

STABILITY, STATES, AND INTERACTIONS  
OF THE ENZYME PMM/PGM  
BY NMR AND MULTIVARIATE ANALYSIS

---

A Dissertation

Presented to

the Faculty of the Graduate School  
at the University of Missouri-Columbia

---

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

---

by

JIA XU

Dr. Steven R. Van Doren and Dr. Lesa J. Beamer

Dissertation Supervisors

DECEMBER 2016

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

STABILITY, STATES, AND INTERACTIONS OF THE ENZYME PMM/PGM

BY NMR AND MULTIVARIATE ANALYSIS

presented by Jia Xu,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

---

Professor Steven R. Van Doren

---

Professor Lesa J. Beamer

---

Professor Linda Randall

---

Professor Xiaoqin Zou

---

Professor Peter V. Cornish

## ACKNOWLEDGEMENTS

My graduate research has been conducted under direct supervision of Dr. Steven R. Van Doren. I sincerely thank him for his invaluable guidance and support. He is very patient, knowledgeable, and always there to listen and to give advice. I have been very fortunate to have him as my advisor.

I would like to give my special thanks to my advisor Dr. Lesa J. Beamer for her support, encouragement, and advice. During the past years, she has been offering inspiring ideas and constructive feedbacks. She also spent a lot of time training my presentation skills.

I would like to thank my committee members: Dr. Linda Randall, Dr. Peter Cornish for their time, effort, and helpful suggestions and discussions. I would also like to thank Dr. Xiaoqin Zou for her advice on science, career development, and personal matters.

I would like to thank Matt Stanley for configuring computation environment, and creating a website distributing TREND software. I would also like to thank Brett Maland for his help with copyright of TREND.

I would like to thank Dr. Shaokai Jiang and Dr. Wycoff Wei for the assistance of using NMR spectrometers.

I also owe gratitude to my coworkers of PMM project. I would like to thank Dr. Akellas Sarma for helpful discussions and conducting NMR titrations of wildtype PMM/PGM; Yirui Wei for conducting NMR titrations of S108C mutant; Alex Hopkins for his help in protein preparation and training me laboratory skills; Dr. Yingying Lee for valuable discussions and conducting biochemical assay including ANS binding and

limited proteolysis; Dr. Ritcha Mehra-Chaudhary for helpful discussions; Kyle Stiers for testing TREND software and providing valuable feedbacks.

I am also grateful for the present and past Van Doren lab members: Dr. Yan Fulcher, Tara Marcink, Dr. Yingchu Zhao, Todd Byrne, Dr. Stephen Prior, and Rama Krishna Koppiseti: Thanks for the good time both inside and outside the lab, I really enjoy every moment we have spent together.

I would also like to thank Dr. Xin Liu and Zhijian Luan for their decades-long friendship with my family, their generous help makes my life in Columbia much easier.

Finally, I would like to thank my family. Thanks to my parents for their unwavering support for so many years. Thanks to my son Lingwen Xu for bringing pure, unbridled joy to my life. Thanks to my wife Chen Chen for her love, understanding, and sacrifice. Without their encouragement and blessings, I would not have been able to finish this work.

This work was supported by NSF grant MCB1409898.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	ii
LIST OF FIGURES .....	xii
LIST OF TABLES .....	xvii
LIST OF ABBREVIATIONS.....	xviii
ABSTRACT.....	xxi
CHAPTER I: Introduction .....	1
I.1    Phosphomanomutase / Phosphoglucomutase (PMM/PGM).....	1
I.2    Technical Background of Nuclear Magnetic Resonance .....	3
I.2.1    NMR basics .....	3
I.2.2    Source of NMR signals.....	3
I.2.3    Time-domain and frequency-domain NMR signal detection .....	4
I.2.4    Chemical shift.....	5
I.2.5    Bloch equations .....	6
I.2.6    One-dimensional NMR experiment.....	6
I.2.7    Multi-dimensional NMR .....	7
I.3    Multivariate analysis .....	9
I.4    References .....	9

CHAPTER II: Phosphorylation in the Catalytic Cleft Stabilizes and Attracts Domains of a Phosphohexomutase .....	12
II.1 Abstract .....	12
II.2 Introduction .....	13
II.3 Materials and methods .....	16
II.3.1 Enzyme preparation .....	16
II.3.2 Accessibility measurements.....	17
II.3.3 NMR spectroscopy.....	17
II.3.4 Rapid HX by NMR .....	17
II.3.5 Quantification of rates of HDX.....	18
II.3.6 Electrostatic field calculations .....	19
II.3.7 Modeling HX protection from MD simulations .....	19
II.4 Results .....	20
II.4.1 Phosphorylation decreases accessibility to probes and enhances stability .	20
II.4.2 Effects of dephosphorylation on NMR spectra.....	22
II.4.3 Slowing of HX by phosphorylation, especially in D3 and adjacent face of D4 .....	23
II.4.4 Phosphorylation attracts D4, while dephosphorylation repels D4.....	28
II.4.5 MD suggests compaction due to phosphorylation and more freedom without it.....	29

II.4.6	Phosphorylation effects on simulated backbone mobility and HX.....	31
II.5	Discussion .....	33
II.5.1	Structural compaction by phosphorylation .....	34
II.5.2	Catalytic relevance of compaction by phosphorylation and loosening by dephosphorylation .....	34
II.5.3	Insights into HX by NMR beyond qualitative HDX-MS .....	36
II.5.4	Long-range effects of phosphorylation on HDX of enzymes.....	38
II.6	Conclusions .....	39
II.7	Acknowledgement.....	40
II.8	References .....	40
II.9	Supporting material .....	45
II.9.1	Supporting figures.....	45
II.9.2	Supporting methods .....	52
II.9.3	Supporting tables .....	59
II.9.4	Supporting movies .....	63
II.9.5	Supporting references .....	63
 CHAPTER III: Multiple Ligand-Bound States of a Phosphohexomutase Based On		
	Principal Component Analysis of NMR Peak Shifts.....	66
III.1	Abstract .....	66
III.2	Introduction .....	67

III.3	Material and Methods.....	70
III.3.1	Enzyme preparation, isotopic labeling, and purification. ....	70
III.3.2	NMR spectroscopy.....	71
III.3.3	NMR spectra interpretation.....	71
III.3.4	Measurement of apparent affinity.....	72
III.3.5	Vectors for shifts of NMR peaks .....	72
III.3.6	PCA clustering of spectra perturbed by ligand, dephosphorylation, or mutation .....	72
III.3.7	CONCISE analysis.....	73
III.3.8	Hydrogen exchange analysis.....	73
III.4	Results .....	74
III.4.1	Ligands affect binding site and a domain-domain interface.....	74
III.4.2	X1P and G16P bind comparatively tightly to wild-type enzyme .....	76
III.4.3	Clustering of ligand binding vectors.....	77
III.4.4	Free and X1P-bound states lie at two extremes of conformational equilibria .....	80
III.4.5	Ligand binding slows hydrogen exchange especially in Domain 2 and Domain 4 .....	82
III.5	Discussion .....	85
III.5.1	Highest affinity for the bis-phosphorylated intermediate and an inhibitor. 85	

III.5.2	pSer108 has little effect on ligand affinities, except for the G16P intermediate .....	85
III.5.3	Five distinct states of PMM/PGM identified by CONCISE.....	86
III.5.4	PCA clustering in two and three dimension confirms CONCISE results...	87
III.5.5	CONCISE reveals distinctions between states overlooked by crystallography .....	88
III.6	Summary .....	89
III.7	Acknowledgement.....	89
III.8	References .....	90
III.9	Supporting figures .....	93
CHAPTER IV: Binding Isotherms and Time Courses Readily from Magnetic Resonance .....		100
IV.1	Abstract .....	100
IV.2	Introduction .....	101
IV.3	Experimental section .....	103
IV.3.1	Preprocessing of spectra and images for SVD.....	103
IV.3.2	Extraction of principle components .....	104
IV.4	Results and discussion.....	104
IV.4.1	PCA capture of time courses.....	104
IV.4.2	Fast exchange scenarios.....	105

IV.4.3	Slow exchange scenarios. ....	107
IV.4.4	Intermediate exchange scenarios. ....	109
IV.4.5	Mixtures of regimes. ....	110
IV.4.6	Two-step binding. ....	112
IV.4.7	Nonlinearity and applicability of PCA. ....	113
IV.4.8	Periodic and multiple components from MRI by PCA.....	114
IV.4.9	Tallying meaningful principal components. ....	116
IV.4.10	Limits to applications of PCA to spectra and images. ....	117
IV.4.11	Potential applications to digital data. ....	118
IV.5	Conclusions .....	118
IV.6	Acknowledgement.....	118
IV.7	References .....	119
IV.8	Supporting material .....	123
IV.8.1	Supporting background.....	123
IV.8.2	Supporting experimental section.....	125
IV.8.3	Supporting tables .....	129
IV.8.4	Supporting figures.....	131
IV.8.5	Supporting movies .....	141
IV.8.6	Supporting references .....	142

## CHAPTER V: Tracking Equilibrium and Non-equilibrium Shifts in Data with TREND

.....	143
V.1 Abstract .....	143
V.2 Introduction .....	144
V.3 Methods .....	147
V.3.1 Implementation of TREND.....	147
V.3.2 Conversion of a stack of 2D measurements into a matrix for analysis.....	147
V.3.3 Preprocessing .....	150
V.3.4 Calculating principal components via SVD.....	152
V.3.5 Reconstruction of spectra, images, or movies by PCA.....	153
V.3.6 ICA calculations.....	154
V.4 Results and discussion.....	155
V.4.1 Workflows of TREND .....	155
V.4.2 Examples of ligand binding to two sites detected by NMR.....	157
V.4.3 ICA for confirming components.....	160
V.4.4 Cardiac MRI Movie resolved into components .....	161
V.5 Conclusions .....	163
V.6 Acknowledgments .....	164
V.7 References .....	164
V.8 Supporting material .....	167

V.8.1	Supporting methods .....	167
V.8.2	Supporting figures.....	170
V.8.3	Supporting movies .....	175
V.8.4	Supporting references .....	176
VITA.....		179

# LIST OF FIGURES

## Chapter II

Figure II-1 Catalytic mechanism of PMM/PGM. ....	14
Figure II-2. Phosphorylation effects on accessibility and stability of PMM/PGM .....	21
Figure II-3. Phosphorylation stabilizes or slows HX at more sites than it accelerates. ....	24
Figure II-4. Domains 1 and 4 are electrostatically attracted when Ser108 is phosphorylated, but repelled when it is dephosphorylated.....	28
Figure II-5. Domain 4 opens wider in Apo-deP than Apo-P in MD simulations.. .....	30
Figure II-6. The fluctuations of the MD trajectories provide simulated free energies of hydrogen exchange .....	32
Figure II-7. Expansion of PMM/PGM accompanying its dephosphorylation.....	35
Figure SII-1. Examples of CLEANEX-PM time courses.....	45
Figure SII-2. Kinetics of the apo-P form of PMM/PGM undergoing dephosphorylation	46
Figure SII-3. Effects of (de)phosphorylation on NMR spectra of PMM/PGM.....	47
Figure SII-4. Skyline projections of 2D TROSY spectra of Apo-P before and after dephosphorylation.....	48
Figure SII-5. Dependence of $^{15}\text{N}\{^1\text{H}\}$ heteronuclear, steady-state NOE on phosphorylation status of PMM/PGM.....	48
Figure SII-6. Phosphorylation slows HX and stabilizes domain 3 in particular.....	49
Figure SII-7. Hydrogen exchange results grouped by $\Delta G_{\text{HDX,apparent}}$ suggest the independence of $\Delta G_{\text{HDX,apparent}}$ from size of $k_{rc}$ , implying that the EX2 regime prevails for both Apo-P and Apo-deP .....	49
Figure SII-8. Examples of data fitting of Apo-P.....	50

Figure SII-9. Phosphorylation slows HDX and stabilizes D3 and D4.....	50
Figure SII-10. Radii of gyration ( $R_g$ ) during MD trajectories indicate that the Apo-deP state is more expanded than Apo-P.....	51
Figure SII-11. Correlation of experimentally measured HX protection factors $\ln(P^{exp})$ of staphylococcal nuclease with those simulated from its coordinates.....	51
<b>Chapter III</b>	
Figure III-1. Reactions of PMM/PGM.....	68
Figure III-2. Ligand binding perturbs chemical shifts of wt PMM/PGM residues. ....	75
Figure III-3. Binding affinities of different ligands obtained by protein-observed NMR titrations. ....	77
Figure III-4. Overall effects of ligand binding to PMM/PGM and its S108C mutant.....	78
Figure III-5. Concerted chemical shift changes observed by $^{15}\text{N}$ TROSY spectra .....	80
Figure III-6. Equilibrium shifts of PMM and its complexes characterized by CONCISE analysis of amide resonances .....	81
Figure III-7. X1P binding increases $\Delta G_{HX}$ at more sites than it decreases. ....	83
Figure SIII-1. Calculation of ligand binding CSP vectors used for PCA and CONCISE analysis.....	93
Figure SIII-2. $^1\text{H}$ - $^{15}\text{N}$ TROSY spectra for $E_P$ , $E_{deP}$ , $E_P+G1P$ , and $E_P+G16P$ .....	94
Figure SIII-3. $^1\text{H}$ - $^{15}\text{N}$ TROSY spectra for $E_P+G6P$ , $E_P+M6P$ , $E_P+X1P$ , and S108C mutant .....	95
Figure SIII-4. $^1\text{H}$ - $^{15}\text{N}$ TROSY spectra for S108C+G1P, S108C+G16P, S108C+G6P, and S108C+X1P. ....	96
Figure SIII-5. Ligand binding perturbs chemical shifts of S108C mutant residues .....	97

Figure SIII-6. Scree plot of Fig. III-4 .....	98
Figure SIII-7. Equilibrium shifts of PMM and its complexes characterized by CONCISE analysis of amide resonances .....	98
Figure SIII-8. Comparison of $\Delta G_{HX}$ between $E_P$ and $E_P+X1P$ complex shows ligand-induced stabilization.. .....	99
Figure SIII-9. Putative alternative binding site of S108C+G16P .....	99
<b>Chapter IV</b>	
Figure IV-1. PC1 from SVD of titrations in fast exchange, simulated or measured, represents Langmuir binding isotherms.....	106
Figure IV-2. SVD of titrations featuring slow exchange, in simulated or measured NMR spectra, distills binding isotherms as PC1.....	108
Figure IV-3. Suppressing the intermediate exchange distortion of binding isotherms by applying PCA directly to spectra .....	110
Figure IV-4. Principal components from SVD of spectra agree with the populations estimated earlier by line shape analysis for a titration of two sequential binding events	112
Figure IV-5. SVD extracts the time courses of pulsation in MRI movies .....	115
Figure SIV-1. Chemical exchange regimes of 1:1 ligand binding.....	131
Figure SIV-2. Algorithm developed to apply SVD to series of spectra, images (movies), or lists.....	132
Figure SIV-3. Capture of main time courses of change by PCA of movies or NMR spectra .....	133
Figure SIV-4. Accuracy and precision of binding isotherms can be enhanced by preprocessing with scaling.....	134

Figure SIV-5. Chemical shift changes between apo and bound states of the simulated 2D spectra. ....	135
Figure SIV-6. The routine success in deriving binding isotherms from the intermediate exchange regime results from standard digital resolution and Pareto scaling .....	136
Figure SIV-7. The nonlinearity of the examples is suggested by analyses of their principal components.. ....	138
Figure SIV-8. SVD extracts binding isotherms suitable for fitting in titrations in mixed exchange regimes and regimes that are artificially homogeneous in intermediate exchange .....	139
Figure SIV-9. PCA of free induction decays (FIDs) captures the correct binding isotherms in fast and slow exchange regimes .....	140
Figure SIV-10. Recognizing significant PCs.....	141
<b>Chapter V</b>	
Figure V-1. Work flows to TREND extraction of the main components.....	148
Figure V-2. TREND implementation of PCA (SVD), ICA, and reconstruction. ....	151
Figure V-3. TREND identifies two components in titrations of two binding sites .....	158
Figure V-4. SVD captures from a cardiac MRI movie the time courses of breathing, diastole and systole from a “four-chamber” angle of view.....	162
Figure SV-1. Peak pick lists from the simulated spectra of Fig. V-3B were reconstructed by TREND and plotted as circles.....	170
Figure SV-2. SVD of the unprocessed FIDs from the two site binding titrations identifies two components similar to those from the spectra or the peak lists. ....	171
Figure SV-3. Reconstruction with eight PCs using <i>trendreconstructgui</i> .....	172

Figure SV-4. ICA extracts binding components nearly identical to those from PCA when the number of components is chosen correctly..... 173

Figure SV-5. Validating the number of significant principal components using ICA. The examples have two components..... 174

Figure SV-6. PC4 from the cardiac MRI movie and its scree plot..... 175

# LIST OF TABLES

## Chapter II

Table SII-1. Behaviors of amide TROSY NMR peaks of Apo-P monitored by HDX and the appropriate data fitting method implemented. ....	59
Table SII-2. NMR identification of residues with HX altered by Ser108 phosphorylation, clarifying behaviors within peptides obscured in HDX-MS results .....	60
Table SII-3. Comparison of one and two sets of coefficients in modeling HX protection factors from surrounding contacts and hydrogen bonds for well-characterized proteins .	62
Table SII-4. Comparison of HX protection factors measured and simulated from MD trajectories.....	63

## Chapter III

Table III-1. Perturbation library.....	70
--	----

## Chapter IV

Table SIV-1. Spectral simulation parameters .....	129
Table SIV-2. Contributions of the first 4 principle components of all PCA results. ....	130

## Chapter V

Table V-1. Executable file components of TREND. ....	145
Table V-2. File formats read and reconstructed by TREND .....	149
Table SV-1. Types of biophysical measurements that PCA has interpreted successfully .....	168
Table SV-2. Scaling options that TREND supports .....	169

## LIST OF ABBREVIATIONS

1D	1 Dimensional
2D	2 Dimensional
3D	3 Dimensional
AFFN	ASCII Free Format Numeric
ANS	1-anilinonaphthalene-8-sulfonate
ASDF	ASCII Squeezed Difference Form
Apo-deP	Dephosphorylated, ligand-free PMM/PGM enzyme
Apo-P	Phosphorylated, ligand-free PMM/PGM enzyme
BEST-TROSY	Band-selective Excitation Short-transient Transverse Relaxation-optimized Spectroscopy
BMRB	Biological Magnetic Resonance Data Bank
CHESCA	Chemical Shift Covariance Analysis
CLEANEX-PM	Phase-Modulated Clean Chemical Exchange
CLI	Command-Line Interface
$C_m$	[urea] required for 50% denaturation
CONCISE	COordinated Chemical Shifts bEhavior
CSP	Chemical Shift Perturbation
D1-D4	Domains 1-4
DTT	dithiothreitol
$E_p$	Phosphorylated PMM/PGM enzyme
EPR	Electron Paramagnetic Resonance
$E_{deP}$	Dephosphorylated PMM/PGM enzyme
FID	Free Induction Decay
FA	Factor Analysis
G16P	Glucose 1,6-bisphosphate

G1P	Glucose 1-phosphate
G6P	Glucose 6-phosphate
GCDA	glycochenodeoxycholate
GUI	Graphic User Interface
H-bonds	Hydrogen Bonds
HDX	Hydrogen Deuterium Exchange
HDX-MS	Hydrogen-deuterium exchange mass spectrometry
HSQC	Heteronuclear Single Quantum Correlation
HX	Hydrogen Exchange
JCAMP-DX	Joint Committee on Atomic and Molecular Physical data – Data Exchange format
IC	Independent Component
ICA	Independent Component Analysis
M6P	Mannose 6-phosphate
MD	Molecular Dynamics
MRI	Magnetic Resonance Imaging
MVA	Multivariate Analysis
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PC	Principal Component
PCA	Principal Component Analysis
PDB	Protein Data Bank
PMM/PGM	Phosphomannomutase / Phosphoglucomutase
ppm	parts per million
pSer	phosphoserine
RF	Radio Frequency
R <sub>g</sub>	Radius of gyration

RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SVD	Singular Value Decomposition
TREND	TRack Equilibrium and Non-equilibrium shifts among Data
TROSY	Transverse Relaxation-Optimized Spectroscopy
UV-Vis	Ultraviolet-Visible Spectroscopy
wt	wild type
X1P	Xylose 1-phosphate

## ABSTRACT

Phosphomannomutase/phosphoglucomutase (PMM/PGM) contributes to the infectivity of an opportunistic human pathogen *Pseudomonas aeruginosa* by participating in carbohydrate biosynthesis. As a phospho-transfer enzyme, PMM/PGM catalyzes reversible conversion between 1- and 6- phosphosugars via a bisphosphorylated hexose intermediate rotating 180° in the catalytic cleft. Although PMM/PGM is well studied both structurally and kinetically, the mechanisms of its intramolecular and intermolecular communication are less well understood, especially in solution. Multiple solution NMR techniques are used in this dissertation's work to reveal information on PMM/PGM interactions at the atomic level. NMR-detected hydrogen exchange in combination with molecular dynamics and electrostatic calculations found phosphorylation of active Ser108 residue stabilizes PMM/PGM by attracting domains together. Responses of PMM/PGM to various phosphosugars were characterized by NMR-detected titrations. The large set of assigned peaks were analyzed by various types of principal component analysis (PCA) to derive binding isotherms, over-arching relationships among phosphosugar ligand binding reactions, and equilibrium shifts of PMM/PGM during its catalytic cycle. PCA was also found to be able to extract binding isotherms directly from non-interpreted 2D NMR spectra of complexes forming in solution. Procedures were identified that are reliable for obtaining the binding isotherms, regardless of the spectral peaks being in the fast, slow, or intermediate exchange regimes, or mixtures thereof. Applying PCA to time-domain NMR data also yields binding isotherms from titrations in fast or slow exchange. The algorithm readily extracts from an MRI movie its time

courses, such as breathing and heart rate in chest imaging. To enable the community to exploit these new capabilities, we have developed the multi-platform software named TREND to track equilibrium and non-equilibrium population shifts among 2D data frames. The principal components obtained represent the main changes among the data frames. Besides binding isotherms and time courses, the main changes extracted by TREND can be any variety of population shifts. TREND can reconstruct the series of measurements from selected principal components. TREND supports multiple data formats, including raw NMR data, images, movies, lists, and spreadsheet files. The software can also be used for data clustering or noise filtering.

# **CHAPTER I**

## **Introduction**

### **I.1 Phosphomannomutase / Phosphoglucosmutase (PMM/PGM)**

*P. aeruginosa* is an opportunistic human pathogen that expresses a series of cell surface polysaccharides. It exists ubiquitously and causes lung infections in ventilator-assisted cystic fibrosis, chronic obstructive pulmonary disease, and pneumonia (1-3). The enzyme (PMM/PGM), encoded by *algC* gene, belongs to the  $\alpha$ -D phosphohexomutase superfamily, members of which are found in all organisms from *E. coli* to humans (4). It contributes to the infectivity of *P. aeruginosa* by participating in the biosynthesis of several bacterial exoproducts, including alginate, lipopolysaccharide, and rhamnolipid (5-8). During biological processes, it catalyzes the reversible conversion of 1- to 6-phosphosugars, including mannose and glucose with equal efficiency (9-12).

The three dimensional structure of PMM/PGM was determined using X-ray crystallography by Regni et al (4). It is a monomeric protein with 463 residues. Four domains containing both  $\alpha$ -helices and  $\beta$ -sheets encompass a large, positive charged cleft and form an overall “heart” shape. Although the four domains of PMM/PGM pack compactly, domains 1-3 share a common topology featuring a four-stranded  $\beta$ -sheets between two helices, while domain 4 is structurally different from the other domains and classified as one of the TATA-box binding protein-like superfamily (13). Domain 4 contains a four-stranded antiparallel  $\beta$ -sheet between two  $\alpha$ - helices, as well as two short  $\beta$ -strands, and is well folded in solution even as isolated fragment (14, 15).

Several important residues located in the active site of PMM/PGM play important roles in the function of the enzyme. The highly conserved catalytic phosphoserine 108 (pSer108) is the essential phosphoryl donor and acceptor. A conserved metal binding loop (residues 242-246) chelates a  $Mg^{2+}$  ion necessary for maximum enzyme activity. In addition, a sugar binding loop (residues 324-328) contains residues predicted to distinguish between related sugar substrates. The distal phosphate-binding site including residue Arg421 interacts with the phosphate group of the sugar ligand (4, 16). It has been demonstrated that the catalysis of PMM/PGM proceeds via a bisphosphorylated intermediate, which is reoriented by  $180^\circ$  in the active site of the enzyme in between the two successive phosphoryl transfers. This intermediate places the phosphophoryl group adjacent to the catalytic phosphoserine 108. Structural and experimental studies have demonstrated conformational change is critical for ligand binding and catalysis (12, 17-20). Upon binding of ligands such as glucose 1-phosphate (G1P), glucose 6-phosphate (G6P), glucose 1,6-bisphosphate (G16P), and xylose 1-phosphate (X1P), domain 4 moves inward and narrows the active cleft (12, 18). (X1P is an inhibitor analogue of G1P lacking C6 and the O6 hydroxyl group which is not able to participate phosphoryl group).  $^{15}N$  NMR relaxation reveals rotational mobility of domain 4 (21). Domain 4 is drawn closer to domains 1 and 3 by electrostatic attraction between phosphoryl group and positive charged side of domain 4, increasing stability while decreasing accessibility of the enzyme (22). It was proposed that flexibility of PMM/PGM facilitates the reorientation of the intermediate, coupling the structural dynamics with its catalytic mechanism (23).

PMM/PGM has been well characterized both structurally and kinetically, making it a good model system for the understanding of dynamics and conformational change and their roles in function of large multi-domain enzymes. However, details of inter- and intramolecular communication of this large enzyme still remain to be elucidated. In addition, other features of this enzyme, including the highly reversible catalytic mechanism, and the high solubility and stability of the protein itself and its ligands, make PMM/PGM quite suitable for biophysical characterization during active catalysis, an exciting long-term goal. Therefore, multiple experimental NMR and theoretical were applied in order to find interrelationships between protein structure, dynamics and catalysis.

## **I.2 Technical Background of Nuclear Magnetic Resonance**

### **I.2.1 NMR basics**

Nuclear magnetic resonance has been one of the most important analytical techniques since its development in the 1940s. NMR spectroscopy, together with X-ray crystallography and cryo-electron microscopy, are the only techniques that determine 3D structures of macromolecules at atomic or near-atomic resolution. NMR spectroscopy is also a powerful tool for investigating kinetics and dynamics.

### **I.2.2 Source of NMR signals**

NMR probes intrinsic spin properties of certain atomic nuclei. The spin of the nuclei is usually denoted as spin quantum number  $I$ , which defines the number of stationary spin states of a nucleus:  $\text{levels} = (2I+1)$ . Since NMR requires existence of nuclear spin, nuclei with 0 spin quantum numbers is not detectable. For biomolecular NMR, nuclei with  $I=1/2$  spins are routinely used, such as  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{19}\text{F}$  and  $^{31}\text{P}$ . In the

presence of external magnetic field, the degeneracy of energy levels of nuclei with non-zero spins is broken, causing Zeeman splitting. NMR is the spectroscopy of split energy levels, which is proportional to the magnitude of the magnetic field:

$$\Delta E = h\gamma B_0 \quad \text{Equation I-1}$$

where  $h$  is Planck's constant,  $\gamma$  is the gyromagnetic ratio of the nucleus, and  $B_0$  is strength of the magnetic field. The frequency of the corresponding electromagnetic waves (light) can be derived by Planck's law:

$$\nu = \frac{\omega}{2\pi} = \frac{\gamma B_0}{2\pi} \quad \text{Equation I-2}$$

When a magnetic field is applied to an oscillating magnetic dipole moment, a torque is exerted and causes precession of the magnetic moment around the applied field, much like a top precessing around the gravitational field. This is called Larmor precession at the Larmor frequency, which is given by

$$\omega_0 = -\gamma B_0 \quad \text{Equation I-3}$$

### **I.2.3 Time-domain and frequency-domain NMR signal detection**

For a system with the same kind of nuclei such as  $^1\text{H}$  with no external magnetic present, its total magnetic moment is zero, because the spin polarizations are randomly and uniformly distributed. However, all spins begin Larmor precession when an external field  $B_0$  is turned on. As a result, the macroscopic nuclear magnetization gradually increases exponentially until thermal equilibrium is established, with all spins precessing about the direction of  $B_0$  at the Larmor frequency  $\omega_0$ . However, experimental study of spin magnetization along the field is impractical because its magnitude is four orders less than diamagnetism (24). In order to detect NMR signal, a radio-frequency (RF) pulse is

applied perpendicular to the direction of  $B_0$  and the magnetization along the direction of RF is measured. The net magnetization moment perpendicular to  $B_0$  with Larmor frequency  $\omega_0$ , which is called transverse magnetization, starts decaying back to zero once the RF is turned off. The change of transverse magnetization is recorded by the NMR spectrometer and referred to as the free induction decay (FID), which is the time-domain NMR signal. In modern NMR spectrometers, the FID signal is converted into a series of discrete values using an analog-to-digital converter (ADC). The FID can then be converted by Fourier transformation into a frequency-domain NMR spectrum, which is readily interpretable.

#### **I.2.4 Chemical shift**

The resonance frequency of a given nuclear spin is dictated not only by the external magnetic field but is also affected more subtly by the local electronic environment, causing slight differences in frequencies described by Eq. I-4:

$$\omega_{obs} = \gamma B_0(1 - \sigma) \quad \text{Equation I-4}$$

where  $\sigma$  is the shielding constant of the observed nuclei. This difference, which is known as chemical shift, is the most accessible information provided by NMR. Chemical shift is usually measured in a manner independent of the magnetic field and relative to a standard reference in the units of parts per million (ppm):

$$\delta = \frac{\omega - \omega_{ref}}{\omega_0} \times 10^6 \quad \text{Equation I-5}$$

$\delta$  is a dimensionless ratio characteristic that is independent of magnetic field strength.

Since chemical shift depends heavily on the local electromagnetic environment surrounding the nucleus, it can be used as a probe to detect nucleus-specific information.

For example, the chemical shift perturbation (CSP) is very sensitive to structural changes of biomolecules. Consequently, it can be used to study protein-ligand binding or protein-protein interactions.

### I.2.5 Bloch equations

NMR phenomena can be described by the density matrix formalism, which is accurate and adequate (25), or alternatively by the Bloch equations (26). The Bloch equations provide simpler descriptions and can be used to predict FIDs and frequency domain spectra, but at the cost of ignoring scalar couplings (26) .

Bloch equations in the rotating reference frame are given by:

$$\frac{dM_x(t)}{dt} = -\Omega M_y(t) + \omega_1 \sin \phi M_z(t) - R_2 M_x(t)$$

$$\frac{dM_y(t)}{dt} = \Omega M_x(t) - \omega_1 \cos \phi M_z(t) - R_2 M_y(t)$$

$$\frac{dM_z(t)}{dt} = \omega_1 [-\sin \phi M_x(t) + \cos \phi M_z(t)] - R_1 [M_z(t) - M_0] \quad \text{Equations I-6}$$

where  $M_x$ ,  $M_y$ ,  $M_z$  are x, y, and z Cartesian components of the magnetization, which are time-dependent functions.  $M_0$  is the magnetization value at time 0.  $R_1$  and  $R_2$  are longitudinal and transverse relaxation constants.  $\Omega$  is the Larmor frequency.  $\phi$  is the phase of field. The Bloch model is a valuable tool to visualize simple NMR experiments. In this study, Bloch equations are used to model line shapes of NMR spectra for multivariate analysis.

### I.2.6 One-dimensional NMR experiment

A simple one-pulse experiment contains only a single RF pulse on one channel. The RF pulse excites transverse magnetization and precession of the net magnetic

moments starts. After the pulse is switched off, the precessing magnetization generates the complex FID signal  $s(t)$  during the acquisition period. The time-dependent signal  $s(t)$ , which can be described by Bloch equations, is recorded by the NMR spectrometer as digital signals at regular intervals. Fourier transformation (FT) converts the time-domain FID signal into complex frequency-domain 1D spectrum:

$$S(\omega) = \int_0^{\infty} s^+(t) \exp(-i\omega t) dt = v(\omega) + i\mu(\omega) \quad \text{Equation I-7}$$

where  $v(\omega)$  and  $\mu(\omega)$  represent the absorptive and dispersive components of the line shapes, respectively. In chapter III we will use absorptive line shapes to simulate 1D NMR spectra using the function:

$$v(\omega) = \lambda M_0 \frac{R_2}{R_2^2 + (\Omega - \omega)^2} \quad \text{Equation I-8}$$

In more complex NMR pulse sequences, RF pulses are applied to perturb magnetizations. Useful information on spin systems can be obtained by comparing spectra with or without perturbations. One-dimensional NMR techniques provide extremely useful information for small molecules. However, the application of 1D NMR is usually limited in the field of macromolecules because of the complex, highly overlapped spectra.

### **I.2.7 Multi-dimensional NMR**

The resolution can be greatly improved by extending NMR spectra to more dimensions. A simple two-dimensional NMR spectrum includes four different components: preparation, evolution, mixing, and acquisition.

In the preparation step, desired non-equilibrium state of the spin system is prepared by applying a single or more complexes sequences of RF pulses. During the

evolution period  $t_1$ , transverse magnetization is allowed to freely evolve in the presence of chemical shift and scalar coupling. Coherence is transferred between spins and cross-peaks are generated during mixing period of the pulse sequence. The FID is recorded as a function of  $t_2$  during the detection period. These steps are repeated with  $t_1$  incremented from initial to maximum  $t_1$  value, constructing two-dimensional matrix as a function of two independent time domains ( $t_1, t_2$ ). Rows of this matrix are FID signals directly detected during the  $t_2$  period. Columns in the data matrix resemble FIDs in oscillating sinusoidally with chemical shift, but respond *indirectly* to chemical shift as functions of  $t_1$ . Applying Fourier-transformations to  $t_2$  and  $t_1$  time domains generates the two-dimensional NMR spectrum with F2 and F1 frequency dimensions for the respective nuclei measured.

Three-dimensional NMR experiments are performed with pulse sequences constructed as a combination of two 2D NMR pulse sequences. The time-domain signal of a 3D NMR experiment is recorded during the acquisition period  $t_3$  as a function of the two indirect evolution times  $t_1$  and  $t_2$ , , generating a 3D time-domain data matrix. Applying 3D Fourier transformation to this data matrix generates the corresponding 3D frequency spectrum, with dimensions denoted as F1, F2, and F3. Similar to a 2D NMR experiment, different dimensions in a 3D experiment can represent identical or different types of nuclei. The experiments in which all dimensions represent the same kind of nuclei (usually  $^1\text{H}$ ) are called as homonuclear multi-dimensional spectra. Otherwise the experiments are referred to as heteronuclear multi-dimensional spectra. 3D NMR spectra can be represented as a 3D cube, however it is usually viewed as a stack of 2D slices (planes) separated by a third frequency axis. A series of 2D experiments can be acquired

as a single 3D experiment while the third dimension is not frequency dimension. A 3D experiment collected in this manner is actually a combination of multiple 2D experiments and hence called pseudo-3D experiment. Four-dimensional NMR consists of three independent evolution times ( $t_1$ ,  $t_2$ , and  $t_3$ ) and the acquisition time  $t_4$ . 4D NMR spectra are used to solve extreme cases of overlap where peaks are still overlapped even in 3D NMR spectra.

### **I.3 Multivariate analysis**

Multivariate analysis (MVA) encompasses a family of mathematical techniques that analyzes and interprets large data sets. MVA examines large amount of variables simultaneously and extracts their key features (27). Many different types of multivariate analysis methods have been developed for various purposes. Some of these methods are popular and do not need human supervision, including factor analysis (FA), principal component analysis (PCA), independent component analysis (ICA), and k-means cluster analysis (28). Multivariate analysis have been used for NMR data for decades (29). In this work, PCA and ICA are used as quantitative chemometrics tools to analyze a series of 2D measurements and extract the main trend(s) out of them. The implementation of PCA and ICA will be described in chapter IV and V.

### **I.4 References**

1. Ballok, A. E., and G. A. O'Toole. 2013. Pouring Salt on a Wound: *Pseudomonas aeruginosa* Virulence Factors Alter  $\text{Na}^+$  and  $\text{Cl}^-$  Flux in the Lung. *Journal of Bacteriology* 195:4013-4019.
2. Li, X. J., Q. Li, L. Y. Si, and Q. Y. Yuan. 2011. Bacteriological Differences Between Patients with Acute Exacerbation of COPD and Community-Acquired Pneumonia. *Respir Care* 56.

3. Govan, J. R. W., and V. Deretic. 1996. Microbial pathogenesis in cystic fibrosis - Mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol. Rev.* 60:539-574.
4. Regni, C., P. A. Tipton, and L. J. Beamer. 2002. Crystal structure of PMM/PGM: an enzyme in the biosynthetic pathway of *P. aeruginosa* virulence factors. *Structure* 10:269-279.
5. Ye, R. W., N. A. Zielinski, and A. M. Chakrabarty. 1994. Purification and characterization of phosphomannomutase/phosphoglucomutase from *Pseudomonas aeruginosa* involved in biosynthesis of both alginate and lipopolysaccharide. *J Bacteriol* 176:4851-4857.
6. Cryz, S. J., Jr., T. L. Pitt, E. Furer, and R. Germanier. 1984. Role of lipopolysaccharide in virulence of *Pseudomonas aeruginosa*. *Infect Immun* 44:508-513.
7. May, T. B., D. Shinabarger, R. Maharaj, J. Kato, L. Chu, J. D. DeVault, S. Roychoudhury, N. A. Zielinski, A. Berry, R. K. Rothmel, and et al. 1991. Alginate synthesis by *Pseudomonas aeruginosa*: a key pathogenic factor in chronic pulmonary infections of cystic fibrosis patients. *Clin Microbiol Rev* 4:191-206.
8. Davey, M. E., N. C. Caiazza, and G. A. O'Toole. 2003. Rhamnolipid surfactant production affects biofilm architecture in *Pseudomonas aeruginosa* PAO1. *J Bacteriol* 185:1027-1036.
9. Ray, W. J., Jr., and G. A. Roscelli. 1964. A Kinetic Study of the Phosphoglucomutase Pathway. *J Biol Chem* 239:1228-1236.
10. Shackelford, G. S., C. A. Regni, and L. J. Beamer. 2004. Evolutionary trace analysis of the alpha-D-phosphohexomutase superfamily. *Protein Sci* 13:2130-2138.
11. Naught, L. E., and P. A. Tipton. 2001. Kinetic Mechanism and pH Dependence of the Kinetic Parameters of *Pseudomonas aeruginosa* Phosphomannomutase/Phosphoglucomutase. *Arch Biochem Biophys* 396:111-118.
12. Regni, C., L. E. Naught, P. A. Tipton, and L. J. Beamer. 2004. Structural basis of diverse substrate recognition by the enzyme PMM/PGM from *P. aeruginosa*. *Structure* 12:55-63.
13. Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247:536-540.
14. Schramm, A. M., D. Karr, R. Mehra-Chaudhary, S. R. Van Doren, C. M. Furdui, and L. J. Beamer. 2010. Breaking the covalent connection: Chain connectivity and the catalytic reaction of PMM/PGM. *Protein science : a publication of the Protein Society* 19:1235-1242.
15. Wei, Y., T. Marcink, J. Xu, A. Sirianni, A. S. Sarma, S. Prior, L. Beamer, and S. Van Doren. 2014. Chemical shift assignments of domain 4 from the phosphohexomutase from *Pseudomonas aeruginosa* suggest that freeing perturbs its coevolved domain interface. *Biomol NMR Assign* 8:329-333.
16. Shankar, S., R. W. Ye, D. Schlichtman, and A. M. Chakrabarty. 1995. Exopolysaccharide alginate synthesis in *Pseudomonas aeruginosa*: enzymology

- and regulation of gene expression. *Adv Enzymol Relat Areas Mol Biol* 70:221-255.
17. Naught, L. E., and P. A. Tipton. 2005. Formation and reorientation of glucose 1,6-bisphosphate in the PMM/PGM reaction: transient-state kinetic studies. *Biochemistry* 44:6831-6836.
  18. Regni, C., A. M. Schramm, and L. J. Beamer. 2006. The Reaction of Phosphohexomutase from *Pseudomonas aeruginosa*: Structural Insights Into A Simple Processive Enzyme. *J Biol Chem* 281:15564-15571.
  19. Regni, C., G. S. Shackelford, and L. J. Beamer. 2006. Complexes of the enzyme phosphomannomutase/phosphoglucomutase with a slow substrate and an inhibitor. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 62:722-726.
  20. Schramm, A. M., D. Karr, R. Mehra-Chaudhary, S. R. Van Doren, C. M. Furdui, and L. J. Beamer. 2010. Breaking the covalent connection: Chain connectivity and the catalytic reaction of PMM/PGM. *Protein Sci* 19:1235-1242.
  21. Sarma, A. V., A. Anbanandam, A. Kelm, R. Mehra-Chaudhary, Y. Wei, P. Qin, Y. Lee, M. V. Berjanskii, J. A. Mick, L. J. Beamer, and S. R. Van Doren. 2012. Solution NMR of a 463-residue phosphohexomutase: domain 4 mobility, substates, and phosphoryl transfer defect. *Biochemistry* 51:807-819.
  22. Xu, J., Y. Lee, Lesa J. Beamer, and Steven R. Van Doren. 2015. Phosphorylation in the Catalytic Cleft Stabilizes and Attracts Domains of a Phosphohexomutase. *Biophysical Journal* 108:325-337.
  23. Lee, Y., M. T. Villar, A. Artigues, and L. J. Beamer. 2014. Promotion of Enzyme Flexibility by Dephosphorylation and Coupling to the Catalytic Mechanism of a Phosphohexomutase. *Journal of Biological Chemistry* 289:4674-4682.
  24. Levitt, M. H. 2001. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. Wiley.
  25. Cavanagh, J., W. J. Fairbrother, A. G. Palmer Iii, M. Rance, and N. J. Skelton. 2007. CHAPTER 2 - THEORETICAL DESCRIPTION OF NMR SPECTROSCOPY. In *Protein NMR Spectroscopy (Second Edition)*. Academic Press, Burlington. 29-113.
  26. Cavanagh, J., W. J. Fairbrother, A. G. Palmer Iii, M. Rance, and N. J. Skelton. 2007. CHAPTER 1 - CLASSICAL NMR SPECTROSCOPY. In *Protein NMR Spectroscopy (Second Edition)*. Academic Press, Burlington. 1-28.
  27. Rencher, A. C., and W. F. Christensen. 2012. Front Matter. In *Methods of Multivariate Analysis*. John Wiley & Sons, Inc. i-xxv.
  28. Ghahramani, Z. 2004. Unsupervised Learning. In *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures*. O. Bousquet, U. von Luxburg, and G. Rätsch, editors. Springer Berlin Heidelberg, Berlin, Heidelberg. 72-112.
  29. Edlund, U., and H. Grahn. 1991. Multivariate data analysis of NMR data. *J Pharm Biomed Anal* 9:655-658.

## **CHAPTER II**

### **Phosphorylation in the Catalytic Cleft Stabilizes and Attracts Domains of a Phosphohexomutase**

Adapted with permission from Jia Xu, Yingying Lee, Lesa J Beamer, Steven R Van Doren, *Biophys. J.* 2015 Jan 20;108(2):325-37. Copyright 2015 Elsevier

#### **II.1 Abstract**

Phosphorylation can modulate activities of enzymes. The phosphoryl donor in the catalytic cleft of  $\alpha$ -D-phosphohexomutases is transiently dephosphorylated while the reaction intermediate completes a 180° reorientation within the cleft. The phosphorylated form of 52 kDa bacterial phosphomannomutase / phosphoglucomutase is less accessible to dye or protease, more stable to chemical denaturation, and widely stabilized against NMR-detected hydrogen exchange across the core of domain 3 to juxtaposed domain 4 (each by  $\geq 1.3$  kcal/mol) and parts of domains 1 and 2. However, phosphorylation accelerates hydrogen exchange in specific regions of domains 1 and 2, including a metal-binding residue in the active site. Electrostatic field lines reveal attraction across the catalytic cleft between the locale of phosphoSer108 and domain 4, contrasted by repulsion when dephosphorylated. Molecular dynamics (MD) simulated the dephosphorylated form to be expanded due to enhanced rotational freedom of domain 4. The contacts and fluctuations of the MD trajectories enabled correct simulation of more than 80% of sites that undergo either protection or deprotection from hydrogen exchange due to phosphorylation. Electrostatic attraction in the phosphorylated enzyme accounts for: (i) domain 4 drawing closer to domains 1 and 3, (ii) decreased accessibility, and (iii) increased stability within these domains. The “electrostriction” due to phosphorylation may help capture substrate, while the opening of the cleft upon transient dephosphorylation allows rotation of the intermediate. The long-range effects of

phosphorylation on hydrogen exchange parallel reports on protein kinases, suggesting a conceptual link among these multi-domain, phosphoryl transfer enzymes.

## II.2 Introduction

Phosphorylation regulates enzymes with diverse consequences ranging from long-range allosteric activation of glycogen phosphorylase to million-fold inhibition of isocitrate hydrogenase (1). Phosphorylation of the activation loop of many eukaryotic protein kinases triggers electrostatic rearrangement of the loop and assembly of the internal regulatory spine of hydrophobic residues in the N- and C-lobes of these proteins, thereby supporting interaction with substrate and activity (2). Phosphorylation has been proposed to stabilize proteins by decreasing the range of motions, typically with global conformational adjustments  $< 2 \text{ \AA}$  (3). Exceptions to this stabilization have been noted (4).

$\alpha$ -D-phospho-hexomutases are also sensitive to phosphorylation. Their reaction cycles oscillate between serine-phosphorylated and dephosphorylated forms (5-7) (Fig. II-1). The serine donates a phosphoryl group to the phosphosugar substrate. While transiently dephosphorylated the serine accepts a phosphoryl group from the bisphosphorylated intermediate (Fig. II-1). The phosphorylated enzyme binds and unbinds the phosphorylated sugar while the bisphosphorylated intermediate reorients by  $180^\circ$  while associated with the dephosphorylated enzyme (6, 8, 9). An attractive member of the enzyme superfamily for biophysical characterization of effects of phosphorylation on catalytic switching is phosphomannomutase / phosphoglucomutase (PMM/PGM) from *Pseudomonas aeruginosa*.

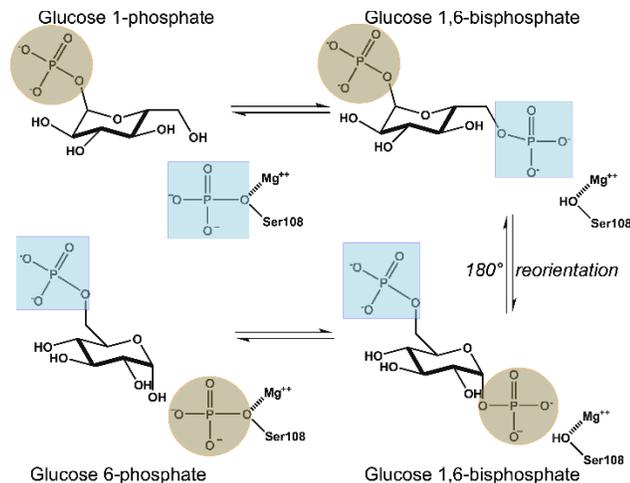


Figure II-1 Catalytic mechanism of PMM/PGM. The enzyme transforms G1P to G6P via G1,6P, which undergoes rotation in the active site.

The phosphomutase activity of PMM/PGM contributes to the infectivity of *P. aeruginosa* (in ventilator-assisted pneumonia, chronic obstructive pulmonary disease, and cystic fibrosis (10-12)) through the biosynthesis of sugar precursors of lipopolysaccharide, exopolysaccharides Pel and Psl, rhamnolipids, and alginate (13-16). Ser108 of PMM/PGM is phosphorylated at the base of its deep, positively charged catalytic cleft formed by its four domains (Fig. II-3B). Upon binding of a phosphosugar, domain 4 (D4) rotates to close the catalytic cleft (9, 17). Structures containing a bisphosphorylated intermediate suggested a half-open state of the enzyme (9), which may accommodate the 180° reorientation of the intermediate required by the catalytic cycle (8). <sup>15</sup>N NMR relaxation indicates partial independence of the rotational mobility of D4 (18), consistent with crystallographic snapshots of varying conformers of D4 (9). Amide NMR spectra suggest a mixture of phosphorylated and dephosphorylated enzyme forms (18). The catalytic cleft was widened by the S108A mutation (6), suggesting a role for phosphorylation of Ser108 in conformational adjustment.

Conformational effects of phosphorylation can be investigated by hydrogen exchange (HX, broadly referring from fast to slow time scales), which depends upon the stability of the structural environment and longevity of hydrogen bonds (H-bonds). Transient openings of H-bonds facilitate access of hydroxide ions to amide protons to catalyze exchange. Previously, hydrogen-deuterium exchange mass spectrometry (HDX-MS) observed percentage changes in the deuteration of peptide fragments of PMM/PGM, suggesting its dephosphorylated form (Apo-deP) to be more flexible than the phosphorylated form (Apo-P), consistent with a role in facilitating the 180° rotation of the intermediate (19). However, limitations inherent to HDX-MS precluded characterization of important aspects of HDX in this enzyme, including regions critical to function. Much of the active site could not be characterized because (i) the peptide containing catalytic phosphoSer108 was missing from mass spectra, and (ii) affected residues in four other peptides encompassing the active site could not be pinpointed due to peptide lengths of 24 to 35 residues. Residue-specific information was also missing elsewhere. Moreover, HDX-MS data was not available for time scales of sec or days (20). Consequently, the HDX-MS results did not provide detailed kinetic and thermodynamic insights required for biophysical understanding (21).

We implemented NMR to measure hydrogen-deuterium exchange (HDX) and msec-scale HX of 463-residue PMM/PGM to pinpoint the effects of phosphorylation on hydrogen exchange, its rate constants, and free energies. These methods show that phosphorylation causes both slowing *and* acceleration of HX around the active site and in domains 1 and 2 (D1 and D2), little change at many residues, and slowing of HX of around 100 residues, widely and remotely. The most protected of these sites from HDX

identified by NMR report the global folding stability (22). This provides a lower limit on the phosphoryl stabilization of the fold of PMM/PGM and its domains, where traditional denaturation experiments fail for lack of reversible folding. Based on changes in free energy of HX and density of protected residues, the core of domain 3 (D3) appears to be most protected by phosphorylation, followed by the proximal face of D4. A biophysical accounting for the latter remote pattern of protection by phosphorylation is provided by electrostatic fields and molecular dynamics (MD) trajectories. Comparative calculations for Apo-P and Apo-deP suggest that compaction results from electrostatic attraction introduced by phosphorylation of catalytic Ser108, i.e. “electrostriction”. Use of surrounding contacts and either hydrogen bonding or  $\text{RMSF}^{-1}$  from the MD trajectories (23, 24) proved to be qualitatively highly consistent with the locations of altered HX protection accompanying dephosphorylation. The solution and computational lines of evidence reveal what X-ray diffraction constrained by crystal packing (19) could not: electrostatic changes due to phosphorylation appear to narrow the active site of PMM/PGM, compact the fold, and slow HX in D3 and D4. These changes may facilitate the capture and accommodation of phosphohexose substrates during the multi-step reaction of the enzyme.

## **II.3 Materials and methods**

### **II.3.1 Enzyme preparation**

His-tagged PMM/PGM from *P. aeruginosa* was prepared (18). Samples were labeled with  $^2\text{H}$  and  $^{15}\text{N}$  for NMR. The enzyme was phosphorylated to 85-90% or dephosphorylated to 90-95% by incubating with excess glucose 1,6-bisphosphate (G16P) or glucosamine 1-phosphate at 4 °C for 18 h, respectively (19).

### II.3.2 Accessibility measurements

Apo-P and Apo-deP (10  $\mu$ M), with 90% and 0% levels of phosphorylation respectively, were incubated with 1 mM ANS for 1 h in 50 mM MOPS (pH 7.4) at 25° C. Fluorescence was excited at 365 nm. The fluorescence emission of free ANS in buffer was subtracted away (25).

30  $\mu$ M Apo-P and Apo-deP, at 87% and 10% levels of phosphorylation respectively, were treated with proteinase K in 50 mM MOPS (pH 7.4) at a 300: 1 (w/w) ratio at 25° C, until terminated with 3 mM PMSF. Bands on SDS-PAGE (25) were quantified using ImageJ.

### II.3.3 NMR spectroscopy

BEST-TROSY (26) and other NMR spectra were collected at 35 °C on a Bruker Avance III 800 MHz spectrometer with TCI cryoprobe. The samples were in 50 mM MOPS (pH 7.4 at 25 °C), 1 mM MgCl<sub>2</sub>, and 10 mM DTT. The spectra were processed with NMRPipe (27) and peaks interpreted using Sparky (28) using the reported assignments (18). An in-house Python script mapped Sparky-assigned peak shifts to NMRPipe peak tables and is available as a Supplemental file. NMRPipe provided the volume and height of each peak.

### II.3.4 Rapid HX by NMR

HX occurring on the ms scale was observed using the CLEANEX-PM pulse sequence enhanced with TROSY detection (29, 30), but with interleaved acquisition (see Supporting files). The solvent exchange rates were fitted to the equation:

$$\frac{I}{I_0} = \frac{k_{ex}}{R_{1A} + k_{ex} - R_{1B}} \times \{\exp(-R_{1B}\tau_m) - \exp[-(R_{1A} + k_{ex})\tau_m]\} \quad \text{Equation II-1}$$

where  $I/I_0$  represents normalized peak heights,  $R_{1A}$  the transverse and longitudinal relaxation during the spinlock,  $k$  the rate constant of HX, and  $R_{1B}$  the relaxation rate constant for the water (Fig. SII-2). Radiation damping complicated measurement of  $R_{1B}$ . Since trial values of  $R_{1B}$  from 0.1 to 0.001  $s^{-1}$  provide the same fitted rate constants (31), a value of 0.01  $s^{-1}$  was used.

### II.3.5 Quantification of rates of HDX

The uniformity of phosphorylation and dephosphorylation and keeping the cysteine residues reduced for a week prevented mixtures and bi-exponential changes. To initiate HDX, the  $^2H/^{15}N$ -labeled samples for HDX were concentrated to about 5 mM in  $H_2O$ -based buffer and then diluted 5-fold into  $D_2O$ -based buffer at 22 °C. Addition of 10 mM DTT and Ar over the samples then sealed in Shigemi tubes maintained reproducible TROSY spectra for one week (judged in  $H_2O$ ). 20 min of dead time for these handling steps limited measurable  $k_{obs}$  to  $< 5 \times 10^{-2} \text{ min}^{-1}$ . A series of at least 57 BEST-TROSY spectra (26) was acquired at 35 °C, continuously during the first 24 h and several per day thereafter until ~8000 min had elapsed. After the first 24 h, samples were maintained at 35 °C in a water bath between acquisitions. Exponential decay rates,  $k_{obs}$ , were fitted to the peak heights (Figs. II-2D), with correction for the residual 20%  $H_2O$  described by Eq. II-S1 in Supporting Material. Monitoring of gradual concurrent dephosphorylation and correction of the affected rate constants of HDX ( $k_{ex}$ ) are detailed in Supporting Material.

