APPLICATION OF HYDROGEN DEUTERIUM EXCHANGE MASS SPECTROMETRY IN PROTEIN-LIGAND AND PROTEIN-PROTEIN INTERACTIONS

by

SIQI GUAN

PATRICK A. FRANTOM, COMMITTEE CHAIR
JACK DUNKLE
LAURA REED
JOHN B. VINCENT
STEPHEN A. WOSKI

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Chemistry
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2016

ABSTRACT

Proteins are not static objects. They have a great variety of internal motions with different amplitudes and different timescales. These internal motions play an important role in catalytic processes. Therefore, the existence of an intimate relationship between protein dynamics and protein function is widely accepted. Due to the significance of protein dynamics, techniques have been developed to study protein dynamics including nuclear magnetic resonance (NMR) spectroscopy, electron paramagnetic resonance (EPR) spectroscopy, and mass spectrometry (MS). Hydrogen-deuterium exchange mass spectrometry (HDX-MS), a combination of HPLC and MS, has become a common and sensitive tool to probe protein structural flexibility and solution dynamics.

In this dissertation, HDX-MS was applied to study dynamic changes of proteins due to substrate binding and protein-protein interactions. The GT-A glycosyltransferase glucosyl-3-phosphoglycerate synthase from *Mycobacterium tuberculosis* (*Mt*GpgS) catalyzes the first step of biosynthesis of 6-*O*-methylglucose lipopolysaccharides (MGLPs). The HDX-MS data revealed that the two substrates UDP-glucose (UDPG) and 3-phosphoglycerate (3PGA) can bind to *Mt*GpgS independently, disagreeing with the previous proposal that 3PGA can only bind to *Mt*GpgS after UDPG. Moreover, 3PGA was found to bind to or allosterically affect the UDPG binding site.The GT-B glycosyltransferase MshA from *Corynebacterium glutamicum* (*Cg*MshA) catalyzes the initial step of mycothiol biosynthesis. A large conformational change was observed in *Cg*MshA on nucleotide binding by superimposing APO structure of *Cg*MshA and complex structure with UDP. HDX-MS was utilized to study conformational changes of

*Cg*MshA on substrate binding from the aspect of dynamics, providing a complementary to static structures. HDX-MS was also employed to study dynamic changes of protein complex $SufBC_2D$ from *Escherichia coli* on ADP/Mg^{2+} binding. The crystal structure of $SufBC_2D$ complex has been determined, while little dynamic information is known. So HDX-MS was applied to study dynamic changes of the $SufBC_2D$ complex. The HDX-MS data revealed that SufC has a significant conformational change, which may be required by ATP binding and hydrolysis. Moreover, SufB and SufD are detected to have dynamic changes due to SufC conformational changes. These dynamic changes suggest that SufB-SufD protomer may have a conformational change in order to provide a suitable conformation for Fe-S cluster assembly.

DEDICATION

I dedicate my dissertation work to my family and close friends. Their teachings, encouragements, and prayers kept me going through the challenges of graduate school and life. I also dedicate this dissertation to my husband, Shian, for providing me motivation and support.

LIST OF ABBREVIATIONS AND SYMBOLS

[Fe-S]	iron sulfur cluster
µL	microliter
µM	micromolar
3PGA	3-phosphoglycerate
A	alanine
ABC	ATP-binding cassette
ACN	acetonitrile
ADP	adenosine diphosphate
Af_INO	inositol-1-phosphate from archaeoglobus fulgidus
AFM	atomic force microscopy
Ala	alanine
Arg	arginine
Asp	aspartate
ATP	adenosine triphosphate
ATR-FTIR	attenuated total reflection fourier transform infrared spectroscopy
BME	β-mercaptoethanol
C	cysteine
cAMP	cyclic adenosine monophosphate
CID	collision induced dissociation
cGMP	cyclic guanosine monophosphate

C. glutamicum	*Corynebacterium glutamicum*
*Cg*MshA	MshA from *Corynebacterium glutamicum*
Cys	cysteine
D	aspartate
DHF	dihydrofolate
DHFR	dihydrofolate reductase
DNase	deoxyribonuclease
D_2O	deuterium oxide
DTT	dithiothreitol
E	glutamate
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EPR	electron paramagnetic resonance
ESI	electrospray ionization
FA	formic acid
Fe	iron
g	gram
G	glycine
GG	glucosylglycerate
GlcNAc-Ins-P	3-phospho-1-D-myo-inosityl-2-acet-amido-2-deoxy-α-D-glucopyranoside
Glu	glutamic acid
Gln	glutamine
GPG	glucosyl-3-phosphoglycerate

GpgP	glucosyl-3-phosphoglycerate phosphatase
GpgS	glucosyl-3-phosphoglycerate synthase
GT	glycosyltransferase
H	histidine
HDX	hydrogen deuterium exchange
HDX-MS	hydrogen deuterium exchange-mass spectrometry
His	histidine
HPLC	high performance liquid chromatography
Ile	isoleucine
I1P	1-L-myo-inositol-1-phosphate
IR	infrared spectroscopy
Isc	iron sulfur cluster
ITC	isothermal titration calorimetry
K	lysine
L	leucine
LB	Luira-Bertani (medium)
LDH	lactate dehydrogenase
Lys	lysine
m/z	mass to charge ratio
$m_{0\%}$	water control
$m_{100\%}$	full deuteration control
MGLP	6-*O*-methylglucose lipopolysaccharide
$min^{-1}mM^{-1}$	per minute per millimolar

mM	millimolar
MMP	3-*O*-methylmannose polysaccharide
MS	mass spectrometry
MSH	mycothiol
M. tuberculosis	*Mycobacterium tuberculosis*
*Mt*GpgS	glucosyl-3-phosphoglycerate synthase from *Mycobacterium tuberculosis*
NADH	nicotinamide adenine dinucleotide
$(NH_4)_2SO_4$	ammonium sulfate
nm	nanometer
NMR	nuclear magnetic resonance
Nif	nitrogen fixation
OD_{600}	optical density at 600 nm
PCR	polymerase chain reaction
PDB	protein database
PDE	phosphodiesterase
PEP	phosphoenolpyruvate
PLP	pyridoxal-phosphate
PMPS	polymethylated polysaccharides
PMSF	phenylmethanesulfonyl fluoride
PPIs	protein protein interactions
PK	pyruvate kinase
Q	glutamine
R	arginine

RT	retention time
S	sulfur
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Ser	serine
SMC	structural maintenance of chromosome
Suf	sulfur formation
T	threonine
Tcep	tris (2-carboxyethyl) phosphine
TEA	triethanolamine
THF	tetrahydrofolate
Thr	threonine
Tris	tris (hydroxymethyl) aminomethane
Tyr	tyrosine
UDP	uridine diphosphate
UDPG	uridine diphosphate glucose
UDP-GlcNAC	uridine diphosphate glucose-*N*-acetylglucosamine
V	valine
Val	valine
Y	tyrosine
z	charge

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Patrick A. Frantom for the continuous support of my Ph.D research and study. His guidance and encouragement helped me during the research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

I would like to thank my committee members Dr. Jack Dunkle, Dr. Laura Reed, Dr. John B. Vincent, and Dr. Stephen A. Woski, for their valuable advice and suggestion throughout my graduate career. Also, I would like to thank Dr. Carolyn J. Cassady as my previous committee member for advice in my IRR and ORP.

Especially I would like to thank and acknowledge the late Dr. Laura S. Busenlehner for her guidance and valuable advice in the hydrogen deuterium exchange mass spectrometry technique.

The financial support of National Institute of Health is gratefully acknowledged. I also appreciate the support from The University of Alabama and Department of Chemistry. I also would like to thank my past and present members of Frantom Laboratory, Dr. Ashley Casey, Dr. Garima Kumar, Jordyn Johnson, Dokyong Kim, Wen Chen, and Juliana Conte for their continuous support and encouragement.

I also would like to thank Dr. Qiaoli Liang for her training and assistance in high performance liquid chromatography and mass spectrometer. I am also thankful to Dr. Wayne Outten (University of South Carolina, Columbia) and Guangchao Dong for providing us protein stocks that enabled us to work on the project.

Finally, I would like to thank my wonderful parents for their blessings and support. I would like to thank my husband, Shian, for his unending love and patience. I would also like to thank my brother, Haoyu, for his love and support. Lastly, I would acknowledge all my friends for always being here with me.

CONTENTS

ABSTRACT ii

DEDICATION iv

LIST OF ABBREVIATIONS v

ACKNOWLEDGEMENTS x

LIST OF TABLES xvii

LIST OF FIGURES xviii

LIST OF ILLUSTRATIONS xxi

CHAPTER 1 INTRODUCTION 1

1.1 Significance of Protein Dynamics 1

1.2 Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) 2

1.2.1 Backbone Amide Hydrogens 4

1.2.2 Linderstrom-Lang Model 5

1.2.3 Factors Affecting H/D Exchange Rate 7

1.2.4 Procedures of HDX-MS Experiments 7

1.2.5 Analysis of HDX-MS Data 11

1.3 HDX-MS Studies of Protein-Ligand and Protein-Protein Interactions 16

1.4 Glycosyltransferase Enzymes 18

1.4.1 Retaining and Inverting Glycosyltransferase 18

1.4.2 GT-A and GT-B fold 20

1.5 Suf System Assembling Fe-S Clusters 22

1.5.1 Iron-Sulfur Clusters 22

1.5.2 Fe-S Cluster Biogenesis Pathways.....22

1.6 HDX-MS Applied in Dynamics of Protein and Protein Complex..... 24

1.7 References..... 26

CHAPTER 2 INSIGHTS INTO STRUCTURE AND DYNAMICS OF RETAINING GLYCOSYLTRANSFERASE GLUCOSYL-3-PHOSPHOGLYCERATE SYNTHASE FROM *MYCOBACTERIUM TUBERCULOSIS* 34

2.1 Introduction..... 34

2.1.1 Significance of MGLPs for *Mycobacterium tuberculosis* 34

2.1.2 Biosynthesis of MGLPs 36

2.1.3 Classification of *Mt*GpgS.....37

2.1.4 Crystal Structure of *Mt*GpgS 37

2.1.5 Conclusion 39

2.2 Materials and Methods..... 40

2.2.1 Materials 40

2.2.2 Amplification and Overexpression of *Mt*GpgS 40

2.2.3 Purification of *Mt*GpgS.....40

2.2.4 Site-directed Mutagenesis of *Mt*GpgS 41

2.2.5 Steady State Assays 41

2.2.6 Determination of Kinetic Parameters.....42

2.2.7 Sequencing of *Mt*GpgS by MS/MS 43

2.2.8 Continuous HDX-MS Experiments 44

2.2.9 Analysis of HDX Data 46

2.3 Results..... 46

2.3.1 Steady State Kinetics of Wild Type *Mt*GpgS and Variants 46

2.3.2 Peptide Map of *Mt*GpgS 49

2.3.3 Solvent Accessibility Changes on Substrate Binding 50

2.3.4 Dynamic Changes of *Mt*GpgS on UDPG Binding 54

2.3.5 Dynamic Changes of *Mt*GpgS on 3PGA Binding 54

2.4 Discussion 59

2.4.1 Roles of Active Site Residues in *Mt*GpgS 59

2.4.2 Insight into Dynamic Changes of *Mt*GpgS 61

2.4.3 Substrate Binding Mechanism of *Mt*GpgS 62

2.4.4 Role of Loop Flexibility in Substrate Recognition and Binding 64

2.5 References 66

CHAPTER 3 INSIGHTS INTO CONFORMATIONAL CHANGES OF RETAINING GLYCOSYLTRANSFERASE MSHA FROM *CORYNEBACTERIUM GLUTAMICUM*..... 69

3.1 Introduction 69

3.1.1 Significance of Mycothiol (MSH) 69

3.1.2 Biosynthetic Pathway of MSH.....70

3.1.3 Crystal Structure of *Cg*MshA 71

3.1.4 Conformational Changes of *Cg*MshA on Nucleotide Binding 72

3.1.5 Conclusion 75

3.2 Materials and Methods 76

3.2.1 Materials 76

3.2.2 Overexpression of *Cg*MshA 76

3.2.3 Purification of *Cg*MshA.....77

3.2.4 Site-directed Mutagenesis of *Cg*MshA 77

3.2.5 Production of I1P 78

3.2.6 Steady State Assays and Determination of Kinetic Parameters.....79

3.2.7 Sequencing of *Cg*MshA by Tandem MS/MS 79

3.2.8 Continuous HDX-MS Experiments of *Cg*MshA 95

3.3 Results..... 82

3.3.1 Steady State Kinetics of Wild Type *Cg*MshA and Variants.....82

3.3.2 Peptide Map of *Cg*MshA 84

3.3.3 Solvent Accessibility Changes on Substrate Binding.....85

3.3.4 Dynamic Changes of *Cg*MshA on Substrate Binding 90

3.4 Discussion..... 98

3.4.1 Substrate Binding Mechanism of *Cg*MshA 99

3.4.2 Insights into Dynamic Changes of *Cg*MshA 99

3.4.3 Roles of Hydrogen Bonds in Substrate Binding.....100

3.4.4 Interdomain Flexibility in GT-B Gycosyltransferases.....101

3.5 References..... 103

CHAPTER 4 CONFORMATIONAL DYNAMICS OF *ESCHERICHIA COLI* SufBC$_2$D COMPLEX ON ADP/Mg^{2+} BINDING 106

4.1 Introduction 106

4.1.1 Machinery of Suf Pathway.....106

4.1.2 Crystal Structure of SufBC$_2$D Complex 108

4.1.3 SufC Dimer Model.....110

4.1.4 Conclusion 111

4.2 Materials and Methods..... 112

4.2.1 Materials and Instruments.....112

4.2.2 Sequencing of $SufBC_2D$ Complex with Tandem MS/MS 112

4.2.3 Continuous HDX-MS Experiments of $SufBC_2D$ Complex 113

4.3 Result 115

4.3.1 Peptide Maps of $SufBC_2D$ Complex.....115

4.3.2 Solvent Accessibility Changes of SufC on ADP/Mg^{2+} Binding.....118

4.3.3 Dynamic Changes of SufC on ADP/Mg^{2+} Binding 119

4.3.4 Dynamic Changes of SufB/SufD Protomer on ADP/Mg^{2+} Binding.....123

4.4 Discussion 126

4.4.1 HDX-MS Data cannot Support the Proposed SufC Dimer.....126

4.4.2 Insight into Dynamic Changes of SufB/SufD Protomer.....128

4.5 References 130

CHAPTER 5 CONCLUSIONS AND FUTURE WORK 133

5.1 Summary 133

5.2 Impact of HDX-MS Study 135

5.3 Future Work 136

5.4 References 138

LIST OF TABLES

2.1 Sequences of forward and reverse primers of each *Mt*GpgS Substitution..... 42

2.2 Elution method of MS/MS for *Mt*GpgS..... 45

2.3 Elution method of continuous HDX-MS experiments..... 45

2.4 Kinetic parameters of wild type *Mt*GpgS and variants..... 48

3.1 Sequences of forward and reverse primers of each *Cg*MshA Substitution 78

3.2 Elution method of MS/MS sequencing..... 81

3.3 Elution method of continuous HDX-MS experiments..... 82

3.4 Kinetic parameters of wild type *Cg*MshA and variants..... 83

4.1 Elution method of tandem MS/MS for *E. coli* $SufBC_2D$ complex 114

4.2 Elution method of continuous HDX-MS experiments for *E. coli* $SufBC_2D$ complex.... 115

LIST OF FIGURES

1.1 Structures indicating the conformational changes of the Met20 loop in DHFR 2

1.2 Different types of hydrogen in a protein's structure 4

1.3 Representation of a typical HDX reaction 6

1.4 Linderstrom-Lang model of hydrogen deuterium exchange 6

1.5 The effect of pH on backbone amide hydrogen exchange rate 8

1.6 The effect of temperature on backbone amide hydrogen exchange rate 8

1.7 Tandem MS/MS 9

1.8 General procedure of a HDX-MS experiment 10

1.9 HDX data shown in Bruker Compass DataAnalysis 4.1 13

1.10 Mass shift of isotope envelopes 14

1.11 Deuterium incorporation curves 14

1.12 Example of HDX data analyzed by HD Examiner 15

1.13 Overall fold of glycosyltransferases 21

2.1 Overall crystal structure of MtGpgS monomer in complex with UDP, 3PGA, and Mg^{2+} 38

2.2 Active site residues mapped in crystal structure of *Mt*GpgS 39

2.3 Michaelis-Menten plots of wild type *Mt*GpgS 47

2.4 Substituted residues of *Mt*GpgS 48

2.5 Peptide map of *Mt*GpgS 49

2.6 Solvent accessibility changes of *Mt*GpgS on UDPG binding 51

2.7 Solvent accessibility changes of *Mt*GpgS on 3PGA binding 52

2.8 Deuterium incorporation curves of peptides from *Mt*GpgS on UDPG binding 53

2.9 Dynamic Changes of peptides on UDPG binding 53

2.10 Deuterium incorporation curves of peptides from *Mt*GpgS on 3PGA binding 55

2.11 Dynamic Changes of peptides on 3PGA binding 57

2.12 Residues involved in substrates binding 58

2.13 Substrate inhibition of *Mt*GpgS by 3PGA versus UDPG 63

3.1 Crystal structure of *Cg*MshA dimer in the absence of substrates 72

3.2 Alignment of the N- and C-terminal domain of *Cg*MshA before and after nucleotide binding 73

3.3 Crystal structure of *Cg*MshA dimer in closed conformation 74

3.4 Active site residues 74

3.5 Substituted residues of *Cg*MshA 84

3.6 Peptide map of *Cg*MshA 85

3.7 Solvent accessibility changes of *Cg*MshA on UDP-GlcNAc binding 86

3.8 Solvent accessibility changes of *Cg*MshA on I1P binding 87

3.9 Deuterium incorporation curves of peptides from *Cg*MshA 88

3.10 Dynamic changes of *Cg*MshA on UDP-GlcNAc binding 92

3.11 Deuterium incorporation curves of peptides from *Cg*MshA on I1P binding 93

3.12 Dynamic changes of *Cg*MshA on I1P binding 94

3.13 Comparison of deuterium incorporation curves between *Cg*MshA-UDP-GlcNAc and *Cg*MshA-UDP complexes 95

3.14 Differences in dynamic changes upon UDP and UDP-GlcNAc binding 97

4.1 Overall crystal structure of *E. coli* SufBC_2D complex 109

4.2 Overall crystal structure of *E. coli* SufC 110

4.3 Generated dimer model of SufC 111

4.4 Peptide maps of *E. coli* $SufBC_2D$ complex 116

4.5 Solvent accessibility changes of SufC on ADP/Mg^{2+} binding 117

4.6 Dynamic changes of catalytic α/β domain of SufC on ADP/Mg^{2+} binding..... 121

4.7 Deuterium incorporation curves of peptides located at catalytic α/β domain of SufC ... 122

4.8 Dynamic changes of α-helical domain of SufC on ADP/Mg^{2+} binding 122

4.9 Deuterium incorporation curves of peptides located at α-helical domain of SufC..... 123

4.10 Dynamic changes of SufB/SufD protomer 124

4.11 Deuterium incorporation curves of SufB peptides..... 124

4.12 Deuterium incorporation curves of SufD peptides 125

4.13 SufC dimer model labeled with peptides showing dynamic changes..... 127

4.14 Deuterium incorporation curves of peptides from SufC dimer interface 128

LIST OF ILLUSTRATIONS

1.1 Reaction schematic of retaining and inverting glycosyltransferase 17

1.2 A direct displacement S_N2-like reaction for inverting glycosyltransferase 19

1.3 Double displacement mechanism for retaining glycosyltransferase 19

1.4 Proposed S_Ni mechanism for retaining glycosyltransferase 21

1.5 The most common forms of Fe-S clusters 23

1.6 Comparison of Nif, Isc, and Suf systems 24

2.1 Structure of MGLPs 35

2.2 Formation of MGLPs 36

2.3 Schematic of PK/LDH coupled assay measuring enzymatic activity of *Mt*GpgS 43

3.1 Structure of MSH 69

3.2 Biosynthesis pathway of MSH 71

3.3 Schematic of PK/LDH coupled assay measuring enzymatic activity of CgMshA 80

4.1 Schematic of Suf pathway 107

CHAPTER 1

INTRODUCTION

1.1 Significance of Protein Dynamics

Proteins are known to be inherently flexible, and this flexibility generally plays an important role in protein function. That is, proteins do not simply occupy one structure, they are usually described by equilibrating sets of time-scale structures. [1-2] The transitions between different structures require dynamic fluctuations also called dynamics. So protein dynamics are crucial to conformational and structural changes of proteins. [3] Protein dynamics involves a great variety of internal motions having different amplitudes and different time scales, which affect a wide range of functions, such as catalytic process of enzymes, signaling/regulation, and thermostability. [4] Thus the investigation of protein dynamics is particularly important for understanding protein function, since they have an intimate relationship.

Dihydrofolate reductase (DHFR) provides a good example of insights into the link between protein dynamics and enzymatic function. [5] It catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF). By applying numerous experimental and theoretical approaches, substantial backbone and side chain motions were revealed to play essential roles in cofactor and substrate binding and in the catalytic cycle. [5] Moreover, analysis of available crystal structures illustrates that the large-scale conformational change of DHFR is concentrated in its Met20 active site loop (residues 9-24). The Met20 loop undergoes fluctuations over a wide range of timescales, and different DHFR states along the catalytic pathway are defined as open, closed, and occluded based on Met20 loop conformations (Figure 1.1). [6] Transitions between

these three states of DHFR are required by its catalytic process. Another example showing the relationship between protein dynamics and enzymatic activity is cyclic nucleotide phosphodiesterase (PDE). PDE plays a major role in cell signaling by hydrolyzing cAMP and cGMP, both of which are intracellular messengers. [7] PDE5 is one of five numbers of PDE superfamily and contains tandem GAF domain that is a ubiquitous motif present in PDE, adenylyl cyclases, bacterial transcription factor FhlA, and hundreds of other signaling and sensory proteins from all three kingdoms of life. [8] The sedimentation velocity studies revealed that PDE5 undergoes a large scale conformational change of the entire catalytic dimer on cGMP binding to GAF domain, which may represent the structural basis for PDE5 allosteric activation. [7] Both examples illustrate that there is an intimate relationship between protein dynamics and function.

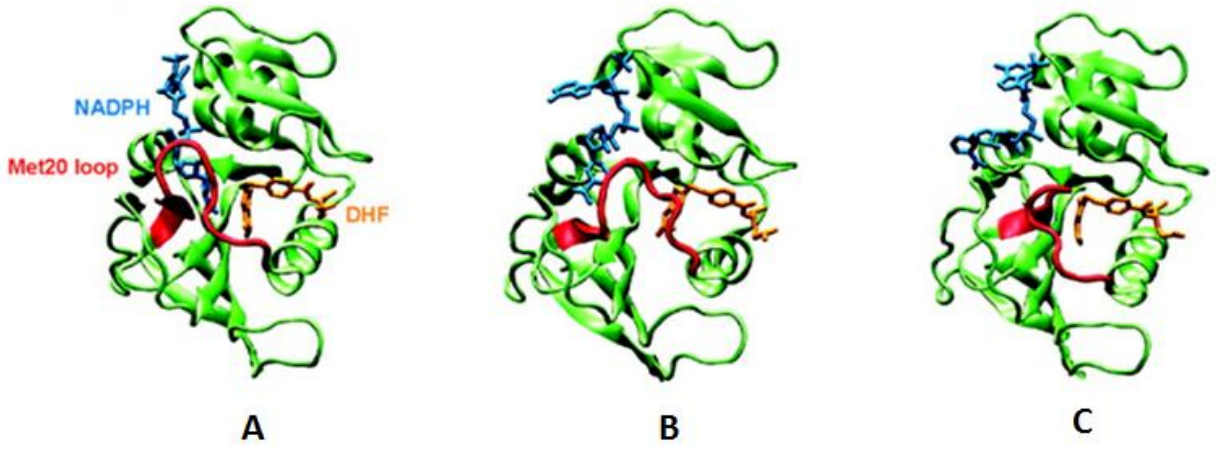


Figure 1.1. Structures displaying the conformational changes of the Met20 loop in DHFR. A, B, and C represent DHFR structures in closed, open, and occluded conformations, respectively. [7]

1.2 Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS)

Due to the significance of protein structure and dynamics contributing to protein function, techniques have been developed in order to study structure and dynamics of proteins.

Traditionally, X-ray crystallography provides a high-resolution static snapshot of a protein, but it is usually unable to reveal structural information in highly dynamic regions. [9] Techniques such as nuclear magnetic resonance (NMR) spectroscopy, fluorescence spectroscopy, and electron paramagnetic resonance (EPR) spectroscopy have been previously utilized in studying conformational changes of proteins. [3, 10-16] Even though these techniques provide insight into dynamics of proteins, they may require proteins to be modified or engineered. [9] For example, a series of single residue mutants on proteins is engineered to incorporate active reporter probes when applying NMR and EPR based approaches. [9] However, the modification may perturb a protein's structure and function. Furthermore, the NMR based approaches require high concentrations of protein samples, along with size limitation that the molecular weight of the protein is less than 30 kDa. So, NMR spectroscopy is not used for larger proteins and those difficult to over-express.

Compared with NMR spectroscopy, MS has less stringent sample requirements. The proteins at low concentrations or with high molecular weights can be characterized with MS. MS has been applied in various fields such as proteomics, drug discovery, and clinical testing. LC-MS, a combined system of HPLC and MS, becomes an attractive experimental system to characterize protein dynamics and conformational changes since the combination of HPLC technology and MS enhances the spatial resolution required for the study of large proteins. [17-18] Some MS based techniques to characterize protein dynamics have been described, including hydrogen deuterium exchange mass spectrometry (HDX-MS). [19-20] It is widely used in probing soluble protein conformational dynamics, as well as dynamic changes of proteins due to modification, ligand binding, and protein-protein interaction. [21]

Linderstrøm-Lang was the first to introduce hydrogen deuterium exchange and to

describe the relationship between the H/D exchange rate and protein dynamics by proposing the Linderstrøm-Lang model, which will be explained in 1.4.2. [22-23] The discovery of HPLC separation techniques combined with proteolysis by Rosa and Richards improved the spatial resolution. [24] Then, the separation techniques were further enhanced by Englander et al in the early 1980s. [25] Zhang and Smith found a new tool for protein elucidation, combining H/D exchange with mass spectrometry. [26] The HDX-MS technique was then improved on different aspects such as digestion condition. [27-29] Currently the HDX-MS technique has become a commonly used tool to study structure and dynamics of proteins in solution.

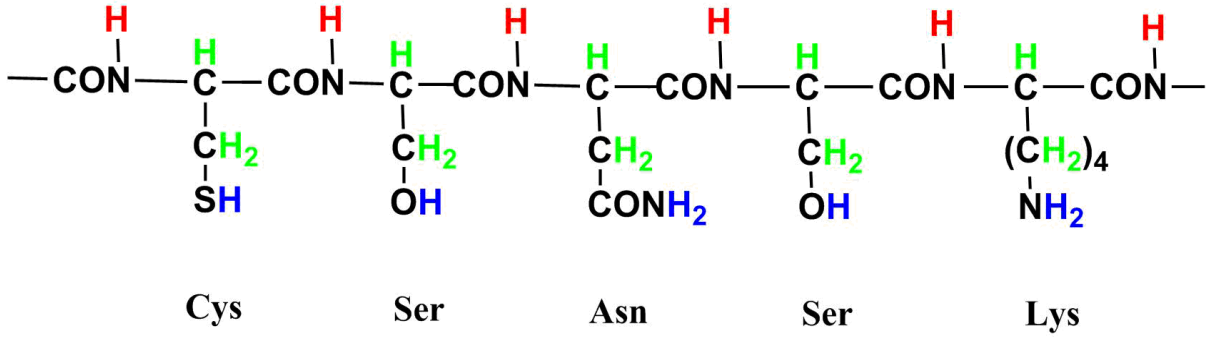


Figure 1.2. Different types of hydrogen in a protein's structure. Red color represents backbone amide hydrogen. Green color represents non-exchangeable hydrogen. Blue color represents side chain hydrogen.

1.2.1 Backbone Amide Hydrogens

Proteins contain a large number of hydrogens that can exchange with deuterium, but only hydrogens located on peptide amide linkages are used for H/D exchange studies (Figure 1.2). They can exchange with protons from surrounding water and can be observed in HDX-MS experiment. The backbone amide hydrogens are excellent indicators of protein structure and dynamics since their exchange rates are related to backbone flexibility, protein secondary structure, and solvent accessibility. [26, 30] Some amide hydrogens can exchange with

deuterium relatively rapidly if they are in unfolded states or exposed to solvent (pH 7, 25 ºC). [31] However, some amide hydrogens have significantly slower exchange rate if they are inaccessible to solvent or participating in stable hydrogen bonds (Figure 1.3). [26, 32-33] They can undergo structural fluctuations and conformational flexibility, which can break and reform hydrogen bonds in the native state of protein. [34]

1.2.2 Linderstrøm-Lang Model

The structural fluctuations resulting in open and closed conformations determine the rate of H/D exchange in a folded protein (Figure 1.4). [22, 35] The open and closed conformations are described by the kinetic rates of opening and closing, k_o and k_c, respectively. The exchange rate constant in the fully unfolded form is described as the intrinsic rates of exchange k_i. The exchange rates of hydrogens involved in hydrogen bonding or solvent inaccessible are determined by k_o, k_c, as well as k_i.

$$k_{ex} = k_o k_i / (k_o + k_c + k_i) \qquad \text{Equation 1.1}$$

The two extremes of hydrogen exchange mechanism exist in some proteins, named as EX1 and EX2. [36-37] In EX1, the intrinsic rate of amide hydrogen exchange is much faster than the rate of closing ($k_i >> k_c$). So the rate expression (Equation 1.1) can be simplified to equation 1.2.

$$k_{ex} = k_o k_i / (k_c + k_i) = k_o k_i / k_i = k_o \qquad \text{Equation 1.2}$$

The rate of amide hydrogen exchange is determined by the rate of opening (k_o). In EX2, the rate of closing is much faster than the intrinsic rate of amide hydrogen exchange ($k_i << k_c$). The rate expression is simplified to equation 1.3,

$$k_{ex} = k_o k_i / (k_c + k_i) = k_o k_i / k_c = (k_o / k_c) k_i \qquad \text{Equation 1.3}$$

describing that the rate of amide hydrogen exchange is dependent on the intrinsic rate of exchange and the equilibrium constant between the open and closed state (k_o / k_c).

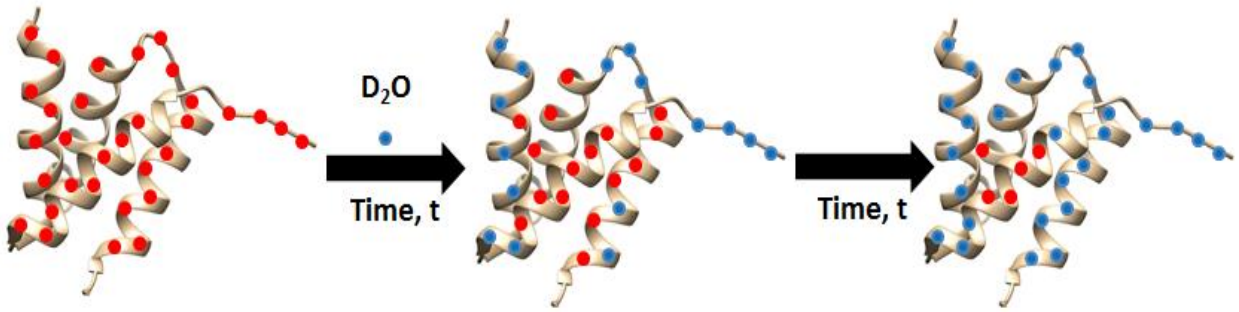


Figure 1.3. Representation of a typical HDX reaction. Backbone amide hydrogens exchange with deuterium rapidly if they are exposed to D_2O. Backbone amide hydrogens inaccessible to solvent or participating in stable hydrogen bonds exchange with deuterium slowly.

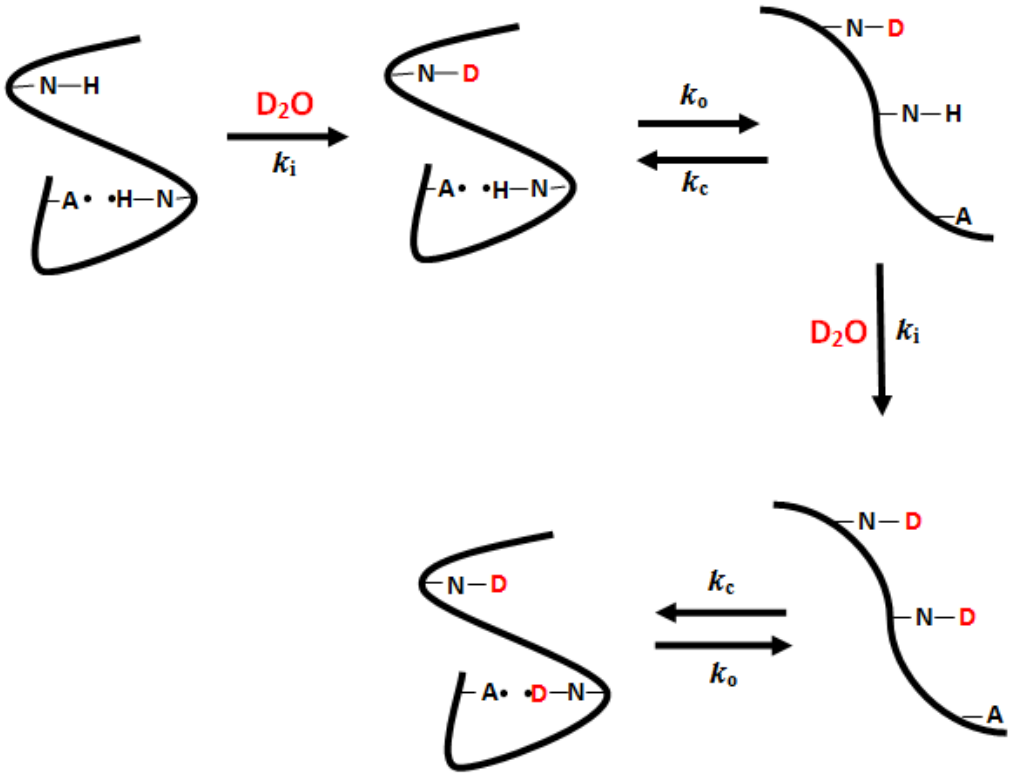


Figure 1.4. Linderstrøm-Lang model of hydrogen deuterium exchange. The structural fluctuations result in open and closed conformations of protein, which are described by the kinetic rates of opening and closing, k_o and k_c, respectively. The exchange rate of amide hydrogen in the fully unfolded form is described by k_i. Backbone amide hydrogens inaccessible to D_2O have slow exchange rate since they undergo from closed to open conformation.

1.2.3 Factors Affecting H/D Exchange Rate

Amide hydrogen exchange kinetics is affected by a variety of factors including temperature and pH. [31, 35] The pH is significantly important for HDX- MS experiments since the rate of hydrogen exchange is very sensitive to pH (Figure 1.5). [32, 36, 38] At neutral pH, the exchange is catalyzed by base abstraction (OD^-). [33] Most HDX-MS assays are performed at pH 7 to label accessible amide hydrogens with deuterium. The hydrogen exchange with solvent can also be catalyzed by acid (D_3O^+), water (D_2O), or buffer solutes. [21] As shown in Figure 1.5, the exchange rate is minimal at pH 2.4, which is commonly used in HDX-MS experiments in order to stop the H/D exchange reaction since it is necessary to quench exchange after the appropriate incubation time. Temperature is another controlling variable in HDX assays. The exchange rate decreases by 10-fold when the temperature drops from 25 ºC to 0 ºC (Figure 1.6). [32, 36] So, the ability to control the exchange rate by changing the temperature and pH is crucial for the HDX-MS experiments, as will be discussed in 1.2.4.

1.2.4 Procedures for HDX-MS Experiments

The HDX-MS experiments can measure deuterium incorporation uptake of an intact protein or peptides from proteolytic digestion. The HDX study of the intact protein reveals global protein structure and stability but cannot provide information for local structures and protein dynamics. The proteolytic digestion of protein allows one to monitor deuterium incorporation uptake of each digested peptide and to provide dynamic information of the protein. Amino acid sequences of peptides are required before initiating any HDX-MS experiment. This is achieved by tandem MS/MS sequencing (Figure 1.7). [39] The protein is digested by an acid protease into small peptides, which are then separated by high performance liquid chromatography (HPLC). The separated peptides are nebulized and ionized by electrospray

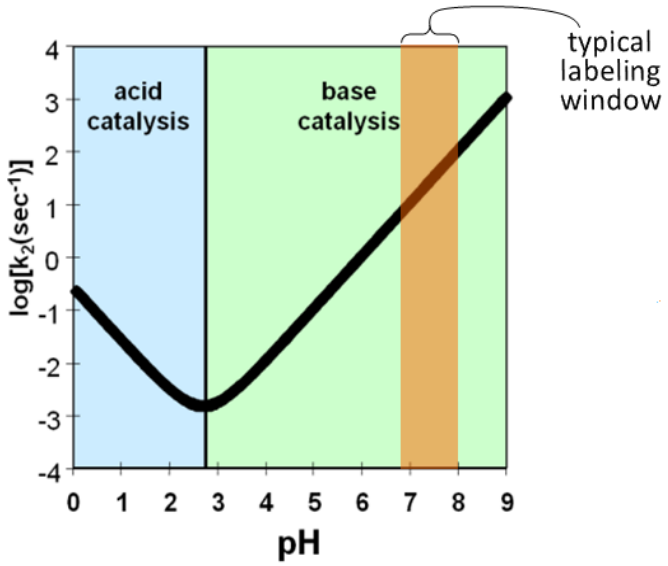


Figure 1.5. The effect of pH on backbone amide hydrogen exchange rate. The exchange rate is minimal at pH 2.3 and increases with increasing of pH. The HDX-MS experiment is usually performed at pH 7. [32, 36, 38]

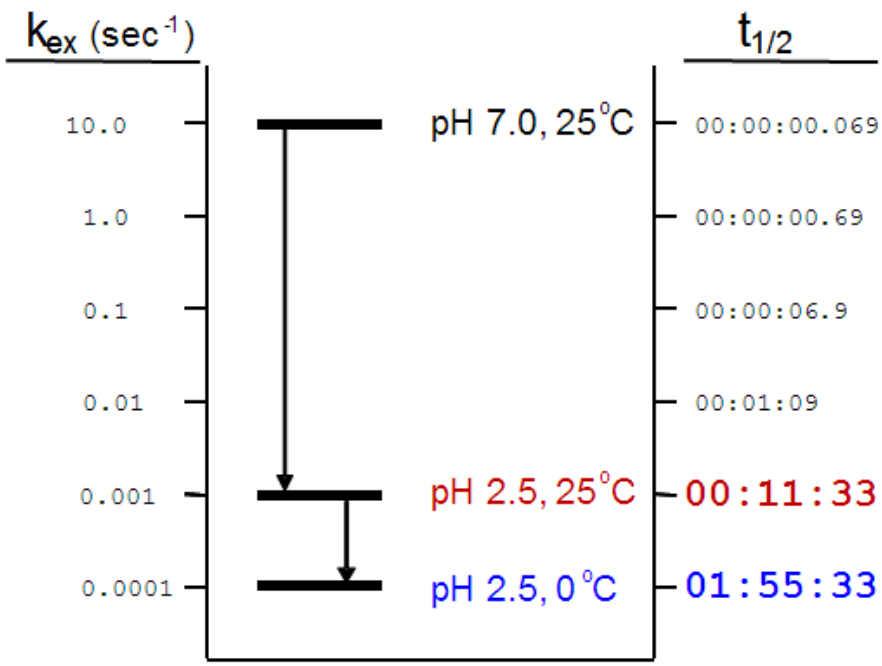


Figure 1.6. The effect of temperature on backbone amide hydrogen exchange rate. Reduction of pH from 7 to 2.5 leads to 1000 fold decrease in exchange rate. Decrease of temperature from 25 ºC to 0 ºC leads to further 10-fold decreasing in exchange rate. [32, 36]

ionization (ESI). The positively charged peptides are directed into a mass spectrometer and precursor ions are selected for collision induced ionization (CID). Finally, the fragmented ions are analyzed to determine the amino acid sequences of peptides by using a software Peaks Client (http://www.bioinfor.com/). Subsequently, the peptide map is built based on the determined sequences of peptides.

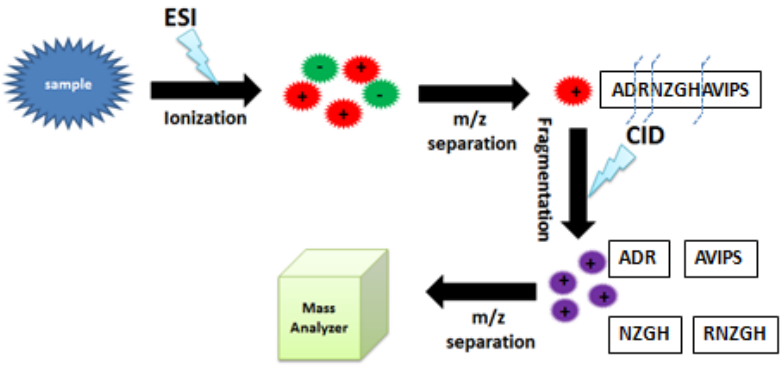


Figure 1.7. Tandem MS/MS. The sample is nebulized and ionized into charged droplets by electrospray ionization (ESI). The electrostatic potential helps to direct positively charged particles into the mass spectrometer. The positive particles with a selected mass, termed as precursor ions, pass through a region where they fall apart to produce fragment ions. It is achieved by collision induced dissociation (CID) by using helium. The product ions are separated according to m/z and analyzed again by the mass spectrometer.

The general procedure of backbone amide HDX-MS experiment is depicted in Figure 1.8 and includes hydrogen deuterium exchange for a specific incubation time, quenching the exchange by lowering pH and temperature, protein digestion by an acid protease such as pepsin, and mass to charge ratio (m/z) determination of peptides by LC-MS. [21] A native protein is incubated in an excess of D_2O (pH 7, 25 ºC) for a set of time points, starting the isotopic exchange. The incubation time usually varies from 15s to 6h, depending on protein structure. In order to trap the deuterium label, the temperature and pH are lowered to 0 ºC and 2.4, respectively, after the incubation time. Decreasing both temperature and pH results in rapid and

sharp reduction in amide hydrogen exchange rate. Moreover, the low pH moderately denatures the protein to help the subsequent digestion. [21] The quenched reaction is digested with pepsin at 0 ºC. Pepsin is an acid protease breaking down protein into smaller peptides. It has maximal activity at low pH. [40] Although pepsin is a non-specific protease, the digestion pattern is consistent at constant pH, temperature and protein to pepsin ratio. Then, the generated peptides are loaded onto a C18 reverse phase HPLC column, which is incubated in an ice bath. The peptides are separated by an acetonitrile gradient. After separation, the peptides are ionized by electrospray ionization (ESI) and introduced into a mass spectrometer. Finally, the mass to charge ratio (m/z) of each peptide at each incubation time point is recorded by a mass analyzer.

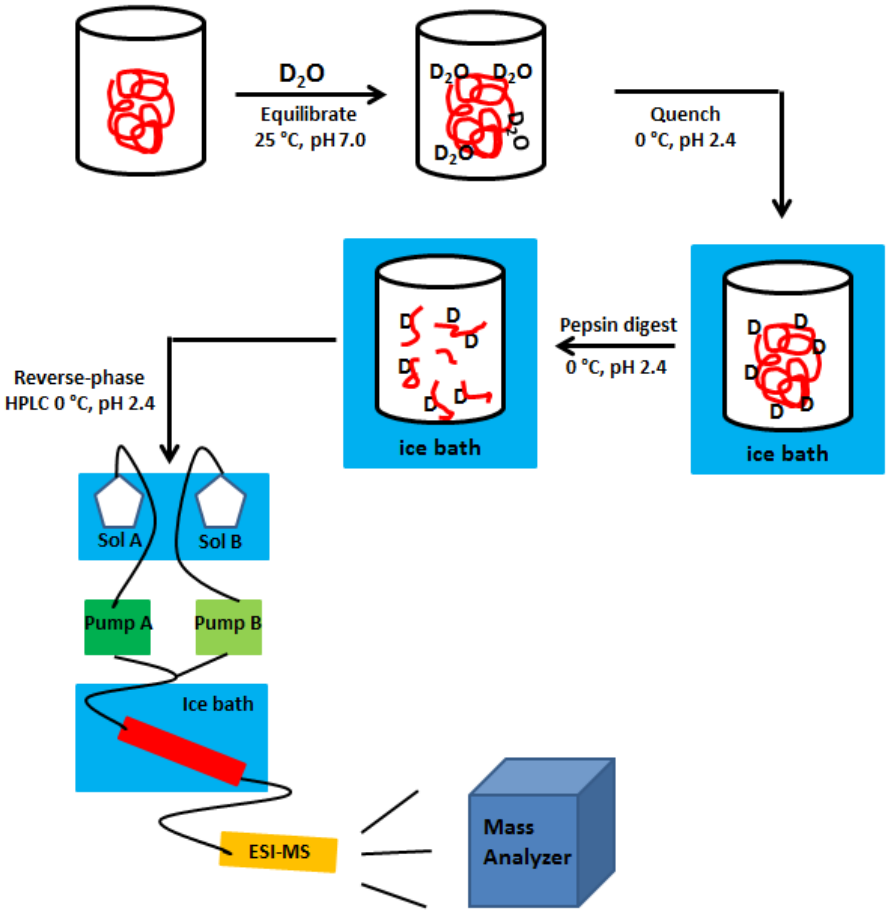


Figure 1.8. General procedure of a HDX-MS experiment.

