, 50, 70, 90, and 110 ms on a Varian 900 MHz spectrometer with a cryogenically cooled probe or a Varian 600 MHz spectrometer.
Molecular dynamics ensembles

Ensembles of GeoCyp and CypA_{V29L} were generated as previously described for CypA.\textsuperscript{125} Briefly, the bound-cis and bound-trans states of GeoCyp were modeled starting from the free state of GeoCyp determined in this work.\textsuperscript{125} The two bound states were then simulated using the Amber99SB*-ILDN force field in explicit TIP3P water for 100ns each at 300K, the two final structures where then used as the starting structures for a chemical shift and NOE replica-averaged restrained simulation.\textsuperscript{125,205,251-253} CamShift was used to back-calculate the chemical shifts from both replicas at each time step. The force constant for the chemical shifts restraints was set to 5.2 kJ/mol, and the force constant for the NOEs was set to 250 kJ·mol\textsuperscript{-1}·nm\textsuperscript{-2} with a bottom flat potential that is zero between 0.3 and 0.5 nm.\textsuperscript{235} Each replica has been evolved through a series of annealing cycles between 300 K and 450 K (100 ps at 300 K, 100 ps during which the temperature increased linearly up to 450 K, 100 ps of constant-temperature molecular dynamics at 450 K, and 300 ps during which the temperature decreased linearly to 300 K). Each replica has been evolved for a total nominal time of 150 ns. The final ensembles comprise all the 300 K structures sampled by both replicas after the first 50 ns. The averaged NOE and chemical shift restraints were added to GROMACS by using PLUMED2 and ALMOST.\textsuperscript{254-256}

Electric field calculations

The electric field in the active site of GeoCyp has been calculated with Gaussian 03 as previously described\textsuperscript{125} using quantum mechanical calculations on the center-of-mass of the GLY-PRO peptide bond, using the partial charges of the force field for the whole GeoCyp. The z-, x- and y- components of the field where defined, following previous work\textsuperscript{125} as the normal to the ring plane defined by the N, C\textalpha{}, and C\textgamma{} atoms of the proline residue; the GLY-C'-'N-PRO peptide bond; and the normal to the such defined xz plane, respectively. The electric field calculated for CypA with the partial charges is in remarkable agreement
with that calculated *ab-initio*. Indeed, the average difference between the two methods is ~1 MV/cm.

**Bacillaceae sequence alignment**

All Bacillaceae cyclophilins in the NCBI RefSeq database [www.ncbi.nlm.nih.gov/refseq] for which a full species name was indicated were included in the analysis. Alignment was performed using Clustal Omega sequence alignment program.\textsuperscript{257}

**Isothermal titration calorimetry**

ITC data were collected using 0.5 mM protein and titrating in up to a 2 molar ratio of either the FGP peptide, the HA peptide, or the HA\textsuperscript{P2A} peptide. Data were collected on a MicroCal iTC200 and fits were performed using MicroCal ITC-ORIGIN.

**Lineshape analysis**

Lineshape data were collected and fit as described in Chapter III. For all fits in Chapter III, spectra were referenced to the proton position of the *trans* amide peak of Gly 10. For all fits in Chapter V, 100 µM of the ^15^N-labeled peptide GSW were added to each sample, and spectra were aligned by reference to one of the amide peaks.

**CPMG-RD simulations**

For each model run, $10^5$ individual atoms were simulated. Atoms were distributed among the three states depicted in Figure 28 as given by the equilibrium concentrations and begun with coherent magnetization. Magnetization was allowed to evolve over a 20 ms simulation time, while stochastically transitioning between states as proscribed by the microscopic rate constants. *trans*-bound and *cis*-bound chemical shifts were arbitrarily set to 0.5 ppm and 1 ppm downfield of the free enzyme chemical shift at 21.1 T, respectively, for all simulations. Refocusing pulses were applied with $\nu_{cpmg}$ of 50-1000 s\textsuperscript{-1} by taking the complex conjugate of the magnetization. The absolute value of the sum of the magnetization all atoms after the simulation time yields the peak intensity, and $R_{ex}$ values are calculated as
for measured data, using Equation 4, for which \( I_0 \) is set to the sum of magnetizations before evolution.
CHAPTER VII
FUTURE DIRECTIONS

We have directly demonstrated the role of non-coherent conformational dynamics in regulating the enzymatic cycle in CypA (Chapter V) and provided evidence that dynamics play a role in regulating function across the cyclophilin family (Chapters II and III). Additionally, we have generated a number of complimentary approaches to studying the mechanisms of allosteric coupling in non-coherent systems, using CypA as a model system.