Each intrinsic rate constant of random coil hydrogen exchange ( $k_{rc}$ ) was obtained using the SPHERE server (<http://www.fccc.edu/research/labs/roder/sphere/sphere.html>). The free energy of amide protection from HX was calculated as:

$$\Delta G_{HX} = -RT \ln(k_{ex}/k_{rc}) \quad \text{Equation II-2}$$

Folding stabilities of each domain and the entire enzyme were estimated from the highest values of  $\Delta G_{HX}$  in each domain and corrected for proline isomerization (22).

### II.3.6 Electrostatic field calculations

Atomic charges and sizes around the metal binding site were parameterized by the Amber tools module MCPB/MTKPP which invokes Gaussian 2009 (32, 33). In the structure of Apo-P (6) (PDB: 1K35),  $Zn^{2+}$  is bound by pSer108 and aspartates 242, 244, and 246. In Apo-deP (PDB: 4MRQ) (19), a bound water replaces Ser108 in coordinating  $Zn^{2+}$ . Other atomic sizes and charges were defined by the Amber ff99SB force field (34). The electrostatic potentials were then computed using the Adaptive Poisson-Boltzmann Solver (35) at 310 K with protein and solvent dielectric constants of 2.0 and 78.5, respectively, and ionic strength of 30 mM. Electrostatic field lines were visualized using VMD (36) as illustrated (37, 38).

### II.3.7 Modeling HX protection from MD simulations

The logarithm of the simulated HX protection factor,  $\ln(P_n^{sim})$ , of an amide group (of residue  $n$ ) can be approximated by linear combination of its number of contacts  $N_n^c$  and a second protection factor element (PFE<sub>2</sub>) measured from a dynamics simulation, e.g. either number of hydrogen bonds  $N_n^h$  or  $RMSF^{-1}$  at that site:

$$\ln(P_n^{sim}) = \beta_c N_n^c + \beta_2 PFE_2 \quad \text{Equation II-3}$$

where the generic term  $\beta_2 PFE_2$  can be either  $\beta_h N_n^h$  or  $\beta_{RMSF} RMSF^{-1}$ , with the weights of the terms being  $\beta_c$ ,  $\beta_h$ , or  $\beta_{RMSF}$  (23, 24). For application below, two independent pairs of weight parameters were optimized for slower and faster regimes of HX:

$$\begin{cases} \beta_c^{\text{more prot}}, \beta_2^{\text{more prot}} & \text{if } \ln(P_n^{\text{exp}}) > 8 \\ \beta_c^{\text{less prot}}, \beta_2^{\text{less prot}} & \text{if } \ln(P_n^{\text{exp}}) < 8 \end{cases}$$

The modeling of protections improved using the alternative sets of coefficients (Table SII-3, Fig. SII-11). The performance of candidate sets of parameters with PMM/PGM is given in Table SII-4. Simulated protection factors were converted into free energies using Eq. II-4:

$$\Delta G_n^{sim} = -RT \ln(P_n^{sim}). \quad \text{Equation II-4}$$

## II.4 Results

We investigated phosphorylation-induced changes between the Apo-P and Apo-deP states of PMM/PGM from *P. aeruginosa* in solution and computationally, in order to understand better the interplay of these enzyme forms with the catalytic cycle.

### II.4.1 Phosphorylation decreases accessibility to probes and enhances stability

Several approaches were used to compare the overall structural flexibility, openness, and stability of Apo-P and Apo-deP in solution. ANS dye has long been used to probe conformational changes of proteins as it fluoresces little in water but much more upon binding to a protein (25, 39). The fluorescence emission of ANS is significantly greater when mixed with Apo-deP than when mixed with Apo-P (Fig. II-2A). This suggests that phosphorylation decreases ANS access to binding pockets. The folding stability of Apo-P and Apo-deP were compared using urea denaturation. The unfolding curves with [urea] of 3.0 M needed for 50% denaturation ( $C_m$ ) of Apo-P suggest its greater stability than Apo-deP with  $C_m$  of 2.8 M (Fig. II-2B). ( $C_m$  values are compared because PMM/PGM cannot be refolded after denaturation, indicating lack of the reversibility needed to obtain  $\Delta G_{fold}$ ). Limited proteolysis by proteinase K degrades Apo-

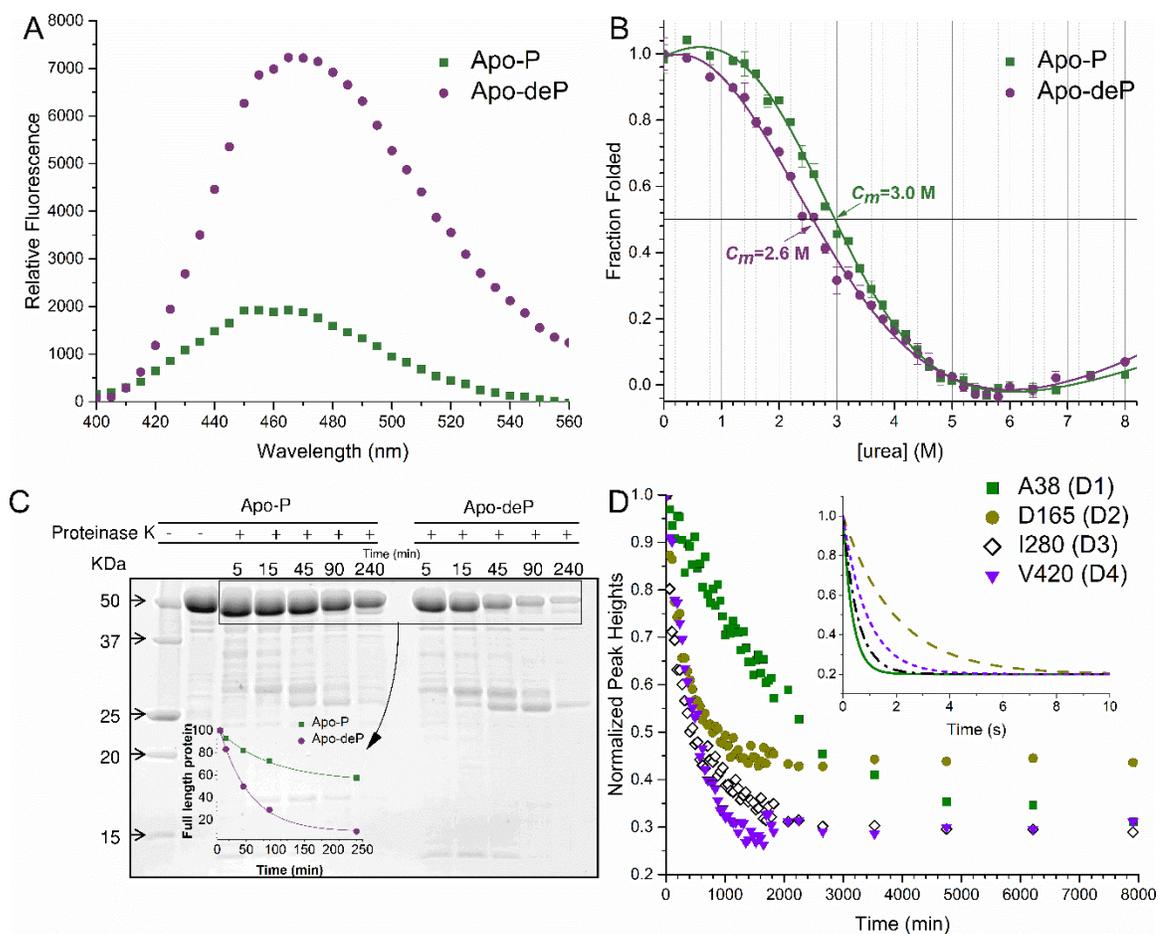


Figure II-2. Phosphorylation effects on accessibility and stability of PMM/PGM. (A) ANS (1-anilidonaphthalene-8-sulfonate) emission spectra with Apo-P and Apo-deP. (B) Denaturation of Apo-P and Apo-deP by incubation in urea for 12 h at 25 °C in 50 mM MOPS, pH 7.4, and 0.1 mM DTT were measured in duplicate. (C) Effect of proteinase K on Apo-P and Apo-deP. Aliquots were removed at times listed and monitored with SDS-PAGE. The inset plots the decay of the intact enzyme. (D) Examples of an amide group in each domain slowed in HX by phosphorylation. Upon change of Apo-P to D<sub>2</sub>O solution, the disappearance of an amide peak of each of the domains from TROSY spectra is plotted. The inset shows sub-sec components of HX simulated using the rate constants fitted to CLEANEX time courses for the same four residues in the Apo-deP form (Fig. SII-1)

deP 6.3-fold faster than Apo-P, based on rate constants fitted to the disappearance of the intact enzyme band on gels of  $2.4 \times 10^{-3} \pm 0.4 \times 10^{-3} \text{ min}^{-1}$  for Apo-P and  $1.5 \times 10^{-2} \pm 0.2 \times 10^{-2} \text{ min}^{-1}$  for Apo-deP (Fig. II-2C). Binding of the small dye, proteolysis by the large enzyme, and urea denaturation suggest that phosphorylation curbs the accessibility and enhances the folding stability of PMM/PGM.

## II.4.2 Effects of dephosphorylation on NMR spectra

Time-dependent changes in NMR spectra and mass spectra (loss of the mass of a phosphoryl group) revealed that PMM/PGM undergoes spontaneous dephosphorylation at physiological temperatures (Fig. SII-2). We sought a residue-by-residue structural comparison of the phosphorylation-dependent changes in accessibility and stability using NMR. Amide peaks representing active Apo-P and Apo-deP forms proved distinguishable for 98 residues with good sensitivity (Fig. SII-2, Table SII-1). At 35 °C, the NMR peaks of Apo-deP appeared and those of Apo-P disappeared (Fig. SII-2, C and D) with a time constant of  $19.0 \pm 1.1$  h ( $1140 \pm 70$  min) in the joint fits of the pairs of peaks of 43 residues.

Most of the amide NMR peaks shifted by dephosphorylation are distributed from D1 through D3 (Fig. SII-3, A and B). Lys219 of D2 plus Arg262 and Trp336 of D3 underwent the largest peaks shifts (with  $\Delta\omega_{\text{HN}} > 0.2$  ppm) despite being more than 19 Å distant from the phosphoryl group of Ser108 (Fig. SII-3B). In addition, dephosphorylation was found to broaden at least 33 amide peaks beyond detection, with other amide peaks of Apo-deP also being broader than their counterparts in Apo-P, especially in D2 and D3, but also D4 (Figs. SII-3 and SII-4). The line broadening probably arises from exchange among structural environments in these domains enabled by dephosphorylation, one manifestation of its long-range effects.

We investigated whether phosphorylation affects the backbone flexibility of PMM/PGM on the scale of ps to ns by measuring  $^{15}\text{N}\{^1\text{H}\}$  NOE relaxation on multiple freshly phosphorylated or dephosphorylated samples. 20 or more loops and termini show locally decreased NOE ratios indicating fast, localized fluctuations (Fig. SII-5). The high

similarity of the NOEs of Apo-P and Apo-deP, without differences beyond the S.D. of the triplicate measurements, suggests no convincing phosphorylation-dependent change on the sub-ns scale. Coupled with the much slower ms dynamics in Apo-deP implied by its NMR line broadening (Fig. SII-3 and SII-4), this suggests the need to test whether the phosphorylation state instead affects more rigidly structured regions, where the conformational resistance may influence dynamics on slow time scales accessible by HX methods.

#### **II.4.3 Slowing of HX by phosphorylation, especially in D3 and adjacent face of D4**

We scrutinized the working hypothesis of dephosphorylation globally mobilizing and increasing the accessibility of the structure of Apo-deP (19) (Fig. II-2), via residue-specific and thermodynamic quantification of HX using NMR. The pH of 7.4 needed for solubility and activity of PMM/PGM, and the temperature of 35 °C that enhances its TROSY NMR spectra, promote HX that is intrinsically comparatively rapid. The majority of amide protons were replaced by deuterons within 20 min of exposure to D<sub>2</sub>O, especially in Apo-deP (Movie II-S1). Upon 1 to 3 h of exchange, however, more than 100 amide groups across Apo-P emerged as slowed in HX (Movie II-S2) relative to Apo-deP, suggesting that multiple locales are stabilized by phosphorylation of the enzyme (Figs. II-2D, II-3, and II-S6).

Comparisons of the effects of phosphorylation status on the thermodynamics of HX, i.e.  $\Delta G_{HX}$ , are newly furnished by these NMR studies of PMM/PGM, both for its individual amide groups and entire domains, provided that its HX occurs in the bimolecular (EX2) limit. In the EX2 regime, the rate of closing of the H-bond ( $k_{cl}$ ) greatly

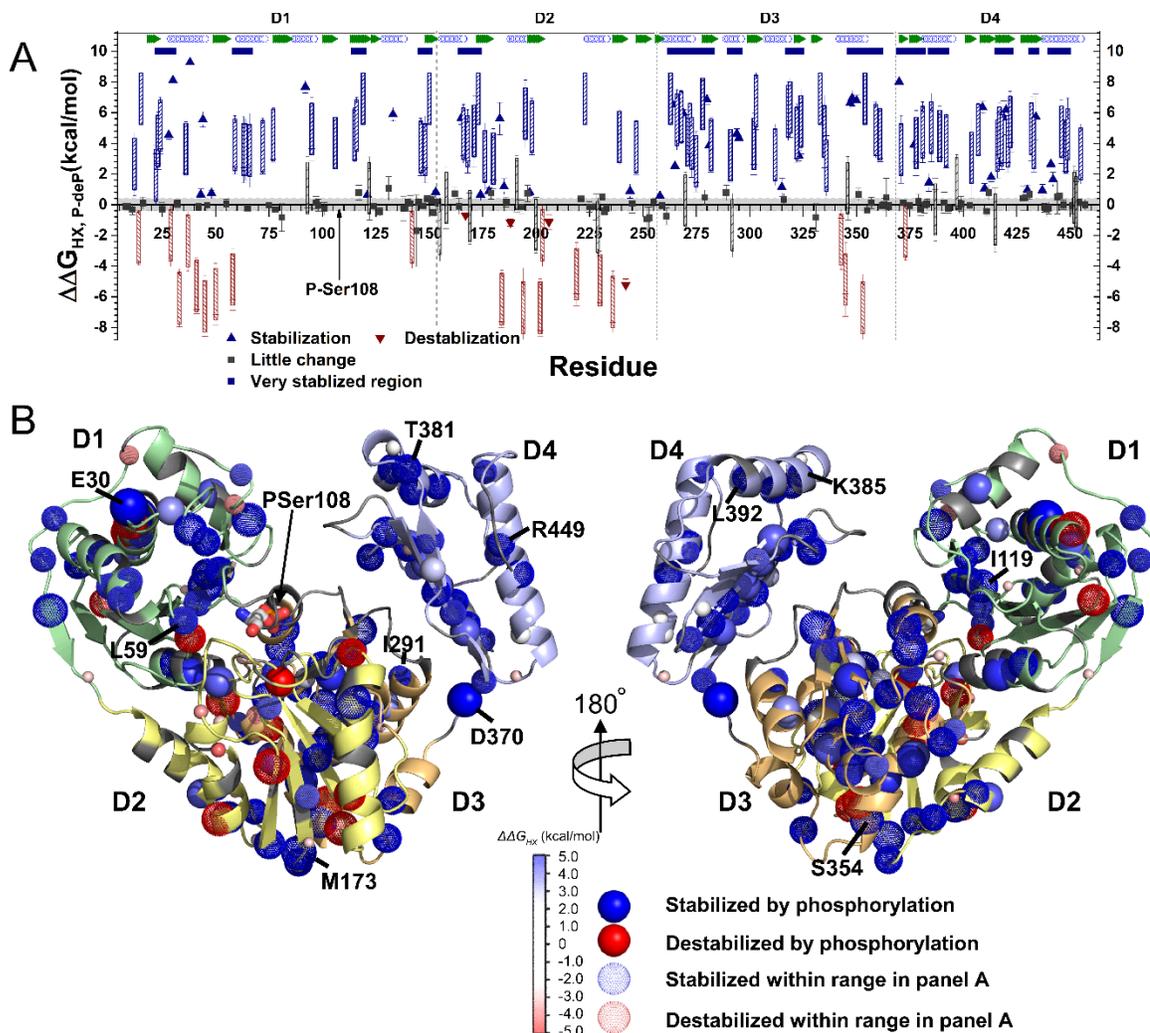
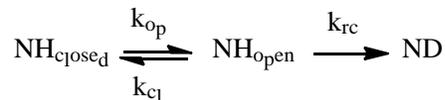


Figure II-3. Phosphorylation stabilizes or slows HX at more sites than it accelerates. (A) Rate constants of HX,  $k_{ex}$ , measured either as HDX by NMR or on the sub-sec scale by CLEANEX-PM (29), were transformed to  $\Delta G_{HX}$  using Eq. II-2 and the  $\Delta\Delta G_{HX, P-deP}$  differences plotted. Blue marks the differences where phosphorylation is significantly stabilizing and red where it is destabilizing. Triangles represent cases where  $k_{ex}$  was measured in both Apo-P and Apo-deP forms. Hatched bars mark cases in which  $k_{ex}$  in one phosphorylation state lay in the intermediate, unmeasured time range where  $1 \text{ s}^{-1} > k_{ex} > 4 \times 10^{-3} \text{ s}^{-1}$ ; the height of the bar indicates  $\Delta\Delta G_{HX, P-deP}$  lies within this range. Sequence segments with  $> 4$  of 10 residues protected by phosphorylation  $> 2$  SD above average are marked with navy blue bars. Locations of strands and helices are given in green and open blue symbols, respectively. (B) Amide groups protected from HX by phosphorylation of Ser108 are marked by spheres with shades of blue while amide groups destabilized by phosphorylation are red. Dotted spheres correspond to residues with the uncertainty ranges marked in (A). The amplitude of phosphorylation-dependent change,  $\Delta\Delta G_{HX}$ , is symbolized by color intensity and radii of the spheres.

exceeds the intrinsic rate of HX of a free amide of that residue,  $k_{rc}$  (40), in the

Linderström-Lang model of HX (41):



EX2 behavior is recognizable from  $\Delta G_{\text{HX}}$  values shared across a wide range of  $k_{\text{rc}}$ , and from a slope of 1 among such residues in log-log plots of  $k_{\text{ex}}$  vs.  $k_{\text{rc}}$  (42, 43). Groups of PMM/PGM residues sharing similar  $\Delta G_{\text{HX,apparent}}$  values do indeed have slopes of 1 in such plots, for both the Apo-P and Apo-deP forms (Fig. SII-7). This strongly suggests that EX2 behavior prevails for both states of the enzyme, from the slowest exchanging residues with  $\Delta G_{\text{HX}} > 10$  kcal/mol through sites of faster exchange with  $\Delta G_{\text{HX}}$  of  $\sim 6$  kcal/mol (Fig. SII-7). The EX2 behavior validates application of Eq. II-2 to PMM/PGM in the estimation of its  $\Delta G_{\text{HX}}$  values from HDX experiments and the stability of H-bonding implied.

Phosphorylation induces stabilization of HX that is significant in segments of D1 and D2, and throughout D3 and D4, as depicted by differences in  $\Delta G_{\text{HX}}$  in Fig. II-3. Phosphorylation stabilizes hydrogen exchange by 1.5 to 9.3 kcal/mol at 35 °C (with slowings of 12-fold to 3.8 million-fold) in  $\sim 23$  residues in D1, 12 residues in D2, 32 in D3, and 28 in D4 (Fig. II-3A). The greatest density of phosphorylation-stabilized residues occurs in the core of D3, throughout its central helix (Leu263 – Ser273) and the two flanking  $\beta$ -strands (Ala278 – Phe282 and Leu322 - Gly324, with ten phosphorylation-stabilized residues between them). In D4, two long, central  $\beta$ -strands (Gly416 - Ala422 and Arg432 - Glu434) that face D3 are similarly stabilized (Fig. II-3B). This suggests that phosphorylation may promote closer packing of D4 against D3.

Additional phosphorylation-dependent stabilization  $\Delta\Delta G_{\text{HX,P-deP}}$  appears to radiate outward from the most stabilized cores of D3 and D4. Next to the stabilized core helix of

D3 are two helices spanning Ile291 - Ser295 and Ala347 – Ser354 that also show phospho-stabilization (Fig. II-3B). Adjacent to the latter helix, a loop from D2 (Asp165 - Met173) is similarly stabilized. In D4, additional stabilized residues are found in a stretch from Asp370 to Thr381 that contains two  $\beta$ -strands, and in the solvent-exposed helices from Lys385 - Leu392 and Leu441 - Arg449 (Fig. II-3B). D1 manifests phosphorylation-dependent stabilization  $\Delta\Delta G_{HX,P-deP}$  in a  $\beta$ -strand that faces the active site and pSer108 (Asn115 - Ile119), in the adjoining loops, and at the N-terminal ends of two helices within the Val23 - Glu30 and Leu59 - Lys66 segments (Fig. II-3B). In total, about 15 residues of D1 within 20 Å of the phosphoryl group of pSer108 are protected by phosphorylation (Fig. II-3B).

Twenty amide groups of Apo-P, mainly in D1 and D3, were found to exchange so slowly that the full amplitude of their  $\Delta G_{HX}$  values cannot be determined due to the gradual, concurrent dephosphorylation of the sample ( $\tau = 19.0$  h, see above and Fig. SII-2). Consequently, it is the *lower bound* for  $\Delta G_{HX}$  that is estimated for these twenty (open squares in Fig. SII-6) after correction for the slow dephosphorylation (Fig. SII-8A). Since the hydrogen exchange of Apo-P and Apo-deP lie in the EX2 regime, the most slowly exchanging sites with the largest  $\Delta G_{HX}$  values can be used to estimate folding stability (44). Applying this method to PMM suggests that *at the very least* a 0.4 kcal/mol greater folding stability of Apo-P ( $\geq 11.0$  kcal/mol) than Apo-deP (10.6 kcal/mol). Local stabilizations by phosphorylation are much greater (Fig. II-3A), especially in D3 and D4 where Apo-P appears stabilized by at least 1.3 kcal/mol over Apo-deP (Fig. SII-9). The increased stability from phosphorylation may accompany increases in the conformational rigidity of the protein. This is consistent with the dozens of peaks of Apo-deP lost or

weakened by line broadening (Fig. SII-3). These observations may be attributable to chemical exchange among conformers at locations in D3 and D1 in the absence of phosphorylation.

In contrast to behaviors above, we find sites in D1 and D2 that are locally destabilized by phosphorylation. Specifically, nine residues in D1 and eleven in D2 show hydrogen exchange accelerated by phosphorylation. At least five of these residues in D1 and seven in D2 are destabilized by  $> 1.5$  kcal/mol (Fig. II-3A). The bulk of the phospho-destabilization in D1 is found within the first helix and in the subsequent two loops (Fig. II-3). Phosphorylation-destabilized residues are scattered throughout D2.

The NMR studies here afford a new opportunity to assess HX around the active site of PMM/PGM. One of the D2 residues destabilized by phosphorylation is Asp242 that coordinates the divalent cation in the active site. HDX by NMR shows that Asp242 is well-protected in Apo-deP with  $\Delta G_{HX,deP} = 7.3 \pm 0.3$  kcal/mol (Figs. II-3A). However, it is strongly destabilized in Apo-P where its HX is rapid and detected by CLEANEX-PM. This might suggest partial repulsion between the phosphoryl and aspartyl carboxylate groups, which could perturb the metal-binding loop. Ala184 of D2 and Ile343 of D3, both within 5 Å of the metal-binding loop, are destabilized in Apo-P as well. In contrast, phosphorylation of Ser108 stabilizes against HX the amide groups of neighboring residues Thr106, Asn115, and Gly116 of D1, as well as Val183 and Arg247 of D2, each within 8 Å of Ser108.

#### II.4.4 Phosphorylation attracts D4, while dephosphorylation repels D4

The pattern of increased protection from HX due to phosphorylation led to the hypothesis that the negative charge of pSer108 at the base of the catalytic cleft could (i) attract the positively charged walls of D1 and D4 surrounding the cleft and (ii) draw

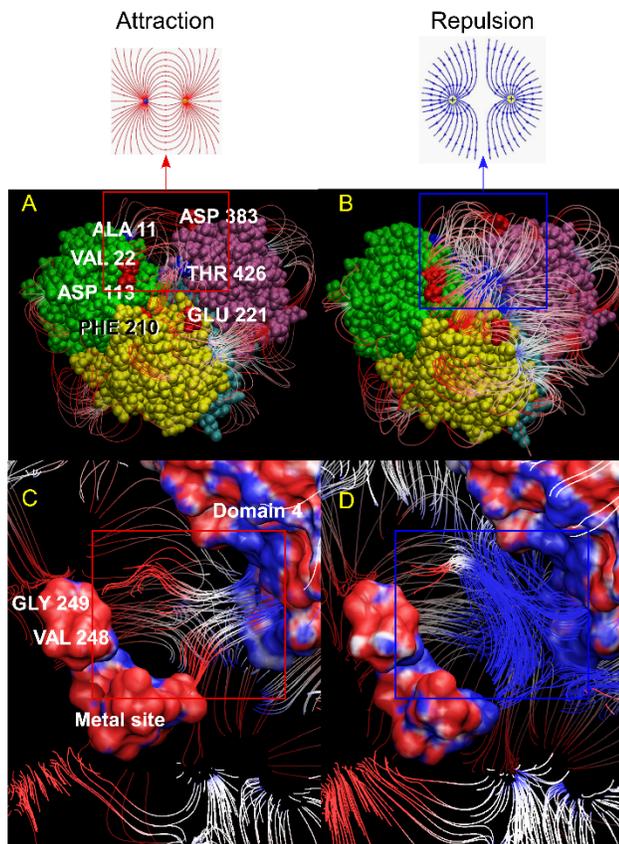


Figure II-4. Domains 1 and 4 are electrostatically attracted when Ser108 is phosphorylated (A,C), but repelled when it is dephosphorylated (B, D). Domains 1 to 4 are colored green, yellow, cyan, and pink, respectively (A, B). (A, C) plot the electrostatic field lines of Apo-P and (B, D) of Apo-deP. (C, D) Expansions of the catalytic cleft highlight attraction and repulsion between D1 and D4 in the red and blue boxes, respectively.

movable D4 closer to phosphoSer108 and D3. We tested this by calculating electrostatic field lines and MD trajectories for structures of both Apo-P and Apo-deP. First, we first carefully simulated the partial charges on the atoms around the divalent cation-binding site, and near Ser108 with and without phosphorylation. Many of the electrostatic

interactions between domains revealed by the field lines are similar between Apo-P and Apo-deP. However, the phosphorylation of Ser108 in D1 profoundly alters at long-range the charge on the face of D4 and negative charge in D1 and D2, as evident in converging field lines to the negative potential near the metal-binding site near Val 248 and Gly249 in D2 (Fig. II-4C). In contrast, Apo-deP switches to electrostatic repulsion between these sites in D2 and D4 (Fig. II-4D), due to the increased positive charge of its catalytic cleft and metal-binding region. This gives rise to the diverging field lines characteristic of repulsion (Fig. II-4D). Several other charged residues on the surface of Apo-P appear to form favorable long-range interactions between D1 and D4 or between D2 and D4, which are absent in Apo-deP. Such favorable field lines in Apo-P which are lost in Apo-deP include those between Ala11 (+) and Asp383 (-), between Thr426 (+) and both Val22 and Asp113 (-), as well as between Thr426 (+) and both Phe210 and Glu221 (-) (Fig. II-4, A and B). Since D4 is attracted to D1 and D2 in Apo-P, D4 is likely to contact the intervening D3 more snugly when it rotates toward D1 and D2.

#### **II.4.5 MD suggests compaction due to phosphorylation and more freedom without it**

To test the working hypotheses that i) D3 and proximal areas of the other domains are stabilized in their H-bonding (higher  $\Delta G_{HX}$ ) and ii) phosphorylation draws D4 with D3, we compared Apo-P and Apo-deP forms by MD. The force field integrates the long-range electrostatics with van der Waals and covalent interactions. The crystal structures of Apo-P and Apo-deP share identical angles of orientation ( $73.2^\circ$ ) and distances of separation ( $28.2 \text{ \AA}$ ) of D4 as defined by Fig. II-5, A and D. The energies and RMSDs from the starting structures stabilized in 5 ns. The distributions of angles and distances

were tabulated from this equilibration through the end of the 120 ns trajectories (Fig. II-5).

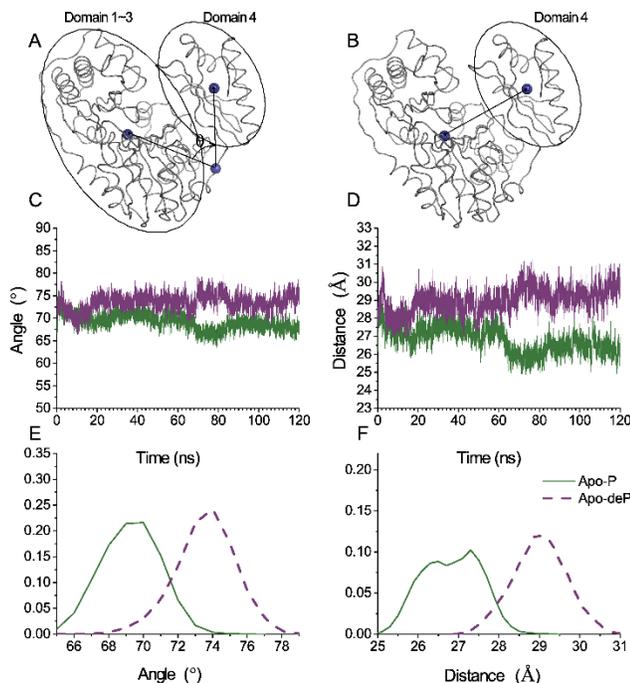


Figure II-5. Domain 4 opens wider in Apo-deP than Apo-P in MD simulations. (A, C, E) plot the angle ( $\theta$ ) between the centers of mass (spheres in A) of D1 – D3 and D4. (B, D, F) mark distances between the center of mass of D4 and the metal site (spheres in B). (C) plots the instantaneous angle between the centers. (D) plots the instantaneous separation between the centers in panel D. (E) This histogram marks the frequency that the angle occurs over the equilibrated portions of the MD trajectories from 5 to 120 ns. (F) This histogram plots marks the frequency at which the separation occurs between 5 and 120 ns. Green represents the Apo-P simulation and purple for Apo-deP.

After 18 ns and 60 ns in the simulation, Apo-deP increased in separation of D4 from D1–D3 and from Apo-P (Fig. II-5). In Apo-P, the angle defined between the centers of mass of D4 and D1–D3 with a vertex at the D3–D4 hinge (Fig. II-5A) averaged  $69.3^\circ \pm 1.7^\circ$  (Fig. II-5, C and E). In Apo-deP, this angle averaged  $73.6 \pm 1.7^\circ$ , i.e. more than  $4^\circ$  wider. In Apo-P, the distances between the catalytic metal site and the center of mass of D4 (Fig. II-5B) averaged  $26.8 \pm 0.7 \text{ \AA}$ , with double maxima around 26.5 and 27.5  $\text{\AA}$  in the histogram, with greater compaction after 60 ns (Fig. II-5, D and F). The distance of

separation in Apo-deP averaged  $29 \pm 0.7 \text{ \AA}$ , i.e. often more than  $2 \text{ \AA}$  greater than Apo-P (Fig. II-5, D and F). D4 of Apo-deP also manifested greater movement relative to D3 by an orthogonal mode not shown. The radius of gyration was also larger in Apo-deP (Fig. SII-10). Although the crystal structures of Apo-P and Apo-deP differ very little (19), the MD simulations imply that Apo-deP samples a wider range of conformations in solution than does Apo-P. MD results for Apo-deP were also reproduced by starting with the Apo-P crystal structure, after removal of its phosphoryl group and recalculating the surrounding partial charges. Further simulations of the S108E mutant (a partial phosphomimetic modeled from the Apo-deP coordinates of PDB code 4MRQ) resulted in a compact form like Apo-P, although with a smaller angle of separation of D4 of  $66$  to  $67^\circ$ . The greater separation and freedom of D4 in Apo-deP simulations strongly support the hypothesis that phosphorylation of Ser108 electrostatically attracts D4, limiting its mobility. This electrostatic compaction by phosphorylation may be called *electrostriction*, a behavior of dielectric materials that is well-known in materials science and acknowledged in biomacromolecules (45).

#### **II.4.6 Phosphorylation effects on simulated backbone mobility and HX**

The range of backbone root-mean-square fluctuations (RMSF) in MD simulations of Apo-P and Apo-deP are generally similar (Fig. II-6A), except as noted here. Phosphorylation enhances the simulated backbone rigidity in D2 near pSer108, the core of D3, part of the D3 interface with D4, and the face of D4 towards the catalytic cleft and D3. Phosphorylation surprisingly appears to increase backbone mobility subtly in parts of D1.

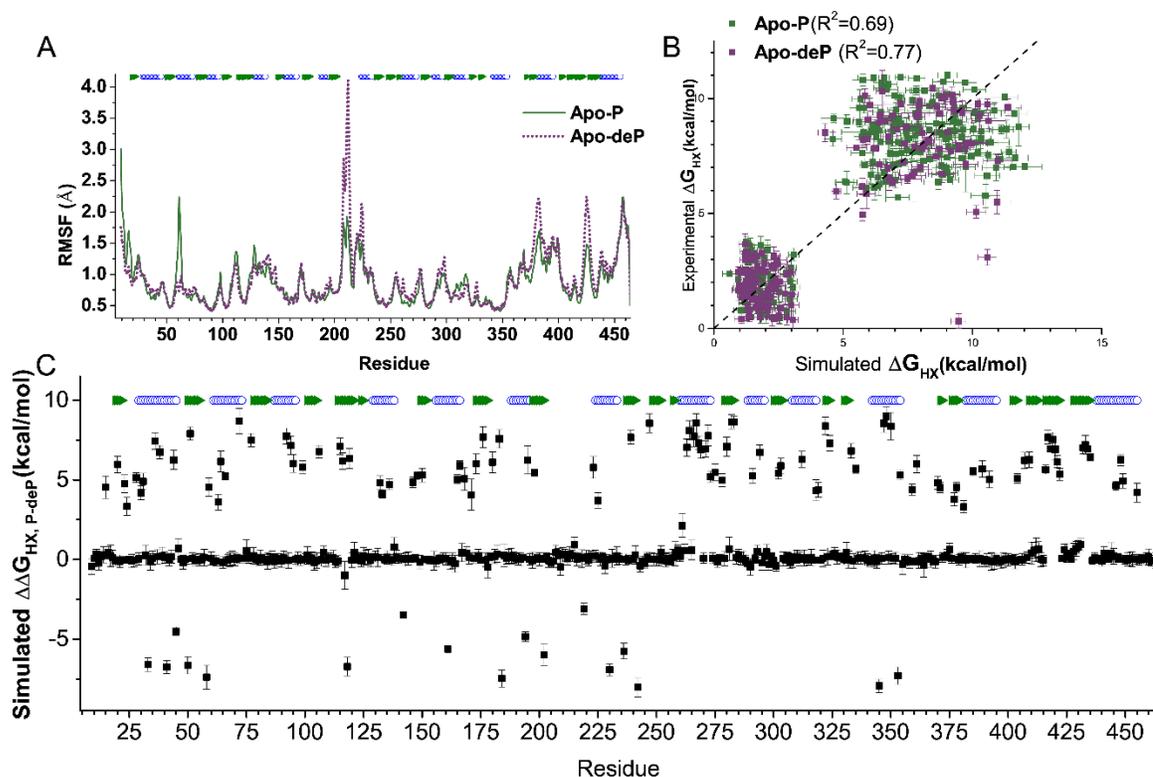


Figure II-6. The fluctuations of the MD trajectories provide simulated free energies of HX and their changes in spatial accord with the measured changes accompanying from phosphorylation. (A) Backbone root-mean squared fluctuation (RMSF) of Apo-P and Apo-deP are compared. (B) Measured  $\Delta G_{HX}$  is plotted against  $\Delta G_{HX}$  simulated from each MD trajectory using  $\ln(p_n^{sim})$  values simulated using Eq. II-3 and the pairs of  $\beta_c$  and  $\beta_{1/RMSF}$  coefficients in the text and at bottom of Table SII-4. (C) The differences in the simulated  $\Delta G_{HX}$  modeled from the MD trajectories of Apo-P and Apo-deP are plotted against the sequence position for comparison with Fig. II-3A.

Correlations of HX protection factors with dynamics simulations via a linear combination of the fluctuating contacts with either H-bonds or  $RMSF^{-1}$  have been demonstrated (23, 24). Using three of the same proteins studied previously, we find that improved correlations can be obtained by optimizing and applying separate sets of coefficients of Eq. II-3 for rapid and slow regimes of HX (Table SII-3), evident through comparison of the previous and new analyses of the HX data of staphylococcal nuclease (Fig. SII-11). We optimized the coefficients of Eq. II-3 for the number of contacts  $N_n^c$  and  $RMSF^{-1}$  as described (24) for agreement with the HDX rates of PMM/PGM across the

120 ns MD trajectories, but with the enhancement of two sets of coefficients. The coefficients obtained are  $\beta_c = 0.22$  and  $\beta_{1/\text{RMSF}} = 0.28$  for the less protected amide groups  $\beta_c = 0.56$  and  $\beta_{1/\text{RMSF}} = 3.94$  for the more protected residues (Table SII-4). We used these coefficients to simulate protection factors and free energies of HX from the MD trajectories for both the rapid and slow cases of HX (Fig. II-6B). Differences in these simulated free energies between the phosphorylated and dephosphorylated states are plotted in Fig. II-6C. Significant increases in  $\ln(P_n^{\text{sim}})$  and  $\Delta G_n^{\text{sim}}$  in Apo-P are based on increased contacts and H-bonding and lower RMSF at those sites in the MD trajectory; see Eq. II-3 above. Though the magnitude of the simulated, large protection factors (free energies in the EX2 regime) by this approach are not quantitatively reliable (23)(Fig. II-6B), the locations and patterns of protection and deprotection by phosphorylation (Fig. II-6C) are in a qualitative sense remarkably consistent with the locations measured to have a strong dependence of HX on phosphorylation status (Fig. II-3A). The simulated *increases* of  $\Delta\Delta G_{\text{HX},P\text{-deP}}$  identify more than 80% of the sites with measured increases (Fig. II-6C, II-3A), i.e. 85% of the amide groups with increased protection in D1, 87% in D2, 83% in D3, and 81% in D4. The simulated *decreases* identify all 14 residues in D1, D2, and D3 with  $\Delta\Delta G_{\text{HX},P\text{-deP}}$  decreased by  $> 2$  kcal/mol (Fig. II-6C, II-3A). Thus, the differences between the MD trajectories of Apo-P and Apo-deP agree with the distribution not only of HX slowed by phosphorylation (densest between D4 and D3), but also of phosphorylation-hastened HX in D1 – D3.

## II.5 Discussion

The studies herein reveal the extent of increased rigidity of PMM/PGM due to phosphorylation, revising a recent qualitative assessment (19) with residue-specific and

quantitative detail. The computational studies agree with the sites of phosphorylation-increased and decreased protection from hydrogen exchange and furnish strong evidence that the structural compaction is electrostatic in nature.

### **II.5.1 Structural compaction by phosphorylation**

Dephosphorylation partially frees D4 in MD simulations, typically increasing its separation from the center of PMM/PGM by 1 to 4 Å (Fig. II-5, D and F). This is consistent with the 1 Å larger radius of gyration of the dephosphorylated form suggested by SAXS (19). Expansion of Apo-deP agrees with the increase of simulated backbone fluctuations at the D3 - D4 interface (Fig. II-6), faster HX at this interface, destabilization of D3 and two  $\beta$ -strands of D4 facing D3 (Fig. II-3), and NMR line broadening evidence of conformational exchange in these regions (Fig. SII-3). The lack of expansion of Apo-deP in its crystal structure (19) was attributed to the restriction of the crystal lattice. Without this constraint, all of the solution measurements and simulations agree upon Apo-deP being loosened and expanded. The compaction of Apo-P may now be attributed to the electrostatic attraction of the positive face of D4 for pSer108 and adjoining electronegative surfaces of D2 (Fig. II-4C), i.e. electrostriction. The closer packing of D4 against D3 and superficially with D1 (Figs. II-7B and II-8A) must result from this, accounting for the longer-lived H-bonds and enhanced stability in these regions.

### **II.5.2 Catalytic relevance of compaction by phosphorylation and loosening by dephosphorylation**

The phosphoryl group at Ser108 of PMM/PGM forms H-bonds with hydroxyl groups of substrates (17) and is transferred to and from the hexose phosphate substrates during catalysis (5, 6)(Fig. II-1). The evidence presented here suggests that the

phosphoryl group also serves to narrow the catalytic cleft of the enzyme (Fig. II-8). This may help position pSer108 and the substrate for H-bond formation and subsequent phosphoryl transfer to the 1- or 6-hydroxyl group of the substrate.

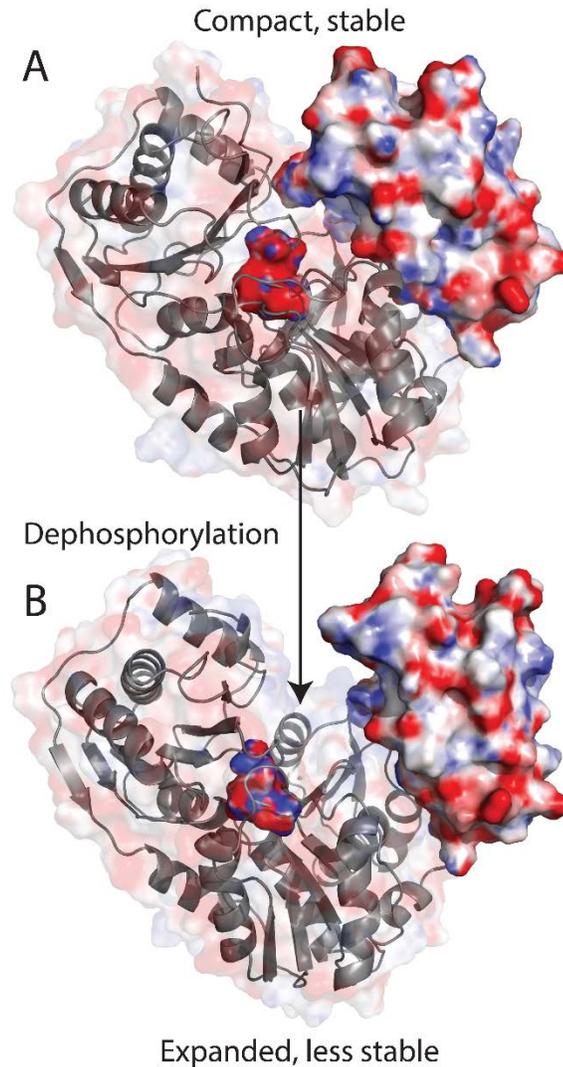


Figure II-7. Expansion of PMM/PGM accompanying its dephosphorylation. The surfaces colored by electrostatics are plotted for D4 and around Ser108 at the base of the cleft. (A) Phosphorylation of Ser108 stabilizes domains of PMM/PGM through electrostatic attraction of D4. Apo-P at 82 ns in the MD trajectory is plotted. (B) Transient dephosphorylation during the catalytic cycle appears to free D4 partially, resulting in an expanded structure with less rigidity in the D3–D4 region. Apo-deP at 117 ns in its MD trajectory is plotted.

The unusual 180° rotation of G16P and mannose 1,6-bisphosphate intermediates, which occurs in the middle of the catalytic cycle *while associated* with PMM/PGM (8),

may also be facilitated by changes in the dynamics of the transiently dephosphorylated enzyme. Measurements of accessibility to probes (Fig. II-2), HX (Figs. II-3 and II-S6), radius of gyration (19), electrostatics (Fig. II-4), and MD simulations (Figs. II-5~7) establish that Apo-deP has a wider catalytic cleft and is looser overall (Fig. II-7), except for small portions of D1 and D2 that may gain rigidity upon dephosphorylation (Figs. II-3 and II-7 and Table SII-2). Expansion and opening of the cleft could help accommodate the catalytically required 180° reorientation of the bisphosphorylated intermediate.

The stabilization of the active site of PMM/PGM by phosphorylation is less than that at distal sites (Fig. II-3). Despite coordination of  $Mg^{2+}$  by three aspartate side chains in the active site, phosphorylation of Ser108 nearby might introduce a degree of repulsion from these three carboxylate groups. Perhaps this accounts for the phospho-destabilization of  $Mg^{2+}$  ligand Asp242, a few neighboring residues, and other residues in D2. Locally diminished stability near the active site due to phosphorylation required for enzyme activity could be reminiscent of observed trends of tradeoffs between stability and function in which clefts and their key residues for catalysis tend to destabilize the protein fold (46, 47).

### **II.5.3 Insights into HX by NMR beyond qualitative HDX-MS**

The NMR studies reveal that the HX behavior within PMM/PGM is more complex and varied than previously appreciated. The most obvious trend of phosphorylation-increased rigidity and stability to HX is clearly long in range, with the distant core of D3 and proximal face of D4 being especially shielded by Ser108 phosphorylation (Fig. II-3). However, the residue-by-residue accounting provided by NMR also implicates ~99 residues with HX unaffected by phosphorylation and two

dozen residues with HX *increased* by phosphorylation (Fig. II-3). This is corroborated by comparison with the changes in HX protection factors (free energies) simulated from the MD trajectories (Fig. II-6C). Among the 237 residues where  $\Delta\Delta G_{HX,P-deP}$  comparisons of HX are available, 48% represent significant slowing by phosphorylation, 10% undergo acceleration by phosphorylation, and 42% have relatively small dependence on phosphorylation (Fig. II-3A). Computational results are consistent with this but suggest a much higher proportion of residues to be unaffected by phosphorylation (Fig. II-6C). Thus, the HDX-MS-based proposal that flexibility increases of PMM/PGM upon dephosphorylation are global (19) should not be construed as universal or comprehensive in light of the NMR and computational data. Six of the peptides monitored by HDX-MS are seen by NMR to have at least two residues each with increased HX upon Ser108 phosphorylation (Table SII-2). Four more peptides had single residues accelerated in HX by NMR. Most of the regions with HX expedited by phosphorylation (in Apo-deP) map to portions of D1 and D2, including Asp242 at the active site (Fig. II-3 and II-6C; Movie II-S1). Observation of the varying and nuanced HX behaviors including local destabilizations within this multi-domain enzyme required the residue resolution and quantification of NMR (enhanced by sub-sec and long time scales). The phosphorylation-induced changes in rigidity implicated are long in range, with increases in stability widespread while decreases are localized.

The NMR methods have also provided the first detailed HX insight on functionally important loops within the active site of PMM/PGM, where HDX-MS did not contribute. For example, HDX by NMR identifies the residues flanking Ser108 in sequence as stabilized by its phosphorylation (Fig. II-3), while mass spectrometry failed

to detect the peptide containing Ser108 when it was phosphorylated (92-117 in Table SII-2). Neighboring residues in D2 and D3 destabilized by phosphoSer108 were already noted (Fig. II-3). NMR also reveals key ligand-binding residues protected by phosphorylation (in two other peptides): (i) the C-terminal half of HDX-MS peptide 14-27 (containing key Arg20 at the active site) and (ii) the middle third of peptide 407-429 near Arg421 and Ser423 that bind the phosphate group of ligands.

NMR has quantified the rates and thermodynamics of HX of specific amide groups, enabling examination of the stabilities of individual H-bonds, domains, and the enzyme as a whole. This revealed phosphorylation-dependent increases in stability that are locally large for individual H-bonds ( $> 4$  kcal/mol at many sites in D1, D3, and D4), moderate in the folding stability of D3 and D4 ( $\geq 1.3$  kcal/mol), and more subtle for the enzyme as a whole ( $\geq 0.4$  kcal/mol) (Figs. II-3, II-S6, and II-S9). These and localized variations in stability (Figs. II-3 and II-6; Movies II-S1 and II-S2) offer a complex but unified biophysical portrayal. Combining HX methods and dynamics simulations (Fig. II-3 and II-6) may have wider applicability to appreciating varied effects of phosphorylation on proteins.

#### **II.5.4 Long-range effects of phosphorylation on HDX of enzymes**

Phosphorylation of Ser108 slowed HDX to the furthest reaches of D2 – D4 (Fig. II-3). Moreover, the addition of two negative charges at the site of phosphorylation is seen to attract D4 in both static and dynamic simulations, resulting in domain rotation about the flexible hinge between D3 and D4 (Figs. II-5~7). These results are reminiscent of those due to phosphorylation in other systems. For example, phosphorylation of the activation segment of p38 MAP kinase at two sites or protein kinase A (PKA) at a single

site analogously slowed HDX of loops at long-range and locally in the activation segment, and increased the folding stability of PKA (48, 49). In the protein kinase ERK2, dual phosphorylation of the activation segment also induced long-range changes in HDX (mostly small decreases) as well as increases in four peptides suggesting functionally relevant mobility in its ATP-binding loop and activation segment (50). The results on PMM/PGM suggest that the trend of phosphorylation restricting motions at long-range, while enabling some motions closer to the active site, may be shared among some members of protein kinase and  $\alpha$ -D-phosphohexomutase superfamilies, which are mechanistically very different phosphoryl transfer enzymes. Mobility appears to modulate substrate access to protein kinases and the 180° reorientation of the intermediate in dephosphorylated PMM/PGM.

## II.6 Conclusions

Calculations strongly suggest that phosphorylation of Ser108 in domain 1 (D1) electrostatically attracts D4, narrowing the catalytic cleft and breadth of PMM/PGM. This electrostriction accounts for the properties of the phosphorylated form (Apo-P) demonstrated above: less accessibility to probes, enhanced folding stability, stabilization of many H-bonds by > 4 kcal/mol, much slowing of HX with stability increases > 1.3 kcal/mol for D3 and the adjacent face of D4, and increased rigidity in the stabilized D3 – D4 interface. Narrowing of the active site cleft due to phosphorylation probably aids capture of phosphohexose substrates while opening upon dephosphorylation aids rotation of the bisphosphorylated intermediate. NMR and MD simulation paint a more complex, varied, and detailed landscape of HX changes than does peptide-resolved HDX-MS. The proposal of flexibility increases in the dephosphorylated form (Apo-deP) being global

(19) must be tempered by the large number of unaffected sites and some sites with dephosphorylation-induced *slowing* of HX, especially in D1 and D2, which is also supported by MD trajectories. The rigidity increases upon phosphorylation are widespread, being most concentrated in D3 and D4 that draw closer upon phosphorylation. The ~100 sites with phosphorylation-promoted stability increases, as well as the stability decreases in D1 and D2, are truly long-range effects that reach the most distant flanks of the enzyme. Such behaviors parallel HDX-MS characterizations of protein kinases, suggesting analogous long-range effects of phosphorylation in both superfamilies of multi-domain enzymes.

## II.7 Acknowledgement

We thank A. Hopkins for enzyme preparations and S. Jiang for assistance with interleaving of the CLEANEX pulse sequence.

This work was supported by National Science Foundation grants MCB 0918389 and 1409898. National Institutes of Health grant S10RR022341 contributed to the purchase of the 800 MHz NMR spectrometer.