The HDX-MS data must be corrected for natural isotopic distribution and loss of deuterium label due to back-exchange during the peptides separation in aqueous solutions. These two controls are termed as $m_{0\%}$ and $m_{100\%}$ controls. [21, 40] $m_{0\%}$ control is used to determine natural isotopic pattern for each peptide, by incubating protein in water instead of D_2O. The full deuteration control $m_{100\%}$ is to measure the maximum amount of deuterium incorporated by each peptide under the given experimental condition.

1.2.5 Analysis of HDX-MS data

The HDX-MS data output from the mass spectrometer is analyzed using Bruker Compass DataAnalysis 4.1 (Figure 1.9). A total ion chromatogram contains all generated peptides of protein with information of m/z, z, and retention time (RT) (Figure 1.9A). A particular peptide can be extracted according to its m/z and RT, which has been determined in tandem MS/MS. Figure 1.9B shows the extracted ion chromatogram of the particular peptide. The mass spectra under the extracted ion chromatogram are averaged to generate the isotope envelope of the peptide, which is depicted in Figure 1.9C. With the increase of incubation time, the isotope envelope of the peptide will have mass shift (Figure 1.10). From $m_{0\%}$ control to $m_{100\%}$ control, the isotopic distribution shifts to higher mass, indicating that more amide hydrogens exchange with deuterium with increase of incubation time.

As shown in Figure 1.9C, the area under the peak can be used to calculate centroid mass, which is employed to calculate the amount of deuterium incorporated in the particular peptide.

$$\boldsymbol{D = N\ \frac{(mt - m0\%)}{(m100\% - m0\%)}} \qquad \text{Equation 1.4}$$

where N is the number of exchangeable backbone amide hydrogens except for the N-terminal amide and any proline residues. m_t, $m_{0\%}$, and $m_{100\%}$ are centroid masses of the same peptide in the partially deuterated at time point t, non-deuterated, and the fully deuterated samples. [24, 41-42]

The numbers of deuterium incorporated in all generated peptides at different time points are calculated with this equation. The amount of deuterium of each peptide is averaged from three independent HDX-MS experiments. The amount of deuterium incorporated is plotted versus incubation time by fitting to a sum of order rate term described in Equation 1.5,

$$\boldsymbol{D = N - \sum_{i=1}^{N} exp(-ki * t)} \qquad \text{Equation 1.5}$$

where N is the number of amides that exchange at a given rate k_i for the isotopic exchange time t. [26, 43] The kinetic curves are plotted by using KaleidaGraph (Synergy Software) , fitting to one or two exponentials depending on the exchange process and the number of data points collected. [21] Figure 1.11 gives an example of kinetic plots of peptides fitting to single and double exponential equations. Figure 1.11A shows 34.64% of amide hydrogens of the peptide exchange with deuterium at the rate of 0.14 min^{-1}. Figure 1.11B describes that 35.79% of amide hydrogens exchange at the rate of 1.02 min^{-1} and 38.79% amide hydrogens have exchange rate of 0.016 min^{-1}.

HDX-MS data can be analyzed manually by employing software such as MagTran (Amgen Inc.) to calculate the centroid masses of each peptide at each time point. However, the vast amount of data generated from a HDX-MS experiment requires a huge amount of time. In order to perform data analysis efficiently, some software platforms have been developed to aid in HDX-MS data analysis. They include Deuterator, HX-Express, HD Desktop, DEX, and HD Examiner. [44-47] HD Examiner (Sierra Analytics) is a complete software solution for bottom up H/D exchange MS data analysis. It provides powerful functionality to show the deuterium incorporation level of a protein, by inputting peptide sequences and peptide list with charge and retention time. The experimental HDX-MS data is analyzed by HD Examiner and fit to theoretical isotope envelopes of peptides (Figure 1.12). Comparison of isotopic patterns between

theoretical and experimental envelopes gives confidence scores for each fit. As shown in Figure 1.12, the two isotope envelopes fit very well, suggesting a high confidence in the amount of deuterium incorporated calculated by HD Examiner.

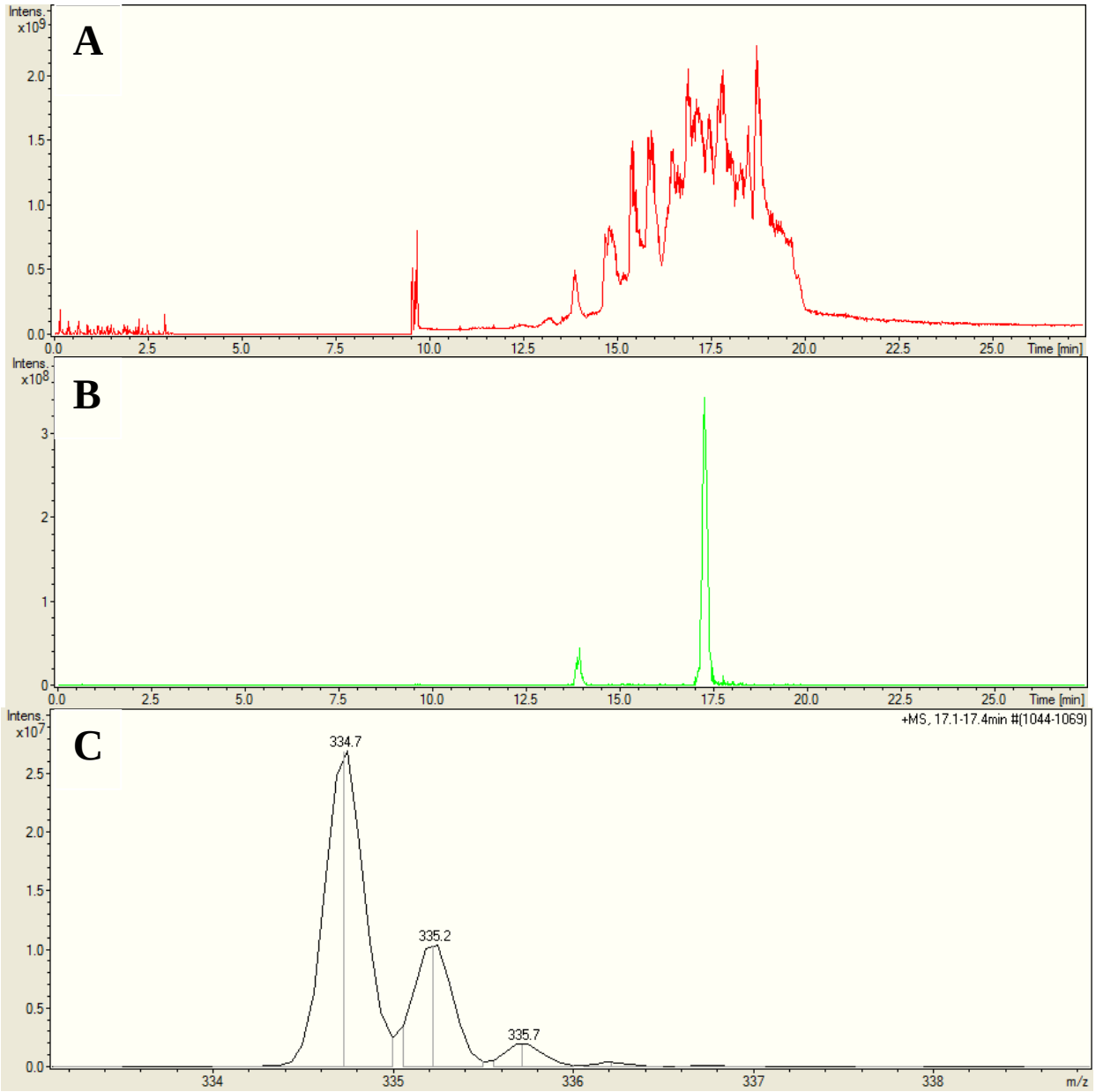


Figure 1.9. HDX-MS data shown in Bruker Compass DataAnalysis 4.1. (A) Total ion chromatogram containing all peptides with m/z, z, and retention time (RT). (B) Extracted ion chromatogram with m/z 334.7 as an example. (C) The peaks under the extracted ion chromatogram are averaged to produce the isotope envelop of the peptide.

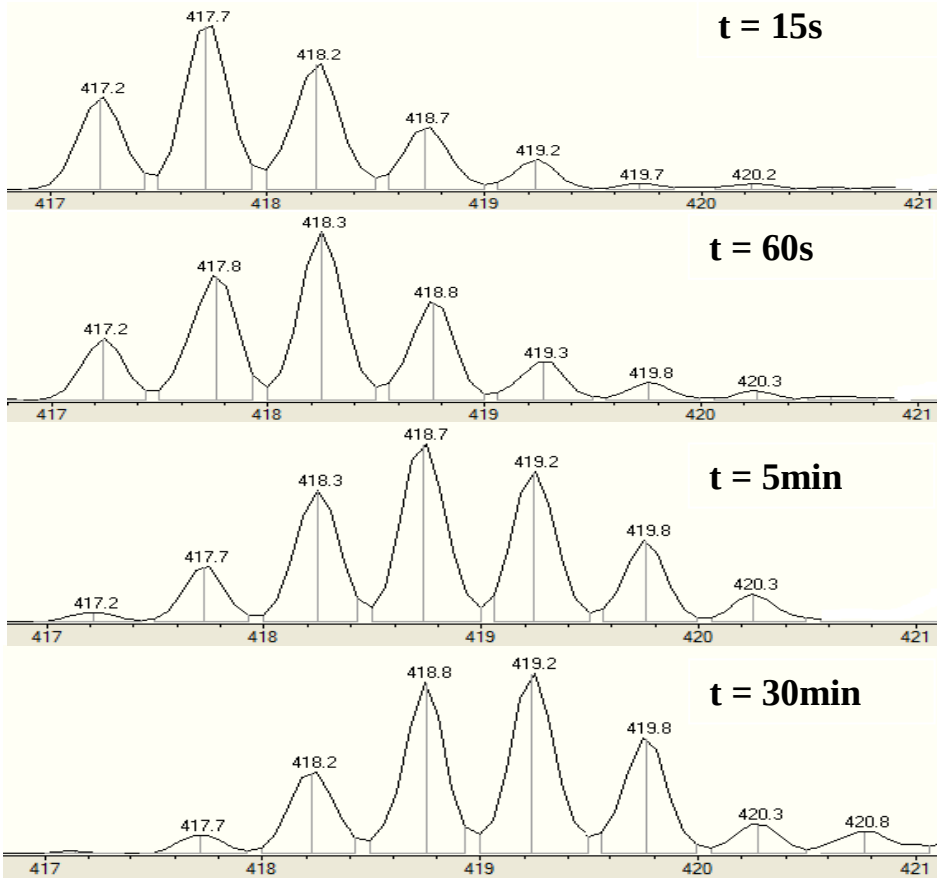


Figure 1.10. Mass shift of isotope envelopes of one searched peptide. Isotope envelopes have mass shift with increase of incubation time. From non-deuterated control to fully deuterated control, the isotopic pattern shifts to higher mass.

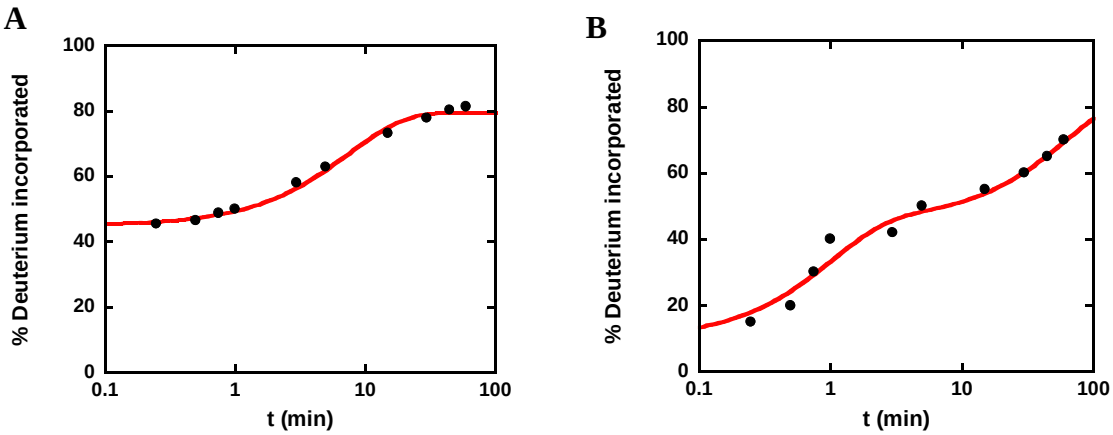


Figure 1.11. Examples of Deuterium incorporation Curves. The percentage of deuterium incorporated is plotted versus incubation time by fitting to (A) single double exponential equation and (B) double exponential equation.

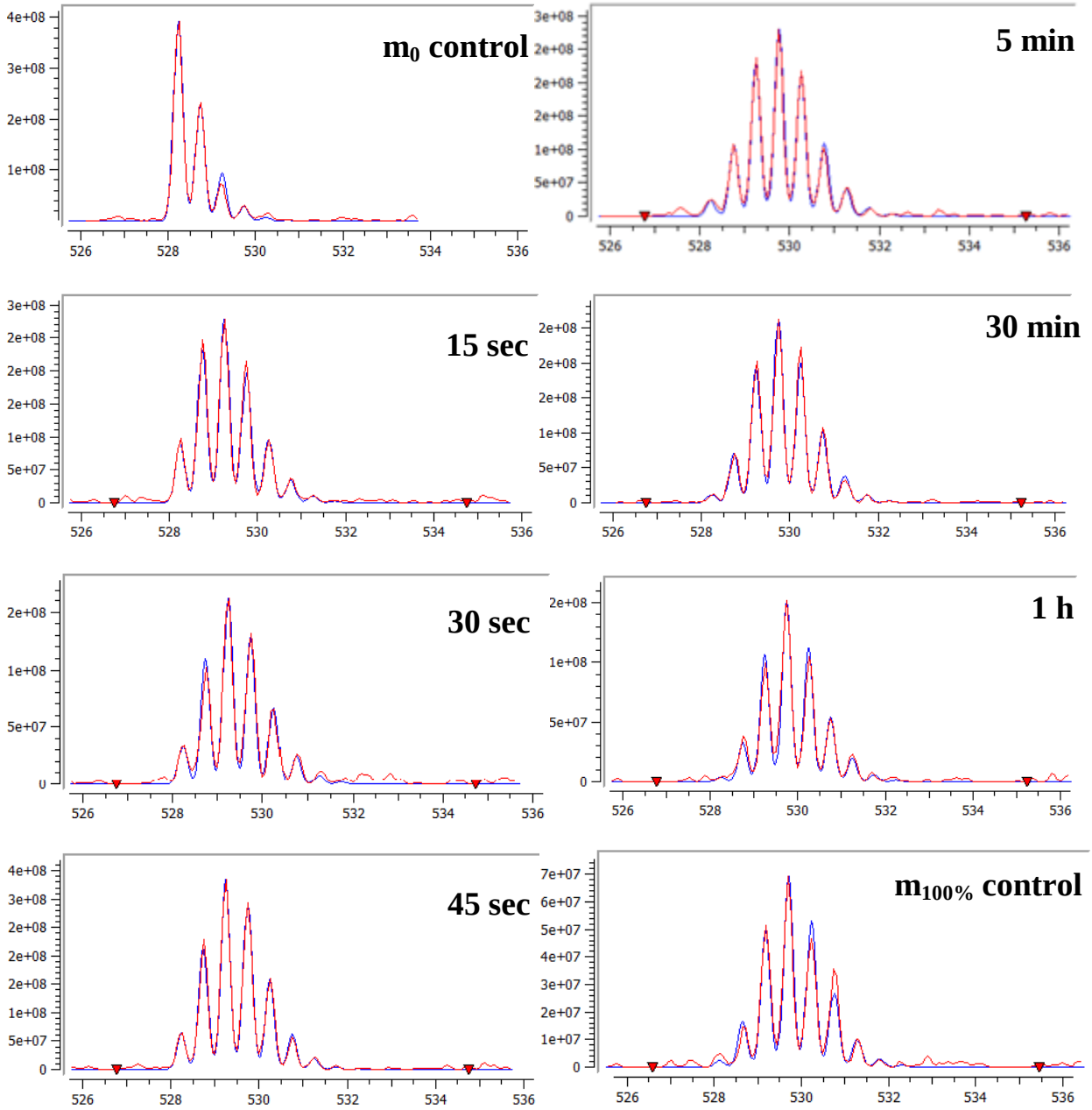


Figure 1.12. Example of HDX-MS data analyzed by HD Examiner. These isotopic distribution patterns are from the same peptide at different time points varying from 15s to 1h along with m_0 control and m_{100} control. x axis is m/z ratio and y axis is intensity of peak. The experimental isotope envelopes of peptide are shown in red color and the theoretical isotope envelopes calculated by HD Examiner are in blue.

1.3 HDX-MS Studies of Protein-Ligand and Protein-Protein Interactions

Proteins are involved in most biological processes of the livings. Some important aspects required to understand protein function include protein sequence and structure, kinetics, evolutional history of conserved sequences, the binding of ligands, and interactions with other proteins. [48] Many proteins have interactions with ligands in order to perform functions properly. These ligands have different types, including ions, substrates, and cofactors. [48]Ligand binding is crucial to some proteins since it can alter conformation of the protein by affecting the three-dimensional shape orientation. The new conformation induced by ligand binding is usually required by catalytic process and enzymatic activity. [49-50] Proteins can also have interactions with other proteins. Proteins can interact with each other through a combination of interactions including van der Waal forces, salt bridges, and hydrogen bonds. [51-52] These interactions can be stable or transient. Transient interactions are temporary in nature and typically require a set of conditions to promote the interaction, such as phosphorylation and conformational changes. [53] The transient interactions are significant in controlling and regulating many cellular processes by affecting both structure and function such as altering kinetic properties of enzymes, allowing ligand binding and active site, changing specificity, and inactivating a protein. [53]

Due to the importance of ligand binding and PPIs for protein function, technologies have been developed to study these interactions. The traditional technologies to study ligand binding include atomic force microscopy (AFM) and isothermal titration calorimetry (ITC). ITC is able to provide information on the thermodynamic contributions of enthalpy and entropy changes to free energies of ligand binding. [54] AFM quantifies directly the range and magnitude of the interaction forces between protein and ligands. [55] Moreover, various methods and approaches

have been used to study PPIs, such as co-immunoprecipitation, pull-down assay, cross-linking protein interaction analysis, and far-western blot analysis. [56] However, these techniques provide little conformational and mechanistic information. So, HDX-MS becomes a useful tool to study dynamic changes of protein on ligand binding and PPIs. [57]

This dissertation focuses on studying dynamic and conformational changes of proteins due to ligand binding and PPI using HDX-MS technique. Chapters 2 and 3 discuss the conformational changes of two glycosyltransferases from *Mycobacterium tuberculosis* and *Corynebacterium glutamicum* on substrate binding by applying HDX-MS. Chapter 4 focuses on developing HDX-MS to probe dynamic changes and protein-protein interactions within a protein complex from *Escherichia coli* involved in Fe-S cluster assembly pathway.

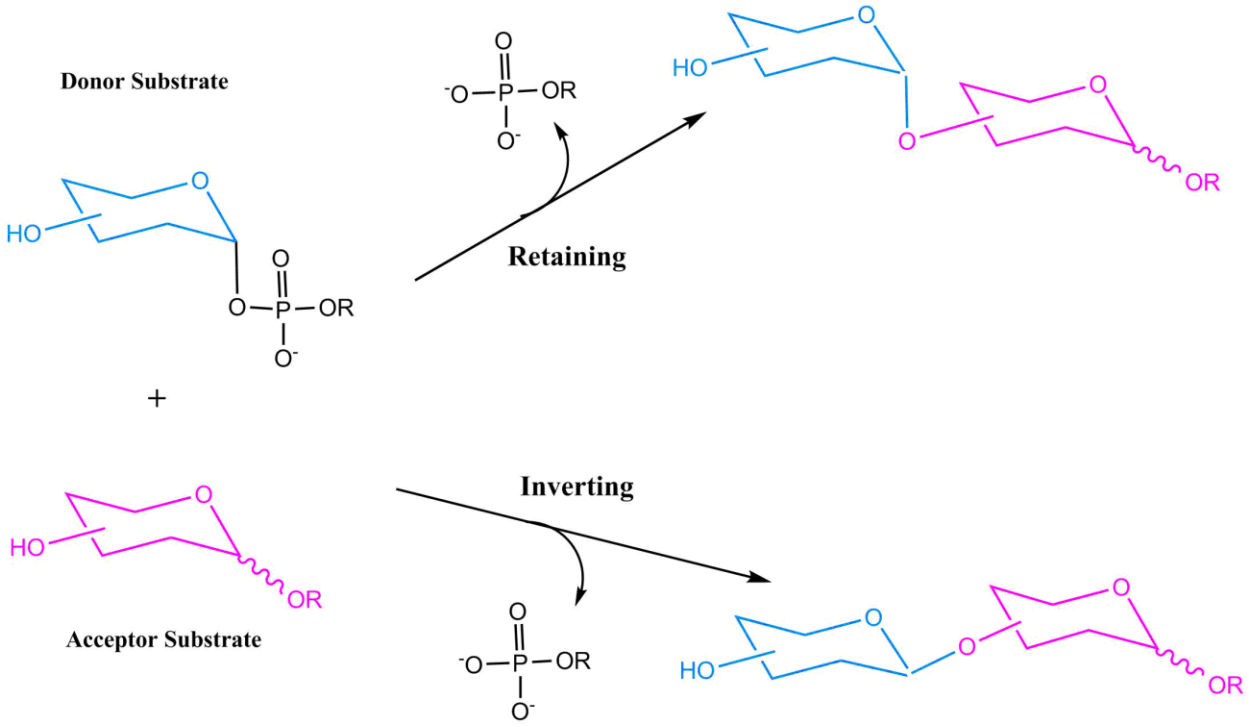


Illustration 1.1. Reaction schematic of retaining and inverting glycosyltransferase. Glycosyltransferase catalyzes transfer of sugar moiety from donor substrate to acceptor substrate with retention or inversion. R: a nucleoside, a nucleoside monophosphate, a lipid phosphate, or phosphate.

1.4 Glycosyltransferase Enzymes

Glycosyltransferases (GTs) are a group of enzymes that catalyze the transfer of a sugar moiety from a sugar donor to a sugar acceptor with formation of a glycosidic bond. [58] GTs can utilize a range of sugar donors including nucleotide sugar and non-nucleotide donors. The sugar acceptor substrates may be monosaccharide, oligosaccharide, polysaccharide, protein, lipid, small organic molecule, and deoxyribonucleic acid. [59] Even though the GTs have a variety of sugar donors and acceptors, they are highly specific for substrates. Each type of glycosidic linkage needs a unique glycosyltransferase. [60-61]

1.4.1 Retaining and Inverting Glycosyltransferase

GTs can be classified as retaining or inverting depending on the configuration at the anomeric centre of the sugar donor product. (Illustration 1.1) [60, 62] Mechanistic study of inverting glycosyltransferases shows that they use a direct displacement S_N2-like reaction mechanism leading to an inverted anomeric configuration of product. (Illustration 1.2) [63] An active site side chain acts as a base catalyst, deprotonating the incoming nucleophilic sugar acceptor. Deprotonation facilitates nucleophilic substitution reaction leading to new bond formation and nucleotide or phosphate group leaving. [64] In contrast, the mechanism of retaining glycosyltransferases is still not clear. [63] A possible mechanism is a double-displacement reaction involving a covalently bound glycosyl-enzyme intermediate. (Illustration 1.3) [63] The formation of this intermediate requires a stringently conserved active site moiety that might play a role in protecting one side of the substrate. Then, the leaving group might act as a base catalyst that activates sugar acceptor for nucleophilic attack. [65-67] This mechanism explains the retention of configuration in the product. However, the absence of the highly conserved active site suggested the presence of a different mechanism. [63] So an alternate

mechanism was proposed for the retaining glycosyltransferases, known as S_Ni reaction where "i" represents internal reaction. (Illustration 1.4) [67-69] It involves the formation of discrete and short-lived ion pair intermediates. [70-71] The formation of ion pair intermediate requires a back-side nucleophilic "push" from the enzyme. Then, the ion pair undergoes front-side attack by an incoming nucleophile. [63]

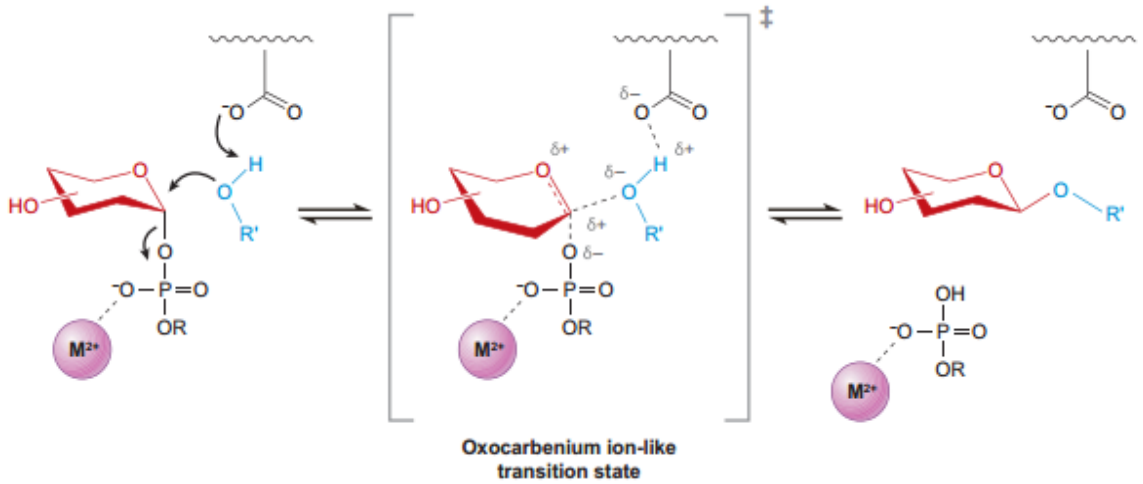


Illustration 1.2. Inverting glycosyltransferase uses a direct displacement S_N2-like reaction. [63]

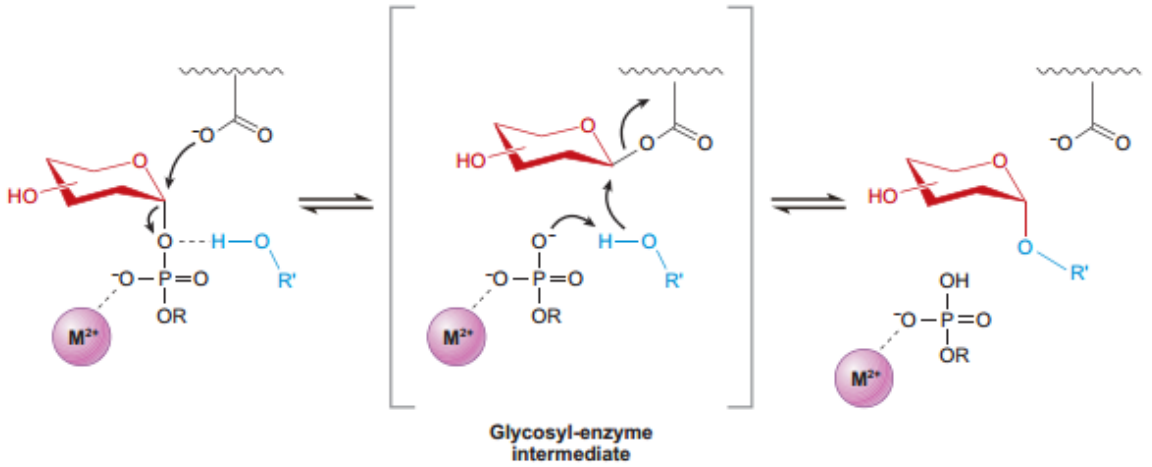


Illustration 1.3. Double-displacement mechanism is proposed for retaining glycosyltransferase. [63] R: a nucleoside, a nucleoside monophosphate, a lipid phosphate, or phosphate; R'OH: an acceptor group, such as another sugar, a protein, or an antibiotic.

1.4.2 GT-A and GT-B fold

As discussed above, GTs can be classified as retaining or inverting with respect to the anomeric configuration of the donor substrate. Moreover, GTs can also be classified based on sequence and structure. Based on amino acid sequence similarities, GTs are classified into 90 families as of 2014. [60, 72] However, structural characterization of representatives of these families has revealed that only two general folds, named GT-A and GT-B, are observed for structurally characterized glycosyltransferases. [60, 73-75] Furthermore, some uncharacterized families are predicted to possess one of these two folds. [63] GT-A fold consists of one Rossmann-like fold, which is a protein structural motif for nucleotide binding. (Figure 1.13A) The Rossmann-like fold is composed of parallel β-sheets linked with α-helices on both sides in the topological order β/α/β. [74, 76-77] Two tightly associated β/α/β domains lead to the formation of a continuous central β-sheet. In GT-A fold, the *N*-terminal domain contains the nucleotide binding site and the *C*-terminal domain recognizes the sugar acceptor substrate. Moreover, GT-A glycosyltransferases possess the DXD motif, which is a determining characteristic of GT-A enzymes. In this motif, the carboxylates usually coordinate a divalent cation and/or a ribose. [78-79] Thus GT-A glycosyltransferases are metal ion dependent.

GT-B folds consist of two Rossmann-like folds, which are less tightly associated. (Figure 1.13B) [80] The *C*-terminal domain of GT-B glycosyltransferases provides the binding site for the nucleotide sugar donor. Unlike the GT-A fold, GT-B glycosyltransferases are metal ion independent. [59] Moreover, GT-B glycosyltransferases have open and closed conformations, which is a characteristic of GT-B fold. The GT-B glycosyltransferase is in an open state in the absence of donor but it has another conformation upon sugar donor binding. The conformational transition of the two states can provide a binding pocket for substrates. [59]

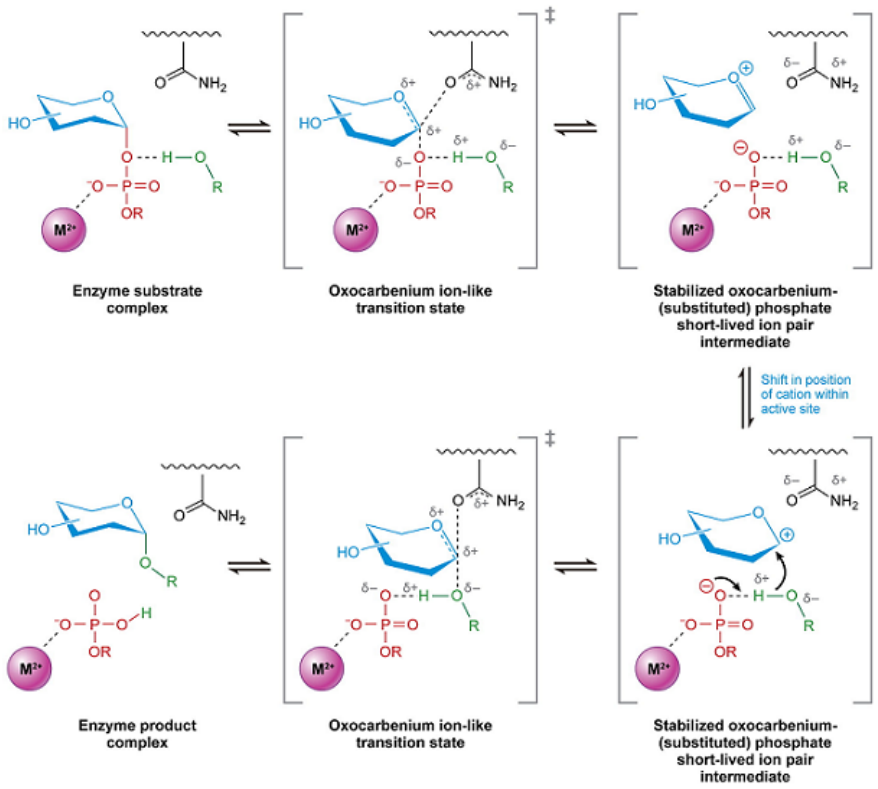


Illustration 1.4. Proposed S_Ni mechanism for retaining glycosyltransferase. [63] R: a nucleoside, a nucleoside monophosphate, a lipid phosphate, or phosphate; R'OH: an acceptor group, such as another sugar, a protein, or an antibiotic.

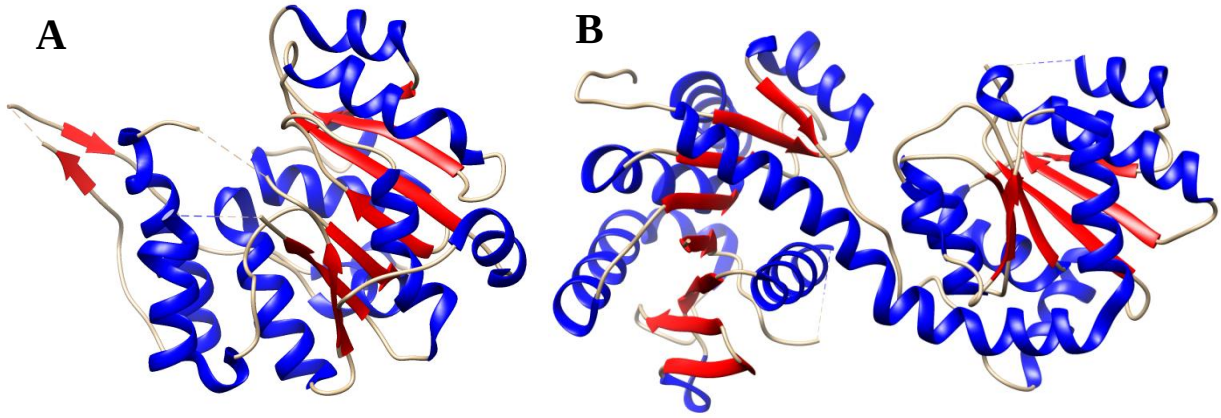


Figure 1.13. Overall fold of glycosyltransferases. Blue color represents β strands and red color represents α helices. (A) GT-A glycosyltransferase represented by retaining GpgS from *mycobacterium tuberculosis* (PDB: 3e26) (B) GT-B glycosyltransferase represented by retaining MshA from *Corynebacterium glutamicum* (PDB: 3c48)

1.5 Suf System Assembling Fe-S Clusters

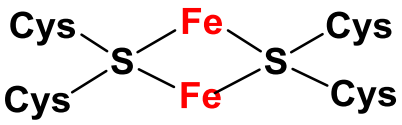
1.5.1 Iron-Sulfur Clusters

Iron plays an important role in biology since it is often at the active site of many redox enzymes dealing with cellular respiration and oxidation/reduction in plants and animals. [81] Another reason for significance of iron in biology is that the iron is able to combine with elemental sulfur to form iron-sulfur (Fe-S) clusters, which are thought to be one of the earliest iron cofactors used in biology. [82] Fe-S clusters play different roles in biology. First, Fe-S clusters are often stable at multiple oxidation states and have relevant redox potentials, so that they have an electron transfer function. [83] Second, some Fe-S clusters are involved in substrate binding and activation in some enzymes. [94] Third, some Fe-S clusters play important roles in gene regulation at both transcriptional and translational levels. [85] In addition to the variety of cluster functions, the Fe-S clusters also have various cluster forms. The most common forms are [2Fe-2S], [3Fe-4S], and [4Fe-4S]. (Illustration 1.5) The forms of Fe-S clusters depend on the number and arrangement of Cys residues provided by proteins. Even though the clusters are usually coordinated to proteins through Cys residues, some other residues can also coordinate clusters at single sites. [86-87]

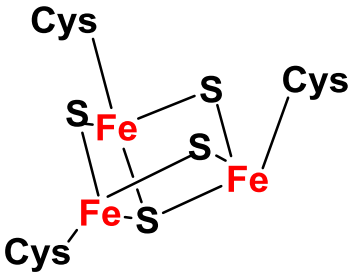
1.5.2 Fe-S Cluster Biogenesis Pathways

Carefully coordinated biogenesis pathways are necessary for Fe-S cluster assembly, since both iron and sulfur are highly reactive and toxic in living cells. [81] Multiple pathways for Fe-S cluster assembly exist in livings, while three distinct pathways have been identified. They are Isc (iron sulfur cluster) system, Nif (nitrogen fixation) system, and Suf (sulfur formation) system. [88] These three systems have similar procedures for Fe-S cluster assembly. (Illustration 1.6) [89] First, the proteins mobilize Fe and S atoms from their donors. Second, the related proteins

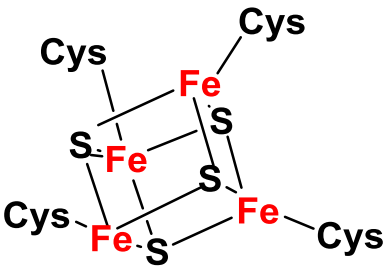
assemble Fe and S atoms in cluster form and transport them. Finally, the assembled Fe-S clusters are transferred to target proteins. [90]



[2Fe-2S] cluster



[3Fe-4S] cluster



[4Fe-4S] cluster

Illustration 1.5. The most common forms of Fe-S clusters.

Even though they have similar assembly processes, these three systems have different taxonomic distributions. The Nif system is primarily found in nitrogen fixing bacteria. The Suf and Isc systems are dominant in bacteria, eukaryotes, and archaea, while the distribution of Isc and Suf is complicated. For example, the Suf pathway seems to be more important in cyanobacteria than Isc pathway. The Isc pathway appears to be major system in *Escherichia coli* compared to the Suf pathway, while the Suf system synthesizes clusters in *E. coli* instead of the Isc system when iron or sulfur metabolism is disrupted by iron starvation or oxidative stress. [91-95] Furthermore, the Suf pathway is the only system assembling Fe-S clusters for some bacteria such as *Mycobacterium tuberculosis*. [81]

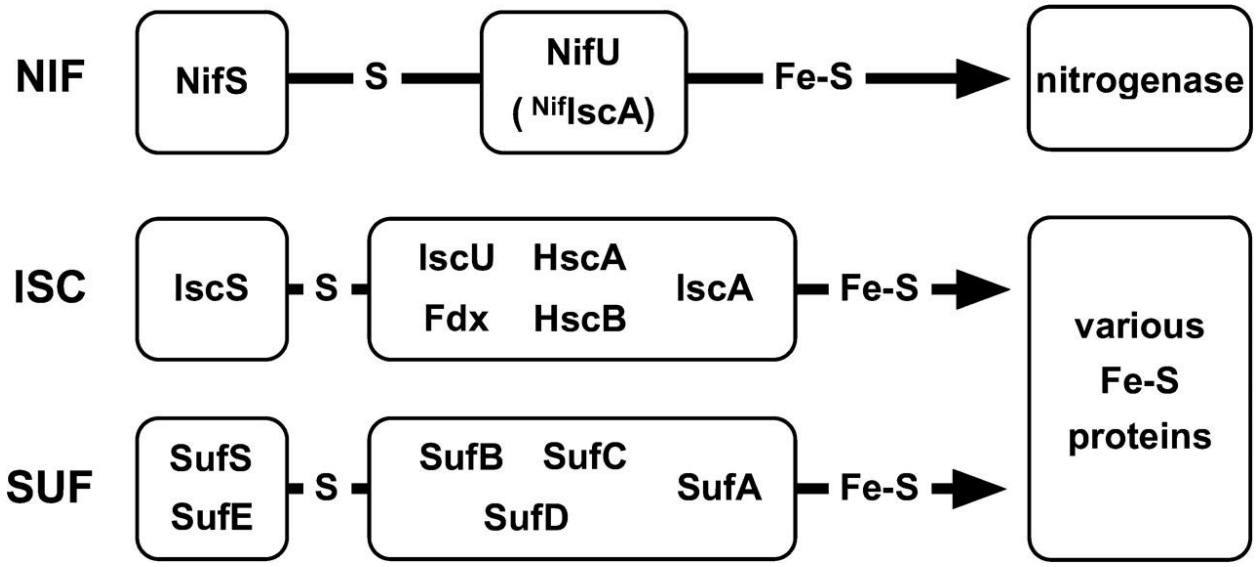


Illustration 1.6. Comparison of the Nif, Isc, and Suf systems. They share the same procedures for cluster assembly: mobilization of sulfur atoms, assembly of Fe-S clusters, and transfer of Fe-S clusters. The proteins involved in each procedure are different for each system. [95]

1.6 HDX-MS Applied in Dynamics of Protein and Protein Complex

Protein dynamics is an important aspect for understanding protein function. HDX-MS is an essential technique to study the protein's structure and dynamics, which can provide information for the enzymatic mechanism and protein functions. It is commonly applied in observing the conformational changes of protein due to ligand binding, PPI, and protein modifications. This dissertation focuses on dynamic changes of proteins and protein complex upon ligand binding and PPI. The GT-A glycosyltransferase glucosyl-3-phosphoglycerate synthase from *Mycobacterium tuberculosis* (*Mt*GpgS) possesses a flexible loop (residues 253-262). This loop was detected to have two conformations on substrate binding by observing crystal structures of free and substrate-binding *Mt*GpgS. [96] The GT-B glycosyltransferase MshA from *Corynebacterium glutamicum* (*Cg*MshA) has two conformations upon nucleotide binding and the closed conformation provides binding pocket for two substrates. [97] Even though both glycosyltransferases have been identified to have conformational changes in large or

small scale, their detailed dynamic information are still unknown. HDX-MS was applied to study dynamic changes of the GT-A glycosyltransferases *Mt*GpgS and GT-B glycosyltransferase *Cg*MshA on substrate binding, which will be discussed in chapters 2 and 3, respectively. The $SufBC_2D$ complex, the scaffold responsible for Fe-S clusters assembly in Suf pathway, is composed of proteins SufB, SufC, and SufD. The cross-linking experiments showed that SufC can form a dimer in the presence of ATP and Mg^{2+}. Moreover, SufB and SufD subunits were proposed to have conformational changes due to SufC dimerization. [98] However, more studies are required to fully characterize SufC function and mechanism of Fe-S cluster assembly. So HDX-MS experiments were applied to study dynamic changes of SufC on ADP/Mg^{2+} binding within $SufBC_2D$ complex and conformational changes of SufB and SufD due to their interactions with SufC.

1.7 References

[1] Henzler-Wildman, K., and Kern, D. Dynamic personalities of proteins. *Nature.* 2007;450:964-972.

[2] Boehr, D. D., Dyson, H. J., and Wright, P. E. An NMR perspective on enzyme dynamics. *Chem Rev*. 2006;106:3055-3079.

[3] Antoniou, D., Basner, J., Nunez, S., and Schwartz, S. D. Computational and theoretical methods to explore the relation between enzyme dynamics and catalysis. *Chem Rev.* 2006;106:3170-3187.

[4] Ma, B., and Nussinov, R. Enzyme dynamics point to stepwise conformation selection in catalysis. *Curr Opin Chem Biol.* 2010;14:652-659.

[5] Schnell, J. R., Dyson, H. J., and Wright, P. E. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biomol Struct.* 2004;33:119-140.

[6] Arora, K., and Brooks III, C. L. Functionally important conformations of the Met20 loop in dihydrofolate reductase are populated by rapid thermal fluctuations. *J. Am. Chem. Soc.* 2009;131:5642-5647.

[7] Matte, S. L., Laue, T. M., and Cote, R. H. Characterization of conformational changes and protein-protein interactions of rod photoreceptor phosphodiesterase (PDE6). *J. Biol. Chem.* 2012;287:20111-20121.

[8] Ho, Y. J., Burden, L. M., and Hurley, J. H. Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO J*. 2000;19:5288-5299.

[9] Kahsai, A. W., Rajagopal, S., Sun, J., and Xiao, K. Monitoring protein conformational changes and dynamics using stable-isotope labeling and mass spectrometry. *Nat. Protoc.* 2014;9:1301-1319.

[10] Volkman, B. F., Lipson, D., Wemmer, D. E., and Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science.* 2001;291:2419-2433.

[11] Tzeng, S. R., and Kalodimos, C. G. Protein dynamics and allostery: an NMR view. *Curr Opin Struct Biol.* 2011;21:62-67.

[12] Rozovsky, S., Jogl, G., Tong, L., and McDermott, A. E. Solution-state NMR investigations of triosephosphate isomerase active site loop motion: ligand release in relation to active site loop dynamics. *J Mol Biol*. 2001;310:271-280.

[13] Yang, H., Luo, G., Karnchanaphanurach, P., Louie, T. M., Rech, I., Cova, S., Xun, L., and Xie, X. S. Protein conformational dynamics probed by single-molecule electron transfer. *Science.* 2003;302:262-266.

[14] Weiss, S. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat Struct Biol*. 2000;7:724-729.