**Alternative regulation of CypA dynamics**

We have generated multiple mutants that alter the active site dynamics and catalytic activity of CypA in vitro. These studies have been fruitful in mapping pathways of allosteric communication in CypA and in developing a mechanistic understanding of the mechanisms by which this communication occurs. The studies also raise the possibility of utilizing these of active site distal regions for in vivo targeting, either of native CypA or of CypA engineered to be targeted.

A tightly binding and effective inhibitor exists for CypA, cyclosporine A, which targets the binding site of the majority of human cyclophilins with a low nanomolar dissociation constant and complete ablation of target binding and isomerization. Here, however, we have identified multiple mutants, which increase affinity for multiple substrates and correspondingly decrease the rate of turnover. These attributes may permit the disentanglement of the relative roles of binding and prolyl isomerization for each of the many biological roles of CypA, which are currently poorly understood and cannot be addressed using cyclosporine or previously generated active site mutants, which reduce both binding and catalysis. Of particular interest would be an inducible system in which a shift in binding and catalysis could be transiently induced via addition of a small molecule or light, allowing for temporal investigation of the role of CypA binding and catalysis in various capacities including in protein folding, in viral infectivity, and in signal transduction.
This approach could be achieved through two complimentary methods. *In silico* based screening for small molecules has significant limitations, but given the identification of multiple distal sites within CypA that are coupled to the dynamic motions within the active site, the *in silico* approach may permit direct targeting of these sites for potential small molecule binding. While the particular functional outcome of binding would currently be difficult to predict, we would expect alteration to active site dynamics upon shifting the conformational sampling of these distal sites via binding. Promising *in silico* hits would need to be validated in their binding to CypA and probed for their influence on both active site dynamics and on the isomerization cycle.

In addition to searching for small molecule binding to CypA<sup>WT</sup>, we also propose that CypA can be engineered to exhibit shifted functionality upon small molecule addition. Specifically, numerous protein systems have been rendered inducible by light or small molecule via introduction of a second, inducible protein domain into an active site distal loop coupled to the active site. An appealing candidate for this approach is a recently engineered fusion of an FK506-binding protein (FKBP12) and the FKBP12-rapamycin binding protein (FRB), dubbed uniRapR. FRB binds FKBP12 with high affinity only in the presence of the small molecule rapamycin and stabilizes FKBP12 significantly. Incorporation of the uniRapR fusion protein into an active site distal loop has been shown to introduce significant conformational mobility into the active sites of multiple kinases, ablating activity; addition of rapamycin stabilizes the construct, returning activity. Utilizing this system, we imagine incorporation of uniRapR into one of the two active site distal loops in CypA, shown here to be coupled to active site dynamics. This incorporation would be expected to shift the active site conformational landscape, and likely enzymatic function, of CypA upon addition of rapamycin, allowing inducible alteration to CypA binding and catalysis. The uniRapR construct is designed to be incorporated into a flexible loop via linkers of variable length; therefore, a number of insertion sites and linkers would likely need to be screened in order
to identify a stable construct for which dynamics are nonetheless allosterically translated across the linkers.

Identification of either a small molecule that binds CypA\(^{WT}\) or development of an inducible fusion protein would allow for direct probing of the allosteric pathways in a single enzyme, as opposed to the comparative analysis performed here. Additionally, either small molecule based approach would allow for probing of the role of CypA in biological systems with high temporal resolution (i.e., observing the direct effects of shifting the balance of binding and catalysis in CypA without the secondary, compensatory effects associated with longer term overexpression of a given mutant).

**Application to other protein systems**

As conformational dynamics have only been characterized for a relatively small number of systems, the extent to which non-coherent, localized motions versus globally coherent motions exist and are functionally relevant throughout all proteins is currently not well characterized. As such, we would like to apply the methods described herein to additional, unrelated protein systems to determine the degree to which the findings we have presented are more generally applicable to other systems.