## II.8 References

1. Johnson, L. N., and R. J. Lewis. 2001. Structural Basis for Control by Phosphorylation. *Chem. Rev.* 101:2209-2242.
2. Taylor, S. S., M. M. Keshwani, J. M. Steichen, and A. P. Kornev. 2012. Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367:2517-2528.
3. Xin, F., and P. Radivojac. 2012. Post-translational modifications induce significant yet not extreme changes to protein structure. *Bioinformatics* 28:2905-2913.
4. Johnson, L. N., and M. O'Reilly. 1996. Control by phosphorylation. *Curr. Opin. Struct. Biol.* 6:762-769.
5. Naught, L. E., and P. A. Tipton. 2001. Kinetic Mechanism and pH Dependence of the Kinetic Parameters of *Pseudomonas aeruginosa*

- Phosphomannomutase/Phosphoglucomutase. *Archives of Biochemistry and Biophysics* 396:111-118.
6. Regni, C., P. A. Tipton, and L. J. Beamer. 2002. Crystal structure of PMM/PGM: an enzyme in the biosynthetic pathway of *P. aeruginosa* virulence factors. *Structure* 10:269-279.
  7. Ray, W. J., Jr., and G. A. Roscelli. 1964. A Kinetic Study of the Phosphoglucomutase Pathway. *J Biol Chem* 239:1228-1236.
  8. Naught, L. E., and P. A. Tipton. 2005. Formation and reorientation of glucose 1,6-bisphosphate in the PMM/PGM reaction: transient-state kinetic studies. *Biochemistry* 44:6831-6836.
  9. Regni, C., A. M. Schramm, and L. J. Beamer. 2006. The Reaction of Phosphohexomutase from *Pseudomonas aeruginosa*: Structural Insights Into A Simple Processive Enzyme. *J Biol Chem* 281:15564-15571.
  10. Ballok, A. E., and G. A. O'Toole. 2013. Pouring Salt on a Wound: *Pseudomonas aeruginosa* Virulence Factors Alter Na<sup>+</sup> and Cl<sup>-</sup> Flux in the Lung. *J. Bacteriol.* 195:4013-4019.
  11. Li, X. J., Q. Li, L. Y. Si, and Q. Y. Yuan. 2011. Bacteriological Differences Between Patients with Acute Exacerbation of COPD and Community-Acquired Pneumonia. *Respir. Care* 56.
  12. Govan, J. R. W., and V. Deretic. 1996. Microbial pathogenesis in cystic fibrosis - Mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiological Reviews* 60:539-574.
  13. Ye, R. W., N. A. Zielinski, and A. M. Chakrabarty. 1994. Purification and characterization of phosphomannomutase/phosphoglucomutase from *Pseudomonas aeruginosa* involved in biosynthesis of both alginate and lipopolysaccharide. *J. Bacteriol.* 176:4851-4857.
  14. King, J. D., D. Kocíncová, E. L. Westman, and J. S. Lam. 2009. Review: Lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa*. *Innate Immunity* 15:261-312.
  15. Olvera, C., J. B. Goldberg, R. Sanchez, and G. Soberon-Chavez. 1999. The *Pseudomonas aeruginosa* algC gene product participates in rhamnolipid biosynthesis. *FEMS Microbiology Letters* 179:85-90.
  16. Remminghorst, U., and B. H. Rehm. 2006. Bacterial alginates: from biosynthesis to applications. *Biotechnol. Lett.* 28:1701-1712.
  17. Regni, C., L. E. Naught, P. A. Tipton, and L. J. Beamer. 2004. Structural basis of diverse substrate recognition by the enzyme PMM/PGM from *P. aeruginosa*. *Structure* 12:55-63.
  18. Sarma, A. V., A. Anbanandam, A. Kelm, R. Mehra-Chaudhary, Y. Wei, P. Qin, Y. Lee, M. V. Berjanskii, J. A. Mick, L. J. Beamer, and S. R. Van Doren. 2012. Solution NMR of a 463-residue phosphohexomutase: domain 4 mobility, substates, and phosphoryl transfer defect. *Biochemistry* 51:807-819.
  19. Lee, Y., M. T. Villar, A. Artigues, and L. J. Beamer. 2014. Promotion of Enzyme Flexibility by Dephosphorylation and Coupling to the Catalytic Mechanism of a Phosphohexomutase. *J. Biol. Chem.* 289:4674-4682.
  20. Coales, S. J., S. Y. E, J. E. Lee, A. Ma, J. A. Morrow, and Y. Hamuro. 2010. Expansion of time window for mass spectrometric measurement of amide

- hydrogen/deuterium exchange reactions. *Rapid Commun. Mass Spectrom.* 24:3585-3592.
21. Mayne, L., Z.-Y. Kan, P. Sevugan Chetty, A. Ricciuti, B. Walters, and S. W. Englander. 2011. Many Overlapping Peptides for Protein Hydrogen Exchange Experiments by the Fragment Separation-Mass Spectrometry Method. *J. Am. Soc. Mass Spectrom.* 22:1898-1905.
  22. Huyghues-Despointes, B. M., J. M. Scholtz, and C. N. Pace. 1999. Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat. Struct. Biol.* 6:910-912.
  23. Vendruscolo, M., E. Paci, C. M. Dobson, and M. Karplus. 2003. Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. *J. Am. Chem. Soc.* 125:15686-15687.
  24. Kieseritzky, G., G. Morra, and E.-W. Knapp. 2006. Stability and fluctuations of amide hydrogen bonds in a bacterial cytochrome c: a molecular dynamics study. *J. Biol. Inorg. Chem.* 11:26-40.
  25. Cellini, B., R. Montioli, A. Paiardini, A. Lorenzetto, F. Maset, T. Bellini, E. Oppici, and C. B. Voltattorni. 2010. Molecular defects of the glycine 41 variants of alanine glyoxylate aminotransferase associated with primary hyperoxaluria type I. *Proc Natl Acad Sci U S A* 107:2896-2901.
  26. Lescop, E., P. Schanda, and B. Brutscher. 2007. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* 187:163-169.
  27. Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6:277-293.
  28. Goddard, T. D., and D. G. Kneller. 2000. SPARKY. University of California, San Francisco, San Francisco.
  29. Hwang, T. L., P. C. van Zijl, and S. Mori. 1998. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *Journal of Biomolecular NMR* 11:221-226.
  30. Hernández, G., and D. M. LeMaster. 2003. Relaxation compensation in chemical exchange measurements for the quantitation of amide hydrogen exchange in larger proteins. *Magn. Reson. Chem.* 41:699-702.
  31. Bertini, I., K. Ghosh, A. Rosato, and P. R. Vasos. 2003. A high-resolution NMR study of long-lived water molecules in both oxidation states of a minimal cytochrome c. *Biochemistry* 42:3457-3463.
  32. Peters, M. B., Y. Yang, B. Wang, L. Fusti-Molnar, M. N. Weaver, and K. M. Merz, Jr. 2010. Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *J Chem Theory Comput* 6:2935-2947.
  33. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E.

- Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. 2009. Gaussian 09, Revision B.01. Wallingford CT.
34. Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* 65:712-725.
  35. Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98:10037-10041.
  36. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33-38, 27-38.
  37. Yamasaki, K., T. Daiho, S. Danko, and H. Suzuki. 2013. Roles of long-range electrostatic domain interactions and K<sup>+</sup> in phosphoenzyme transition of Ca<sup>2+</sup>-ATPase. *J Biol Chem* 288:20646-20657.
  38. Craddock, T. J., J. A. Tuszynski, and S. Hameroff. 2012. Cytoskeletal signaling: is memory encoded in microtubule lattices by CaMKII phosphorylation? *PLoS Comput Biol* 8:e1002421.
  39. Schonbrunn, E., S. Eschenburg, K. Luger, W. Kabsch, and N. Amrhein. 2000. Structural basis for the interaction of the fluorescence probe 8-anilino-1-naphthalene sulfonate (ANS) with the antibiotic target MurA. *Proc Natl Acad Sci U S A* 97:6345-6349.
  40. Bai, Y., T. R. Sosnick, L. Mayne, and S. W. Englander. 1995. Protein folding intermediates: native-state hydrogen exchange. *Science* 269:192-197.
  41. Linderstrøm-Lang, K. 1955. Deuterium exchange between peptides and water. *Chem. Soc. (London) Spec. Publ.* 2:1-20.
  42. Yan, S., S. D. Kennedy, and S. Koide. 2002. Thermodynamic and kinetic exploration of the energy landscape of *Borrelia burgdorferi* OspA by native-state hydrogen exchange. *Journal of Molecular Biology* 323:363-375.
  43. Liang, X., G. I. Lee, and S. R. Van Doren. 2006. Partially unfolded forms and non-two-state folding of a beta-sandwich: FHA domain from Arabidopsis receptor kinase-associated protein phosphatase. *Journal of Molecular Biology* 364:225-240.
  44. Huyghues-Despointes, B. M., U. Langhorst, J. Steyaert, C. N. Pace, and J. M. Scholtz. 1999. Hydrogen-exchange stabilities of RNase T1 and variants with buried and solvent-exposed Ala → Gly mutations in the helix. *Biochemistry* 38:16481-16490.
  45. Marky, L. A., and D. W. Kupke. 2000. Enthalpy-entropy compensations in nucleic acids: Contribution of electrostriction and structural hydration. In *Methods Enzymol.* G. K. A. Michael L. Johnson, editor. Academic Press. 419-441.

46. Beadle, B. M., and B. K. Shoichet. 2002. Structural bases of stability-function tradeoffs in enzymes. *J. Mol. Biol.* 321:285-296.
47. Liang, X., A. Arunima, Y. Zhao, R. Bhaskaran, A. Shende, T. S. Byrne, J. Fleeks, M. O. Palmier, and S. R. Van Doren. 2010. Apparent tradeoff of higher activity in MMP-12 for enhanced stability and flexibility in MMP-3. *Biophys. J.* 99:273-283.
48. Sours, K. M., S. C. Kwok, T. Rachidi, T. Lee, A. Ring, A. N. Hoofnagle, K. A. Resing, and N. G. Ahn. 2008. Hydrogen-Exchange Mass Spectrometry Reveals Activation-Induced Changes in the Conformational Mobility of p38 $\alpha$  MAP Kinase. *J. Mol. Biol.* 379:1075-1093.
49. Steichen, J. M., G. H. Iyer, S. Li, S. A. Saldanha, M. S. Deal, V. L. Woods, and S. S. Taylor. 2010. Global Consequences of Activation Loop Phosphorylation on Protein Kinase A. *J. Biol. Chem.* 285:3825-3832.
50. Hoofnagle, A. N., K. A. Resing, E. J. Goldsmith, and N. G. Ahn. 2001. Changes in protein conformational mobility upon activation of extracellular regulated protein kinase-2 as detected by hydrogen exchange. *Proc. Natl. Acad. Sci. U.S.A.* 98:956-961.

## II.9 Supporting material

### II.9.1 Supporting figures

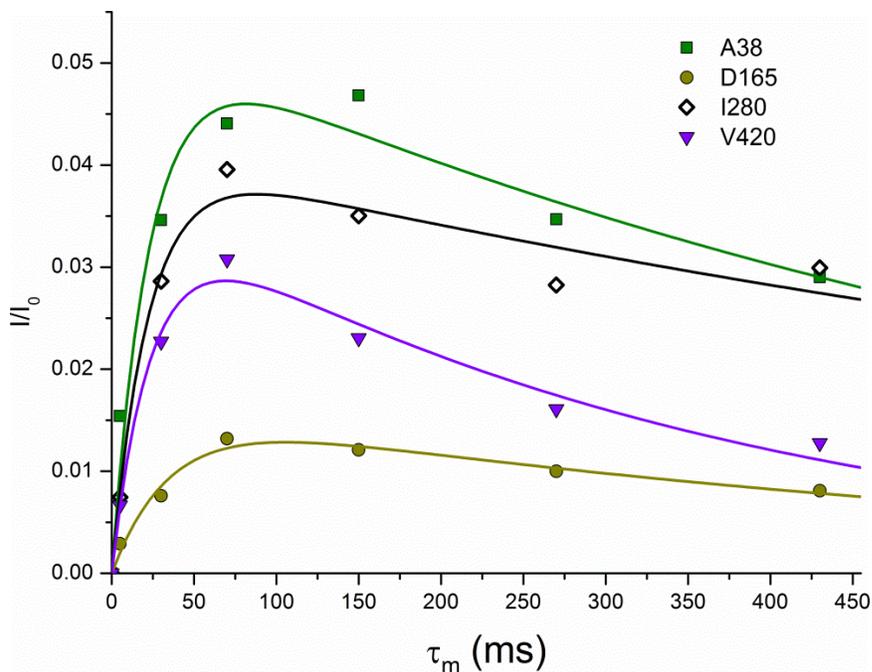


Figure SII-1. Examples of CLEANEX-PM time courses for the Apo-deP form of PMM/PGM. The fitted rate constants,  $k_{ex}$ , of Ala38, Asp165, Ile280, and Val420 are  $2.8 \pm 0.8 \text{ s}^{-1}$ ,  $0.5 \pm 0.1 \text{ s}^{-1}$ ,  $1.9 \pm 0.3 \text{ s}^{-1}$ , and  $1.1 \pm 0.2 \text{ s}^{-1}$ , respectively. These rate constants were used to simulate sub-second components of hydrogen exchange in the inset of Fig. II-2D. The spin-lock during the mixing period used a radio frequency field strength of 6.25 kHz. To maintain sample conditions consistent among the spectra during acquisition, the CLEANEX spectra were acquired in an interleaved manner and stored as a pseudo 3D spectrum. The interleaved pulse sequence and script developed to acquire and process the spectra are available as other Supplemental files.

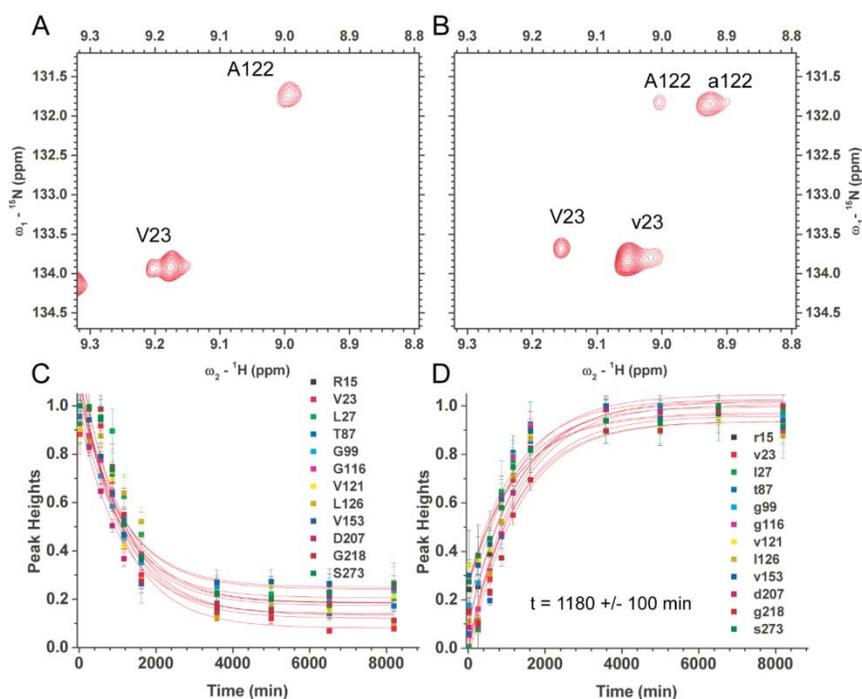


Figure SII-2. Kinetics of the Apo-P form of PMM/PGM undergoing dephosphorylation. (A) Peaks of Val23 and Arg122 of Apo-deP 15 min after preparation are shown. (B) Peaks of the same sample after 60 h of dephosphorylation are shown. (C) The time course of the disappearance of 12 normalized peaks representing the phosphorylated form is plotted. (D) The time course of the appearance of the peaks of the dephosphorylated form is plotted for the same 12 residues. These 12 used for illustration are a subset of the 43 actually used in global fits of the dephosphorylation kinetics that yielded a time constant of  $1140 \pm 70$  min or  $19.0 \pm 1.1$  h (where the uncertainty quoted is four-fold the fitting uncertainty). The TROSY NMR spectra were acquired at 35 °C, pH 7.5, 10 mM DTT, and 800 MHz.

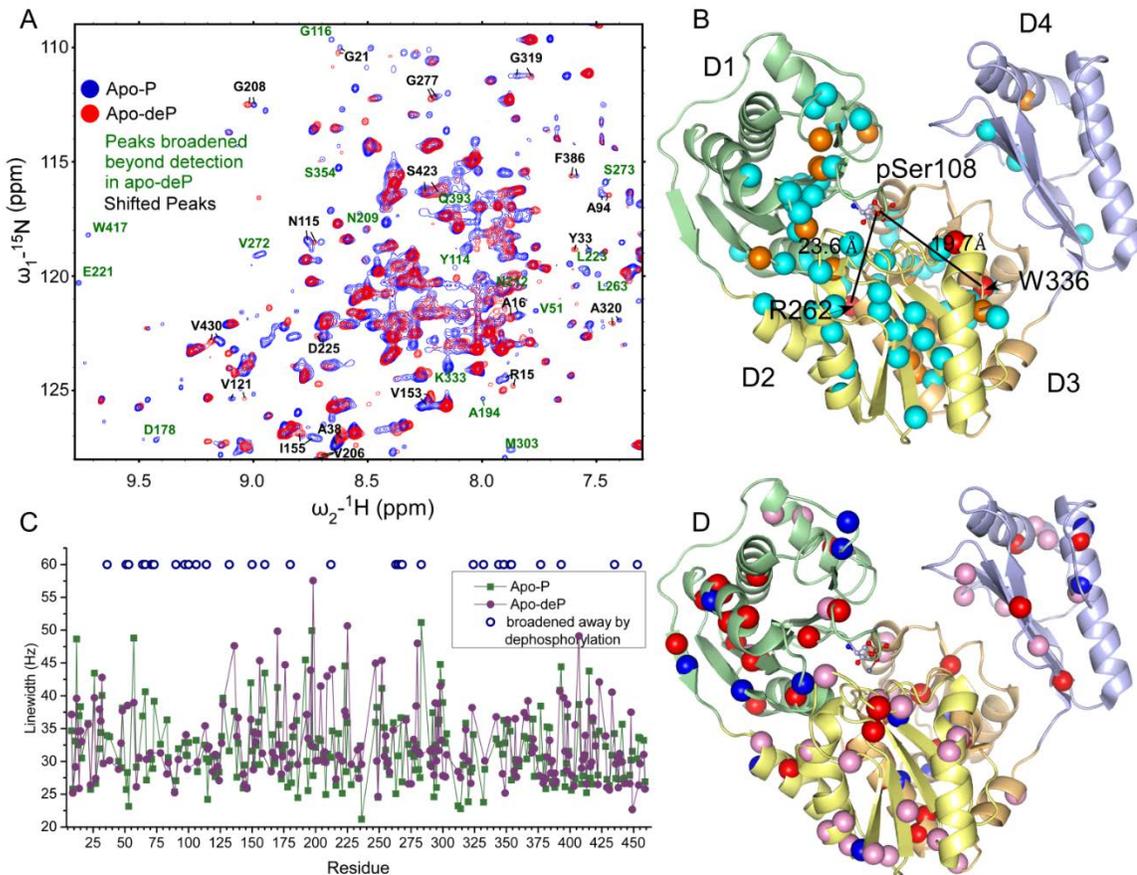


Figure SII-3. Effects of (de)phosphorylation on NMR spectra of PMM/PGM. (A)  $^{15}\text{N}$  TROSY spectra of phosphorylated PMM/PGM (Apo-P; blue contours) and dephosphorylated PMM/PGM (Apo-deP; red) are superimposed. Dephosphorylation broadened beyond detection the peaks with green labels. The greater number of peaks of Apo-P reflects their being sharper and more readily detected. Conditions were 1 mM enzyme, pH 7.4, 308 K, and 800 MHz. (B) Locations of chemical shift perturbations introduced by dephosphorylation are marked on the crystal structure of Apo-P (PDB: 1K35) with red spheres for  $\Delta\omega_{\text{HN}} > 0.2$  ppm, orange for  $\Delta\omega_{\text{HN}} > 0.1$  ppm, and cyan for  $\Delta\omega_{\text{HN}} > 0.03$  ppm, where  $\Delta\omega_{\text{HN}} = ((\Delta\omega_{\text{H}})^2 + (\Delta\omega_{\text{N}}/5)^2)^{1/2}$ . The locations of Arg262 and Trp336 with the peaks most shifted by dephosphorylation are labeled along with their distances from pSer108. (C) Peaks broadened away by dephosphorylation are marked by open circles at top. The line widths plotted are each a composite of  $^1\text{H}$  and  $^{15}\text{N}$  line widths at half-height:  $((\Delta\nu_{1/2,\text{H}})^2 + (\Delta\nu_{1/2,\text{N}})^2)^{1/2}$ . (D) Amide peak broadenings by dephosphorylation are mapped on the structure in red where broadened away and pink where broadened  $> 20\%$ , whereas blue marks amide peaks  $> 20\%$  sharper in Apo-deP.

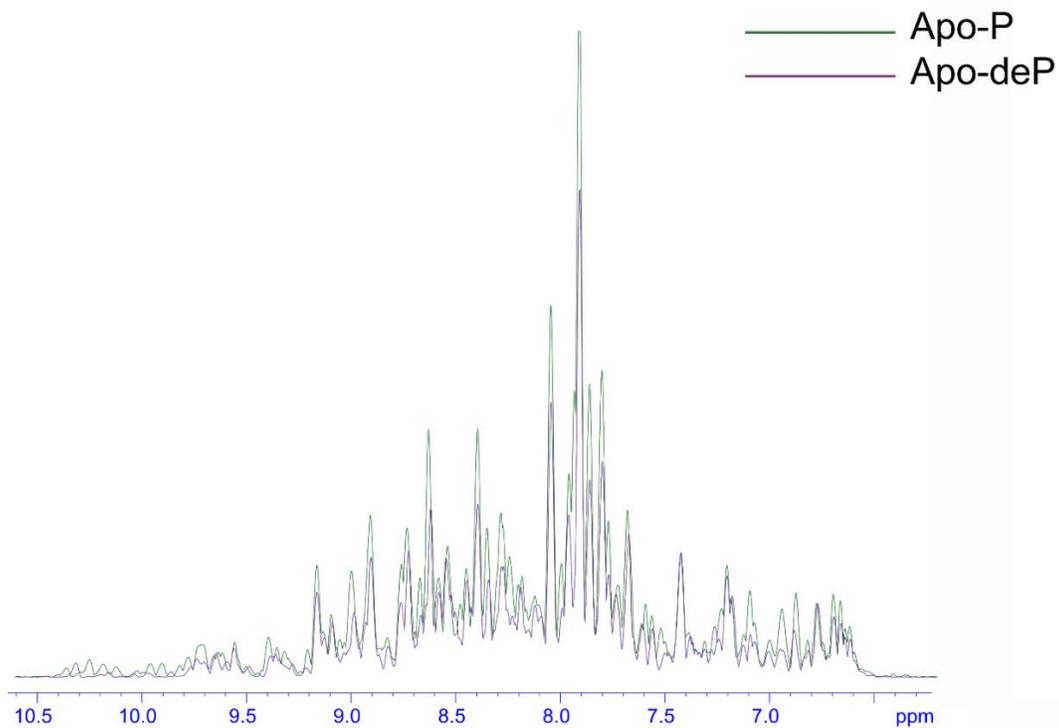


Figure SII-4. Skyline projections of 2D TROSY spectra of Apo-P before (green) and after more than 3 days at 35 °C where it has undergone dephosphorylation to Apo-deP (purple).

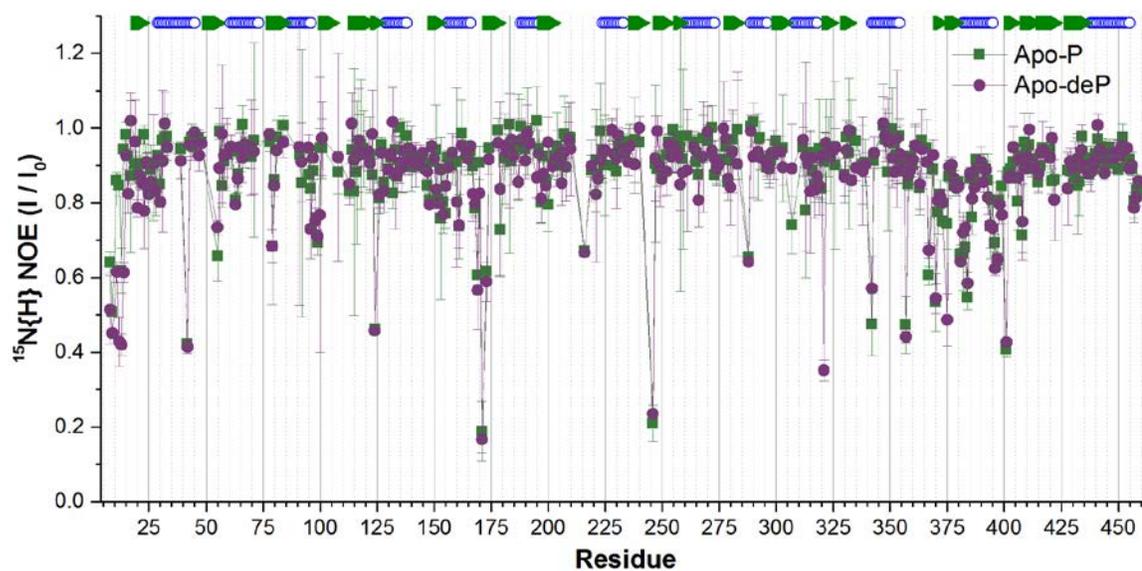


Figure SII-5. Dependence of  $^{15}\text{N}\{^1\text{H}\}$  heteronuclear, steady-state NOE on phosphorylation status of PMM/PGM. Peak height with  $^1\text{H}$  signal saturated is designated I and without saturation as  $I_0$ . The measurements were performed in triplicate at 800 MHz and 35 °C. The values plotted are mean  $\pm$  SD ( $n=3$ ) for Apo-P in purple circles and Apo-deP in green squares. The locations of  $\beta$ -strands are marked with green arrows and helices with blue circles.

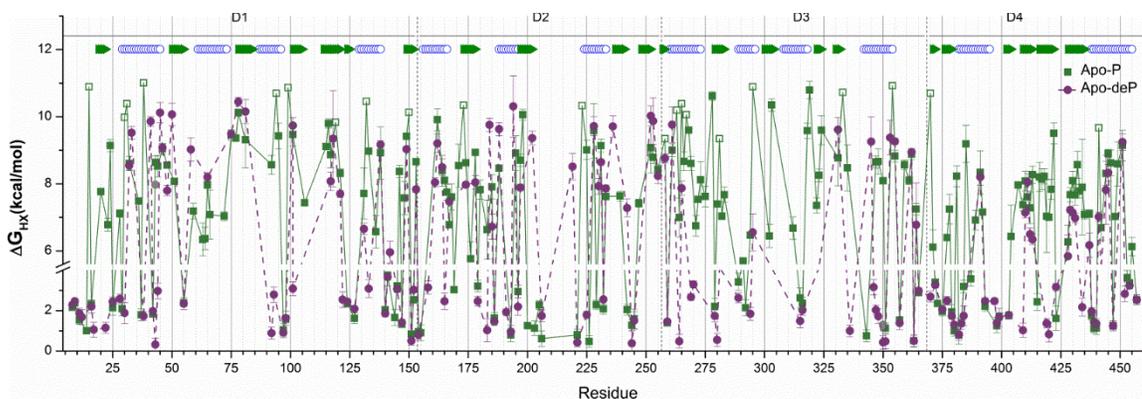


Figure SII-6. Phosphorylation slows HX and stabilizes domain 3 in particular.  $\Delta G_{HX}$  values from slow HDX appear above the vertical break, while the small  $\Delta G_{HX}$  values of rapid HX, measured on the sub-sec scale by CLEANEX-PM NMR (1), appear below the break. Open squares represent lower bounds of the  $\Delta G_{HX}$  estimates for the especially slowly exchanging amide groups of Apo-P. The locations of strands and helices are marked at the top by triangles and blue circles, respectively.

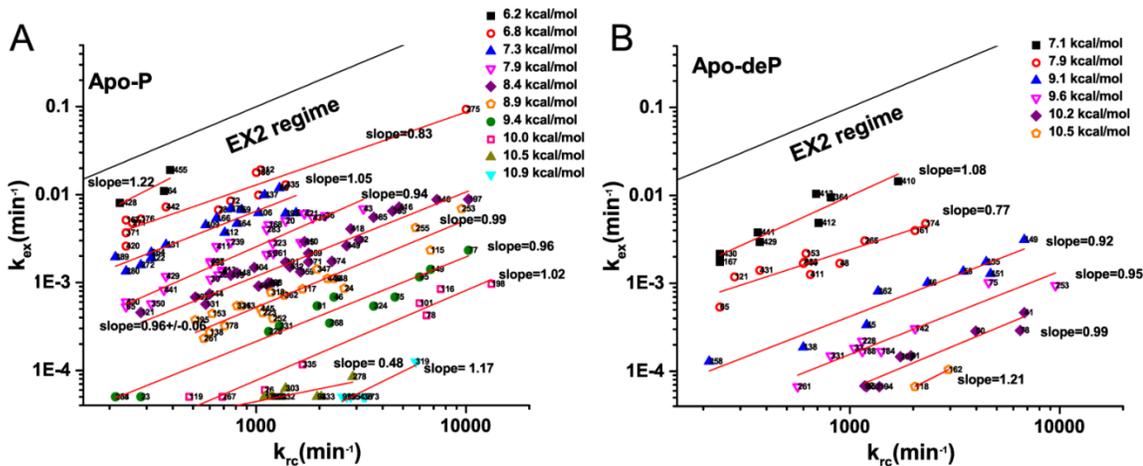


Figure SII-7. Hydrogen exchange results grouped by  $\Delta G_{HDX,apparent}$  suggest the independence of  $\Delta G_{HDX,apparent}$  from size of  $k_{rc}$ , implying that the EX2 regime prevails for both Apo-P (A) and Apo-deP (B). The reference line depicts the slope of 1.0 typifying the theoretical behavior of the EX2 regime. EX1 has a slope of zero. Residue numbers are listed by each point.

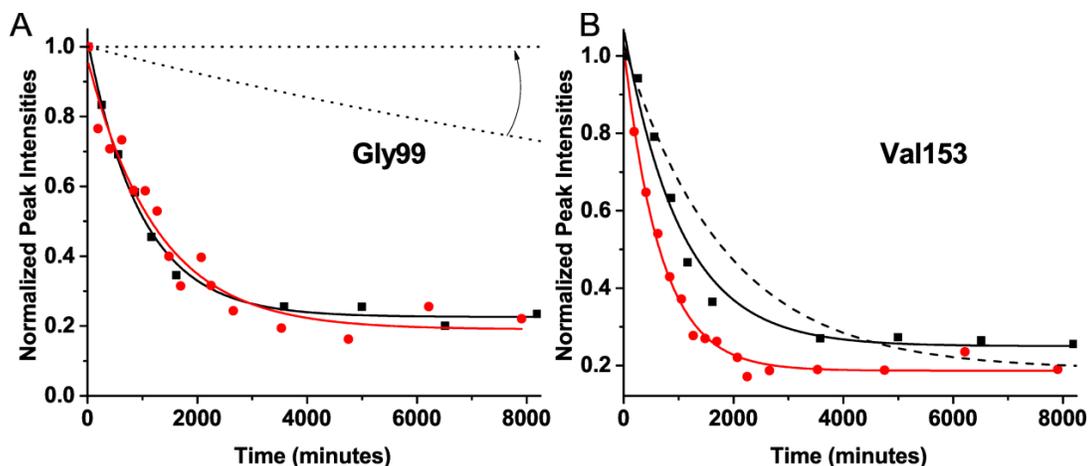


Figure SII-8. Examples of data fitting of Apo-P. Since Apo-P undergoes dephosphorylation, its  $k_{ex}$  of HDX is calculated as  $k_{ex} = k_{obs} - k_{deP}$ . Black squares plot the apparent deuterium exchange curve in  $D_2O$ , providing  $k_{obs}$ . Red circles plot the decay of the peak in  $H_2O$  due to dephosphorylation, providing  $k_{deP}$ . Dashed lines mark the computed curve for  $k_{ex}$  from which effects of dephosphorylation are removed. (A) In the case of Gly99, the similarity of the two measured decays implies that  $k_{ex}$  approaches zero. The upper limit upon its  $k_{ex}$  is represented by the lower dashed line. Yet this cannot be experimentally distinguished from  $k_{ex} = 0$  represented by the horizontal line, delimiting the other end of the range of potential  $k_{ex}$  values. (B) A simple case is illustrated for Val153 where  $k_{ex} = k_{obs} - k_{deP}$ .

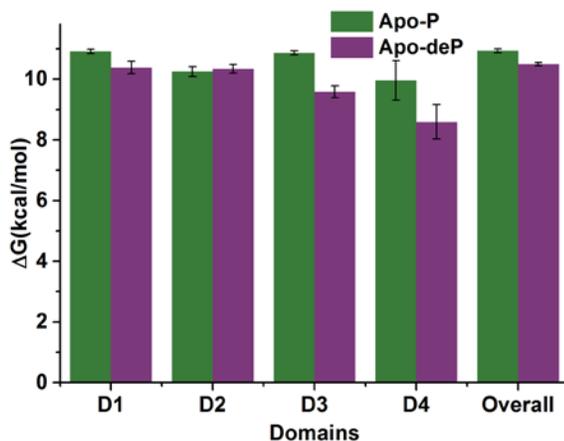


Figure SII-9. Phosphorylation slows HDX and stabilizes D3 and D4. The three highest  $\Delta G_{HDX}$  values of each domain, or the enzyme as a whole, suggest the folding stability of that domain and its standard deviation by the method of ref (2) employed. The folding stabilities of Apo-P have been corrected for proline isomerization according to the method (2). The stabilities of Apo-P and its domains 1, 3, and 4 should be regarded as lower limits because the true size of their largest  $k_{ex}$  values (used for these  $\Delta G_{fold}$  estimates) is masked by the rate of dephosphorylation.

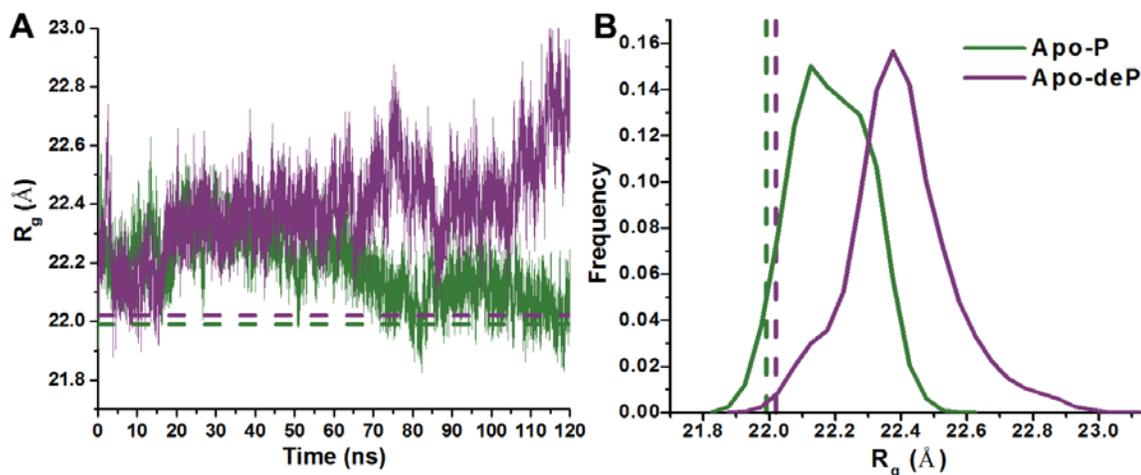


Figure SII-10. Radii of gyration ( $R_g$ ) during MD trajectories indicate that the Apo-deP state (purple) is more expanded than Apo-P (green). Dashed lines represent the radii of the crystal structures of Apo-deP and Apo-P (PDB codes 4MRQ and 1K35, respectively). Panel (A) plots the time courses of  $R_g$  in the MD trajectories and (B) the histogram of  $R_g$  in the trajectories.

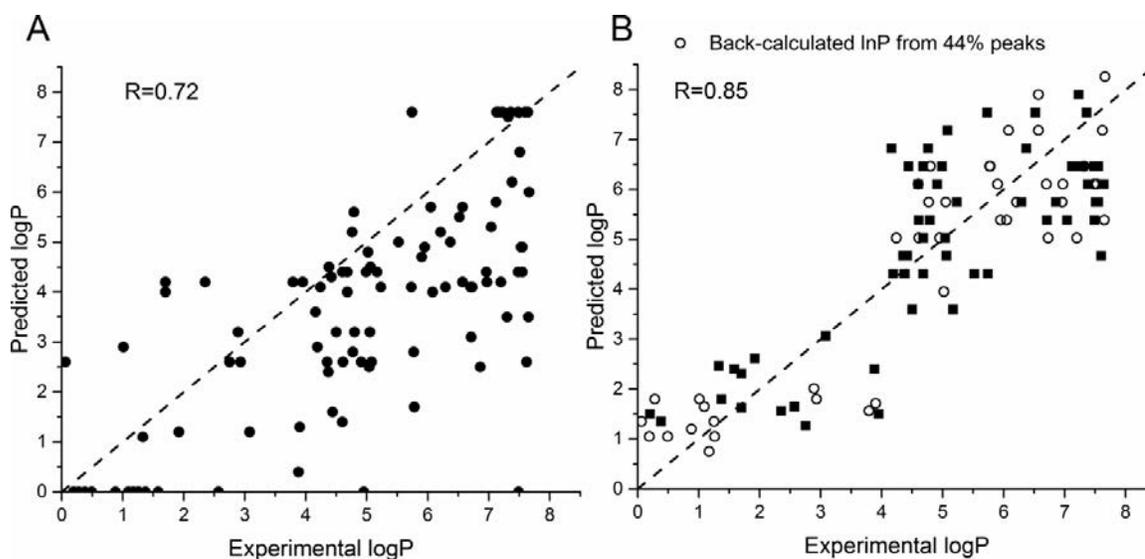


Figure SII-11. Correlation of experimentally measured HX protection factors  $\ln(P^{exp})$  of staphylococcal nuclease with those simulated from its coordinates using (A) a single set of  $\beta_C$  and  $\beta_H$  coefficients as reported (3) or (B) two sets of coefficients optimized for rapid and slow regimes of HX using 44% of its residues (black squares); see Eq. II-3. The correlation is effective for the randomly chosen 31% of its residues not used in the optimization (open circles). (The other 25% of residues were not detected). Use of the two sets of coefficients (B) clearly improves the correlation of  $\ln(P^{sim})$  compared with use of a single pair of coefficients (A); see Table SII-3.

## II.9.2 Supporting methods

### II.9.2.1 $^{15}\text{N}\{^1\text{H}\}$ NOE

$^{15}\text{N}\{^1\text{H}\}$  steady-state NOE relaxation was measured with a Bruker Avance III 800 MHz NMR spectrometer using a pulse sequence with TROSY-enhanced  $^1\text{H}$  resolution and sensitivity (4). In order to measure the protein under identical conditions, saturated and unsaturated  $^{15}\text{N}\{^1\text{H}\}$  NOE spectra were collected in an interleaved manner. Each of the triplicate measurements of pairs of spectra of Apo-P was performed on freshly phosphorylated enzyme within 14 h of the phosphorylation. The triplicate measurements of spectra of Apo-deP were performed on freshly dephosphorylated enzyme in 17 h each. All samples were protected from oxidation with 10 mM DTT, Ar, and sealing with Shigemi tube plunger and parafilm.

### II.9.2.2 Correction of HDX decays for lingering solvent

Since the HDX series were initiated by 5-fold dilution with  $\text{D}_2\text{O}$ -based buffer, 20% of solvent was still  $^1\text{H}_2\text{O}$ . The initial mixture decays as:

$A(t) = A_0 \exp(-k_{obs} \times t) + A_\infty$ , where the eventual peak size  $A_\infty = 0.2 \cdot A_0$  reflects the original  $\text{H}_2\text{O}$  still present at 20%. The ~20 min dead time for sample handling,  $t_d$ , delayed acquisition of the first BEST-TROSY spectrum. The exponential decay following the dead time was fitted as:

$$A(t') = A'_0 \exp(-k_{obs} \times t') + A'_\infty \quad \text{Equation II-S1}$$

where  $A'_0 = A(t_d) = A_0 \exp(-k_{obs} \times t_d) + A_\infty$  and  $A'_\infty$  required fitting for each decaying peak because:

$$A'_\infty = \frac{A_\infty}{A_0 \exp(-k_{obs} \times t_d) + A_\infty} > 0.2 \times A'_0$$

$A'_\infty$  exceeding 20% of  $A'_0$  is evident in Fig. II-2D.

### II.9.2.3 Rate of dephosphorylation and correction of affected rates of HDX

In order to monitor the longevity of the phosphorylated form PMM/PGM and quantify the loss of its phosphorylation as the rate constant  $k_{deP}$ , control NMR experiments were set up in H<sub>2</sub>O and acquired very similarly to those in D<sub>2</sub>O. Decay rates in H<sub>2</sub>O solvent were obtained by fitting a single exponential decay rate  $k_{deP}$  to the peak heights (Figs. II-2D and II-S2C). Fitting peak heights and volumes gave equivalent results using Eq. II-1. However, 12 peaks of Apo-P and 10 peaks of Apo-deP underwent small, scattered shifts of the peak centers that introduced modest scatter to the decays; this was rectified using manual picking and volume integration of these peaks. Uncertainties were estimated from spectral noise and fitting errors. Global fitting of the decays of 43 peaks characteristic of Apo-P together with the concurrent increases of the peaks of Apo-deP of the same 43 residues, measured in spectra in H<sub>2</sub>O (Fig. SII-2, C and D), provided the most reliable estimate of  $k_{deP}$ .

The rate constants of HDX ( $k_{ex}$ ) of Apo-P form were corrected for its dephosphorylation by subtracting apparent  $k_{deP}$  from the observed rate constant of overall decay after dilution into D<sub>2</sub>O:  $k_{ex} = k_{obs} - k_{deP}$  (Fig. SII-8). For 98 Apo-P peaks without overlap with their counterparts of Apo-deP, each peak was corrected by its apparent  $k_{deP}$  values fitted independently to the decay of the corresponding peak measured in H<sub>2</sub>O spectra. Most of the individually measured decays from dephosphorylation conform to the shared time constant of 19 h for dephosphorylation (Fig. SII-2, C and D). However, 14 peaks were subject to an additional faster decay process in H<sub>2</sub>O with a time constant of ~4.5 h which we fitted independently to generate pseudo  $k_{deP}$  values for use in

correcting  $k_{ex}$ . Consequently, the decay rates of each individual peak in H<sub>2</sub>O spectra were used in correcting the  $k_{obs}$  values in D<sub>2</sub>O to the  $k_{ex}$  of HDX for the 98 amide peaks of Apo-P in the first two categories listed in Table SII-1. The globally fitted best estimate of  $k_{deP}$  was used to correct the  $k_{ex}$  of 44 peaks characteristic of Apo-P but overlapped with the corresponding peaks of Apo-deP, which are listed in the third and fourth categories in Table SII-1.

The  $k_{ex}$  rate constants of some well-protected residues appear to be nearly zero (Fig. SII-8A and SII-6, open squares). Therefore, upper bounds on  $k_{ex}$  were estimated in these most-protected cases. The smallest  $k_{obs}$  fitted to HDX spectra was  $5.56 \times 10^{-5} \text{ min}^{-1}$ , suggesting that the slowest  $k_{ex}$  in the enzyme should be less than  $5.56 \times 10^{-5} \text{ min}^{-1}$ .

In the HDX experiments, 13 residues with overlapped Apo-P and Apo-deP peaks in the TROSY spectra displayed  $k_{ex}$  unaffected by phosphorylation. *No correction* for rate of dephosphorylation was needed for these (fifth category of Table SII-1), as justified in the following section deriving behaviors possible with overlapped peaks.

#### II.9.2.4 Fitting of HDX decays of Apo-P peaks where overlapped Apo-deP peaks could develop during the time course

It can be demonstrated that the HDX decays of the overlapped NMR peaks of the phosphorylated and dephosphorylated enzyme forms can be described by this expression encompassing a wide range of scenarios:

$$Pe^{-(k_{ex,P}+k_{deP})t} + Qe^{-k_{ex,deP}t} \quad \text{Equation II-S2}$$

$$\text{where } P = \frac{a(k_{ex,P}-k_{ex,deP})}{k_{ex,P}-k_{ex,deP}+k_{deP}} \text{ and } Q = \frac{(1-a)(k_{ex,P}-k_{ex,deP})+k_{deP}}{k_{ex,P}-k_{ex,deP}+k_{deP}},$$

where  $a$  is the fraction of the enzyme in the phosphorylated state Apo-P,  $k_{ex,P}$  is the rate constant of H(D)X of Apo-P,  $k_{ex,deP}$  the rate constant of H(D)X of Apo-deP, and  $k_{deP}$  the

rate constant for the dephosphorylation.  $a = 0.9$  in the experiments reported herein, minimizing the contribution of the second term. We denote the important comparison  $k_{ex,P} - k_{ex,deP}$  as  $\Delta k_{ex}$ . Table SII-1 summarizes the proper choices of data fitting deduced from use of Eq. II-S2 in different settings.

Eq. II-S2 indicates that when the  $\Delta k_{ex}$  difference in the HDX rates of the two phosphorylation states is small, then the first term of Eq. II-S2 becomes small. When  $\left|\frac{P}{Q}\right| < 0.2$ , the term  $Pe^{-(k_{ex,P}+k_{deP})t}$  becomes negligible compared to  $Qe^{-k_{ex,deP}t}$ .

Therefore, the kinetic equation for this case of overlapped peaks simplifies to the single exponential decay  $Qe^{-k_{ex,deP}t}$  needing no correction because  $k_{ex,deP}$  was fitted and  $k_{ex,deP} \approx k_{ex,P}$  (Table SII-1). This situation was recognized at residues where  $k_{obs}$  of Apo-P is similar to  $k_{ex,deP}$  of Apo-deP.

In cases where phosphorylation causes  $k_{ex,P}$  and  $k_{ex,deP}$  to differ enough that  $|\Delta k_{ex}|$  is enlarged compared to the dephosphorylation rate  $k_{deP}$ , the first term of Eq. II-S2 becomes large compared to the second term. When  $\left|\frac{Q}{P}\right| < 0.2$ ,  $Qe^{-k_{ex,deP}t}$  becomes negligible compared to  $Pe^{-(k_{ex,P}+k_{deP})t}$ . In such cases, a single exponential decay with rate constants  $k_{ex,P} + k_{deP}$  was observed. This was corrected by subtracting away  $k_{deP}$ .

It can be shown that in the bi-exponential regime (where  $0.2 < \left|\frac{P}{Q}\right| < 5$ ), inappropriately fitting of a single-exponential decay to the bi-exponential decay introduces tolerably little error to resulting estimates of  $\Delta G_{HX}$ .

#### II.9.2.5 Chemical Denaturation

The chemical denaturation of Apo-P and Apo-deP was performed using urea as described (5) but with modifications. 1.3  $\mu$ M enzyme stock solutions of proteins in 50

mM MOPS, pH 7.4, and 0.1 mM DTT were mixed with this same buffer containing 10 M urea to obtain final urea concentrations ranging up to 8 M. The mixtures were incubated for 12 h. (Longer incubation times give similar denaturation curves.) Trp fluorescence emission was excited at 280 nm and detected at 320 nm using a BioTek Synergy MX plate reader.

#### **II.9.2.6 Molecular dynamics simulations**

The crystal structures of the Apo-P and Apo-deP forms of PMM/PGM (PDB codes 1K35 and 4MRQ, respectively) provided the initial structures. All crystallographic waters were included in the simulations. Hydrogen and missing side chain atoms were added using xleap (6). The Amber ff99SB force field was applied to all residues except the metal binding site (7). The region around the metal-binding site was parameterized as described for the electrostatic calculations. The protonation of the enzyme was set by neutral pH with the pSer108 deprotonated, consistent with the catalytic mechanism (8). The simulations were carried out using GROMACS 4.5.5 (9, 10) on the high performance computing facility at the University of Missouri Bioinformatics Consortium. The enzyme coordinates were solvated in a 10 Å truncated octahedral box containing about 13,000 TIP3P waters and neutralized with Na<sup>+</sup> ions. Moderate Langevin dynamics with a collision frequency of  $\gamma = 1 \text{ ps}^{-1}$  were applied. The cutoff for non-bonded interactions was 10 Å. The SHAKE algorithm was applied (11). After initial minimization, stepwise heating to 10, 50, 100, 150, 200, 250, and 310 K (12) was performed. The system was equilibrated under constant volume for 100 ps and then under constant pressure for another 100 ps. The 60-ns production simulations used periodical boundary condition in

an NPT ensemble. The temperature was maintained at 310 K and pressure 1 atm using Berendsen thermostat and Parrinello-Rahman barostat algorithms (13, 14), respectively.

### II.9.2.7 HX protection simulated from MD trajectories

Previous efforts to simulate HX protection, i.e.  $\ln(P_n^{sim})$ , from simulations and structures assumed that the contributions of contacts and hydrogen bonding or  $\text{RMSF}^{-1}$  to the protection are identical for all residues (15-17). However, as each residue could have a unique environment, the contributions may vary. Therefore, we clustered all residues into two groups, those with more or less protection from measured hydrogen exchange. We optimized independent sets of  $\beta_c$  and  $\beta_H$  or  $\beta_{1/\text{RMSF}}$  for these two clusters of residues, using the same type of optimization of the difference between  $\ln(P_n^{sim})$  from unrestrained MD simulations and experimentally measured  $\ln(P_n^{exp})$  as described by Kieseritzky et al (17), i.e. minimizing this RMS deviation:

$$\Delta \ln P = \sqrt{\frac{1}{N} \sum_{n=1}^N (\ln(P_n^{sim}) - \ln(P_n^{exp}))^2} \quad \text{Equation II-S3}$$

Two alternative sets of  $\beta_c$  and  $\beta_H$  for quickly or slowly exchanging sites turned out to better simulate the protections and minimize the overall  $\Delta \ln P$  than did a single set of coefficients, for a set of well-characterized model proteins and their structures (Table SII-3) and for the PMM/PGM enzyme and its MD trajectories (Table SII-4).

In the tallying of structural interactions for the optimization process, contacts  $N_n^c$  between two residues were defined when the distances between amide nitrogen and heavy atoms are closer than 6.5 Å (16). Hydrogen bonds  $N_n^h$  were considered as formed for H-O distances shorter than 2.4 Å and N-H-O angles greater than 140°. The errors were calculated as the standard deviations of the time-averaged protection factor

elements (PFEs, e.g.  $N_n^c$  and  $N_n^h$ ) counted from the trajectories. Since well-protected amide groups should be observed by NMR-detected HDX,  $\beta_c^{\text{less prot}}$  and  $\beta_H^{\text{less prot}}$  or  $\beta_{1/\text{RMSF}}^{\text{less prot}}$  were used to calculate  $\ln(P_n^{\text{sim}})$  of residues that were not detected by HDX or CLEANEX-PM in plotting Fig. II-6C. The good correlations between  $\ln(P_n^{\text{sim}})$  and  $\ln(P_n^{\text{exp}})$  (Tables II-S3 and S4, Fig.SII-11 and 6B) suggest that  $\ln(P_n^{\text{sim}})$  could be used as a proxy for  $\ln(P_n^{\text{exp}})$  for the numerous amide groups of PMM/PGM in the unmeasured time regime of HX slower than 12 s and faster than 20 min; compare Fig. II-3A and II-6C.

### II.9.3 Supporting tables

Table SII-1. Behaviors of amide TROSY NMR peaks of Apo-P monitored by HDX and the appropriate data fitting method implemented.

Category	Count	Data Fitting
Total peaks of Apo-P tracked by HDX	155 (68%) *	
Residues with well-resolved peaks for Apo-P and Apo-deP	87 (38.2%)	Single exponential $k_{obs} = k_{ex,P} + k_{deP}$ †
Only Apo-P peaks present as Apo-deP peaks were missing	11 (4.8%)	Single exponential $k_{obs} = k_{ex,P} + k_{deP}$ †
<b>Overlapped peaks</b>	57 (36.7%)	
Apo-deP peak decayed too quickly to be seen by HDX-NMR. ‡	31 (13.6%)	Single exponential § $k_{obs} = k_{ex,P} + k_{deP}$
Phosphorylation changes $k_{ex}$ significantly: $\Delta k_{ex} < -3.6k_{deP}$ or $\Delta k_{ex} > 12.5k_{deP}$ , $\left \frac{Q}{P}\right  < 0.2$ ¶	13 (5.7%)	Single exponential $k_{obs} = k_{ex,P} + k_{deP}$ §
Little effect of phosphorylation: $ \Delta k_{ex}  < 0.22k_{deP}$ , $\left \frac{P}{Q}\right  < 0.2$	13 (5.7%)	Single exponential $k_{obs} = k_{ex,P}$ $= k_{ex,deP}$
$0.2 < \left \frac{P}{Q}\right  < 5$	0 (0%)	Bi-exponential See Eq. II-S2

\* % of the total amide peaks monitored are given in parentheses

† Apparent  $k_{deP}$  values for each of these individual residues (individually fitted to the decay of each NMR peak in H<sub>2</sub>O) were used to accommodate the anomalously fast decay of 14 of these 98 peaks (a group with time constant of  $4.5 \pm 0.1$  h and “ $k_{deP}$ ” of  $\sim 3.7 \times 10^{-3} \text{ min}^{-1}$ ).

‡  $k_{ex,deP}$  was rapid, apparently  $> 0.05 \text{ min}^{-1}$

§ The globally fitted value of  $k_{deP}$  was used to correct the  $k_{ex}$  values of Apo-P.

¶  $\Delta k_{ex} = k_{ex,P} - k_{ex,deP}$ . P and Q are defined by Eq. II-S2.

Table SII-2. NMR identification of residues with HX altered by Ser108 phosphorylation, clarifying behaviors within peptides obscured in HDX-MS results

	<b>Peptide residues *</b>	<b>Stabilized by phosphorylation</b>	<b>Destabilized by phosphorylation</b>	<b>Little change<sup>†</sup></b>
<b>Partly overlapping active site</b>	92-117 <sup>‡</sup>	92, 95,106,115,116	Not detected	93,97,98, 101,117
	205-239	223,239	206,219,230,236	221,225,228,229, 231-233
	240-266	244,247,263-266	242	245,252,253,255, 258,259,261
	267-293	268,269,271-273, 275,278,280-282,291	Not detected	270,279,289,292
	323-346	323,324,333,335,336	343, 345	331, 346
<b>Far from active site</b>				
	28-44	28,30,36,38,44	33,37,41	29,32,42,43
	45-52	Not detected	45,50	46,48
	53-78	59,63,64,66,72,77	58	55,65,75,78
	169-191	171,173,174,176, 178,180,183,185	184, 188	169,179,186,191
	192-204	195,197,198,	194,202,203	193,196,200
	378-406	378,381,384, 385,389, 392,404	Not detected	379,380,382,383,387, 391,393,397,398,399,403
	407-429	407,409,410,413, 416-422	Not detected	411,412,415,428,429
442-463	442,446,448,449,455	Not detected	444,445,447, 451-454,457	

\* Peptides were generated by pepsinolysis. Percentage increase in deuteration of peptide as a whole was measured by mass spectrometry (18).

† Insufficient evidence of a difference greater than 1 kcal/mol

‡ The phosphopeptide evaded detection by HDX-MS.

Table SII-3. Comparison of one and two sets of coefficients in modeling HX protection factors from surrounding contacts and hydrogen bonds for well-characterized proteins

Protein	One pair of coefficients				Two pairs of coefficients					
	$\beta_C^*$	$\beta_H^\dagger$	$\Delta \ln P^*$	Pearson's correlation coefficient	less protected residues		more protected residues		$\Delta \ln P^\ddagger$	Pearson's correlation coefficient
					$\beta_C^*$	$\beta_H^\dagger$	$\beta_C^*$	$\beta_H^\dagger$		
Ubiquitin <sup>§</sup>	0.35 <sup>¶</sup>	2.00 <sup>¶</sup>	3.73	0.75						
	0.58 <sup>l</sup>	2.30 <sup>l</sup>	2.33	0.75	0.48	1.06	0.75	1.39	1.72	0.88
Staphylococcal nuclease**	0.35 <sup>¶</sup>	2.00 <sup>¶</sup>	2.07	0.64						
	0.27 <sup>l</sup>	1.80 <sup>l</sup>	1.71	0.64	0.13	0.79	0.35	1.26	1.19	0.85
bacterial cytochrome c <sup>††</sup>	0.35 <sup>¶</sup>	2.00 <sup>¶</sup>	3.04	0.58						
	0.60 <sup>§§</sup>	1.59 <sup>§§</sup>	1.86 <sup>§§</sup>	0.62 <sup>§§</sup>	0.51	0.36	0.65	0.28	1.57	0.77

\* Coefficient for the number of contacts surrounding each residue (15)

† Coefficient for the number of hydrogen bonds to each backbone amide group (15)

‡ Difference in natural logarithm of simulated and experimental protection factor P (15); smaller indicates better agreement.

¶ Optimized in ref (16)

<sup>l</sup> Parameters optimized for measured HX protection factors following the method of ref (17).

<sup>§§</sup> From ref (17)

<sup>§</sup> Protection factors were calculated from rates of HX of ref (19). PDB: 1UBI

\*\* HX protection factors from ref (3). PDB: 1SNO. Hydrogen atoms were added to the coordinates using Sybyl 7.3.

<sup>††</sup> HX protection factors are digitized from Fig. 2 of ref (20), PDB: 1K3G (first conformer).