[15] Hubbell, W. L., Cafiso, D. S., and Altenbach, C. Identifying conformational changes with site-directed spin labeling. *Nat. Struct. Biol.* 2000;7:735-739.

[16] Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L., and Khorana, H. G. Requirement of rigid body motion of transmembrane helices for light activation of rhodopsin. *Science.* 1996:274:768-770.

[17] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 1999;17:994-999.

[18] Ong, S. E., and Mann, M. Mass spectrometry-based proteomics turns quantitative. Nat. Chem. Biol. 2005;1:252-262.

[19] Eyles, S. J., and Kaltashov, I. A. Methods to study protein dynamics and folding by mass spectrometry. *Methods*. 2004;34:88-99.

[20] Kaltashov, I. A., and Eyles, S. J. Studies of biomolecular conformations and conformational dynamics by mass spectrometry. *Mass Spectrom Rev*. 2002;21:37-71.

[21] Busenlehner, L. S., and Armstrong, R. N. Insights into enzyme structure and dynamics elucidated by amide H/D exchange mass spectrometry. *Arch Biochem Biophys.* 2004;433:34-46.

[22] Englander, S. W., Mayne, L., Bai, Y., and Sosnick, T. R. Hydrogen exchange: the modern legacy of Linderstrøm-Lang. *Protein Sci.* 1997;6:1101-1109.

[23] Hvidt, A., and Linderstrom-Lang, K. Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. *Biochim Biophys Acta.* 1954;14:574-575.

[24] Rosa, J. J., and Richards, F. M. An experimental procedure for increasing the structural resolution of chemical hydrogen-exchange measurements on proteins: Application to ribonuclease S peptide. *J Mol Biol.* 1979;133:399-416.

[25] Englander, J. J., Rogero, J. R., and Englander, S. W. Protein hydrogen exchange studied by the fragment separation method. *Anal Biochem.* 1985;147:234-244.

[26] Zhang, Z., and Smith, D. L. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* 1993;2:522-531.

[27] Deng, Y., Pan, H., and Smith, D. L. Selective isotope labeling demonstrates that hydrogen exchange at individual peptide amide linkages can be determined by collision-induced dissociation mass spectrometry. *J Am Chem Soc.* 1999;121:1966-1967.

[28] Wang, L., Pan, H., and Smith, D. L. Hydrogen-exchange mass spectrometry: optimization of digestion conditions. *Mol Cell Proteomics.* 2002;1:132-138.

[29] Cravello, L., Lascoux, D., and Forest, E. Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins. *Mass Spectrom*. 2003;21:2387-2393.

[30] Englander, S. W., Mayne, L., and Rumbley, J. N. Submolecular coorperativity produces multi-state protein unfolding and refolding. *Biophys Chem*. 2002;101-102:67-65.

[31] Bai, Y., Milne, J. S., Mayne, L., and Englander S. W. Primary structure effects on group hydrogen exchange. *Proteins*. 1993;17:75-86.

[32] Hvidt, A. Deuterium exchange between ribonuclease and water. *Biochim. Biophys. Acta.*

[33] Maity, H., Lim, W. K., Rumbley, J. N., and Englander S. W. Protein hydrogen exchange mechanism: local fluctuations. *Protein Sci*. 2003;12:153-160.

[34] Konermann, L., Pan, J., and Liu, Y. H. Hydrogen exchange mass spectrometry for studying protein structure and refolding. *Chem Soc Rev*. 2011;40:1224-1234.

[35] Molday, R. S., Englander, S. W., and Kallen, R. G. Primary structure effects on peptide group hydrogen exchange. *Biochemistry.* 1972;11:150-158.

[36] Englander, S. W., and Kallenbach, N. R. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys*. 1983;16:521-655.

[37] Roder, H., Wagner, G., and Wuethrich Kurt. Amide protein exchange in proteins by EX1 kinetics: studies of the basic pancreatic trypsin inhibitor at variable p2H and temperature. *Biochemistry.* 1985;24:7396-7407.

[38] Jacob, R. E., and Engen, J. R. Hydrogen exchange mass spectrometry: are we out of quicksand? *J Am Soc Mass Spectrom*. 2012;23:1003-1010.

[39] Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S., and Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A.* 1986;83:6233-6237.

[40] Wales, T. E., and Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev.* 2006;25:158-170.

[41] Asuru, A. P., Anand, M., and Busenlehner, L. S. Dissection of porphyrin-induced conformational dynamics in the heme biosynthesis enzyme ferrochelatase. *Biochemistry.* 2012;51:7116-7127.

[42] Busenlehner, L. S., Branden, G., Namslauer, I., Brzezinskiand, P., and Armstrong, R. N. Structural elements involved in protein translocation by cytochrome c oxidase as revealed by backbone amide hydrogen-deuterium exchange of the E286H mutant. *Biochemistry.* 2008;47:73-83.

[43] Thevenon-Emeric, G., Kozlowski, J., Zhang, Z., and Smith, D. L. Determination of amide hydrogen exchange rates in peptides by mass spectrometry. *Anal Chem.* 1992;20:2456-2458.

[44] Liu, S., Liu, L., Uzuner, U., Zhou, X., Gu, M., Shi, W., Zhang, Y., Daiand, S. Y., and Yuan, J. S. HDX-analyzer: a novel package for statistical analysis of protein structure dynamics. *BMC Bioinformatics.* 2011;12 Suppl 1:S34.

[45] Weis, D. D., Engenand, J. R., and Kass, I. J. Semi-automated data processing of hydrogen exchange mass spectra using HX-Express. *J Am Soc Mass Spectrom.* 2006;17:1700-1703.

[46] Pascal, B. D., Willis, S., Lauer, J. L., landgraf, R. R., West, G. M., Marciano, D., Novick, S., Goswami, D., Chalmersand, M. J., and Griffin, P. R. HDX workbench: software for the analysis of H/D exchange MS data. *J Am Soc Mass Spectrom.* 2012;23:1512-1521.

[47] Pascal, B. D., Chalmers, M. J., Busby, S. A., Mader, C. C., Southern, M. R., Tsinoremasand, N. F., and Griffin, P. R. The Deuterator: software for the determination of backbone amide deuterium levels from H/D exchange MS data. *BMC Bioinformatics*. 2007;8:156.

[48] Dunn, M. F. (April 2010) Protein–Ligand Interactions: General Description. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0001340.pub2

[49] Frimurer, T. M., Peters, G. H., Peters, G. H., Iversen, L. F., Andersen, H. S., Moller, N. P. H., and Olsen, O. H. Ligand-induced conformational changes: improved predications of ligand binding conformations and affinities. *Biophys J.* 2003;84:2273-2281.

[50] Keskin, O. Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies. *BMC Struct Biol.* 2007;7:31.

[51] Berg, J. M., Tymoczkoand, J. L., and Stryer, L. *Biochemistry*. 5th ed. 2002, New York: W.H. Freeman. 1 v. (various pagings).

[52] Percy, A. J., Rey, M., Burnsand, K. M., and Schriemer, D. C. Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-a review. *Anal Chim Acta*, 2012;721:7-21.

[53] Keskin, O., Ma, B., Rogale, K., Gunasekaranand, K., and Nussinov, R. Protein-protein interactions: organization, cooperativity and mapping in a bottom-up Systems Biology approach. *Phys Biol*, 2005;2:S24-35.

[54] Olsson, T. S. G., Williams, M. A., Pitt, W. R., and Ladbury, J. E. The thermodynamics of protein-ligand interaction and solvation: insights for ligand design. *J Mol Biol.* 2008;384:1002-1017.

[55] Lee, C. K., Wang, Y. M., Huang, L. S., and Lin, S. Atomic force microscopy: determination of unbinding force, off rate and energy barrier for protein-ligand interaction. *Micron.* 2007;38:446-461.

[56] Phizicky, E. M., and Fields, S. Protein-protein interactions: methods for detection and analysis. *Microbiol Rev*. 1995;59:94-123.

[57] Lalonde, S., Ehrhardt, D. W., Loque, D., Chen, J., Rheeand S. Y., and Frommer, W. B. Molecular and cellular approaches for the detection of protein-protein interactions: latest techniques and current limitations. *Plant J*, 2008;53:610-635.

[58] Breton, C., Snajdrova, L., Jeanneau, C., Koe,a J., and Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology.* 2006;16:29R-37R.

[59] Rini, J., Esko, J., and Varki, A. (2009) in Essentials of Glycobiology (Varki, A., Cummings, R., Esko, J., Freeze, H., Stanley, P., Bertozzi, C. R., Hart, G., Etzler, M. E., editors. , eds) 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[60] Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol.* 2003;328:307–317.

[61] Galan, M. C., Venot, A. P., and Boons, G. J. Glycosyltransferase activity can be modulated by small conformational changes of acceptor substrates. *Biochemistry.* 2003;4: 8522–8529.

[62] Tarbouriech, N., Charnock, S. J., and Davies, G. J. Three-dimensional structures of the mn and mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: A comparison with related NDP-sugar glycosyltransferases. *J Mol Biol.* 2001;314:655-661.

[63] Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. Glycosyltransferases: Structures, functions, and mechanisms. *Annu Rev Biochem*. 2008;77:521-555.

[64] McMurry, J. E. Organic Chemistry (3rd edition). 1992.

[65] Zechel, D. L., and Withers, S. G. Glycosidase mechanisms: Anatomy of a finely tuned catalyst. *Acc Chem Res.* 2000;33:11–18.

[66] Davies, G. S., and Withers, S. G. Comprehensive biological catalysis. In: Sinnott ML, ed. London: Academic. 1998;119-208.

[67] Cowdrey, W. A., Hughes, E. D., Ingold, C. K., Masterman, S., and Scott, A. D. Reaction kinetics and the walden inversion. Part VI. Relation of steric orientation to mechanism in substitutions involving halogen atoms and simple or substituted hydroxyl groups. *J Chem Soc.* 1937;1253–1271.

[68] Hughes, E. D., Ingold, C. K., and Whitfield, I. C. The walden inversion in the replacement of hydroxyl by halogen. *Nature.* 1941;147:206-207.

[69] Lewis, E. S., Boozer, C. E. The kinetics and stereochemistry of the decomposition of secondary alkyl chlorosulfites. *J Am Chem Soc.* 1952;74:308-311.

[70] Sinnott, M. L., and Jencks, W. P. Solvolysis of D-glucopyranosyl derivatives in mixtures of ethanol and 2, 2, 2-trifluroethanol. *J Am Chem Soc*. 1980;102:2026-2032.

[71] Klein, H. W., Im, M. J., and Palm, D. Mechanism of the phosphorylase reaction. *Eur J Biochem*. 1986;157:107-114.

[72] Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat, B. A classification of nucleotide-diphospho-sugar based on amino acid sequences similarities. *Biochem J.* 1997;326*:*929–939.

[73] Bourne, Y., and Henrissat, B. Glycoside hydrolases and glycosyltransferases: families and functional modules. *Curr Opin Struc Biol.* 2001;11:593–600.

[74] Unligil, U. M., and Rini, J. M. Glycosyltransferase structure and mechanism. *Curr Opin Struc Biol.* 2000;10:510–517.

[75] Hu, Y. N., and Walker, S. Remarkable Structural Similarities between Diverse Glycosyltransferases. *Chem Biol.* 2002;9:1287–1296.

[76] Charnock, S. J., and Davies, G. J. Structure of the nucleotide-diphosphosugar transferase, SpsA from *Bacillus subtilis* in native and nucleotide complexed forms. *Biochemistry*. 1999;38:6380-6385.

[77] Breton, C., and Imberty, A. Structure/function studies of glycosyltransferases. *Curr Opin Struct Biol*. 1999;9:563-571.

[78] Wiggins, C. A. R., and Munro, S. Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc Natl Acad Sci U S A*. 1998;95:7945-7950.

[79] Breton, C., Bettler, E., Joziasse, D. H., Geremia, R. A., and Imberty, A. Sequence/function relationships of prokaryotic and eukaryotic galactosyltransferases. *J Biochem*. 1998;123:1000-1009.

[80] Vrielink, A., Ruger, W., Driessen, H. P. C., and Freemont, P. S. Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of substrate uridine diphosphoglucose. *EMBO J*. 1994;13:3413-3422.

[81] Ayala-Castro, C., Saini, A., and Outten, F. W. Fe-S assembly pathways in bacteria. *Microbiology and Molecular Biology Reviews*. 2008:72:110-125.

[82] Wachtershauser, G. Groundworks for an evolutionary biochemistry: the iron-sulphur world. *Prog. Biophys. Mol. Biol.* 1992:58:85-201.

[83] Beinert, H., and Kiley, P. J. Fe-S proteins in sensing and regulatory functions. *Curr. Opin. Microbiol.* 1999:3:152-157.

[84] Ruzicka, F. J., and Beinert, H. The soluble "high potential" type iron-sulfur protein from mitochondria is aconitase. *J. Biol. Chem*. 1978:253:2514-2517.

[85] Kiley, P. J., and Beinert, H. The role of Fe-S proteins in sensing and regulation in bacteria. *Curr. Opin. Microbiol.* 2003:6:181-185.

[86] Moulis, J. M., Davasse, V., Golinelli, M. P., Meyer, J., and Quinkal, I. The coordination sphere of iron-sulfur clusters: lessons from site-directed mutagenesis experiments. *J. Biol. Chem*. 1996:1:2-14.

[87] Malkin, R., and Rabinowitz, J. C. The reconstitution of clostridial ferredoxin. *Biochem. Biophys. Res. Commun.* 1966:23:822-827.

[88] Johnson, D. C., Dean, D. R., Smith, A. D., and Johnson, M. K. Structure, function, and formation of biological iron-sulfur clusters. *Annu. Rev. Biochem.* 2005:74:247-281.

[89] Tokumoto, U., Kitamura, S., Fukuyama, K., and Takahashi, Y. Interchangeability and distinct properties of bacteria Fe-S cluster assembly systems: functional replacement of the *isc* and *suf* operons in *Escherichia coli* with the *nifSU*-like operon from *Helicobacter pylori*. *J. BioChem.* 2004:136:199-209.

[90] Fontecave, M., de Choudens, S. O., and Barras, B. Py. F. Mechanism of iron-sulfur cluster assembly: the SUF machinery. *J. Biol. Inorg. Chem.* 2005:10:713-721.

[91] Nachin, L., E1 Hassouni, M., Loiseau, L., Expert, D., and Barras, F. SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of the SufC, an orphan ABC ATPase. *Mol. Microbiol.* 2001:39:960-972.

[92] Nachin, L., Loiseau, L., Expert, D., and Barras, F. SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. *Embo. J.* 2003:22:427-437.

[93] Outten, F. W., Djaman, O., and Storz, G. A *suf* operon requirement for Fe-S cluster assembly during iron starvation in *Escherichia coli*. *Mol. Microbiol.* 2004:25:861-872.

[94] Takahashi, Y., and Tokumoto, U. A third vacterial system for the assembly of iron-sulfur clusters with homologues in archaea and plastids. *J. Biol. Chem.* 2002:277:28380-28383.

[95] Imlay, J. A. Iron-sulphur clusters and the problem with oxygen. *Mol. Microbiol.* 2006:59:1073-1082.

[96] Urresti, S., Albesa-Jove, D., Schaeffer, F., Pham, H. T., Kaur, D., Gest, P., van der Woerd, M. J., Carreras-Gonzalez, A., Lopez-Fernandez, S., Alzari, P. M., Brennan, P. J., Jackson, M., and Guerin, M. E . Mechanistic insights into the retaining glucosyl-3-phosphoglycerate synthase from mycobacteria. *J Biol Chem.* 2012;287:24649-24661.

[97] Vetting, M. W., Frantom, P. A., and Blanchard, J. S. Structural and enzymatic analysis of MshA from *Corynebacterium glutamicum*. *J Biol Chem*. 2008;283:15834-15844.

[98] Hirabayashi, K., Yuda, E., Tanaka, N., Katayama, S., Iwasaki, K., Matsumoto, T., Kurisu, G., Outten, F. W., Fukuyama, K., Takahashi, Y., and Wada, K. Functional dynamics revealed by the structure of the SufBCD complex, a novel ATP-binding cassette (ABC) protein that serves as a scaffold for iron-sulfur cluster biogenesis. *J Biol Chem.* 2015;290:29717-29731.

CHAPTER 2

INSIGHTS INTO STRUCTURE AND DYNAMICS OF RETAINING GLYCOSYLTRANSFERASE GLUCOSYL-3-PHOSPHOGLYCERATE SYNTHASE FROM *MYCOBACTERIUM TUBERCULOSIS*

2.1 Introduction

2.1.1 Significance of MGLPs for *Mycobacterium tuberculosis*

Tuberculosis is a widespread and infectious disease caused by various strains of mycobacterium. It is a serious threat to human health worldwide, accounting for over one million deaths annually. [1] With the advent of the antibiotic era, tuberculosis was treatable, and at one point eradication was even believed to be possible. [2] However, tuberculosis has reemerged as a human health threat due to the emergence of extensively drug-resistant strains that are almost untreatable with current therapies. [3] In recent years, it is even becoming prevalent in the world due to the lack of development of novel drugs against *M. tuberculosis*. [3] In order to resist the resurgence of tuberculosis, new drugs for *M. tuberculosis* are urgently needed.

Some unique carbohydrates produced by *M. tuberculosis* have become a source of potential targets for new drugs against the tuberculosis. [4] These unique carbohydrates are important for the growth of mycobacteria. One type of unusual carbohydrate is polymethylated polysaccharides (PMPS). PMPS are cytoplasmic polysaccharides of intermediate size composed of 10-20 sugar units. Many of sugar units are partially methylated thus leading to PMPS having a slight hydrophobicity. [4-5] Mycobacteria produce two classes of PMPS. One class is the 6-*O*-methylglucose lipopolysaccharides (MGLPs), and the second is the 3-*O*-methylmannose polysaccharides (MMPs). MGLPs are made up of 16-20 hexose units that have esterified acetate,

octanoate, propionate, isobutyrate, and succinate (Illustration 2.1). [6] MMPs are linear unbranched chains of 11-14 sugar units, each terminated by a single mannose at the nonreducing end and by a methyl aglycon at the reducing end. [7] Both MGLPs and MMPs are able to form stable 1:1 complex with long-chain fatty acids and acyl-coenzyme A derivatives *in vitro*, leading to the suggestion that they may be an important regulator of lipid metabolism in mycobacteria. Moreover, MGLPs can sequester acyl-coenzyme A, protecting them from degradation in the cytoplasm. [4, 8-9] Due to the important role of MGLPs in the existence and growth of mycobacteria, enzymes involved in MGLP biosynthetic pathway have stimulated the interest of scientists from the perspective of potential drug targets. [10]

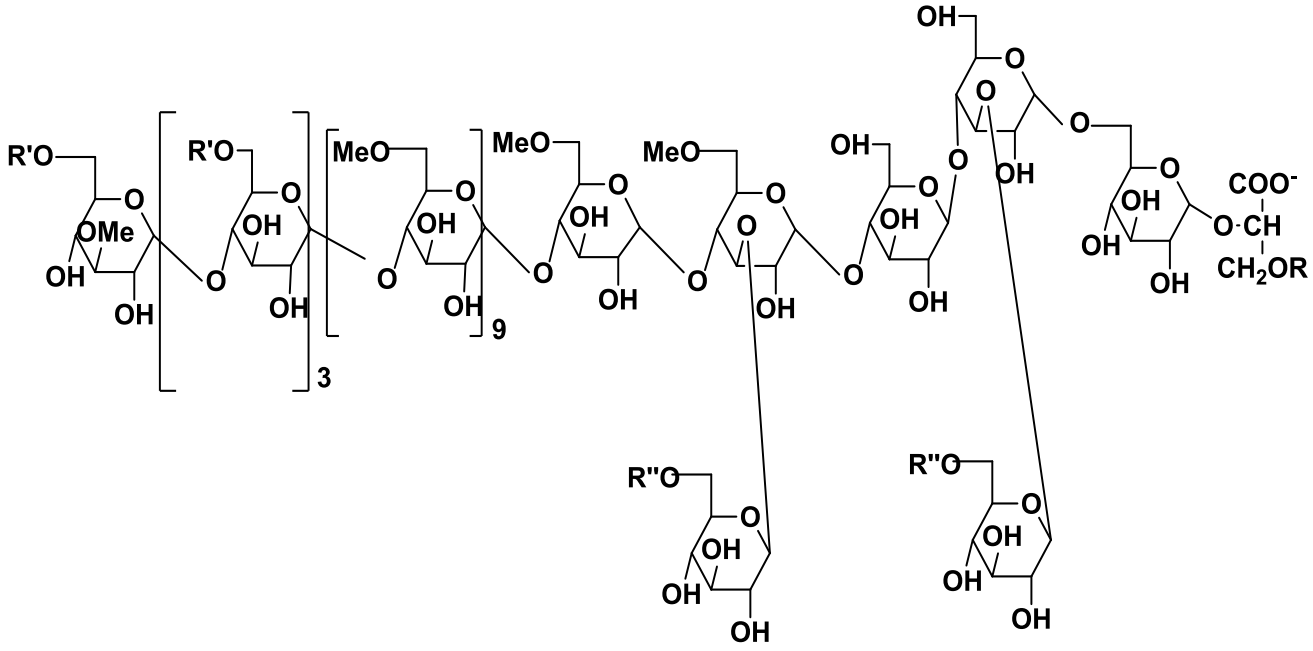


Illustration 2.1. Structure of MGLPs, in which R, R′ and R″ are acyl groups. R and R″ represent octanoate and succinate respectively. R′ represent acetate, propionate, or isobutyrate. Made with ChemDraw Professional 15.0

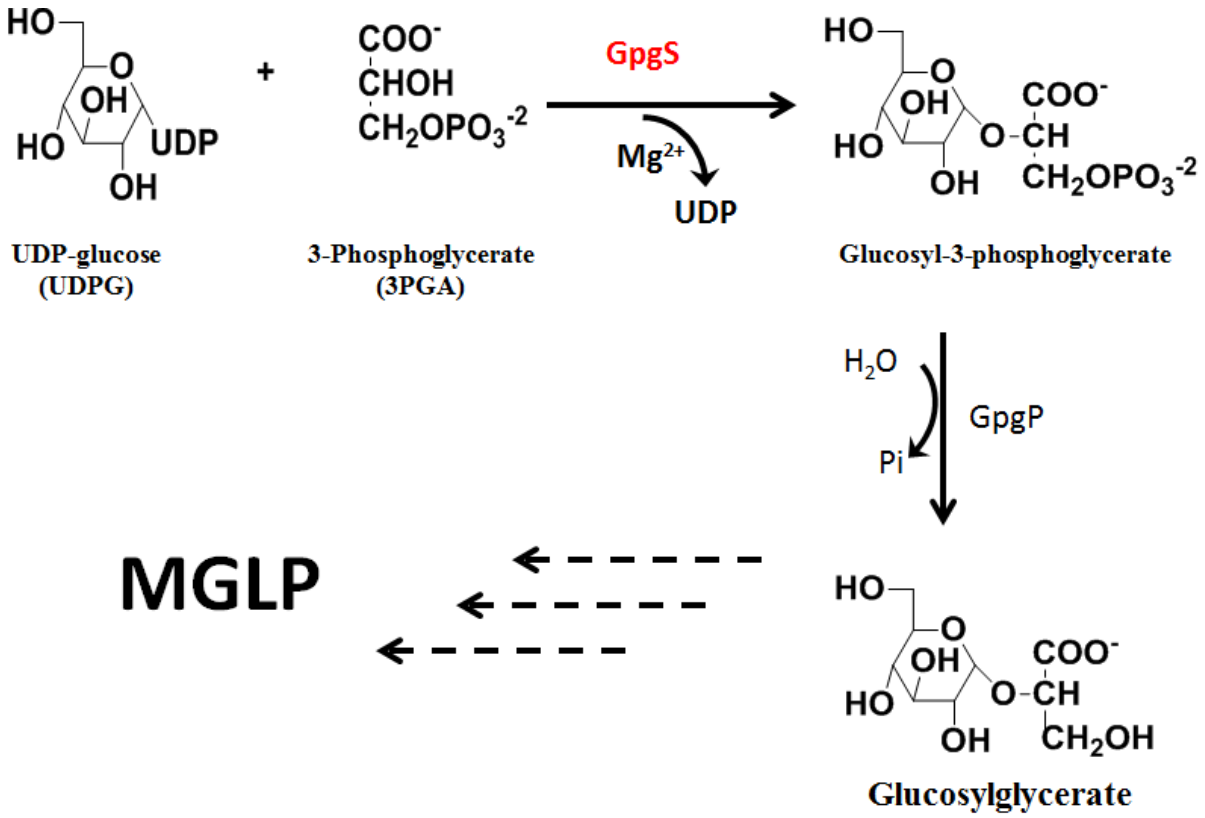


Illustration 2.2. Formation of MGLPs. Glucose is transferred to 3PGA by GpgS. The formed glucosyl-3-phosphoglycerate is dephosphorylated to GG by GpgP. With more reactions, MGLP is generated. Made with ChemDraw Professional 15.0

2.1.2 Biosynthesis of MGLPs

In the pathway of MGLP biosynthesis, glucosylglycerate (GG) is the putative precursor for MGLP synthesis. [11-13] GG is considered essential for *M. tuberculosis* growth and has been identified to be important in other organisms. [14-16] The biosynthesis of GG has a two-step process (Illustration 2.2). [17] In the first step, glucosyl-3-phosphoglycerate synthase (GpgS) transfers the glucose moiety from donor UDP-glucose (UDPG) to acceptor 3-phosphoglycerate (3PGA). The glucose adds to hydroxyl group in carbon 2 of 3PGA and forms glucosyl-3-phosphoglycerate (GPG). In the second step, the phosphate group of GPG is removed by glucosyl-3-phosphoglycerate phosphatase (GpgP) and GG is generated. Then, GG goes through

reactions catalyzed by unknown enzymes and MGLP is produced. As one of two known enzymes involved in MGLP biosynthesis pathway, GpgS from *M. tuberculosis* (*Mt*GpgS) is attractive to be studied. Indeed, the interruption of gene encoding GpgS in *Mycobacterium smegmatis* resulted in bacteria with a severely deficient growth rate. [18-19]

2.1.3 Classification of *Mt*GpgS

*Mt*GpgS belongs to family of glycosyltransferase enzymes. As discussed in chapter 1, GTs are classified into 90 families based on amino acid sequence similarities. [20] *Mt*GpgS is assigned to the GT81 family, which was recently created. [20-22] Moreover, it is a retaining glycosyltransferase due to the retention of configuration at the anomeric center of glycosyl groups. [20] Based on the structural characterization, GTs have two general folds, as discussed in chapter 1. *Mt*GpgS was revealed to possess a GT-A like domain with DXD motif that is a characteristic of GT-A fold. [10, 23]

2.1.4 Crystal Structure of *Mt*GpgS

The crystal structure of *Mt*GpgS has been reported to possess a GT-A like fold, which contains a single Rossmann fold domain (Figure 2.1). [10] The structure of *Mt*GpgS can be subdivided into two closely associated sub-domains, an *N*-terminal nucleotide domain (residues 45-137), resembling the Rossmann-like fold, and a *C*-terminal domain (residues 138-281), consisting of a mixture of α/β secondary structure. [10] Like other GT-A glycosyltransferases, the *N*-terminal domain recognizes the nucleotide donor substrate, and the *C*-terminal domain is involved in the acceptor substrate recognition.

Some residues of *Mt*GpgS are proposed to contribute to substrate binding (Figure 2.2). The *N*-terminal domain contains a $D^{134}X^{135}D^{136}$ motif, in which Asp134 forms a stable hydrogen-bonded network with Lys114 and Glu232. This network provides one side for the

nucleotide binding pocket. The second aspartate coordinates with metal cation Mg^{2+} which has interactions with the α and β phosphates of UDP. The phosphate moiety of UDP is further stabilized by polar interactions between the α phosphate and side chain of Tyr229 and between the α and β phosphate groups and side chain of Arg261. [10] The residues Val 186 and Thr187 are involved in acceptor recognition. The side chain of Thr187 forms a hydrogen bond with the carboxyl group of 3PGA. [10] A loop from residues 253-262 was identified to play an important role in substrate binding and catalysis since it adopts two conformations in dimer. [23] In this loop, residues Arg256 and Arg261 were observed to form hydrogen bonds with side chain of Asp136 and Tyr229, respectively. These hydrogen bonds may contribute to the two conformations of the loop in dimer. [23]

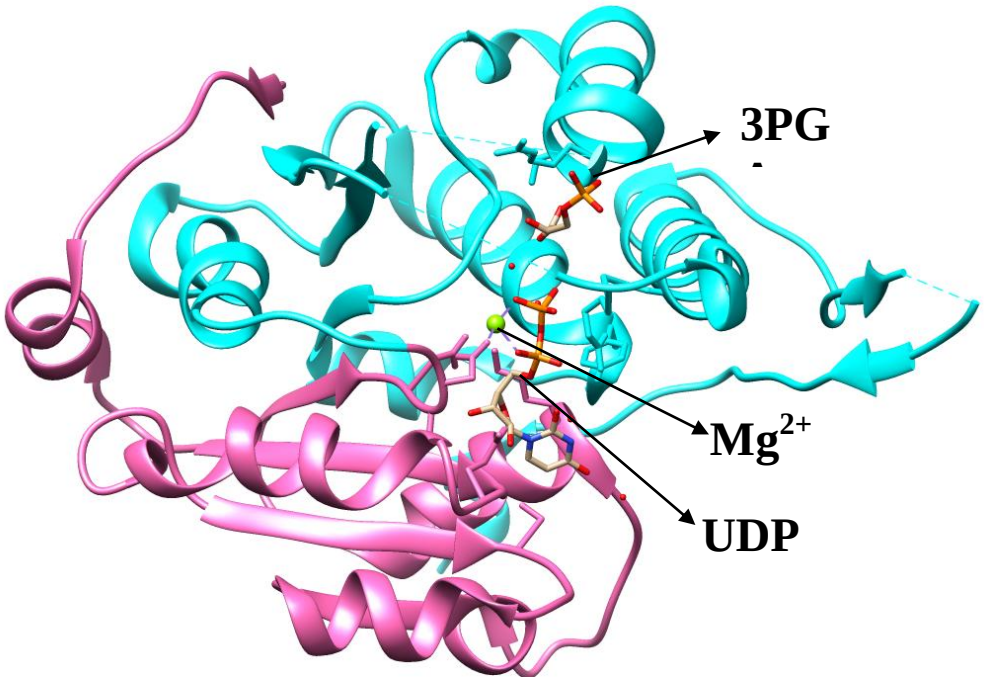


Figure 2.1. Crystal structure of *Mt*GpgS monomer in complex with UDP, 3PGA, and Mg^{2+}. Pink color represents the *N*-terminal domain and cyan color represents the *C*-terminal domain. Made with Chimera. PDB: 3e25

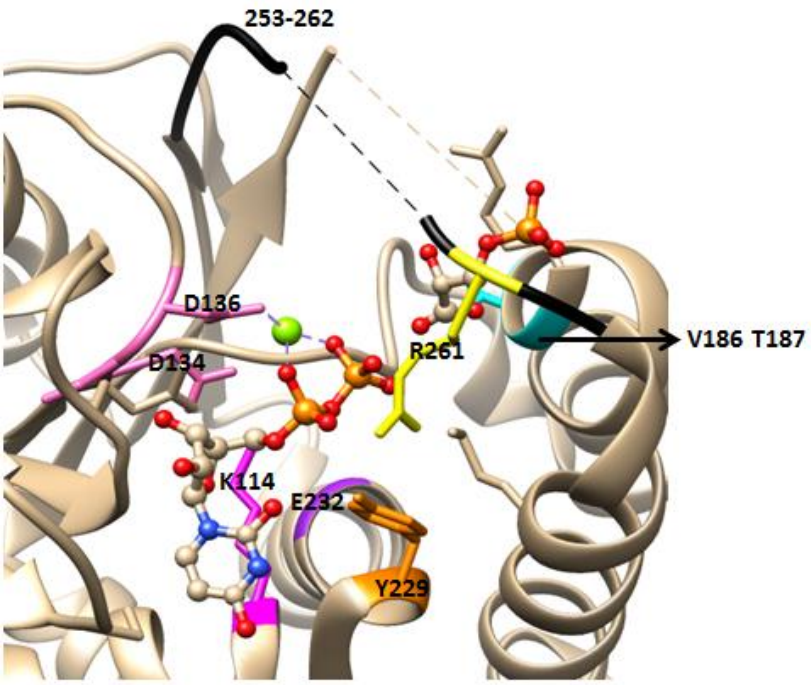


Figure 2.2. Active site residues mapped in crystal structure of *Mt*GpgS. Made with Chimera. PDB: 3e25

2.1.5 Conclusion

Although much information is known about the structure of *Mt*GpgS, a dynamic characterization of this enzyme is still lacking. By observing *Mt*GpgS APO and complex structures, only the loop 253-262 was observed to have a conformational change on ligand binding. [23] However, *Mt*GpgS is expected to have more dynamic and conformational changes on substrate binding, which cannot be detected by static structures. Moreover, the structural differences between GT-A and GT-B glycosyltransferases have been detected by comparing crystal structures, but the differences in dynamic and conformational changes on substrate binding are still unknown. Therefore, HDX-MS was applied to study dynamic changes of *Mt*GpgS on UDPG and 3PGA binding. The HDX-MS data can provide important clues into the substrate binding mechanism and conformational changes of *Mt*GpgS. Moreover, the dynamic clarification of *Mt*GpgS promotes the dynamic investigation of GT-A glycosyltransferases.

2.2 Materials and Methods

2.2.1 Materials

Primers for amplification and mutagenesis of Rv1208 were obtained from Eurofins Genomics (Huntsville, AL). UDP-Glucose was purchased from Merck Millipore (Billerica, MA). 3PGA, NADH, phosphoenolpyruvate (PEP), deuterium oxide (D_2O), and porcine pepsin (3200-4500 u/mg) were purchased from Sigma-Aldrich (St. Louis, MO). The coupling enzymes pyruvate kinase (PK) and lactate dehydrogenase (LDH) were obtained from Roche Diagnostics (Indianapolis, IN). The HisTrap HP and the Q Sepharose HP column were obtained from GE Healthcare (Piscataway, NJ). Microbore C18 reverse phase column with internal diameter between 0.5-2.0 mm and a guard column were bought from Phenomenex (Torrance, CA). Competent *E. coli* cells (BL21(DE3)pLysS and XL-10Gold) were purchased from Invitrogen (Grand Island, NY). All other reagents were obtained from VWR (Radnor, PA).

2.2.2 Amplification and Overexpression of *Mt*GpgS

Amplification of the gene Rv1208 encoding GpgS has been described previously except that gene Rv1208 was cloned into plasmid pET28. [10] The amplified Rv1208-pET28 was transformed into BL21(DE3)pLysS *E. coli* cells and grown in LB media with 0.03 mg/ml of kanamycin and 0.035 mg/ml of chloramphenicol at 37 °C. When OD_{600} reached 0.8, 0.5 mM isopropyl-b-D1-thiogalactopyranoside was added. The LB media was incubated while shaking for 8 hours at 30 °C. Finally the cells were harvested by centrifugation at 6330 ×g for 10 minutes at 4 °C.

2.2.3 Purification of *Mt*GpgS

Cell pellets were resuspended in lysis buffer (20 mM sodium phosphate (pH 7.4), 0.5 M sodium chloride, and 20 mM imidazole), lysozyme (0.25 mg/ml), phenylmethylsulfonyl fluoride

(1mM, PMSF), DNase (10 µg/ml) and few crystals of magnesium chloride. The resuspended cells were sonicated, and the supernatant was collected by centrifugation at 34540 ×g for 30 minutes at 4 ºC. *Mt*GpgS was purified by using Ni^{2+} HisTrap column (5 ml) with the linear gradient of 20 column volumes of elution buffer (20 mM sodium phosphate (pH 7.4), 0.5 M sodium chloride, and 1 M imidazole). The purity of protein was checked by SDS-PAGE gel, and the fractions containing *Mt*GpgS were dialyzed against 20 mM sodium phosphate (pH 6.99). *Mt*GpgS was further purified by using anion exchange Q Sepharose HP column with the linear gradient of 20 column volumes of elution buffer (20 mM sodium phosphate (pH 6.99) and 1 M sodium chloride). The purest fractions were pooled and dialyzed against 20 mM sodium phosphate (pH 7.4) and 200 mM sodium chloride. Finally, *Mt*GpgS was concentrated by ultracentrifugation to the concentration of about 200 µM and stored with 20% glycerol in -20 ºC.

2.2.4 Site-directed Mutagenesis of *Mt*GpgS

Oligonucleotide primer of each substitution was designed and listed in table 2.1. These primers were ordered from Eurofins Genomics. The substitutions in Rv1208-pET28 plasmid were created by using QuikChange Lightning site-directed mutagenesis kit (Stratagene). The mutated Rv1208-pET28 plasmids were transformed into XL-10 Gold *E. coli* cells and grown in LB media containing 0.03 mg/ml kanamycin with shaking overnight at 37 ºC. The Rv1208-pET28 plasmids were purified and sequenced (Eurofins Genomics, Huntsville, AL) in order to confirm the presence of target mutation. Then the plasmids with correct mutations were transformed into BL21(DE3)pLysS *E. coli* cells. The overexpression and purification of *Mt*GpgS variants was similar to that of wild type *Mt*GpgS.

2.2.5 Steady State Assays

The condition for kinetic assays of wild type *Mt*GpgS was 50 mM of triethanolamine

(TEA) (pH 7.8), 100 μM of NADH, 250 μM of phosphoenol pyruvate (PEP), 20 mM of magnesium chloride, 20 units of pyruvate kinase (PK), 55 units of lactate dehydrogenase (LDH), and saturating concentration of each substrate (UDP-Glucose and 3-PGA) in a 1 ml reaction at 25 ˚C. The reaction was initiated by addition of enzyme *Mt*GpgS (Illustration 2.3). The released UDP has a reaction with PEP and the pyruvate consumes NADH. The velocity of formation of UDP was measured by the consumed amount of NADH at 340 nm (ε=6.22 $mM^{-1}cm^{-1}$).

Table 2.1. Sequences of forward and reverse primers of each *Mt*GpgS Substitution.

*Mt*GpgS Substitution Primer	Primer Sequence (5′ - 3′)
D134A Forward	- catcgtggtgttcatcgcctcagacctgatcaacc -
D134A Reverse	- ggttgatcaggtctgaggcgatgaacaccacgatg -
D136A Forward	- gtgttcatcgactcagccctgatcaacccgcac -
D136A Reverse	- gtgcgggttgatcagggctgagtcgatgaacac -
T187A Forward	- gcgggagggtcgccgagctggtg -
T187A Reverse	- caccagctcggcgaccctcccgc -
R256A Forward	- caggtcaacttgggcgttgcggcgcaccgt -
R256A Reverse	- acggtgcgccgcaacgcccaagttgacctg -
R261A Forward	- gggcgcaccgtaacgcgcccctagacga -
R261A Reverse	- tcgtctaggggcgcgttacggtgcgccc -

2.2.6 Determination of Kinetic Parameters

Kinetic parameters were determined by fitting initial velocities of reactions to the Michaelis-Menten equation (Equation 2.1) using Kaleidagraph (Synergy Software, Reading, PA).

$$v = \frac{k_{\text{cat}}[S]}{K_m + [S]} \qquad \text{Equation 2.1}$$

In the equation, v is the velocity of reaction corresponding to varied substrate concentration, $[S]$ is the concentration of varied substrate, k_{cat} is the maximum velocity of the reaction, and K_m is the Michaelis-Menten constant.

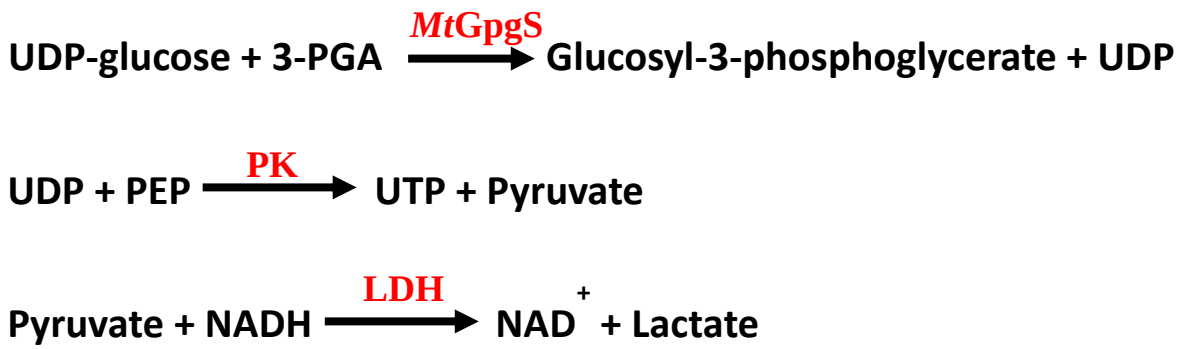


Illustration 2.3. Schematic of PK/LDH coupled assay measuring enzymatic activity of *Mt*GpgS.

2.2.7 Sequencing of *Mt*GpgS by MS/MS

The peptide map of *Mt*GpgS was determined before the continuous HDX-MS experiments. The stock *Mt*GpgS was diluted to 20 µM with 20 mM of TEA (pH 7.8) and 200 mM of NaCl. Fresh 5 mg/ml of pepsin was dissolved in 0.01 M of potassium phosphate (pH 7). The HDX reaction was initiated with 50 µl of 20 µM *Mt*GpgS to 50 µl of quench solution (0.1 M potassium phosphate, pH 2.4). The digestion started with addition of 6 µl of 5 mg/ml pepsin and lasted for 5 minutes. The final concentrations of *Mt*GpgS and pepsin were 10 µM and 0.3 mg/ml respectively. The generated peptides were loaded onto a 2 mm x 50 mm C18 reverse-phase column, which had been equilibrated with solvent A (98% H_2O, 2% acetonitrile, and 0.4% formic acid). The peptides were separated by solvent B (98% acetonitrile, 2% H_2O, and 0.4% formic acid) with a linear gradient 0% to 50% for 20 minutes (Table 2.2). The separated peptides were sequenced using a HCT Ultra PTM Discovery ion trap mass spectrometer. The ion trap scanning range was set at 100 – 2000 m/z in positive ion mode. The maximum accumulation

time was set as 100 ms with an average of 5 spectra. The nebulizer pressure, drying gas flow rate, and capillary temperature were set as 28 psi, 7 L/min, and 350 ºC, respectively. The number of precursor ions selected for CID was set as 6. The MS/MS data was analyzed using Peaks Client software, and the peptide map of *Mt*GpgS was determined.

2.2.8 Continuous HDX-MS Experiments

The continuous HDX-MS experiments were initiated by the addition of 5 µl of 200 µM free *Mt*GpgS into 45 µl of 99.5% D_2O in a PCR tube. D_2O was preheated at 25 ºC water bath. The protein was incubated in D_2O at 25 ºC for various time points (15s – 1h). The isotopic exchange was quenched by decreasing pH to 2.4 and temperature to 0 °C with addition of 50 µl of quench solution in ice bath. The protein was digested on ice by 6 µl of 5 mg/ml pepsin for 5 minutes. The digested peptides were separated by HPLC with elution method listed in Table 2.3. The mass spectrometer was set in positive mode with ion trap scanning range 300 – 1500 m/z. The nebulizer pressure, drying gas flow rate, and capillary temperature were set as 28 psi, 7 L/min, and 250 ºC, respectively. The HDX-MS experiments of *Mt*GpgS bound by UDPG or 3PGA had the same procedures with free *Mt*GpgS except the protein sample used to incubate in D_2O.100 µl of 200 µM free *Mt*GpgS was incubated with 10 µl of 100 mM UDPG and 4 µl of 1 M $MgCl_2$ for 30 min, in order to make sure UDPG and Mg^{2+} bound to *Mt*GpgS. Similarly, 100 µl of 200 µM free *Mt*GpgS was incubated with 10 µl of 100 mM 3PGA for 30 min to make 3PGA bound to *Mt*GpgS. Then, these protein complexes were incubated with D_2O, starting the isotopic exchange as described above. All protein samples were prepared individually and one set of continuous HDX-MS experiments was run within one day.

The two control experiments corresponding to no deuteration ($m_{0\%}$), and full deuteration ($m_{100\%}$) were also run. The m_0 control was run as the normal sample except that 5 µl of 200 µM

*Mt*GpgS was dissolved H_2O instead of D_2O. The m_{100} control was run with the same procedures except that *Mt*GpgS was incubated in D_2O at 50°C for 8 hours to make sure all possible amide hydrogens exchanged with deuterium. Then, the reaction was quenched and digested as described above.

Table 2.2. Elution method of MS/MS for *Mt*GpgS.

Elution Time (min)	% Solvent A	% Solvent B
0.00	100	0
1.00	100	0
21.00	50	50
22.00	0	100
32.00	0	100
33.00	100	0
40.00	100	0

Table 2.3. Elution method of continuous HDX-MS experiments.