As a beginning to this approach, we have focused some preliminary work on a functionally related group of enzymes for which dynamics is necessarily involved in the enzymatic process, namely DNA endonucleases and glycosylases for which flipping bases out of the DNA helix is required for damage recognition and base excision.\(^{262,263}\) We have thus generated and purified \(^{15}\)N-labeled homologous DNA-3-methyladenine glycosylase II from humans (AAG), *E. coli* (AlkA), and *G. kaustophilus* (DMG2), as well Endonuclease V from *E.coli* (Endo V). As shown in Figure 48, each of these proteins yields a well-dispersed spectrum that should be amenable to further characterization. Comparison of dynamics among multiple glycosylases may also permit us to examine the degree of conservation of dynamics among another protein family, as we have done for the cyclophilins.
In addition to the enzymes mentioned above, we would like to apply the methods developed in this manuscript to additional systems, with a particular focus on proteins for which known allosteric regulation has been previously characterized, yet the particular molecular mechanism of allosteric communication is unknown. A number of significant hurdles must be overcome to apply the approach outlined in Chapter V to any novel systems, including the glycosylases and endonuclease described above. Any proteins must be expressible in *E. coli* to permit labeling and be able to be refolded to permit amide exchange after deuteration. A protein must be sufficiently stable and small enough to be amenable to NMR spectroscopy (~30 kD or less for $^{15}$N-dynamics experiments, although some larger proteins may be accessible to $^{13}$C-based experiments). Proteins must also exhibit motions on the µs-ms timescale and not self-associate, which is an issue even for multiple members of the cyclophilin family, as shown in Figure 20 and Figure 24. Due to these significant restrictions on potential systems, the development of a new system in which to apply these approaches will likely involve screening a number of potential candidates in order to identify one that satisfies the requirements. However, for those systems that are amenable to the NMR approaches outlined here, putative pathways of allosteric communication could be identified as for CypA. Given the role of allosteric regulation in a large number of protein systems, identification of these pathways could reveal novel sites to be targeted in protein engineering or by structure-based drug design. Additionally, the MD ensemble based approach is more broadly applicable, and could be used independently of the CPMG-RD approach to understand allosteric coupling in non-coherent systems.
Figure 48. $^{15}$N-HSQC spectra of AAG, AlkA, DMG2, and EndoV
REFERENCES


(114) Carver, J.P. and Richards, R.E. A general two-site solution for the chemical exchange produced dependence of T2 upon the carr-Purcell pulse separation. Journal of Magnetic Resonance 6, 89-105 (1972).


(157) Mulder, F.A.; Hon, B.; Mittermaier, A.; Dahlquist, F.W., and Kay, L.E. Slow internal dynamics in proteins: application of NMR relaxation dispersion spectroscopy to...


(171) Bhabha, G.; Ekiert, D.C.; Jennewein, M.; Zmasek, C.M.; Tuttle, L.M.; Kroon, G.;
Dyson, H.J.; Godzik, A.; Wilson, I.A., and Wright, P.E. Divergent evolution of protein
(2013).

(172) Eisenmesser, E.Z.; Bosco, D.A.; Akke, M., and Kern, D. Enzyme dynamics during

(173) Schlegel, J.; Armstrong, G.S.; Redzic, J.S.; Zhang, F., and Eisenmesser, E.Z.

(174) Venters, R.A.; Farmer, B.T., 2nd; Fierke, C.A., and Spicer, L.D. Characterizing the use
of perdeuteration in NMR studies of large proteins: 13C, 15N and 1H assignments of


(176) Bosco, D.A.; Eisenmesser, E.Z.; Clarkson, M.W.; Wolf-Watz, M.; Labeikovsky, W.;
Millet, O., and Kern, D. Dissecting the microscopic steps of the cyclophilin A

(177) Fraser, J.S.; Clarkson, M.W.; Degnan, S.C.; Erion, R.; Kern, D., and Alber, T. Hidden
(2009).

(178) van den Bedem, H.; Bhabha, G.; Yang, K.; Wright, P.E., and Fraser, J.S. Automated
identification of functional dynamic contact networks from X-ray crystallography. *Nat

(179) Li, G. and Cui, Q. What is so special about Arg 55 in the catalysis of cyclophilin A?

(180) Agarwal, P.K.; Geist, A., and Gorin, A. Protein dynamics and enzymatic catalysis:
investigating the peptidyl-prolyl cis-trans isomerization activity of cyclophilin A.

(181) Agarwal, P.K. Role of protein dynamics in reaction rate enhancement by enzymes. *J

(182) Ramanathan, A.; Savol, A.J.; Langmead, C.J.; Agarwal, P.K., and Chennubhotla, C.S.