Table SII-4. Comparison of HX protection factors measured and simulated from MD trajectories,  $\Delta \ln P$ , of PMM/PGM, using one or two pairs of coefficients for protection factor elements (PFE)

Phosphorylation state	PFE1, no. of contacts	PFE2	less protected residues		more protected residues		$\Delta \ln P$	Pearson's correlation coefficient
			$\beta_c$	$\beta_2$	$\beta_c$	$\beta_2$		
<b>Apo-P</b>	$N_n^c$	$N_n^h$	0.35	2.0*	0.35	2.0*	6.7	0.34
<b>Apo-deP</b>							5.5	0.36
<b>Apo-P</b>	$N_n^c$	$N_n^h$	0.22	0.68	0.80	2.50	3.2	0.82
<b>Apo-deP</b>							2.6	0.89
<b>Apo-P</b>	$N_n^c$	RMSF <sup>-1</sup>	0.22	0.28 <sup>†</sup>	0.56	3.94 <sup>†</sup>	3.1	0.83
<b>Apo-deP</b>							2.7	0.88

\* The single pair of coefficients optimized by ref (16) was applied and is listed twice.

<sup>†</sup> These two pairs of coefficients were used to generate Fig. II-6, B and C.

#### II.9.4 Supporting movies

Movie II-S1. Time course of hydrogen-deuterium exchange of Apo-deP. Yellow is >90% exchanged. Orange is 10–90% exchanged. Red is <10% exchanged.

Movie II-S2. Time course of hydrogen-deuterium exchange of Apo-P. Yellow is >90% exchanged. Orange is 10–90% exchanged. Red is <10% exchanged.

#### II.9.5 Supporting references

1. Hwang, T. L., P. C. van Zijl, and S. Mori. 1998. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *Journal of Biomolecular NMR* 11:221-226.
2. Huyghues-Despointes, B. M., J. M. Scholtz, and C. N. Pace. 1999. Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat. Struct. Biol.* 6:910-912.
3. Skinner, J. J., W. K. Lim, S. Bedard, B. E. Black, and S. W. Englander. 2012. Protein hydrogen exchange: testing current models. *Protein Sci.* 21:987-995.
4. Zhu, G., Y. Xia, L. K. Nicholson, and K. H. Sze. 2000. Protein Dynamics Measurements by TROSY-Based NMR Experiments. *J. Magn. Reson.* 143:423-426.

5. Fulcher, Y. G., and S. R. Van Doren. 2011. Remote exosites of the catalytic domain of matrix metalloproteinase-12 enhance elastin degradation. *Biochemistry* 50:9488-9499.
6. Case, D. A., T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman. 2012. AMBER 12. University of California, San Francisco.
7. Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* 65:712-725.
8. Naught, L. E., and P. A. Tipton. 2001. Kinetic Mechanism and pH Dependence of the Kinetic Parameters of *Pseudomonas aeruginosa* Phosphomannomutase/Phosphoglucomutase. *Archives of Biochemistry and Biophysics* 396:111-118.
9. Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43-56.
10. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 7:306-317.
11. Ryckaert, J.-P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* 23:327-341.
12. Xu, J., G. Yin, and W. Du. 2011. Distal mutation modulates the heme sliding in mouse neuroglobin investigated by molecular dynamics simulation. *Proteins* 79:191-202.
13. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684-3690.
14. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52:7182-7190.
15. Vendruscolo, M., E. Paci, C. M. Dobson, and M. Karplus. 2003. Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. *J. Am. Chem. Soc.* 125:15686-15687.
16. Best, R. B., and M. Vendruscolo. 2006. Structural Interpretation of Hydrogen Exchange Protection Factors in Proteins: Characterization of the Native State Fluctuations of CI2. *Structure* 14:97-106.
17. Kieseritzky, G., G. Morra, and E.-W. Knapp. 2006. Stability and fluctuations of amide hydrogen bonds in a bacterial cytochrome c: a molecular dynamics study. *J. Biol. Inorg. Chem.* 11:26-40.

18. Lee, Y., M. T. Villar, A. Artigues, and L. J. Beamer. 2014. Promotion of Enzyme Flexibility by Dephosphorylation and Coupling to the Catalytic Mechanism of a Phosphohexomutase. *J. Biol. Chem.* 289:4674-4682.
19. Bougault, C., L. Feng, J. Glushka, E. Kupče, and J. H. Prestegard. 2004. Quantitation of rapid proton-deuteron amide exchange using hadamard spectroscopy. *J. Biomol. NMR* 28:385-390.
20. Bartalesi, I., A. Rosato, and W. Zhang. 2003. Hydrogen exchange in a bacterial cytochrome c: a fingerprint of the cytochrome c fold. *Biochemistry* 42:10923-10930.

## **CHAPTER III**

### **Multiple Ligand-Bound States of a Phosphohexomutase Based On Principal Component Analysis of NMR Peak Shifts**

#### **III.1 Abstract**

Enzymes sample an ensemble of conformers in solution that recognize substrates with specificity. NMR is uniquely suited to characterize ensembles in solution since chemical shifts are hypersensitive to conformational changes and are easy to quantify accurately. Phosphomannomutase / phosphoglucosmutase (PMM/PGM) is a 52 kDa multi-domain enzyme that catalyzes phosphoryl transfer across its phosphomannose and phosphoglucose substrates. The equilibrium shifts of PMM/PGM during its catalytic cycle are still not fully understood. In this work, responses of PMM/PGM and an inactive mutant to various phosphosugars were characterized by NMR-detected titrations. Binding affinities of substrates were estimated by principal component 1-based binding isotherms. Shared over-arching relationships among phosphosugar ligand association reactions were assessed by applying Principal Component Analysis (PCA) on unfiltered peak lists of all titrations. COordinated Chemical Shifts bEhavior (CONCISE) analysis quantitatively identified five different states along a continuum suggested by colinear shifts of NMR peaks. Synergistic effects of enzyme phosphorylation and phosphosugar binding spear to drive the equilibrium from the open, dephosphorylated state to the free phosphorylated state, and on to closed states bound to a monophosphosugar, bisphosphosugar, or monophospho analogue inhibitor. NMR-detected hydrogen exchange reveals that binding of this inhibitor stabilizes domains 3 and 4 significantly, establishing a globally unified structure with enhanced folding stability.

## III.2 Introduction

Conformational change and substrate binding play central roles in enzyme catalysis, where coexistence of multiple substates, including structural conformations, protonation states, or tautomers, is ubiquitous (1). However, many studies failed to observe conformational change upon ligand binding, presumably due to limitations of techniques such as crystal packing effects (2). NMR chemical shifts are extremely useful to study protein-ligand interactions, because they can be measured accurately, and are sensitive to subtle structural or dynamics changes (3). Chemical shifts observed by solution NMR can provide detailed information on interchanging enzyme conformations during the catalytic cycle in native-like conditions, and can be used as a powerful tool to study catalytic mechanisms at the atomic level.

Enzymes in the  $\alpha$ -D-phosphohexomutase superfamily are ubiquitous and important in carbohydrate biosynthesis. Among the proteins in this superfamily, phosphomannomutase / phosphoglucomutase (PMM/PGM) contributes to the virulence of the opportunistic human pathogen *Pseudomonas aeruginosa*, which produces life-threatening infections in patients with cystic fibrosis or burn injuries. PMM/PGM participates in biosynthesis of virulence factors such as lipopolysaccharide, rhamnolipids, Pel and Psl polysaccharides, and alginate (4-10). In solution, bacterial PMM/PGM is a 463-residue monomer comprised of four domains. It catalyzes a reversible phosphoryl group transfer between glucose- or mannose-based substrates, depending on the biosynthetic pathway involved. The enzyme requires  $Mg^{2+}$  and a highly conserved phosphoserine-108 for activity (11). Its reaction mechanism entails two successive phosphoryl transfer steps (Fig. III-1). First, the phosphoryl group of phosphoSer108

(pSer108) is transferred to the monophosphorylated substrate, forming a bisphosphorylated intermediate which reorients by 180° within the active site. Then the phosphoryl group is transferred back to Ser108 from the intermediate, generating the phosphorylated enzyme.

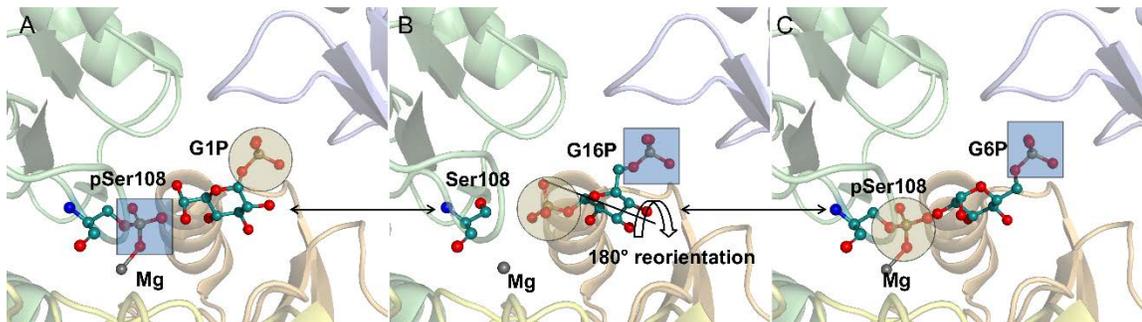


Figure III-1. Reactions of PMM/PGM. The enzyme reversibly transforms G1P (A, PDB ID: 1P5D) to G6P (C, PDB ID: 1P5G) via an intermediate G16P, which undergoes 180° reorientation (B, PDB ID: 2FKM). Domain 1 to domain 4 are colored as pale green, pale yellow, light orange, and light blue, respectively. Note this color code also applies to all subsequent figures with crystal structures, including Figs. III-2, III-4, III-6, III-7, and Figs. SIII-4, SIII-7, SIII-9.

The effects of phosphorylation on PMM/PGM were previously studied by various biophysical approaches, including hydrogen exchange behaviors monitored by mass spectrometry (HDX-MS) and NMR (HDX-NMR), which suggest that phosphorylation of Ser-108 in the catalytic cleft widely stabilizes PMM/PGM while dephosphorylation tends to increase structural flexibility for the most part (12, 13). Ligand binding of PMM/PGM has also been previously studied, both structurally and enzymatically. Structural studies revealed domain 4 of PMM/PGM rotates to close the catalytic cleft upon binding of phosphosugar substrates (14, 15). Among these ligands, xylose-1-phosphate (X1P), a substrate analog, is known to shift the conformational ensemble of PMM/PGM towards a closed and less flexible ligand-bound state (12).

Comparison of crystallographic structures showed ligand induced changes vary between open, half-open, and closed conformations (16). It failed to capture the global

changes due to phosphorylation of catalytic residue Ser108 (12), which were revealed by hydrogen exchange detected either by NMR or mass spectrometry (13). The hydrogen exchange results raise the question of whether ligand-induced conformational changes of PMM/PGM are oversimplified by crystallography.

In protein titrations with ligands, NMR can be used to measure progressive shifting of peaks in the fast-exchange regime (17). Chemical shifts in the fast exchange regime reflect population-weighted averages of the spectral contributions of the states in rapid equilibrium (3). Principal component analysis (PCA) applied to chemical shift changes has been used to gain insight from the shifts in population (18). PCA is a widely used tool of unsupervised statistics and chemometrics that can reduce dimensionality, extract patterns, and perform cluster analysis on large data sets (19, 20). PCA was previously used on amide  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts to monitor the dependence of protein reactions on time, pH, or [ligand] (21-23). PCA can reduce noise in spectra to improve binding isotherm fitting (24). Chemical shift covariance analysis (CHESCA), which is a combination of PCA and dendrogram analyses, is a well-established tool for mapping allosteric networks by analyzing chemical shifts. CHESCA employs PCA to cluster residues and/or spectra that mainly contribute to either binding or allosteric equilibria (25, 26). Another chemical shift based statistical analysis approach called COordinated Chemical Shifts bEhavior (CONCISE), which evaluates the collective response of proteins to perturbations, such as ligand binding, mutation, or post-translational modification is applied to the peak picks of all NMR peaks (27). This has the effect of filtering out minor changes that do not reflect conformational equilibrium

shift. The ability of PCA to extract main patterns and remove noise makes it a powerful approach to decode complicate systems.

In this work, a small library of perturbations, including binding of phosphosugar species involved in the catalytic cycle and inhibitor (Table III-1, Fig. III-4), was constructed to test ligand binding effects on PMM/PGM using titrations. A variety of PCA methods were carried out on the large set of  $^1\text{H}$ ,  $^{15}\text{N}$  chemical shift data, suggesting a progression among global states of conformational equilibria of PMM/PGM in solution.

Table III-1. Perturbation library

<b>Enzyme</b>	<b>Perturbation</b>	<b>Enzyme</b>	<b>Perturbation</b>
wildtype	G1P	wildtype	dephosphorylation
wildtype	G16P	S108C	G1P
wildtype	G6P	S108C	G16P
wildtype	M6P	S108C	G6P
wildtype	X1P	S108C	X1P

\*Structures of sugar substrates are shown in Fig. III-4

### III.3 Material and Methods

#### III.3.1 Enzyme preparation, isotopic labeling, and purification.

$^2\text{H}$ ,  $^{15}\text{N}$  labeled His-tagged PMM/PGM and its S108C mutants were expressed and purified according to the protocol described in ref (28) and ref (12). The phosphorylation states of  $^2\text{H}$ ,  $^{15}\text{N}$  labeled PMM/PGM samples used for G1P, G6P, G16P, M6P, and X1P titrations were estimated to be  $53 \pm 9\%$ ,  $75 \pm 8\%$ ,  $77 \pm 9\%$ ,  $77 \pm 10\%$ , and  $66 \pm 12\%$ , respectively, using well-resolved peaks of the phosphorylated ( $E_p$ ) and dephosphorylated ( $E_{dep}$ ) forms of the enzyme as described (13).  $\alpha$ -D-glucose 1-phosphate, glucose 1,6-bisphosphate, glucose 6-phosphate, and mannose 6-phosphate

were purchased from Sigma-Aldrich (St Louis, MO, USA). Xylose-1-phosphate was synthesized by Professor Thomas Mawhinney (University of Missouri).

### III.3.2 NMR spectroscopy

Isotope-labeled samples for NMR acquisition were prepared with a concentration of 0.5 to 1.0 mM (determined by Bradford assay) in 50 mM MOPS (pH 7.4), 1 mM MgCl<sub>2</sub> with 5% D<sub>2</sub>O (v/v). Sugars were titrated into <sup>2</sup>H, <sup>15</sup>N PMM/PGM at 310 K on a Bruker Avance III 800 MHz spectrometer with TCI cryoprobe. A series of 2D <sup>15</sup>N BEST-TROSY spectra were collected with sugar-phosphate additions up to 8-fold excess.

### III.3.3 NMR spectra interpretation

The spectra were processed with NMRPipe (29) and analyzed using Sparky (30). The peaks were interpreted with the chemical shift assignments reported (28). Chemical shift perturbations (CSPs) of amide peaks were calculated as the radius:

$$\Delta\delta_{NH} = \sqrt{\left(\left(\frac{\Delta\delta_N}{5}\right)^2 + \Delta\delta_H^2\right)} \quad \text{Equation III-1}$$

where  $\delta N$  and  $\delta H$  are the changes in <sup>15</sup>N and <sup>1</sup>H dimensions in the unit of ppm. The <sup>15</sup>N frequency changes were scaled down by a factor of 5 to normalize them to the <sup>1</sup>H scale. All TROSY spectra were referenced to the free (“apo”) state by minimizing the root-mean-square deviation of 60 peaks with smallest chemical shift perturbations, by different means but for the same need as ref (27) .

$$RMSD = \sqrt{\sum_{i=1}^{N_{res}} \left( \left( \frac{\delta_{Ni} - \delta_{Ni}^{ref}}{5} \right)^2 + (\delta_{Hi} - \delta_{Hi}^{ref})^2 \right)} \quad \text{Equation III-2}$$

### III.3.4 Measurement of apparent affinity

Binding isotherms of all titrations were calculated by applying principal component analysis to lists of the peaks picked from each of the spectra in that titration (23). This was implemented with the program called TREND (18). Each binding isotherm was fit to Eq. III-3

$$p_{free} = \frac{\Delta\delta_{NH,obs}}{\Delta\delta_{NH,max}} = \frac{([P]_t + [L]_t + K_D) - \sqrt{([P]_t + [L]_t + K_D)^2 - 4[P]_t[L]_t}}{2[P]_t} \quad \text{Equation III-3}$$

where  $p_{free}$  is the fraction of free protein (of total concentration  $[P]_t$ ) that is not bound to ligand.  $[L]_t$  is the total ligand concentration.  $\Delta\delta_{NH,obs}$  is the observed protein chemical shift.

### III.3.5 Vectors for shifts of NMR peaks

A CSP vector tracks the shift of a peak from its position in the free state to its position after near-saturating addition of ligand; it is calculated as the difference between their coordinates:  $(x_{end} - x_{ini}, y_{end} - y_{ini})$ . Examples of calculating CSP vectors for wt PMM/PGM and its S108C mutant are shown in Fig. SIII-1. All shifts of NMR peaks from the peak lists from the titrations were converted into CSP vectors. This constructed a new set of lists based on the CSP vectors instead of the original peak positions. The CSP vector of each residue in a titration was referenced to the free (apo) state at (0, 0).

### III.3.6 PCA clustering of spectra perturbed by ligand, dephosphorylation, or mutation

PCA analysis using TREND (18) was performed on CSP vectors for each set of ligand-induced spectral changes listed in Table III-1. Lists of CSP vectors for each peak were converted into a large 2D matrix according to procedure described in ref (18, 23).

Each column represents a list of CSP vectors, while the  $^1\text{H}$  and  $^{15}\text{N}$  coordinates of each vector are arranged in an interleaved manner. The columns of the input matrix were centered and autoscaled (25). Clustering of perturbations were visualized by 2D and 3D biplots.

### III.3.7 CONCISE analysis

COordiNate ChemIcal Shift bEhavior (CONCISE) was used to monitor linear trajectories of CSP vectors (27) and measure the equilibrium position of each ligand-bound (and free, dephosphorylated “E<sub>dep</sub>”) state. CONCISE analysis was carried out on CSP vectors for all ligand-bound states of wt PMM/PGM and S108C mutant according to (27). The direction of largest covariance of each residue was identified by principal component analysis (PCA) independently. Standard deviations (SD) of the first and second principal components (PC1 and PC2) were used to calculate linearity of each residue along all states. Residues with poor linearity ( $\text{SD}_{\text{PC1}}/\text{SD}_{\text{PC2}} < 3.0$ ) and/or with small perturbations ( $\text{PC1} < 0.05$  ppm) were discarded from the analysis to filter out resonances affected by nearest neighbors and reduce systematic error (27).

### III.3.8 Hydrogen exchange analysis

The rates of rapid hydrogen exchange (HX) of amide groups were detected by an interleaved version of a TROSY-detected CLEANEX-PM NMR pulse sequence (13, 31). The hydrogen exchange rates on the rapid time scale were obtained by fitting to the equation  $\frac{I}{I_0} = \frac{k_{ex}}{R_{1A} + k_{ex} - R_{1B}} \times [\exp(-R_{1B}\tau_m) - \exp[-(R_{1A} + k_{ex})\tau_m]]$ , where  $I/I_0$  stands for normalized peak heights,  $R_{1A}$  is a combination of transverse and longitudinal relaxation,  $k$  is the amide HX rate constant, and  $R_{1B}$  is the water relaxation constant

during the mixing time. Since  $R_{IB}$  values ranging from 0.1 to 0.001 s<sup>-1</sup> give little difference in the fitted  $k_{ex}$  values, 0.01 s<sup>-1</sup> was used (32).

HX rates of slow exchanging amides were characterized according to the previously described protocols (13). H<sub>2</sub>O-based buffer dissolved <sup>2</sup>H/<sup>15</sup>N samples were concentrated to ~5 mM and the diluted to ~1 mM in D<sub>2</sub>O-based buffer at room temperature. Exponential decay rate constants in D<sub>2</sub>O ( $k_{obs}$ ) were obtained by fitting peak heights to a single exponential function. Due to turbidity developing over time, hydrogen-deuterium exchange rates ( $k_{ex}$ ) were obtained by subtracting decay rates measured in a control NMR experiments using H<sub>2</sub>O instead of D<sub>2</sub>O ( $k_{control}$ ):  $k_{ex} = k_{obs} - k_{control}$ . The details were described in ref (13).

Experimental protection factors of amide hydrogens were calculated as free energy differences between the measured folded state and the random coil reference state:

$$\Delta G = -RT \ln \left( \frac{k_{ex}}{k_{rc}} \right) \quad \text{Equation III-4}$$

Estimates of intrinsic rate constants of random coil HX ( $k_{rc}$ ) were obtained using the SPHERE server (<http://www.fccc.edu/research/labs/roder/sphere/sphere.html>).

## III.4 Results

### III.4.1 Ligands affect binding site and a domain-domain interface

NMR-detected titrations were performed with five sugar-phosphate ligands on wt and S108C-inactivated PMM/PGM from *P. aeruginosa* (Table III-1). Figs. SIII-2 to 4 show the <sup>15</sup>N TROSY NMR spectra at 800 MHz of free and ligand-bound states of wt and S108C-substituted PMM/PGM. The amide chemical shifts were updated with the aid of the incremental peak shifts from ligand additions, typically in fast exchange, in the

TROSY spectra of the titrations. The peak shifts introduced by ligand binding on wt PMM/PGM are plotted in Fig. III-2A. In general, these shifts of backbone NMR peaks

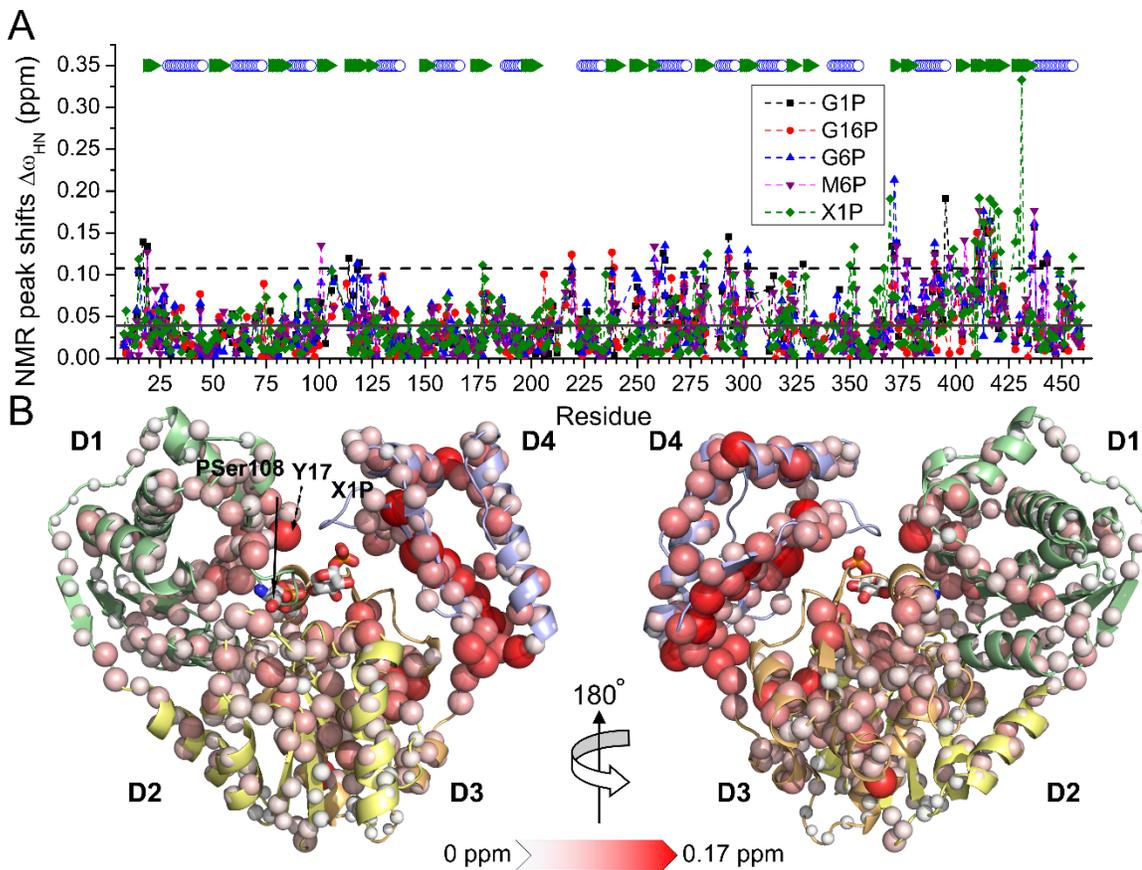


Figure III-2. Ligand binding perturbs chemical shifts of wt PMM/PGM residues (A). Plot of  $^1\text{H}$ ,  $^{15}\text{N}$  NMR peak shifts introduced by G1P (black squares), G16P (red circles), G6P (blue up triangles), M6P (purple down triangles), and X1P (green diamonds). The mean NMR peaks shifts are marked by a black line, while the dashed line represents threshold that is two  $\sigma$  above the average value. (B). Locations of mean NMR peak shift changes introduced by multiple ligand bindings are marked with spheres on the crystal structure of PMM:X1P complex (PDB ID: 2H5A). The magnitudes of NMR peak shifts are represented by white to red color gradient and radii of the spheres.

introduced by phosphosugars of differing structures and charge share similar patterns with minor differences. The largest perturbations are observed around residues 17-20 and 100-120 in domain 1, 207-220 in domain 2, 258-302 in domain 3, and most of domain 4 (D4). When the averages of these peaks shifts are mapped onto the crystal structure (Fig. III-2B), it is obvious that in domains 1-3 (D1-3), the residues with the largest shifts of

NMR peaks (red spheres with large radii) are located in the vicinity of the sugar-phosphate binding site. For example, Tyr17, which participates in protein-ligand contacts, shows significant shift (Fig. III-2B) (33). The entirety of D4 has large CSPs, consistent with the previous finding that D4 rotates upon ligand binding (14). Xylose 1-phosphate (X1P) binding (green diamonds in Fig. III-2A) shows larger deviations than other ligands around residues 345-355, 370-375, and 425-430. Most of these positions reside in the face of D4 toward the active site and D3, implying X1P binding has a greater impact on D4.

The peak shifts introduced to S108C-impaired enzyme are similar to those of the wt enzyme (Fig. SIII-4). This suggests that this substitution for the catalytic serine does not alter the rotation of D4. Glucose 1,6-phosphate (G16P) binding shows unique and large peak shifts at D3 residues 330-340 in the S108C mutant (red circles in Fig. SIII-5A).

#### **III.4.2 X1P and G16P bind comparatively tightly to wild-type enzyme**

Binding isotherms were obtained from principal component 1 (PC1) from PCA of <sup>15</sup>N TROSY peak lists from titrations in fast exchange (23) (Fig. III-3). In order to reveal the effects on the enzyme overall, all peaks were analyzed together without preference among peaks. Ligand binding affinities were easily obtained by fitting binding isotherms Eq. III-3. As shown in Fig. III-3, the binding affinities of G6P ( $K_D = 122 \pm 14 \mu\text{M}$ ) and M6P ( $K_D = 105 \pm 7 \mu\text{M}$ ) for wt enzyme are similar but a little higher than for G1P ( $K_D = 285 \pm 26 \mu\text{M}$ ). However, G16P ( $K_D = 25 \pm 4 \mu\text{M}$ ) and X1P ( $K_D = 27 \pm 13 \mu\text{M}$ ) bind to the wt enzyme with the highest affinities observed. Similar to wt PMM/PGM, dissociation constants of G1P, G6P, and X1P from the S108C-substituted enzyme are  $198 \pm 48 \mu\text{M}$ ,

61 ± 37 μM, and 103 ± 21 μM, respectively. However, the affinity of the mutant enzyme for G16P, the bisphosphoryl intermediate, is decreased, with  $K_D$  of 839 ± 221 μM.

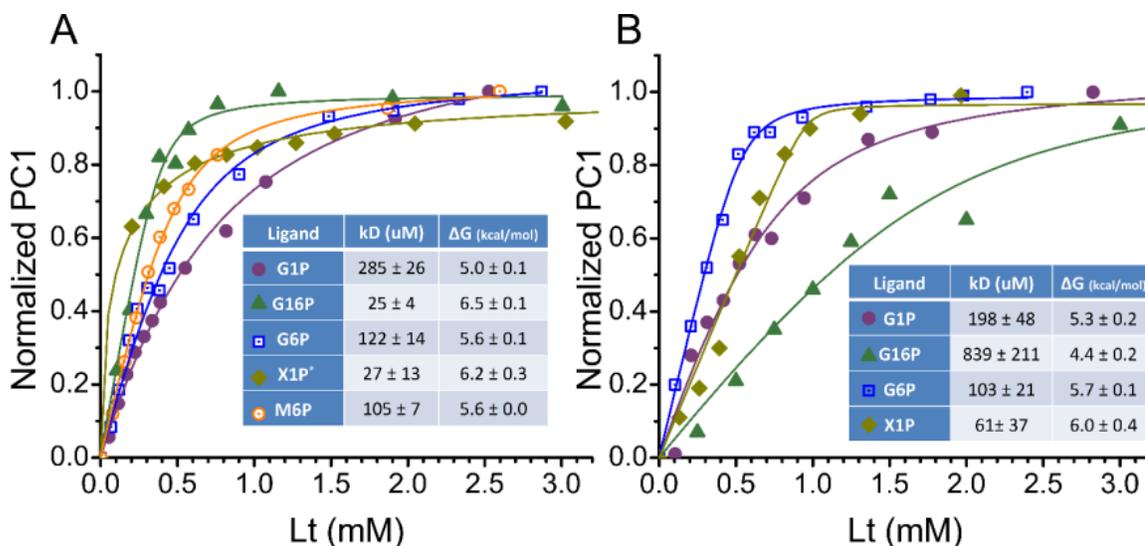


Figure III-3. Binding affinities of different ligands obtained by protein-observed NMR titrations of wt PMM/PGM (A) and the S108C mutant (B). \*X1P binding data was adapted from a titration in slow exchange regime in which  $N^{15}$  TROSY spectra collected at 298 K (see Fig. 2C, D in ref (23)).

For X1P binding to wt PMM/PGM at 308 K, there is a mixture of peaks undergoing fast, intermediate, or slow exchange. This complicated deriving a binding isotherm from the peak lists. Consequently, the affinity X1P was obtained by applying PCA to peak heights of an X1P titration at 298 K (23), where its amide peaks are in the slow exchange regime.

### III.4.3 Clustering of ligand binding vectors

PCA was carried out on CSP vectors of all perturbations (Table III-1) by singular value decomposition (SVD) in a very similar manner to CHESCA (18, 25), except for using CSP vectors instead of scalar combined chemical shifts (CCS) (26); see methods. The results are visualized by PCA biplot, which provides separation and clustering (34).

Fig. III-4 shows the interrelationships among the spectra perturbed by ligands or

dephosphorylation (Table III-1) using either two principal components (Fig. III-4F) or three (Fig. III-4G). In the 3D space constructed by the first three components (accounting for 74% of total variance; Fig. III-S6), all ligand-bound states lie within a cone (Fig. III-4G), suggesting similar effects due to phosphosugars binding to PMM/PGM. The vectors

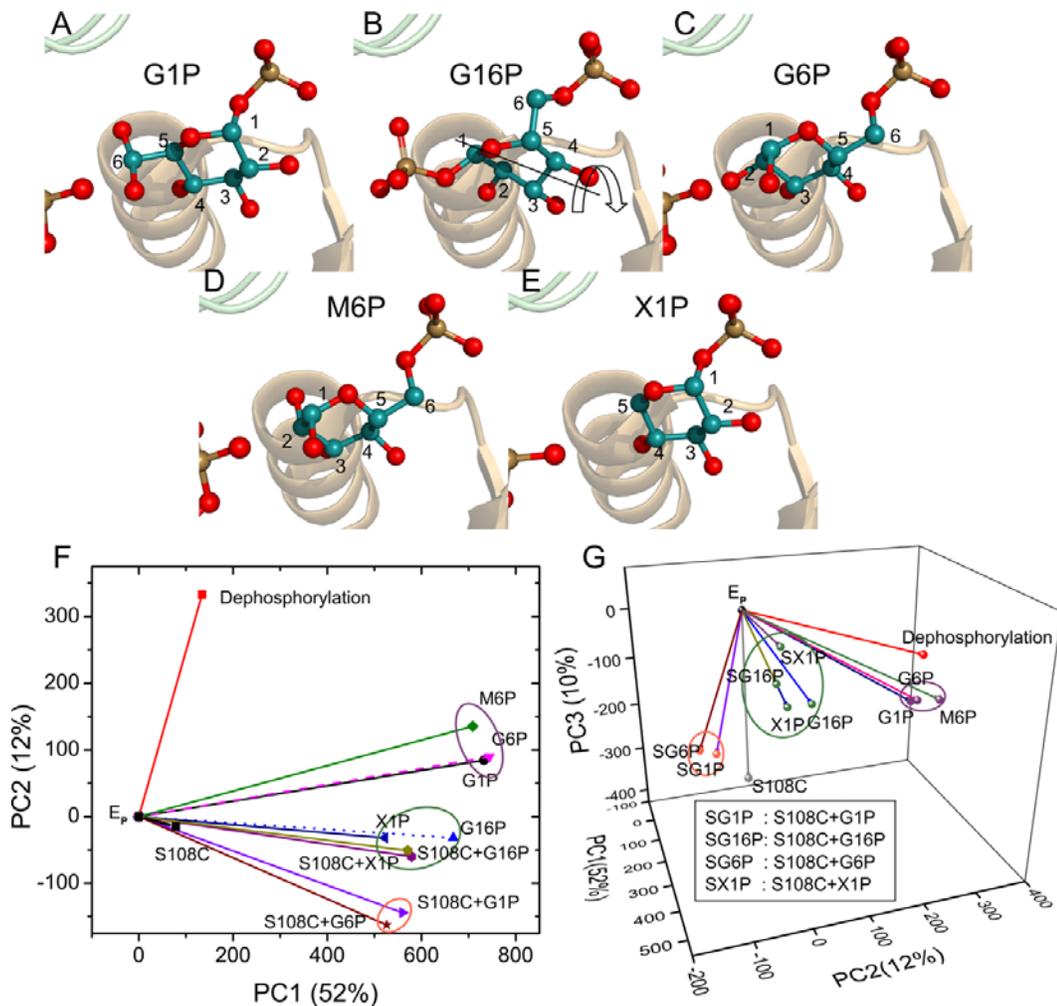


Figure III-4. Overall effects of ligand binding to PMM/PGM and its S108C mutant.(A-E): Structures of ligands: (A) glucose 1-phosphate (G1P), (B) glucose 1,6-bisphosphate (G16P), (C) glucose 6-phosphate (G6P), (D) mannose 6-phosphate (M6P), (E) xylose 1-phosphate (X1P). Ligands are shown in the active site of PMM/PGM complexes with PDB ID: (A) 1P5D, (B) 1P5G, (C) 2FKM, (D) 1PCM, (E) 2H5A, respectively. Numbers refer to the positions of sugar C atoms. G16P can rotate 180° in the active site as indicated by the arrow. (F) Biplot of PC1 and PC2. Purple, green, and orange circles mark different clusters. (G) Biplot of PC1-PC2-PC3 space. Clusters highlighted in (F) are shown in purple, green, and orange spheres.

for the S108C mutant and dephosphorylation point to directions outside the cone, suggesting these two perturbations are quite different from the binding of phosphosugars to the enzyme (Fig. III-4G). Although the S108C mutation removes pSer108 to establish a dephosphorylated state, it shows little similarity with  $E_{deP}$  in the biplot. This is indicated by the nearly perpendicular angle between the two vectors, which suggests statistical independence of the S108C lesion from  $E_{deP}$  (Fig. III-4F,G). Adding an extra dimension (PC3) helps resolve S108C from the other sources of perturbations to the spectra (Fig. III-4G).

Three clusters of phosphosugar binding can be identified in both the 2D and 3D plots of principal components (highlighted by the ellipses in Fig. III-4F,G). The clusters are visualized best in the projected PC1-PC2 plane. In the 2D biplot (Fig. III-4F), the responses to binding the substrates G1P, G6P, and M6P are very similar and hence can be clustered together (purple ellipse). The G1P and G6P states are almost identical in another cluster. The effects of G1P and G6P binding to the S108C mutant can be assigned to another cluster (orange ellipse). Thus, it seems that binding of monophosphosugar substrates (G1P, G6P) to wt PMM/PGM and S108C mutant are different, as indicated by the distribution of their clusters marked by purple and orange ellipses. This suggests the active pSer108 plays a role in the binding of monophosphosugar substrates.

In contrast, binding of G16P and X1P to both wt PMM/PGM and the S108C mutant belong to the same cluster, despite their structural and electrostatic differences. This third cluster in the biplot could suggest the mutation of pSer108 to Cys108 might not affect association with G16P and X1P as much. However, the S108C lesion impairs

affinity for G16P by 2 kcal/mol (Fig. III-3). This mutation introduced disorder to this key loop in the active site in the crystal structure (28), probably impeding association of G16P, which normally includes phosphoryl transfer to Ser108.

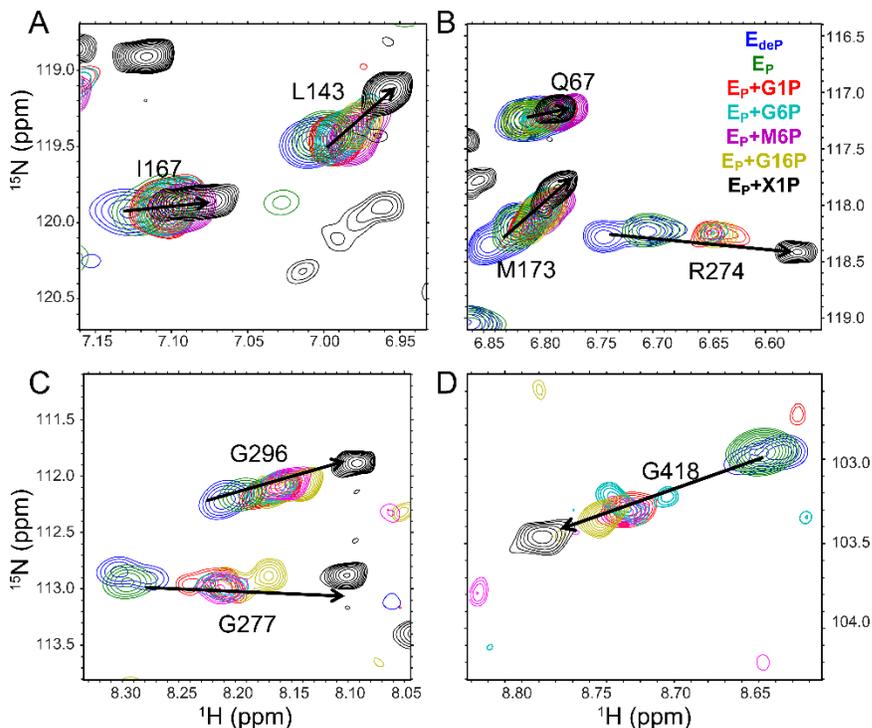


Figure III-5. Concerted chemical shift changes observed by  $^{15}\text{N}$  TROSY spectra. Blue contours represent dephosphorylated PMM/PGM ( $E_{\text{dep}}$ ); green contours represent phosphorylated PMM/PGM ( $E_{\text{P}}$ ); red contours represent G1P complex ( $E_{\text{P}}+\text{G1P}$ ), cyan contours represent the G6P complex ( $E_{\text{P}}+\text{G6P}$ ); purple contours represent the M6P complex ( $E_{\text{P}}+\text{M6P}$ ); yellow contours represent G16P complex ( $E_{\text{P}}+\text{G16P}$ ); black contours represent the X1P complex ( $E_{\text{P}}+\text{X1P}$ ). Examples in domains 1, 2, 3, and 4 are shown in subfigure A), B), C), and D), respectively.

#### III.4.4 Free and X1P-bound states lie at two extremes of conformational equilibria

In order to evaluate the conformational equilibria of PMM/PGM upon the application of various perturbations, a chemical shift based statistical tool, Coordinated Chemical shifts behavior (CONCISE) was performed on linear chemical shift trajectories. This serves to suppress minor effects, while retaining the main perturbations. The CONCISE approach suggests the “density of probabilities” of each state in the equilibria

(27). Unlike other chemical shift-based PCA analysis (21-23, 25, 35), CONCISE applies PCA individually on each residue and utilizes only PC1. PC1 captures the largest variance in the experiment and is regarded as reporting of equilibrium position of each residue (27). Several examples of amide peaks showing a high degree of linearity can be seen in Fig. III-5. Population distributions of each state along the major conformational equilibrium are evaluated based upon their projections on the PC1 direction (marked by black arrows in Fig. III-5), while variances in other directions (such as curvatures found in the trajectories of Leu143, Arg274, and G277) are not considered.

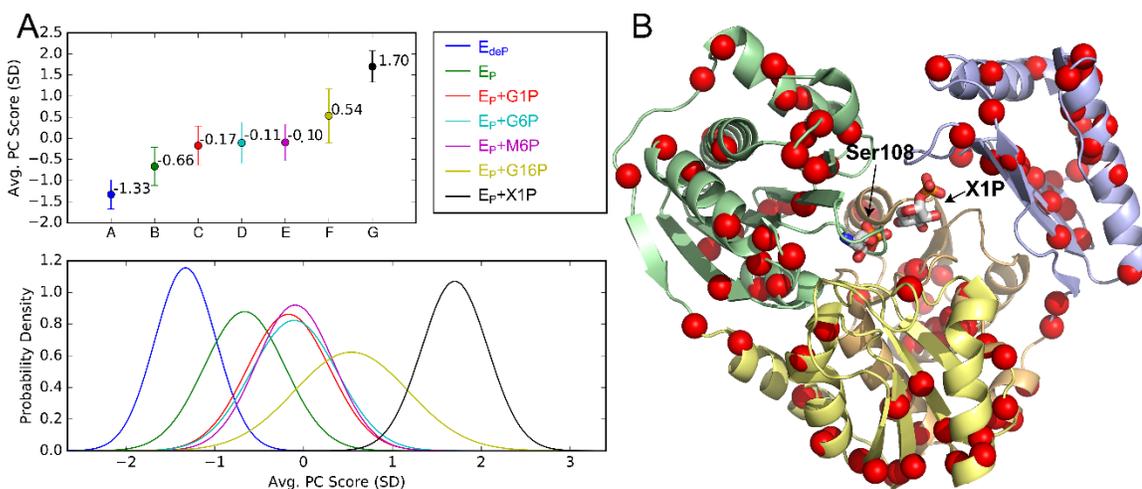


Figure III-6. Equilibrium shifts of PMM and its complexes characterized by CONCISE analysis of amide resonances. (A). Average PC scores and the corresponding probability densities of different states; (B) Linearly shifted residues used for CONCISE analysis are marked as red spheres in crystal structure of PMM:X1P complex (PDB ID: 2H5A).

CONCISE identified 110 residues with linear chemical shift trajectories upon various perturbations in PMM/PGM (Fig. III-6). These 110 residues with linear shifts residues are distributed over the whole enzyme structure (Fig. III-6B), suggesting the participation of the entire protein. The equilibrium values measured for  $E_{dep}$ ,  $E_p$ ,  $E_p+G1P$ ,  $E_p+G6P$ ,  $E_p+M6P$ ,  $E_p+G16P$ , and  $E_p+X1P$  are -1.33, -0.66, -0.17, -0.11, -0.1, 0.54, 1.7

standard units. Normalized to a scale of 0 to 100%, these equilibrium positions are 0, 22%, 38%, 40%, 41%, 62%, and 100%, respectively.

These values correspond to the positions of the peaks in the population density plot (lower panel of Fig. III-6A). In that plot,  $E_{\text{deP}}$  (purple blue) and  $E_{\text{P+X1P}}$  (black) define two extremes of equilibrium positions, which can be expressed as 0% and 100%. The position of  $E_{\text{P}}$  (green) is next to  $E_{\text{deP}}$  at 22%. Peaks of enzyme bound with monophosphosugar substrates (G1P, G6P, and M6P) are overlapped and located at ~40% (red, cyan, and purple) of the equilibrium. The wide peak of  $E_{\text{P+G16P}}$  (yellow) resides between the peaks of the monophosphosugar-bound states and the X1P-bound state at a position of 60%. The broader distribution of the  $E_{\text{P+G16P}}$  state suggests a mixture of substates in the complex with this intermediate.

CONCISE performed on the S108C mutant identified 111 linearly shifting peaks (Fig. SIII-7). The equilibrium states measured from these 111 place the free state (blue) and X1P-bound (purple) states at the two extremes 0% and 100%, very much like with wt PMM/PGM. Peaks of enzyme bound to monophosphosugar substrates, including S108C+G1P (green) and S108C+G6P (red), are overlapped and located at a position of about 32%. The peak for the S108C complex with G16P is wide, as in the wt enzyme. The G16P probability density is likewise centered at 74% between the peaks of monophosphosugar-bound states and the peak of the X1P complex positioned (Fig. SIII-7).

#### **III.4.5 Ligand binding slows hydrogen exchange especially in Domains 2 and 4**

The effects of ligand binding on stability and flexibility of PMM/PGM were evaluated by comparing the hydrogen exchange behavior with and without ligand binding.

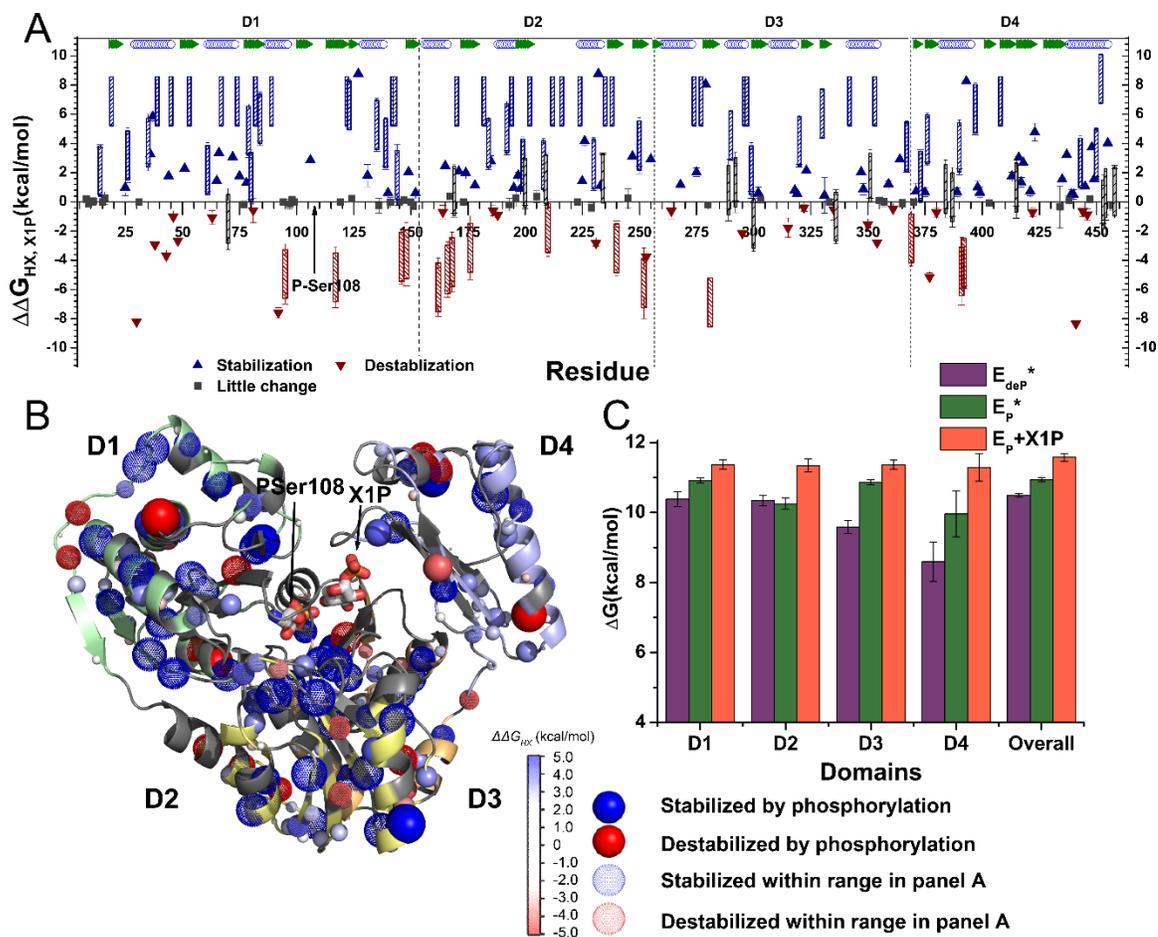


Figure III-7. X1P binding increases  $\Delta G_{HX}$  at more sites than it decreases. (A) Free energy difference ( $\Delta\Delta G_{HX, X1P}$ ) between X1P-bound complex and apo state. Triangles indicate cases where  $k_{ex}$  was measured in both  $E_P$  and  $E_P + X1P$  forms. Hatched bars represent cases in which  $k_{ex}$  in one state is unmeasurable and estimated in the range  $1 \text{ s}^{-1} > k_{ex} > 4 \times 10^{-3} \text{ s}^{-1}$ ; (B)  $\Delta\Delta G_{HX, X1P}$  values are mapped onto a crystal structure (PDB ID: 2H5A) as spheres. Amide groups stabilized from HX by X1P binding are marked by blue spheres, while amide groups destabilized (mobilized) by X1P binding are in red. Dotted spheres correspond to residues with uncertainty ranges marked as hatched bars in (A). The magnitudes of  $\Delta\Delta G_{HX, X1P}$  are symbolized by color gradient as well as radii of the spheres. (C). Folding stabilities of each domain, or the enzyme as a whole suggest X1P binding mainly stabilizes D2 and D4 compared with  $E_P$ . \*Data of  $E_P$  and  $E_{deP}$  were adapted from Fig. S9 in ref (13).

Thermodynamic quantification of NMR-detected hydrogen exchange (HX) was performed on the wt complex with X1P and compared with previously reported free energy of HX determined for the free state with phosphorylation ( $E_P$ ) and without ( $E_{deP}$ ) (13). A pH of 7.4 and temperature of 35 °C ensure that HX of the X1P complex occurs in the bimolecular EX2 regime, where residues with the largest  $\Delta G_{HX}$  can be used to

estimate folding stability (36). Both subsecond HX and slow hydrogen-deuterium exchange behaviors of X1P complex were measured, i.e. by CLEANEX-PM or HDX-NMR, respectively. This revealed rate constants and  $\Delta G_{HX}$  for 201 residues.  $\Delta G_{HX}$  values of E<sub>P</sub>-X1P were plotted together with E<sub>P</sub> for comparison (Fig. SIII-8). Seventy-four amide residues in E<sub>P</sub>-X1P decay too slowly to be quantified by HDX-NMR. Consequently their lower bounds were estimated by setting  $k_{ex}$  as  $5.56 \times 10^{-5} \text{ min}^{-1}$ , which is the slowest  $k_{ex}$  value that could be measured experimentally in the E<sub>P</sub> and E<sub>deP</sub> forms of PMM/PGM (13). On the other hand, only 20 groups in E<sub>P</sub> needed estimation with this lower bound.

$\Delta\Delta G_{HX}$  values calculated on a residue-by-residue basis as differences between  $\Delta G_{HX}$  of E<sub>P</sub>-X1P and E<sub>P</sub> are plotted in Fig. III-7A and mapped onto the crystal structure of PMM/PGM (spheres in Fig. III-7B). The comparison shows that amide groups stabilized by X1P binding are interspersed throughout all four domains of PMM/PGM. Residues in the vicinity of bound ligand do not show higher stabilization than remote residues, indicating the long-range nature of effects of ligand binding. Apart from the overall stabilization by binding of this inhibitor, 25 residues located at positions at least 10 Å from the binding site appear to be destabilized ( $\Delta\Delta G_{HX} < 2 \text{ kcal/mol}$ ), i.e. with accessibility to HX increased by the X1P bound (red spheres). Proteins that undergo a mixture of slowing, acceleration, and no change of hydrogen exchange upon ligand binding have been discussed (37). These authors pointed out that a variety of effects of ligand binding upon the open excited states at different sites in a protein can accompany the overall stabilization of the ground state by the ligand. This can account for the mix of hydrogen bonds being stabilized, destabilized, or unaffected by ligand binding (37). The

strategy of estimating global protein stability using the amide groups most protected from hydrogen exchange (36) is not affected by localized accelerations of hydrogen exchange. The estimates of folding stability by this method for each domain and the entire enzyme suggest that the X1P-bound state is globally stabilized by  $\sim 0.6$  kcal/mol over the  $E_p$  state. D2 and D4 are most stabilized (Fig. III-7C). D3, which contains the sugar-binding loop, is strongly stabilized by phosphorylation (13), but shows relatively little stabilization by X1P binding.

## **III.5 Discussion**

### **III.5.1 Highest affinity for the bis-phosphorylated intermediate and an inhibitor**

NMR titrations of various sugars revealed a continuum of binding affinities of different phosphosugars for wt PMM/PGM and its inactivated S108C mutant (which replaces the catalytic phosphoSer108). Titrations of ligands into wt PMM/PGM indicate that X1P and G16P are the tightest binding ligands, with binding energies of about  $\sim 6.5$  kcal/mol. Monophosphosugar substrates bind to wt PMM/PGM with less affinity. The free energies of G1P, G6P, and M6P binding are 5.0, 5.6, and 5.6 kcal/mol, respectively. The substitution of pSer108 in S108C does not significantly change binding affinities of G1P and G6P, whose binding energies are 5.3 and 5.7 kcal/mol. X1P binding to S108C is still tight, showing little difference from wt, with a  $\Delta G$  value of 6.0 kcal/mol.

### **III.5.2 pSer108 has little effect on ligand affinities, except for the G16P intermediate**

The comparison between titrations of wt PMM/PGM and S108C mutant suggests that phosphorylation of Ser108 does not influence binding affinities of monophosphosugars such as G1P and G6P, although the negative charge phosphorylation of Ser108 was shown to make PMM/PGM more compact and stable by attracting

domains together. Crystallographic studies show that phosphate groups of all monophosphosugars bind in essentially the same site, while their sugar rings of 1-phosphate and 6-phosphate are flipped relative to each other by 180° (Fig. SIII-1) and ref (11, 15). Their phosphate groups are regarded as potential determinants of ligand binding (11). This conclusion is reinforced by this study, as pSer108 is ruled out as a significant determinant of affinity for monophosphosugars.

However, the free energy of binding of G16P changed considerably from 6.5 kcal/mol in wt PMM/PGM to 4.4 kcal/mol in the S108C mutant. The unique shifts of the NMR peaks of residues 250, 258, and residues 325, 328 in the S108C complex with G16P may suggest an altered mode of binding G16P (Figs. SIII-4, SIII-9). The disorder introduced by the S108C mutation to the active site loop from His109 through Tyr114 (28) may impair affinity for G16P, as one of its phosphate groups is normally positioned near residue 108.