Elution Time (min)	% Solvent A	% Solvent B
0.00	100	0
1.00	100	0
13.00	50	50
14.00	0	100
24.00	0	100
25.00	100	0
35.00	100	0

2.2.9 Analysis of HDX-MS Data

The software HD Examiner was applied to analyze HDX-MS data. The protein sequence in fasta format, peptide list with m/z ratio and RT, and .mzxml format files of all HDX-MS experiments with continuous incubation time and two controls were input into HD Examiner. The software in fasta searched the experimental isotopic envelope of each peptide according to its m/z ratio, sequence, and RT, which was fit to theoretic isotopic envelope. The centroid of each peptide was calculated and employed to calculate the amount of deuterium incorporated, which was plotted versus incubation time by fitting to exponential equation. The dynamic plots of free *Mt*GpgS, *Mt*GpgS-UDPG-Mg^{2+}, and *Mt*GpgS-3PGA were compared to investigate the conformational changes of *Mt*GpgS on substrate binding.

2.3 Results

2.3.1 Steady State Kinetics of Wild Type *Mt*GpgS and Variants

The PK-LDH coupled assay was employed to determine the kinetic parameters of wild type *Mt*GpgS by the consumption of NADH. The k_{cat} value is 50 ± 2 min^{-1}. The K_m values for UDP-glucose and 3-PGA are 0.80 ± 0.06 mM and 0.20 ± 0.03 mM, respectively (Figure 2.3). As discussed in 2.1.4, the $D^{134}X^{135}D^{136}$ motif and the loop 253-262 were proposed to contribute to substrate binding. So, the residues D134, D136, R256, and R261 were selected to study their specific roles in *Mt*GpgS activity. Moreover, the residue T187 was proposed to be involved in acceptor recognition. Therefore, in order to understand the roles of these active site residues, five mutations D134A, D136A, T187A, R256A, and R261A were designed and constructed on *Mt*GpgS (Figure 2.4). Alanine was utilized to substitute target residues since it is small and neutral. The kinetic parameters of five *Mt*GpgS variants were determined by PK-LDH assays and shown in Table 2.4. Substitution of any aspartate residue from DXD motif, which plays an

important role in UDP and Mg^{2+} binding, results in undetectable activity. The T187A variant exhibits a 100-fold decrease in the value of k_{cat} relative to wild type *Mt*GpgS, while its K_m values for UDPG and 3PGA do not show obvious changes. Both residues R256 and R261 are from a flexible loop (253-262), which is proposed to be critical in substrate binding and catalysis. [23] The k_{cat} and K_m values of R256A variant are not obtainable due to high K_m values, but k_{cat}/K_m values with varying UDPG and 3PGA concentrations were determined. The k_{cat}/K_{UDPG} and k_{cat}/K_{3PGA} values are reduced by 636- and 2388-fold, respectively, relative to wild type *Mt*GpgS. However, the R261A variant does not show a dramatic kinetic difference with wild type. The k_{cat} value only decreases by 2-fold, and K_m values corresponding to UDPG and 3PGA increase by 2- and 5-fold, respectively.

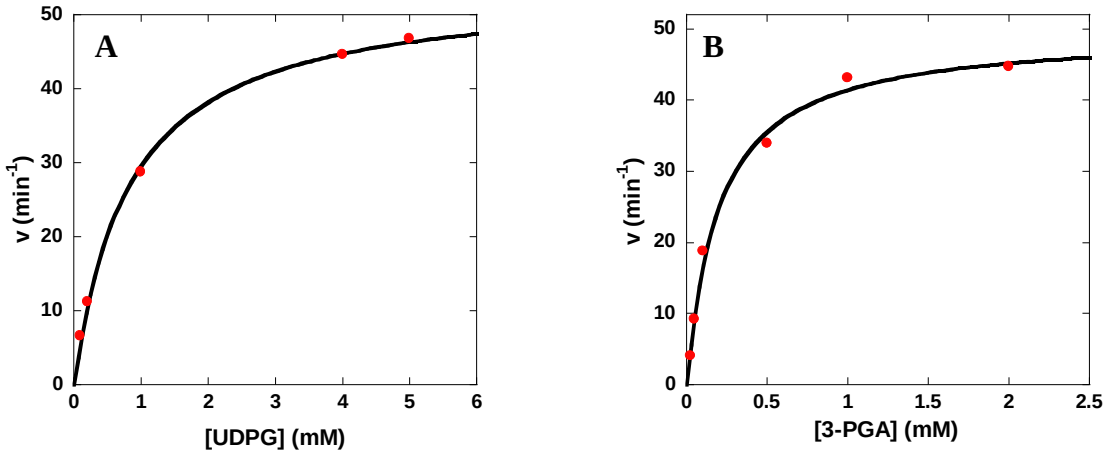


Figure 2.3. Michaelis-Menten plots of wild type *Mt*GpgS. The initial velocity for UDPG and 3PGA were shown in (A) and (B) respectively.

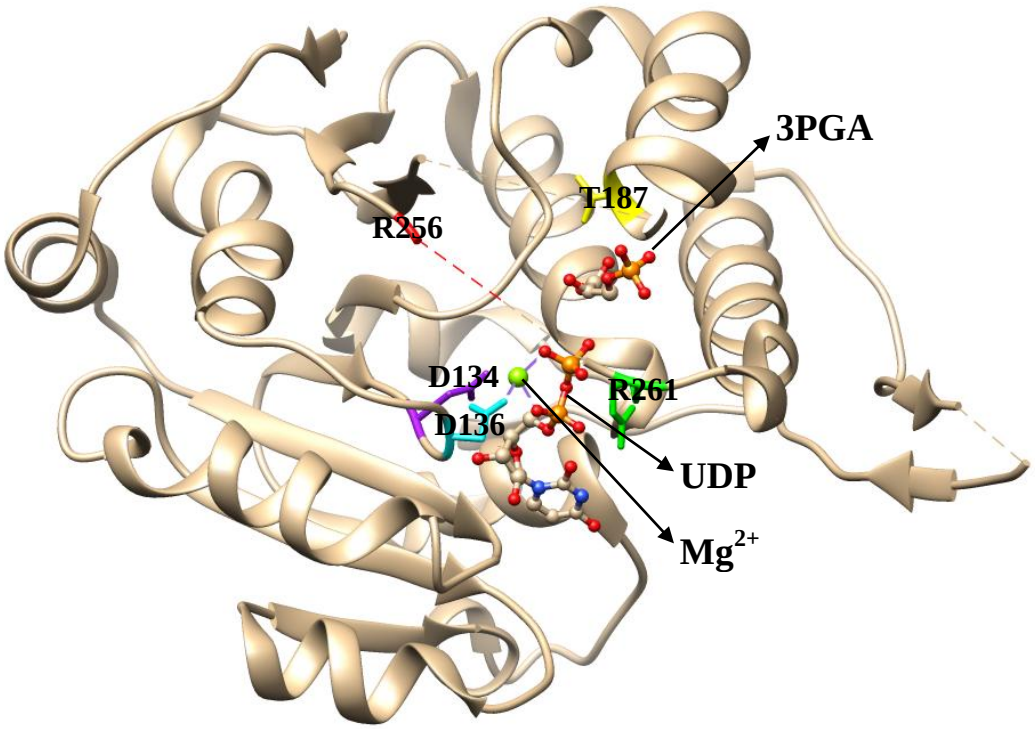


Figure 2.4. Substituted residues of *Mt*GpgS. Selected residues for substitution were labeled in *Mt*GpgS crystal structure with UDP, Mg^{2+}, and 3PGA. Made with Chimera. PDB: 3e25.

Table 2.4. Kinetic parameters of wild type *Mt*GpgS and variants. ND means not detectable.

Enzyme	k_{cat} (min^{-1})	K_{UDPG} (mM)	K_{3PGA} (mM)	k_{cat} / K_{UDPG} ($min^{-1}mM^{-1}$)	k_{cat} / K_{3PGA} ($min^{-1}mM^{-1}$)
Wild Type	50 ± 2	0.80 ± 0.06	0.20 ± 0.03	70 ± 5	215 ± 25
D134A	ND	ND	ND	ND	ND
D136A	ND	ND	ND	ND	ND
T187A	0.70 ± 0.03	0.48 ± 0.07	0.74 ± 0.09	1.43 ± 0.16	0.85 ± 0.09
R256A	ND	ND	ND	0.11 ± 0.01	0.09 ± 0.01
R261A	23 ± 1	1.5 ± 0.1	1.1 ± 0.1	14.9 ± 0.8	18.5 ± 1.3

2.3.2 Peptide Map of *Mt*GpgS

Tandem MS/MS technology and Peaks Client software were used to generate the peptide map of *Mt*GpgS (Figure 2.5). By using the elution method described in Table 2.2, about 95% of peptides are detected. In order to minimize back-exchange of amide hydrogens, the gradient time was reduced to 12 min, which is listed in Table 2.3. The reduction of gradient time results in loss of some peptides. As shown in Figure 2.5, about 83.6% of peptides are determined by tandem MS/MS.

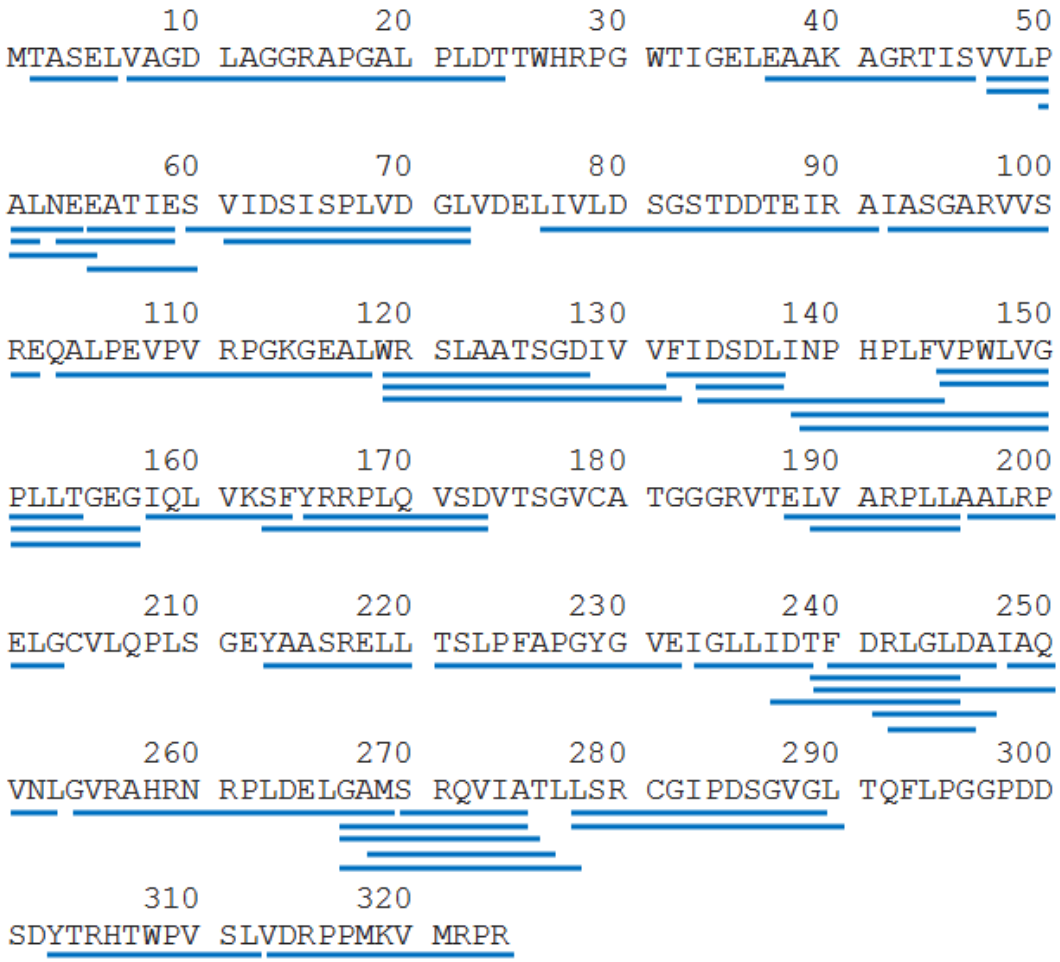


Figure 2.5. Peptide map of *Mt*GpgS. *Mt*GpgS was digested by pepsin and generated peptides were subject to collision induced dissociation (CID). MS/MS data was analyzed using software Peaks Client.

2.3.3 Solvent Accessibility Changes on Substrate Binding

The deuteration level at 15s can be used to estimate the relative D_2O accessibility of a protein's backbone amide hydrogens. In order to observe the difference in solvent accessibility between free protein and protein-ligand complex, the solvent accessibility of free *Mt*GpgS was compared with *Mt*GpgS-UDPG and *Mt*GpgS-3PGA complex. Figure 2.6 (A) shows the percentage difference in solvent accessibility between free *Mt*GpgS and the *Mt*GpgS-UDPG complex. Most peptides have decreases in solvent accessibility on UDPG binding, except peptides 7-24 and 233-239, which show minor increases in solvent accessibility. Considerable protection by UDPG is observed for peptides 47-52, 50-54, 103-118, 133-144, and 164-173. They have percentage difference greater than 5% and are labeled on *Mt*GpgS crystal structure (Figure 2.6 B). These peptides show significant decreases in solvent accessibility, suggesting that some backbone amide hydrogens are buried due to the binding of UDPG. As shown in Figure 2.6, these peptides are surrounding the binding site of UDPG, confirming that these peptides provide the binding site for UDPG and may have interactions with the substrate in order to stabilize it.

The difference plot between free *Mt*GpgS and the *Mt*GpgS-3PGA complex shows that more peptides are protected by 3PGA compared to the *Mt*GpgS-UDPG complex (Figure 2.7 A).Peptides showing decreases in solvent accessibility larger than 5% are labeled in the crystal structure of *Mt*GpgS. As shown in Figure 2.7 (B), the peptides expected to compose the 3PGA binding site are protected, including peptides 189-195, 254-268, and 269-277. They show significant decreases in solvent accessibility on 3PGA binding. Moreover, the peptides surrounding the UDPG binding pocket also show protection by 3PGA, including peptides 50-54, 55-60, 77-91, 92-102, 133-144, and 164-173. So, the binding of 3PGA induces decreases in

solvent accessibility of both binding sites for UDPG and 3PGA. Therefore, it is proposed that 3PGA can bind to or can allosterically affect the UDPG binding site.

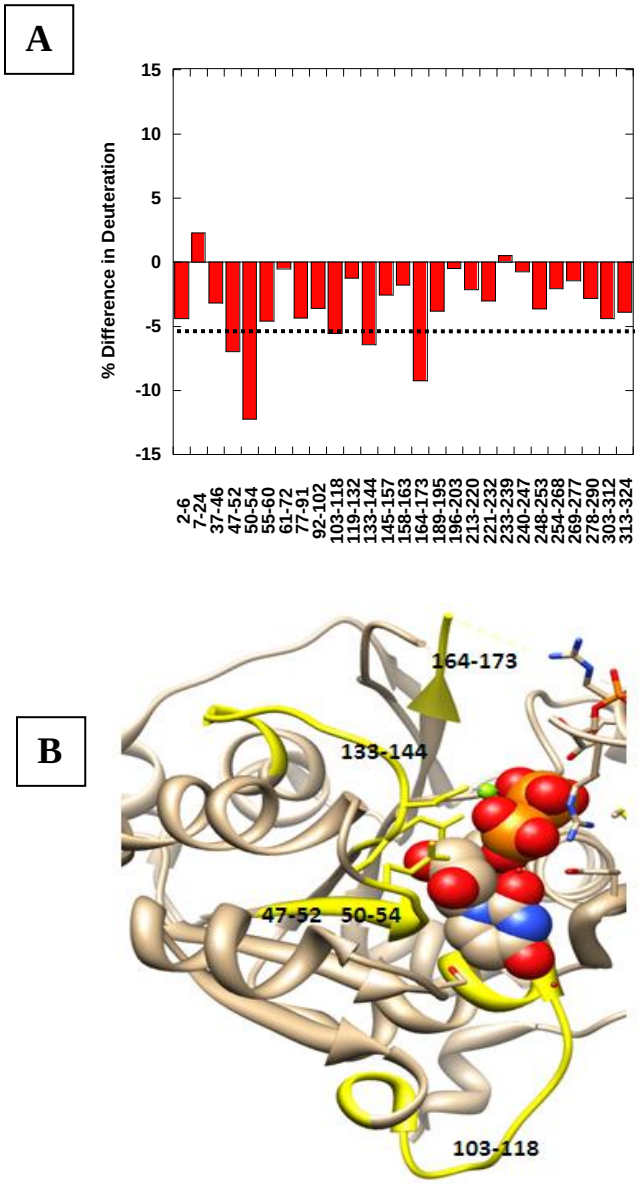


Figure 2.6. Solvent accessibility changes of MtGpgS on UDPG binding. (A) % Difference in solvent accessibility between free *Mt*GpgS and *Mt*GpgS-UDPG complex. It was calculated by subtracting percentage of deuterium incorporated of *Mt*GpgS-UDPG complex from that of free *Mt*GpgS. The positive and negative values of y axis correspond to increase and decrease in solvent accessibility, respectively. (B) Peptides with significant decrease in solvent accessibility were mapped in crystal structure of *Mt*GpgS, with UDP in sphere format. Made with Chimera. PDB: 3e25

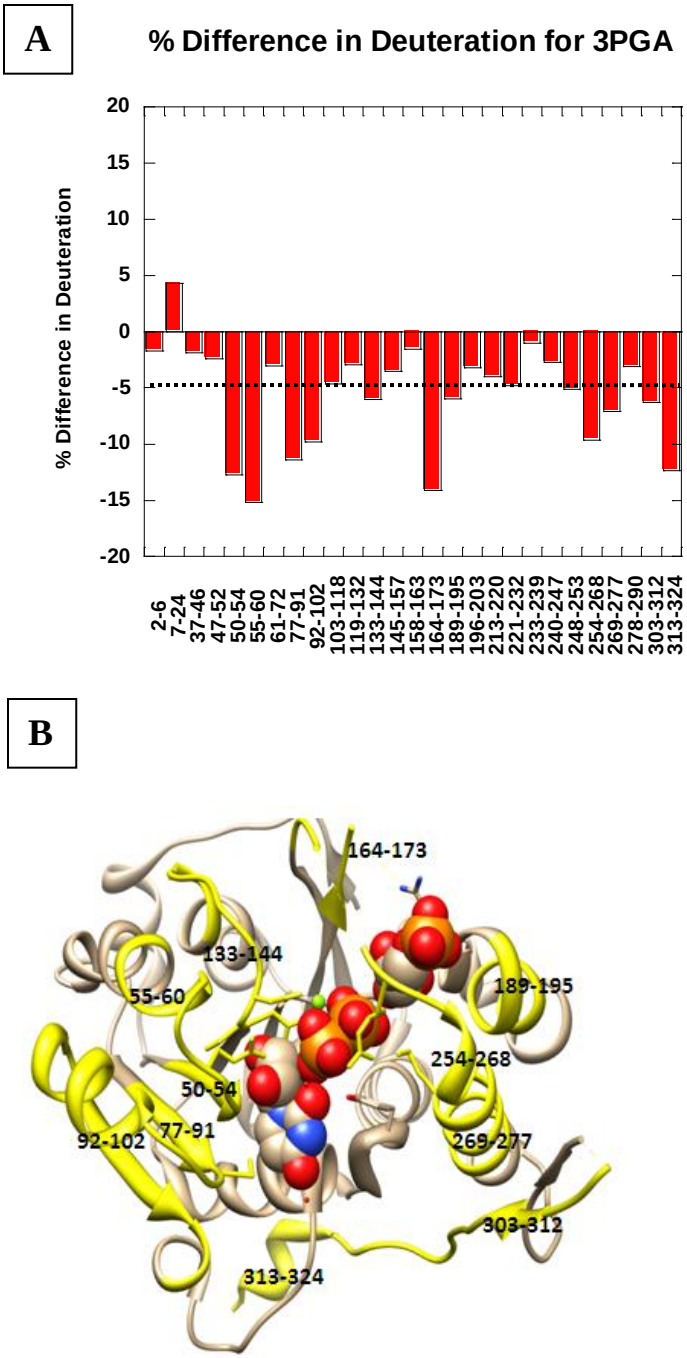


Figure 2.7. Solvent accessibility changes of MtGpgS on 3PGA binding (A) % Difference in solvent accessibility between free *Mt*GpgS and *Mt*GpgS-3PGA complex. (B) Peptides with significant decrease in solvent accessibility were mapped in crystal structure of *Mt*GpgS, with UDPand 3PGA in sphere format. Made with Chimera. PDB: 3e25

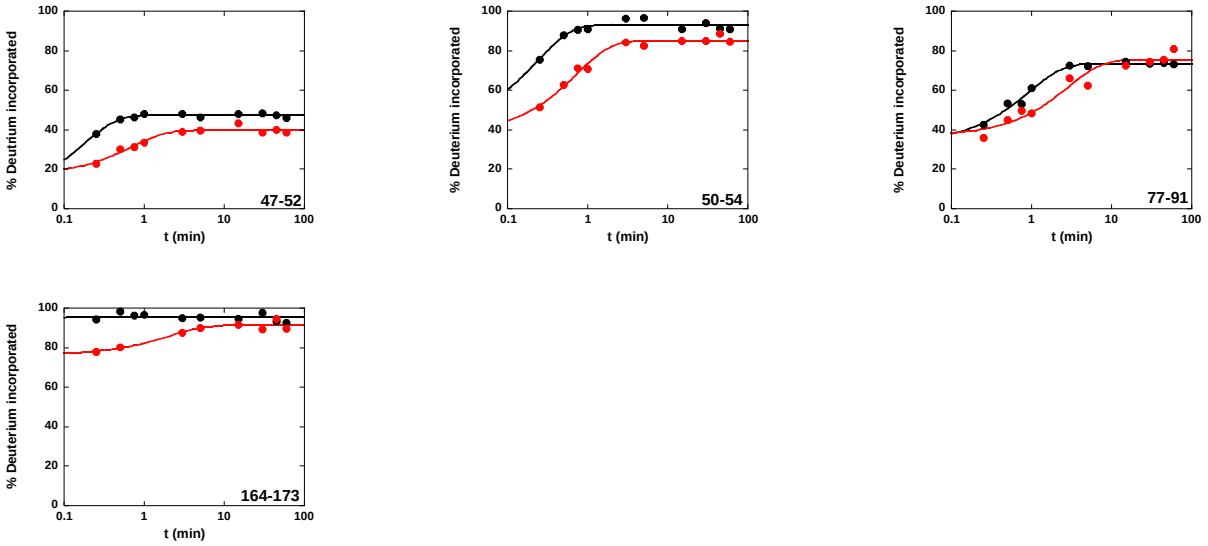


Figure 2.8. Deuterium incorporation curves of peptides from *Mt*GpgS on UDPG binding. Black curve represents percentage of deuterium incorporation of free *Mt*GpgS, and red curve represents that of *Mt*GpgS-UDPG complex. Made with KaleidaGraph.

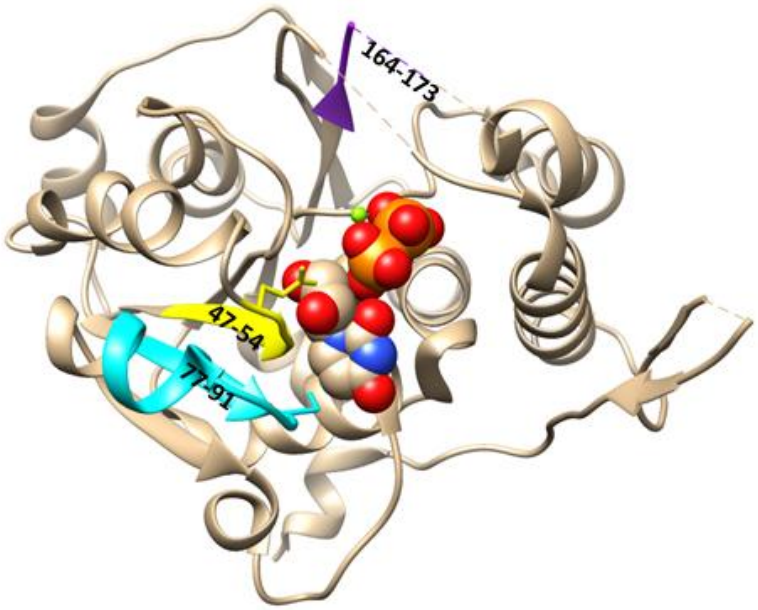


Figure 2.9. Dynamic Changes of peptides on UDPG binding. Peptides showing dynamic changes on UDPG binding were labeled on crystal structure of *Mt*GpgS dimer with UDP marked as sphere. Draw with Chimera. PDB: 3e25

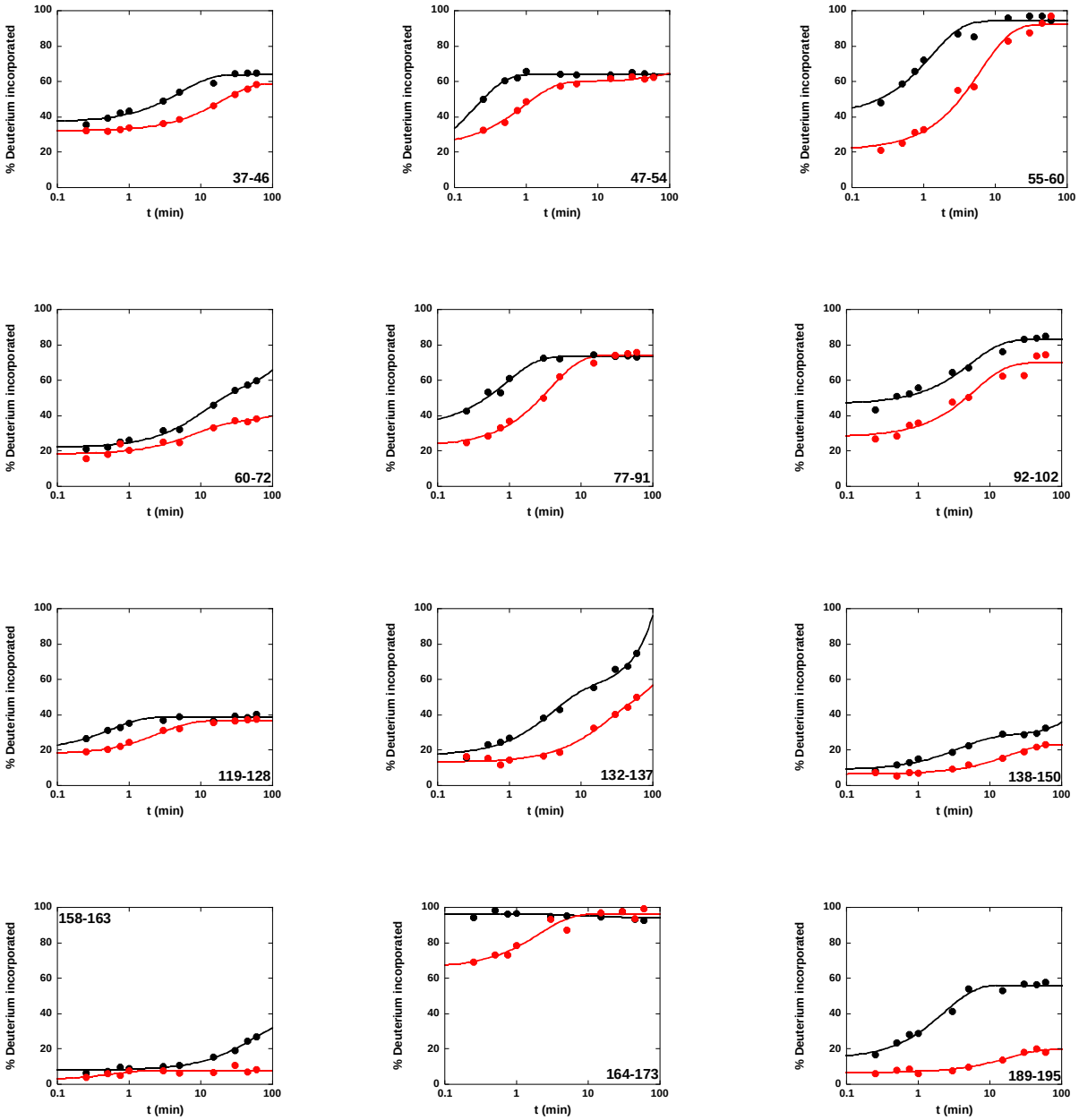
2.3.4 Dynamic Changes of *Mt*GpgS on UDPG Binding

The dynamic changes of *Mt*GpgS on substrate binding could be detected by comparing deuterium incorporation curves of free *Mt*GpgS with substrate-bound *Mt*GpgS. The solvent accessibility changes of *Mt*GpgS on UDPG binding show that the peptides surrounding the binding site of UDPG have decreases due to the binding of UDPG. So, the UDPG binding is expected to induce dynamic changes in these peptides. The HDX-MS data shows three peptides 47-52, 50-54, and 77-91 have dynamic changes on UDPG binding (Figure 2.8). These peptides are proposed to be involved in UDPG binding. The side chain of Glu54 is predicted to form a hydrogen bond with the hydroxyl group of the ribose (Figure 2.9). [10] This is supported by decreased dynamics of the region from residue 47 to 54. Peptides 47-52 and 50-54 show reduction in exchange rate by 4-fold and 3-fold, respectively. Another residue Ser81 is predicted to form a hydrogen bond with the uracil moiety, which is supported by 3-fold reduction in exchange rate for peptide 77-91. A flexible loop from residues 164 to 173, missing in the crystal structure, shows about 20% decrease in deuterium incorporation. It indicates that the loop loses some conformational flexibility due to the binding of UDPG. Overall, the binding of UDPG does not induce large conformational changes of *Mt*GpgS, suggesting that *Mt*GpgS may have a necessary conformation for UDPG binding.

2.3.5 Dynamic Changes of *Mt*GpgS on 3PGA Binding

While the binding of UDPG does not induce major conformational changes of *Mt*GpgS, the deuterium incorporation curves of peptides listed in Figure 2.10 reveal that *Mt*GpgS has significant dynamic changes on 3PGA binding. Peptides showing dynamic changes on 3PGA binding are labeled on the crystal structure of *Mt*GpgS dimer (Figure 2.11). This figure shows the binding of 3PGA induces dynamic changes of peptides located at both binding sites of

UDPG and 3PGA. So, the dynamic changes of *Mt*GpgS upon 3PGA binding are consistent with the conclusion drawn from changes in solvent accessibility that 3PGA can bind to or allosterically affect the UDPG binding site.



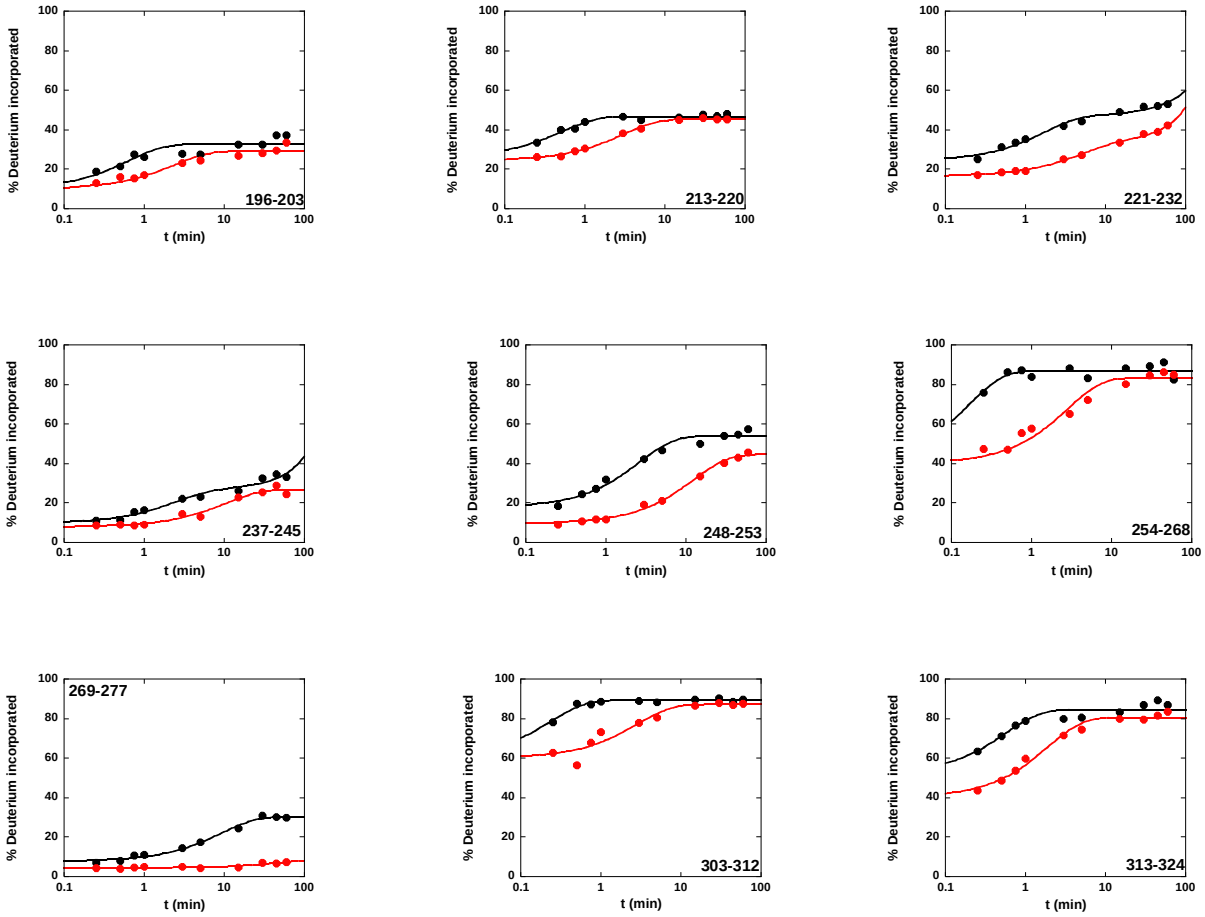


Figure 2.10. Deuterium incorporation curves of peptides from *Mt*GpgS. Black curve represents percentage of deuterium incorporation of free *Mt*GpgS, and red curve represents that of *Mt*GpgS-3PGA complex. Made with KaleidaGraph.

Peptides located at the *N*-terminal domain show decreasing dynamics on 3PGA binding. Some regions located at the *N*-terminal domain and proposed to contribute to the UDPG binding have changes in dynamics. Peptide 47-54 containing the residue Glu54 shows a decrease in exchange rate by 5-fold, supporting the formation of the hydrogen bond between the side chain of Glu54 and the hydroxyl group of the ribose (Figure 2.12). Peptide 77-91 shows a decrease in exchange rate by 4-fold, supporting the importance of the residue Ser81. The dynamic changes of these two regions are consistent with the dynamic changes of *Mt*GpgS on UDPG binding. Another region of UDPG binding site showing dynamic changes is located at peptide 132-137.

The DXD motif (residues 134-136) is predicted to be important for UDPG and Mg^{2+} binding, supported by 5-fold reduction in exchange rate of the peptide 132-137. The decreasing dynamics of these regions support the conclusion that 3PGA can bind to or allosterically affect the UDPG binding site. In addition to the UDPG binding site showing dynamic changes, some other regions from the *N*-terminal domain also have dynamic changes on 3PGA binding. Peptide 37-46 has a 4-fold reduction in exchange rate, suggesting it has decreasing dynamics on 3PGA binding. Peptides 55-60 and 60-72 compose an α-helix within the Rossmann-like fold, showing 4-fold decrease in exchange rate and 20% decrease in deuterium incorporation, respectively. Moreover, peptide 92-102 neighbor to the peptide 77-91 also shows 20% decrease in deuterium incorporation, and peptide 119-128 located at an α-helix within the Rossmann-like fold shows 5-fold reduction in exchange rate. The dynamic changes of these peptides support the conclusion of the decreasing dynamics of the *N*-terminal domain induced by 3PGA binding.

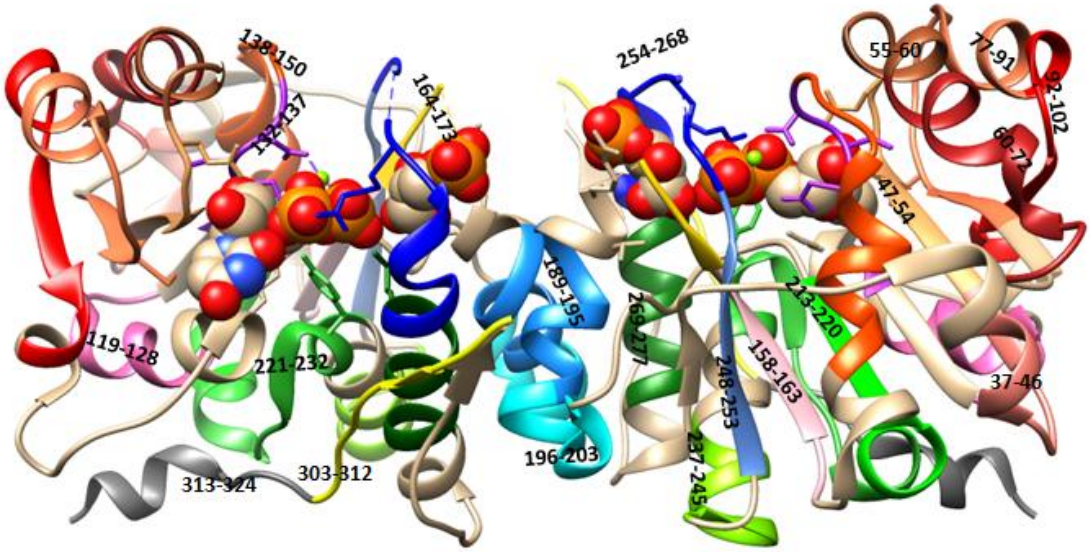


Figure 2.11. Dynamic Changes of peptides on 3PGA binding. Peptides showing dynamic changes on 3PGA binding were labeled on crystal structure of *Mt*GpgS dimer with UDP and 3PGA which were acted as sphere. Made with Chimera. PDB: 3e25

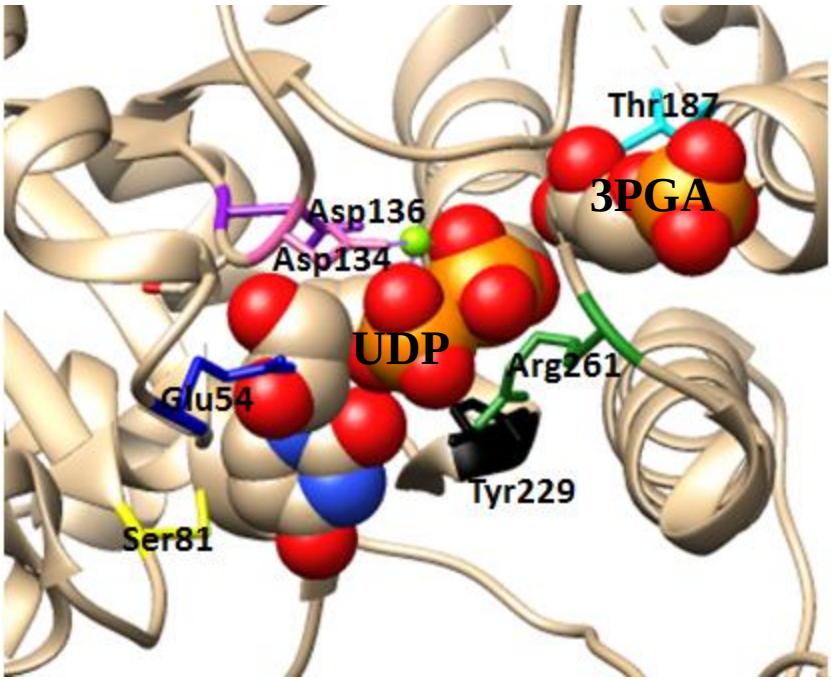


Figure 2.12. Residues involved in substrates binding. Residues contributing to substrates binding were labeled in structure of *Mt*GpgS. Made with Chimera. PDB: 3e25

Dynamic changes of the *C*-terminal domain induced by 3PGA binding are also observed. As shown in Figure 2.11, the interface between two *Mt*GpgS monomers has decreasing dynamics, supported by the dynamic changes located at the interface. Peptides 189-195 and 196-203 show decreases in exchange rate by 9-fold and 3-fold, respectively. The two residues Val186 and Thr187 are proposed to form interactions with 3PGA (Figure 2.12). The loop containing these two residues is missing in HDX-MS experiments. But, decreasing dynamics of the peptide 189-195 may provide an evidence for the dynamic change of this missing loop. Another two peptides 237-245 and 269-277 also show significant decreases in exchange rate, suggesting that the dimer interface has decreasing dynamics on 3PGA binding. In addition to the dynamic changes of the dimer interface, two regions proposed to contribute to UDPG binding are identified to have dynamic changes. Peptide 221-232 containing Tyr229 has 3-fold decrease in exchange rate, supporting the proposal that the side chain of Tyr229 may have an interaction

with the phosphate group of UDPG. The flexible loop 253-262 is identified to have two conformations on UDPG binding by obsersing crystal structures of *Mt*GpgS. [23] It is supported by the dynamic change of peptide 254-268. So, the binding of 3PGA induces the loss of conformational flexibility of this loop. Moreover, the decreasing dynamics of the peptide 254-268 supports the possibility that the side chain of Arg261 may form an interaction with UDPG. Besides the dynamic changes happening in the UDPG binding site and dimer interface, more peptides located at the *C*-terminal domain show decreasing dynamics. Peptides 213-220 and 248-253, composed of β-sheets from the *C*-terminal domain, show decreases in exchange rate by 6-fold and 5-fold, respectively. Furthermore, the end of the *C*-terminal domain also shows decreasing dynamics, supported by dynamic changes of peptides 303-312 and 313-324. The flexible loops were reported to be important for the *Mt*GpgS substrate binding and activity, supported by the HDX-MS data. [23] Two loops are detected to have dynamic changes, including the loop 253-262 described above and the loop 165-184. The dynamic change of the loop 165-184 is supported by 30% decrease in deuterium incorporation of peptide 164-173. Overall, the dynamic changes of *Mt*GpgS on 3PGA binding suggest that the binding of 3PGA can induce decreasing dynamics of substrate binding sites, dimer interface, as well as the surface of both domains, supporting the conclusion that 3PGA can induce the dynamic changes of the UDPG binding pocket allosterically or by binding to it.

2.4 Discussion

2.4.1 Roles of Active Site Residues in *Mt*GpgS

The crystal structure of *Mt*GpgS has been determined, and two conserved architectures are proposed to play a role in substrate binding. [10] The DXD motif is conserved in GT-A enzymes and significant for substrate binding. So, the residues D134 and D136 contained in the

DXD motif of *Mt*GpgS are predicted to be important for UDPG and Mg^{2+} binding. Another conserved architecture in GT-A enzyme is a flexible loop (residues 253-262 in *Mt*GpgS). This loop has been implicated to contribute to catalytic activity and substrate specificity of several GT-A enzymes including *Mt*GpgS. So the two residues R256 and R261 may be essential to the catalytic activity. Moreover, the residue D178 is proposed to be involved in 3PGA recognition. Based on their importance, the variants of these five residues were constructed in order to study their specific roles in substrate binding and enzymatic catalysis.

The DXD motif is a determining characteristic in GT-A glycosyltransferase. [24-25] *Mt*GpgS is identified to have conserved $Asp^{134}Ser^{135}Asp^{136}$ motif, in which both aspartate residues are involved in UDP and Mg^{2+} binding. [10] The D134 residue forms a hydrogen-bonded network with K114 and E232, and these three residues are predicted to have interactions with UDPG. The kinetic assay of the D134A variant shows that its kinetic parameters cannot be detected, indicating that residue D134 is essential to the activity of *Mt*GpgS. The Frantom lab has reported that the substitution of K114 with alanine results in insoluble enzyme, indicating this residue is important for integrity of *Mt*GpgS. The E232A variant results in a fully functioning enzyme, suggesting this residue is not critical to the *Mt*GpgS activity. [26] So within this hydrogen-bonded network, only the residue D134 is indispensable to the function of *Mt*GpgS. The second aspartate was reported to coordinate with Mg^{2+} which bridges the oxygens of the UDP α and β phosphates. [10] The undetectable activity of D136A variant is consistent with the role of D136 playing in ion binding, which is required by the activity of *Mt*GpgS.

3PGA is predicted to form hydrogen bonds with the main chain nitrogen of V186 and side chain of T187. [10] The T187A variant shows about 100-fold decrease in reaction velocity, suggesting the important role of T187 in *Mt*GpgS catalysis. However, the K_m values for UDPG

and 3PGA have minor changes relative to wild type *Mt*GpgS, indicating that both substrates UDPG and 3PGA can bind to *Mt*GpgS as normal even though T187 was substituted by alanine. The substitution of T187 is proposed to lead to the reorientation of 3PGA which has an effect on reaction velocity.

The loop 253-262 was reported to be the only one significant structural change of *Mt*GpgS upon substrate binding. [10] Residues R256 and R261 were selected for site-directed mutagenesis since the side chain of Arg256 participates in Mg^{2+} coordination and R261 makes electrostatic interaction with the α-phosphate of the UDP. [23] The substitution of R256 results in undetectable k_{cat} and K_m values of the R256A variant. So the k_{cat}/K_m values for UDPG and 3PGA were calculated and shown to be reduced by 636-fold and 2388-fold, respectively. It indicates that R256 residue is essential to the binding of UDPG and 3PGA. It also confirms the significance of the flexible loop 253-262 which has structural changes when substrates bind to *Mt*GpgS. Compared with R256A variant, R261A variant shows smaller difference in k_{cat} and K_m values. It suggests that residue R261 is dispensable for catalysis of *Mt*GpgS.