(183) Doshi, U.; McGowan, L.C.; Ladani, S.T., and Hamelberg, D. Resolving the complex
role of enzyme conformational dynamics in catalytic function. *Proc Natl Acad Sci U S


APPENDIX A

DESCRIPTION OF COMPUTER CODE

All code has been deposited online at https://github.com/mjholliday

Fit_cpmg_batch.exe

A routine allowing fitting of CPMG RD data from one or multiple residues to the Carver Richards equation, the slow limit equation, and/or the fast limit equation. Run by:

>Fit_cpmg_batch.exe Example.inp

Where Example.inp is an input file as described below

Input File Format

Each set of data is defined to be associated with a given atom and group
- Each atom has a single dw
- Each group has a single kex and PA

Multiple line types are allowed, at least one valid data line and an output line are required. Lines not beginning with line type keywords will be ignored

Line type: data file

Line format: D_FILE data_file_# atom_# group_# filename res_num field_strength temperature atom_type

Line example: D_FILE 3 2 2 ./cpmg_data.txt 12 900 10 N

Line description: Input any number of data files. Each line of each data file should list a vcpmg frequency in Hz and a R2 value. For each file, specify the file number (must be different for all data files), atom number, and group number. Program will fit an R0 for every data file, a dw for every atom, and a PA & kex for every group. Field strength is given in MHz, for proton.

Line type: initial conditions

Line description: Entering initial conditions is optional, will default to ic’s listed below for any that are not specified set is the data_file_# for parameter R0, atom_# for parameter dw, or group_# for PA or kex. Boundary conditions are also optional. If one boundary condition is specified, both must be. Fit will initially be attempted w/o boundary conditions, and they will be enforced if fit exceeds boundary conditions.

Line format: IC set param ic boundary_condition_lower(optional) boundary_condition_upper(optional)

Line example: IC 2 kex 1500
Line example: IC 1 dw 2 0.1 5

Line type: option

Line description: Various options, listed below

Line format: OPTION opt
opt may be set to:
- COMPARE_MODEL_FIT: only works if multiple groups are specified, will also calculate fit with all data/atoms a single group, with only one kex/PA then compare the fits using both F-test and AIC
- CALC_SLOW_EQ: will also generate fit to slow equation, and compare to CR using AIC
- CALC_FAST_EQ: will also generate fit to fast equation, and compare to CR using AIC
- FIT_GROUPS_SEPARATE: will run minimization separately for each group (requires no atom to be part of several groups, will return error if so.
- WEIGHTS X: Weights data by -1/(Rex^X), default is no weighting (X=0). Rex is estimated by the difference btw largest and smallest R2 values within a data set. All data for a given atom are scaled by the same value, the data set with the largest Rex.
Line type: output file
Line description: File to which output from fitting routine will be printed
Line format: O_FILE ./output.txt

Line type: print file
Line description: Will print data and fits to an eps file if included
Line format: P_file ./output.eps
Flag –norm will normalize all plots to the same Rex

Example Input Files

Simple_fit.inp:
#Will fit data from a single residue to the Carver-Richards Equation, the slow
#equation, and the fast equation, and compare the fits
D_FILE 1 1 1 ./Res45_10d_900.dat 45 900 10 N
D_FILE 2 1 1 ./Res45_10d_600.dat 45 600 10 N

OPTION CALC_SLOW_EQ
OPTION CALC_FAST_EQ
OPTION WEIGHTS 1

O_FILE ./45_10d.out
P_FILE ./45_10d.eps –norm

Multi_fit.inp:
#Will fit data from a multiple residues to the Carver-Richards Equation and
#compare the fit when fitting each residue separately to the fit when all are
#fit together
D_FILE 1 1 1 ./Res45_10d_900.dat 45 900 10 N
D_FILE 2 1 1 ./Res45_10d_600.dat 45 600 10 N
D_FILE 3 2 2 ./Res55_10d_900.dat 55 900 10 N
D_FILE 4 2 2 ./Res55_10d_600.dat 55 600 10 N

OPTION COMPARE_MODEL_FIT
OPTION FIT_GROUPS_SEPARATE
OPTION WEIGHTS 1

O_FILE ./45_55_10d.out
P_FILE ./45_55_10d.eps -norm

CPMG_SIM.m

An octave function, will simulate CPMG RD trace for simple two-state exchange or
for the full CypA catalytic cycle as described in Chapter IV:

function [dat,vcpmg,track_t,track] = FGP_cpmg_sim(K,Cyp_Conc=1e-3,FGP_Conc=6e-
3,w_bound=[45,90],time_T2=20,ncyc=[2:2:20],SIMPLE_CPMG=0,dead_enzyme=0,Natoms=1e5
)