### **III.5.3 Five distinct states of PMM/PGM identified by CONCISE**

Based upon the statistical analysis of CSP vectors, CONCISE analysis seems to distinguish various states of PMM/PGM and its S108C mutant in solution. There are five distinguishable groups of equilibrium positions of wt PMM/PGM of relative positioning around 0% ( $E_{deP}$ ), 22% ( $E_P$ ), ~40% (G1P, G6P, and M6P complexes), 60% ( $E_P+G16P$ ), and 100% ( $E_P+X1P$ ) (Fig. III-6). Similarly, CONCISE identified four groups in titrations of the S108C mutant, whose relative equilibrium positions are: 0% (Apo-S108C), 32% (G1P and G6P complexes), 74% (S108C-G16P), and 100% (S108C+X1P). In both the wt and S108C mutant, the population density of the G16P complexes is broadened, suggesting a mixture of multiple substates potentially caused by different

phosphorylation states of Ser108 or reorientation of G16P in the active site. The equilibrium positioning of the states of wt PMM/PGM (Fig. III-6) and its S108C mutant (Fig. SIII-7) appear very similar. However, wt PMM/PGM has two apo states (phosphorylated and dephosphorylated). In other words, the presence of the pSer does not change the pattern of progressive conformational transitions in response to binding of phosphosugars, implying that the phosphorylation of Ser108 may not play an important role in the global response of PMM/PGM to ligand binding.

#### **III.5.4 PCA clustering in two and three dimension confirms CONCISE results**

CONCISE analysis of only PC1 of peaks that shift linearly reflects the major conformational change (27). PCA performed on unfiltered lists of CSP vectors may provide information that was filtered out by CONCISE. Compared to 1D results from CONCISE, biplots of PCA depict similarities and differences between various perturbations (Table III-1) in a less quantitative but better resolved manner with second or third dimension. Groups of states identified by CONCISE are also recognized as clusters in PCA biplots (Fig. III-4). The apo states ( $E_P$ ,  $E_{deP}$ , and Apo S108C) are distinctly from ligand-bound states. Binding of monophosphosugars to wt PMM/PGM and the S108C mutant form two different clusters (purple and orange circles in Fig. III-4A, B). Interestingly, binding of X1P and G16P to wt PMM/PGM or the S108C mutant converges to another single cluster (green circle in Fig. III-4A, B), suggesting their similarities. Perhaps binding of X1P or G16P binding can overcome part of the disruptive effects of the S108C lesion. Thus, comparing PCA-identified clusters with CONCISE-measured states confirms the major equilibrium states.

### III.5.5 CONCISE reveals distinctions between states overlooked by crystallography

CONCISE was first applied to analyze allostery in the C-subunit of the cAMP-dependent protein kinase A (PKA-C) (27). Binding of ligands has been shown to shift PKA-C from an apo, open state towards a fully closed ligand-bound form. CONCISE distinguished these multiple states and quantified their population densities based on statistical analysis of chemical shifts (27). The equilibrium positions of different states were found to correspond to their protein conformations, as observed in crystal structures (38). However the case of PMM/PGM appears more complex. Crystal structures revealed only three different conformational states for PMM/PGM: the open apo state, a half-closed bound to G16P, and a closed state that is highly similar among all other ligand complexes (14). The evidences for dephosphorylation-induced dynamic opening of the catalytic cleft came more recently from studies of hydrogen exchange, small angle X-ray scattering, and molecular dynamics simulations (12, 13)., The CONCISE approach now suggests the progressive shifts among three different ligand-bound states. This cannot be understood through the earlier viewpoint of simple open-to-closed transition. There are perceptible differences among the closed states, which are potentially nuanced, such as perhaps in the proportions of time spent with the catalytic cleft fully closed.

Since G16P and X1P share similar affinities, cluster by PCA of the spectra of their complexes with PMM/PGM, and can each donate a phosphoryl group to Ser108, it may be that both G16P and X1P decrease the flexibility of the enzyme active site by suppressing the appearance of the transiently dephosphorylated or open intermediates. X1P may lie at the extreme of CONCISE-measured equilibrium because it lacks the ability of the 180° rotation possessed by G16P. X1P might rigidify the active site more,

which is consistent with its overall slowing of hydrogen exchange. Phosphorylation and binding of monophosphosugars may push the equilibrium from the open, more flexible, and less stable dephosphorylated state to a more fully closed, rigid, stabilized and phosphorylated state progressively.

### **III.6 Summary**

We have characterized responses of PMM/PGM to ligand binding by PCA and closely related CONCISE analysis of chemical shift perturbation vectors. Chemical shift-based analysis revealed multiple ligand-bound states not resolved by crystal structures. We have found that phosphorylation of Ser108 does not influence affinities for monophosphosugars. The effects of ligand-binding are long range, according to both the analyses of shifted NMR peaks and hydrogen exchange protection. The synergistic effects of phosphorylation and substrate recognition progressively shift the equilibrium states along a continuum from the dephosphorylated state towards the state with an inhibitory phosphosugar in the active site. Our study provides new insights on understanding the conformational ensemble of a major class of phosphoryl-transfer enzymes.

### **III.7 Acknowledgement**

Dr. Akellas V. Sarma performed the NMR titrations of wt PMM/PGM. Yirui Wei performed the NMR titrations of the S108C mutant. Alex H. Hopkins expressed and purified wt PMM/PGM.

### III.8 References

1. Bar-Even, A., R. Milo, E. Noor, and D. S. Tawfik. 2015. The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness. *Biochemistry* 54:4969-4977.
2. Cooper, A., and D. T. Dryden. 1984. Allostery without conformational change. A plausible model. *Eur. Biophys. J.* 11:103-109.
3. Williamson, M. P. 2013. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* 73:1-16.
4. Ballok, A. E., and G. A. O'Toole. 2013. Pouring Salt on a Wound: *Pseudomonas aeruginosa* Virulence Factors Alter Na<sup>+</sup> and Cl<sup>-</sup> Flux in the Lung. *J. Bacteriol.* 195:4013-4019.
5. Li, X. J., Q. Li, L. Y. Si, and Q. Y. Yuan. 2011. Bacteriological Differences Between Patients with Acute Exacerbation of COPD and Community-Acquired Pneumonia. *Respir. Care* 56.
6. Govan, J. R. W., and V. Deretic. 1996. Microbial pathogenesis in cystic fibrosis - Mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol. Rev.* 60:539-574.
7. Ye, R. W., N. A. Zielinski, and A. M. Chakrabarty. 1994. Purification and characterization of phosphomannomutase/phosphoglucomutase from *Pseudomonas aeruginosa* involved in biosynthesis of both alginate and lipopolysaccharide. *J. Bacteriol.* 176:4851-4857.
8. King, J. D., D. Kocíncová, E. L. Westman, and J. S. Lam. 2009. Review: Lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa*. *Innate Immunity* 15:261-312.
9. Olvera, C., J. B. Goldberg, R. Sanchez, and G. Soberon-Chavez. 1999. The *Pseudomonas aeruginosa* algC gene product participates in rhamnolipid biosynthesis. *FEMS Microbiol. Lett.* 179:85-90.
10. Remminghorst, U., and B. H. Rehm. 2006. Bacterial alginates: from biosynthesis to applications. *Biotechnol. Lett.* 28:1701-1712.
11. Regni, C., L. E. Naught, P. A. Tipton, and L. J. Beamer. 2004. Structural basis of diverse substrate recognition by the enzyme PMM/PGM from *P. aeruginosa*. *Structure* 12:55-63.
12. Lee, Y., M. T. Villar, A. Artigues, and L. J. Beamer. 2014. Promotion of Enzyme Flexibility by Dephosphorylation and Coupling to the Catalytic Mechanism of a Phosphohexomutase. *J. Biol. Chem.* 289:4674-4682.
13. Xu, J., Y. Lee, L. J. Beamer, and S. R. Van Doren. 2015. Phosphorylation in the catalytic cleft stabilizes and attracts domains of a phosphohexomutase. *Biophys. J.* 108:325-337.
14. Regni, C., A. M. Schramm, and L. J. Beamer. 2006. The Reaction of Phosphohexomutase from *Pseudomonas aeruginosa*: Structural Insights Into A Simple Processive Enzyme. *J. Biol. Chem.* 281:15564-15571.
15. Regni, C., G. S. Shackelford, and L. J. Beamer. 2006. Complexes of the enzyme phosphomannomutase/phosphoglucomutase with a slow substrate and an inhibitor. *Acta Crystallograph Sect F Struct Biol Cryst Commun* 62:722-726.
16. Chuang, G. Y., R. Mehra-Chaudhary, C. H. Ngan, B. S. Zerbe, D. Kozakov, S. Vajda, and L. J. Beamer. 2010. Domain motion and interdomain hot spots in a

- multidomain enzyme. *Protein science : a publication of the Protein Society* 19:1662-1672.
17. Furukawa, A., T. Konuma, S. Yanaka, and K. Sugase. 2016. Quantitative analysis of protein–ligand interactions by NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* 96:47-57.
  18. Xu, J., and S. R. Van Doren. 2017. TRacking Equilibrium and Non-equilibrium Shifts in Data with TREND, *Biophys. J.*, <http://dx.doi.org/10.1016/j.bpj.2016.12.018>
  19. Ricciardi-Castagnoli, P., and F. Granucci. 2002. Opinion: Interpretation of the complexity of innate immune responses by functional genomics. *Nat Rev Immunol* 2:881-889.
  20. Ma, S., and Y. Dai. 2011. Principal component analysis based methods in bioinformatics studies. *Brief Bioinform* 12:714-722.
  21. Sakurai, K., and Y. Goto. 2007. Principal component analysis of the pH-dependent conformational transitions of bovine beta-lactoglobulin monitored by heteronuclear NMR. *Proc. Natl. Acad. Sci. U. S. A.* 104:15346-15351.
  22. Konuma, T., Y. H. Lee, Y. Goto, and K. Sakurai. 2013. Principal component analysis of chemical shift perturbation data of a multiple-ligand-binding system for elucidation of respective binding mechanism. *Proteins* 81:107-118.
  23. Xu, J., and S. R. Van Doren. 2016. Binding Isotherms and Time Courses Readily from Magnetic Resonance. *Anal. Chem.* 88:8172-8178.
  24. Arai, M., J. C. Ferreón, and P. E. Wright. 2012. Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J. Am. Chem. Soc.* 134:3792-3803.
  25. Selvaratnam, R., S. Chowdhury, B. VanSchouwen, and G. Melacini. 2011. Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc. Natl. Acad. Sci. U. S. A.* 108:6133-6138.
  26. Boulton, S., M. Akimoto, R. Selvaratnam, A. Bashiri, and G. Melacini. 2014. A tool set to map allosteric networks through the NMR chemical shift covariance analysis. *Sci Rep* 4:7306.
  27. Cembran, A., J. Kim, J. Gao, and G. Veglia. 2014. NMR mapping of protein conformational landscapes using coordinated behavior of chemical shifts upon ligand binding. *Phys Chem Chem Phys* 16:6508-6518.
  28. Sarma, A. V., A. Anbanandam, A. Kelm, R. Mehra-Chaudhary, Y. Wei, P. Qin, Y. Lee, M. V. Berjanskii, J. A. Mick, L. J. Beamer, and S. R. Van Doren. 2012. Solution NMR of a 463-residue phosphohexomutase: domain 4 mobility, substates, and phosphoryl transfer defect. *Biochemistry* 51:807-819.
  29. Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6:277-293.
  30. Goddard, T. D., and D. G. Kneller. SPARKY 3. University of California, San Francisco.
  31. Hwang, T. L., P. C. van Zijl, and S. Mori. 1998. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical

- EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* 11:221-226.
32. Bertini, I., K. Ghosh, A. Rosato, and P. R. Vasos. 2003. A high-resolution NMR study of long-lived water molecules in both oxidation states of a minimal cytochrome c. *Biochemistry* 42:3457-3463.
  33. Schramm, A. M., R. Mehra-Chaudhary, C. M. Furdui, and L. J. Beamer. 2008. Backbone flexibility, conformational change, and catalysis in a phosphohexomutase from *Pseudomonas aeruginosa*. *Biochemistry* 47:9154-9162.
  34. Chapman, S., P. Schenk, K. Kazan, and J. Manners. 2002. Using biplots to interpret gene expression patterns in plants. *Bioinformatics* 18:202-204.
  35. Dawson, J. E., P. J. Farber, and J. D. Forman-Kay. 2013. Allosteric Coupling between the Intracellular Coupling Helix 4 and Regulatory Sites of the First Nucleotide-binding Domain of CFTR. *PLoS ONE* 8:e74347.
  36. Huyghues-Despointes, B. M. P., J. M. Scholtz, and C. N. Pace. 1999. Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat. Struct. Biol.* 6:910-912.
  37. Konermann, L., A. D. Rodriguez, and M. A. Sowole. 2014. Type 1 and Type 2 scenarios in hydrogen exchange mass spectrometry studies on protein-ligand complexes. *Analyst* 139:6078-6087.
  38. Masterson, L. R., A. Mascioni, N. J. Traaseth, S. S. Taylor, and G. Veglia. 2008. Allosteric cooperativity in protein kinase A. *Proc. Natl. Acad. Sci. U. S. A.* 105:506-511.

### III.9 Supporting figures

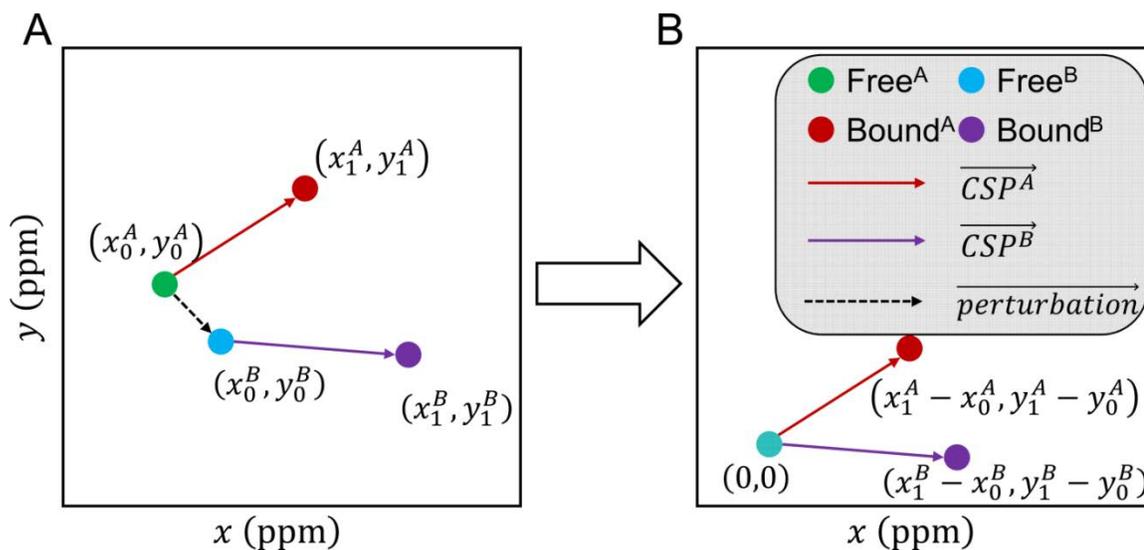


Figure SIII-1. Calculation of ligand binding CSP vectors used for PCA and CONCISE analysis. In  $^1\text{H}$ ,  $^{15}\text{N}$  TROSY spectra reported in this work, x axis represents  $^1\text{H}$  chemical shifts; y axis represents  $0.2 \times ^{15}\text{N}$  chemical shifts. A and B label two example species, for example, wildtype and S108C mutant. Green and cyan spheres represent initial states of A and B species, while brown and purple spheres stand for end (ligand-bound) states. Ligand binding CSP vectors of A and B states are marked by brown and purple arrows, while the perturbation between A and B state (in this case S108C mutation) is represented by blacked dashed arrow.

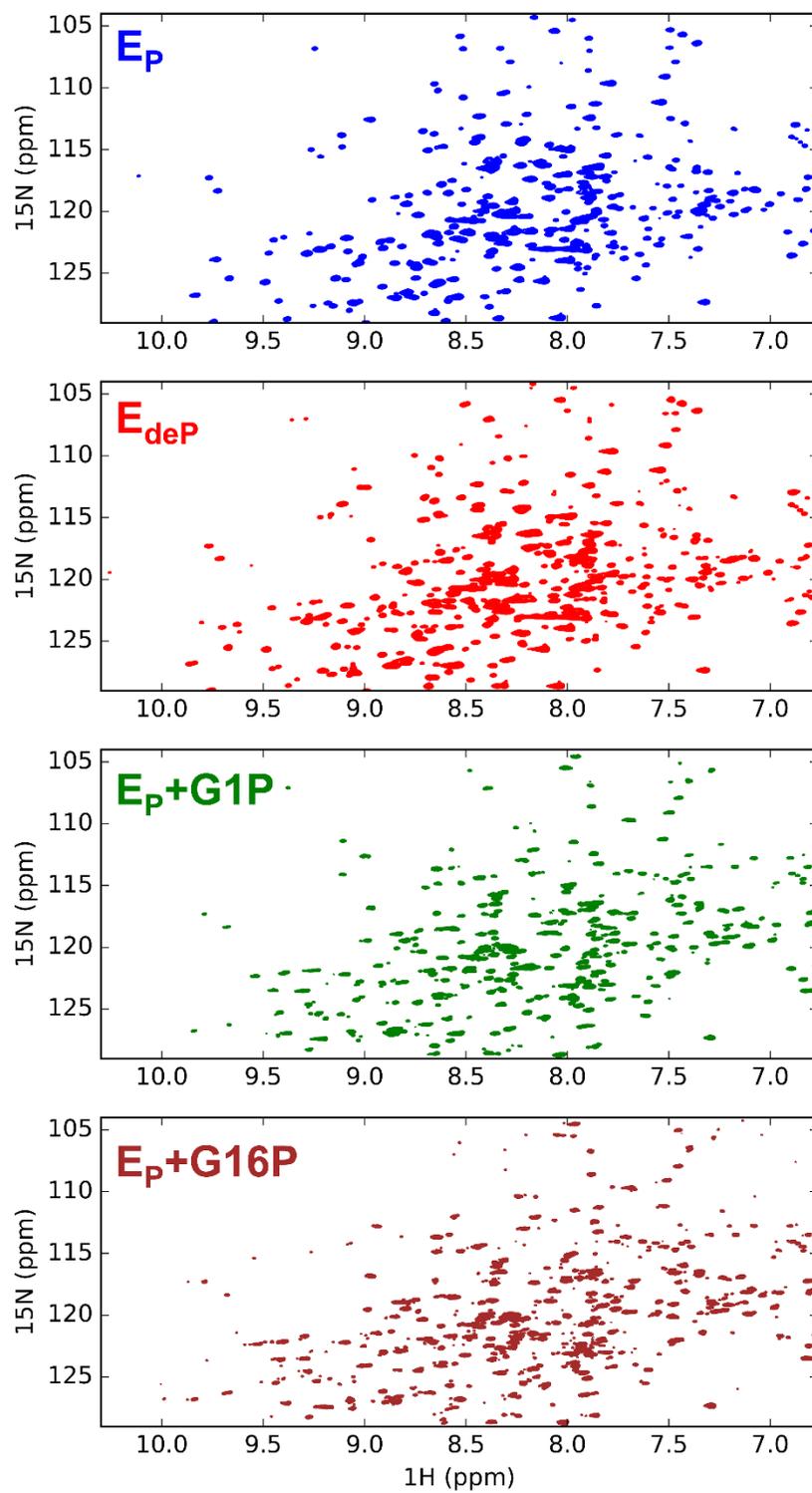


Figure SIII-2.  $^1\text{H}$ - $^{15}\text{N}$  TROSY spectra for  $E_p$  (blue),  $E_{\text{deP}}$  (red),  $E_p + \text{G1P}$  (green), and  $E_p + \text{G16P}$  (brown).

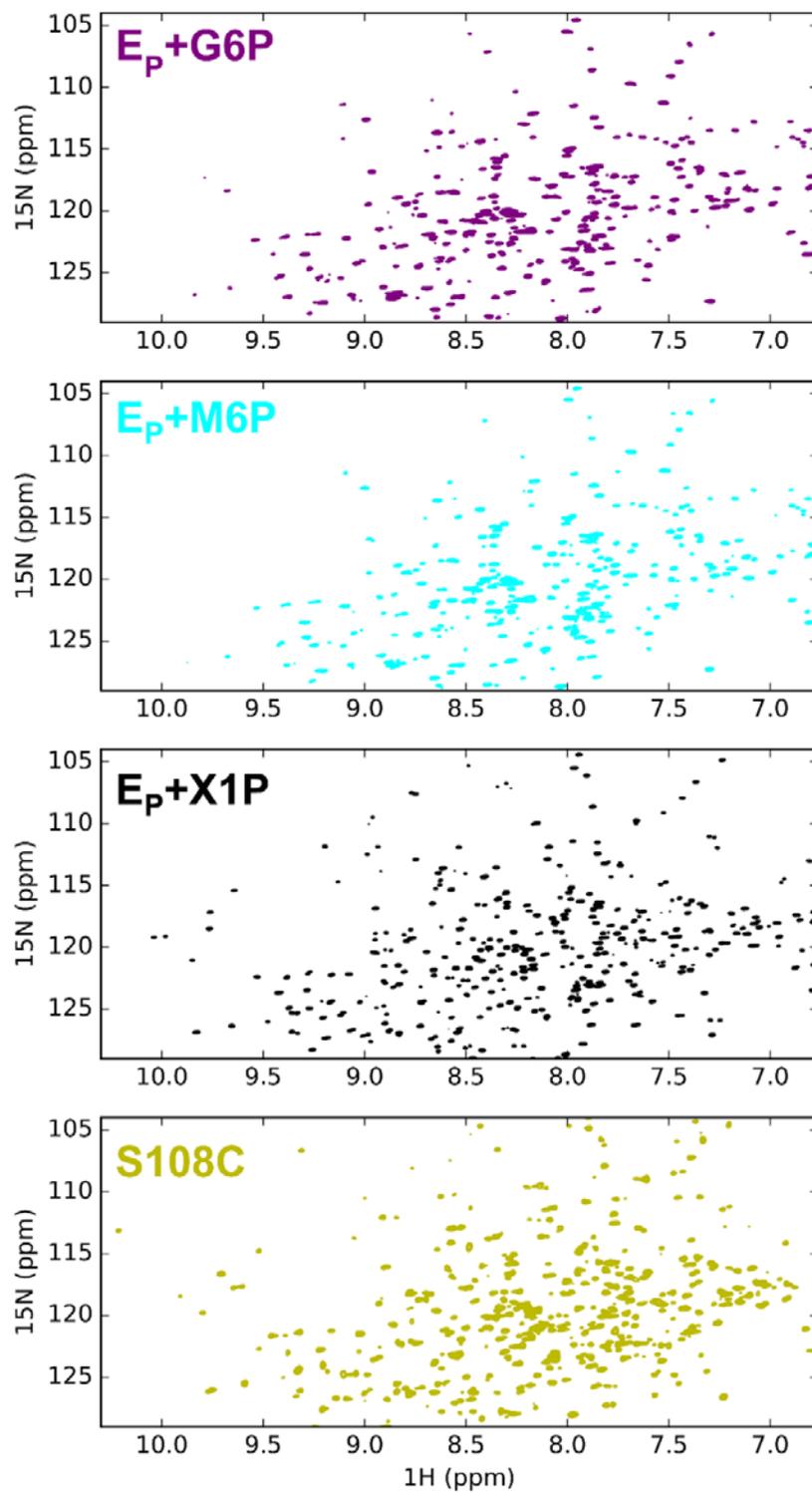


Figure SIII-3.  $^1\text{H}$ - $^{15}\text{N}$  TROSY spectra for  $E_p + \text{G6P}$  (purple),  $E_p + \text{M6P}$  (cyan),  $E_p + \text{X1P}$  (black), and S108C mutant (yellow)

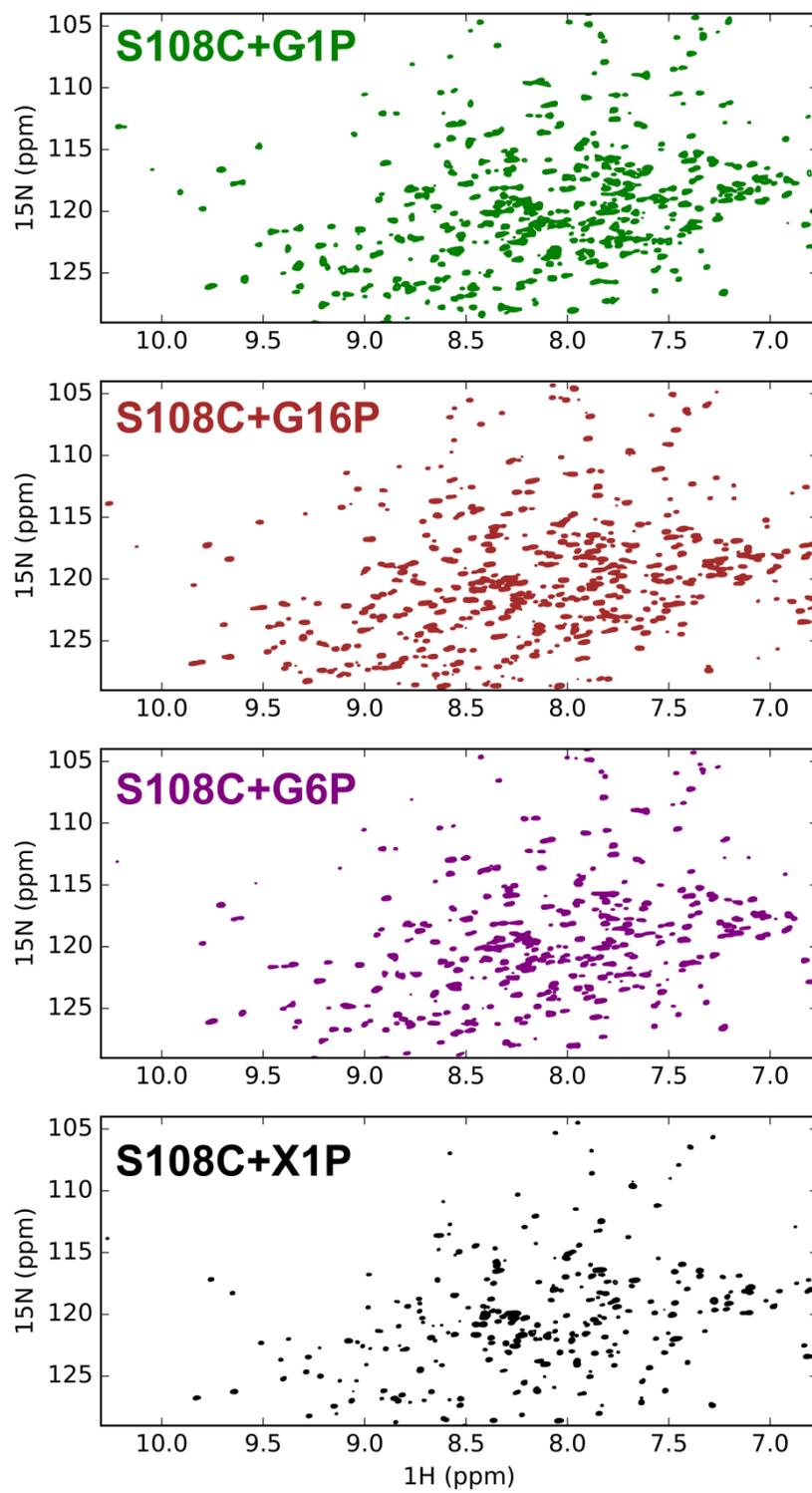


Figure SIII-4.  $^1\text{H}$ - $^{15}\text{N}$  TROSY spectra for S108C+G1P (green), S108C+G16P (brown), S108C+G6P (purple), and S108C+X1P (black).

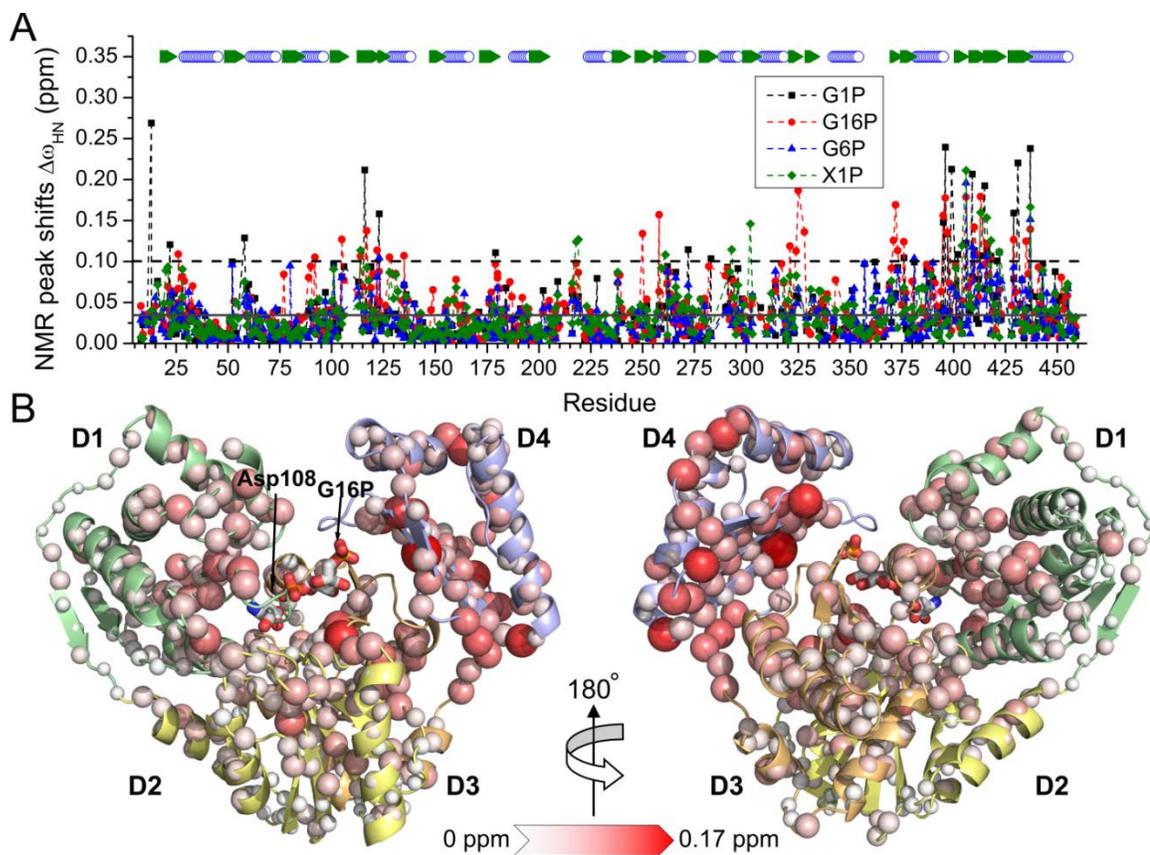


Figure SIII-5. Ligand binding perturbs chemical shifts of S108C mutant residues (A). Plot of  $^1\text{H}$ ,  $^{15}\text{N}$  NMR peak shifts introduced by G1P (black squares), G16P (red circles), G6P (blue up triangles), and X1P (green diamonds). (B). Locations of NMR peak shifts introduced by ligand binding are marked with spheres on the crystal structure of PMM:G16P complex (PDB ID: 2FKM). The magnitude of NMR peak shifts is marked by white to red color gradient as well as radii of the spheres.

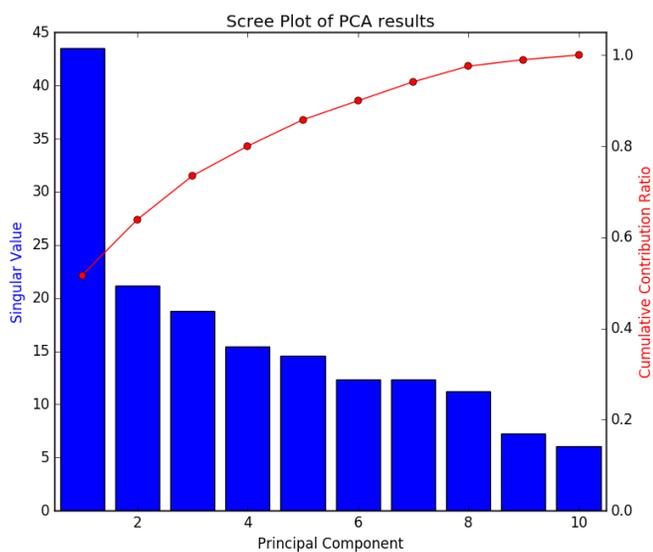


Figure SIII-6. Scree plot of Fig. III-4 (F, G).

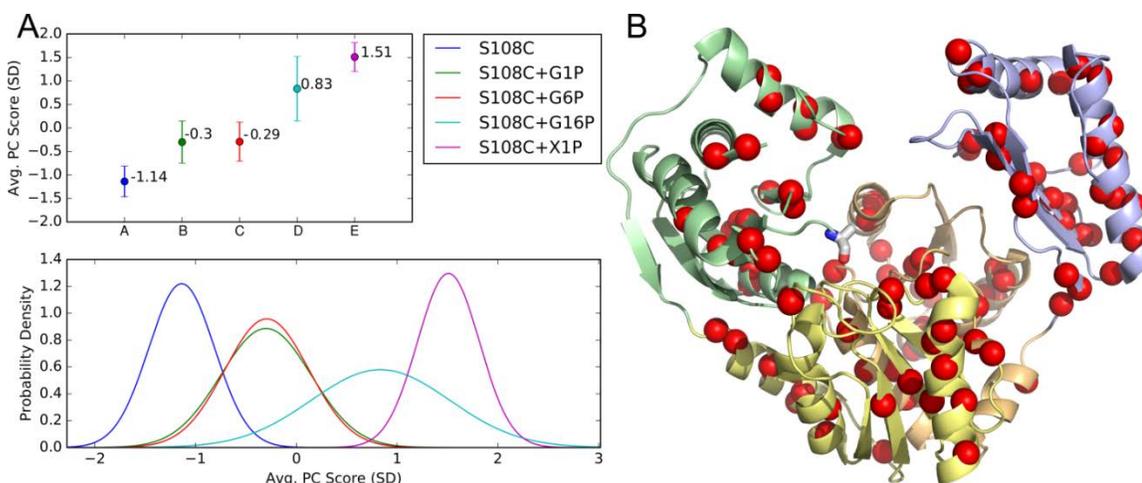


Figure SIII-7. Equilibrium shifts of PMM and its complexes characterized by CONCISE analysis of amide resonances. (A). Average PC scores and the corresponding probability densities of different states; (B) Linearly shifted residues used for CONCISE analysis are marked as red spheres in crystal structure of PMM:X1P complex (PDB ID: 2H5A).

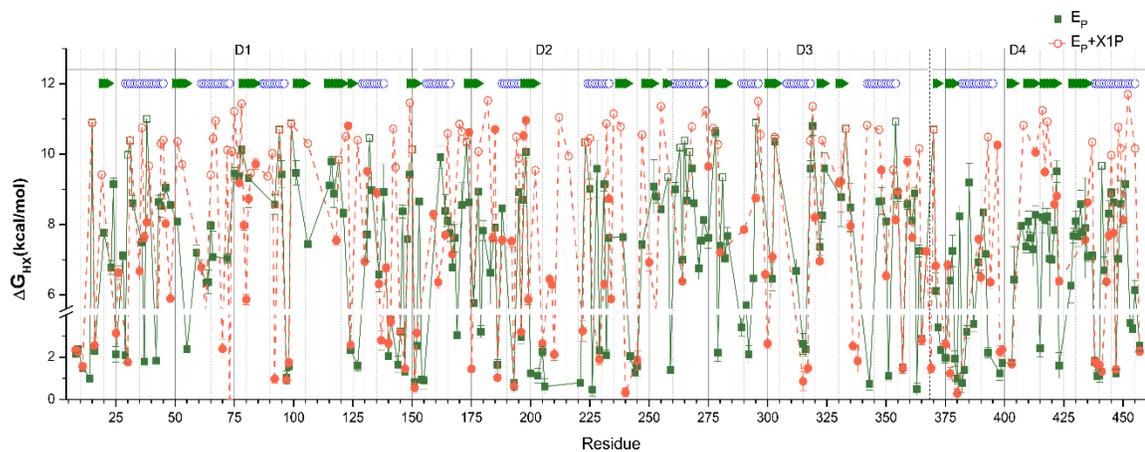


Figure SIII-8. Comparison of  $\Delta G_{HX}$  between  $E_P$  and  $E_P+X1P$  complex shows ligand-induced stabilization.  $\Delta G_{HX}$  values obtained by HDX experiments are shown above the break, while the small  $\Delta G_{HX}$  values measured by CLEANEX-PM NMR are shown below the break. Lower bounds of the  $\Delta G_{HX}$  that were estimated for slowly exchanging amide residues were shown in open squares. Strands and helices are marked by green triangles and blue circles at the top.

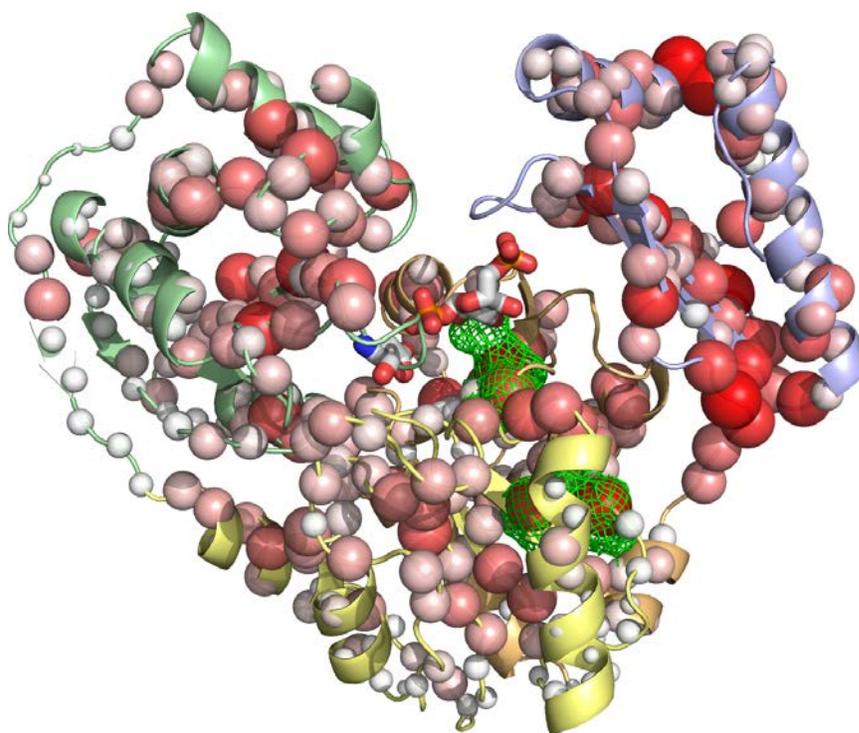


Figure SIII-9. Putative alternative binding site of S108C+G16P, residues highlighted by green mesh show unique high CSP values in Figure SIII-5A (PDB ID: 2FKM).

**CHAPTER IV**  
**Binding Isotherms and Time Courses**  
**Readily from Magnetic Resonance**

Reprinted with permission from Jia Xu and Steven R. Van Doren, *Anal. Chem.* 2016, 88 (16), pp 8172-8178 Copyright @ 2016 American Chemical Society

**IV.1 Abstract**

Evidence is presented that binding isotherms, simple or biphasic, can be extracted directly from noninterpreted, complex 2D NMR spectra using principal component analysis (PCA) to reveal the largest trend(s) across the series. This approach renders peak picking unnecessary for tracking population changes. In 1:1 binding, the first principal component captures the binding isotherm from NMR-detected titrations in fast, slow, and even intermediate and mixed exchange regimes, as illustrated for phospholigand associations with proteins. Although the sigmoidal shifts and line broadening of intermediate exchange distorts binding isotherms constructed conventionally, applying PCA directly to these spectra along with Pareto scaling overcomes the distortion. Applying PCA to time-domain NMR data also yields binding isotherms from titrations in fast or slow exchange. The algorithm readily extracts from magnetic resonance imaging movie time courses such as breathing and heart rate in chest imaging. Similarly, two-step binding processes detected by NMR are easily captured by principal components 1 and 2. PCA obviates the customary focus on specific peaks or regions of images. Applying it directly to a series of complex data will easily delineate binding isotherms, equilibrium shifts, and time courses of reactions or fluctuations.

## IV.2 Introduction

Affinity measurements are essential in understanding molecular recognition and in assessing drug discovery. Time courses of chemical and biological transformations are of wide interest. A theme shared in monitoring either equilibria or kinetics is to describe the shifts in population, the central interest of this Article. We propose to marshal a classic method of chemometrics to follow such shifts more generally.

In the case of ligand associations, a preferred spectral approach has been heteronuclear NMR, due to its information on binding site and suitability over a range of affinities.(1-5) Typically, the ligand-binding equilibrium is monitored by shifts of NMR peaks.(1, 2, 4) Arriving at affinities, however, has meant traveling through slow bottlenecks of spectral peak picking to obtain binding isotherms, usually assignment of the peaks, and global fitting of a binding isotherm consistent with the shifts of multiple peaks of the protein or macromolecule.(6) Despite the advantages of this approach and rapidity of modern collection of spectra,(7, 8) the time invested in interpreting these spectra is a barrier to wider and faster applications. Below, we propose an improved strategy that bypasses the selection of favorable peaks in spectra and favorable features in images for analysis.

The stepwise population changes due to ligand binding in a titration are usually accompanied by changes in NMR peaks that depend on the exchange regime, i.e., the time scale of chemical exchange relative to the chemical shift differences between free and bound states. Behaviors of fast, slow, and intermediate exchange regimes are depicted in Fig. SIV-1. Peak shifts in the fast exchange regime are favored for modeling binding isotherms.(4, 9) In the slow exchange regime, peaks representing the free state

can disappear and reappear elsewhere in the bound state, complicating peak assignments. In intermediate exchange, the nonlinearity of chemical shift changes from titrations can corrupt binding isotherms with sigmoidal distortion, resulting in skewed and unreliable fits of the association(4) (Fig. SIV-1).

Principal component analysis (PCA) reduces the dimensionality of data to reveal a simpler set of shared features or patterns. It is efficient, robust, and widely applied in chemometrics, analytical spectroscopy, and imaging.(10, 11) PCA is often implemented using singular value decomposition (SVD). The approach has only occasionally been applied to reactions monitored by 2D NMR spectra.(12-16) These included resolution of time-dependent(12) or pH-dependent components (using CS-PCA).(13) PCA filtered noise out of spectra to improve global fits of binding.(15) SVD of peak heights from in-cell NMR spectra of proteins associating suggested the binding site.(16) The SVD of these NMR studies was applied to peak pick lists,(13-16) rather than to the stack of 2D NMR spectra “unfolded” into a stack of vectors, which avoided peak lists and worked well on sparse 2D NMR spectra.(12) In NMR-detected titrations, the applicability of PCA is regarded at this writing as limited to the fast exchange regime.(14, 17, 18) The need for wide applicability to complex scenarios such as binding of multiple ligands, mixtures of chemical exchange regimes, and changing linewidths was articulated.(14) The work herein responds to this need.

PCA can be computed by either SVD or eigenvector decomposition of covariance, aiming at maximization of variance with minimization of correlation and redundancy (see the Supporting Information for more detail). PCA computes new orthogonal components that are linear combinations of the original experimental variables, with the first principal

component (PC1) reporting the largest variance. Jolliffe asserts that PCA is often useful for data deviating from Gaussian distributions and linear relationships of observed variables to underlying components.(19)

Magnetic resonance imaging (MRI) of brain and diseased tissues presents opportunities for chemometrics, such as comparing and registering images spatially, temporally, and metabolically.(20-24) Resolution of trends of change between the frames of a stack of congruent images or 2D spectra can be undertaken by three-way multiple image analysis such as “unfold”-PCA, which simplifies the 3D stack into two dimensions for standard PCA.(12, 25)

We demonstrate how to extend unfold-PCA to extract binding isotherms successfully from 2D NMR spectra of ligand titrations in slow exchange and problematic intermediate exchange by introducing preprocessing steps. Moreover, the improved approach needs no peak picking or peak assignments. The algorithm is even successful in deriving binding isotherms from the unprocessed free induction decays (FIDs) from titrations in fast or slow exchange. When a second binding process has been detected spectrally, PCA can also derive it as the second component of the reaction. Likewise, this enhancement of unfold-PCA is general enough to extract multiple and periodic time-varying components from MRI movies. Applying PCA directly to a series of spectra or images saves much time in handling them and in resolving the processes present.

## **IV.3 Experimental section**

### **IV.3.1 Preprocessing of spectra and images for SVD**

Each spectrum or image in the series of measurements is collected and processed under identical conditions, except for the experimental variable changed (concentration,

time, pH, etc.). Each 2D spectrum or image ( $F1 \times F2$  points) is rearranged as a 1D vector arrayed over the experimental variable (25) (Fig. SIV-2). Each vector is compressed, by deleting unchanging positions, in order to expedite computational manipulations of the matrix  $X'$ . Low intensity regions of the vectorized spectra were usually filtered out prior to SVD. Alternative choices of *no scaling*, *autoscaling*, and *Pareto scaling* (26) of the rows of  $X'$  were compared. The rows were mean-centered.(11)

### IV.3.2 Extraction of Principle Components

SVD of  $X'$  can be expressed as:

$$X'_{mn} = U_{mn} S_{nn} V_{nn}^T \quad \text{Equation IV-1}$$

where  $U$  and  $V^T$  are orthogonal matrices,  $S$  is a diagonal matrix, and subscripts denote sizes of matrices. The eigenvectors of  $X'^T \cdot X'$  constitute the matrix  $V^T$  containing the singular vectors of interest, such as PC1 as the first row with the largest trend (Fig. SIV-2) and PC2 as the second row with the second largest trend. PC1 may depend on time,(27) [ligand],(15) or other conditions.(13) The simulations of NMR spectra used for part of the testing PCA applied directly to them are described in the Supporting Information.

## IV.4 Results and discussion

### IV.4.1 PCA capture of time courses.

We extended the unfold-PCA strategy of converting a 3D stack of 2D NMR spectra (perturbed by the experimental variable) into a 2D array of vectors for SVD.(12) To improve performance, we inserted preprocessing steps for data compression, noise filtration, and scaling options (Fig. SIV-2). We automated these processing and

calculation procedures for multiple data formats.(28) This algorithm avoids user selection of features in the data (Fig. SIV-2). Its ability to capture main trends is introduced using time-lapse images of a sunset or multiplying bacteria (Fig. SIV-3). The trajectory of the setting sun is marked by PC1 (Fig. SIV-3A,B). The exponential growth in bacteria is represented by PC1, despite their motility (Fig. SIV-3C,D). Applying the same PCA approach to time-lapse 2D NMR spectra captures a reaction progress curve as PC1. Changes in  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra have been used to track dephosphorylation or phosphorylation rates.(29, 30) PCA applied directly to time-lapse TROSY spectra of a phosphoryl transfer enzyme reveals the time course of dephosphorylation (Fig. SIV-3E,F). The kinetics derived from unsupervised PCA of entire spectra echo those obtained from global fitting of carefully selected peak height changes (29) but with new ease.

#### **IV.4.2 Fast exchange scenarios.**

PCA was demonstrated on peak pick lists of titrations with NMR peaks in the fast exchange regime, where the shifts of the peak positions are linear combinations of the basis spectra and suffice to indicate population change.(13, 14, 16) However, applying PCA directly to noninterpreted spectra means that more information is considered: not only selected peak positions but also line shapes (widths, heights, volumes, etc.) throughout the spectrum. *Autoscaling*(32) and Pareto scaling(26) perform acceptably when applying the improved algorithm to fast exchange (Fig. SIV-4A,B). *Autoscaling* is, however, more accurate and precise for fast exchange, especially with the threshold for retention of spectral points set to 3- to 7-fold the noise level (Fig. SIV-4A,B).

The list-based and improved spectrum-based implementations of PCA reproduce conventional results in obtaining binding isotherms. An example of 1:1 protein–ligand binding in the fast exchange regime with  $K_D$  set to 270  $\mu\text{M}$  is shown with the simulated

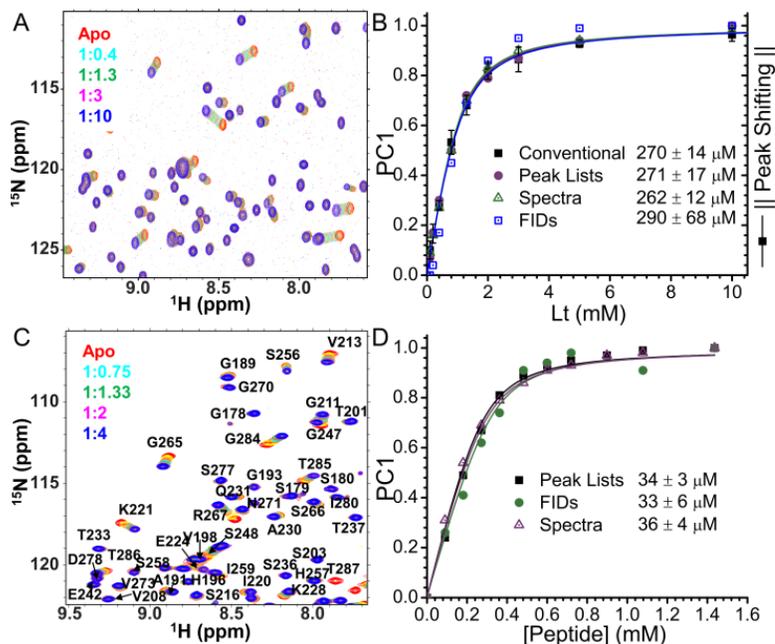


Figure IV-1. PC1 from SVD of titrations in fast exchange, simulated or measured, represents Langmuir binding isotherms. (A) Simulated spectral shifts in the fast exchange regime. The colors of the contours progress with ligand additions up to 10-fold excess. (B) Binding isotherms were obtained by applying SVD to the simulated spectra without peak picking (triangles), peak pick lists (circles), or the simulated raw FIDs (open squares). Black squares mark conventional, global fitting of the shifts of individual peaks.  $\|\cdot\|$  denotes normalization of the peak shifts. (C) Superposed  $^{15}\text{N}$  HSQC spectra of a phosphoprotein-binding FHA domain (600  $\mu\text{M}$ ) titrated with a phosphopeptide from a protein kinase exhibit fast exchange behavior.(31)(D) Binding isotherms were derived from the titration shown in (C) by applying SVD directly to the spectra (open triangles), lists of the peaks of each spectrum (squares), or FIDs (circles). The  $K_D$  of  $40 \pm 5 \mu\text{M}$  globally fitted to the peak shifts of multiple amide peaks(31) is closest to the  $K_D$  fitted to PC1 of the spectra.

titration of Fig. IV-1A. Application of PCA to lists of all peaks provides an accurate binding isotherm as PC1 plotted vs [ligand]. Fitting to standard Eq. IV-S4 places  $K_D$  at  $271 \pm 17 \mu\text{M}$  (Fig. IV-1B). This indicates that PCA of all peak positions, whether shifted by the ligand or not, matches conventional global fitting of only the big shifts of well-resolved peaks. It is more convenient and thorough to apply the improved unfold-PCA

algorithm directly to the spectra (Fig. SIV-2). The binding isotherm captured as PC1 in this way reproduces the true populations (Fig. IV-1B). This is also illustrated for the titration of a phosphoprotein binding domain with a phosphoThr peptide in fast exchange<sup>(31)</sup> (Fig. IV-1C). PC1 direct from the spectra delineates the binding isotherm fitted by  $K_D$  of  $36 \pm 4 \mu\text{M}$  (Fig. IV-1D), which closely resembles the binding isotherms and  $K_D$  of  $40 \pm 5 \mu\text{M}$  globally fitted previously to the shifts of multiple amide peaks.<sup>(31)</sup> PCA of lists of the spectral peaks picked from the titration provides PC1 fitted by a similar  $K_D$  of  $34 \pm 3 \mu\text{M}$  (Fig. IV-1D).

Parseval's theorem suggests that signals in time and frequency domains can be considered equivalent.<sup>(33)</sup> With this in mind, PCA of the unprocessed FIDs was also evaluated (Fig. IV-1). PC1 derived from the array of FIDs from the simulation of fast exchange managed to obtain a binding isotherm with nearly correct affinity but larger uncertainty, i.e.,  $K_D$  of  $290 \pm 68 \mu\text{M}$  (Fig. IV-1B). This outcome is promising for PCA overcoming the high level of noise added to the simulated example (S/N of 5 at the median peak height). PCA of the sets of FIDs from the protein titration with phosphoThr peptide in fast exchange<sup>(31)</sup> generated a binding isotherm with  $K_D$  close to the  $33 \pm 6 \mu\text{M}$  obtained by other methods (Fig. IV-1D). The smaller uncertainties when applying PCA after Fourier transformation might reflect increased sensitivity from integration of the signals or from better signal resolution.