2.4.2 Insight into Dynamic Changes of *Mt*GpgS

As discussed in 2.3.4 and 2.3.5, *Mt*GpgS may have a necessary conformation for UDPG binding, but it goes through large conformational changes upon 3PGA binding. The binding of 3PGA induces dynamic changes of both UDPG and 3PGA binding sites. As shown in Figure 2.10, the peptides 47-54, 77-91, 132-137, 221-232, and 254-268 display significant decreases in exchange rate. These peptides contain important residues that may form interactions with UDPG and Mg^{2+}, including Glu54, Ser81, Asp134, Asp136, Tyr229, and Arg256 (Figure 2.12). The decreasing dynamics of these peptides confirms the significance of these residues for substrate binding, which is also consistent with the kinetics of *Mt*GpgS variants. The undetected activity of

D134A and D136A variants supports the role of DXD motif in *Mt*GpgS activity, which is also confirmed by the decrease in exchange rate of the peptide 132-137. The reduced dynamics of the peptide 254-268 is also consistent with the low activity of the R256A variant, indicating the essential role of R256 in *Mt*GpgS catalysis. Moreover, the peptides 189-195, 196-203, 237-245, and 269-277 also show decreases in dynamics, indicating the dynamics of dimer interface is reduced upon 3PGA binding. It suggests that the dynamic changes of dimer interface may be necessary for *Mt*GpgS activity.

The catalytic site of *Mt*GpgS was proposed to be preformed before donor and acceptor binding, since very few differences were found between the conformations of the free and complexed *Mt*GpgS in the crystal structure. [23] However, the HDX-MS data disagrees with the proposal. The free *Mt*GpgS indeed provides a necessary confirmation for UDPG binding since the UDPG binding does not induce any significant changes in *Mt*GpgS. While *Mt*GpgS displays significant changes in dynamics on 3PGA binding. The peptides involved in substrate binding sites and located at dimer interface show decreases in dynamics. These changes in dynamics are proposed to be required by 3PGA binding and *Mt*GpgS activity.

2.4.3 Substrate Binding Mechanism of *Mt*GpgS

Kinetic and structural studies have reveled that most GT- A glycosyltransferases follow an ordered mechanism that nucleotide sugar donor binds first prior to the binding of the acceptor. [23] Isothermal titration calorimetry (ITC) was applied to study substrate binding reactions of *Mt*GpgS. The ITC data demonstrated that binding of UDPG and 3PGA are sequential. [23] However, the HDX-MS data disagrees with the ITC data. The solvent accessibility of peptides surrounding the 3PGA binding site decreases due to the binding of 3PGA, indicating that the residues involved in the 3PGA binding pocket are protected by 3PGA. Moreover, the binding of

3PGA induces significant dynamic changes of *Mt*GpgS in the absence of UDPG. These evidences support that 3PGA can bind to *Mt*GpgS independently.

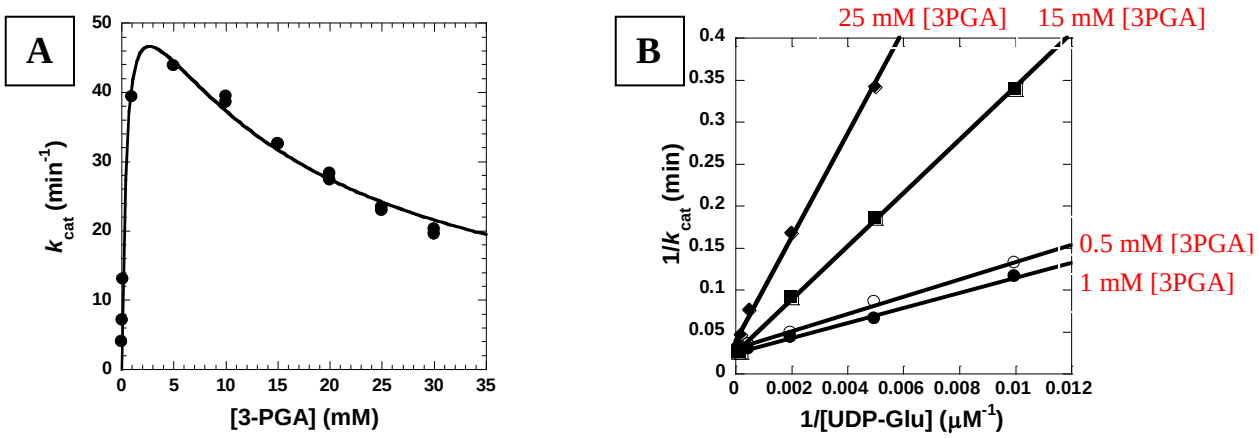


Figure 2.13. Substrate inhibition of *Mt*GpgS by 3PGA versus UDPG. (A) x axis is the concentration of 3PGA, and y axis is the velocity of reaction. (B) x axis is 1/[UDPG] and y axis is $1/k_{cat}$. In each reaction, the concentration of 3PGA is a constant marked in red. [26] Made with KaleidaGraph.

The decrease in solvent accessibility of *Mt*GpgS due to the binding of 3PGA suggests that the peptides involved in both UDPG and 3PGA binding pockets are protected by 3PGA. It supports that 3PGA can bind to or allosterically affect the UDPG binding site. Furthermore, the deuterium incorporation curves of *Mt*GpgS-3PGA complex relative to free *Mt*GpgS display that the peptides surrounding the UDPG binding site have decreasing dynamics on 3PGA binding. The decreases in solvent accessibility and dynamics points out a possibility that 3PGA may compete to the UDPG binding site with UDPG. This proposal is confirmed by the substrate inhibition mechanism of *Mt*GpgS by 3PGA versus UDPG (Figure 2.13). [26] The kinetic plot shown in Figure 2.13 (A) indicates that high concentration of 3PGA inhibits the activity of *Mt*GpgS since the velocity of reaction decreases due to the increase of 3PGA concentration. The

plot shown in Figure 2.13 (B) suggests that 3PGA is a competitive inhibitor. The four curves intersect at one point when the concentration of UDPG is infinitely high, indicating 3PGA cannot inhibit the activity of *Mt*GpgS when UDPG is saturated. Moreover, the reaction velocity increases when the concentration of 3PGA increases from 0.5mM to 1mM, while the velocity decreases when the concentration of 3PGA continues to increase from 15mM to 25mM. Both two plots showed that 3PGA is a competitive inhibitor versus UDPG, which is also supported by the HDX-MS data.

2.4.4 Role of Loop Flexibility in Substrate Recognition and Binding

Loop mobility around the active site upon ligand binding has been reported for some members of GT families. [27] The two loops of inverting GT β-(1, 4)-galactosyltransferase were revealed to have conformational changes by comparing substrate bound and unbound crystal structures. [28] The active site of retaining GT α-(1, 4)-galactosyltransferase LgtC is almost buried by two loops that play an important role in organizing substrates before reactions. Essential dynamics analysis (EDA) revealed a limited movement in one loop and a larger conformational change in another loop. [29] In GT-A superfamily, several loops of N-acetyl-α-galactosaminyl-transferase were identified to have conformational changes. Especially, the loop (residues 361-377) that is folded back over the active site might be involved in positioning substrates into active site since it undergoes a large movement on substrate binding. [27] So it is proposed that loop mobility around the active site may be a general feature of GT-A superfamily, which required further testing of other members of this superfamily. [27] The HDX-MS data of *Mt*GpgS displays that the loops indeed play a significant role in substrate recognition and binding, supporting the proposal. The loops 106-112, 133-142, and 165-184 are identified to contribute to the recognition of substrate UDPG, since the peptides involving these three loops

show significant decreases in solvent accessibility on nucleotide binding (Figure 2.6 A). The binding site of UDPG has been preformed before binding, thus very few dynamic changes of *Mt*GpgS are observed on UDPG binding. Furthermore, the loop mobility plays a key role in recognition and binding of 3PGA. Five loops of *Mt*GpgS show significant decreases in solvent accessibility and dynamics on 3PGA binding (Figure 2.7A and Figure 2.8). Two loops (residues 165-184 and 253-262) are located around the binding site of 3PGA. They might be responsible for recognizing and positioning 3PGA in the active site. The loops 106-112 and 133-142 are involved in nucleotide binding site. They show decreases in solvent accessibility and dynamics because 3PGA can bind to nucleotide binding site in the absence of UDP-GlcNAc. The loop 303-316 is located at the end of *C*-terminal domain and far away from the active site. Therefore, the dynamic changes of these loops confirm that the loop mobility is necessary for substrate binding and catalysis of *Mt*GpgS.

2.5 References

[1] Smith, L. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev.* 2003;16:463-496.

[2] Dye, C. Global epidemiology of tuberculosis. *Lancet.* 2006;367:938-940.

[3] Kalscheuer, R., Syson, K., Veeraraghavan, U., weinrick, B., Biermann, K. E., Liu, Z., Sacchettini, J. C., Besra, G., Bornemann, S., and Jacobs Jr, W. R. Self-poisoning of *mycobacterium tuberculosis* by targeting GlgE in an α-glucan pathway. *Nat Chem Biol.* 2010;6:376-384.

[4] Jackson, M., Brennan, P. Polymethylated polysaccharides from Mycobacterium species revisited. *J Biol Chem.* 2009;284:1949-1953.

[5] Maitra, S.K., Ballou, C. E. Heterogeneity and refined structures of 3-O-methyl-D-mannose polysaccharides from *Mycobacterium smegmaits*. *J Biol Chem.* 1977;252:2459-2469.

[6] Lee, Y. C., and Ballou, C. E. 6-*O*-methyl-D-glucose from Mycobacteria. *J Biol Chem*. 1977;239:PC3602-PC3603.

[7] Maitra, S. K., and Ballou, C. E. Heterogeneity and refined structures of 3-*O*-methyl-D-mannose polysaccharides from *Mycobacterium smegmatis*. *J Biol Chem*. 1977;252:2459-2469.

[8] Stadthagen, G., Sambou, T., Guerin, M., Barilone, N., Boudou, F., Kordulakova, J., Charles, P., Alzari, P. M., Lemassu, A., Daffe, M., Puzo, G., Gicquel, B., Riviere, M., and Jackson, M. Genetic basis for the biosynthesis of methylglucose lipopolysaccharides in *Mycobacterium tuberculosis*. *J Biol Chem.* 2007;282:27270-27276.

[9] Empadinhas, N., Albuquerque, L., Mendes, V., Macedo-Ribeiro, S., and Costa, M. S. Identification of the mycobacterial glucose-3-phosphoglycerate synthase. *FEMS Microbiol Lett.* 2008;280:195-202.

[10] Pereira, P. J. B., Empadinhas, N., Albuquerque, L., Sa-Moura, B., Costa, M. S., and Macedo-Ribeiro, S. *Mycobacterium tuberculosis* glucosyl-3-phosphoglycerate synthase: structure of a key enzyme in methylglucose lipopolysaccharides biosynthesis. *Plos One.* 2008;3:e3748.

[11] Tuffal, G., Albigot, R., Riviere, M., and Puzo, G. Newly found 2-N-acetyl-2,6-dideoxy-β-glucopyranose containing methyl glucose polysaccharides in M.bovis BCG: revised structure of the mycobacterial methyl glucose lipopolysaccharides. *Glycobiology.* 1998;8:675-684.

[12] Smith, W. L., and Ballou, C. E. The 6-*O*-methylglucose-containing lipopolysaccharides of *Mycobacterium Phlei*. Locations of the neutral and acidic acyl groups. *J Biol Chem.* 1973;248:7118-7125.

[13] Saier, M. H. Jr., and Ballou, C. E. The 6-*O*-methylglucose-containing lipopolysaccharides of *Mycobacterium Phlei*. Identification of D-glyceric acid and 3-*O*-methyl-D-glucose in the polysaccharide. *J Biol Chem.* 1968;243:992-1005.

[14]Klahn, S., Steglich, C., Hess, W. R., and Hagemann, M. Glucosylglycerate: a secondary compatible solute common to marine cyanobacteria from nitrogen-poor environments. *Environ Microbiol.* 2010;12:83-94.

[15] Goude, R., Renaud, S., Bonnassie, S., Bernard, T., and Blanco, C. Glutamine, glutamate, and α-glucosylglycerate are the major osmotic solutes accumulated by *Erwinia chrysanthemi* strain 3937. *Appl Environ Microbiol.* 2004;70:6535–6541.

[16] Robertson, D. E., Lai, M. C., Gunsalus, R. P., and Roberts, M. F. Composition, Variation, and Dynamics of Major Osmotic Solutes in *Methanohalophilus* Strain FDF1. *Appl Environ Microbiol.* 1992;58*:*2438–2443.

[17] Kamisango, K., Dell, A., and Ballou, C. E. Biosynthesis of the mycobacterial O-methylglucose lipopolysaccharide. Characterization of putative intermediates in the initiation, elongation, and termination reactions. *J Biol Chem.* 1987;262:4580-4586.

[18] Kaur, D., Pham, H., Larrouy-Maumus, G., Riviere, M., Vissa, V., Guerin, M. E., Puzo, G., Brennan, P. J., and Jackson, M. Initiation of methylglucose lipopolysaccharide biosynthesis in mycobacteria. *PLos ONE.* 2009;4:e5447.

[19] Sassetti, C. M., Boyd, D. H., and Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol.* 2003;48:77-84.

[20] Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol.* 2003;328:307–317.

[21] Costa, J., Empadinhas, N., Gonclaves, L., Lamosa, P., Santos, H., and Costa, M. S. Characterization of the biosynthetic pathway of glucosylglycerate in the archaeon *Methanococcoides burtonii*. *J Bacteriol*. 2006;188:1022–1030.

[22] Empadinhas, N., Albuquerque, L., Mendes, V., Macedo-Ribeiro, S., and Costa, M. S. Identification of the mycobacterial glucosyl-3-phosphoglycerate synthase. *FEMS Microbiol Lett.* 2008;280: 195–202.

[23] Urresti, S., Albesa-Jove, D., Schaeffer, F., Pham, H. T., Kaur, D., Gest, P., van der Woerd, M. J., Carreras-Gonzalez, A., Lopez-Fernandez, S., Alzari, P. M., Brennan, P. J., Jackson, M., and Guerin, M. E . Mechanistic insights into the retaining glucosyl-3-phosphoglycerate synthase from mycobacteria. *J Biol Chem.* 2012;287:24649-24661.

[24] Wiggins, C. A. R., and Munro, S. Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc Natl Acad Sci U S A*. 1998;95:7945-7950.

[25] Breton, C., Bettler, E., Joziasse, D. H., Geremia, R. A., and Imberty, A. Sequence/function relationships of prokaryotic and eukaryotic galactosyltransferases. *J Biochem*. 1998;123:1000-1009.

[26] Kumar, G., Guan, S. Q., and Frantom, P. A. Biochemical characterization of the retaining glycosyltransferase glucosyl-3-phosphoglycerate synthase from Mycobacterium tuberculosis. *Arch Biochem Biophys*. 2014;564:120-127.

[27] Milac, A. L., Buchete, N. V., Fritz, T. A., Hummer, G., and Tabak, L. A. Substrate-induced conformational changes and dynamics of UDP-*N*-acetylgalactosamine: polypeptide *N*-acetylgalactosaminyltransferase-2. *J Mol Biol*. 2007;373:439-451.

[28] Gunasekaran, K., Ma, B., and Nussinov, R. Triggering loops and enzyme function: identification of loops that trigger and modulate movements. *J Mol Biol*. 2003;332:143-159.

[29] Snajdrova, L., Kulhanek, P., Imberty, A., and Koca, J. Molecular dynamics simulations of glycosyltransferase LgtC. *Carbohydr Res.* 2004;339:995-1006.

CHAPTER 3

INSIGHTS INTO CONFORMATIONAL CHANGES OF RETAINING GLYCOSYLTRANSFERASE MSHA FROM *CORYNEBACTERIUM GLUTAMICUM*

3.1 Introduction

3.1.1 Significance of Mycothiol (MSH)

In most organisms, proper cellular function requires maintenance of appropriate reducing environment in cells. It is usually achieved by synthesis and cellular balance of low-molecular-weight thiols. [1] Glutathione is the major thiol in Gram-negative bacteria and eukaryotes, while in the *Actinomycetales*, mycothiol (MSH) is the predominant thiol. [2] MSH is composed of a cysteine residue with an acetylated amino group linked to glucosamine, which is then linked to inositol (Illustration 3.1). *Actinomycetales* deficient in MSH biosynthesis are identified to be more susceptible to oxidants and antimicrobial agents, making MSH important for the existence and growth of *Actinomycetales.* [3]

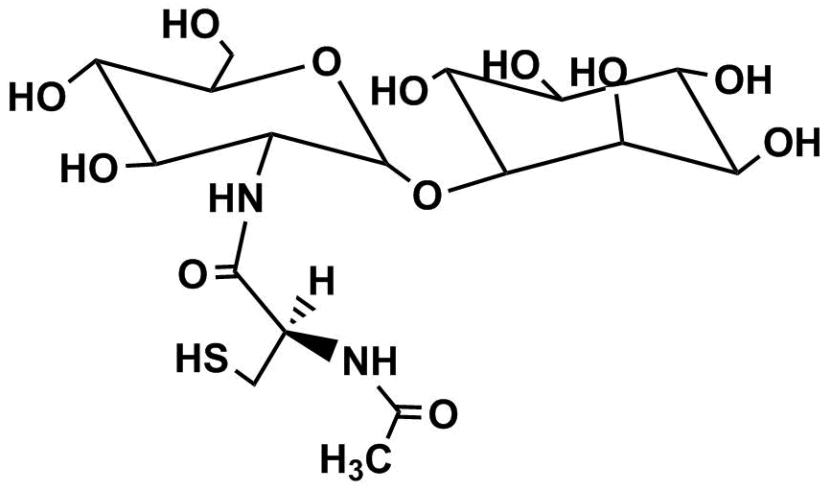


Illustration 3.1. Structure of MSH. Made with ChemDraw Professional 15.0

The *Actinomycetales* family has high diversity in function and is important to humans. On the positive side, *Corynebacterium* species are often utilized in industrial large-scale production of amino acids. [4] On the negative side, some members of *Actinomycetales* are the causative agents of various diseases, such as *Mycobacterium tuberculosis* that accounts for over one million deaths annually. [5] MSH is the primary thiol found in mycobacteria and the important role MSH plays in mycobacteria suggests that the MSH metabolism may provide potential targets for new drugs against *M. tuberculosis*. [6] This is supported by studies of *Mycobacterium smegmatis* mutants blocked in MSH biosynthesis. These mutants were shown to have enhanced sensitivity to rifampin and other antibiotics. [3, 7]

3.1.2 Biosynthetic Pathway of MSH

The biosynthetic pathway of MSH contains at least four enzymes: MshA, MshB, MshC, and MshD (Illustration 3.2). [8] MshA catalyzes the initial step of MSH biosynthesis, transferring *N*-acetylglucosamine from UDP-*N*-acetylglucosamine (UDP-GlcNAc) to 1-L-myo-inositol 1-phosphate (11P) and producing 3-phospho-1-D-myo-inosityl-2-acet-amido-2-deoxy-α-D-glucopyranoside (GlcNAc-Ins-P). [9-10] GlcNAc-Ins-P is dephosphorylated by an unknown phosphatase, producing GlcNAc-Ins. [9] Then GlcNAc-Ins is deacetylated by MshB and subsequently cysteinylated by MshC, producing Cys-GlcN-Ins. [11-14] Finally, MSH is generated by acetylation of Cys-GlcN-Ins. [15] Gene disruption studies of these four enzymes were carried out in order to investigate their significance. The mutants blocked in MSH biosynthesis were reported to be not viable in *M. tuberculosis* (Erdman strain). [16] Of the four genes, *mshA* and *mshC*, encoding MshA and MshC respectively, were identified to be vital to the production of MSH and viability of bacteria. [16-17] However, the interruption of genes *mshB* and *mshD*, encoding enzymes MshB and MshD, respectively, could be complemented by either

the product of interrupted synthesis or a promiscuous cellular activity. [18-19] Therefore, the enzymes MshA and MshC become important potential drug targets for the treatment of tuberculosis and infections by other *Actinomycetales*.

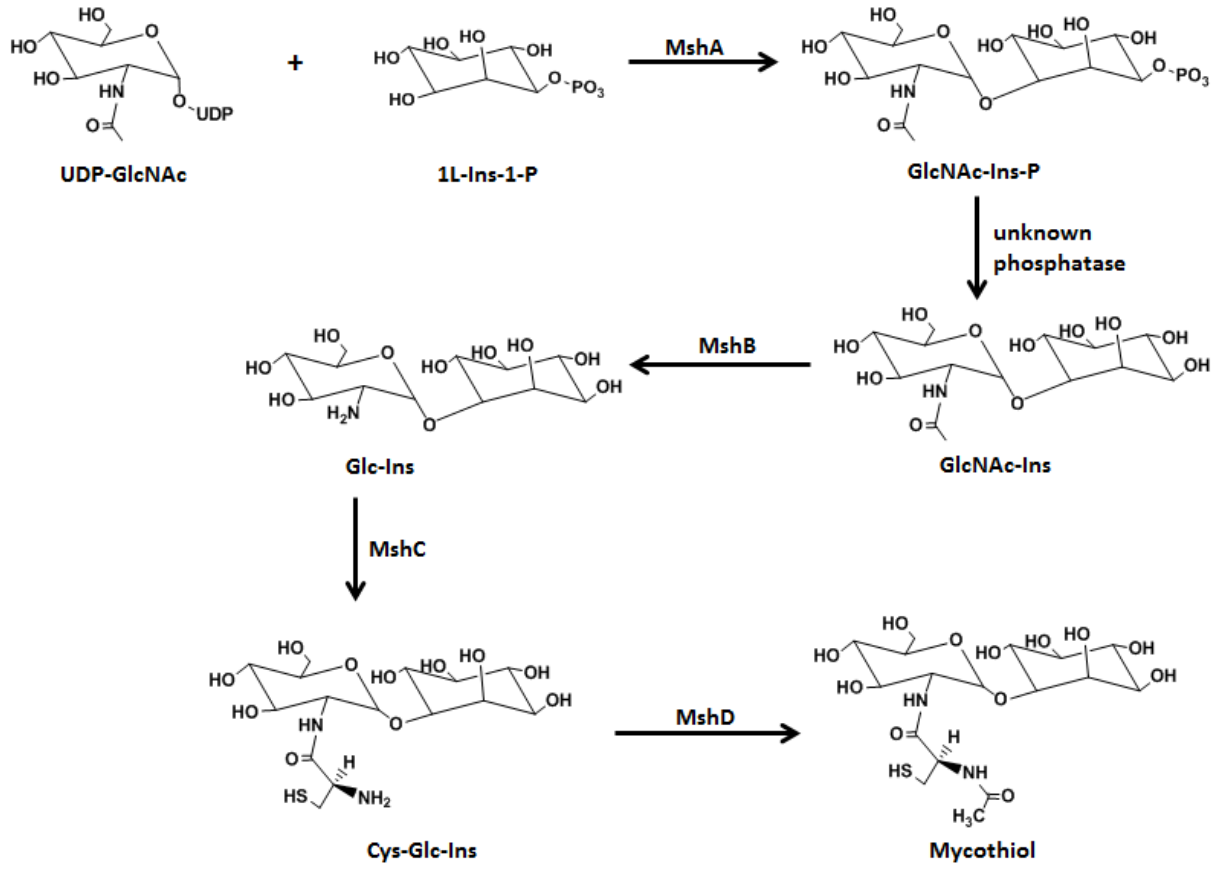


Illustration 3.2. Biosynthesis pathway of MSH. Made with ChemDraw Professional 15.0

3.1.3 Crystal Structure of *Cg*MshA

Based on sequence similarity, MshA is grouped into the GT-4 family of glycosyltransferases. As discussed in Chapter 1, GTs can be termed retaining and inverting based on the configuration of the anomeric centre of sugar donor. MshA is identified to be the retaining glycosyltransferase. Based on MshA structure, it is identified to have the GT-B fold. [20] As shown in Figure 3.1, MshA from *C. glutamicum* (*Cg*MshA) monomer contains two domains.

The *N*-terminal domain (residues 1-196 and 391-418) contains eight β-sheets bounded by six α-helices. The *C*-terminal domain (residues 197- 390) contains six β-sheets bounded by eight α-helices. The β-sheets of *C*-terminal domain are parallel, whereas β-sheets of *N*-terminal domain are a mixture of parallel and anti-parallel strands. The two domains of *Cg*MshA are connected by the hinge residues 193-208. As shown in *Cg*MshA structure, both domains contain Rossmann-like folds. Gel filtration chromatography data showed that *Cg*MshA is a dimer in solution. [20] This is supported by crystal structure of *Cg*MshA dimer, which shows the dimer interface is completely composed of residues from the *N*-terminal domains of two *Cg*MshA monomers.

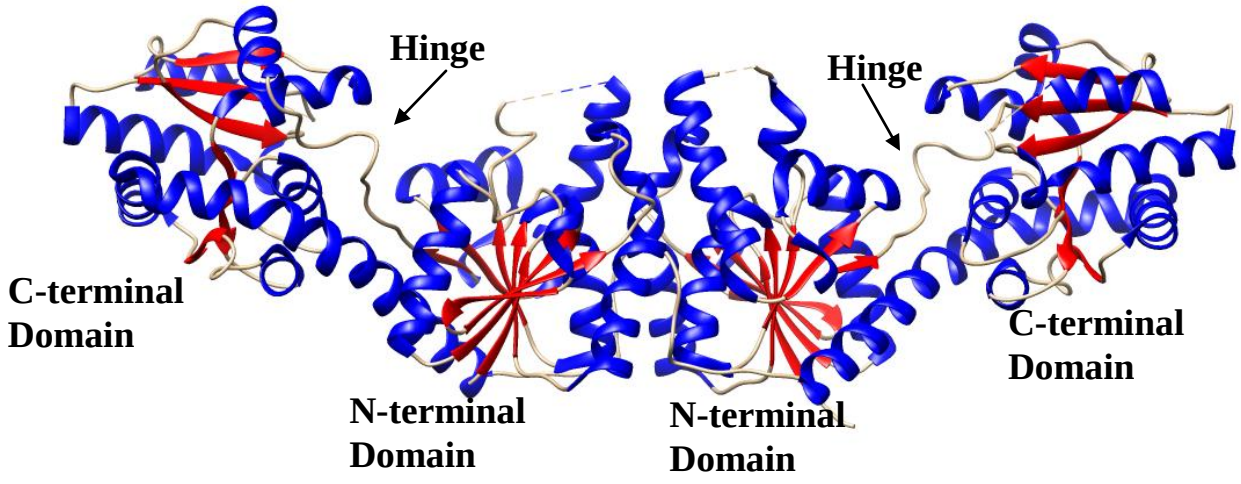


Figure 3.1. Crystal structure of *Cg*MshA dimer in the absence of substrates. Blue color represents β strands and red color represents α helices. The *N*- and *C*-terminal domains are connected by the flexible loops labeled as hinge. The interface of dimer is composed of residues involved in *N*-terminal domain. Made with Chimera. PDB: 3c48.

3.1.4 Conformational Changes of *Cg*MshA on Nucleotide Binding

By superimposing the APO structure of *Cg*MshA and the structure of *Cg*MshA bound by the product UDP, a large conformational change was observed in *Cg*MshA upon the nucleotide binding. This conformational change brings the binding sites for substrates UDP-GlcNAc and I1P into close proximity (Figure 3.2). [20] The *C*-terminal domain has a 97º rotation relative to

the *N*-terminal domain and yields a V-shape oligomer. [20] The axis of the rotation is almost parallel to α-helices (residues 357-408) and intersects the hinge residues connecting two domains. Therefore, the nucleotide binding induces the formation of a closed form of *Cg*MshA. Figure 3.3 shows the closed form of *Cg*MshA dimer bound by UDP and I1P. The binding sites for both substrates are located at the cleft between the *N*- and *C*-terminal domains. The majority of residues having interactions with UDP are located in the *C*-terminal domain. [20]

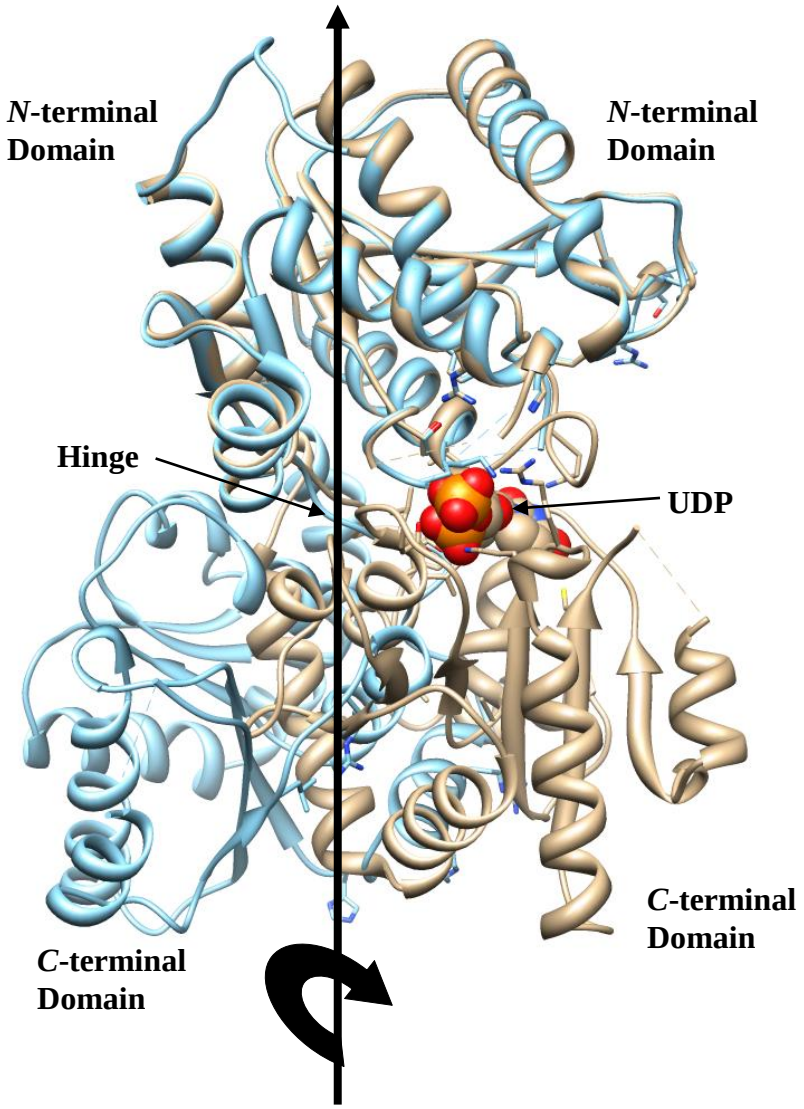


Figure 3.2. Alignment of the *N*- and *C*-terminal domain of *Cg*MshA before and after nucleotide binding. Free *Cg*MshA was labeled in blue color. *Cg*MshA bound by UDP was in tan color. Made with Chimera. PDB: 3c48 and 3c4q.

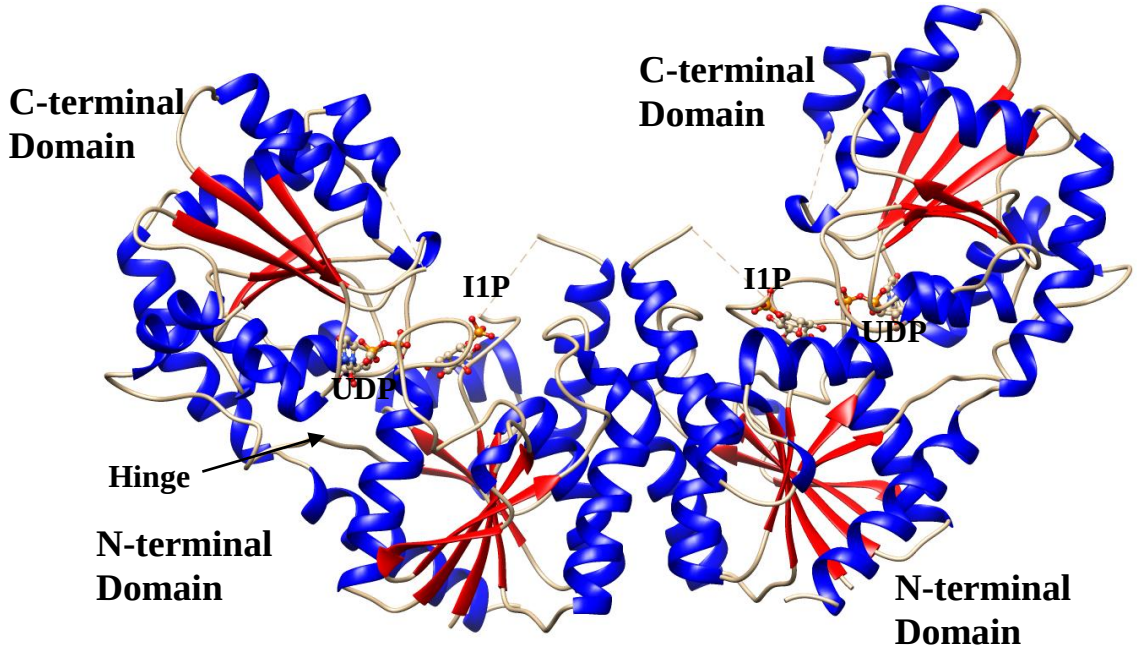


Figure 3.3. Crystal structure of *Cg*MshA dimer in closed conformation. Blue color represents β strands and red color represents α helices. The formation of V-shape oligomer due to the binding of UDP provides binding sites for substrates UDP and I1P. Made with Chimera. PDB: 3c4v.

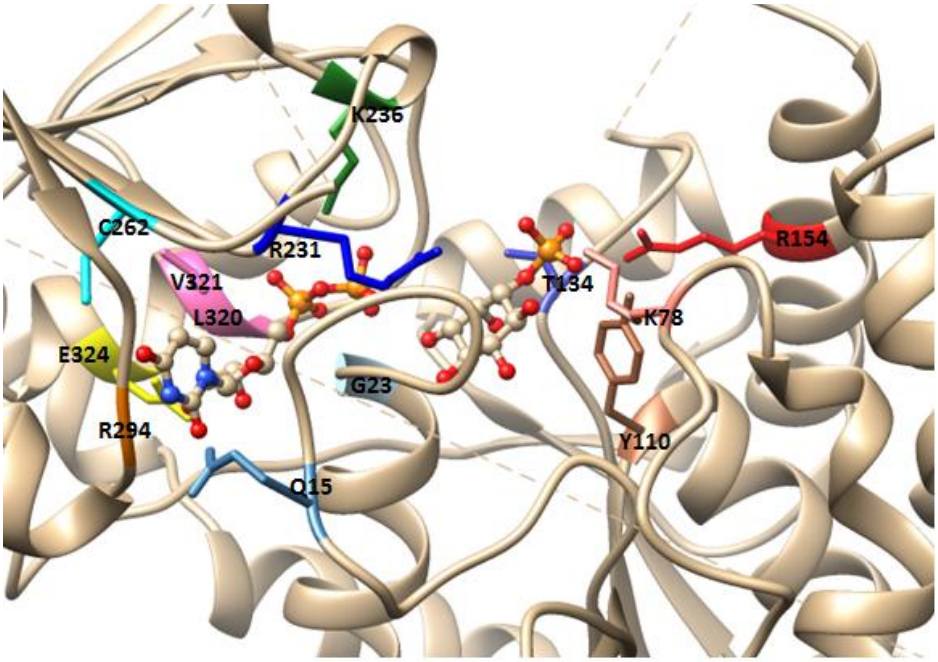


Figure 3.4. Active site residues. The residues contributing to substrate binding were labeled in *Cg*MshA structure with UDP and I1P. Made with Chimera. PDB: 3c4v.

Some residues were proposed to have interactions with UDP and I1P and may play an important role in substrate binding (Figure 3.4). [20] As shown in Figure 3.4, the residues located at the *C*-terminal domain mainly contribute to the UDP binding. The uracil moiety of UDP may form polar interactions with the main-chain atoms of Arg294 and the side chain of Cys262. The ribose of UDP may have an interaction with the side chain of Glu324. And, the pyrophosphate group may be stabilized by interacting with the main-chain atoms of Leu320, Val321, and the side chains of Arg231 and Lys236. Upon nucleotide binding and domain rotation, some residues from the *N*-terminal domain also contribute to UDP binding. In the new conformation, Gln15 may form a hydrogen bond with the uracil moiety, and the backbone amide of Gly23 may be hydrogen bonded to the pyrophosphate group. No significant structural changes were identified between crystal structures of *Cg*MshA-UDP-I1P and *Cg*MshA-UDP complexes, but several key residues are required for the binding of I1P. [20] The side chains of Lys78 and Arg154 can form electrostatic interactions with the phosphate moiety of I1P. Moreover, the side chains of Tyr10 and Thr134 may form hydrogen bonds with the phosphate group of I1P.

3.1.5 Conclusion

Based on the APO structure of *Cg*MshA and the structure of *Cg*MshA with the substrates, the large conformational change of *Cg*MshA upon nucleotide binding was identified. [20] However, the dynamic information of *Cg*MshA on substrate binding is still lacking. The comparison of crystal structures of *Cg*MshA-UDP and *Cg*MshA-UDP-I1P complexes reveals that the binding of I1P does not induce significant structural changes, but dynamic changes of *Cg*MshA on the I1P binding are expected. Moreover, the investigation of *Cg*MshA dynamics can provide important information for other GT-B glycosyltransferases. So HDX-MS was employed to study dynamic changes of *Cg*MshA on substrate binding. The positively charged residues

His9, Lys78, and Arg154 were predicted to form electrostatic interactions with the phosphate moiety of I1P. Moreover, the residue Arg231 was proposed to form a polar interaction with the phosphate group of UDP. [20] Therefore, site-directed mutagenesis has been utilized to probe how these residues participate in catalysis.

3.2 Materials and Methods

3.2.1 Materials

Primers for mutagenesis of *Cg*MshA gene were obtained from Eurofins Genomics (Huntsville, AL). UDP-GlcNAc was purchased from Calbiochem (Billerica, MA). Uridine-5'-diphosphate disodium salt hydrate (UDP), NAD^+, NADH, glucose-6-phosphate, phosphoenolpyruvate (PEP), deuterium oxide (D_2O), and porcine pepsin (3200-4500 u/mg) were purchased from Sigma-Aldrich (St. Louis, MO). The coupling enzymes pyruvate kinase (PK) and lactate dehydrogenase (LDH) were obtained from Roche Diagnostics (Indianapolis, IN). The HisTrap HP, Q Sepharose HP and Phenyl Sepharose HP column were obtained from GE Healthcare (Piscataway, NJ). Microbore C18 reverse phase column with internal diameter between 0.5-2.0 mm and a guard column were bought from Phenomenex (Torrance, CA). Competent *E. coli* cells (BL21(DE3)pLysS and XL-10Gold) were purchased from Invitrogen (Grand Island, NY). All other reagents were obtained from VWR (Radnor, PA).

3.2.2 Overexpression of *Cg*MshA

The plasmid MshA-pET29a was transformed into BL21(DE3)pLysS *E. coli* cells and grown in LB media with 0.03 mg/ml of kanamycin, 0.035 mg/ml of chloramphenicol, and 5 mg/ml glucose at 37°C overnight. Then, overnight media was induced into auto-induction media (2 mM $MgSO_4$, 25 mM Na_2HPO_4, 19.6 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 14 mM disodium succinate, 97 mM glycerol, 190 mM lactose, 17 mM glucose, 0.1 g/ml yeast extract,

and 0.4 g/ml tryptone) with 0.1 mg/ml of kanamycin and 0.035 mg/ml of chloramphenicol at 25°C, 280 rpm until OD_{600} of media reached 12. Finally, the cells were harvested by centrifugation at 6330 xg for 10 minutes at 4°C.

3.2.3 Purification of *Cg*MshA

The cell pellets were resuspended in lysis buffer (100 mM TEA (pH 7.8), 200 mM ammonium sulfate, 10% glycerol, and 15 mM imidazole), lysozyme (0.25 mg/ml), and DNAase (10 µg/ml). The resuspended cells were sonicated, and the supernatant was collected by centrifugation at 34540 xg for 30 minutes at 4ºC. *Cg*MshA was purified by using Ni^{2+} HisTrap column (5 ml) with a linear gradient of 15 mM to 300 mM imidazole of 20 column volumes. The purity of protein was checked by SDS-PAGE gel, and the fractions containing *Cg*MshA were mixed with 1 M of ammonium sulfate at 4°C for 30 minutes. After the centrifugation of fractions with SS-34 rotor at 4500 xg and 4°C, the fractions containing *Cg*MshA were loaded onto 20 ml phenyl sepharose column which was equilibrated with buffer B (20 mM TEA (pH 7.8), 1 M ammonium sulfate, 10% glycerol, 0.5 mM EDTA, and 1 mM DTT). The bound protein was eluted out with a linear gradient of 1 M to 0 M $(NH_4)_2SO_4$ of 20 column volumes. The purest fractions were dialyzed against 20 mM TEA (pH 7.8). Finally *Cg*MshA was concentrated by ultracentrifugation to the concentration of about 200 µM and stored in -20°C.

3.2.4 Site-directed Mutagenesis of *Cg*MshA

Oligonucleotide primers of each substitution were designed and are listed in Table 3.1. The mutations in MshA-pET29a plasmid were created by using QuikChange Lightning site-directed mutagenesis kit (Stratagene). The detailed procedures of amplifying *Cg*MshA variants are the same to *Mt*GpgS variants that have been described in 2.2.4. Then, the overexpression and purification of *Cg*MshA variants are similar to the wild type *Cg*MshA.

Table 3.1 Sequences of forward and reverse primers of each *Cg*MshA Substitution.

*Cg*MshA Substitution Primer	Primer Sequence (5′ - 3′)
H9A Forward	-cgtagctatgatttccatggccacctctccattgcagcag -
H9A Reverse	-ctgctgcaatggagaggtggccatggaaatcatagctacg -
K78A Forward	- cgtatgaggggctttccgcagaggagcttcctactc -
K78A Reverse	-gagtaggaagctcctctgcggaaagcccctcatacg -
R154A Forward	- cggagtcggaggcggctcgcatttgtgagc-
R154A Reverse	-gctcacaaatgcgagccgcctccgactccg -
R231A Forward	-tagtggcttttgtgggtgcgttgcagccgtttaagg -
R231A Reverse	-ccttaaacggctgcaacgcacccacaaaagccacta -

3.2.5 Production of I1P

The gene for inositol-1-phosphate synthase is from *Archaeoglobus fulgidus* and termed as *Af_INO*. The ligation and amplification of *Af_INO*-pET23a plasmid has been described in references 20 and 21. The amplified *Af_INO*-pET23a plasmid was transformed into BL21(DE3)pLysS *E. coli* cells and grown in LB media with 0.03 mg/ml ampicillin, 0.035 mg/ml chloramphenicol, and 5 mg/ml glucose at 37 ºC overnight. The *Af_INO* enzyme was expressed in auto-induction media containing 0.1 mg/ml ampicillin and 0.035 mg/ml chloramphenicol. The auto-induction media containing cells were incubated at 37 ºC and shaken for 18h. Finally the cells were harvested by centrifugation at 6330 ×g for 10 minutes at 4°C.

The cell pellets were resuspended in buffer A (50 mM Tris, pH 8.5) and sonicated. The supernatant was collected by centrifugation at 17,000 rpm for 30 min. The collected supernatant was heated in 80 ºC water bath for 5 min and then cooled to 4 ºC. Precipitated proteins were removed by centrifugation. The new collected supernatant was loaded onto a Q anion exchange

column which was equilibrated with the buffer A. The target protein was eluted out with a linear gradient of 0 M to 0.25 M $(NH_4)_2SO_4$ of 10 column volumes. The fractions containing target protein were further purified by utilizing a phenyl sepharose column with a linear gradient of 1 M to 0 M $(NH_4)_2SO_4$. The purity of the target protein was checked by SDS-PAGE, and the pure fractions were dialyzed in buffer A. Pure *Af_INO* protein was concentrated by ultrafiltration and stored in -20 ºC.

The synthesis of I1P required *Af_INO*, glucose-6-phosphate, and NAD^+. The starting condition for I1P synthesis was 125 mM glucose-6-phosphate, 0.625 mM $ZnCl_2$, 1.25 mM NAD^+, 3.6 mg *Af_INO*, and 50 mM Tris, pH 7.5. The reaction was incubated in 85 ºC water bath. Every 45 min 0.5 mM NAD^+ and 2 mg *Af_INO* were added into the reaction. The progress of I1P synthesis was monitored by PK/LDH coupled assays explained in 3.2.6. The concentration of I1P was detected by decreasing amount of NADH. Finally the reaction was stopped when the I1P concentration reached 125 mM.

3.2.6 Steady State Assays and Determination of Kinetic Parameters

The enzymatic activity of *Cg*MshA was measured by PK/LDH coupled assays at 340 nm (Illustration 3.3). The standard condition of assay was 50 mM TEA (pH 7.8), 10 mM $MgCl_2$, 200 µM NADH, 500 µM PEP, 20 units of pyruvate kinase (PK), and 55 units of lactate dehydrogenase (LDH), and saturating concentration of each substrate (UDP-GlcNAc and I1P) in a 1ml reaction at 25 ˚C. The reactions were initiated by the addition of enzyme. The determination of kinetic parameters has been described in 2.2.6.