K = Microscopic rate constants defining the catalytic cycle: [kab, kba, kbc, kcb,
 kcd, kdc] If SIMPLE_CPMG==1, only kbc and kcb will be used
Cyp_Conc = concentration of CypA in M
FGP_Conc = concentration of peptide in M
w_bound = chemical shift difference of trans and cis from unbound: [w_trans, w_cis]
time_T2 = delay time, in ms
ncyc = number of refocusing pulses to be applied, can be a single value or vector
of values
SIMPLE_CPMG = flag, when set to 1, will simulate a simple 2-state exchange, using
the values of kbc and kcb as the forward and reverse rate constants, repectively
dead_enzyme = flag, when set to 1, will simulate data with kbc=kcb=0
Natoms = number of atoms to simulate for each ncyc value
dat = vector of simulated R2 values for each vcpmg (s^-1)
vcpmg = cpmg refocusing frequencies (s^-1)
track = state of atom #1 throughout the simulation
track_t = time corresponding to each track value (ms)

**Fit_Lineshapes.m**

An octave function, will fit lineshape data, along with dissociation constant, isomerization rate, and bound chemical shift values to determine microscopic rate constants $k_{ab}$, $k_{ba}$, $k_{bc}$, $k_{cb}$, $k_{cd}$, and $k_{dc}$ as described in Chapter III. Data are 2D extracted hsqcs comprising lineshapes that will be summed over nitrogen to generate a 1D proton lineshape to fit. Data must be stored in:

```
./dat/MUT/FGPconcentration_CYPconcentration/hsqc_RES#_RESconformation.txt
```

e.g.: ./data/WT/1000_50/hsqc_6_trans.txt

Data folder must contain data for residues 6 and 7, cis and trans, and also contain an ‘hsqc_11_ref.txt’ file that contains reference data for GSW peptide reference peak

Output file is in Octave format, with parameters as listed below

```
function [] = Fit_Lineshapes(MUT, KD, kiso, freq_bound, F_Out, Cyp_Conc, FGP_Conc, iterations, acquisition_time, acquisition_points, zero_fill)

MUT = string protein/mutant name (default, “WT”)
KD = Dissociation constant to fit (M), if set to 0, will not use in fit
kiso = isomerization rate, at 1 mM peptide, 20 uM protein (s^-1), if set to 0, will not use in fit
freq_bound = vector of proton frequencies, in Hz, of bound residues, if set to 0, will not use in fit
F_out = output file (Default “temp.out”)
Cyp_Conc = Cyp concentrations to use, in uM, first value must be free peptide, with Cyp_Conc==0 (default = [0, 5, 10, 20, 5, 10, 20, 50, 100])
FGP_Conc = Peptide concentrations to use, in uM (default = [1000, 500, 500, 500, 1000, 1000, 1000, 1000, 1000])
iterations = number of fits to perform (default = 200)
acquisition_time = collected FID time (in s, default 1)
acquisition_points = number of points acquired in FID (default 14045)
zero_fill = if zero filling is applied, number of points (default 2^16)
```

Output File Parameters (N = number of iterations)

```
PARS_OUT = fit parameters, a 10xN matrix of parameters, with each column corresponding to [kab; kba; kbc; kcb; kcd; kdc; w_6_bound_trans, w_6_bound_cis, w_7_bound_trans, w_7_bound_cis]
r2_OUT = r^2 values for each fit
f_OUT = XxN matrix of fits, where X is the number of data points fit
full_data_vector = linearized 1D proton data set to which fit is performed
data_freq_linear = frequencies corresponding to each data point in full_data_vector
Populations_OUT = frequencies corresponding to each data point in full_data_vector
```

```
Pars_OUT = fit parameters, a 10xN matrix of parameters, with each column corresponding to [kab; kba; kbc; kcb; kcd; kdc; w_6_bound_trans, w_6_bound_cis, w_7_bound_trans, w_7_bound_cis]
r2_OUT = r^2 values for each fit
f_OUT = XxN matrix of fits, where X is the number of data points fit
full_data_vector = linearized 1D proton data set to which fit is performed
data_freq_linear = frequencies corresponding to each data point in full_data_vector
```

```
Pars_OUT = fit parameters, a 10xN matrix of parameters, with each column corresponding to [kab; kba; kbc; kcb; kcd; kdc; w_6_bound_trans, w_6_bound_cis, w_7_bound_trans, w_7_bound_cis]
r2_OUT = r^2 values for each fit
f_OUT = XxN matrix of fits, where X is the number of data points fit
full_data_vector = linearized 1D proton data set to which fit is performed
data_freq_linear = frequencies corresponding to each data point in full_data_vector
```