#### **IV.4.3 Slow exchange scenarios.**

Binding isotherms can be constructed conventionally in the slow exchange regime (with slower  $k_{off}$  and higher affinities) from changes of peak volumes or heights but with more difficulty and rarity. Tracking the appearance of bound state peaks is preferred (4)

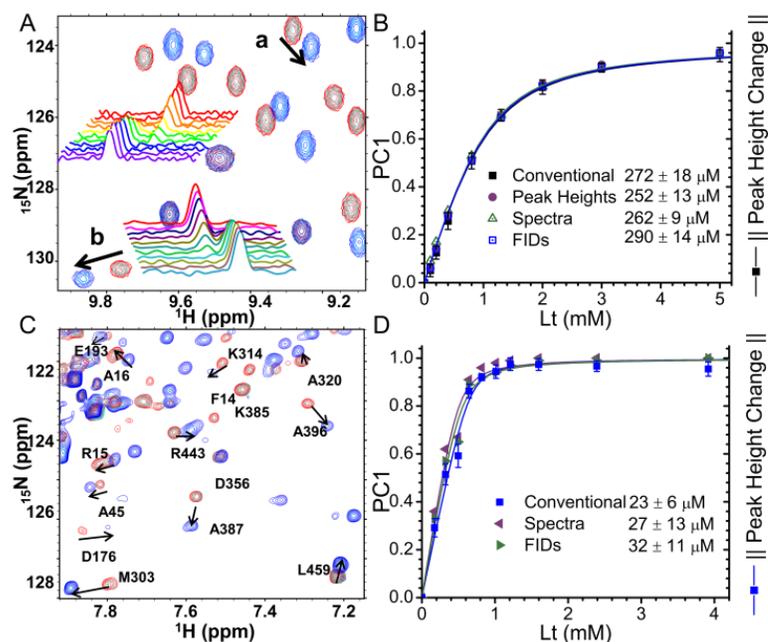


Figure IV-2. SVD of titrations featuring slow exchange, in simulated or measured NMR spectra, distills binding isotherms as PC1. (A) Overlay of HSQC spectra simulated with slow exchange. Protein ligand ratios of 1:0, 1:1.3, and 1:10 are represented by red, cyan, and darker blue, respectively. Insets are 1D slices of peak pairs indicated by black arrows. (B) PC1 derived from the simulated series of spectra (triangles) in panel A provides binding isotherms equivalent to plotting heights of disappearing peaks of the free state (black squares). PC1 was also calculated from peak lists (circles) or the FIDs (open squares). (C) Spectra from a slow exchange titration of an enzyme with an inhibitor.  $^{15}\text{N}$  TROSY spectra of PMM (52 kDa, 800 MHz, 25 °C) titrated with X1P are superposed and contain amide peaks in slow exchange. PMM/X1P ratios of 1:0, 1:0.6, and 1:8 are represented by red, cyan, and blue, respectively. (D) PC1 of either the spectra or FIDs from this titration captures the binding isotherm. Standard global fitting of peak heights is shown with blue symbols for comparison.

but can be complicated by challenging peak assignments and peak attenuation by line broadening. PCA of the simulated titration ( $K_D$  set at 270  $\mu\text{M}$ ) in the slow exchange regime derives a binding isotherm as PC1 that is virtually indistinguishable ( $K_D$  of  $262 \pm 9 \mu\text{M}$ ) from the simulated populations (Fig. IV-2B). SVD of the series of spectra derives robust binding isotherms from titrations in slow exchange. The fits to them are precise with all three options of scaling, provided that with autoscaling the threshold for data inclusion is kept  $\leq 7$ -fold the noise level (Fig. SIV-4E,F). PC1 extracted from simulated FIDs provides a binding isotherm resembling the simulated populations, with slight

deviations in points and fitted  $K_D$  of  $290 \pm 14 \mu\text{M}$  (Fig. IV-2B). PCA was applied to the entirety of crowded  $^{15}\text{N}$  TROSY spectra of the 52 kDa PMM enzyme titrated by its inhibitor xylose 1-phosphate (X1P), exhibiting slow exchange behavior (Fig. IV-2C). The binding isotherm globally fitted to the increasing peak heights of several selected bound state peaks estimates  $K_D$  at  $23 \pm 6 \mu\text{M}$ . (The blue curve in Fig. IV-2D summarizes many normalized peak heights fitted.) The points of PC1 obtained directly from the spectra are fitted by  $K_D$  of  $27 \pm 13 \mu\text{M}$  and PC1 from FIDs by  $K_D$  of  $32 \pm 11 \mu\text{M}$  (Fig. IV-2D). These PC1-derived binding isotherms match well those obtained from conventional global fitting of bound peak heights but with the advantages of minimal data handling or interpretation.

#### **IV.4.4 Intermediate Exchange Scenarios.**

Intermediate exchange is most problematic for estimating affinities due to its sigmoidal plots of NMR peak shifts (4) vs [ligand] (Fig. SIV-1F and IV-3B). These nonlinear shifts can be fitted erroneously with deviations up to 2 orders of magnitude from actual.(4) It can also be misconstrued as evidence of cooperativity.

In intermediate exchange, both line shapes and peak positions appear to be critical for capturing population change. As a simple and extreme case, NMR spectra of a titration were simulated with intermediate exchange broadening in all peaks in the 1H dimension. The application of standard *autoscaling* (32) in the algorithm of Fig. SIV-2 falls short of the accuracy and precision needed (see purple box in Fig. SIV-4C,D). For obtaining a binding isotherm of high accuracy and precision from intermediate exchange behavior, *Pareto scaling* of the rows is required and improved by the threshold remaining

small (Fig. SIV-4C,D). Though the shifts of all peaks are sigmoidal (Fig. IV-3A,B), PCA of the Pareto-scaled, linearized spectra avoids any such distortion of PC1; it is best fitted

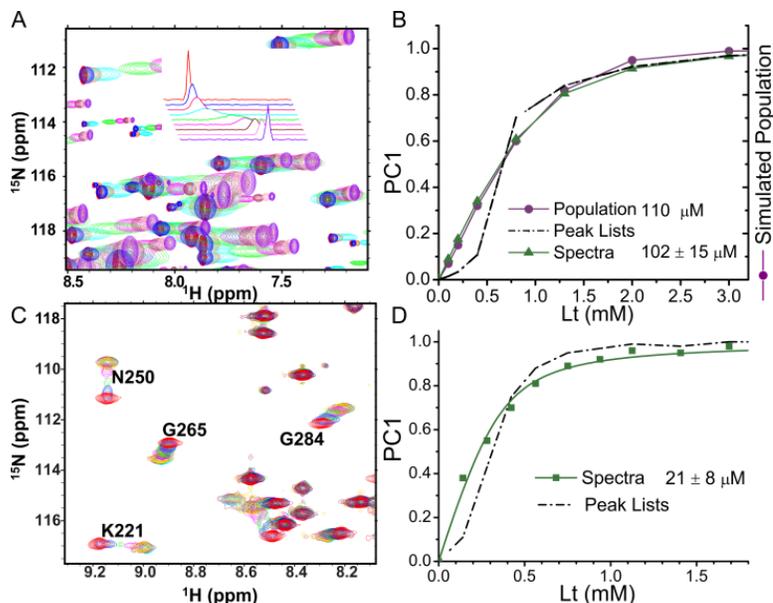


Figure IV-3. Suppressing the intermediate exchange distortion of binding isotherms by applying PCA directly to spectra. (A) HSQC spectra simulated to be intermediate to fast in exchange for  $^1\text{H}$  chemical shift changes and line shapes. The inset shows slices through a shifted and broadened peak. (B) In intermediate to fast exchange, the ligand-induced peak shifts deviate sigmoidally from a 1:1 binding isotherm when applying PCA to the peak pick lists (dashed line). The lag is suppressed in PC1 (green triangles) from SVD of Pareto-scaled spectra. (C) A region of the  $^{15}\text{N}$  HSQC spectrum of the FHA domain titrated with a phosphopeptide displays intermediate-fast exchange behavior at the peaks of four amino acids labeled. (D) PC1 of the spectra yields a binding isotherm fitted by  $K_D$  of  $21 \pm 8 \mu\text{M}$ , which agrees with the  $K_D$  of  $20 \mu\text{M}$  measured by isothermal titration calorimetry.(31)

by a  $K_D$  of  $102 \pm 15 \mu\text{M}$  that agrees with the simulated  $K_D$  (Fig. IV-3B). Pareto scaling with a low threshold increases the weighting of weak peaks broadened by intermediate exchange and appears to move the data closer to a Gaussian (Fig. SIV-6), the distribution optimal for PCA.(19)

#### IV.4.5 Mixtures of regimes.

It is much more typical of titrations with NMR peaks in intermediate exchange to be accompanied by other peaks in fast or slow exchange. We simulated a titration with a

mixture of all three regimes and 34% of the peaks in intermediate exchange (Fig. SIV-8A). The sigmoidal shifts of the latter are enough to cause PCA of the lists of all picked peaks to extract PC1 which is sigmoidal and unacceptable as a binding isotherm (Fig. SIV-8B). The application of PCA to these *spectra* instead (with Pareto scaling for accuracy) successfully captures the simulated population change as PC1 with fitted  $K_D$  within 7% of the simulated value (Fig. SIV-8B). When using only peaks in intermediate exchange from this simulation (Fig. SIV-8C), the sigmoidal distortion of PC1 from PCA of peak lists worsens, but PCA of the Pareto-scaled spectra still suppresses distortion of PC1, as is evident from fitted  $K_D$  within 13% of the actual value (Fig. SIV-8D).

$^{15}\text{N}$  HSQC spectra of an FHA domain titrated with a phosphoThr peptide(31) exhibit intermediate-fast exchange (Fig. IV-3C). Though numerous unaffected peaks are also present, fitting of the PC1-derived binding isotherm matches the  $K_D$  of  $20 \pm 3 \mu\text{M}$  measured independently by isothermal titration calorimetry (Fig. IV-3D). PCA is not recommended for application to FIDs with intermediate exchange broadening because of the skewing of PC1 that results (Fig. SIV-9E,F).

Applying unfold-PCA to spectra along with the preprocessing recommended herein (Fig. SIV-2 and IV-S4) reliably defines the binding isotherms. This is much easier than seeking  $K_D$  through fitting of line shapes or competition experiments(4) requiring prior knowledge of relative ligand affinity. Use of PCA does not change the need for [protein] to be 0.2 to 0.8 of  $K_D$  for best accuracy in fitting  $K_D$  and within 10-fold for acceptable accuracy.(5, 9) When affinities are too tight to use this range (evident as an

abrupt transition), competition can then be introduced to weaken the affinity of interest into the concentration range where it can be fitted accurately.(4, 5, 15)

#### IV.4.6 Two-Step Binding.

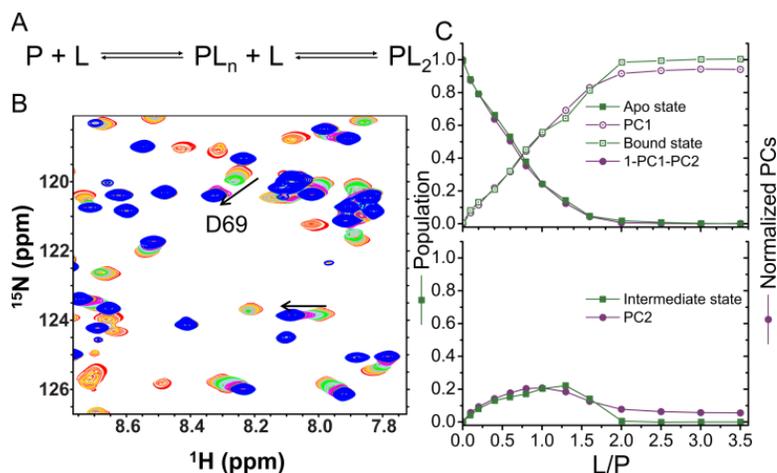


Figure IV-4. Principal components from SVD of spectra agree with the populations estimated earlier by line shape analysis (34) for a titration of two sequential binding events. (A) Scheme of the two-step binding mechanism hypothesized. (B) Chicken liver bile acid binding protein with disulfide bridge was titrated with GCDA and underwent intermediate exchange broadening, as is evident for two peaks marked with arrows in the superposed HSQC spectra.(34) (B) HSQC spectra of this protein titrated with GCDA, specifically ligand/protein ratios of 0, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 1.3, 1.6, 2.0, 2.5, 3.0, and 3.5, with contours ranging from red to blue. Black arrows indicate peaks in intermediate exchange.(34) (C) Comparison between normalized PCs (purple) and populations of the states P, PL, and PL<sub>2</sub> previously calculated using line shape analysis (green, adapted from Figure 3e in ref 34 with permission, copyright 2010 John Wiley & Sons).

Next, we attempted resolution of two binding events, reactions determined to be sequential.(34)In the course of multiple ligand binding, mixed exchange regimes are likely to complicate previous strategies of analysis. Cogliati et al. reported a challenging mixture of exchange regimes in the two-step binding of two molecules of sodium glycochenodeoxycholate (GCDA) to bile acid binding protein(34) (Fig. IV-4A). The titrations display a mixture of fast, slow, and intermediate exchange regimes accompanying the complex binding (Fig. IV-4B). The authors exploited line shape analysis to selected amide NMR peaks undergoing intermediate exchange broadening;

see those marked with black arrows in Fig. IV-4B.(34) This enabled them to estimate the proportions of the apo (P), intermediate (PL), and ligand-saturated (PL<sub>2</sub>) states through the course of titrations(34) (green in Fig. IV-4C).

The application of SVD directly to the same spectra without peak picking and with Pareto scaling results in PC1 accounting for 61% of the variances and PC2 accounting for 12% (Table IV-S2). PC1 approximates the disappearance of the apo state P. The quantity  $1 - PC1$  (not shown) resembles but slightly exceeds the formation of the fully bound state PL<sub>2</sub> (Fig. IV-4C). PC2 resembles the rise and fall of the population of the singly ligated intermediate PL, once PC2 is normalized to the scale of PC1 (Fig. IV-4C). Since the population changes of P and PL<sub>2</sub> are highly correlated ( $R = -0.93$ ) and hence statistically related, it is mathematically unrealistic to distinguish these two correlated components by PCA, a decorrelation technique.

When no ligand is present ( $L/P = 0$ ) or the bile acid binding protein is saturated with the GCDA ligand (e.g.,  $L/P = 3.5$ ), PC1 and PC2 sum to 1.0 in agreement with the proportions of PL and PL<sub>2</sub> summing to 1.0. Consequently, the sum of PC1 and PC2 is renormalized to 1.0. This implies that PL<sub>2</sub> should be modeled by  $1 - PC1 - PC2$ , which matches well the fractional concentrations of PL<sub>2</sub> estimated previously (34) (Fig. IV-4C).

#### **IV.4.7 Nonlinearity and Applicability of PCA.**

Are the nonlinear peak shifts of the peaks in intermediate exchange (see Figs. IV-3, IV-4, and SIV-8) suitable for PCA? Neither SVD nor covariance calculations require Gaussian distributions.(19)The series of NMR spectra and time-lapse images analyzed in this study all have a degree of the nonlinear character (non-normal distributions) exemplified more dramatically by a chaotic system (Fig. SIV-7). This may result from

the spectra and images containing more components than lists of their peaks or features. It would require multiple PCs to capture most of the greater complexity to reconstruct the original measurements (with matrix  $U$  in Eq. IV-1). However, for this study's more modest goal of extracting the largest population shifts among the spectra or images, the nonlinearity (Fig. SIV-7) does not interfere in the largest PCs capturing the main processes. When these largest trends are abstracted from matrix  $V^T$  (Eq. IV-1), they robustly withstand nonlinearity. The central limit theorem generates an approximation of normality for most data sets, as they have the large size required by the theorem. The scaling of the data matrix of spectra appears to shift it toward a normal-like distribution (Fig. SIV-6). Thus, discovering the main trends requires far fewer PCs from matrix  $V^T$  than needed for faithful reconstruction of nonlinear spectra and images using matrix  $U$ .

#### **IV.4.8 Periodic and multiple components from MRI by PCA.**

We tested the fitness of this SVD approach for wider applications to measurements paralleling macromolecular NMR spectra in being complex and responsive to coordinated processes, e.g., MRI movies. The SVD algorithm extracts from an MRI movie of brain fluctuations (35) the periodic flow of cerebral spinal fluid as PC1 (Fig. IV-5A,B). PC1 from the full breadth of the movie frames appears similar to the reported modulation of image intensities within the box confined to the third ventricle (36) (Fig. IV-5A,B). PC1 represents the 5 cycles of respiration, each with 2.5 s of inspiration and 2.5 s of expiration, similarly to the conventional plot of the localized intensities of the MRI signal(36) (Movie IV-S1). PC1 being smoother than the local

intensity changes may reflect the integration of more covarying data and the noise filtering that is intrinsic to PCA.

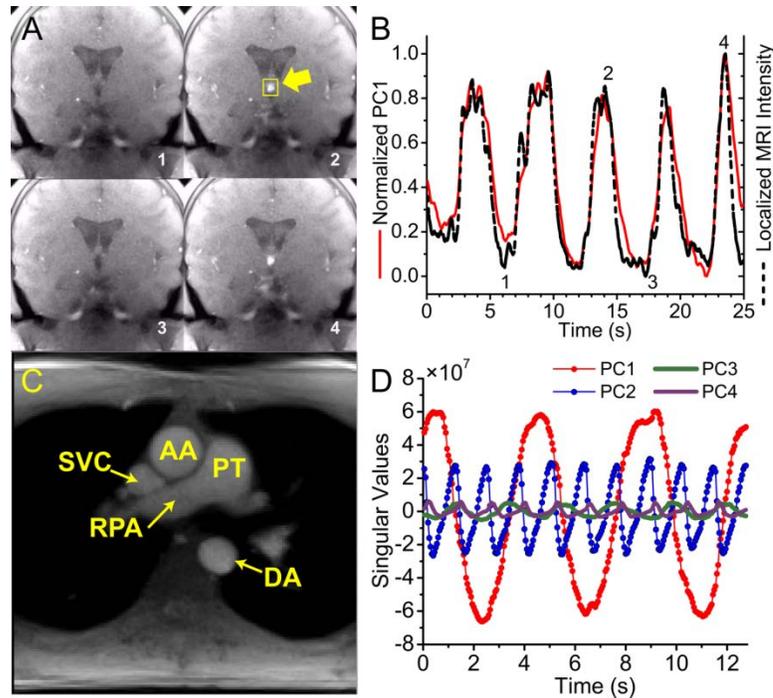


Figure IV-5. SVD extracts the time courses of pulsation in MRI movies of cross sections through the brain (35) or chest.(37) (A) Frames from the brain imaging (Movie IV-S1, adapted from ref 35 with permission, copyright BiomedNMR/CC-BY-SA-3.0) feature cerebral spinal fluid flow most apparent within the box pointed out by an arrow in frame 2.(35, 36) (B) PC1 from the movie captures five cycles of breathing, plotted with the red line. Signal intensities within the boxed central region with the arrow in the third ventricle are plotted with the black dashed line. (C) A frame from the movie of ref 37 (adapted with permission, copyright 2014 John Wiley & Sons) is labeled AA for ascending aorta, DA for descending aorta, PT for pulmonary trunk, RPA for right pulmonary artery, and SVC for superior vena cava. (D) The time courses of the four PCs generated by unsupervised SVD are plotted and suggest four types of periodic fluctuations. This movie (37) is synchronized with plotting of its PC1 and PC2 in Movie IV-S2.

We also applied this PCA approach to an MRI movie of a chest cross-section(37) through the large arteries (the aorta and pulmonary trunk) and vein (superior vena cava) each connected to the heart (Fig. IV-5C). The aorta, pulmonary trunk, and superior vena cava pulse in unison upon contraction of the heart, while chest dimensions undulate more slowly with breathing(37) (Movie IV-S2). Applying unfold-PCA to the

standard magnitude view of the MRI movie easily extracts four time courses as PC1 to PC4. PC1 represents breathing with three cycles of inspiration and expiration (red in Fig. IV-5D and Movie IV-S2). PC2 represents the pulsation of the major arteries and superior vena cava upon heart contraction for ten consecutive heart beats; the troughs mark the expansion of the vessels (blue in Fig. IV-5D and Movie IV-S2). The process represented by PC3 is unclear but is synchronized to breathing and repeats at exactly twice the frequency of PC1 and breathing. Movie reconstruction(28) using only PC3 suggests subtle fluctuations in the pulmonary trunk (not shown), which ties to the lungs. PC4 is clearly synchronized to the cardiac cycle. Movie reconstruction(28) reveals that PC4 affects the pulmonary trunk the most and the aorta slightly. The crests of PC4 (Fig. IV-5D) probably represent contraction of the heart (systole) because they are narrow and immediately precede the bolus of blood that appears in the arteries (troughs in PC2). The broad troughs of PC4 probably represent the relaxation of the heart known as diastole, with its rapid filling and subsequent slower filling phases; these are evident as the steeper and more gradual slopes at the bottom of the troughs (Fig. IV-5D). Thus, the strategy of applying PCA directly to the series of images resolves multiple concurrent processes. Two PCs are as intuitive as breathing and heart beat while another PC represents phases of the cardiac cycle.

#### **IV.4.9 Tallying meaningful principal components.**

Determining the number of meaningful PCs can become important when there are concurrent processes. Scree plots of the contributions of PCs are widely trusted and give especially clear suggestions of the significant PCs for the peak lists and movies that we analyzed. Additional strategies of counting significant PCs were proposed (e.g., singular

values and RMSD)(15, 38) but appear inconclusive in all applications of unfold-PCA to the series of spectra and images that we have examined, except to highlight the ubiquity of nonlinear behavior (Fig. SIV-7). Even for a simple titration with NMR peaks in slow to intermediate exchange, using the percentage of the variances accounted for cannot judge the adequacy of the single component (Fig. SIV-6). The criterion that a PC be smooth (high autocorrelation),(15) however, appears more reliable for recognizing a meaningful component, when coupled with some understanding of the processes. For example, in 1:1 protein–ligand binding, the hyperbolic PC1 curve represents the binding isotherm regardless of the proportion of variance contributed by PC1. This inspection of PC1 works for the slow-intermediate exchange example (Fig. SIV-6). When more than one significant component is present, the shapes of lesser PCs need to be checked.(15) In analyses of protein–ligand titrations with two reactions (see Fig. IV-4), PC1 and PC2 are smooth and clearly larger than other PCs (Fig. SIV-10).

#### **IV.4.10 Limits to applications of PCA to spectra and images.**

We have encountered instances of deterioration or failure of the improved unfold-PCA algorithm. PCs were corrupted when spectral windows, signal averaging, management of water suppression, or gain were not uniform. This is usually overcome by applying SVD to peak pick lists. SVD of unprocessed FIDs diminished by simulated intermediate exchange failed to represent the binding isotherms of the titrations (Fig. SIV-9F). This is avoided by Fourier transformation. When SVD is applied to 1D spectra of abnormally low digital resolution, the accuracy of the binding isotherm deteriorates (Fig. SIV-6). However, PCA appears remarkably reliable in representing at least two processes from a series of 2D measurements.

#### **IV.4.11 Potential applications to digital data.**

Unfold-PCA, improved by preprocessing steps described, can process many kinds of series of comparable spectra and images. It makes most sense to apply it to data that are complex but that respond to one or more concerted processes, for the purpose of finding the main trends. Macromolecular NMR and MRI provide good examples. Plotting the course of protein folding intermediates recorded by expedited NMR spectra (39) is another potential application. Potential applications may extend to other series of 2D measurements such as spectra, gels, and imaging of microarrays (40), chromatographic separations, (41) electrochemistry,(42) and chemical biology signals.(43, 44)

#### **IV.5 Conclusions**

The application of this PCA strategy (enhanced by preprocessing) to a series of spectra or MRI images offers convenience and wide applicability to characterizing concerted processes. Such applications will expand the accessibility of affinities, equilibria, kinetics, and time-evolving processes. This will include noninterpreted, unassigned, and overlapped features in spectra and movies, which may number two or more concurrent processes. For example, NMR studies will be enabled to elucidate binding isotherms masked by intermediate exchange and/or two or more concurrent processes.

#### **IV.6 Acknowledgement**

We are grateful to J. Frahm and his group for real-time MRI movies and to H. Molinari and L. Ragona for spectra of titrations of chicken bile acid binding protein. We thank Y. Fulcher and L. Beamer for discussion, Beamer for PMM, and T. Mawhinney for

synthesizing X1P. This work was supported by NSF grant MCB1409898. Spectrometer purchases were supported in part by NIH grants RR022341 toward the 800 MHz and GM57289 toward the 600 MHz system.

## IV.7 References

1. Hajduk, P. J., J. R. Huth, and S. W. Fesik. 2005. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* 48:2518-2525.
2. Shuker, S. B., P. J. Hajduk, R. P. Meadows, and S. W. Fesik. 1996. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* 274:1531-1534.
3. Shortridge, M. D., D. S. Hage, G. S. Harbison, and R. Powers. 2008. Estimating protein-ligand binding affinity using high-throughput screening by NMR. *J Comb Chem* 10:948-958.
4. Williamson, M. P. 2013. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* 73:1-16.
5. Fielding, L. 2007. NMR methods for the determination of protein-ligand dissociation constants. *Progress in Nuclear Magnetic Resonance Spectroscopy* 51:219-242.
6. Lowe, A. J., F. M. Pfeffer, and P. Thordarson. 2012. Determining binding constants from H-1 NMR titration data using global and local methods: a case study using [n]polynorbornane-based anion hosts. *Supramol Chem* 24:585-594.
7. Gal, M., P. Schanda, B. Brutscher, and L. Frydman. 2007. UltraSOFAST HMQC NMR and the repetitive acquisition of 2D protein spectra at Hz rates. *J Am Chem Soc* 129:1372-1377.
8. Amero, C., P. Schanda, M. A. Dura, I. Ayala, D. Marion, B. Franzetti, B. Brutscher, and J. Boisbouvier. 2009. Fast two-dimensional NMR spectroscopy of high molecular weight protein assemblies. *J Am Chem Soc* 131:3448-3449.
9. Markin, C., and L. Spyropoulos. 2012. Increased precision for analysis of protein-ligand dissociation constants determined from chemical shift titrations. *Journal of Biomolecular NMR* 53:125-138.
10. Trygg, J., E. Holmes, and T. Lundstedt. 2007. Chemometrics in metabonomics. *J Proteome Res* 6:469-479.
11. Adams, M. J. 2004. *Chemometrics in Analytical Spectroscopy*. Royal Society of Chemistry, Cambridge.
12. Jaumot, J., V. Marchan, R. Gargallo, A. Grandas, and R. Tauler. 2004. Multivariate curve resolution applied to the analysis and resolution of two-dimensional [1H,15N] NMR reaction spectra. *Anal Chem* 76:7094-7101.
13. Sakurai, K., and Y. Goto. 2007. Principal component analysis of the pH-dependent conformational transitions of bovine beta-lactoglobulin monitored by heteronuclear NMR. *Proc Natl Acad Sci U S A* 104:15346-15351.
14. Konuma, T., Y. H. Lee, Y. Goto, and K. Sakurai. 2013. Principal component analysis of chemical shift perturbation data of a multiple-ligand-binding system for elucidation of respective binding mechanism. *Proteins* 81:107-118.

15. Arai, M., J. C. Ferreon, and P. E. Wright. 2012. Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J Am Chem Soc* 134:3792-3803.
16. Majumder, S., C. M. DeMott, D. S. Burz, and A. Shekhtman. 2014. Using singular value decomposition to characterize protein-protein interactions by in-cell NMR spectroscopy. *Chembiochem* 15:929-933.
17. Furukawa, A., T. Konuma, S. Yanaka, and K. Sugase. 2016. Quantitative analysis of protein–ligand interactions by NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* 96:47-57.
18. Selvaratnam, R., S. Chowdhury, B. VanSchouwen, and G. Melacini. 2011. Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc Natl Acad Sci U S A* 108:6133-6138.
19. Jolliffe, I. T. 2002. *Principal Component Analysis*. Springer-Verlag, New York.
20. Witjes, H., A. W. Simonetti, and L. Buydens. 2001. Peer Reviewed: Better Brain Imaging with Chemometrics. *Analytical Chemistry* 73:548 A-556 A.
21. Nika, V., P. Babyn, and H. Zhu. 2014. Change detection of medical images using dictionary learning techniques and principal component analysis. *J Med Imaging (Bellingham)* 1:024502.
22. Wu, J., G. Gong, Y. Cui, and R. Li. 2016. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. *Journal of magnetic resonance imaging : JMRI*:n/a-n/a.
23. Kawaguchi, H., H. Shimada, F. Kodaka, M. Suzuki, H. Shinotoh, S. Hirano, J. Kershaw, Y. Inoue, M. Nakamura, T. Sasai, M. Kobayashi, T. Suhara, and H. Ito. 2016. Principal Component Analysis of Multimodal Neuromelanin MRI and Dopamine Transporter PET Data Provides a Specific Metric for the Nigral Dopaminergic Neuronal Density. *PLoS One* 11:e0151191.
24. Huizinga, W., D. H. J. Poot, J. M. Guyader, R. Klaassen, B. F. Coolen, M. van Kranenburg, R. J. M. van Geuns, A. Uitterdijk, M. Polfliet, J. Vandemeulebroucke, A. Leemans, W. J. Niessen, and S. Klein. 2016. PCA-based groupwise image registration for quantitative MRI. *Medical Image Analysis* 29:65-78.
25. Huang, J., H. Wium, K. B. Qvist, and K. H. Esbensen. 2003. Multi-way methods in image analysis-relationships and applications. *Chemometrics and Intelligent Laboratory Systems* 66:141-158.
26. van den Berg, R. A., H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde, and M. J. van der Werf. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics* 7:142.
27. Casanovas, O., M. Jaumot, A. B. Paules, N. Agell, and O. Bachs. 2004. P38SAPK2 phosphorylates cyclin D3 at Thr-283 and targets it for proteasomal degradation. *Oncogene* 23:7537-7544.
28. Xu, J., and S. R. Van Doren. 2017. TRacking Equilibrium and Non-equilibrium Shifts in Data with TREND, *Biophys J*, <http://dx.doi.org/10.1016/j.bpj.2016.12.018>

29. Xu, J., Y. Lee, L. J. Beamer, and S. R. Van Doren. 2015. Phosphorylation in the catalytic cleft stabilizes and attracts domains of a phosphohexomutase. *Biophys J* 108:325-337.
30. Mayzel, M., J. Rosenlow, L. Isaksson, and V. Y. Orekhov. 2014. Time-resolved multidimensional NMR with non-uniform sampling. *J Biomol NMR* 58:129-139.
31. Ding, Z., H. Wang, X. Liang, E. R. Morris, F. Gallazzi, S. Pandit, J. Skolnick, J. C. Walker, and S. R. Van Doren. 2007. Phosphoprotein and phosphopeptide interactions with the FHA domain from Arabidopsis kinase-associated protein phosphatase. *Biochemistry* 46:2684-2696.
32. Noda, I., and Y. Ozaki. 2004. *Two-dimensional Correlation Spectroscopy - Applications in Vibrational and Optical Spectroscopy*. Wiley, West Sussex, England.
33. Cavanagh, J., W. J. Fairbrother, A. G. Palmer Iii, and N. J. Skelton. 2007. Preface to the First Edition. In *Protein NMR Spectroscopy (Second Edition)*. J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, and N. J. Skelton, editors. Academic Press, Burlington. vii-x.
34. Cogliati, C., L. Ragona, M. D'Onofrio, U. Günther, S. Whittaker, C. Ludwig, S. Tomaselli, M. Assfalg, and H. Molinari. 2010. Site-Specific Investigation of the Steady-State Kinetics and Dynamics of the Multistep Binding of Bile Acid Molecules to a Lipid Carrier Protein. *Chemistry – A European Journal* 16:11300-11310.
35. Dreha-Kulaczewski, S., A. A. Joseph, K. D. Merboldt, H. C. Ludwig, J. Gartner, and J. Frahm. 2014. [https://commons.wikimedia.org/wiki/File:Dreha-Kulaczewski\\_JNeurosci\\_CSF\\_flow\\_Supplementary\\_movie1.webm](https://commons.wikimedia.org/wiki/File:Dreha-Kulaczewski_JNeurosci_CSF_flow_Supplementary_movie1.webm)
36. Dreha-Kulaczewski, S., A. A. Joseph, K. D. Merboldt, H. C. Ludwig, J. Gartner, and J. Frahm. 2015. Inspiration is the major regulator of human CSF flow. *J Neurosci* 35:2485-2491.
37. Joseph, A., J. T. Kowallick, K. D. Merboldt, D. Voit, S. Schaetz, S. Zhang, J. M. Sohns, J. Lotz, and J. Frahm. 2014. Real-time flow MRI of the aorta at a resolution of 40 msec. *Journal of magnetic resonance imaging : JMRI* 40:206-213.
38. Lee, J. M., C. K. Yoo, S. W. Choi, P. A. Vanrolleghem, and I. B. Lee. 2004. Nonlinear process monitoring using kernel principal component analysis. *Chem Eng Sci* 59:223-234.
39. Rennella, E., and B. Brutscher. 2013. Fast real-time NMR methods for characterizing short-lived molecular states. *Chemphyschem : a European journal of chemical physics and physical chemistry* 14:3059-3070.
40. Rao, A. N., C. K. Rodesch, and D. W. Grainger. 2012. Real-time fluorescent image analysis of DNA spot hybridization kinetics to assess microarray spot heterogeneity. *Anal Chem* 84:9379-9387.
41. Teisseyre, T. Z., J. Urban, N. W. Halpern-Manners, S. D. Chambers, V. S. Bajaj, F. Svec, and A. Pines. 2011. Remotely detected NMR for the characterization of flow and fast chromatographic separations using organic polymer monoliths. *Anal Chem* 83:6004-6010.
42. Britton, M. M., P. M. Bayley, P. C. Howlett, A. J. Davenport, and M. Forsyth. 2013. In Situ, Real-Time Visualization of Electrochemistry Using Magnetic Resonance Imaging. *J Phys Chem Lett* 4:3019-3023.

43. Mizukami, S., R. Takikawa, F. Sugihara, Y. Hori, H. Tochio, M. Walchli, M. Shirakawa, and K. Kikuchi. 2008. Paramagnetic relaxation-based  $^{19}\text{F}$  MRI probe to detect protease activity. *J Am Chem Soc* 130:794-795.
44. Zhu, X., X. Chi, J. Chen, L. Wang, X. Wang, Z. Chen, and J. Gao. 2015. Real-time monitoring in vivo behaviors of theranostic nanoparticles by contrast-enhanced T1 imaging. *Anal Chem* 87:8941-8948.

## IV.8 Supporting material

### IV.8.1 Supporting background

#### IV.8.1.1 Chemical Exchange Regimes.

Chemical exchange in NMR refers to interconversion between two or more states differing perceptibly in magnetic environment for a chemical equilibrium maintained during the NMR spectral acquisition. The exchange rate,  $k_{ex}$ , is given by:

$$k_{ex} = k_{on}[L] + k_{off} \quad \text{Equation IV-S1}$$

for 1:1 protein ligand binding. The chemical shift difference between two exchanging states or “sites” is denoted as  $\Delta\omega$ . The exchange regimes are defined relative to  $\Delta\omega$  (see Fig. SIV-1):

$$k_{ex} \gg \Delta\omega \quad \text{Fast exchange}$$

$$k_{ex} \approx \Delta\omega \quad \text{Intermediate exchange}$$

$$k_{ex} \ll \Delta\omega \quad \text{Slow exchange} \quad \text{Equations IV-S2}$$

In the fast exchange regime, the  $\Delta\omega_{obs}$  of progressively shifted peaks (Fig. SIV-1A) are mined by conventional NMR to construct binding isotherms (requiring concentrations suitable for estimating association or dissociation constants accurately) (1, 2). In the slow exchange regime, peaks representing the free state can disappear and reappear elsewhere as peaks of the bound state (Fig. SIV-1C). In the intermediate exchange regime, line broadening followed by sharpening can be observed as the ligand binding progresses to saturation (Fig. SIV-1E).

#### IV.8.1.2 1:1 Protein-Ligand Binding.

1:1 protein-ligand binding can be expressed as:



The fraction of the protein or host remaining free,  $p_{free}$ , is well-known (1) to be represented by:

$$p_{free} = \frac{\Delta_{obs}}{\Delta_{max}} = \frac{([P]_t + [L]_t + K_D) - \sqrt{([P]_t + [L]_t + K_D)^2 - 4[P]_t[L]_t}}{2[P]_t} \quad \text{Equation IV-S4}$$

$$p_{bound} = 1 - p_{free} \quad \text{Equation IV-S5}$$

where  $p_{bound}$  is the fraction of the protein (of total concentration  $[P]_t$ ) occupied by the ligand of total concentration  $[L]_t$ .  $K_D$  is the dissociation constant.  $\Delta_{obs}$  is the current observed change (which can be change in peak position, height, or volume) of the free protein or macromolecular host.  $\Delta_{max}$  is the maximum change reached upon saturation.

#### IV.8.1.3 Close Relationships among Covariance, SVD, and PCA.

SVD is a key method of implementing PCA. Another method is to solve the correlation matrix (of covariances) eigenvector decomposition, with the aim of maximizing variance and minimizing redundancy. Covariance is a measurement of strength of correlation between a set of variables. When measuring the relationship between two column vectors X and Y with elements  $x_1, x_2, \dots, x_N$ , and  $y_1, y_2, \dots, y_N$ , respectively, the covariance between X and Y can be calculated as:

$$cov(X, Y) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \quad \text{Equation IV-S6}$$

where  $\bar{x}$ ,  $\bar{y}$  are the respective means of X and Y. The covariance between A and B, which is a scalar value, can also be written in the form of inner product between two vectors:

$$cov(X, Y) = \frac{1}{N-1} (X - \bar{X})(Y - \bar{Y})^T \quad \text{Equation IV-S7}$$

In this study, the processed matrix  $X'$  with  $n$  spectra, whose mean value is zero, can be written as:

$$X' = \begin{bmatrix} D_{11} & D_{21} & \cdots & D_{n1} \\ D_{12} & D_{22} & \cdots & D_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ D_{1m} & D_{2m} & \cdots & D_{nm} \end{bmatrix} \quad \text{Equation IV-S8}$$

As shown in Fig. SIV-2, each column of  $X'$  stands for a vector with  $m$  data points, which are reorganized from 2D data.  $D$  stands for element of matrix  $X'$ . Note  $m \gg n$ . A new matrix  $Y$  with  $n$  row vectors can be defined by  $Y \equiv X'^T$ . The covariance of  $Y$  is the outer product with itself:

$$C_Y = \frac{1}{N-1} Y Y^T \quad \text{Equation IV-S9}$$

which is a square symmetric  $n \times n$  matrix. The  $ij^{\text{th}}$  entry in matrix  $C_Y$  is an inner product between  $i^{\text{th}}$  row vector in  $Y$  and  $j^{\text{th}}$  column vector in  $Y^T$ , which is the covariance between these two vectors (Eq. IV-S9). Since row vectors in  $Y$  are column vectors in  $X'$ , the matrix  $C_Y$  shows the covariance between all possible pairs of reorganized 2D data.  $C_Y$  is the normalized form of  $X'X'^T$ , which is used for solving  $V^T$  in SVD. Hence, the row eigenvectors of  $V^T$  (Fig. SIV-2E) are *principal components* of matrix  $X'$ .

## IV.8.2 Supporting experimental section

### IV.8.2.1 Data Scaling.

Data scaling enlarges small signals relative to large ones by dividing each point by a factor. Scaling is used in chemometric analyses of metabolomics data to adjust the variances to improve statistical pattern recognition (3, 4). We evaluated scaling of the rows of data matrix  $X$  (see Fig. SIV-2) in analyses of NMR spectra of titrations containing weak peaks (Fig. SIV-4). We compared *autoscaling*, *Pareto scaling* (3), and

*no scaling*. Since all 2D data in a time or [ligand]-dependent series are of the same type and expected to change gradually, *no scaling* should be acceptable in most conditions, e.g. real time MRI movies. *No scaling* also suffices when extracting binding isotherms and affinities from titrations in fast and slow exchange regimes (Fig. SIV-4). Important exceptions such as intermediate exchange (Fig. SIV-4) are discussed in the main text. Autoscaling treats all points or peaks as equally important but inflates the measurement error. Pareto scaling reduces the importance of large peaks but enhances the low, broader peaks (3). Given that NMR resonances in intermediate exchange with extensive broadening are unusual, *autoscaling* of rows will usually be the scaling method of choice for titrations.

#### IV.8.2.2 Construction and Evaluation of Tests Cases for PCA.

Simulation parameters for 1D and 2D NMR spectra and FIDs are given in Table SIV-1. Details are discussed in the following sections.

##### IV.8.2.2.1 1D Spectral Simulations.

1D NMR line shapes were simulated according to Kovrigin (5). In the absence of scalar coupling, the modified Bloch-McConnell equations simplify to:

$$\frac{dM^+(t)}{dt} = (i\Omega - R + K)M^+(t) \quad \text{Equation IV-S10}$$

where  $M$  stands for the transverse magnetization vector for the spins in the exchanging system.  $\Omega$  and  $R$  are diagonal matrices of Larmor frequencies and transverse relaxation rate constants, respectively.  $K$  is the matrix of rate constants of each species. Assuming the total magnetization of the system is in a steady state ( $\frac{dM^+(t)}{dt} = 0$ ), the intensities in the spectrum can be calculated:

$$I(\nu) \propto \text{Re}[[1,1](i(\Omega - \nu) - R + K)^{-1}P] \quad \text{Equation IV-S11}$$

where  $\nu$  is the spectral frequency and  $P$  is the column vector of populations of species.

FIDs were calculated by Hilbert transformation followed by inverse Fourier transformation. The simulation parameters are given in Table SIV-1.

#### IV.8.2.3 2D Spectral Simulations of 1:1 Ligand Binding:

Nmrglue (6) was used to simulate  $^{15}\text{N}$  HSQC spectra using Gaussian line shapes. Simulated peak positions, heights, linewidth, and assignments of free and fully bound states were modeled after spectra of a kinase-interacting FHA domain (7) (BMRB: 5564) when bound to the phosphoThr546 peptide from the kinase BAK1(8), the size of the chemical shift perturbations from binding of this peptide are shown in Fig. SIV-5. Free and bound populations were simulated from  $K_D$  and Eq. IV-S4. Gaussian noise was added to each spectrum to set the signal-to-noise ratio (S/N, defined as ratio of peak height to  $2\sigma$ (9)) to 5.0 for peaks of median height at the beginning of the titration. In addition, linewidths were perturbed 5% randomly among spectra in both the  $^1\text{H}$  and  $^{15}\text{N}$  dimensions.

The populations of free and bound states were used to calculate population-weighted chemical shifts in the fast regime exchange or population-weighted peak heights in the slow exchange regime. For simulating 2D spectra in intermediate exchange in the  $^1\text{H}$  dimension (Fig. IV-3), contributions of free and bound states to the peak positions, peak heights and full width at half maximum were modeled after 2 exchanging spins using the parameters listed in Table SIV-1. (In the  $^{15}\text{N}$  dimension, chemical shifts were simulated with fast exchange). To simulate mixed exchange regimes (Fig. SIV-8), 1D  $^1\text{H}$  and  $^{15}\text{N}$  NMR line shapes were simulated for each free/bound peak pair separately.

In order to maximize the number of residues in the intermediate exchange regime,  $k_D$  and  $k_{off}$  values were optimized using the Solver module of Microsoft Excel 2013. The optimized parameters (Table SIV-1) were used to simulate the spectra of Fig. SIV-8. Simulation of 2D spectra that are uniformly in intermediate exchange regime was similar, but that only peaks exhibiting intermediate exchange in both  $^1\text{H}$  and  $^{15}\text{N}$  were retained.

#### IV.8.2.4 Construction of a Nonlinear System for SVD.

The nonlinear system was constructed by embedding vectors, which were generated by Logistic map:  $X_{n+1} = aX_n(1 - X_n)$ , where  $a=4.0$  and  $n=1\ 100$  with unity delay as described in ref (10):

$$E = \begin{bmatrix} X_1 & X_2 & \cdots & X_m \\ X_2 & X_3 & \cdots & X_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ X_k & X_{k+1} & \cdots & X_{m+k} \end{bmatrix} \quad \text{Equation IV-S12}$$

The dimensions of the nonlinear matrix built are  $k \times m$ , where  $k = 1100$  and  $m = 70$ .

### IV.8.3 Supporting tables

Table SIV-1. Spectral simulation parameters <sup>a</sup>

Exchange regimes	Figure	$P_t$ mM	$\nu_{\text{Free}}$ Hz	$\nu_{\text{bound}}$ Hz	$K_D$ $\mu\text{M}$	$k_{\text{on}}$ $\text{M}^{-1}\text{s}^{-1}$	$k_{\text{off}}$ $\text{s}^{-1}$
Fast	S1A	1.0	6000	6400	300	$3.33 \times 10^9$	$1 \times 10^6$
Slow	S1C	1.0	6000	6400	300	$3.33 \times 10^4$	10
Intermediate	S1E	0.3	4800	6400	300	$1 \times 10^6$	300
Intermediate	3A	1.0	5600	6400	110	$9.09 \times 10^5$	100
Mixture of fast, intermediate, slow	S8A,E	0.3	NA <sup>b</sup>	NA <sup>b</sup>	345	$5.51 \times 10^4$	19
Intermediate	S8C	0.3	4800	6400	200	$2.25 \times 10^6$	150

<sup>a</sup> Relaxation rates for both free and bound states were set to  $25 \text{ s}^{-1}$

<sup>b</sup> Chemical shifts perturbations between apo and bound state are shown in Fig. SIV-9E.

Table SIV-2. Contributions of the first 4 principle components of all PCA results.

<b>Figure</b>	<b>Data Type</b>	<b>PCs computed</b>	<b>PC1 (%)</b>	<b>PC2 (%)</b>	<b>PC3 (%)</b>	<b>PC4 (%)</b>	<b>Sum of PC1-PC4 (%)</b>
<b>1B</b>	Peak Lists	11	96	1	1	0	98
<b>1B</b>	Spectra	11	42	15	10	7	74
<b>1B</b>	FIDs	11	12	11	10	10	43
<b>1D</b>	Peak Lists	11	96	3	1	0	100
<b>1D</b>	Spectra	11	13	11	10	9	43
<b>1D</b>	FIDs	11	16	12	10	10	48
<b>2B</b>	Peak Heights	11	71	6	5	4	86
<b>2B</b>	Spectra	11	67	9	4	4	84
<b>2B</b>	FIDs	11	15	10	9	10	44
<b>2D</b>	Spectra	12	65	9	5	3	82
<b>2D</b>	FIDs	12	22	10	8	8	48
<b>3B</b>	Spectra	9	68	14	4	3	89
<b>3D*</b>	Spectra	11	41	16	11	8	76
<b>4C</b>	Peak Lists	16	90	6	1	1	98
<b>4C</b>	Spectra	13	61	12	4	3	80
<b>5B</b>	Movie	500	17	13	3	3	36
<b>5D*</b>	Movie	255	51	21	4	3	79
<b>S3B*</b>	Movie	40	60	12	6	3	81
<b>S3D*</b>	Movie	105	23	12	9	7	51
<b>S3F*</b>	Spectra	31	11	6	3	3	23
<b>S8B</b>	Spectra	11	40	15	10	7	72
<b>S8D</b>	Spectra	11	25	13	12	12	62
<b>S7F*</b>	Matrix <sup>§</sup>	40	4	3	4	3	14

\* Cumulative contributions can be seen in Fig. SIV-7.

§ The chaotic, nonlinear data are the matrix described as Eq. IV-S12.

#### IV.8.4 Supporting figures

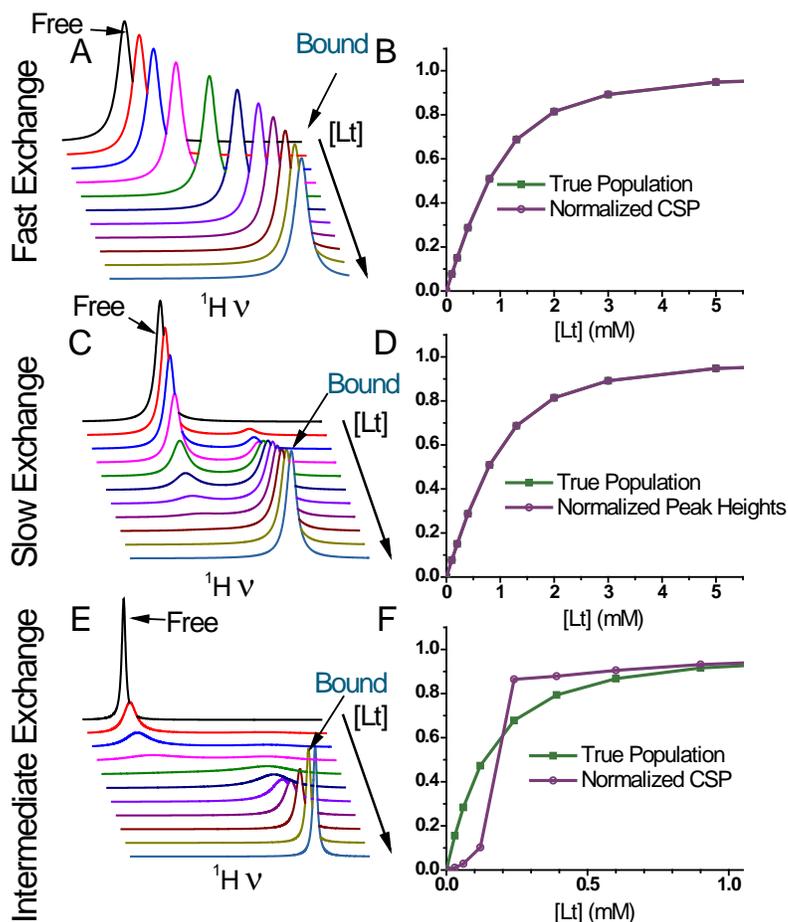


Figure SIV-1. Chemical exchange regimes of 1:1 ligand binding. A) Simulated 1D  $^1\text{H}$  line shapes in the fast exchange regime. B) In fast exchange, the observed chemical shift change is the weighted average of chemical shifts of free and bound states. Fitting these weighted shifts vs. [Ligand] (purple) reports the true population change (dark green). C) Simulated 1D  $^1\text{H}$  line shapes in slow exchange. Peak positions of the free and bound states are fixed. The free peak disappears and reappears as the bound peak. D) In the slow exchange regime, the peak shifts cannot provide the binding isotherm, whereas the changes in peak heights or volumes can. E) Simulated 1D  $^1\text{H}$  line shapes in the slow-intermediate exchange regime. F) Changeover from fast into intermediate exchange causes the CSPs to be less than expected of the weighted average of the free and bound populations. The resulted sigmoidal line is shown in purple. Simulation parameters are given in Table SIV-1.

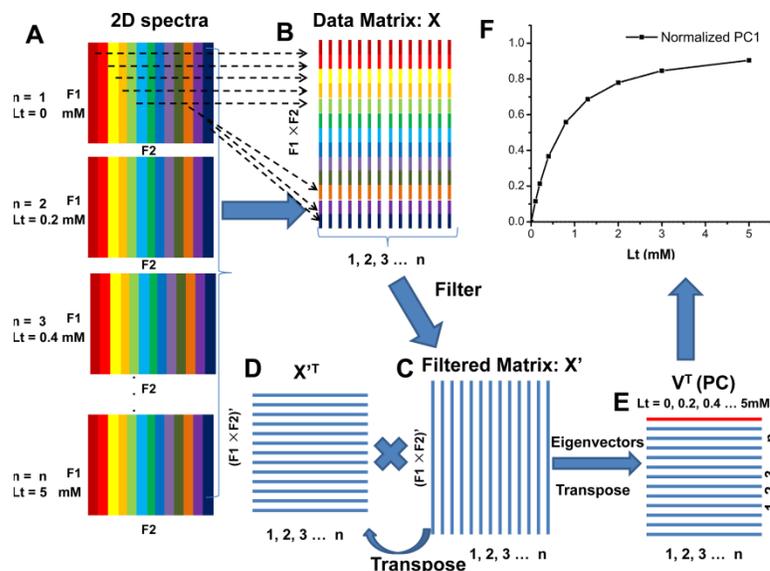


Figure SIV-2. Algorithm developed to apply SVD to series of spectra, images (movies), or lists. The series of 2D measurements is reorganized as a series of 1D columns within the data matrix used for SVD. No peak picking is employed in analyzing spectra or images. The steps shown are implemented as a Python script<sup>27</sup>. (A) A series of 2D spectra from a protein-ligand titration or frames from a movie for example, are symbolized by rainbow coloring of the columns for clarity. (B) The columns are positioned end-to-end (see dashed arrows) into a single 1D vector for each spectrum or frame in forming matrix X. Each column of the 2D measurement is symbolized by a short bar of the corresponding color in matrix X. The matrix is compressed by filtering out invariant rows. (C,D) The filtered X' matrix is transposed to X'<sup>T</sup>. (E) V<sup>T</sup> is a set of transposed orthonormal eigenvectors of the matrix X'<sup>T</sup>•X'. The first row in the matrix V<sup>T</sup> (red) is PC1, the largest trend of change. (F) In a typical titration, PC1 represents the ligand binding isotherm ready for fitting.

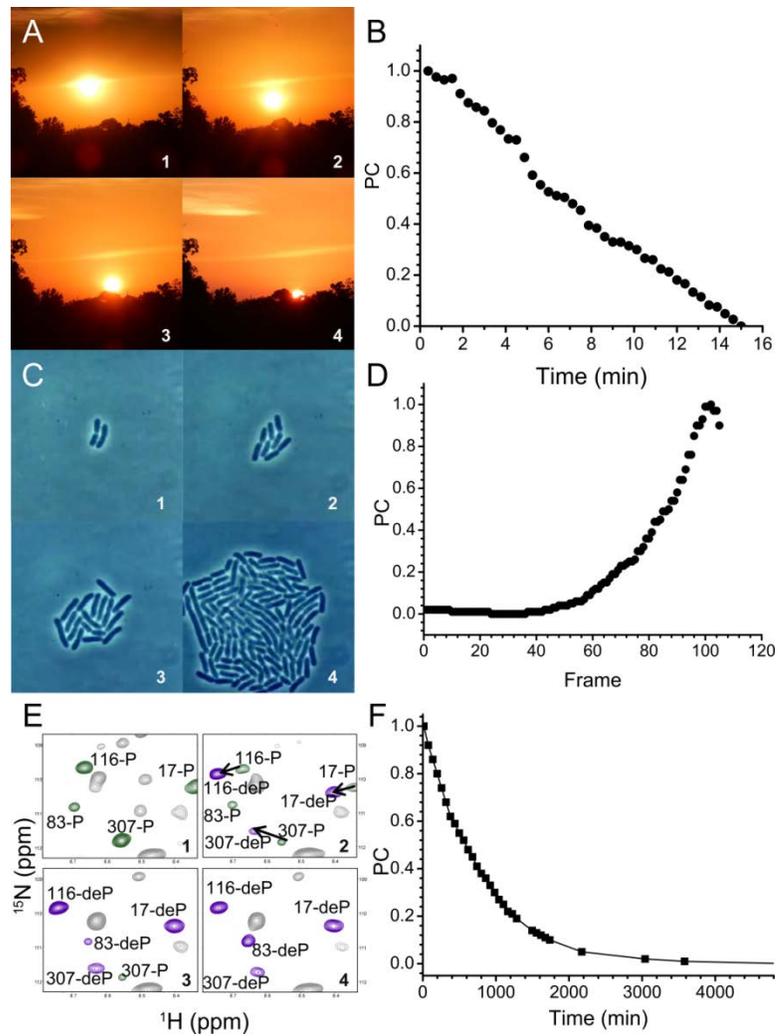


Figure SIV-3. Capture of main time courses of change by PCA of movies or NMR spectra. This is illustrated for time-lapse images of (A, B) a sunset, (C, D) microscopy of bacterial growth, and NMR spectra of (E, F) an enzyme. PC1 extracted from each series (using the algorithm of Figure S2) is shown in panels on the right. (E) NMR spectra detecting dephosphorylation of an enzyme over time color in green the peaks of the phosphorylated form, in purple the contours of the dephosphorylated form, and unshifted peaks with gray. The enzyme is 0.5 mM phosphomanno/glucomutase (PMM) monitored by  $^{15}\text{N}$  TROSY at 35 °C, pH 7.5, 10 mM DTT, and 800 MHz. (F) PCA applied to the entire set of spectra reveals the time course of dephosphorylation observed (11).