3.2.7 Sequencing of *Cg*MshA by Tandem MS/MS

The amino acid sequences of *Cg*MshA generated from pepsin digestion were determined. *Cg*MshA was digested with pepsin and the proteolytic peptides were sequenced using MS/MS

CID. The stock *Cg*MshA was stored in 20% glycerol at a concentration of 100 µM. The procedures of sequencing were similar to that of *Mt*GpgS. 4 µl of the stock *Cg*MshA was mixed with 46 µl of H_2O. The diluted protein was incubated with 50 µl of quench solution (0.1 M potassium phosphate, pH 2.4). The protein was digested on ice with 6 µl of 5 mg/ml pepsin for 5 minutes. The final concentrations of *Cg*MshA and pepsin were 4 µM and 0.3 mg/ml, respectively. The generated peptides were loaded onto the C18 reverse-phase column, which has been equilibrated with solvent A (98% H_2O, 2% ACN, and 0.4 % FA). The peptides were separated by solvent B (98% ACN, 2% H_2O, and 0.4 % FA) with a linear gradient 0% to 50% for 25 minutes (Table 3.2). The separated peptides were sequenced by the ion trap mass spectrometer. Finally the MS/MS data was analyzed by Peaks Client software and the peptide map of *Cg*MshA was determined.

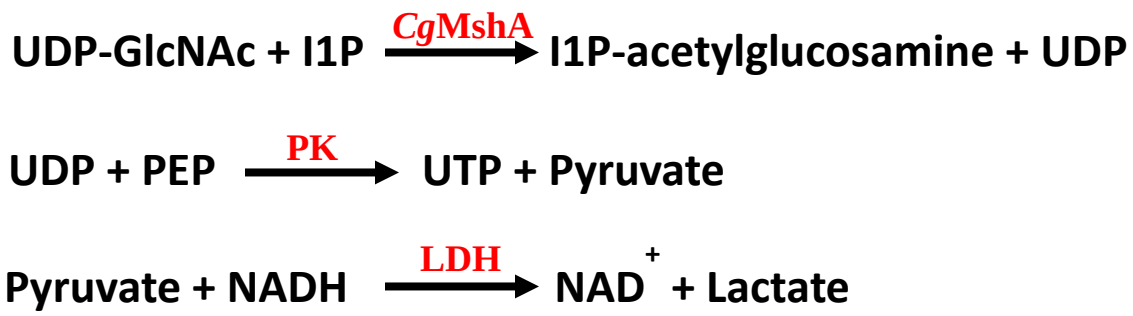


Illustration 3.3. Schematic of PK/LDH coupled assay measuring enzymatic activity of *Cg*MshA.

3.2.8 Continuous HDX-MS Experiments of *Cg*MshA

The D_2O stock and the quench solution were always kept at 25 °C and 0 °C, respectively, during the course of experiment. Separate HDX-MS experiments were performed with free *Cg*MshA, *Cg*MshA-UDP-GlcNAc, *Cg*MshA-I1P, *Cg*MshA-UDP, and *Cg*MshA-UDP-I1P complexes. The *Cg*MshA-UDP-GlcNAc complex was generated by incubating 100 µM *Cg*MshA with 10 mM UDP-GlcNAc for 30 min. Similarly, the *Cg*MshA-I1P and *Cg*MshA-UDP

complexes were generated by incubating 100 µM *Cg*MshA with 10 mM I1P and 10 mM UDP, respectively, for 30 min. And the *Cg*MshA-UDP-I1P complex was generated by incubating 100 µM *Cg*MshA with 10 mM UDP and 10 mM I1P together for 30 min. The HDX reaction was initiated with 46 µl of 99.5% D_2O added to 4 µl of the protein stock or preformed protein complex. The samples were incubated at 25 °C from 15s to 1h. The reaction was quenched by 50 µl of quench solution on ice. Then, the quenched reaction was digested on ice for 5 min using 6 µl of 5 mg/ml pepsin. The digested peptides were separated by ACN with a linear gradient of 0%-50% for 10 min (Table 3.3). The parameters for the mass spectrometer were set to be the same to *Mt*GpgS described in 2.2.8.

Table 3.2. Elution method of MS/MS sequencing.

Elution Time (min)	% Solvent A	% Solvent B
0.00	100	0
1.00	100	0
26.00	50	50
27.00	0	100
37.00	0	100
38.00	100	0
45.00	100	0

Two control experiments $m_{0\%}$ and $m_{100\%}$ were run to correct HDX data. For the $m_{0\%}$, protein stock (4 µl of 100 µM) was incubated with 46 µl of ultrapure water at 25 °C, followed by quenching and digestion as described above. For the $m_{100\%}$, protein stock was incubated with 11-fold excess of D_2O at 50°C for 8 hours. The reaction was quenched using 50 µl of quench solution and transferred to ice immediately. Then quenched solution was digested as described

above. Finally the HDX-MS data of the continuous HDX experiments including two controls were loaded into HD Examiner software, and dynamic curves of all peptides were made by KaleidaGraph.

Table 3.3. Elution method of continuous HDX-MS experiments.

Elution Time (min)	% Solvent A	% Solvent B
0.00	100	0
1.00	100	0
11.00	50	50
12.00	0	100
22.00	0	100
23.00	100	0
30.00	100	0

3.3 Results

3.3.1 Steady State Kinetics of Wild Type *Cg*MshA and Variants

PK/LDH coupled assays were performed to determine the kinetic parameters of wild type *Cg*MshA and its variants. For wild type *Cg*MshA, The k_{cat} value is 17.6 ± 1.1 s^{-1}. The K_m values for UDP-GlcNAc and I1P are 0.48 ± 0.09 mM and 0.22 ± 0.04 mM, respectively. Four *Cg*MshA variants H9A, K78A, R154A, and R231A were designed and constructed in order to understand the roles of these four residues in catalysis (Figure 3.5). The kinetic parameters of four *Cg*MshA variants were determined by PK/LDH coupled assays and are listed in Table 3.4. Based on the crystal structure of *Cg*MshA complex, the residues H9, K78, and R154 are proposed to form electrostatic interactions with the phosphate moiety of I1P. [20]

Table 3.4. Kinetic parameters of wild type *Cg*MshA and variants. *Performed at K_m concentration of I1P. ND means not detectable.

Enzyme	k_{cat} (s^{-1})	$K_{UDP\text{-}GlcNAc}$ (mM)	K_{I1P} (mM)	k_{cat} / $K_{UDP\text{-}GlcNAc}$ ($s^{-1}mM^{-1}$)	k_{cat} / K_{I1P} ($s^{-1}mM^{-1}$)
Wild Type	17.6 ± 1.1	0.48 ± 0.09	0.22 ± 0.04	37 ± 2	60 ± 7
H9A	0.25 ± 0.03	1.15 ± 0.33	0.72 ± 0.15	0.25 ± 0.05	0.41 ± 0.09
K78A	4 ± 1*	9 ± 4*	5 ± 1	0.64 ± 0.04*	0.71 ± 0.02
R154A	ND	ND	ND	ND	ND
R231A	ND	ND	ND	ND	ND

The substitution of R154 results in undetectable activity, suggesting the residue R154 is crucial to the activity of *Cg*MshA. It is possible that I1P cannot bind to *Cg*MshA without the interaction with the residue R154. While the undetectable activity of the R154A variant can also be explained by the high K_m values of substrates required. Both explanations support that the residue R154 plays an essential role in I1P binding. The H9A variant shows reduced activity on the order of a 70-fold reduction in the k_{cat} value as compared to the wild type *Cg*MshA, suggesting an important role of H9 in *Cg*MshA catalysis. However, the K_m values for UDP-GlcNAc and I1P have minor changes relative to wild type *Cg*MshA. This indicates that both substrates can bind to *Cg*MshA normally despite the residue H9 being substituted by alanine. Thus, the substitution of residue H9 probably leads to the reorientation of the substrate I1P, reducing the reaction velocity. The residue K78 is also proposed to form an electrostatic interaction with I1P and may play an important role in I1P binding. This is supported by kinetic parameters of the K78A variant. The substitution of K78 results in a reduction of k_{cat} value but shows increases in K_m values. The side chain of the residue R231 may form an electrostatic

interaction with the pyrophosphate group of UDP. The substitution of R231 leads to undetectable activity of *Cg*MshA, suggesting the residue R231 is essential to the catalysis of *Cg*MshA.

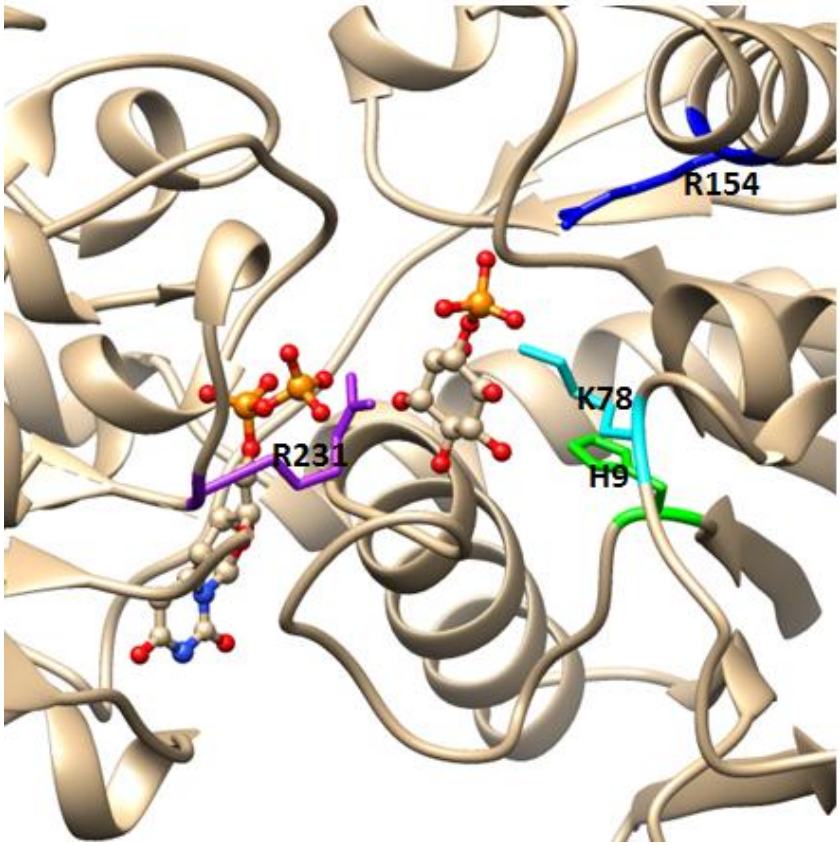


Figure 3.5. Substituted residues of *Cg*MshA. Four residues for substitution were labeled in *Cg*MshA crystal structure bound by UDP and I1P. Made by Chimera. PDB: 3c4v.

3.3.2 Peptide Map of *Cg*MshA

The peptide map of *Cg*MshA was created by analyzing tandem MS/MS sequencing data using the program Peaks Client. The identified peptides with confidence score (10LgD) greater than 20 were chosen. With acetonitrile gradient of 0% to 50% for 25 min, about 92% peptides were determined. In order to minimize the back-exchange, the gradient time was decreased to 10 min, leading to the missing of some peptides. Overall, *Cg*MshA has sequence coverage of 85.2% (Figure 3.6).

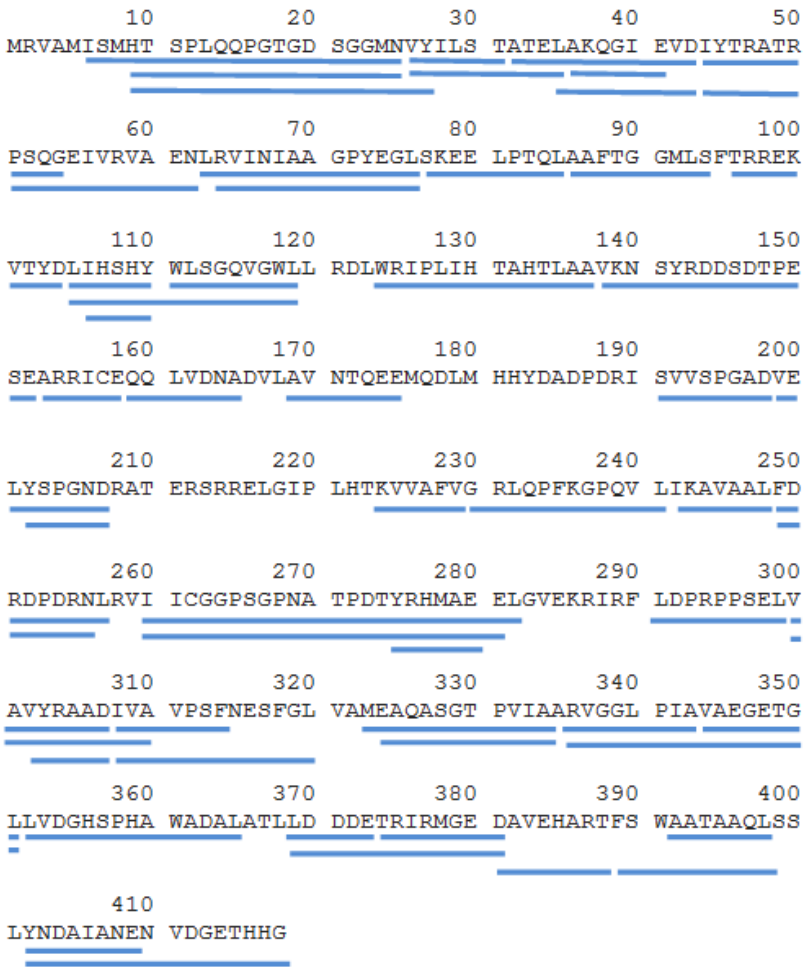


Figure 3.6. Peptide map of *Cg*MshA. *Cg*MshA was digested by pepsin and generated peptides were subject to collision induced dissociation (CID). MS/MS data was analyzed using software Peaks Client.

3.3.3 Solvent Accessibility Changes on Substrate Binding

Substrate binding can induce dynamic changes of proteins performing proper functions. These dynamic changes can be observed by HDX-MS experiments via changes in amount and/or rate of deuterium incorporation for each peptide. The percentage of deuterium incorporated within 15s was used to compare the solvent accessibility of backbone hydrogens for free *Cg*MshA and substrate-bound *Cg*MshA complexes. The difference in solvent accessibility between free *Cg*MshA and *Cg*MshA-UDP-GlcNAc complex is shown in Figure 3.7 (A).

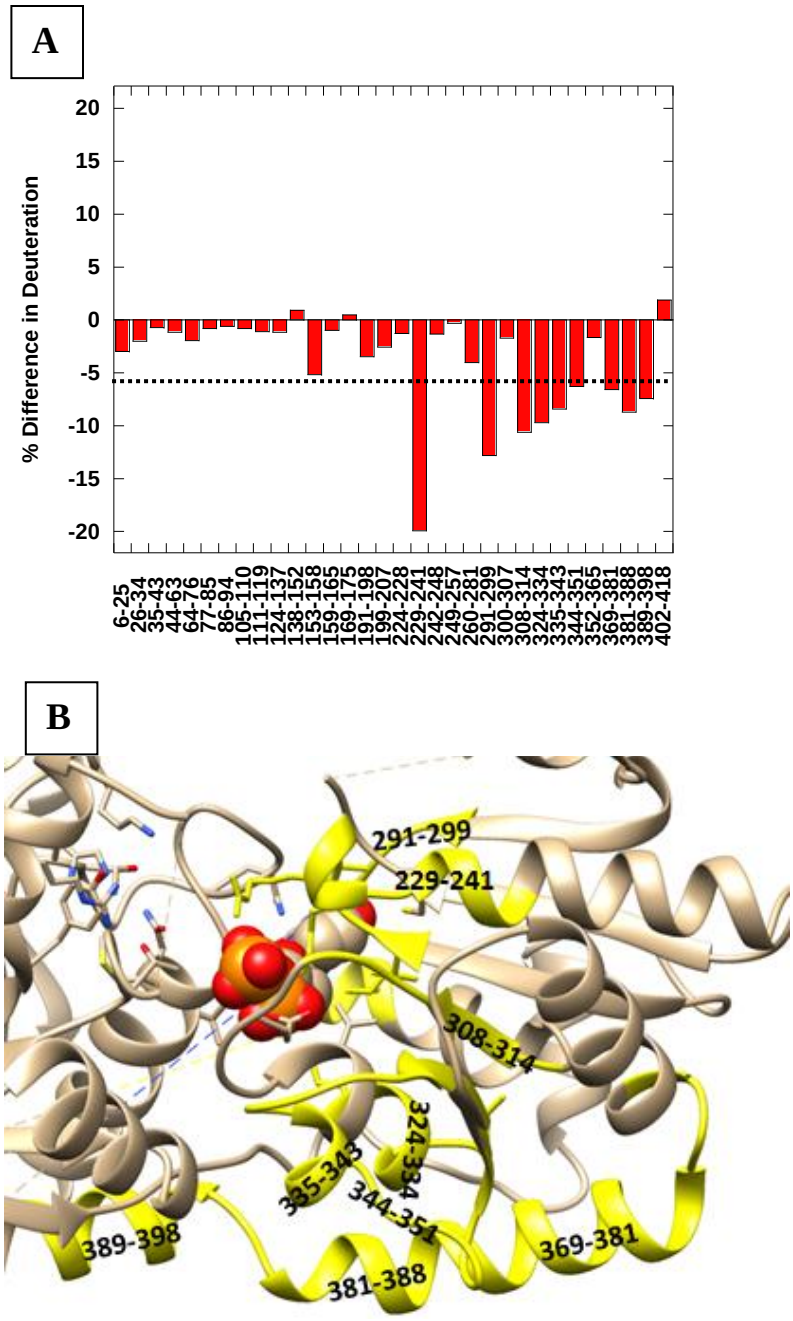


Figure 3.7. Solvent accessibility changes of *Cg*MshA on UDP-GlcNAc binding. (A) % Difference in solvent accessibility between free *Cg*MshA and *Cg*MshA-UDP-GlcNAc complex. (B) Peptides with significant decrease in solvent accessibility were mapped in crystal structure of *Cg*MshA, with UDP molecule in sphere format. Made with Chimera. PDB: 3c4v

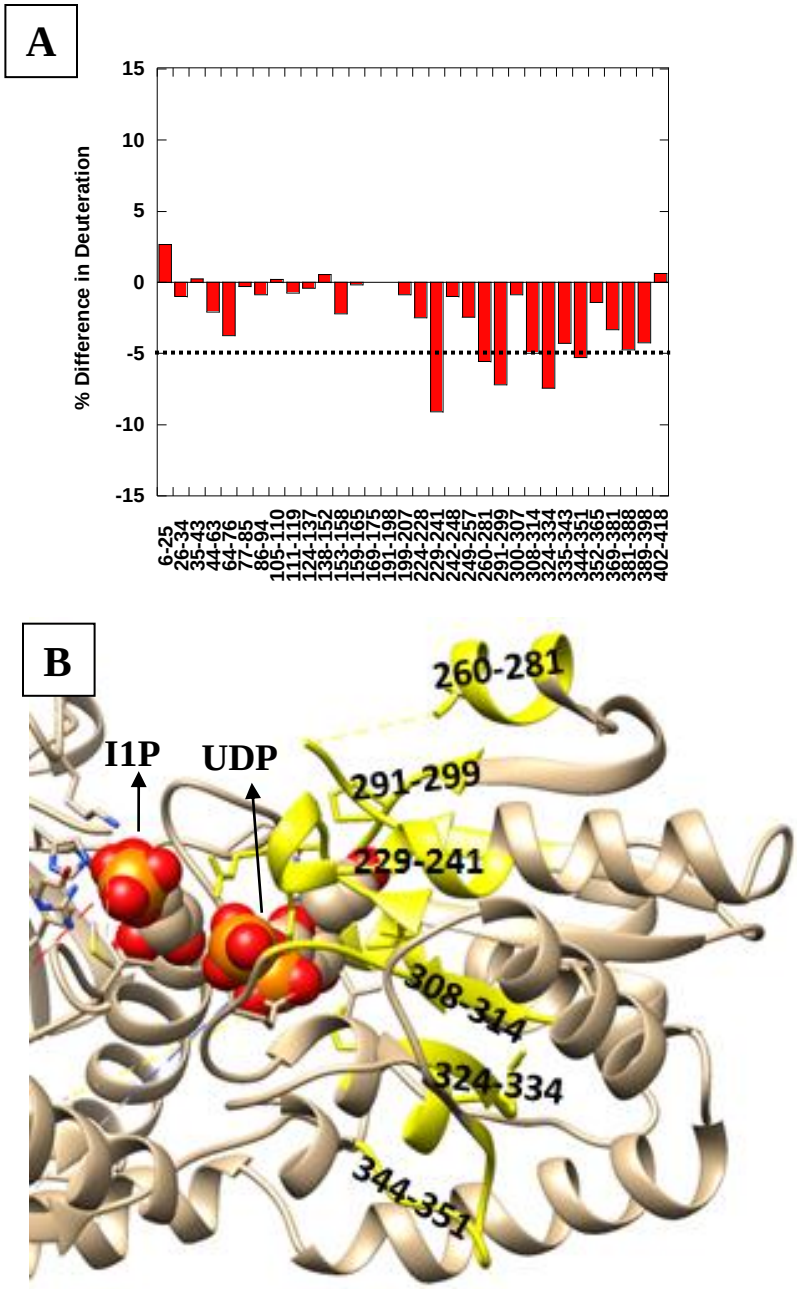
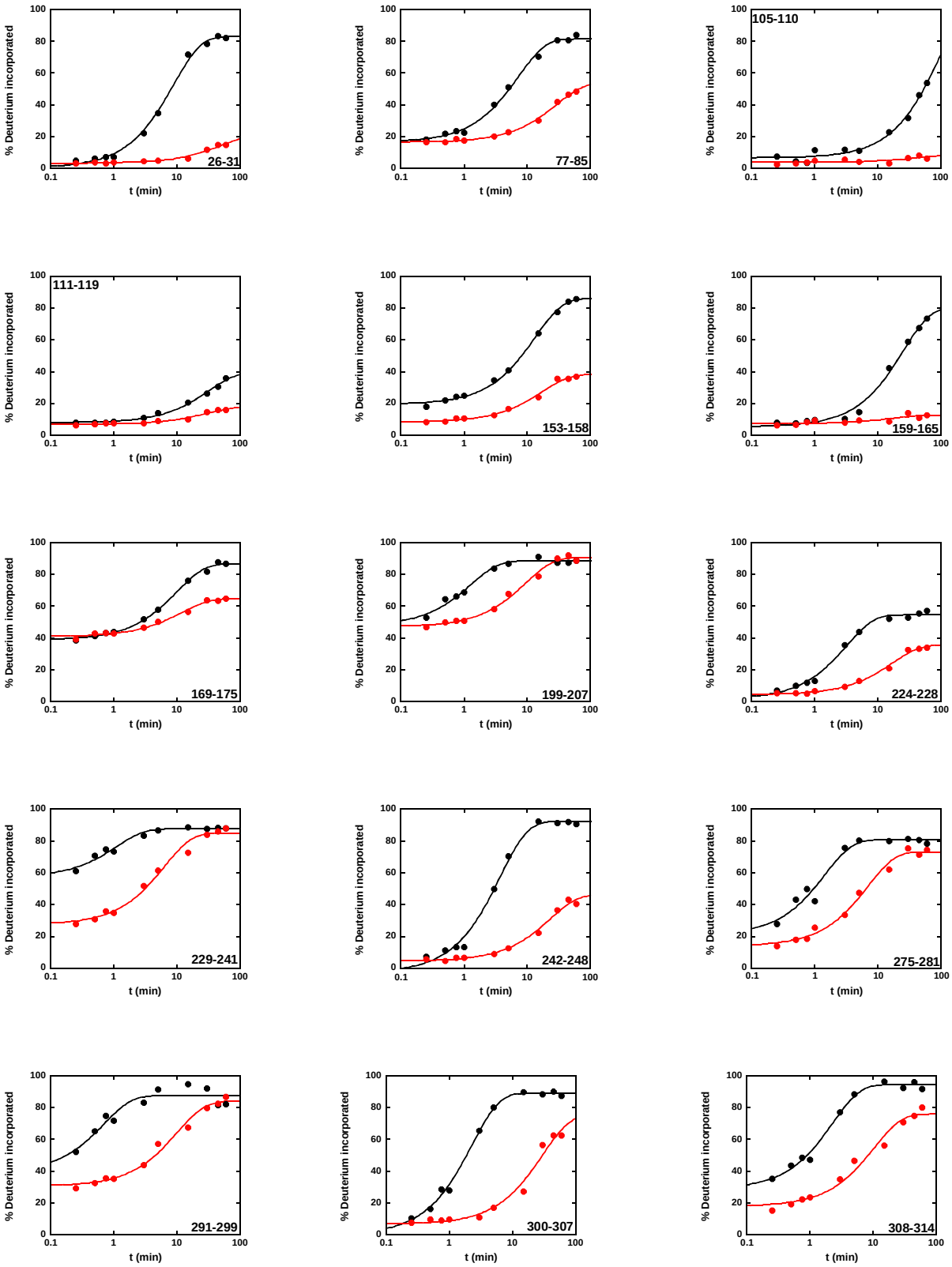


Figure 3.8. Solvent accessibility changes of *Cg*MshA on I1P binding. (A) % Difference in solvent accessibility between free *Cg*MshA and *Cg*MshA-I1P complex. (B) Peptides with significant decrease in solvent accessibility were mapped in crystal structure of *Cg*MshA, with UDP and I1P in sphere format. Made with Chimera. PDB: 3c4v


26-31
77-85
105-110
111-119
153-158
159-165
169-175
199-207
224-228
229-241
242-248
275-281
291-299
300-307
308-314
% Deuterium incorporated
t (min)

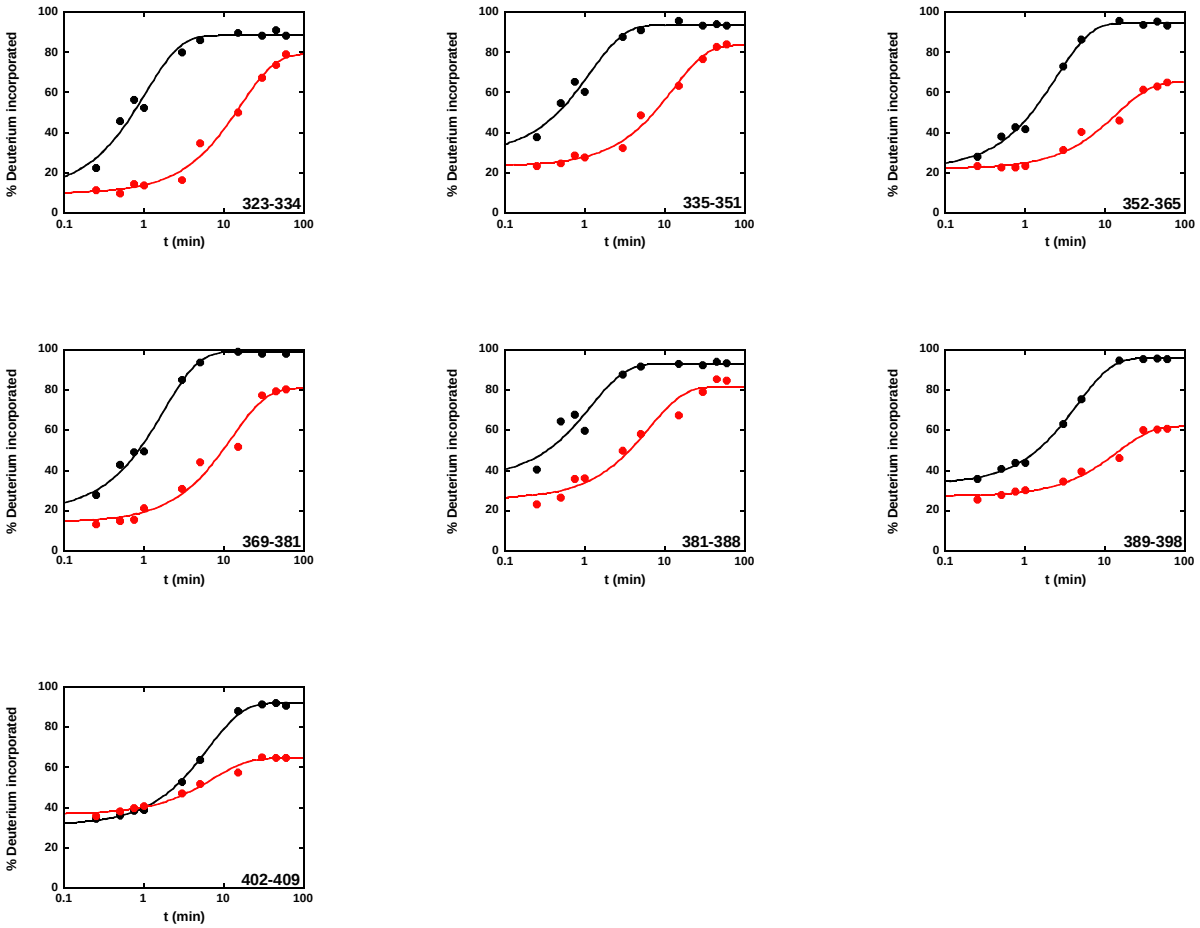


Figure 3.9. Deuterium incorporation curves of peptides from *Cg*MshA on UDP-GlcNAc binding. Black curve represents deuterium incorporation of free *Cg*MshA, and red curve represents that of *Cg*MshA-UDP-GlcNAc complex. Made with KaleidaGraph.

Most peptides of *Cg*MshA are protected by the substrate UDP-GlcNAc, apart from peptides 138-152, 169-175, and 402-418 showing minor increases in solvent accessibility. The peptides showing solvent accessibility differences greater than 5% are labeled in *Cg*MshA structure (Figure 3.7 B). These peptides are located in the *C*-terminal domain of *Cg*MshA, indicating that the *C*-terminal domain provides a binding site for UDP-GlcNAc. This is consistent with the nucleotide binding location in the crystal structure. Four peptides 229-241, 291-299, 308-314, and 324-334 show over 10% decreases in solvent accessibility. These four peptides are located at the UDP binding site and contain residues Arg231, Lys236, Arg294, and

Glu324, which are proposed to have interactions with UDP. The modest protection by UDP-GlcNAc is observed in peptides 335-343, 344-351, 369-381, 381-388, and 389-398, which are surrounding the nucleotide binding site. Therefore, the differences in solvent accessibility upon UDP-GlcNAc binding lead to the conclusion that the substrate UDP-GlcNAc can bind to *Cg*MshA independently and the *C*-terminal domain provides a binding site.

The solvent accessibility of backbone hydrogens between free *Cg*MshA and *Cg*MshA-I1P complex was also compared. As shown in Figure 3.8 (A), most peptides are protected by I1P since these peptides show decreases in solvent accessibility. The peptides showing decreases in solvent accessibility greater than 5% are labeled in the crystal structure of *Cg*MshA (Figure 3.8 B). Peptides 229-241, 260-281, 291-299, 308-314, 324-334, and 344-351 contribute to the I1P binding, since these peptides have decreases in solvent accessibility. Thus, I1P can interact with the *C*-terminal domain of *Cg*MshA in the absence of UDP-GlcNAc.

3.3.4 Dynamic Changes of *Cg*MshA on Substrate Binding

The difference in solvent accessibility on substrate binding shows both UDP-GlcNAc and I1P can bind to *Cg*MshA independently. So, it is attractive to study the dynamic changes of *Cg*MshA induced by both substrates. The significant loss of conformational flexibility is observed from deuterium incorporation for *Cg*MshA in complex with the substrate UDP-GlcNAc (Figure 3.9). The peptides showing dynamic changes on UDP-GlcNAc binding are labeled in the crystal structure of *Cg*MshA dimer (Figure 3.10). These peptides are located in both the *N*-terminal and *C*-terminal domains. The dynamic changes of these peptides support that *Cg*MshA has a major conformational change due to the binding of UDP-GlcNAc.

As observed in Figure 3.7, UDP-GlcNAc binding leads to solvent accessibility decreases of peptides located in the *C*-terminal domain, pointing out the possibility of the *C*-terminal

domain having conformational changes. This is supported by deuterium incorporation changes of peptides from the *C*-terminal domain. As discussed in 3.1.4, the residues Arg231 and Lys236 are proposed to form polar interactions with the pyrophosphate group of UDP-GlcNAc. This is supported by 6-fold decrease in exchange rate of peptide 229-241. Moreover, peptide 291-299 has significant reduction in exchange rate by 14-fold on UDP-GlcNAc binding, supporting the formation of an interaction between the residue Arg294 and the uracil moiety. Another region contributing to UDP-GlcNAc binding is located at peptide 323-334. The residue Glu324 is predicted to form an interaction with the ribose, supported by 15-fold reduction in exchange rate of the peptide 323-334. Several loops involved in the *C*-terminal domain also show loss of conformational flexibility. Peptide 199-207, composing the hinge connecting two domains of *Cg*MshA, shows a decrease in exchange rate by 7-fold. Loop 312-319 is observed to surround the nucleotide in the structure of *Cg*MshA-UDP complex. So, *Cg*MshA is predicted to have a conformational change when UDP-GlcNAc binds to it. This proposal is supported by 4-fold reduction in exchange rate of peptide 308-314. Moreover, peptide 335-351 containing loops 336-339 and 344-349, has a decrease in exchange rate by 10-fold. It indicates the flexibility of loops is important for substrate binding. In the *C*-terminal domain, more regions are detected to have dynamic changes, though they are not located at the binding site. Peptides 224-248, 242-248, and 275-281 composing the Rossmann-like fold, show significant decreases in deuterium incorporation and exchange rate. Moreover, the two helixes, peptides 300-307 and 352-365, have decreases in exchange rate by 13-fold and 6-fold, respectively. Peptides 369-381 and 381-388, located at the surface of the *C*-terminal domain, also have decreases in exchange rate. Overall, the dynamic changes of these peptides support the conclusion that the UDP-GlcNAC binding induces a large conformational change in the *C*-terminal domain.

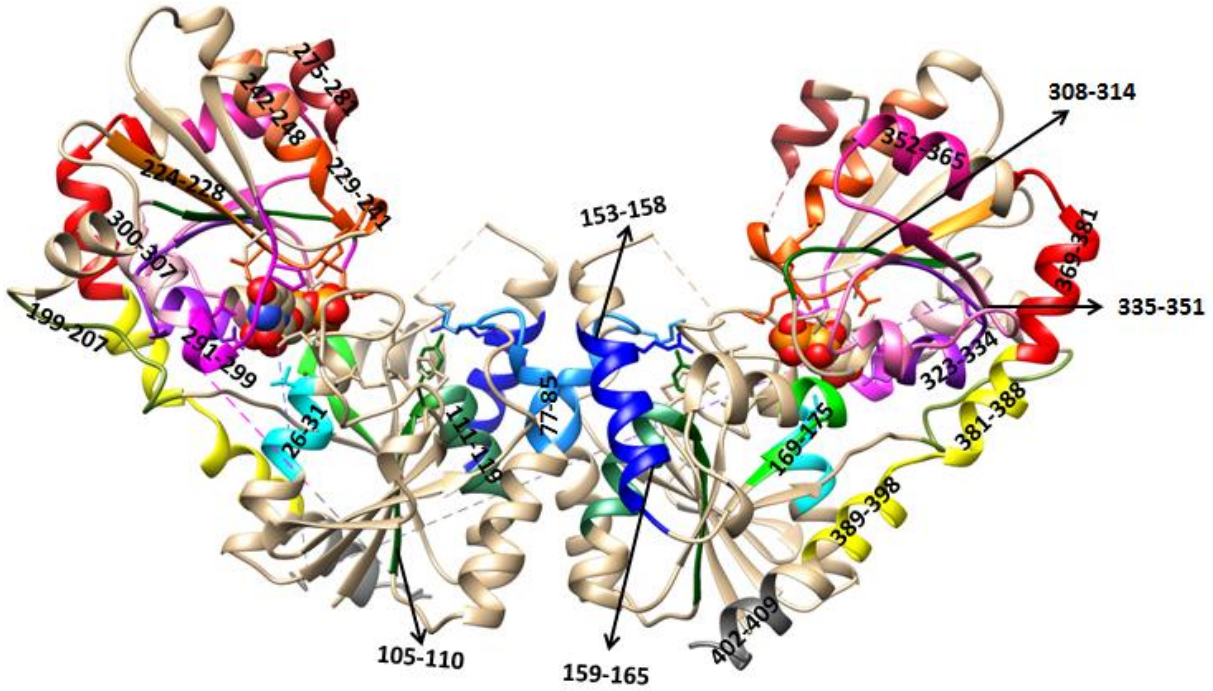
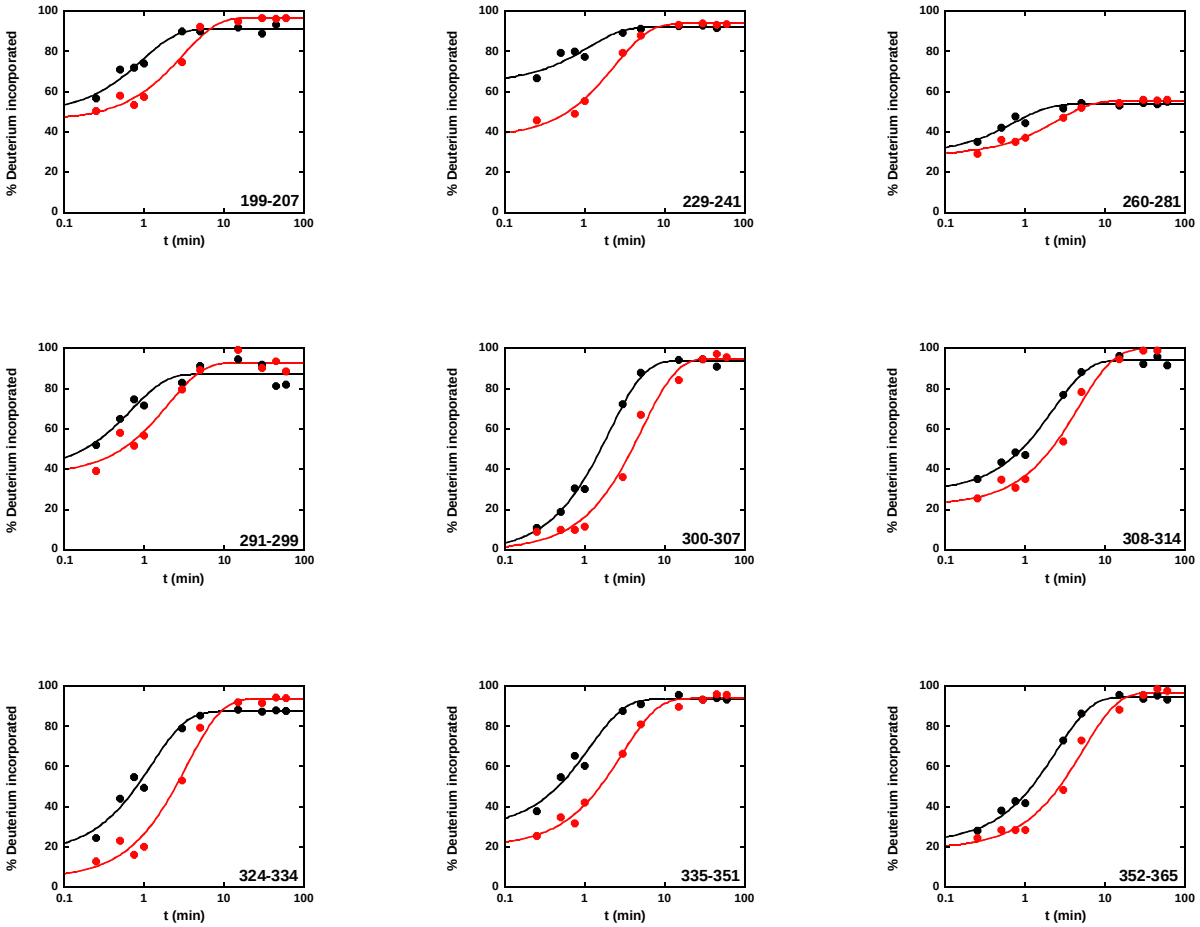


Figure 3.10. Dynamic changes of *Cg*MshA on UDP-GlcNAc binding. Peptides having dynamic changes on UDP-GlcNAc binding were labeled on crystal structure of *Cg*MshA dimer bound by UDP marked in sphere. Made with Chimera. PDB: 3c4v

As shown in Figure 3.10, not only the *C*-terminal domain but also the *N*-terminal domain has dynamic changes on UDP-GlcNAc binding, supporting that the nucleotide binding induces a large conformational change of *Cg*MshA observed in crystal structures of free *Cg*MshA and *Cg*MshA complex. Several regions located at the *N*-terminal domain are detected to have decreasing dynamics, suggesting these regions may have interactions with UDP-GlcNAc. Peptide 26-31 has a decrease in exchange rate by 5-fold, supporting the residue Val26 may form a hydrogen bond with the uracil moiety. Peptides 105-110 and 169-175 may contribute to the stabilization of acetylglucosamine molecule since the residues Tyr110 and Asn171 may form hydrogen bonds with the acetylglucosamine. This is supported by the decreasing dynamics of the peptides 105-110 and 169-175. The crystal structure of *Cg*MshA dimer displays that the *N*-

terminal domains of two *Cg*MshA monomers provide the interface for the dimer. The HDX data suggest that the interface has a dynamic change due to the binding of UDP-GlcNAc. Peptides 77-85 and 111-119 show about 20% decrease in deuterium incorporation, and peptides 153-158 and 159-165 also have significant decreases in amount of deuterium incorporated, indicating that some backbone hydrogens of dimer interface are buried due to the UDP-GlcNAc binding. Furthermore, another two peptides 389-398 and 402-409, located at the surface of the *N*-terminal domain, show about 30% decrease in deuterium incorporation.



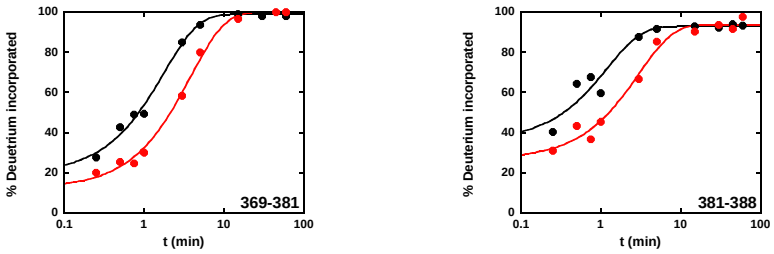


Figure 3.11. Deuterium incorporation curves of peptides from *Cg*MshA on I1P binding. Black curve represents deuterium incorporation of free *Cg*MshA, and red curve represents that of *Cg*MshA-I1P complex. Made with KaleidaGraph.