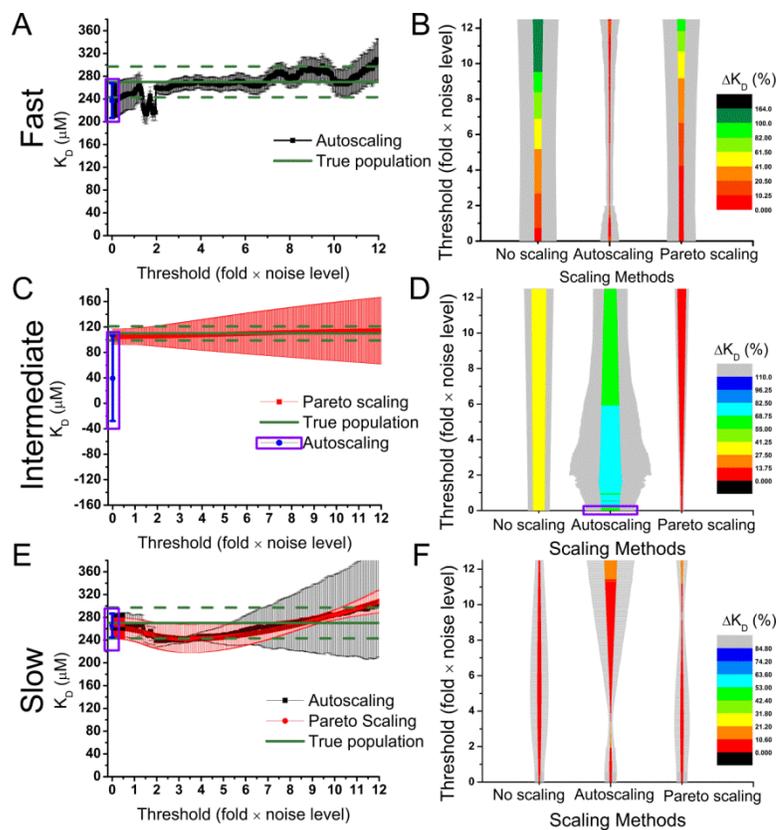


Figure SIV-4. Accuracy and precision of binding isotherms can be enhanced by preprocessing with scaling, as illustrated for chemical exchange regimes encountered in NMR spectra. The exchange regimes are uniformly: fast (A,B), intermediate (C,D), or slow (E,F). (A, C, E) Apparent  $K_D$  fitted is plotted vs. the threshold for inclusion of points in the analysis, where the threshold is figured relative to the noise amplitude. The true  $K_D$ , established by simulation, is shown by solid dark green lines, along with 10% deviations in dashed lines in panels A,C,E. (B, D, F) The relationships between apparent  $K_D$  value, data threshold for scaling, and type of scaling of the row of input matrix  $X'$  are plotted. Fitted  $K_D$  is plotted with color coded by the deviation from the true value. The width of the vertical bar is proportional to the RMSD between  $K_D$  from estimated binding isotherms and the simulated theoretical isotherms, i.e. smaller means more accurate binding isotherms. Uncertainties are plotted as the breadth in gray. The performance available from the previous unfold-PCA algorithm (12) is marked by the purple rectangles.

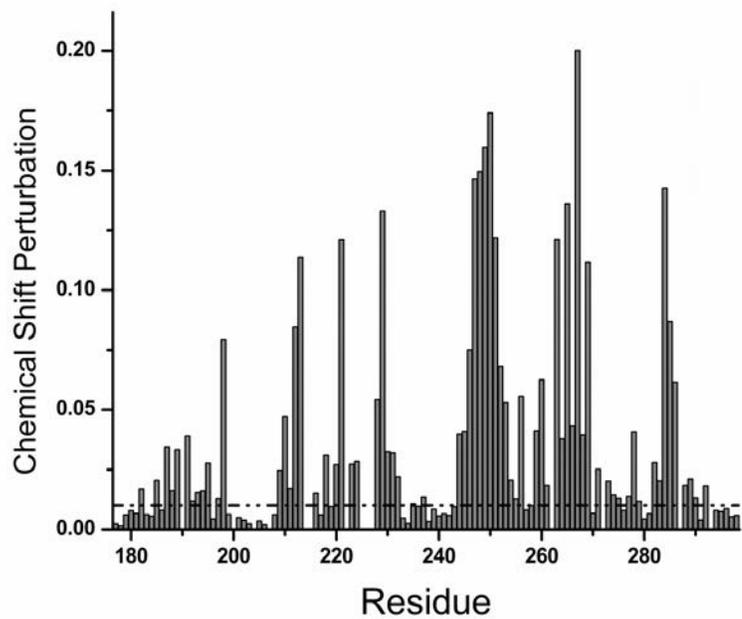


Figure SIV-5. Chemical shift changes between apo and bound states of the simulated 2D spectra.

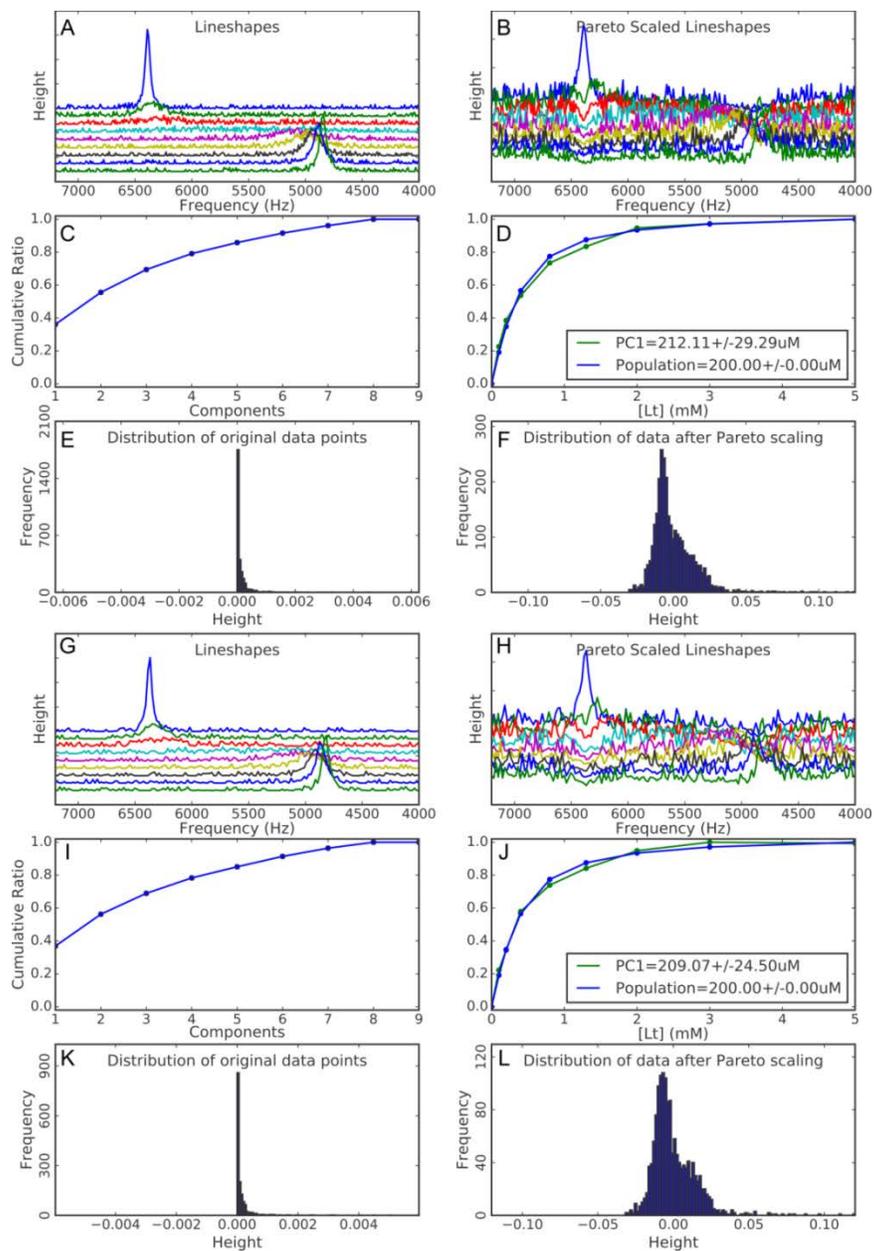
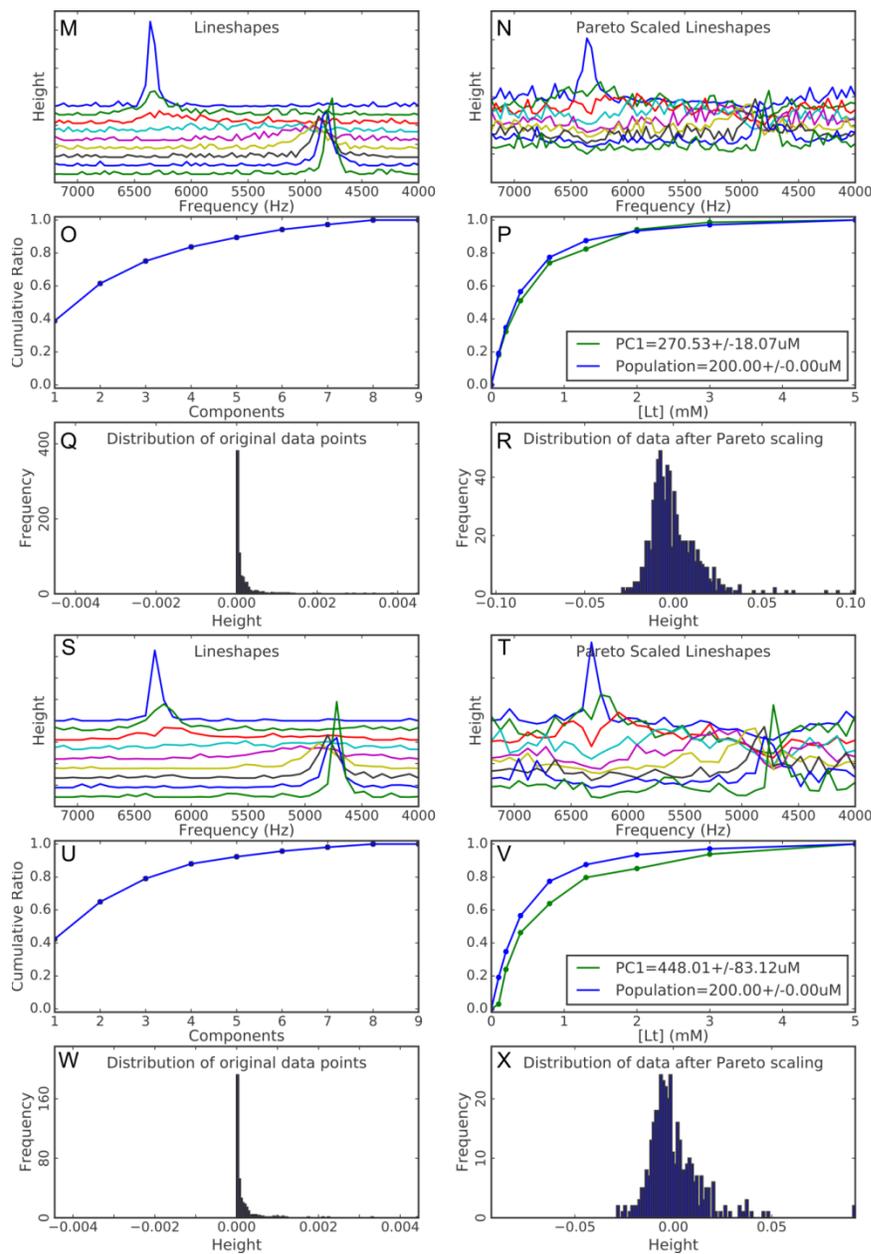


Figure SIV-6. The routine success in deriving binding isotherms from the intermediate exchange regime results from standard digital resolution and Pareto scaling. These move the data matrix toward a normal-like distribution. Reducing digitization of spectra to unrealistically sparse levels imparts deterioration to PC1 representation of the binding isotherm.  $^1\text{H}$  line shapes for 1:1 ligand binding in the slow-intermediate exchange regime are simulated using parameters listed in Table SIV-1. (A, B) use 3000 points for digital resolution of 1.07 Hz/point. (G,H) use 1500 points for digital resolution of 2.13 Hz/point. (M, N) uses 734 points for digital resolution of 4.36 Hz/point.



(S,T) use 371 points for digital resolution of 8.62 Hz/point. (B, H, N, T) Pareto-scaled simulated 1D spectra. (C, I, O, U) plot the proportion of the statistical variance represented by increasing numbers of principal components. (D, J, P, V) plot binding isotherms captured by PC1 (green) along with the simulated binding isotherm (blue). (E, K, Q, W) plot histograms of the noise and peak heights in the simulated spectra. (F, L, R, X) show histograms of these heights after Pareto scaling.

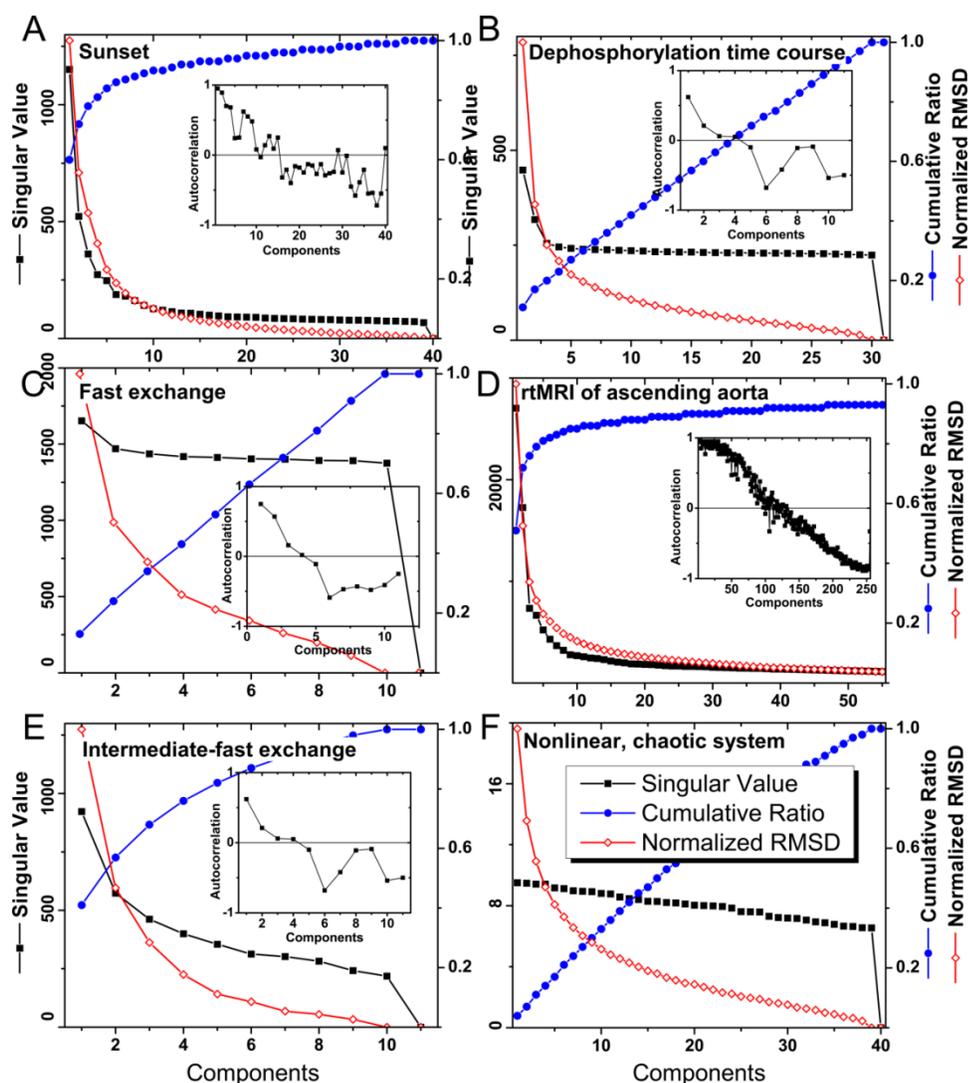


Figure SIV-7. The nonlinearity of the examples is suggested by analyses of their principal components. Black curves show scree plots of singular values. Blue curves plot the collective contributions of that number of principal components to the statistical variances. RMSD values between reconstructed and original spectra are represented by red lines (see ref 27 in main text). Autocorrelation coefficients (insets) that are smaller and / or scattered suggest the insignificance of those components. The first three examples shown are (A) the sunset of Fig. SIV-3A,B, (B) the example of fast exchange in Fig. IV-1C,D, and (C) the NMR spectra of an enzyme undergoing dephosphorylation of Fig. SIV-3E,F. The other cases plotted are (D) the real-time MRI movie of the ascending aorta in the chest cavity of Fig. IV-5 and (E) the FHA domain titrated with a phosphopeptide, exhibiting fast-intermediate exchange in the HSQC spectra in Fig. IV-1C. For reference, (F) plots a standard example of a very nonlinear system based on a logistic map (see Supporting Experimental Section) evaluated similarly.

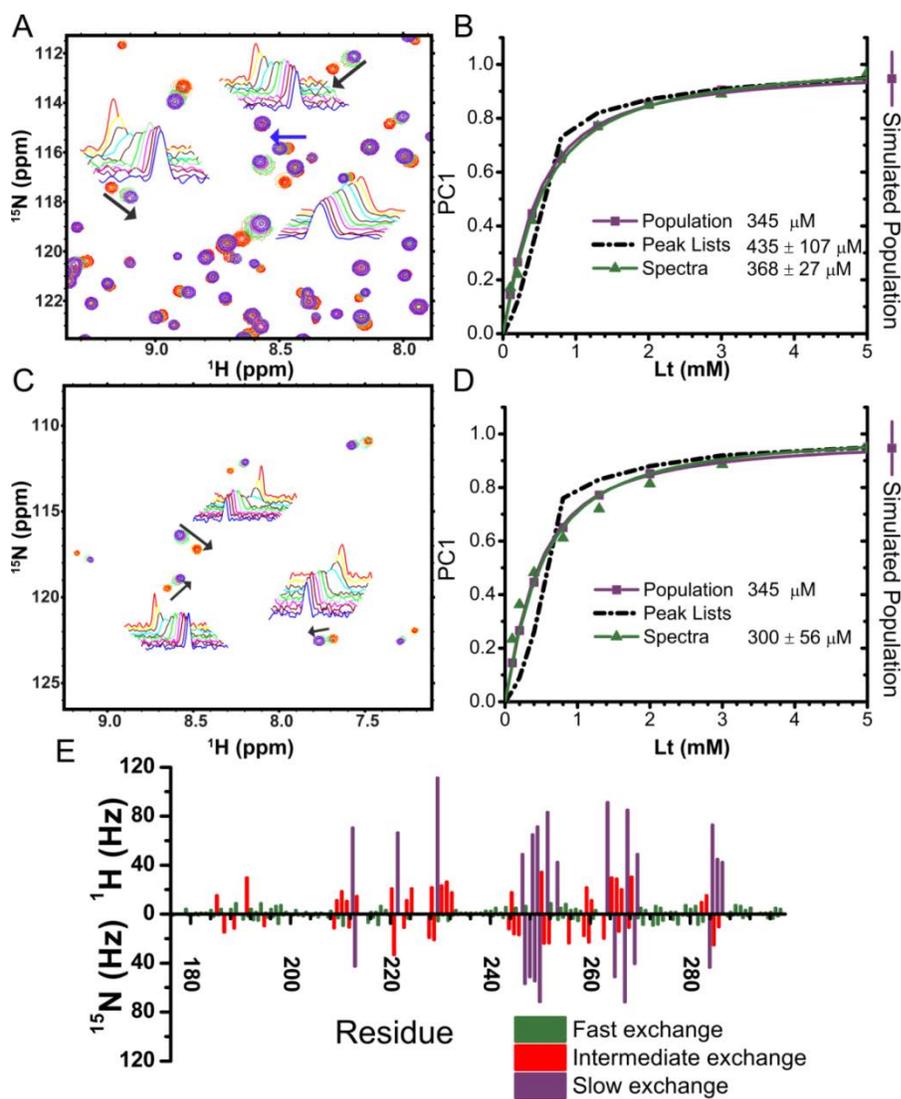


Figure SIV-8. SVD extracts binding isotherms suitable for fitting in titrations in mixed exchange regimes and regimes that are artificially homogeneous in intermediate exchange. A) HSQC spectra simulated to represent a titration series in which fast, slow, and intermediate exchange are present. Line shapes of intermediate (black arrows) and fast exchange (blue arrow) are shown as insets. B) PC1 of spectra suppresses the distortion observed in PCA of peak lists. C) HSQC spectra simulated by selected peaks from panel A that exhibit intermediate exchange behaviors in  $^1\text{H}$  and  $^{15}\text{N}$  dimension. Line shapes are shown as insets. D) PC1 of the simplified spectra, made homogeneous in intermediate exchange, capture the binding isotherm. (E) The chemical shift differences for each residue in the simulated examples shown in (A) ( $Lt = 0.2$  mM) for fast, intermediate, and slow exchange regimes. The upper panel represents the  $^1\text{H}$  dimension while the lower panel shows the  $^{15}\text{N}$  dimension. The heights of columns indicate frequency differences between apo and bound states. Table SIV-1 lists the simulation parameters.

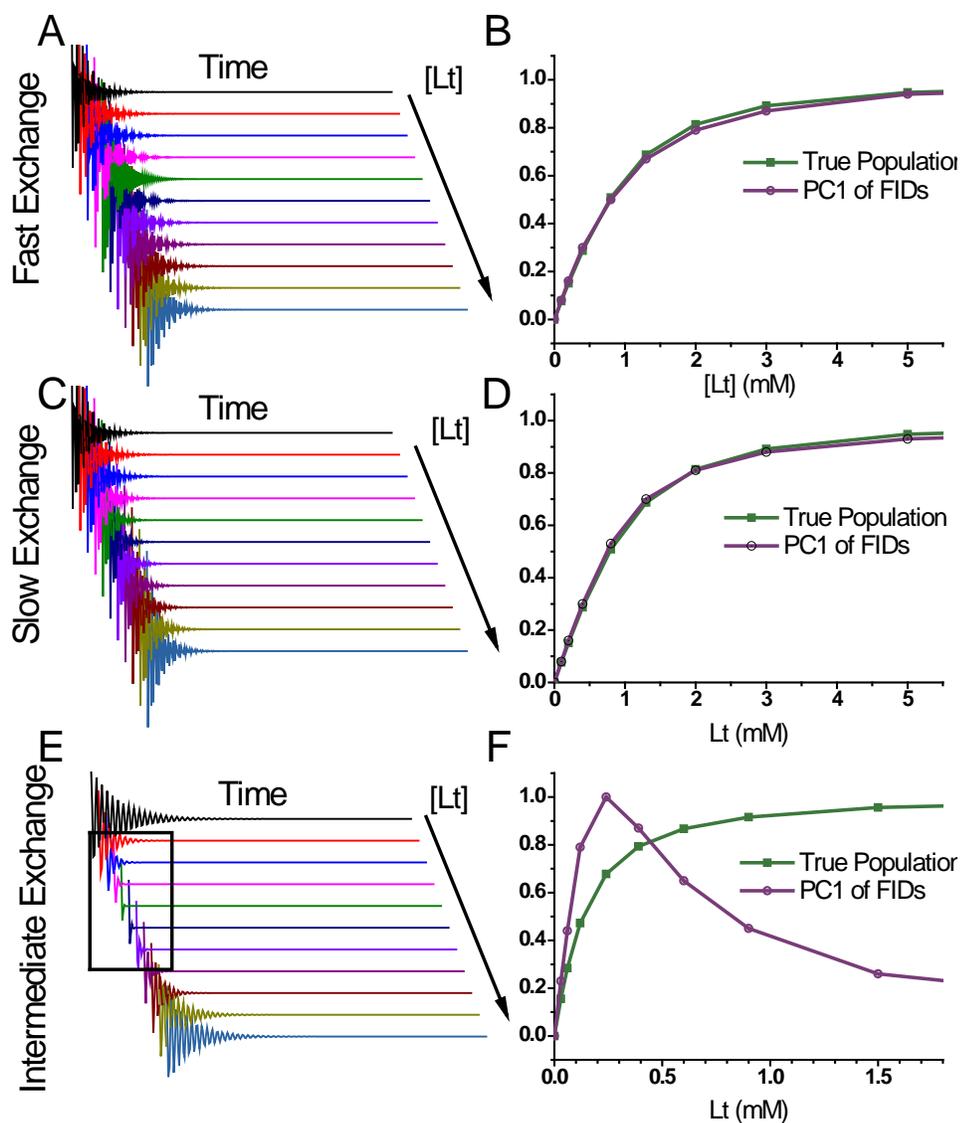


Figure SIV-9. PCA of free induction decays (FIDs) captures the correct binding isotherms in fast and slow exchange regimes. Simulated  $^1\text{H}$  FID envelopes are plotted for A) fast exchange, (C) slow exchange, and (E) intermediate exchange as a function of total [Ligand], Lt. Simulated FIDs were inverse Fourier-transformed from Fig. SIV-1A, C, E, respectively. (B) PC1 derived from the FIDs of (A) provides binding isotherms equivalent to the true population changes in the fast exchange regime. (D) PC1 of FIDs captures the correct binding isotherm in the slow exchange regime. (E)  $^1\text{H}$  FID envelopes in the intermediate exchange regime become significantly weaker, due to the broadening, and then stronger as Lt increases to saturate the host. (F) PC1 of the FIDs with high intermediate exchange broadening cannot represent the binding isotherms.

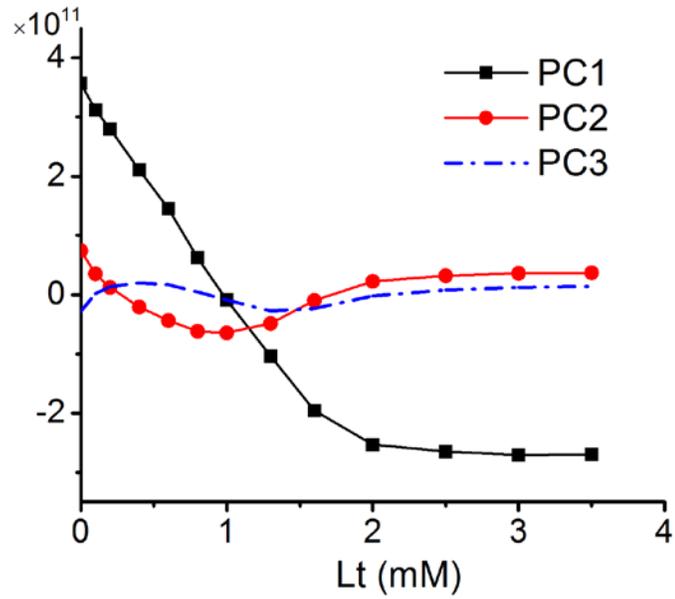


Figure SIV-10. Recognizing significant PCs. The first three PCs from the 2-step binding of GCDA to bile acid binding protein in Fig. IV-4 are plotted. PC1 is colored black, PC2 red, and PC3 with blue dashed lines. PC1 and PC2 are larger and smooth, making them more likely to be meaningful.

#### IV.8.5 Supporting movies

Movie IV-S1. MRI magnitude movie of cerebral spinal fluid flow, with plotting of PCs

Movie IV-S2. MRI magnitude movie of ascending aorta in the chest, with plotting of PCs

#### IV.8.6 Supporting references

1. Williamson, M. P. 2013. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* 73:1-16.
2. Markin, C., and L. Spyropoulos. 2012. Increased precision for analysis of protein–ligand dissociation constants determined from chemical shift titrations. *J. Biomol. NMR* 53:125-138.
3. van den Berg, R. A., H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde, and M. J. van der Werf. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 7:142.
4. Parsons, H. M., C. Ludwig, U. L. Gunther, and M. R. Viant. 2007. Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC Bioinformatics* 8:234.
5. Kovrigin, E. L. 2012. NMR line shapes and multi-state binding equilibria. *J. Biomol. NMR* 53:257-270.
6. Helmus, J. J., and C. P. Jaroniec. 2013. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. *J. Biomol. NMR* 55:355-367.
7. Lee, G., J. Li, J. C. Walker, and S. R. Van Doren. 2003.  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  resonance assignments of kinase-interacting FHA domain from Arabidopsis phosphatase KAPP. *J. Biomol. NMR.* 25:253-254.
8. Ding, Z., H. Wang, X. Liang, E. R. Morris, F. Gallazzi, S. Pandit, J. Skolnick, J. C. Walker, and S. R. Van Doren. 2007. Phosphoprotein and phosphopeptide interactions with the FHA domain from Arabidopsis kinase-associated protein phosphatase. *Biochemistry* 46:2684-2696.
9. Stoyanova, R., and T. R. Brown. 2001. NMR spectral quantitation by principal component analysis. *NMR Biomed.* 14:271-277.
10. Vaidya, P. G., S. P. S. Anand, and N. Nagaraj. 2010. A Nonlinear Generalization of Singular Value Decomposition and Its Applications to Mathematical Modeling and Chaotic Cryptanalysis. *Acta Appl Math* 112:205-221.
11. Xu, J., Y. Lee, Lesa J. Beamer, and Steven R. Van Doren. 2015. Phosphorylation in the Catalytic Cleft Stabilizes and Attracts Domains of a Phosphohexomutase. *Biophys. J.* 108:325-337.
12. Jaumot, J., V. Marchan, R. Gargallo, A. Grandas, and R. Tauler. 2004. Multivariate curve resolution applied to the analysis and resolution of two-dimensional [ $^1\text{H}$ , $^{15}\text{N}$ ] NMR reaction spectra. *Anal. Chem.* 76:7094-7101.

## **CHAPTER V**

### **Tracking Equilibrium and Non-equilibrium Shifts in Data with TREND**

Adapted with permission from Jia Xu, Steven R Van Doren, *Biophys. J.* (2017)

<http://dx.doi.org/10.1016/j.bpj.2016.12.018> Copyright 2017 Elsevier

#### **V.1 Abstract**

Principal component analysis (PCA) discovers patterns in multivariate data that include spectra, microscopy, and other biophysical measurements. Direct application of PCA to crowded spectra, images, and movies (without selecting peaks or features) was shown recently to identify their equilibrium or temporal changes. To enable the community to utilize these capabilities with a wide range of measurements, we have developed multi-platform software named TREND to track equilibrium and non-equilibrium population shifts among 2D data frames. TREND can also carry this out by independent component analysis. We highlight a few examples of finding concurrent processes. TREND extracts dual phases of binding to two sites directly from the NMR spectra of the titrations. In a cardiac movie from magnetic resonance imaging, TREND resolves principal components (PCs) representing breathing and the cardiac cycle. TREND can also reconstruct the series of measurements from selected PCs, as illustrated for a biphasic, NMR-detected titration and the cardiac MRI movie. Fidelity of reconstruction of series of NMR spectra or images requires more PCs than needed to plot the largest population shifts. TREND reads spectra from many spectroscopies in the most common formats (JCAMP-DX and NMR), and multiple movie formats. The TREND package thus provides convenient tools to resolve the processes recorded by diverse biophysical methods.

## V.2 Introduction

Plotting the course of biomolecular or physiological processes typically uses procedures specific to the field. In the case of spectroscopy and imaging, tracking the process can be laborious due to the steps of assigning the peaks of the spectra or features in the images, manually choosing peaks or image features subjectively judged optimal for monitoring the process of interest, and managing complications from any concurrent processes. Spectral overlap and peak broadening (e.g., from chemical exchange in NMR) can prevent correct fitting (1). A more elegant alternative to such efforts is to apply unsupervised, multivariate statistical pattern recognition such as principal component analysis (PCA). PCA has provided insight from series of measurements from diverse techniques of molecular biophysics that include magnetic resonance, vibrational, optical and dichroic spectroscopies; X-ray scattering and diffraction; mass spectrometry; calorimetry; hydrodynamics; atomic force microscopy; electron microscopy; and imaging by fluorescence, Raman or light scattering, as well as functional magnetic resonance imaging (Table SV-1 and references therein). Biophysical studies have often used PCA to determine dependencies of various reactions upon time, concentration or other conditions, e.g. in protein folding (Table SV-1). PCA applied directly to spectra, images, and movies appears to be a convenient and general way to determine the main trends of change among measurement frames that record many localized changes. PCA is much more accommodating of many data distributions than is often appreciated (2). It transforms many measured variables to far fewer and uncorrelated principal components (PCs) that each capture part of the trends of covariation among measured variables (2).

Table V-1. Executable file components of TREND.

Executable file	Roles	Interface
<i>trendmain.exe</i>	Preprocess and compute PCs or ICs	CLI
<i>trendmaingui.exe*</i>	“	GUI
<i>trendplot.exe</i>	Plot selected PCs or ICs with choice of normalization	CLI
<i>trendplotgui.exe*</i>	“	GUI
<i>trendreconstruct.exe</i>	Reconstruct spectra, images, or movies from PCs	CLI
<i>trendreconstructgui.exe*</i>	“	GUI

\* Executable files with GUI are *trendmaingui.app*, *trendplotgui.app*, and *trendreconstructgui.app* for OS X or macOS platforms.

PCA of NMR peak lists was used to track equilibrium transitions of proteins due to pH (3) and binding of partners (4-6). Closely related singular value decomposition (SVD) of NMR peak pick lists was used to reconstruct filtered basis spectra for use in fitting biphasic ligand binding (7) or for identifying binding sites (8). The applications of PCA were very recently extended directly to NMR spectra, images, and movies without choosing any peaks or features for analysis. Moreover, applying PCA directly to NMR spectra makes binding isotherms easily accessible in all chemical exchange regimes, including intermediate exchange where severe broadening and nonlinearity of peak shifts ordinarily mask the true course of molecular association (1, 9). Application of SVD to series of *time*-dependent 2D images or spectra extracted the dominant time course as PC1. The approach also detected multiple time-evolving processes in MRI movies as PCs. Similarly, when two sequential steps of binding were monitored by NMR, PCA detected both binding steps and the intermediate with a single ligand bound (9). Accomplishments with SVD (PCA) have usually been limited to the laboratories that wrote task-specific code to perform the calculations however.

Independent component analysis (ICA) can complement PCA. ICA aims instead to find *independent* components (ICs) (2). The quest of ICA for statistical independence

is more demanding than PCA's aim of correlation coefficients of zero. These objectives are equivalent for Gaussian (normal) distributions. ICA can be regarded as more general than PCA and is effective for non-Gaussian data and situations where PCA fails (2). However, ICA can be very slow to compute compared to PCA, lower in convergence, and require repeated calculations. Like PCA, ICA has been used to reduce dimensionality and filter or separate data in processing signals, images (10), large biological data sets (11), and NMR spectra of mixtures (12-14).

To make these capabilities available for application to a variety of spectroscopic and imaging techniques used in biophysics, we have developed a software package named TREND, for TRacking and resolving Equilibrium and Non-equilibrium population shifts in Data. Its main means of tracking the shifts is PCA implemented with SVD. Its secondary means is an ICA algorithm, which recapitulates the PCA results we examined, provided the correct number of ICs is specified. We first sought to extract binding isotherms, equilibrium shifts and time courses (all potentially with multiple components) from series of NMR spectra. Because of the suitability of PCA for many other kinds of series of 2D digital data frames, we utilized the Python community's support of file I/O in multiple data formats (e.g., movies and spreadsheets) and wrote code for additional spectroscopic formats, enabling wide application (e.g. JCAMP-DX, Sparky peak list). For example, we analyzed a cardiac MRI movie (15) with TREND to isolate multiple aspects of the cardiac cycle and to reconstruct movies from combinations of PCs. This software package can resolve biologically relevant reactions and processes with relative ease from many biophysical sources of complicated spectral and imaging data.

## V.3 Methods

### V.3.1 Implementation of TREND

TREND was written in Python 2.7 and calls NumPy for linear algebra and random number generation. It implements PCA (SVD) with function calls to NumPy. We first wrote TREND for operation at the command line. We added a GUI, supported by Goocy, using function calls to wxPython. Most users will prefer to use the GUI to operate TREND. TREND comprises three programs, each with both interfaces (Table V-1). The executable files *trendmaingui* and *trendmain* compute the PCs or ICs across the 2D series of measurements, create temporary files used by the plotting or reconstruction programs run afterward, and plot the first three components plus benchmarks of their significance. *Trendplotgui* and *trendplot* provide optional plotting that is customizable in terms of the number and choice of normalization of the components. Optional reconstructions of the measurement series are available from *trendreconstructgui* and *trendreconstruct* (Table V-1). Explanations of the flags and parameters for the command line versions are available online in the manual for TREND (16). For convenience of installation, we packaged TREND and the public domain software it depends upon using PyInstaller. Consequently, TREND does not need Python on the host system. Distributions are available for Windows 7 and later, Mac OS X 10.7 and later, and these versions of Linux: Ubuntu 14.04 / Fedora 23, Ubuntu 16.04, and Red Hat 7.1 / CentOS 7.

### V.3.2 Conversion of a stack of 2D measurements into a matrix for analysis

A wide variety of 2D measurements can be read and analyzed by *trendmaingui* and *trendmain*. This includes images or movie frames comprising pixels, 1D and 2D spectra from many spectroscopies, lists of peak positions and heights, and unprocessed

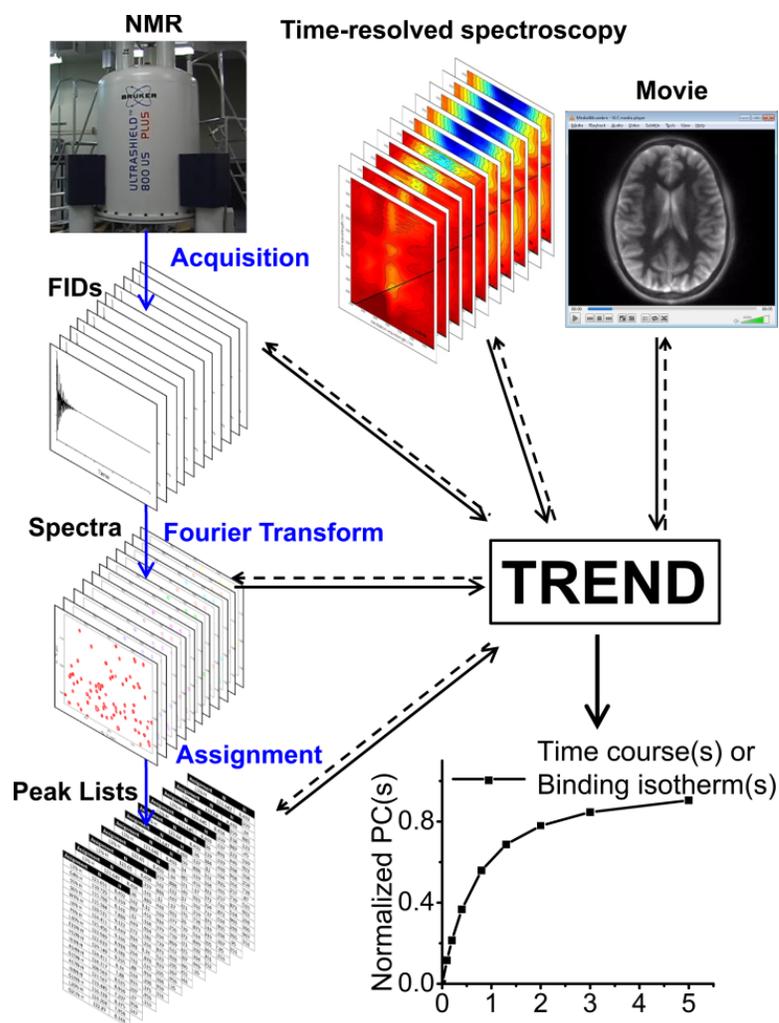


Figure V-1. Work flows to TREND extraction of the main components of change across a series of measurements. TREND reads a series of spectra, peak lists, FIDs, images, or a movie and extracts the dominant trends from them. Spectral or electronic data series in general-purpose JCAMP-DX, text, or spreadsheet formats, NMR formats, and video formats are readable. The dashed lines signify reconstruction of measured data from the principal components chosen.

NMR spectroscopic data in the time domain (free induction decays, FIDs; Fig. V-1). The program reads NMR spectra in NMRpipe, Sparky, and Bruker Topspin formats (as well as FIDs in NMRpipe and Topspin formats) (Table V-2) using code from Nmrglue (18).

*Trendmaingui* and *trendmain* also read Agilent (Varian) VNMRJ format and JCAMP-DX formats of Bruker, Agilent, and Jeol spectrometers. To analyze the measurements from many other kinds of spectroscopy and biophysical measurements (Table SV-1), the program reads the most common JCAMP-DX formats, as well as spreadsheet and text

Table V-2. File formats read and reconstructed by TREND.<sup>\$</sup>

Choice in trendmaingui	Format	Reconstruction support	Comment
<b>NMR Data Formats</b>			
<b>fid</b>	NMRPipe FID	Yes	
<b>ft2</b>	NMRPipe Ft2	Yes	
<b>ucsf</b>	Sparky UCSF	Yes	
<b>brukerfid</b>	Bruker Topspin FID	Yes	fid, ser in /1/pdata/ subfolder *
<b>brukerft2</b>	Bruker Topspin spectra	Yes	1r, 2rr files
<b>agilentfid</b>	VnmrJ, OpenVnmrJ FID	Yes	fid
<b>agilentspectra</b>	VnmrJ, OpenVnmrJ spectra	No	Phasefile <sup>\$</sup>
<b>sparkylist</b>	Sparky peak list	Yes	duplicate peaks not allowed
<b>JCAMP-DX: Joint Committee on Atomic and Molecular Physical data – Data Exchange format</b>			
<b>jcamp</b>	JCAMP-DX	No	Only supports X..(Y+Y) and (XY..XY) ***
<b>Text File Formats</b>			
<b>txt</b>	floating point	Yes	for series of text files
<b>complextxt</b>	complex numbers	Yes	for series of text files
<b>singletxt</b>	complex or floating point	Yes	for single .TXT file containing entire series
<b>Spreadsheet formats</b>			
<b>csv</b>	comma-separated floating point	No	for series of .CSV files
<b>complexcsv</b>	comma-separated complex numbers	No	for series of .CSV files
<b>singlecsv</b>	comma-separated complex or floating point	No	for single .CSV file containing entire series
<b>excel</b>	Excel format	No	for series of Excel files
<b>singleexcel</b>	Excel format with tabs	No	for single file with single or multiple tabs
<b>Images and Movies</b>			
<b>png</b>	images in PNG format	Yes	For series of .PNG files
<b>movie</b>	common video formats	Yes	.ogv, .mp4, .mpeg, .avi, .mov, .webm

<sup>\$</sup> See the online TREND manual (16).

\* Currently the processed spectra must be saved by setting processing number as 1.

\*\*\* JCAMP-DX is a general format for exchanging and archiving data from many instruments, including but not limited to IR, Raman, UV-Vis, Fluorescence, NMR, and EPR. The data stored in JCAMP-DX files can be spectral plots, contours, or peak tables. TREND supports the most common JCAMP-DX formats. The digital data in JCAMP-DX can be AFFN (ASCII FREE FORMAT NUMERIC) form or ASDF (ASCII SQUEEZED DIFFERENCE FORM). TREND supports decoding compressed data, including PAC, SQZ, DIF, SQZDUP, and DIFDUP. Two most common tabular data forms, (X++(Y..Y)) and (XY..XY) are supported. TREND reads a series of JCAMP-DX files, or a single JCAMP-DX file with one or multiple blocks. TREND

supports NTUPLE format (introduced by JCAMP-DX 5.0), which is designed for multi-dimensional techniques with data sets with multiple variables. For example, JCAMP-DX NMR uses NUTPLE to show mixed real/imaginary FID data sets. See for format details in <http://www.jcamp-dx.org/> , [https://badc.nerc.ac.uk/help/formats/jcamp\\_dx/](https://badc.nerc.ac.uk/help/formats/jcamp_dx/) and <http://wwwchem.uwimona.edu.jm:1104/spectra/testdata/index.html>

formats commonly written by instruments (Table V-2). *Trendmaingui* and *trendmain* read NMR peak lists either in the format of Sparky peak lists (19) or plain text files, prior to converting them into column vectors (3, 7). Movies are read in multiple formats (i.e. avi, mov, mp4, ogv, webm) by the MoviePy module into 3D arrays with color layers. *Trendmaingui* and *trendmain* convert the movie frames to gray-scale (8-bit depth) and rearrange them into 2D matrices (Fig. V-2). Time-lapse series of PNG images are read, using the Scipy module of Python, and handled similarly.

In the case of NMR data, spectra very recently emerged as probably the preferred format for application of PCA (9). In the examples below, NMR spectra (collected with a uniform set of parameters) were processed with NMRPipe (19, 20) and converted to the UCSF format of Sparky (21). NMR spectra in UCSF format were read by *trendmaingui* or *trendmain* for conversion into 2D matrices (Fig. V-2). Unprocessed NMR data in the time domain (FIDs) can also be read, processed, and the solvent signal subtracted. (Analysis of time domain data is justified by Parseval's theorem regarding the equivalency of signals in the time and frequency domains (22)).

### V.3.3 Preprocessing

Regardless of original data format, columns from each 2D measurement read are positioned end-to-end into a single 1D vector for convenience (9) (Fig. V-2). These 1D columns are arrayed over the experimental variable (concentration, pH, time, etc.) into

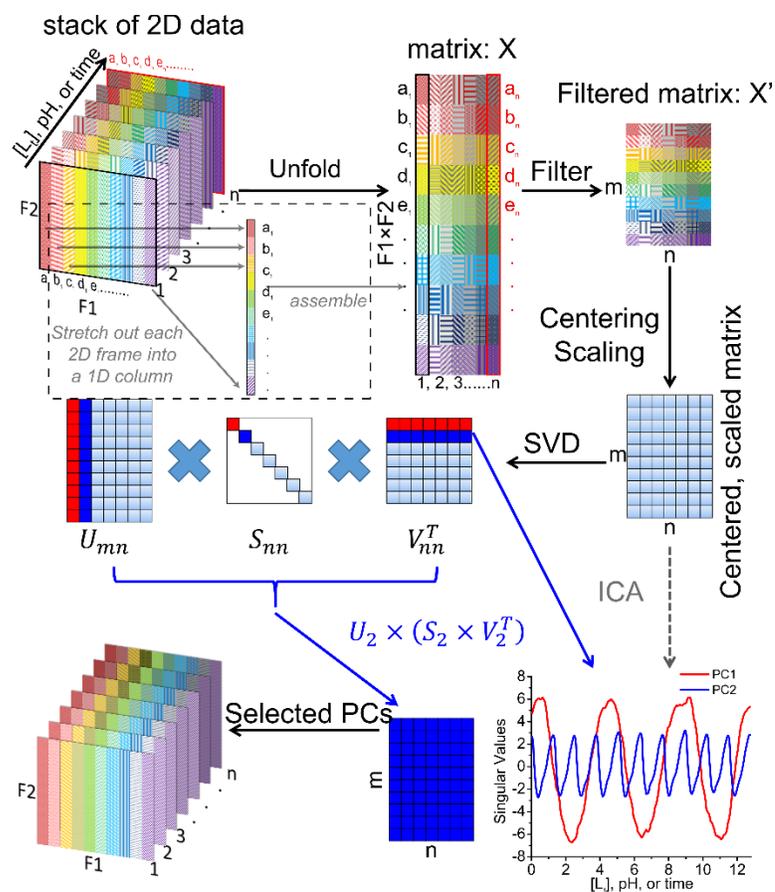


Figure V-2. TREND implementation of PCA (SVD), ICA, and reconstruction. The algorithm reorganizes a series of 2D measurements as a series of 1D columns. The dashed box indicates the unfolding of the first 2D data (marked by black edges) with F2 rows and F1 columns to a long 1D column with F1×F2 points. The data matrix  $X$  is compressed to  $X'$  and used for SVD. (No user selection of spectral peaks or image features is involved). The resulting first several right singular vectors in the decomposed matrix  $V^T$  represented the largest trend(s). ICA can be used to corroborate SVD results. Single or multiple PCs can be used to reconstruct the original data series.

the data matrix  $X$ , which has F1×F2 points in the column dimension and  $n$  points per row for the  $n$  experimental conditions. In order to expedite manipulations of this matrix  $X$  and facilitate calculations on a modest laptop computer, each vector is compressed by removing unchanging positions, resulting in matrix  $X'$  (Fig. V-2). For SVD of spectra, the user is encouraged to use a lower threshold that is 3 to 7-fold the noise level to filter out low intensity regions of the spectra, which compresses matrix  $X'$  further. However, it is better to avoid a lower threshold where intermediate exchange broadening significantly weakens NMR peaks.

As required by PCA and ICA algorithms, the rows of compressed matrix  $\mathbf{X}'$  are centered, and then optionally scaled. Scaling enlarges weaker signals relative to large signals. The options for scaling methods include *autoscaling*, *Pareto scaling*, or *no scaling* (23). *No scaling* appears acceptable in most titrations, but *autoscaling* generally enhances fits to the binding isotherms. *Autoscaling* obviates the systematic scaling of  $^{15}\text{N}$  NMR peak shifts down by several-fold relative to  $^1\text{H}$  shifts that were used in PCA of lists in ref (3). *Autoscaling* also generalizes to  $^1\text{H}$ - $^{13}\text{C}$  correlation spectra. *Pareto scaling* is recommended for NMR titrations with substantial intermediate exchange broadening (9). *Range*, *Vast*, and *Level* scaling (23) are also implemented in *trendmain* but do not work well with NMR spectra. *No scaling* has been used for MRI movies. Column centering and scaling are not necessary in our experience, but are available in *trendmaingui* and *trendmain* as they are sometimes used for PCA (24) and ICA (11). The descriptions of the data scaling and centering methods (23) are listed in Table V-S2

### V.3.4 Calculating principal components via SVD

The compressed, preprocessed matrix  $\mathbf{X}'$  has  $m$  points per column and  $n$  points or experimental conditions in each row, with  $m > n$ .  $\mathbf{X}'$  can be decomposed into three matrices:

$$\mathbf{X}'_{mn} = \mathbf{U}_{mn}\mathbf{S}_{nn}\mathbf{V}'^T_{nn} \quad \text{Equation V-1}$$

where  $\mathbf{U}$  and  $\mathbf{V}'^T$  are orthogonal matrices and  $\mathbf{S}$  is a diagonal matrix that contains the square roots of eigenvalues for vectors in  $\mathbf{U}$  or  $\mathbf{V}$  in descending order. To obtain the trends of change across the measurements, we are interested in  $\mathbf{V}'^T$ , whose rows span  $\mathbf{X}'$  and are called the right singular vectors. The  $\mathbf{V}'^T$  matrix has row vectors  $\mathbf{V}'^T_{nn} = (\mathbf{V}'^T_1, \mathbf{V}'^T_2, \mathbf{V}'^T_3 \dots \mathbf{V}'^T_n)$ . Importantly, the first row in the  $\mathbf{V}'^T$  matrix is PC1 and the second row

PC2, i.e. the two largest trends of change among the series of spectra or images measured. To obtain these PCs that record the relationships among columns in  $X'$  (Fig. 2), it suffices to calculate  $V^T$ . The rows of  $V^T$  are orthonormal eigenvectors of the symmetric matrix  $X'^T X'$  (Fig. 2). (The normalized form of  $X'^T X'$  is equivalent to the covariance matrix, the alternative algorithm for computing PCA (2).) The normalized PC1 values from the first row of  $V^T$  indicate the fractional population of the main change at each measurement in the series of measurements. When obtained from a typical titration of ligand binding, PC1 represents the binding isotherm; a dissociation constant may be fitted to it (9).

### V.3.5 Reconstruction of spectra, images, or movies by PCA

The reconstructed dataset  $X_{reconst}$ , with size of  $m \times n$ , can be calculated as:

$$X_{reconst} = U_a(S_a V_a^T) + U_b(S_b V_b^T) + U_c(S_c V_c^T) + U_d(S_d V_d^T) + U_e(S_e V_e^T) \quad \text{Equation V-2}$$

where a, b, c, d, e... refer to the index of PCs generated by *trendmain* or *trendmaingui* to use in the reconstruction by *trendreconstruct* or *trendreconstructgui*. (Note a, b, c, d, e ... can be nonconsecutive integers. To enable this, the “reconst” box should be selected in *trendmaingui* or the --reconst flag included when using *trendmain*). The  $U$  matrix is used for the reconstruction. It can be rewritten as column vectors:  $U_{mn} = (U_1, U_2, U_3 \dots U_n)$  which lie in the column space of  $X'$ .  $U$  can be calculated similarly to  $V^T$ , by solving eigenvectors of the matrix  $X' X'^T$ . In order to recover the original 2D data series, the preprocessing steps of centering, scaling, and compression (filtering) can be reversed as described in the manual for TREND. The user can choose to reconstruct the centered and scaled matrix, matrix  $X'$ , or matrix  $X$  in the format of the original data (Fig. V-2).

### V.3.6 ICA Calculations

ICA is available in TREND and implemented using scikit-learn (25). Despite the potential generality of ICA, two limitations need to be respected. Since the magnitudes of independent components (ICs) cannot be determined, their contributions cannot be ranked. ICA is also prone to local minima during optimization, requiring comparisons of repeated calculations (10, 11). TREND implements the FastICA algorithm for computational efficiency. FastICA preprocesses data by PCA in order to reduce dimensions and avoid overlearning (26-28). (Overlearning is an underdetermined situation that interferes in obtaining parameters, and which introduces artifacts to ICs (27, 28)).

ICA decomposes the data matrix  $X$  as:

$$X = AS \tag{Equation V-3}$$

where  $A$  is the unknown mixing matrix that is invertible, square, and mixes the components in  $X$ .  $S$  is the matrix containing underlying independent sources. The aim of ICA is to solve for the mixing matrix  $A$  because it contains the ICs that may contain the meaningful trends sought. However,  $A$  and  $S$  both being unknown makes ICA calculations challenging (10). The equation can be rewritten:

$$S = WX = VX_w \tag{Equation V-4}$$

where  $W$  is the unmixing matrix that is calculated as  $A^{-1}$ . In order to simplify and improve convergence of ICA,  $X$  is preprocessed to remove correlations and to normalize it, a process called whitening, which generates  $X_w$ . FastICA implements this whitening step using PCA to calculate the whitened data matrix  $X_w$  as:

$$X_w = \left(D^{-\frac{1}{2}}E^T\right)X \quad \text{Equation V-7}$$

$E$  is the matrix whose columns are normalized eigenvectors of the covariance matrix of  $XX^T$ .  $D$  is the diagonal matrix of the corresponding eigenvalues. The preprocessing with PCA also removes noise and reduces dimensions for ICA. The whitening simplifies the ICA problem to finding the unknown rotation matrix  $V$  that is defined as  $D^{-\frac{1}{2}}E^T$ . In FastICA,  $V$  is estimated by maximizing non-Gaussian character. The equations lead to:

$$A = W^{-1} = (VD^{-\frac{1}{2}}E^T)^{-1} \quad \text{Equation V-6}$$

## V.4 Results and discussion

### V.4.1 Workflows of TREND

For wide application of SVD or ICA to diverse series of 2D measurements, we wrote TREND in Python in order to read and analyze multiple types of data. These include diverse spectra, images, movies, or lists in text or spreadsheet formats available from many modern instruments (Fig. V-1). The spectral formats include widely used JCAMP-DX standards and NMR formats. TREND can also apply PCA or ICA to a single 2D data matrix read in from a text file, spreadsheet file, or multi-block JCAMP-DX file containing multiple spectra (Table V-2). The algorithm of the *trendmain* and *trendmaingui* executable files includes these steps:

- i. Convert each 2D measurement into a 1D vector arrayed by the experimental condition varied, in the data matrix  $X$ .
- ii. Preprocess  $X$  with compression to  $X'$  and optional scaling.

iii. Perform streamlined SVD or ICA to identify components (PC1, PC2, ... or IC1, IC2, ...) representing the major trend(s) varying with the experimental variable (Fig. V-1).