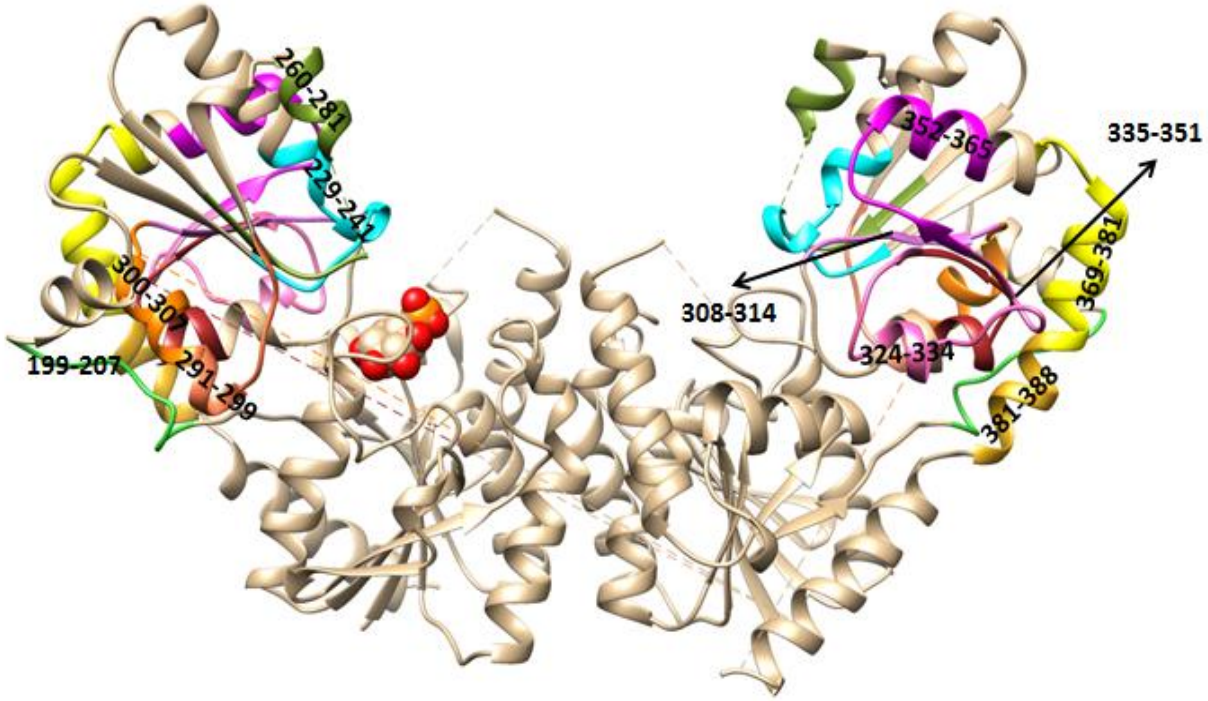
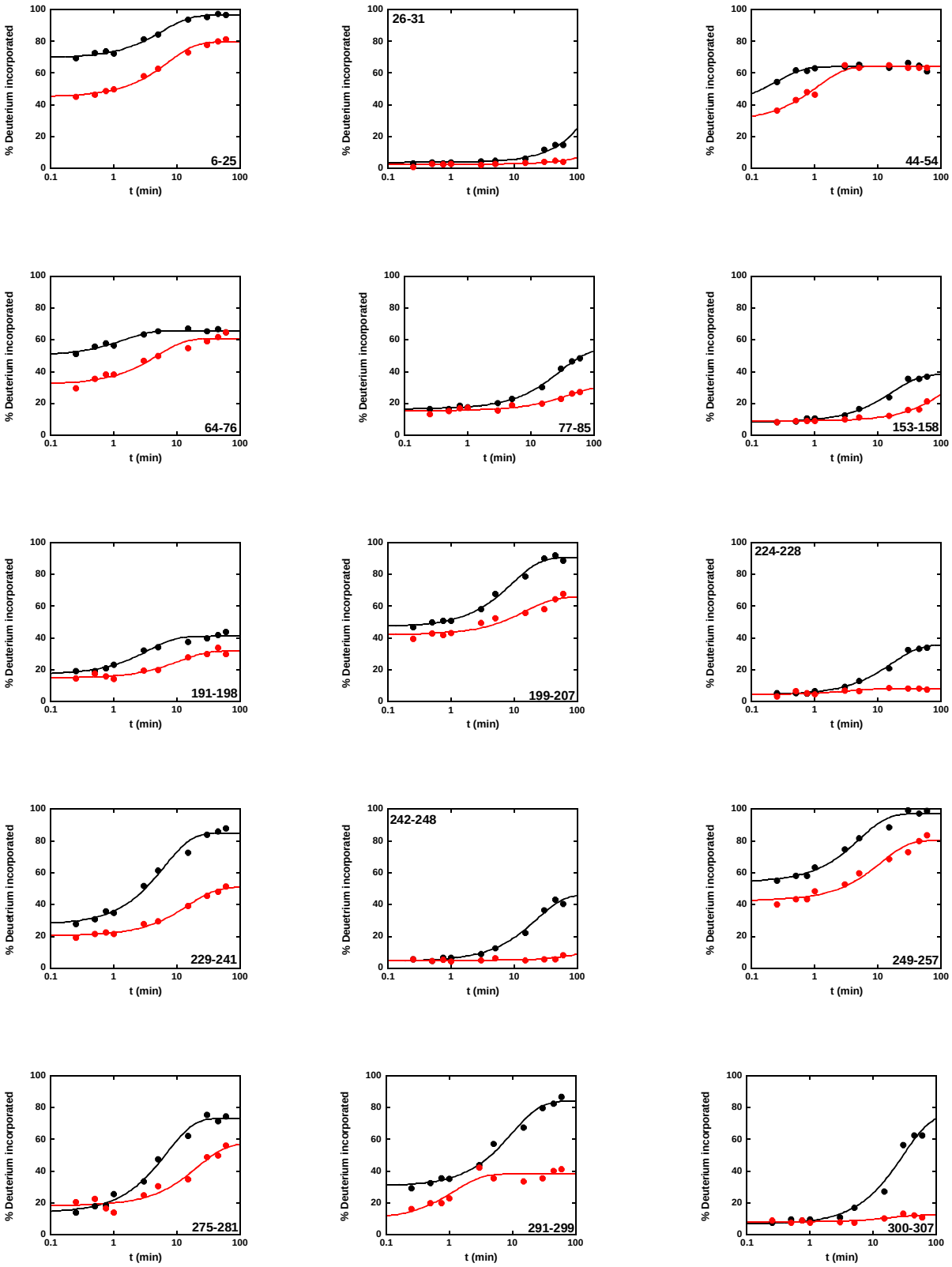


Figure 3.12. Dynamic changes of *Cg*MshA on I1P binding. Peptides having dynamic changes on I1P binding were labeled on crystal structure of *Cg*MshA dimer bound by I1P marked in sphere. Made with Chimera. PDB: 3c4v


6-25
26-31
44-54
64-76
77-85
153-158
191-198
199-207
224-228
229-241
242-248
249-257
275-281
291-299
300-307
% Deuterium incorporated
% Deutrium incorporated
t (min)

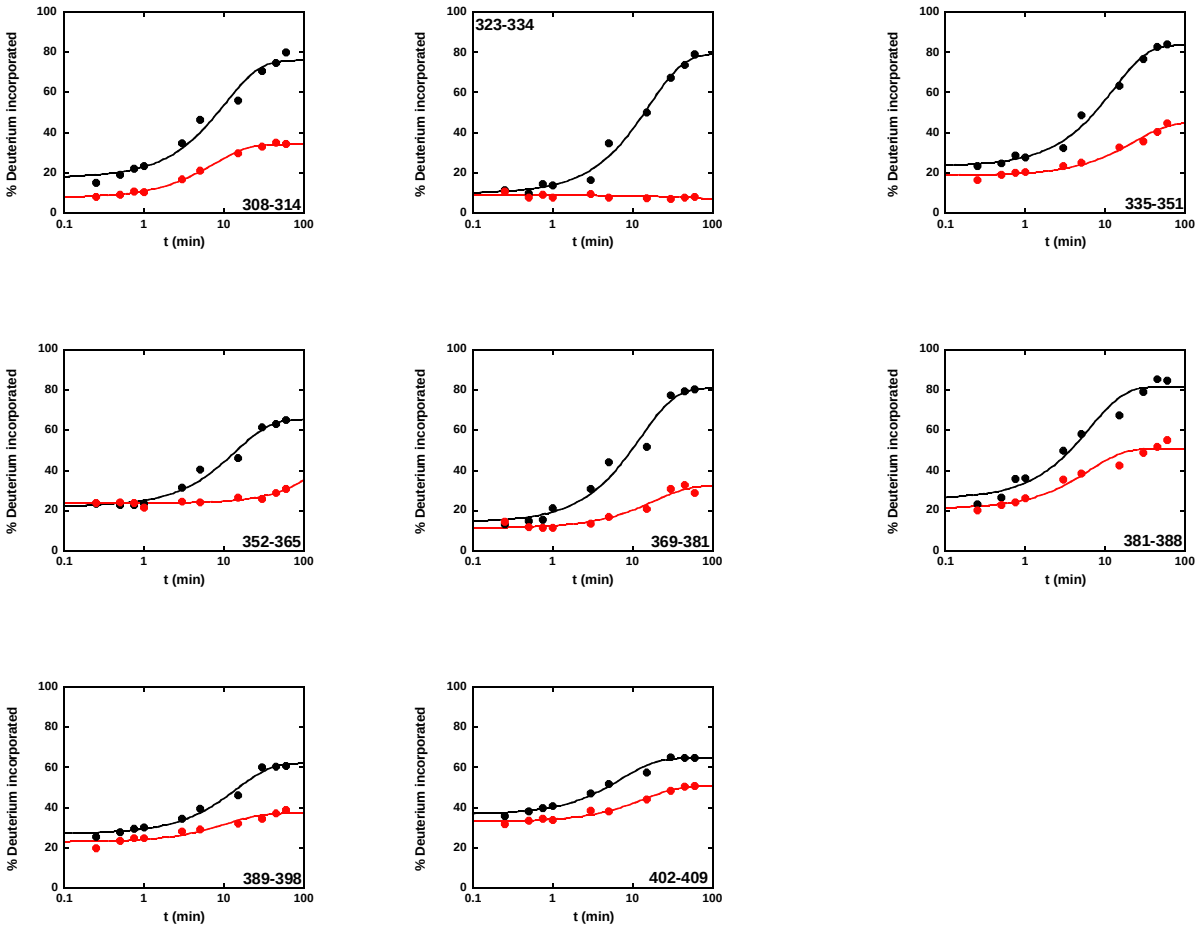


Figure 3.13. Comparison of deuterium incorporation curves between *Cg*MshA-UDP-GlcNAc and *Cg*MshA-UDP complexes. Black curve represents deuterium incorporation of *Cg*MshA-UDP-GlcNAc complex, and red curve represents deuterium incorporation of *Cg*MshA-UDP complex. Made with KaleidaGraph.

The solvent accessibility changes suggest the I1P can bind to the *C*-terminal domain of *Cg*MshA independently, which is strongly supported by the dynamic changes of *Cg*MshA on I1P binding (Figure 3.11). All of these peptides show decreases in exchange rate, suggesting these regions have decreasing dynamics on I1P binding. As shown in Figure 3.12, these peptides are located at the *C*-terminal domain that provides a binding site for I1P. So, the *C*-terminal domain has a dynamic change when I1P binds to it. However, compared with the dynamic changes of *Cg*MshA induced by UDP-GlcNAc, peptides located at the *N*-terminal domain do not have any

dynamic changes on I1P binding. It is very likely that the binding of I1P cannot induce a major conformational change observed from crystal structures. Thus, a conclusion is drawn that both substrates can bind to *Cg*MshA independently and induce dynamic changes of *Cg*MshA, but only UDP-GlcNAc can induce the transition of *Cg*MshA from an open to a closed form.

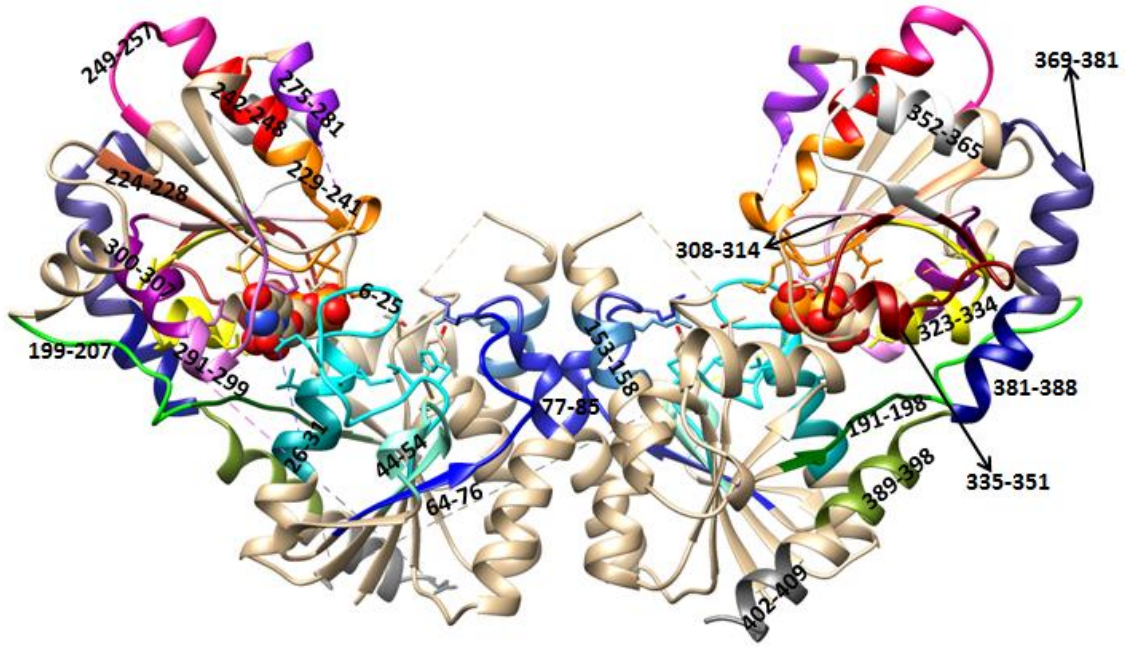


Figure 3.14. Differences in dynamic changes upon UDP and UDP-GlcNAc binding. Peptides showing dynamic changes between *Cg*MshA-UDP-GlcNAc and *Cg*MshA-UDP complexes were labeled in crystal structure of *Cg*MshA dimer. Made with Chimera. PDB: 3c4v

Comparison of crystal structures between *Cg*MshA-UDP complex and *Cg*MshA-UDP-I1P complex displays that the I1P binding does not have effects on *Cg*MshA conformation in the presence of UDP-GlcNAc. [20] However, the dynamic changes of *Cg*MshA on I1P binding are expected in the presence of UDP-GlcNAc. So HDX-MS was employed in order to study the further dynamic changes of *Cg*MshA on I1P binding. HDX-MS experiments of *Cg*MshA bound by UDP were performed first in order to see if UDP can work as UDP-GlcNAc inducing the major conformational change of *Cg*MshA. The deuterium incorporations are compared between free *Cg*MshA and *Cg*MshA-UDP complex. The HDX data shows that UDP can induce the major

conformational change of *Cg*MshA. Interestingly, the differences in deuterium incorporation between *Cg*MshA-UDP and *Cg*MshA-UDP-GlcNAc complexes are observed despite both UDP and UDP-GlcNAc molecules can induce the major conformational change of *Cg*MshA (Figure 3.13). Peptides showing differences in deuterium incorporation are labeled in the *Cg*MshA dimer structure and these peptides are located at both domains (Figure 3.14). These peptides show decreases in exchange rate or deuterium incorporation upon UDP binding compared with *Cg*MshA bound by UDP-GlcNAc. The decreasing dynamics of these peptides may result from the loss of acetylglucosamine molecule. Interestingly, the peptide 6-25 does not show a change in dynamics on UDP-GlcNAc binding while it has 8-fold reduction in exchange rate on UDP binding, suggesting that the loop (residues 7-22) contained by the peptide 6-25 has decreasing dynamics due to the loss of acetylglucosamine. It is very likely that this loop may be involved in the binding of the acetylglucosamine.

Even though the amplitudes of changes induced by UDP-GlcNAc and UDP are different, both of them can induce the major conformational change that is necessary for substrate binding and enzymatic catalysis. So the HDX-MS experiments of *Cg*MshA bound by UDP and I1P were performed and the deuterium incorporation was compared with *Cg*MshA-UDP complex. The HDX data shows that the majority of *Cg*MshA does not have dynamic changes on I1P binding. Only one peptide 191-198 shows 4-fold increase in exchange rate. This peptides is contained the hinge connecting two domains. Thus, the binding of I1P may induce the increase of loop flexibility in the presence of UDP.

3.4 Discussion

MshA is a potential drug target for tuberculosis treatment since disruption of the gene *mshA* leads to MSH biosynthesis being blocked. [17] The structure of *Cg*MshA has been

determined and a conformational change was observed by comparing the structure of free *Cg*MshA and *Cg*MshA-UDP-GlcNAc complex. [20] However, the crystal structures cannot provide dynamic information for proteins. To understand the dynamic changes of enzymes on substrate binding, it is essential to characterize the conformational changes in the active/binding site along with surrounding regions. [22] It also provides insights into reaction mechanism.

3.4.1 Substrate Binding Mechanism of *Cg*MshA

Initial velocity studies were employed to determine whether *Cg*MshA proceeds through a sequential or ping-pong mechanism. [20] The data indicates a sequential mechanism where UDP-GlcNAc almost always binds first followed by I1P. However, the HDX-MS data shows that I1P can interact with the *C*-terminal domain of *Cg*MshA in the absence of UDPG. The peptides 291-299, 229-241, 308-314, and 324-334 may provide a binding site for UDP-GlcNAc, since these peptides have significant decreases in solvent accessibility upon UDP-GlcNAc binding. Moreover, the substrate I1P is identified to bind to the *C*-terminal domain of *Cg*MshA based on the decreases in solvent accessibility and dynamics of *Cg*MshA. However, the crystal structure of *Cg*MshA-UDP-I1P complex shows that the *N*-terminal domain provides a binding site for I1P. Therefore, the substrate I1P can have interactions with the *C*-terminal domain of *Cg*MshA in the absence of UDP-GlcNAc, but this is a non-productive binding. In the presence of UDP-GlcNAc, I1P binds to the formed substrate binding pocket induced by the UDP-GlcNAc binding.

3.4.2 Insights into Dynamic Changes of *Cg*MshA

As discussed in 3.3.4, *Cg*MshA has a major conformational change upon UDP-GlcNAc binding. The peptides located at the binding site have significant decreases in exchange rate or deuterium incorporation, including peptides 224-228, 229-241, 242-248, 291-299, 300-307, 308-

314, and 323-334. The 6-fold reduction in exchange rate of peptide 229-241 is consistent with the undetectable activity of R231A variant, confirming the importance of the residue R231 in *Cg*MshA catalysis. Apart from the dynamic changes in binding site, the surface of the *C*-terminal domain also has dynamic changes, which is supported by decreasing dynamics of peptides 192-207, 275-281, 335-351, 352-365, 369-381, 381-388, and 389-398. These changes support the conclusion that *Cg*MshA goes through a large conformational change on UDP-GlcNAc binding. It is consistent with the rotation of the *C*-terminal domain on nucleotide binding by comparing *Cg*MshA APO and complex structures. Moreover, the peptides 77-85, 153-158, and 159-165, located at the dimer interface, also have decreasing dynamics on UDP-GlcNAc binding. The dynamic changes of these peptides are consistent with the undetectable activity of the R154A variant and high K_m values of the K78A variant, supporting that the R154 and K78 residues play an important role in *Cg*MshA catalysis and substrate binding, respectively.

3.4.3 Roles of Hydrogen Bonds in Substrate Binding

The *Cg*MshA-UDP-GlcNAc complex was not crystallized successfully, so a model of UDP-GlcNAc binding to *Cg*MshA was built. [20] In this model, an important loop composed by residues from 306 to 319 is proposed to contribute to binding and orientation of the sugar moiety GlcNAc, since the model shows that the GlcNAc molecule is bent back over the pyrophosphate moiety and interacts with loop 306-319. Hydrogen bonds are proposed to be formed between 4-OH and the main chain amides of Gly319 and Phe318 and between the 3-OH and the main chain amides of Ser317 and the side chain of Glu316. This model is supported by 4-fold reduction in exchange rate of the peptide 308-314 (Figure 3.9). The loop containing peptide 308-314 has reduction in flexibility due to the hydrogen bonds formed with GlcNAc. Moreover, the side chains of His133 and Asn171 are proposed to form hydrogen bonds with 6-OH of the GlcNAc

moiety. [20] This is supported by the HDX data that peptide 169-175 containing Asn171 has 20% decrease in deuterium incorporation due to the binding of UDP-GlcNAc (Figure 3.9). The hydrogen bonding interactions play an important role in not only GlcNAc moiety binding but also the nucleotide binding. The uracil moiety is stabilized by polar interactions with the main chain of Arg294 and the side chain of Cys262, which is supported by significant decreases in exchange rate of peptide 291-299. The dynamic changes of peptides 323-334 and 229-241 support hydrogen bond formation between Glu324 and the ribose and between pyrophosphate and the side chains of Arg231 and Lys236.

3.4.4 Interdomain Flexibility in GT-B Glycosyltransferases

Crystal structures have revealed open and closed conformations of *Cg*MshA, implying that large-scale conformational dynamics is required by *Cg*MshA catalysis. The HDX data indicates that the peptides located not only at the binding sites but also on the surface have significant reduction in exchange rate and deuterium incorporation. This provides dynamic evidence for the conclusion about the rotation of the *C*-terminal domain and the formation of binding sites. The dynamic changes of *Cg*MshA also provide information on substrate binding and catalysis of other glycosyltransferases in GT-B family. Many GT-B members have been identified to have another conformation on sugar donor binding. *E. coli* MurG adopts a more closed conformation on UDP-GlcNAc binding, which results from the entire *C*-terminal domain rotation. [23] TDP-vancosaminyltransferase (GtfD) and glycogen synthase also have a closed conformation for catalytic process. [24-25] Identification of closed conformation of these glycosyltransferases supports that the GT-B structural family has interdomain flexibility for formation of binding sites and transfer of sugar moiety. However, the identification of closed conformation of these GT-B glycosyltransferases are only based on comparison between APO

and complexes structures; the dynamic characterization is lacking. Therefore dynamic changes of *Cg*MshA detected by HDX-MS provide dynamic information for a GT-B glycosyltransferase undergoing substrate binding-induced transition from an open to a closed state. The dynamic information of *Cg*MshA is particularly important for understanding protein dynamics in GT-B glycosyltransferases and for designing inhibitors against *Cg*MshA.

3.5 References

[1] Hand, C. E., and Honek, J. F. Biological chemistry of naturally occurring thiols of microbial and marine origin. *J Nat Prod*. 2005;68:293-308.

[2] Newton, G. L., Arnold, K., Price, M. S., Sherrill, C., Delcardayre, S. B., Aharonowitz, Y., Cohen, G., Davies, J., Fahey, R. C., and Davis, C. Distribution of thiols in microorganisms: mycothiol is a major thiol in most actinomycetes. *J Bacteriol.* 1996;178:1990-1995.

[3] Rawat, M., Newton, G. L., Ko, M., Martinez, G. J., Fahey, R. C., and Av-Gay, Y. Mycothiol-deficient *Mycobacterium smegmatis* mutants are hyper-sensitive to alkylating agents, free radicals, and antibiotics. *Antimicrob Agents Chemother*. 2002;46:3348-3355.

[4] Zahoor, A., Lindner, S. N., and Wendisch, V. F. Metabolic engineering of *Corynebacterium glutamicum* aimed at alternative carbon sources and new products. *J Comput Struct Biotechnol.* 2012;3:1–11.

[5] Dye, C., Scheele, S., Dolin, P., Pathania, V., and Raviglione, M. C. Global burden of tuberculosis. *JAMA*. 1999;282:677-686.

[6] Young, D. B., and Duncan, K. Prospects for new interventions in the treatment and prevention of mycobacterial disease. *Annu Rev Microbiol*. 1995;49:641-673.

[7] Newton, G. L., Unson, M. D., Anderberg, S. J., Aguilera, J. A., Oh, N. N., delCardayre, S. B., Davies, J., Av-Gay, Y., and Fahey, R. C. Characterization of a *Mycobacterium smegmatis* mutant defective in 1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside and mycothiol biosynthesis. *Biochem Biophys Res Commun*. 1999;255:239-244.

[8] Newton, G. L., Buchmeier, N., and Fahey, R. C. Biosynthesis and functions of mycothiol, the unique protective thiol of *Actinobacteria*. *Microbiol Mol Biol R*. 2008;72:471-494.

[9] Newton, G. L., Ta, P., Bzymek, K. P., and Fahey, R. C. Biochemistry of the initial steps of mycothiol biosynthesis. *J Biol Chem.* 2006;281:33910-33920.

[10] Newton, G. L., Koledin, T., Gorovitz, B., Rawat, M., Fahey, R. C., and Av-Gay, Y. The glycosyltransferase gene encoding the enzyme catalyzing the first step of mycothiol biosynthesis (mshA). J Bacteriol. 2003;185:3476-3479.

[11] Newton, G. L., Av-Gay, Y., and Fahey, R. C. N-Acetyl-1-D-myo-inosityl-2-amino-2-deoxy-alpha-D-glucopyranoside deacetylase (MshB) is a key enzyme in mycothiol biosynthesis. *J Bacteriol.* 2000;182:6958-6963.

[12] Newton, G. L., and Fahey, R. C. Mycothiol biochemistry. *Arch Microbiol*. 2002;178:388-394.

[13] Anderberg, S. J., Newton, G. L., and Fahey, R. C. Mycothiol biosynthesis and metabolism. Cellular levels of potential intermediates in the biosynthesis and degradation of mycothiol in mycobacterium smegmatis. *J Biol Chem.* 1998;273:30391-30397.

[14] Sareen, D., Steffek, M., Newton, G. L., and Fahey, R. C. ATP-dependent L-cysteine: 1D-myo-inosityl 2-amino-2-deoxy-alpha-D-glucopyranoside ligase, mycothiol biosynthesis enzyme MshC, is related to class I cysteinyl-tRNA synthetases. *Biochemistry.* 2002;41:6885-6890.

[15] Koledin, T., Newton, G. L., and Fahey, R. C. Identification of the mycothiol synthase gene (mshD) encoding the acetyltransferase producing mycothiol in actinomycetes. *Arch Microbiol*. 2002;178:331-337.

[16] Sareen, D., Newton, G. L., Fahey, R. C., and Buchmeier, N. A. Mycothiol is essential for growth of Mycobacterium tuberculosis Erdman. *J Bacteriol.* 2003;185:6736-6740.

[17] Buchmeier, N. A., and Fahey, R. C. The *mshA* gene encoding the glycosyltransferase of mycothiol biosynthesis is essential in *Mycobacterium tuberculosis* Erdman. *FEMS Microbiol. Lett.* 2006;264:74-79.

[18] Newton, G. L., Ta, P., and Fahey, R. C. A mycothiol synthase mutant of *Mycobacterium smegmatis* produces novel thiols and has an altered thiol redox status. *J Bacteriol.* 2005;187:7309-7316.

[19] Buchmeier, N. A., Newton, G. L., and Fahey, R. C. A mycothiol synthase mutant of *Mycobacterium tuberculosis* has an altered thiol-disulfide content and limited tolerance to stress. *J Bacteriol.* 2006;188:6245-6252.

[20] Vetting, M. W., Frantom, P. A., and Blanchard, J. S. Structural and enzymatic analysis of MshA from *Corynebacterium glutamicum*. *J Biol Chem*. 2008;283:15834-15844.

[21] Chen, L., Zhou, C., Yang, H., and Roberts, M. F. Inositol-1-phosphate synthase from *Archaeoglobus fulgidus* is a class II aldolase. *Biochemistry*. 2000;39:12415-12423.

[22] Busenlehner, L. S., and Armstrong, R. N. Insights into enzyme structure and dynamics elucidated by H/D exchange mass spectrometry. *Arch Biochem Biophys*. 2005;433:34-36.

[23] Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M., and Walker, S. Crystal structure of the MurG: UDP-GlcNAc complex reveals common structural

principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci U. S. A.* 2003;100:845-849.

[24] Mulichak, A. M., Losey, H. C., Lu, W., Wawrzak, Z., Walsh, C. T., and Garavito, R. M. Crystal structure of vancosaminyltransferase GftD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. *Proc Natl Acad Sci U. S. A.* 2003;100:9238-9243.

[25] Buschiazzo, A., Ugalde, J. E., Guerin, M. E., Shepard, W., Ugalde, R. A., and Alzari, P. M. Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. *EMBO J.* 2004;23:3196-3205.

Chapter 4

CONFORMATIONAL DYNAMICS OF *ESCHERICHIA COLI* SUFBC$_2$D SCAFFOLD COMPLEX ON ADP/MG^{2+} BINDING

4.1 Introduction

4.1.1 Machinery of Suf Pathway

As discussed in Chapter 1, the Suf system is the major biogenesis pathway for Fe-S clusters under oxidative stress in *E. coli*. The Suf system contains six proteins: SufA, SufB, SufC, SufD, SufS, and SufE, encoded by *sufA*, *sufB*, *sufC*, *sufD*, *sufS*, and *sufE*, respectively (Illustration 4.1). SufS is a cysteine desulfurase enzyme using pyridoxal-phosphate (PLP) as a cofactor to mobilize sulfur from L- cysteine. [1-4] SufE binds to SufS tightly and form a stable 1:1 complex. [4] Cys364 of SufS is critical for cysteine desulfurase activity since it forms a persulfide by abstracting a sulfur atom from L-cysteine. [2, 5] This sulfur atom is then transferred from persulfide to Cys51 of SufE. [6-7] The binding of SufE greatly enhances the cysteine desulfurase activity of SufS with regard to that of SufS alone. [4, 6] The high cysteine desulfurase activity of SufS may result from the binding of SufE actively remodeling the active site of SufS. [8]

SufE transfers persulfide to SufBC$_2$D complex, which is a scaffold complex assembling Fe-S clusters. [9-11] SufB, SufC, and SufD are all critical in this pathway since deletion of any of them leads to abolishment of Suf function. [12-14] SufB accepts sulfur transferred from SufE and provides a site for cluster assembly, and SufD is proposed to play a role in iron acquisition. [11] SufC is a novel type of ATP-binding cassette (ABC) ATPase, which also includes ABC

transporters and structural maintenance of chromosome (SMC) proteins. [15-18] In ABC transporters and SMC proteins, ABC ATPase activity drives conformational changes in partner domains required for each function. Therefore, it is likely that the ATPase activity of SufC may induce conformational changes of SufB and SufD subunits, providing a suitable conformation for cluster assembly. [16, 19-20] Moreover, SufC may provide energy for iron acquisition on SufD in the form of ATP hydrolysis. [21-22]

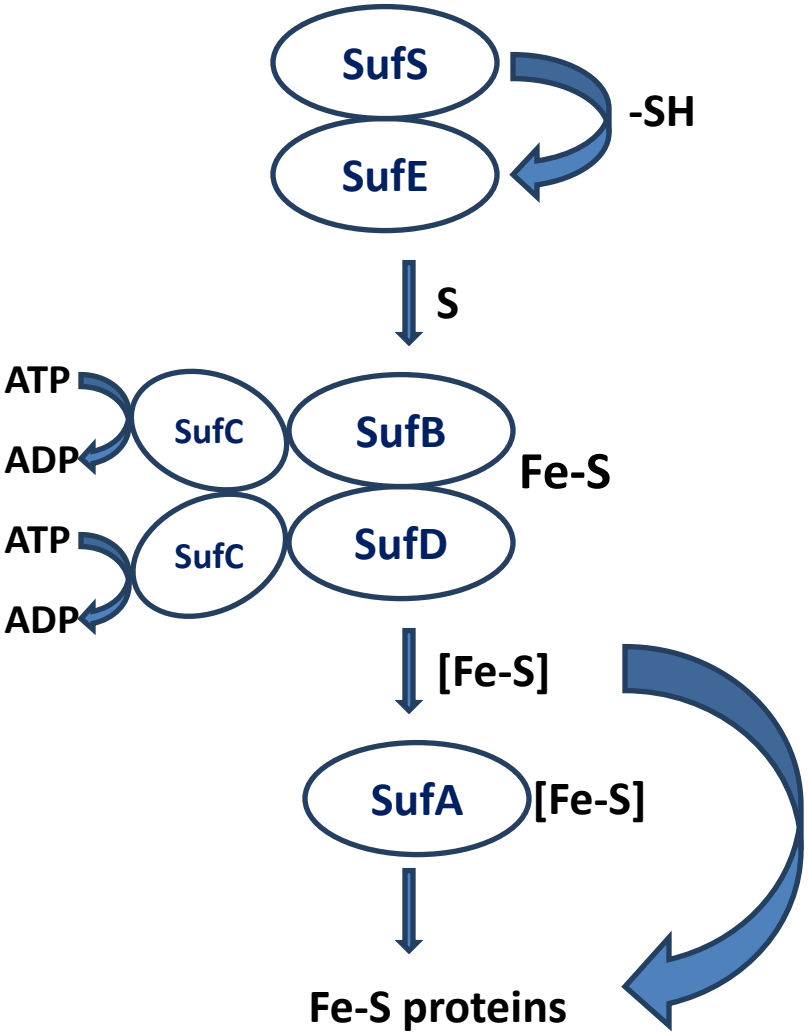


Illustration 4.1. Schematic of Suf pathway. The first step is to mobilize sulfur atoms by SufS-SufE complex. The second step is to assemble [Fe-S] clusters by $SufBC_2D$ scaffold complex. The last step is to transfer formed [Fe-S] clusters to target proteins directly or via SufA.

Within the $SufBC_2D$ complex, SufB, SufC, and SufD exhibit a 1:2:1 (B: C: D) stoichiometry. SufB and SufD are homologues and show limited sequence similarity (17% identity and 37% similarity). They show very similar structures and the mode of SufC binding are strikingly similar between SufB and SufD. [6, 19] The binding of SufB or SufD was reported to enhance the ATPase activity of SufC, since their binding leads to SufC having a suitable conformation for ATP binding and hydrolysis. [23-24] Furthermore, the addition of $SufBC_2D$ complex in *E. coli* can further enhance the cysteine desulfurase activity of SufS-SufE, since the transfer of the sulfur atom from SufE to SufB allows SufE to accept another sulfur atom from SufS. [6, 11, 25] The last protein SufA is a Fe-S carrier protein that transfers clusters to target proteins. It transfers assembled Fe-S clusters from $SufBC_2D$ complex to Fe-S proteins. [26] However, the Fe-S clusters can also be transferred to target proteins directly.

4.1.2 Crystal Structure of $SufBC_2D$ Complex

The $SufBC_2D$ complex is the scaffold for nascent Fe-S cluster assembly, playing a central role in the Suf machinery. SufB, SufC, and SufD can interact with each other, and their different states have been reported in a SufBC subcomplex, a SufCD subcomplex, and a SufBCD ternary complex. [21, 23, 27] Recently, the crystal structure of $SufBC_2D$ complex was determined by x-ray crystallography (Figure 4.1). [28] Each $SufBC_2D$ complex consists of one SufB subunit, two SufC subunits, and one SufD subunit. The SufB and SufD subunits form a heterodimer, where the interface is primarily composed of 25 hydrogen bonds. [28] Each of two SufC subunits binds to SufB and SufD, termed as $SufC_{SufB}$ and $SufC_{SufD}$, respectively. These two SufC subunits are facing each other but are spatially separated.

SufB and SufD have limited sequence similarity, but they have very similar structures. [28] Both of them have the same domain organization: a *N*-terminal helical domain, a core

domain composed of parallel β-strands arranged in a helical architecture, and a *C*-terminal helical domain (Figure 4.1). The *C*-terminal domains of SufB and SufD provide binding sites for two SufC subunits, and the core domains provide interface for SufB/SufD heterodimer. The interface is stabilized by hydrogen bonds.

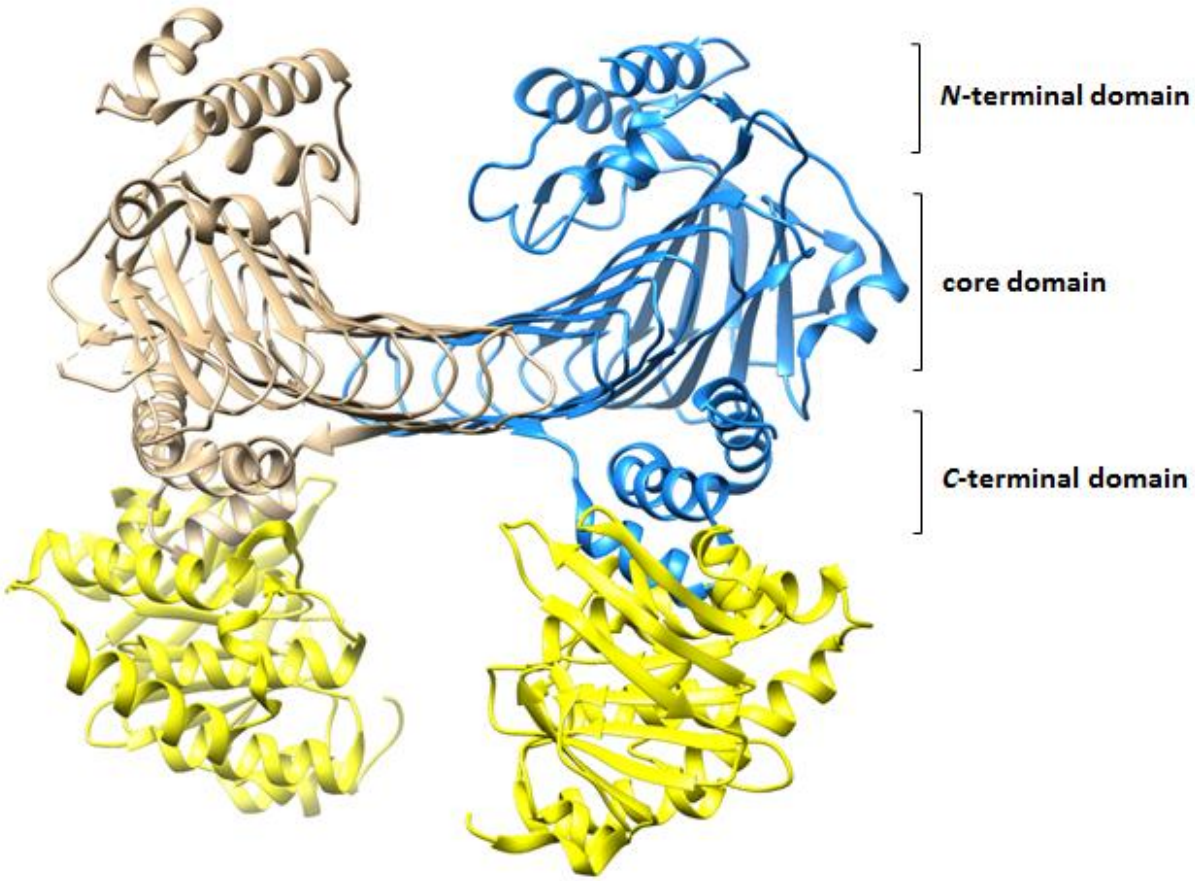


Figure 4.1. Overall crystal structure of *E. coli* $SufBC_2D$ complex. Individual subunits were shown in tan (SufB), blue (SufD), and yellow (SufC). Adapted with Chimera. PDB: 5awf

The SufC subunit has two domains: a catalytic α/β domain and a helical domain (Figure 4.2). [28] It contains highly conserved sets of amino acid residues including Walker A motif, Walker B motif, ABC signature motif, H motif, Q loop, and D loop. All of them are characteristics of ABC type ATPase. [19-20, 29] The Walker A motif was identified to be involved in nucleotide binding and the Walker B motif is likely to be involved in ATP

hydrolysis. [30-32] The Q loop connects the two domains of SufC. As discussed above, the SufB or SufD binding enhances the ATPase activity of SufC, supported by structural changes of SufC within the $SufBC_2D$ complex. [28] A unique salt bridge between Glu171 and Lys152 observed in SufC monomer is cleaved within the $SufBC_2D$ complex, and His203 shifts toward Glu171. [28, 30-31] So the binding of SufB/SufD drives SufC to rearrange the ATP binding site, leading to the enhancement of SufC ATPase activity.

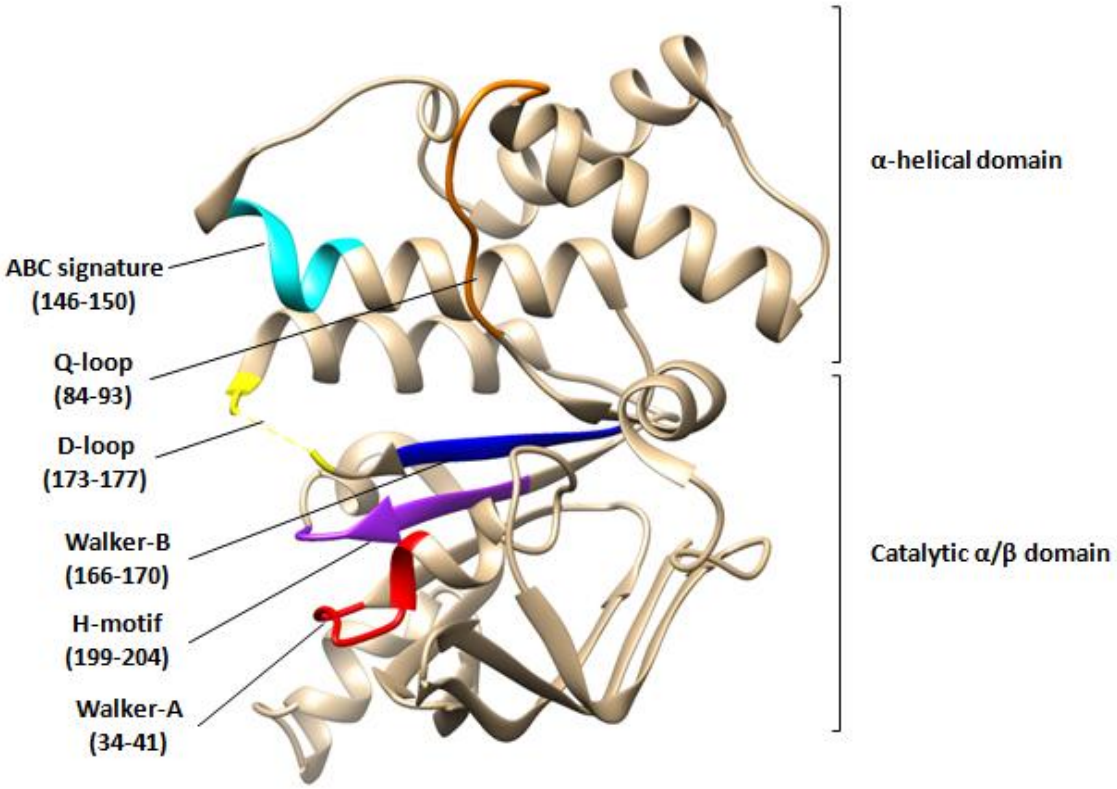


Figure 4.2. Overall crystal structure of *E. coli* SufC. It is divided into two domains: catalytic α/β domain and α-helical domain. The characteristics of ABC-type ATPase were labeled in SufC structure. Adapted with Chimera. PDB: 2d3w

4.1.3 SufC Dimer Model

According to the current consensus model, ABC ATPases can form a transient head to tail dimer in which two nucleotides are sandwiched at dimer interface between the Walker motifs of one subunit and the ABC signature of another subunit. [28] The structure of the SufC dimer is

still not determined, but disulfide cross-linking experiments support that SufC can form a transient dimer within the $SufBC_2D$ complex, in the presence of ATP and Mg^{2+}. [28] A SufC dimer model was generated by superimposing the structure of $SufC_{SufB}$ and $SufC_{SufD}$ onto the dimer structure of ATP-bound HlyB that is ABC ATPase and has very similar topology with SufC (Figure 4.3). [28] In the dimer model, the Walker A and Walker B motifs contained in catalytic α/β domain and ABC signature located at α-helical domain may provide the binding site for ATP.

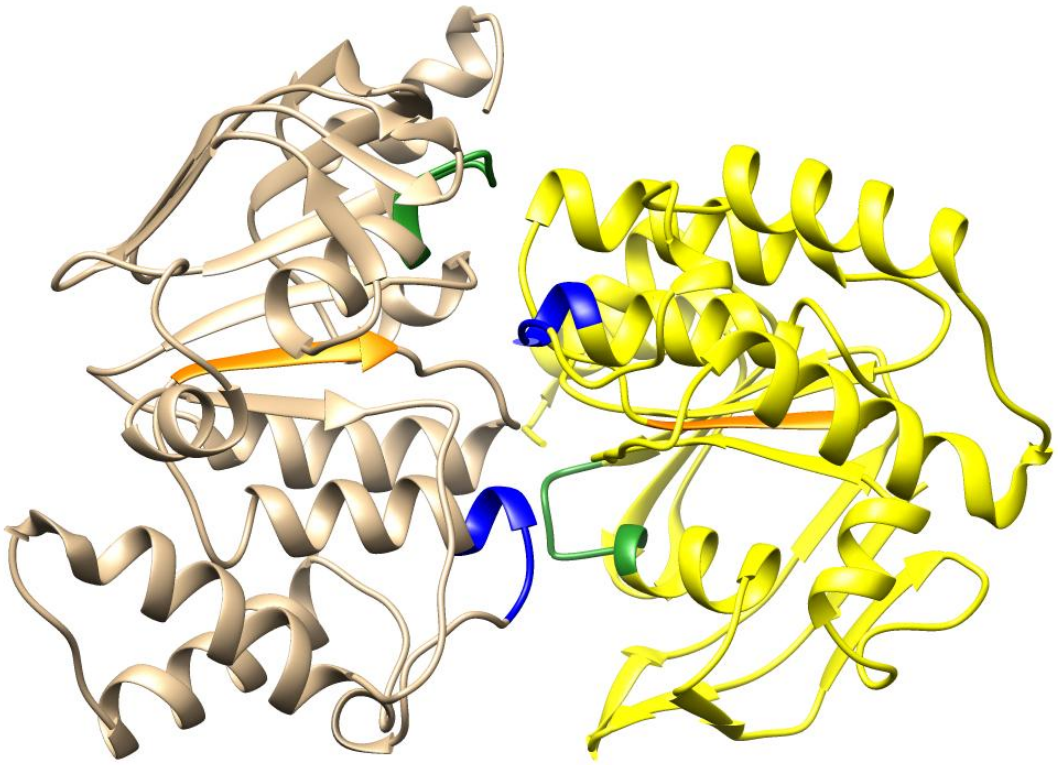


Figure 4.3. Generated dimer model of SufC. Walker A, Walker B, and ABC signature motifs were indicated in green, orange, and blue, respectively. Adapted with Chimera. PDB: 5awf and 1mt0.

4.1.4 Conclusion

As discussed above, the $SufBC_2D$ complex plays a crucial role in the Suf pathway, but the mechanism of Fe-S cluster assembly within the $SufBC_2D$ complex is still enigmatic. The

disulfide cross-linking experiments revealed that the two SufC subunits can form a transient dimer within the $SufBC_2D$ complex, in the presence of ATP/Mg^{2+}. The ATPase activity of SufC may drive the structural changes of SufB-SufD protomer, since fluorescent labeling experiments revealed that SufC dimer formation may lead to exposure of the heterodimer interface of the SufB-SufD protomer. [28] However, more studies are required to fully characterize the SufC function and the assembly mechanism of $SufBC_2D$ complex. The HDX-MS technique was applied to study the dynamic and conformational changes of SufC on ADP/ Mg^{2+} binding. The HDX-MS experiments can also provide insight into conformational changes of the SufB-SufD protomer due to the structural changes of SufC. Furthermore, the investigation of dynamics of $SufBC_2D$ complex may provide more evidence for mechanism of Fe-S cluster assembly.