The TREND package provides additional executable files for plotting the course of selected PCs or ICs or for rebuilding spectra, images, or movie from selected PCs (Table V-1). For convenience, the plotting and reconstruction routines read temporary files just created by *trendmain* or *trendmaingui*; this frees the user from specifying input files, which is optional. The user may operate and customize these computations by a choice of GUI or command line arguments described in documentation for the software.

While we used the Python routine NumPy to implement PCA (SVD) and scikit-learn (25) to implement ICA calculations, corresponding routines are available in R (29), MATLAB (30), and the MATLAB Statistics Toolbox. Recreating the workflows and functions depicted in Figs. V-1 and V-2 in an R or MATLAB environment would require code to parse the file formats of interest, reduce their dimensionality (i.e. “unfold” them), preprocess for readiness for the SVD or ICA routine, and interpret or reconstruct the results in the appropriate format. TREND spares the user this effort with a package that is user-friendly for NMR and other measurements from a variety of instrumentation, including spectroscopies and imaging; see Table V-2 for data formats handled. TREND is free for academics, avoiding the cost of licensing MATLAB. TREND requires < 150 MB of disk space whereas the MATLAB environment occupies 2 to 3 GB. TREND is portable and its installation lacks dependencies, other than the need for internet access upon first usage.

We present examples of uses of TREND that illustrate (i) its performance in resolving two or more processes, which is non-routine by conventional means, and (ii) its wide applicability to trace and reconstruct concurrent, complex transformations recorded by biophysical means such as spectra or imaging.

#### V.4.2 Examples of ligand binding to two sites detected by NMR

Antecedents to TREND's direct application of PCA to spectra and images were previous PCA studies of NMR peak lists. SVD was used to filter noise out of the lists, in turn used to reconstruct clean basis spectra to resolve three pH transitions (3) or two binding events (4, 7). With TREND we demonstrate a direct spectrum-driven approach to the latter examples of two biphasic associations. Fig. V-3A plots a two-site binding scheme, where P and L denote [protein] and [ligand], respectively.  $K_{D1}$  and  $K_{D2}$  are dissociation constants from site 1 and 2.  $PL_{n1}$  and  $PL_{n2}$  are intermediates with ligand at site 1 or 2, where  $n1$  and  $n2$  indicate numbers of ligand molecules that bind cooperatively to site 1 and 2, respectively.  $PL_{n1}L_{n2}$  stands for the fully bound state. Equations 3 to 6 from ref (7) were used to simulate populations of species from the two-site binding scheme in a series of  $^{15}\text{N}$  HSQC spectra (Fig. V-3B) using methods given in Supporting Material. The curvature in the simulated shifts of several peaks (red arrows in Fig. V-3B) accompanies more than one mode of binding (7). PCA on the peak lists (chemical shifts) captures two smooth components, PC1 and PC2 (purple in Fig. V-3C), contributing 90% and 6% of the variance, respectively. The PC1 and PC2 components of the peak lists were recreated using *trendreconstruct*. PC1 captures from the curved trajectories of peak movements the main linear paths of change (Fig. SV-1A,B). PC2 identifies the peak shifts orthogonal to PC1 (Fig. V-S1C). Computing PC1 and PC2 instead directly from the

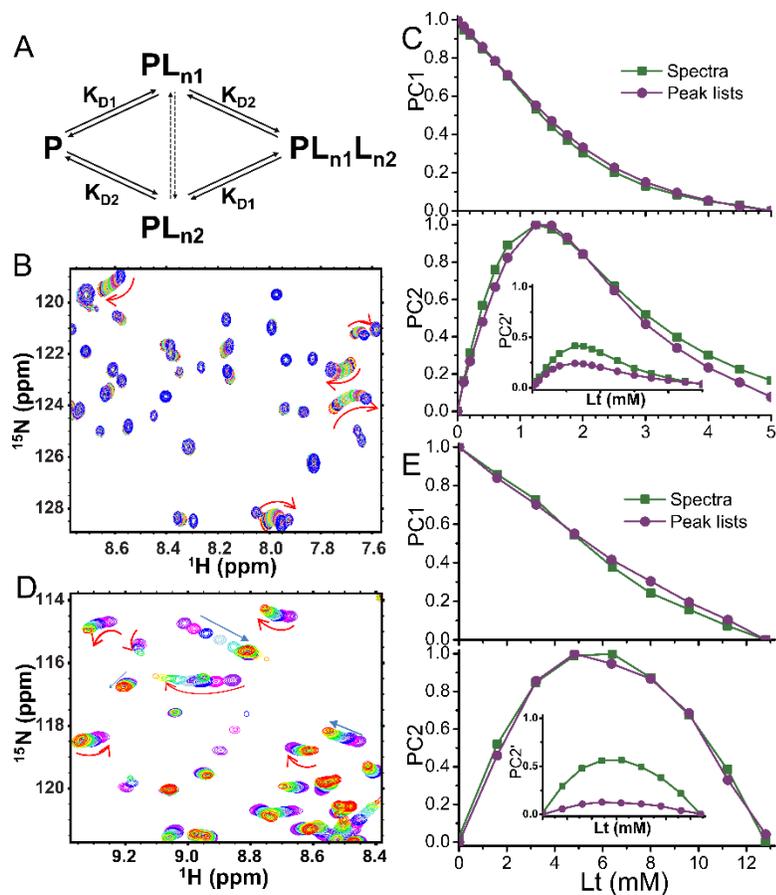


Figure V-3. TREND identifies two components in titrations of two binding sites. (A) Scheme of the two-site binding model. The number of ligands that bind to sites 1 and 2 are  $n_1$  and  $n_2$ , respectively. (B)  $^{15}\text{N}$  HSQC spectra simulated according to the two-site model, as described in SI, are plotted for ligand:protein ratios of 0, 0.1, 0.2, 0.4, 0.6, 0.8, 1.25, 1.5, 1.75, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0 with contour colors that range from red to blue. Both  $n_1$  and  $n_2$  equal 1 (7). See Supporting Information for details of the simulation. In (B) and (D), blue and red arrows indicate linear and curved peak shifting, respectively. Panel (C) compares PCs from the spectra and peak lists from (B). PC1 and PC2 are normalized to the maximum amplitude of each. In the inset, each PC2 is instead normalized by the amplitude of PC1, which is symbolized by the PC2' labeling of the ordinate. (D) Measured  $^{15}\text{N}$  HSQC spectra of  $\beta$ -lactoglobulin titrated with ANS additions of 0, 1.6, 3.2, 4.8, 6.4, 8.0, 9.6, 11.2, and 12.8 mM are plotted with contour colors ranging from purple to red. While  $n_1 \approx 2$ ,  $n_2 \approx 1$  (4). Panel (E) compares PCs from the spectra and peak lists of (D). The purple symbols plot the PCA results from peak lists reported by ref (4). The plots are normalized and labeled as in (C).

simulated HSQC spectra using *trendmain* (green in Fig. V-3C) reproduces their

counterparts extracted from peak lists very well, although each component contributes

much less of the variance (38% and 15%, respectively). However, when PC2 values are

normalized by PC1, there is a systematic difference in amplitude of PC2 consistent with

its percentage of the variances listed above (inset in Fig. V-3C). This simulated two-site binding example and a number of 1:1 ligand-binding examples (9, 31) suggest that normalized PCs extracted from lists of picked peaks in the fast-exchange regime can be reproduced well by applying PCA to the series of spectra. However, PC1 and PC2 extracted by TREND from the FIDs from the simulated two-site binding example are skewed with sigmoidal deviation from the PCs obtained from either the peak lists or spectra (Fig. SV-2A,B). In the investigation of a titration of  $\beta$ -lactoglobulin with 1-anilinoaphthalene-8-sulfonate (ANS), Konuma et al. resolved two binding components using PCA of the assigned peaks from the NMR spectra of the titration (4). They observed curved trajectories (red arrows in Fig. V-3D) and linear trajectories (blue arrows in Fig. V-3D), suggesting the presence of multiple binding sites. Fast exchange behavior supported reliable PCA of the chemical shift data in peak lists, which provided binding isotherms (4). TREND extracted the PCs from the spectra (green in Fig. V-3E) and unprocessed FIDs from the titration (green in Fig. SV-2A,B). These PCs are compared with the previously reported binding isotherms (purple in Fig. V-3E). The binding populations of ref (4) are reproduced well by the normalized PC1 and PC2 derived from the spectra despite the  $t_1$ -noise present (Fig. SV-3), and less well by PC1 and PC2 obtained from the FIDs. (The residual solvent signal was subtracted on-resonance from the FIDs using the *trendmaingui* option of a convolution difference window (32). In cases of especially poor solvent suppression, this subtraction might not be enough for reliable PCs.) When choosing the form of NMR data to analyze, application of TREND directly to spectra appears to be the most consistently accurate.

Reconstruction of the spectra of the ANS titration with *trendreconstructgui* using only PC1 and PC2 introduces artifacts that are ghosts of the peaks from each spectrum of the titration (not shown). The cumulative contribution ratio (reported by *trendmaingui*) saturates at eight PCs, suggesting eight to suffice to represent the series of spectra. Using eight PCs in the reconstruction removed the ghosts of peaks and reproduced well the spectra and their biphasic trajectories of peak shifts upon additions of ANS (Fig. SV-3). The need for eight or more PCs is typical of the need for faithful reconstruction of series of spectra and images. Nonlinearity is typical of such series and spreads their variances across many PCs; see Fig. S7 of ref (9). This spreading of variances to many PCs could account for the need for many PCs for faithful reconstruction. Inspection of the reconstructed and original spectra finds both fast and fast-intermediate exchange regimes (Fig. SV-3). Application of PCA directly to the spectra, followed by reconstruction, accommodated this mixture of behaviors, as recently proposed (9).

#### **V.4.3 ICA for Confirming Components**

TREND supports optional use of ICA. If the number or significance of PCs obtained comes into question, ICA can be used to test the significance and validity of the PCs. It is also conceivable that ICA may be able to resolve components from some experiments that are not resolvable by PCA. ICA of peak pick *lists* from the two-site binding example of Fig. V-3B yields independent components (ICs) equivalent to PC1 and PC2 (Fig. SV-4). We tested ICA with various numbers  $K$  of trial components with series of *spectra* containing  $N$  true components. When  $K \leq N$ , ICA derives components that are very similar to those from PCA (Fig. SV-5). However when  $K > N$ , which means trying to extract more “independent components” than true components, ICA always fails

in our experience, as evident from components that are jagged and meaningless (Fig. SV-5E,F). Consequently, we propose that this failure of ICA can be used to count the meaningful components. The ICA should be repeated with incrementally higher  $K$  trial components. The lowest value of  $K$  at which ICA fails implies  $K - 1$  significant components; see Fig. SV-5 for two examples of the iterative process. The drawback of ICA validation of components is in repeating FastICA calculations  $N + 1$  times for each trial number of components, preferably with three to five repetitions of each, in order to escape local minima. Though the process is repetitive, it requires no prior knowledge of the number of components. Deciding the PCs that are significant may be quicker by identifying the PCs that contribute the most to scree plots (the convention) and which have large autocorrelation coefficients (smoothness) (7). However, recapitulation of PCs by ICs may engender more confidence in the reproducibility of the analysis.

#### **V.4.4 Cardiac MRI Movie Resolved into Components**

Real-time imaging by MRI generates complex movies that are suitable to showcase the capabilities of TREND. An MRI movie of a slice through the four chambers of the heart (15, 33) was analyzed by TREND. A movie for each of the first four individual PCs was reconstructed using *trendreconstructgui*, aiding interpretation of the PCs. PC1 follows the time course of breathing where the trough represents inhalation (Fig. V-4A; Movie VS-1). Fig. V-4B plots a frame from the PC2 movie (Movie VS-2) where the left ventricle is relaxed and open, known as diastole. Fig. V-4C plots a frame from the PC2 movie where the left ventricle and heart overall are contracted in systole. The time course of PC2 follows the alternation between the crests representing diastole and narrow troughs representing systole (Fig. V-4A). In the crests of PC2, the phases of

rapid filling and subsequent slower filling of the ventricles can be observed. (Ref. (34) provides an overview of the cardiac cycle). The troughs of PC3 coincide with the isovolumetric contraction phase that begins systole (Fig. V-4A). The left ventricle and atrium walls and interiors alternate in appearance in the PC3 movie (Movie VS-3). Bright density between the left ventricle and atrium in the PC3 movie at the troughs in the PC3 time course suggests the closed state of the left atrioventricular (mitral) valve. Coinciding with this is detectable rotation of the right atrium and ventricle. The PC4 movie represents sudden overall rotations of the heart (Movie VS-4). The time courses indicate synchronization of these rotations (PC4) with both the cardiac cycle (PC2) and each inspiration of a breath (PC1); see Fig. SV-6A. The rotations appear largest when a breath begins and ends. These observations illustrate the ability of TREND to resolve and aid interpretation of concurrent processes.

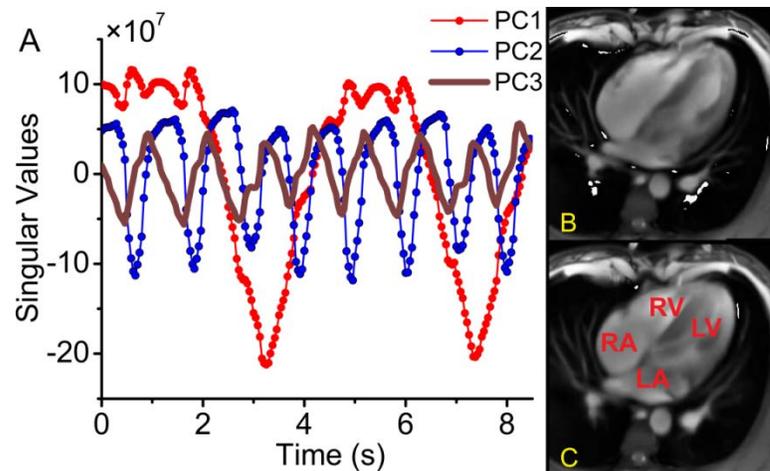


Figure V-4. SVD captures from a cardiac MRI movie the time courses of breathing, diastole and systole from a “four-chamber” angle of view. A) PC1 represents respiration. PC2 tracks the oscillation between diastole and systole. B) This frame in the reconstructed PC2 movie is during diastole with the open cavities especially evident in the ventricles (Movie VS-2). C) In this frame from the same movie, all four chambers are contracted (systole). Abbreviations: RA, right atrium; RV, right ventricle; LV, left ventricle; and LA, left atrium.

A movie reconstructed from all four of these PCs using *trendreconstructgui* captures the major morphological changes of the cardiac cycle (Movie VS-5), but is not

as smooth and nuanced as the original (15, 33). *Trendmaingui* reports autocorrelation coefficients exceeding 0.7 for the first 44 PCs, suggesting their information content. Inspection of the scree plot and the cumulative contribution plot generated by *trendmaingui* indicates that the first four PCs account for about 69% of the statistical variance across the movie, ten PCs account for 85%, and 20 account for 93% (Fig. SV-6B). Reconstruction of the cardiac MRI movie using the first 10 PCs imparts much increased realism to the depiction of the turbulent blood flow in the cardiac chambers and smoothness to the cardiac movement (Movie VS-6). Doubling the PCs to the first 20 enhances the fidelity further but more subtly (Movie VS-7). Omission of PC1 removes the largest background of breathing changes to the chest cavity, while preserving the cardiac cycle portrayal (Movie VS-8).

In reconstruction of other movies and NMR spectra, we also observed the faithfulness of the reconstruction to increase with number of PCs. Eight or more PCs may often be desirable for satisfying reconstruction of a measurement series. The scree plot and secondarily the autocorrelation coefficients appear useful for anticipating the number of PCs beneficial for reconstruction of the measurement series.

## **V.5 Conclusions**

Direct application of PCA (or ICA) to 2D measurements using TREND will expand the accessibility of equilibrium and time-evolving processes measured by spectra and imaging. No curation, selection, assignment, or resolution of specific spectral peaks or image features is necessary using this unsupervised statistical approach. TREND can be applied “on-the-fly” on an instrument host computer during data collection in order to assess if the process or reaction has progressed far enough. Multiple concurrent processes,

measured by biophysical techniques, have been readily resolved into principal or independent components. Movies and spectra can be reconstructed with TREND from the user's choice of principal components. These capabilities will introduce new convenience and insight to analyses of spectrally detected reactions and imaging-detected processes studied by biophysics, physiology, and other disciplines.

## V.6 Acknowledgments

We are grateful to K. Sakurai, T. Konuma, and Y. Goto for spectra of the ANS titration of  $\beta$ -lactoglobulin; J. Frahm and his group for real-time MRI movies; M.D. Stanley for setting up the TREND website; A.G. Roberts, K. Stiers, and reviewers for beta-testing; and Y. Fulcher for discussion of PCA. Supported by NSF grant MCB1409898.

## V.7 References

1. Williamson, M. P. 2013. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* 73:1-16.
2. Jolliffe, I. T. 2002. *Principal Component Analysis*. Springer-Verlag, New York.
3. Sakurai, K., and Y. Goto. 2007. Principal component analysis of the pH-dependent conformational transitions of bovine beta-lactoglobulin monitored by heteronuclear NMR. *Proc. Natl. Acad. Sci. U. S. A.* 104:15346-15351.
4. Konuma, T., Y. H. Lee, Y. Goto, and K. Sakurai. 2013. Principal component analysis of chemical shift perturbation data of a multiple-ligand-binding system for elucidation of respective binding mechanism. *Proteins* 81:107-118.
5. Majumder, S., C. M. DeMott, D. S. Burz, and A. Shekhtman. 2014. Using singular value decomposition to characterize protein-protein interactions by in-cell NMR spectroscopy. *ChemBioChem* 15:929-933.
6. Cembran, A., J. Kim, J. Gao, and G. Veglia. 2014. NMR mapping of protein conformational landscapes using coordinated behavior of chemical shifts upon ligand binding. *Phys Chem Chem Phys* 16:6508-6518.
7. Arai, M., J. C. Ferreon, and P. E. Wright. 2012. Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J. Am. Chem. Soc.* 134:3792-3803.
8. Cobbert, J. D., C. DeMott, S. Majumder, E. A. Smith, S. Reverdatto, D. S. Burz, K. A. McDonough, and A. Shekhtman. 2015. Caught in Action: Selecting Peptide

- Aptamers Against Intrinsically Disordered Proteins in Live Cells. *Scientific Reports* 5:9402.
9. Xu, J., and S. R. Van Doren. 2016. Binding Isotherms and Time Courses Readily from Magnetic Resonance. *Anal. Chem.* 88:8172-8178.
  10. Shlens, J. 2014. A Tutorial on Independent Component Analysis. In ArXiv e-prints.
  11. Yao, F., J. Coquery, and K.-A. Lê Cao. 2012. Independent Principal Component Analysis for biologically meaningful dimension reduction of large biological data sets. *BMC Bioinformatics* 13:24.
  12. Nuzillard, D., S. Bourg, and J. M. Nuzillard. 1998. Model-free analysis of mixtures by NMR using blind source separation. *J. Magn. Reson.* 133:358-363.
  13. Ladroue, C., F. A. Howe, J. R. Griffiths, and A. R. Tate. 2003. Independent component analysis for automated decomposition of in vivo magnetic resonance spectra. *Magn. Reson. Med.* 50:697-703.
  14. Monakhova, Y. B., A. M. Tsikin, T. Kuballa, D. W. Lachenmeier, and S. P. Mushtakova. 2014. Independent component analysis (ICA) algorithms for improved spectral deconvolution of overlapped signals in H-1 NMR analysis: application to foods and related products. *Magn. Reson. Chem.* 52:231-240.
  15. Zhang, S., A. A. Joseph, D. Voit, S. Schaez, K.-D. Merboldt, C. Unterberg-Buchwald, A. Hennemuth, J. Lotz, and J. Frahm. 2014. Real-time magnetic resonance imaging of cardiac function and flow—Recent progress. *Quantitative Imaging in Medicine and Surgery* 4:313-329.
  16. Xu, J., S. R. Van Doren. <https://trendmizzou.gitbooks.io/trend-manual/content/>. 2016.
  17. Xu, J., S. R. Van Doren. <http://biochem.missouri.edu/trend>. 2016.
  18. Helmus, J. J., and C. P. Jaroniec. 2013. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. *J. Biomol. NMR* 55:355-367.
  19. Goddard, T. D., and D. G. Kneller. 2000. SPARKY. University of California, San Francisco, San Francisco.
  20. Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6:277-293.
  21. Lee, W., M. Tonelli, and J. L. Markley. 2015. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31:1325-1327.
  22. Cavanagh, J., W. J. Fairbrother, A. G. Palmer Iii, and N. J. Skelton. 2007. Preface to the First Edition. In *Protein NMR Spectroscopy (Second Edition)*. J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, and N. J. Skelton, editors. Academic Press, Burlington. vii-x.
  23. van den Berg, R. A., H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde, and M. J. van der Werf. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 7:142.
  24. Selvaratnam, R., S. Chowdhury, B. VanSchouwen, and G. Melacini. 2011. Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc. Natl. Acad. Sci. U. S. A.* 108:6133-6138.
  25. <http://scikit-learn.org/stable/modules/decomposition.html#ica>. Independent component analysis (ICA).

26. Hyvärinen, A., and E. Oja. 2000. Independent component analysis: algorithms and applications. *Neural Networks* 13:411-430.
27. Hyvärinen, A., J. Karhunen, and E. Oja. 2002. Practical Considerations. In *Independent Component Analysis*. John Wiley & Sons, Inc. 269-271.
28. Särelä, J., and R. Vigario. 2004. Overlearning in marginal distribution-based ICA: Analysis and solutions. *J Mach Learn Res* 4:1447-1469.
29. <https://mran.microsoft.com/packages/>.
30. Moore, B., <https://www.mathworks.com/matlabcentral/fileexchange/38300-pca-and-ica-package>.
31. Jolliffe, I. T. 2002. *Principal Component Analysis*. Springer-Verlag, New York.
32. Marion, D., M. Ikura, and A. Bax. 1989. Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data. *Journal of Magnetic Resonance* (1969) 84:425-430.
33. [http://www.biomednmr.mpg.de/index.php?option=com\\_content&task=view&id=132&Itemid=39#IRT](http://www.biomednmr.mpg.de/index.php?option=com_content&task=view&id=132&Itemid=39#IRT). Real-time MRI. Biomedizinische NMR Forschungs GmbH am Max-Planck-Institut für biophysikalische Chemie, Göttingen.
34. Klabunde, R. E., <http://www.cvphysiology.com/Heart%20Disease/HD002b.htm>. *Cardiovascular Physiology Concepts*. Indianapolis.

## V.8 Supporting material

### V.8.1 Supporting methods

#### V.8.1.1 2D Spectral Simulations of Two Site Ligand Binding

$^{15}\text{N}$  HSQC spectra were simulated with Gaussian line shapes using NmrGlue (1). The total [protein],  $P_t$ , was set to  $50\ \mu\text{M}$  and  $K_{D1}$  and  $K_{D2}$  to  $32$  and  $10.2\ \mu\text{M}$ , respectively, per ref (2). These simulations in fast exchange (Fig. V-3A,B) are similar to simulations of 1:1 ligand binding; see SI in ref (3). However, free, intermediate, and bound populations were simulated using Equations V-3 to V-5 given in ref (2). Chemical shifts of peaks were calculated as population weighted averages of the simulated free, fully-bound, and two hidden intermediate states. Peaks of the free and fully-bound states were modeled after the FHA domain (BMRB: 5564) and its peptide-bound form. Peak positions of hidden intermediate states were generated by random perturbation from the linear path connecting the pairs of free and bound states. Gaussian noise was added to each spectrum to set the S/N to 5.0 for peaks of median intensities in the free state spectrum (3). Linewidths were also randomly perturbed by 5% in both dimensions. The corresponding FIDs were calculated by inverse Fourier transformation of simulated spectra.

### V.8.1.2 Supporting tables

Table SV-1. Types of biophysical measurements that PCA has interpreted successfully. The purposes of the PCA usages are indicated by the letter symbols defined in the legend. These sources of data are amenable to analysis by TREND. Most of the techniques can supply files in JCAMP-DX, spreadsheet, or text format, each readable by TREND. Images from microscopy and other imaging methods can be read into TREND either as a series of PNG images or in one of several movie formats listed in Table V-2. NMR spectra are read in one of the formats listed in Table V-2.

<b>Biophysical Technique</b>	<b>Objective of PCA</b>	<b>References</b>
<b>Spectroscopy</b>		
<b>Circular Dichroism</b>	A, B	(4, 5)
<b>Electron Paramagnetic Resonance</b>	B	(6)
<b>Electronic Absorption</b>	B	(7)
<b>Fluorescence, fluorescence correlation</b>	A, B	(4, 8, 9)
<b>Linear Dichroism</b>	B	(10)
<b>Magnetic Optical Rotatory Dispersion</b>	B	(11)
<b>Multi-wavelength</b>	B	(12)
<b>Nonlinear IR</b>	F	(13)
<b>Raman</b>	H, I	(14)
<b>UV-Vis</b>	B, D, E, G	(15-17)
<b>X-ray absorption</b>	A	(18)
<b>Imaging</b>		
<b>Atomic Force Microscopy</b>	B, G	(19)
<b>Electron Energy Loss Spectrum</b>	C	(20)
<b>Fluorescence</b>	B, G, H	(21, 22)
<b>Functional Magnetic Resonance Imaging</b>	B, D	(23)
<b>Raman</b>	C	(24)
<b>Scattered-Light</b>	G	(25)
<b>Transmission Electron Microscopy</b>	C	(26)
<b>X-ray scattering techniques</b>		
<b>Small-angle X-ray Scattering</b>	B, D	(27, 28)
<b>Wide-angle X-ray Scattering</b>	B	(29)
<b>X-ray Free Electron Laser</b>	C	(30)
<b>X-ray Diffraction</b>	B	(31, 32)
<b>Other techniques</b>		
<b>Chromatographic Data</b>	H	(33)
<b>Differential Scanning Calorimetry</b>	B, I	(34)
<b>Mass Spectrometry</b>	I	(35)

A: Plot concentration-dependent curve  
 C: Noise filtering  
 E: Find pH dependent curve  
 G: Extract important components  
 I: Classification

B: Plot time course  
 D: Determine meaningful components  
 F: Plot temperature-dependent curve  
 H: Find basis data set (deconvolution)

Table SV-2. Scaling options that TREND supports. Element  $x_{ij}$  represents the data point in the  $i_{th}$  row and  $j_{th}$  column of matrix  $X'$ .  $\tilde{x}_{ij}$  stands for scaled value of  $x_{ij}$ ,  $\bar{x}_i$  and  $s_i$  represent the mean value and standard deviation, respectively, of the  $i_{th}$  row.  $\bar{x}_j$  and  $s_j$  represent mean value and standard deviation, respectively, of the  $j_{th}$  column. Van Den Berg et al. define these formulae and discuss usage of scaling (36).

Scaling option	Definition	Formula (row scaling)	Formula (column scaling)
none	<i>do nothing</i>	$\tilde{x}_{ij} = x_{ij}$	$\tilde{x}_{ij} = x_{ij}$
noscaling	<i>Centering</i>	$\tilde{x}_{ij} = x_{ij} - \bar{x}_i$	$\tilde{x}_{ij} = x_{ij} - \bar{x}_j$
auto	<i>Autoscaling</i>	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i}$	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$
pareto	<i>Pareto scaling</i>	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}}$	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{s_j}}$
vast	<i>Vast scaling</i>	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i} \cdot \frac{\bar{x}_i}{s_i}$	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \cdot \frac{\bar{x}_j}{s_j}$
range*	<i>Range scaling</i>	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{x_{i_{max}} - x_{i_{min}}}$	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{x_{j_{max}} - x_{j_{min}}}$
level	<i>Level scaling</i>	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\bar{x}_i}$	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\bar{x}_j}$

\* *trendreconstruct* does not support reconstruction of dataset that was pretreated with *Range scaling*.

## V.8.2 Supporting figures

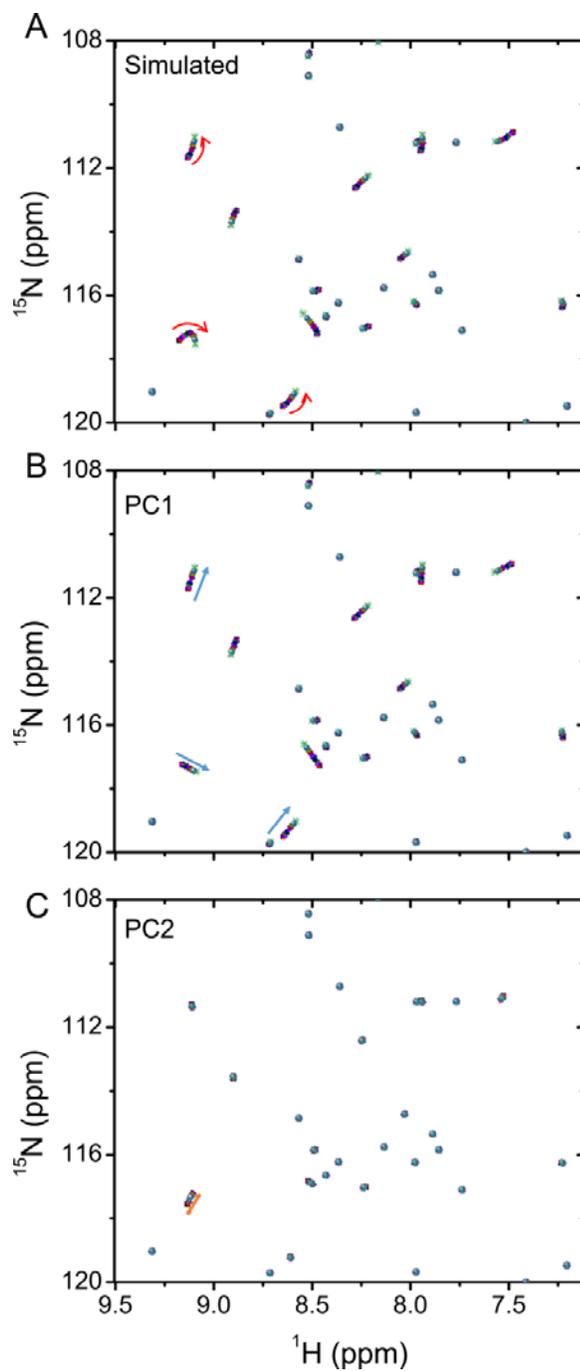


Figure SV-1. Peak pick lists from the simulated spectra of Fig. V-3B were reconstructed by TREND and plotted as circles. A) The peak lists from the simulated two-site ligand-binding titrations of Fig. V-3B are plotted and overlaid. TREND's reconstruction of the PC1 and PC2 portions of the peak lists are plotted in (B) and (C), respectively.

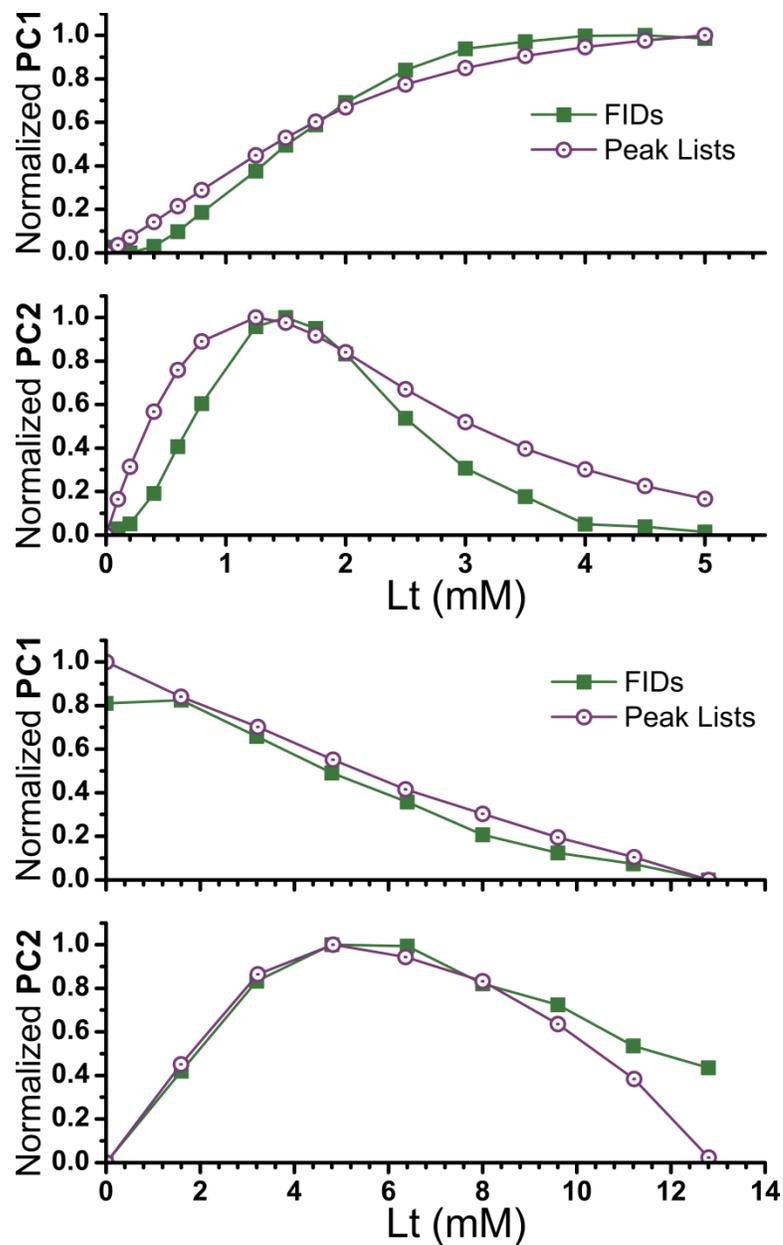


Figure SV-2. SVD of the unprocessed FIDs from the two site binding titrations identifies two components similar to those from the spectra or the peak lists. (Upper panels) PC1 and PC2 derive from the simulated spectra of Fig. V-3B, c.f. Fig. V-3C. (Lower panels) PC1 and PC2 derive from Fig. V-3D regarding ANS binding titrations, c.f. Fig. V-3E.

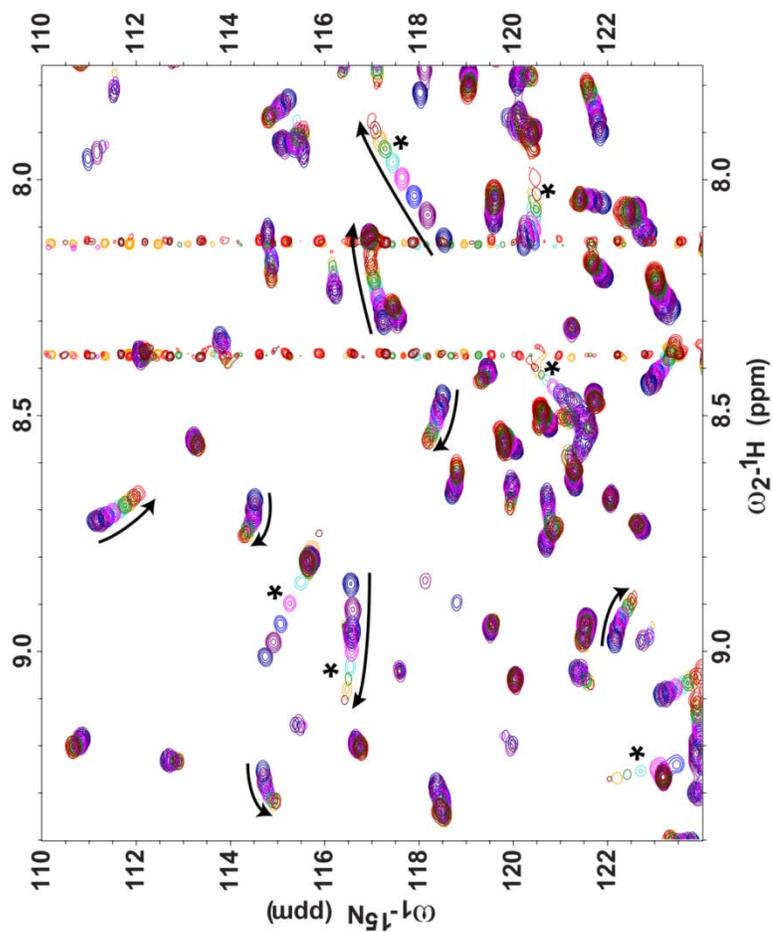


Figure SV-3. Reconstruction with eight PCs using *trendreconstructgui* reproduces the spectra of Konuma et al. (2013) for the titration of bovine  $\beta$ -lactoglobulin with ANS (37). Arrows mark part of the curved trajectories of peak shifts upon succession of additions of ANS from blue contours at no addition to red contours at 12.8 mM ANS. Some instances of fast-intermediate exchange broadening are pointed out with an asterisk (\*).

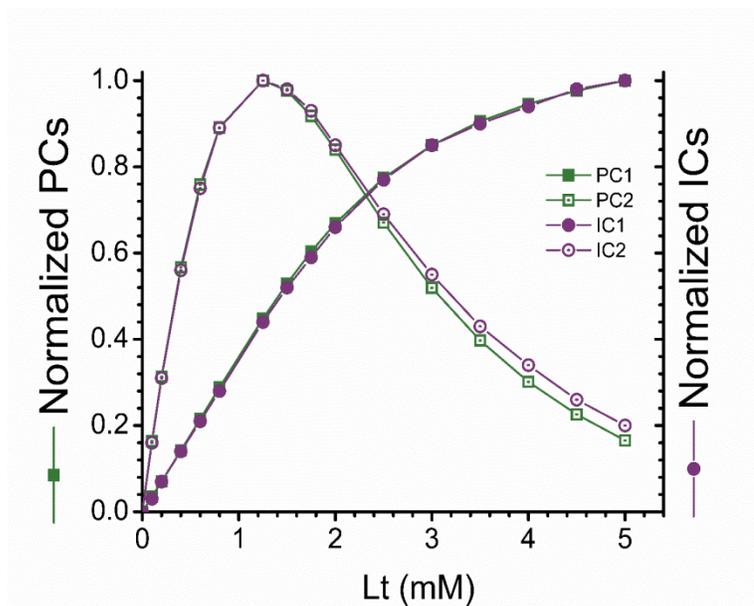


Figure SV-4. ICA extracts binding components nearly identical to those from PCA when the number of components is chosen correctly. IC1, IC2, PC1, and PC2 obtained from peak lists from the simulated two-site binding titration of Fig. V-3B.

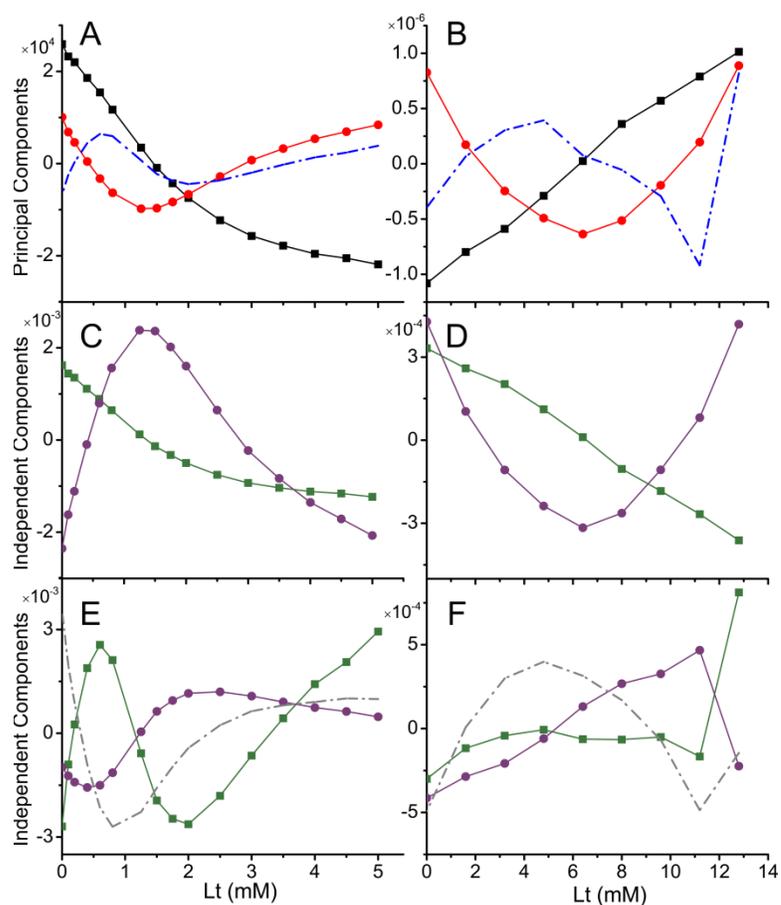


Figure SV-5. Validating the number of significant principal components using ICA. The examples have two components. The failure of ICA suggests the independent components that are meaningful in titrations. (A, C, E) report on the simulated two site binding events of Fig. V-3B. (B, D, F) report on ANS binding in Fig. V-3D. The first row (A, B) colors PC1 black, PC2 red, and PC3 with blue dashed lines. (C, D) plot the IC1 and IC2 (green or purple, respectively) when calculating *two* independent components (inherently unordered). The third row (E, F) shows computed IC1 to IC3. These three ICs have lost the shape of IC1 and IC2 when only two ICs were computed (panels D to I). This loss of the shapes of IC1 and IC2 indicates the failure of ICA when choosing three components, implying that only two components are significant and needed.

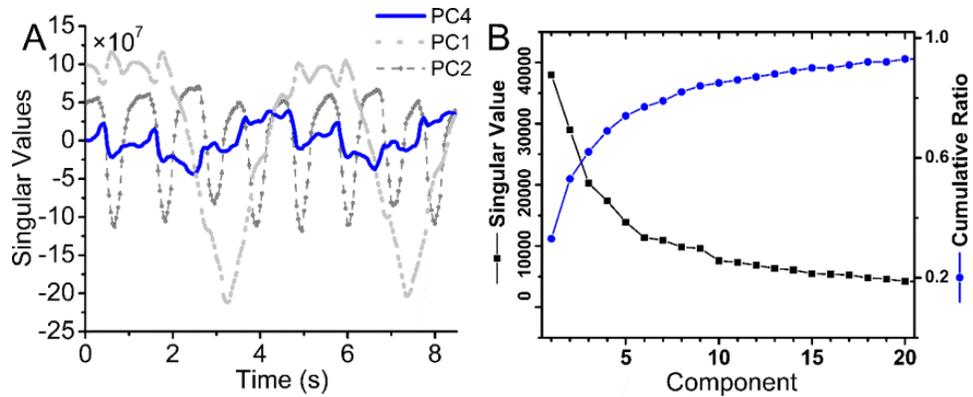


Figure SV-6. (A). PC4 (blue) from the cardiac MRI movie of Fig. V-4 reports sudden overall rotations of the heart which synchronize with both the cardiac cycle (gray dashed curve) and each inspiration of a breath (gray dotted curve). (B) The scree plot for the analysis of Fig. V-4 is black. The cumulative contribution ratio is blue.

### V.8.3 Supporting movies

Movie V-S1. Movie reconstructed from PC1 from the cardiac MRI movie of Fig. V-4 (38), showing the effects of breathing on the chest cavity.

Movie V-S2. Movie reconstructed from PC2 from the cardiac movie of Fig. V-4, showing cycling between contraction (systole) and relaxation (diastole).

Movie V-S3. Movie reconstructed from PC3 from the cardiac movie of Fig. V-4, which is likely to show isovolumetric changes of the cardiac cycle.

Movie V-S4. Movie reconstructed from PC4 from the cardiac movie of Fig. V-4, revealing sudden overall rotations of the heart.

Movie V-S5. Movie reconstructed from PCs 1 through 4 calculated from the cardiac MRI movie.

Movie V-S6. Movie reconstructed from PCs 1 through 10 from the cardiac MRI movie.

Movie V-S7. Movie reconstructed from PCs 1 through 20 from the cardiac MRI movie.

Movie V-S8. Movie reconstructed from PCs 2 through 20 from the cardiac MRI movie.

#### V.8.4 Supporting references

1. Helmus, J. J., and C. P. Jaroniec. 2013. Nmrplug: an open source Python package for the analysis of multidimensional NMR data. *J. Biomol. NMR* 55:355367.
2. Arai, M., J. C. Ferreon, and P. E. Wright. 2012. Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J. Am. Chem. Soc.* 134:3792-3803.
3. Xu, J., and S. R. Van Doren. 2016. Binding Isotherms and Time Courses Readily from Magnetic Resonance. *Anal. Chem.* 88:8172-8178.
4. Gualfetti, P. J., O. Bilsel, and C. R. Matthews. 1999. The progressive development of structure and stability during the equilibrium folding of the alpha subunit of tryptophan synthase from *Escherichia coli*. *Protein Sci.* 8:1623-1635.
5. Ruther, A., M. Pfeifer, V. A. Lorenz-Fonfria, and S. Ludeke. 2014. Reaction Monitoring Using Mid-Infrared Laser-Based Vibrational Circular Dichroism. *Chirality* 26:490-496.
6. Kakitani, Y., R. Fujii, Y. Koyama, H. Nagae, L. Walker, B. Salter, and A. Angerhofer. 2006. Triplet-state conformational changes in 15-cis-spheroidene bound to the reaction center from *Rhodobacter sphaeroides* 2.4.1 as revealed by time-resolved EPR spectroscopy: strengthened hypothetical mechanism of triplet-energy dissipation. *Biochemistry* 45:2053-2062.
7. Kim-Shapiro, D. B., S. B. King, C. L. Bonifant, C. P. Kolibash, and S. K. Ballas. 1998. Time resolved absorption study of the reaction of hydroxyurea with sickle cell hemoglobin. *Biochim. Biophys. Acta* 1380:64-74.
8. Isin, E. M., and F. P. Guengerich. 2007. Multiple sequential steps involved in the binding of inhibitors to cytochrome p450 3A4. *J. Biol. Chem.* 282:6863-6874.
9. Frank, G. A., M. Goomanovsky, A. Davidi, G. Ziv, A. Horovitz, and G. Haran. 2010. Out-of-equilibrium conformational cycling of GroEL under saturating ATP concentrations. *Proc. Natl. Acad. Sci. U. S. A.* 107:6270-6274.
10. Shapiro, D. B., R. M. Esquerra, R. A. Goldbeck, S. K. Ballas, N. Mohandas, and D. S. Kligler. 1996. A study of the mechanisms of slow religation to sickle cell hemoglobin polymers following laser photolysis. *J. Mol. Biol.* 259:947-956.
11. Esquerra, R. M., R. A. Goldbeck, D. B. Kim-Shapiro, and D. S. Kligler. 1998. Spectroscopic evidence for nanosecond protein relaxation after photodissociation of myoglobin-CO. *Biochemistry* 37:17527-17536.
12. Hendler, R. W., S. K. Bose, and R. I. Shrager. 1993. Multiwavelength analysis of the kinetics of reduction of cytochrome aa3 by cytochrome c. *Biophys. J.* 65:1307-1317.
13. Chung, H. S., M. Khalil, and A. Tokmakoff. 2004. Nonlinear infrared spectroscopy of protein conformational change during thermal unfolding. *J. Phys. Chem. B* 108:15332-15342.
14. Uy, D., and A. E. O'Neill. 2005. Principal component analysis of Raman spectra from phosphorus-poisoned automotive exhaust-gas catalysts. *J Raman Spectrosc* 36:988-995.

15. Zapata, A. L., M. R. Kumar, D. Pervitsky, and P. J. Farmer. 2013. A singular value decomposition approach for kinetic analysis of reactions of HNO with myoglobin. *J. Inorg. Biochem.* 118:171-178.
16. Hendriks, J., and K. J. Hellingwerf. 2009. pH Dependence of the Photoactive Yellow Protein Photocycle Recovery Reaction Reveals a New Late Photocycle Intermediate with a Deprotonated Chromophore. *J. Biol. Chem.* 284:5277-5288.
17. J. C. Martínez, N. A. C., J. L. González, T. Cordova. 2012. Alternative Metodology for Gold Nanoparticles Diameter Characterization Using PCA Technique and UV-VIS Spectrophotometry. *Nanoscience and Nanotechnology* 2:184-189.
18. Wasserman, S. R., P. G. Allen, D. K. Shuh, J. J. Bucher, and N. M. Edelstein. 1999. EXAFS and principal component analysis: a new shell game. *J Synchrotron Radiat* 6:284-286.
19. Kalinin, S. V., B. J. Rodriguez, S. Jesse, A. N. Morozovska, A. A. Bokov, and Z. G. Ye. 2009. Spatial distribution of relaxation behavior on the surface of a ferroelectric relaxor in the ergodic phase. *Appl. Phys. Lett.* 95.
20. Lichtert, S., and J. Verbeeck. 2013. Statistical consequences of applying a PCA noise filter on EELS spectrum images. *Ultramicroscopy* 125:35-42.
21. Seo, J., Y. An, J. Lee, T. Ku, Y. Kang, C. Ahn, and C. Choi. 2016. Principal component analysis of dynamic fluorescence images for diagnosis of diabetic vasculopathy. *Journal of biomedical optics* 21:46003.
22. Cohen, A. E., and W. E. Moerner. 2007. Principal-components analysis of shape fluctuations of single DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* 104:12622-12627.
23. Hansen, L. K., J. Larsen, F. Å. Nielsen, S. C. Strother, E. Rostrup, R. Savoy, N. Lange, J. Sidtis, C. Svarer, and O. B. Paulson. 1999. Generalizable Patterns in Neuroimaging: How Many Principal Components? *NeuroImage* 9:534-544.
24. Hashimoto, A., Y. Yamaguchi, L. D. Chiu, C. Morimoto, K. Fujita, M. Takedachi, S. Kawata, S. Murakami, and E. Tamiya. 2015. Time-lapse Raman imaging of osteoblast differentiation. *Sci Rep-Uk* 5.
25. Rector, D. M., R. F. Rogers, J. S. Schwaber, R. M. Harper, and J. S. George. 2001. Scattered-light imaging in vivo tracks fast and slow processes of neurophysiological activation. *NeuroImage* 14:977-994.
26. Furnival, T., R. K. Leary, and P. A. Midgley. 2016. Denoising time-resolved microscopy image sequences with singular value thresholding. *Ultramicroscopy.*
27. Kim, T. W., C. Yang, Y. Kim, J. G. Kim, J. Kim, Y. O. Jung, S. Jun, S. J. Lee, S. Park, I. Kosheleva, R. Henning, J. J. van Thore, and H. Ihee. 2016. Combined probes of X-ray scattering and optical spectroscopy reveal how global conformational change is temporally and spatially linked to local structural perturbation in photoactive yellow protein. *PCCP* 18:8911-8919.
28. Perez, J., P. Vachette, D. Russo, M. Desmadril, and D. Durand. 2001. Heatinduced unfolding of neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. *J. Mol. Biol.* 308:721-743.
29. Malmerberg, E., Z. Omran, J. S. Hub, X. W. Li, G. Katona, S. Westenhoff, L. C. Johansson, M. Andersson, M. Cammarata, M. Wulff, D. van der Spoel, J. Davidsson, A. Specht, and R. Neutze. 2011. Time-Resolved WAXS Reveals

- Accelerated Conformational Changes in Iodoretinal-Substituted Proteorhodopsin. *Biophysical journal* 101:1345-1353.
30. Haldrup, K. 2014. Singular value decomposition as a tool for background corrections in time-resolved XFEL scattering data. *Philos T R Soc B* 369.
  31. Boetker, J. P., J. Rantanen, L. Arnfast, M. Doreth, D. Rajjada, K. Loebmann, C. Madsen, J. Khan, T. Rades, A. Mullertz, A. Hawley, D. Thomas, and B. J. Boyd. 2016. Anhydrate to hydrate solid-state transformations of carbamazepine and nitrofurantoin in biorelevant media studied in situ using time-resolved synchrotron X-ray diffraction. *Eur. J. Pharm. Biopharm.* 100:119-127.
  32. Oka, T., N. Yagi, T. Fujisawa, H. Kamikubo, F. Tokunaga, and M. Kataoka. 2000. Time-resolved x-ray diffraction reveals multiple conformations in the M-N transition of the bacteriorhodopsin photocycle. *Proc. Natl. Acad. Sci. U. S. A.* 97:14278-14282.
  33. Macnaughtan, D., L. B. Rogers, and G. Wernimont. 1972. Principal-component analysis applied to chromatographic data. *Analytical Chemistry* 44:1421-1427.
  34. Maggio, R. M., L. Cerretani, C. Barnaba, and E. Chiavaro. 2012. Application of Differential Scanning Calorimetry-Chemometric Coupled Procedure to the Evaluation of Thermo-Oxidation on Extra Virgin Olive Oil. *Food Biophys* 7:114123.
  35. Idborg, H., P. O. Edlund, and S. P. Jacobsson. 2004. Multivariate approaches for efficient detection of potential metabolites from liquid chromatography/mass spectrometry data. *Rapid Commun. Mass Spectrom.* 18:944-954.
  36. van den Berg, R. A., H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde, and M. J. van der Werf. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 7:142.
  37. Konuma, T., Y. H. Lee, Y. Goto, and K. Sakurai. 2013. Principal component analysis of chemical shift perturbation data of a multiple-ligand-binding system for elucidation of respective binding mechanism. *Proteins* 81:107-118.
  38. [http://www.biomednmr.mpg.de/index.php?option=com\\_content&task=view&id=132&Itemid=39#IRT](http://www.biomednmr.mpg.de/index.php?option=com_content&task=view&id=132&Itemid=39#IRT). Real-time MRI. Biomedizinische NMR Forschungs GmbH am Max-Planck-Institut für biophysikalische Chemie, Göttingen.

## VITA

Jia Xu was born and grew up in Jiangsu, China. He enrolled in the graduate program at the University of Missouri to pursue a PhD degree in Biochemistry under the guidance of Dr. Steven R. Van Doren and Dr. Lesa J. Beamer. Prior to beginning his doctorate studies, Xu served as an NMR assistant manager at Renmin University of China. He earned a master of science in chemistry from Renmin University of China (advisor Dr. Weihong Du) and a bachelor's degree in chemistry from Nanjing University (advisor Dr. Hongbin Du). In September 2016, Jia Xu completed his doctoral studies.