4.2 Materials and Methods

4.2.1 Materials and Instruments

E. coli $SufBC_2D$ complex was provided by Dr. F. Wayne Outten, University of South Carolina. Deuterium oxide (D_2O), porcine pepsin (3200-4500 u/mg), and protease XIII from *Aspergillus saitoi* (≥ 0.6 u/mg) were purchased from Sigma-Aldrich (St. Louis, MO). Other reagents were purchased from VWR. All buffers were prepared using ultrapure water (18 MΩ at 25℃). An Agilent 1100 HPLC with a 1/16″ internal diameter tubing was connected to guard column and microbore C18 reverse phase column (Phenomenex) for peptides separation. A gas tight syringe (Hamilton) was utilized to load protein sample into an external injector (Rheodyne). A Bruker HCT Ultra PTM Discovery ion trap mass spectrometer was applied for peptides detection.

4.2.2 Sequencing of $SufBC_2D$ Complex with Tandem MS/MS

In order to improve the sequence coverage of peptides, the peptide maps of $SufBC_2D$

complex digested by pepsin and protease XIII were combined. The $SufBC_2D$ complex stock was 170 µM. It was diluted to 150 µM by 25 mM Tris (pH 7.5), 150 mM NaCl, and 10 mM BME. The sequencing of $SufBC_2D$ complex was initiated with the addition of 2 µl of 150 µM $SufBC_2D$ complex into 43 µl H_2O. Addition of 1M urea and 2 mM tris (2-carboxyethyl) phosphine (Tcep) solutions was applied to denature the protein and break disulfide bonds, respectively. The addition of urea and Tcep solutions enhanced protein digestion efficiency. Then 2 µl of 10 mg/ml pepsin was added to digest protein complex and the digestion lasted for 5 minutes. The generated peptides were loaded into HPLC and eluted by C18 reverse column with a linear gradient of acetonitrile solvent for 50 min (Table 4.1 A). The digestion of $SufBC_2D$ complex by protease XIII was almost identical to pepsin, except that 5 µl of 12 mg/ml protease XIII was applied for protein digestion, and the elution methods for peptides separation were different (Table 4.1 B). Finally, the separated peptides were detected by ESI ion trap mass spectrometer.

4.2.3 Continuous HDX-MS Experiments of $SufBC_2D$ Complex

After determining the peptide maps of $SufBC_2D$ complex, continuous HDX-MS experiments of free $SufBC_2D$ and ADP/Mg^{2+} bound $SufBC_2D$ complex were performed. The $SufBC_2D$ complex bound by ADP and Mg^{2+} was preformed by incubating 150 µM $SufBC_2D$ complex with 4 mM $MgCl_2$ and 1 mM ADP on ice for 30 min. A HDX experiment was initiated by incubating 2 µl of 150 µM $SufBC_2D$ or preformed $SufBC_2D$ complex in 23 µl of 99.9% pure D_2O at 25℃ for specific incubation time (15s, 30s, 45s, 60s, 5min, 10min, 30min, 60min, 2h, 4h). The H/D exchange reaction was quenched by adding 20 µl of quench solution (0.1 M potassium phosphate, pH 2.4) and incubating on ice. Then 5 µl of 10 M urea and 1 µl of 100 mM Tcep were added into reaction in order to enhance digestion efficiency. The protein complex was digested by 2 µl of 10 mg/ml pepsin or 5 µl of 12 mg/ml protease XIII for 5 min. The generated

peptides were loaded onto C18 reverse phase column equilibrated by buffer A (98% H_2O, 2% acetonitrile, and 0.4% formic acid) and separated by a linear gradient of buffer B (98% acetonitrile, 2% H_2O, and 0.4% formic acid) for 17min (Table 4.2).

Table 4.1. Elution method of tandem MS/MS for *E. coli* $SufBC_2D$ complex. (A) The *E. coli* $SufBC_2D$ complex was digested by pepsin; (B) The *E. coli* $SufBC_2D$ complex was digested by protease XIII.

A

Elution Time (min)	% Solvent A	% Solvent B
0.00	100	0
2.00	90	10
50.00	70	30
55.00	50	50
60.00	0	100
63.00	100	0
65.00	100	0

B

Elution Time (min)	% Solvent A	% Solvent B
0.00	100	0
40.00	50	50
41.00	0	100
51.00	0	100
52.00	100	0
60.00	100	0

Two control experiments corresponding to no deuteration ($m_{0\%}$) and fully deuteration ($m_{100\%}$) were also performed. The $m_{0\%}$ control was performed as the normal HDX experiment except that 2 µl of 150 µM $SufBC_2D$ complex was incubated in H_2O instead of D_2O. The $m_{100\%}$ control was performed in the same procedures except that the $SufBC_2D$ complex was incubated in D_2O at 25 °C for 12 hours to make sure all possible amide hydrogens exchanged with deuterium. These two controls were utilized to correct HDX data.

Table 4.2. Elution method of continuous HDX-MS experiments for *E. coli* $SufBC_2D$ complex.

Elution Time (min)	% Solvent A	% Solvent B
0.00	100	0
17.00	50	50
20.00	0	100
24.00	0	100
25.00	100	0
30.00	100	0

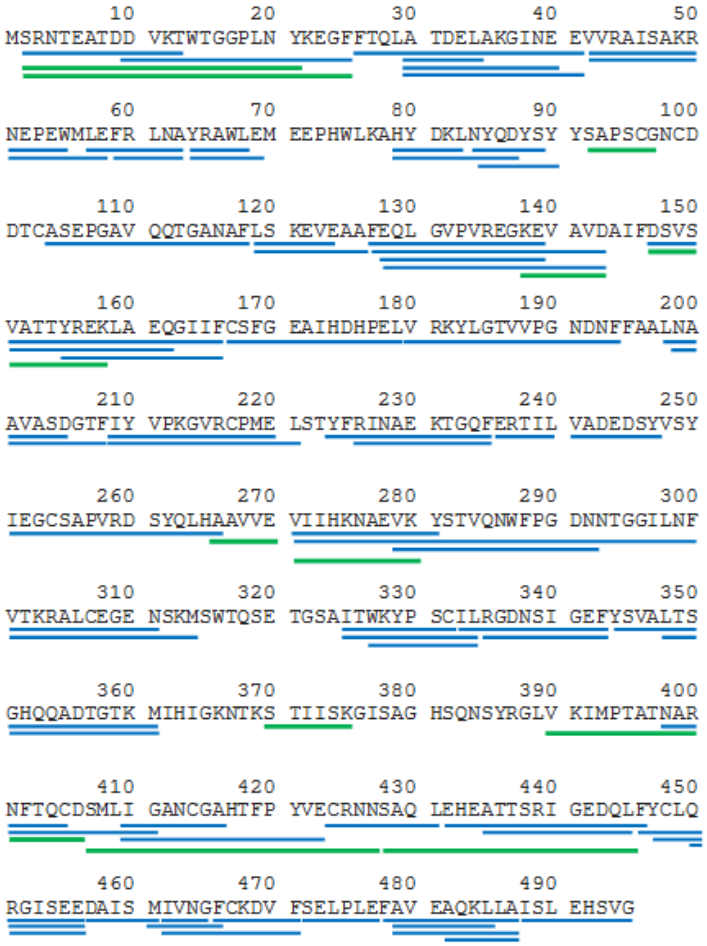
4.3 Result

4.3.1 Peptide Maps of $SufBC_2D$ Complex

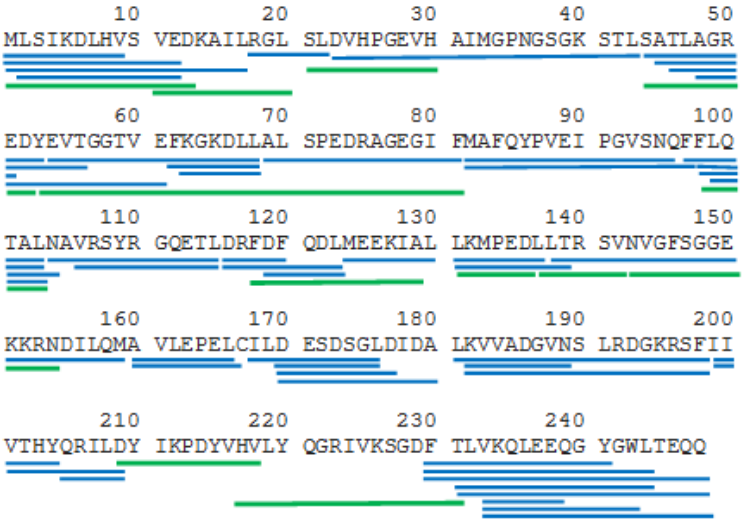
The $SufBC_2D$ complex was digested by both pepsin and protease XIII in order to obtain sufficient sequence coverage for SufB, SufC, and SufD. Tandem MS/MS was employed to detect digested peptides and Peaks Client software was utilized to determine sequences of detected peptides. Pepsin was the major acid protease to digest $SufBC_2D$ complex and protease XIII was used to find missing peptides. By combining peptides created by both proteases, the peptide maps of SufB, SufC, and SufD were determined (Figure 4.4). SufB, SufC, and SufD were

identified tohave sequence coverage of 87.8%, 100%, and 86.5%, respectively. The peptide maps of $SufBC_2D$ complex were employed in continuous HDX-MS experiments.

A



B



C

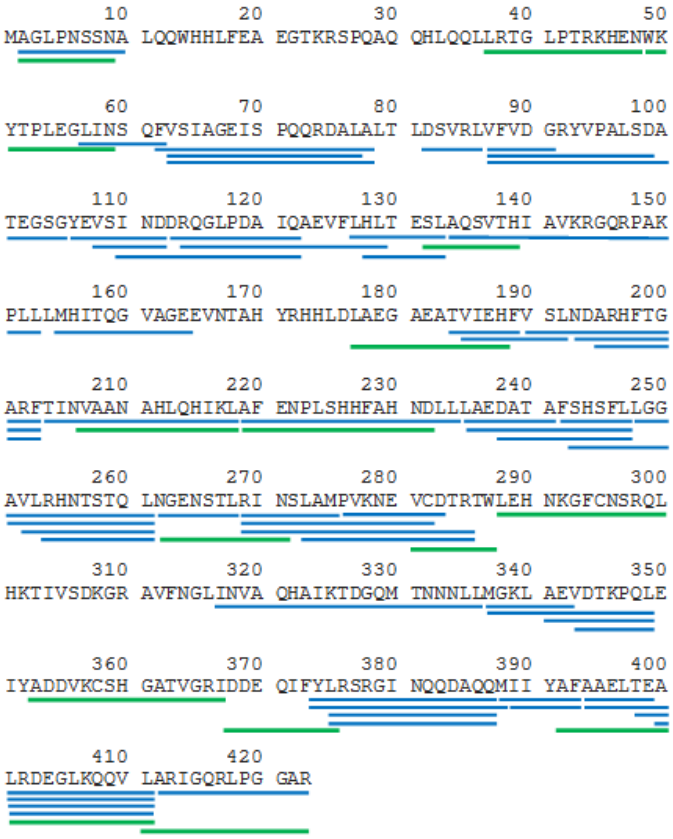
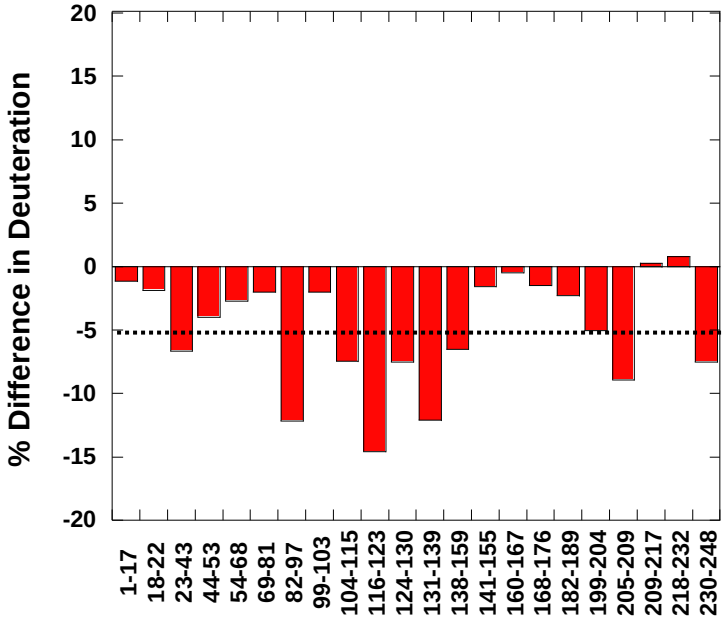


Figure 4.4. Peptide maps of (A) SufB, (B) SufC, and (C) SufD. Peptides digested by pepsin are underlined in blue, and peptides digested by protease XIII are underlined in green.

A



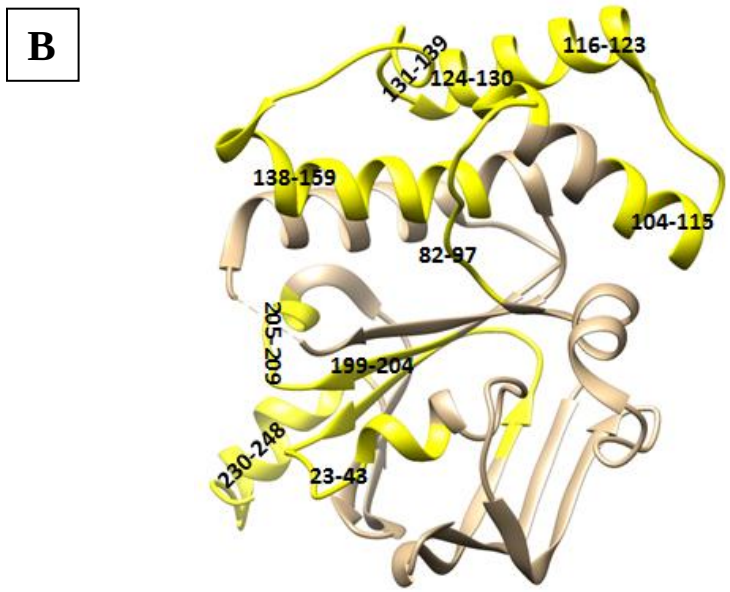


Figure 4.5. Solvent accessibility changes of SufC on ADP/Mg^{2+} binding. (A) Difference in solvent accessibility of free SufC relative to SufC bound by ADP/Mg^{2+}. (B) Peptides showing difference in deuteration ≥ 5% were labeled in crystal structure of SufC monomer. Made with Chimera. PDB: 2d3w

4.3.2 Solvent Accessibility Changes of SufC on ADP/Mg^{2+} Binding

The deuterium incorporation at 15s is used to estimate the relative solvent accessibility of proteins. So, comparison of free SufC and ADP/Mg^{2+}- bound SufC complex shows most peptides decrease in solvent accessibility due to the binding of ADP/Mg^{2+}, apart from peptides 209-217 and 218-232 showing minor increases (Figure 4.5 A). The peptides showing changes in solvent accessibility greater than 5% are labeled in structure of SufC monomer (Figure 4.5 B). Both domains show changes in solvent accessibility. In the catalytic α/β domain, peptides 23-43, 199-209, and 230-248 have a reduction in solvent accessibility due to ADP/Mg^{2+} binding. In the α-helical domain, the region form residues 104 to 159 show a decrease in solvent accessibility, supported by changes of peptides 104-115, 116-123, 124-130, 131-139, and 138-159. The three peptides 82-97, 116-123 and 131-139 have over 10% decrease in solvent accessibility, in which the peptide 82-97 contains the Q loop connecting two domains of SufC. Its significant decrease

in solvent accessibility indicates that SufC may have a conformational change on ADP/Mg^{2+} binding, which is also supported by the solvent accessibility changes of both domains. Moreover, the peptides 23-43, 82-97, 138-159, and 199-209 located at the interface of SufC dimer model show decreases in solvent accessibility, supporting the model of SufC dimer. The crystal structure of ATP-bound SufC complex is still not determined so the binding site of ATP is not identified. However, the Walker motifs of SufC are proposed to provide a binding site for ATP, since Walker A and Walker B motifs are observed to be involved in nucleotide binding in ABC ATPase family. This proposal is supported by changes to solvent accessibility in SufC. The peptide 23-43 containing the Walker A motif has a significant reduction in solvent accessibility due to the ADP/Mg^{2+} binding. Based on the SufC dimer model, the ABC signature contained in the α-helical domain may also contribute to the ADP/Mg^{2+}binding. This is supported by the decrease in solvent accessibility of the peptide 138-159. Therefore, the changes of solvent accessibility support the SufC dimer model that the Walker A motif and the ABC signature may provide a binding site for ADP/Mg^{2+}.

4.3.3 Dynamic Changes of SufC on ADP/Mg^{2+} Binding

By comparing deuterium incorporation curves of ADP/Mg^{2+}- bound SufC with free SufC, both domains of SufC were observed to have dynamic changes on ADP/Mg^{2+} binding. Figure 4.6 depicts the $SufBC_2D$ complex structure labeled with peptides that are located at the catalytic α/β domain and show dynamic changes. The deuterium incorporation curves of these peptides are shown in Figure 4.7. Peptide 1-17 shows a tiny decrease in exchange rate but about 40% decrease in amount of deuterium incorporated due to ADP/Mg^{2+} binding. The reduction in deuterium incorporation suggests that the peptide 1-17 may have a structural change leading to four backbone hydrogens being buried. This structural change may contribute to the binding of

ADP. The peptide 44-53 has a 7-fold decrease in exchange rate, suggesting its dynamics are decreased by ADP binding. Its decreasing dynamics may be due to its location close to the Walker A motif. Moreover, the region from residues 62 to 97 shows decreasing dynamics, supported by the dynamic changes of peptides 62-68, 69-81, and 82-97. The peptide 69- 81 is located at the interface between SufC and SufB/SufD protomer. It shows a 40% decrease in deuterium incorporation due to the binding of ADP/Mg^{2+}. The sharp decrease in exchange rate may result from the formation of hydrogen bonds between SufC and SufB-SufD protomer. The decrease in exchange rate of the peptide 69-81 suggests the interface between SufC and SufB-SufD protomer has decreasing dynamics, which is also supported by the dynamic changes of peptides from SufB/SufD promoters located at the interface (discussed in 4.3.4). The decreasing dynamics of the catalytic domain is also supported by the changes of peptide 82-97 containing the Q loop that connects two domains of SufC. This peptide shows not only a significant decrease in solvent accessibility but also a decrease in exchange rate by 2-fold. The dynamic decrease of the Q loop indicates the loss of conformational flexibility due to the binding of ADP/Mg^{2+}. Overall, the catalytic α/β domain shows decreasing dynamics on ADP/Mg^{2+} binding. Unexpectedly, the Walker motifs contributing to nucleotide binding do not show dynamic changes.

The α-helical domain of SufC also has changes in dynamics, supported by decreases in deuterium incorporation of peptides located at the α-helical domain upon ADP/Mg^{2+} binding (Figure 4.8 and 4.9). The region of residues 104-139 has decreasing dynamics, supported by dynamic changes of peptides 104-115, 116-123, 124-130, and 131-139. The three peptides 116-123, 124-130, and 131-139 show reductions in exchange rate by 3-fold, 3-fold, and 4-fold, respectively. The peptide 104-115 located at the interface between SufC and SufB/SufD

protomer, shows a decrease in exchange rate by 2-fold. The decreasing dynamics of the peptides 104-115 and 69-81 suggests that the interface between SufC and SufB/SufD indeed has lower dynamics due to the binding of ADP and Mg^{2+}. Moreover, the peptide 160-167 neighboring the Walker B motif shows a 20% decrease in deuterium incorporation. And peptides 182-189 and 199-204 show decreases in exchange rate by 2-fold and 10-fold, respectively. The dynamic changes of these peptides point out the possibility that the α-helical domain of SufC has conformational changes upon ADP/Mg^{2+} binding. Therefore, the SufC is concluded to undergo a conformational change within $SufBC_2D$ complex when ADP and Mg^{2+} bind to it, leading to the interface between SufC and SufB/SufD protomer having decreasing dynamics.

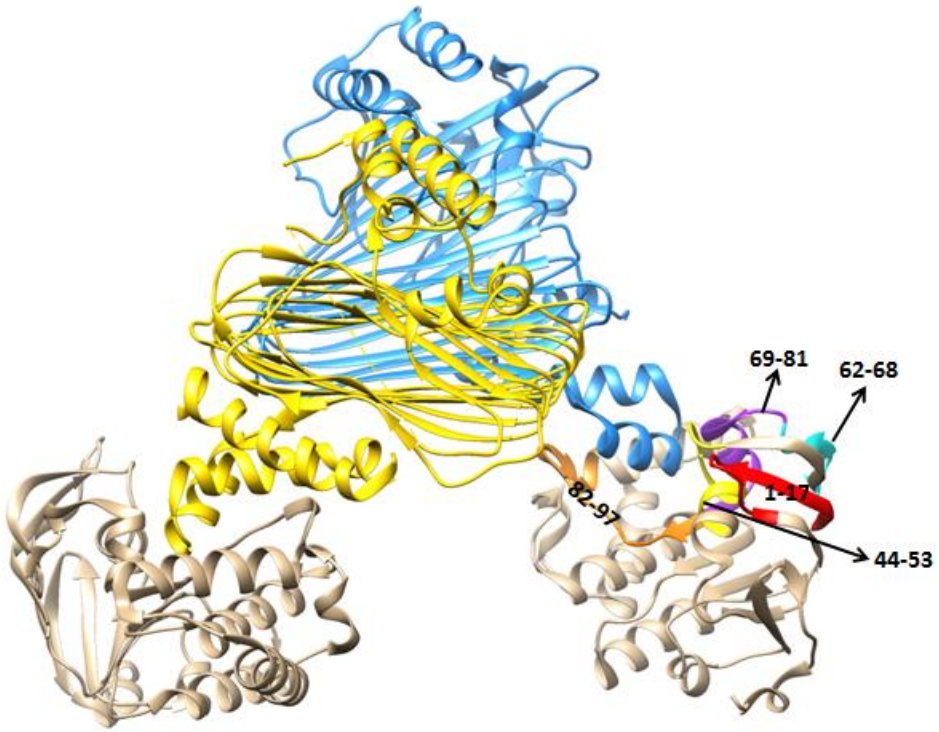


Figure 4.6. Dynamic changes of catalytic α/β domain of SufC on ADP/Mg^{2+} binding. Peptides from catalytic α/β domain of SufC showing dynamic changes were labeled in $SufBC_2D$ complex structure. The SufB, SufC, and SufD subunits were indicated in gold, tan, and blue respectively. Adapted with Chimera. PDB: 5awf

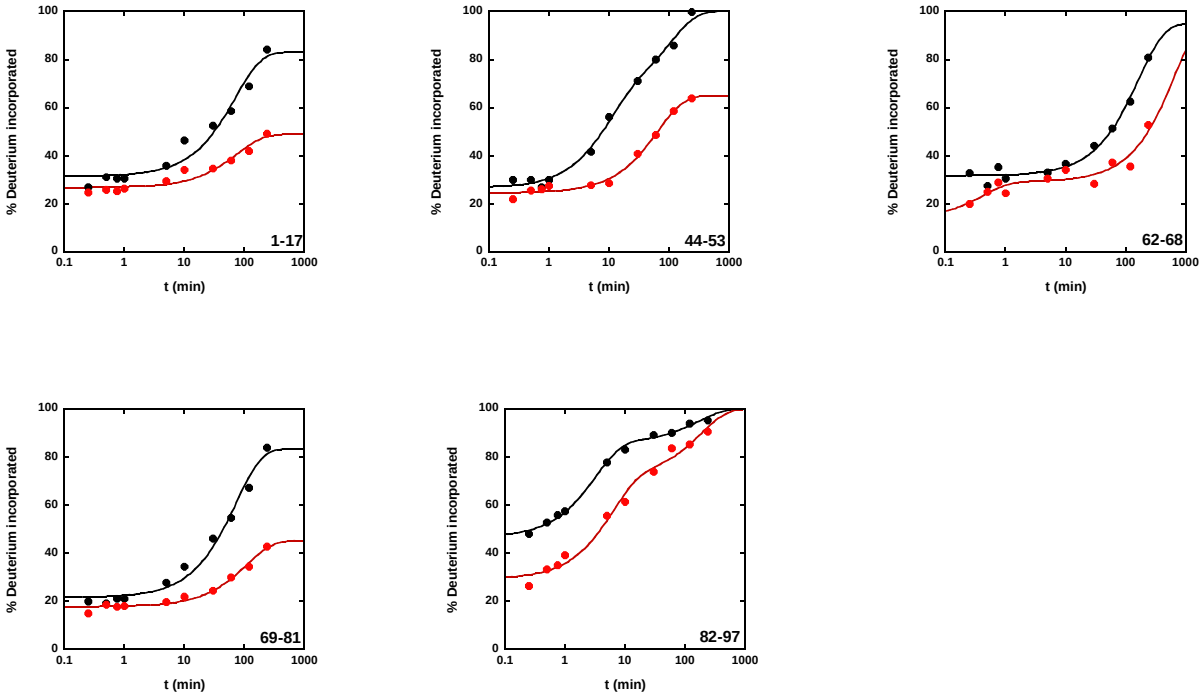


Figure 4.7. Deuterium incorporation curves of peptides located at catalytic α/β domain of SufC. The black and red curves represent percentage of deuterium incorporation of SufC peptides in the absence and presence of ADP/Mg^{2+}, respectively. Made with KaleidaGraph.

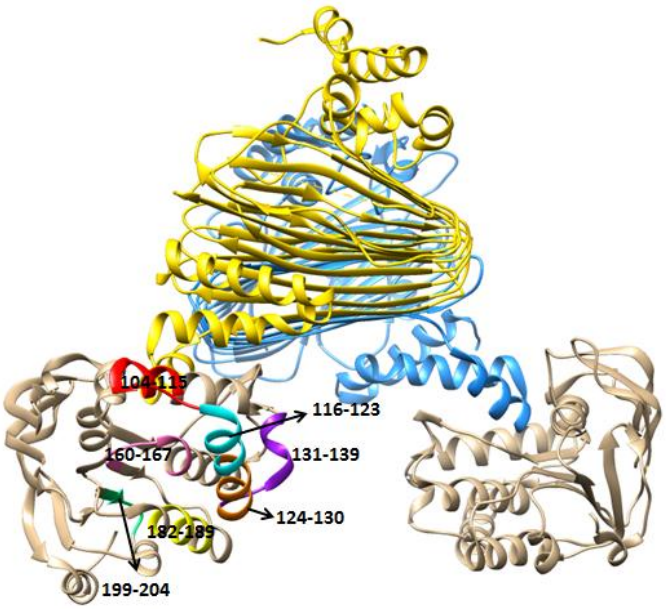


Figure 4.8. Dynamic changes of α-helical domain of SufC on ADP/Mg^{2+} binding. Peptides of α-helical domain showing dynamic changes were labeled in $SufBC_2D$ complex structure. The SufB, SufC, and SufD subunits were indicated in gold, tan, and blue respectively. Adapted with Chimera. PDB: 5awf

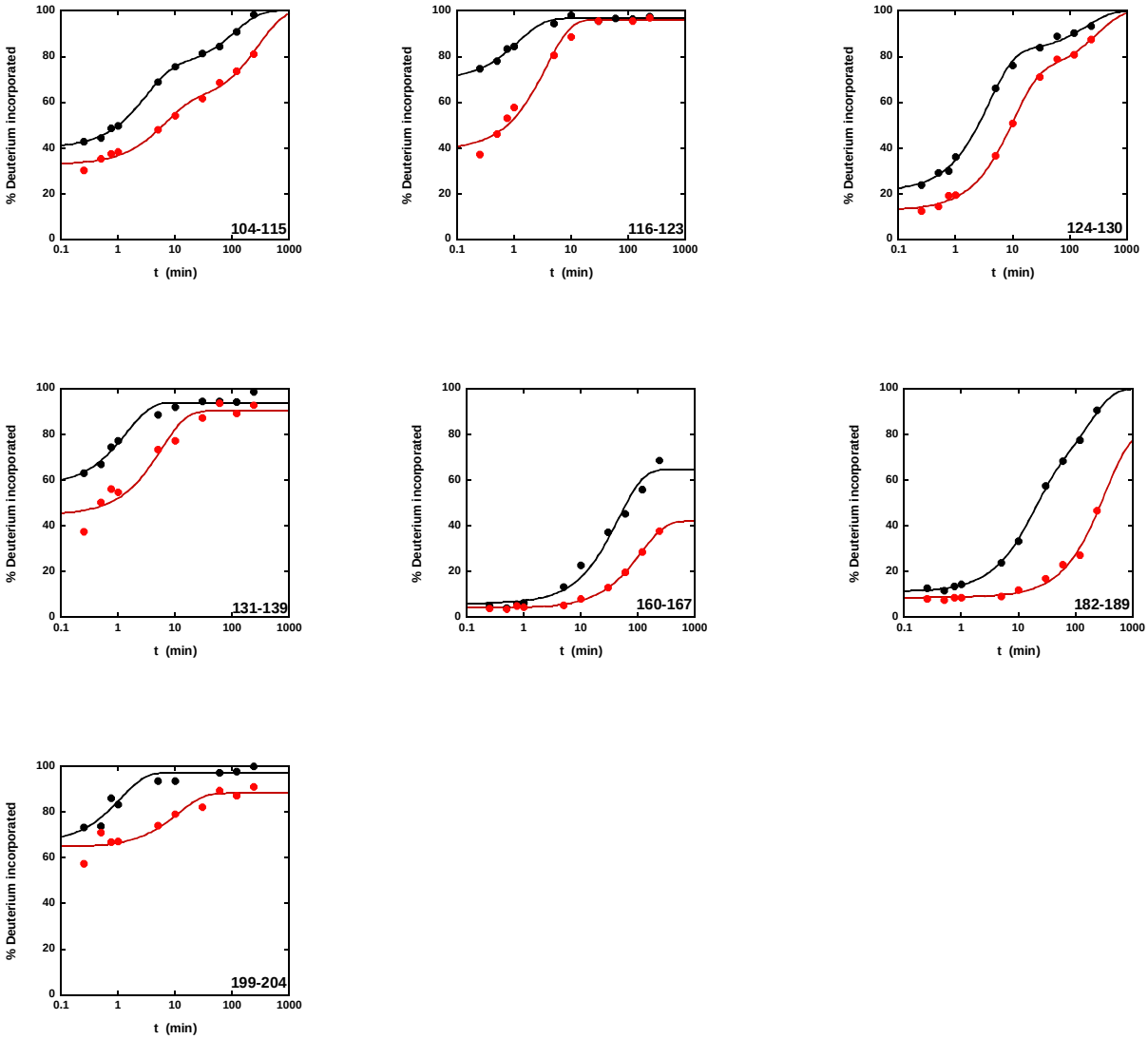


Figure 4.9. Deuterium incorporation curves of peptides located at α-helical domain of SufC. The black and red curves represent percentage of deuterium incorporation of SufC peptides in the absence and presence of ADP/Mg^{2+}, respectively. Made with KaleidaGraph.

4.3.4 Dynamic Changes of SufB/SufD Protomer on ADP/Mg^{2+} Binding

The HDX-MS data shows that the ADP/Mg^{2+} binding induces conformational changes of SufC, resulting in the decreasing dynamics of the interface between SufC and SufB/SufD protomer. So, the SufB/SufD protomer is expected to have conformational changes due to the SufC motions. This speculation is supported by the dynamic changes of SufB/SufD protomer

(Figure 4.10). Deuterium incorporation curves of peptides from SufB and SufD are shown in Figure 4.11 and Figure 4.12, respectively.

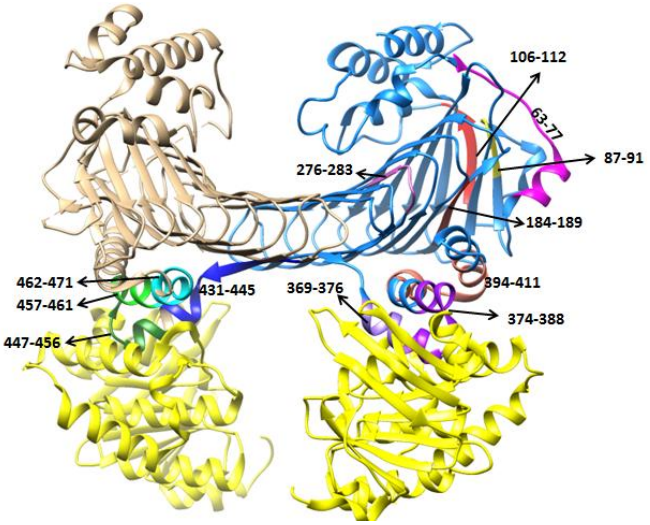


Figure 4.10. Dynamic changes of SufB/SufD protomer. Peptides of SufB/SufD subunits showing dynamic changes were labeled in $SufBC_2D$ complex structure. The SufB, SufC, and SufD were indicated in tan, yellow, and blue respectively. Adapted with Chimera. PDB: 5awf

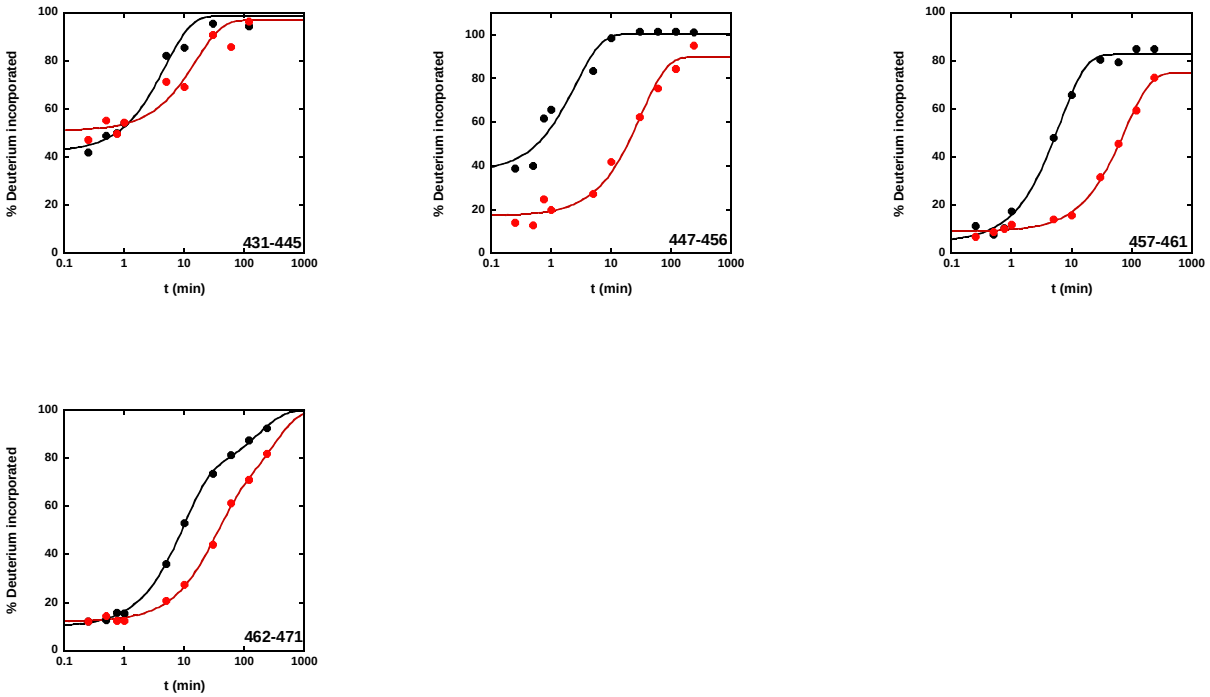


Figure 4.11. Deuterium incorporation curves of SufB peptides. The black and red curves represent percent of deuterium incorporation of SufB in the absence and presence of ADP/Mg^{2+}, respectively. Made with KaleidaGraph.

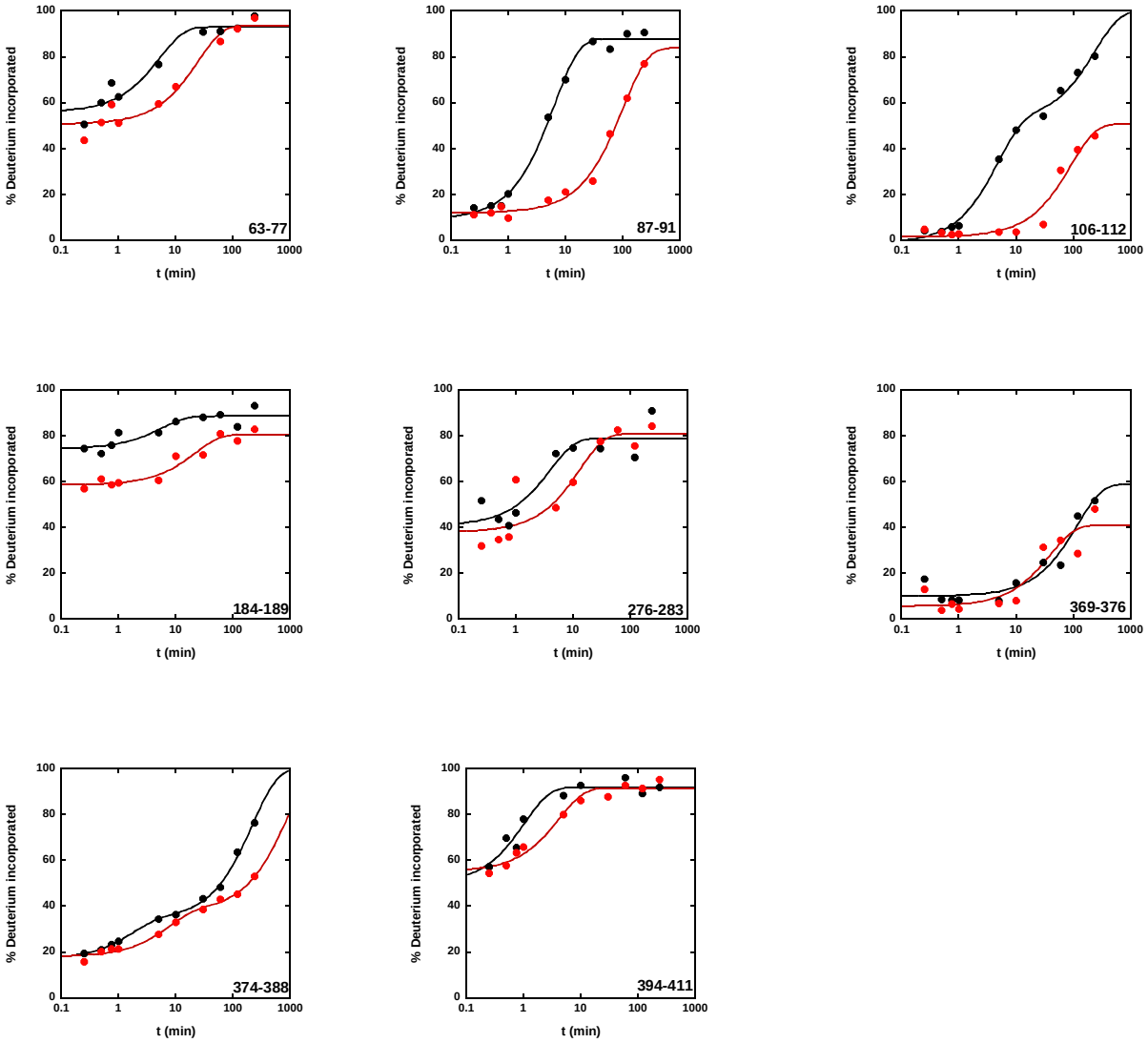


Figure 4.12. Deuterium incorporation curves of SufD peptides. The black and red curves represent percent of deuterium incorporation of SufD in the absence and presence of ADP/Mg^{2+}, respectively. Made with KaleidaGraph.

In SufB subunit, peptides 431-445, 447-456, 457-461, and 462-471, located at the interface between SufB and SufC, display significant decreases in exchange rate by 3-fold, 12-fold, 12-fold, and 3-fold, respectively. This is consistent with the conclusion that the interface between SufB and SufC has decreases in dynamics due to the ADP/Mg^{2+} binding. Compared with the SufB subunit, SufD has more peptides showing dynamic changes. In SufD subunit, the

peptides located at the interface between SufD and SufC also show decreases in dynamics. Peptide 369-376 has a 20% reduction in deuterium incorporation and peptide 374-388 has a 4-fold decrease in exchange rate. Moreover, peptide 394-411 shows a 4-fold decrease in exchange rate, supporting that the interface between SufC and SufD has decreasing dynamics on ADP/Mg^{2+} binding. Furthermore, several peptides far away from the interface with SufC also show dynamic changes. Peptide 63-77 containing a loop shows 5-fold decrease in exchange rate, indicating the loss of loop flexibility. Peptides 87-91, 106-112, 184-189, and 276-283 are located at the core domain of SufD. All of them show dynamic changes upon ADP/Mg^{2+} binding. The peptide 106-112 has 50% reduction in deuterium incorporation, and the peptides 87-91, 184-189, and 276-283 show decreases in exchange rate by 16-fold, 4-fold, and 3-fold, respectively. The dynamic changes of these peptides indicate that some conformational changes may happen on SufD in order to provide a suitable conformation for Fe-S cluster assembly or iron acquisition. Overall, the SufB/SufD protomer does not show as large a conformational change as SufC subunit. It has decreasing dynamics in the interface with SufC and shows dynamic changes in core domains, which may play a role in cluster assembly.

4.4 Discussion

4.4.1 HDX-MS Data cannot Support the Proposed SufC Dimer

E. coli SufC has very similar overall topologies with *E. coli* HlyB, a typical ABC-type ATPase. Based on dimer structure of HlyB bound by ATP, the SufC dimer model was proposed and built (Figure 4.3). [28] As discussed above, in ABC-type ATPase family, Walker A and Walker B motifs contained in catalytic α/β domain are involved in nucleotide binding. This point is strengthened by changes in solvent accessibility of SufC peptides. The peptide 23-43 contains Walker A motif (residues 34 – 41) and has reduction in solvent accessibility, suggesting it may

be involved in ADP binding. The peptide 138-159 containing ABC signature (residues 146-150) shows 6.5% decrease in solvent accessibility. The decreases in solvent accessibility of these two peptides support that the Walker A motif and ABC signature of SufC are likely to play an important role in nucleotide binding. This seems to be consistent with the SufC dimer model that the two nucleotides are sandwiched at the dimer interface between Walker motifs of one monomer and ABC signature of another monomer. [28] The peptide 82-97 shows over 10% decrease in solvent accessibility, confirming the important role of Q-loop (residues 84 – 93) in ADP binding. Overall, the significant decreases in solvent accessibility of Walker A, ABC signature, and Q loop show that the residues of these characteristics may be indispensable for ATPase activity of SufC.

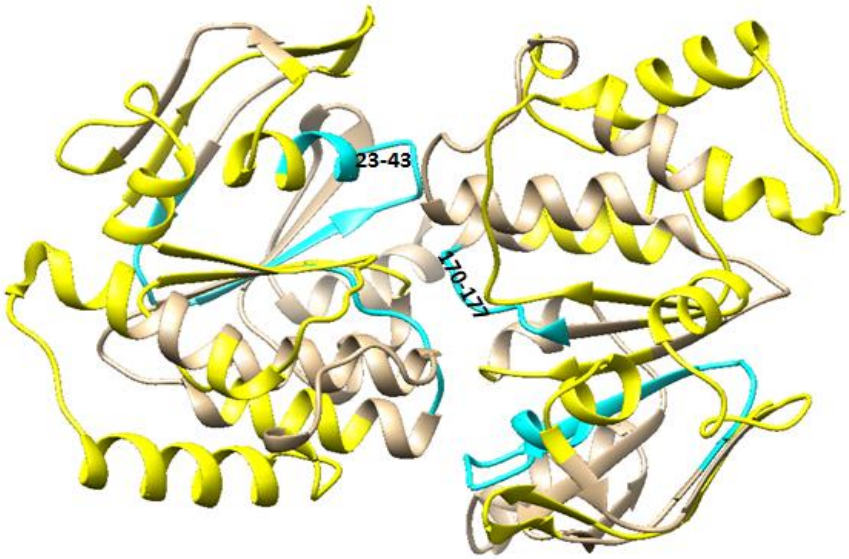


Figure 4.13. SufC dimer model labeled with peptides showing dynamic changes and indicated by yellow. Adapted with Chimera. PDB: 5awf and 1mt0.

Previous studies revealed that the SufC monomers can form a transient dimer in the presence of ATP and Mg^{2+}. However, the HDX data does not correlate with a formation of SufC dimer of this type by observing dynamic changes of peptides from SufC. All the peptides

showing dynamic changes are labeled in SufC dimer model (Figure 4.13). The peptides located at both domains are detected to have dynamic changes, indicating that the two SufC subunits may have conformational changes on ADP/Mg^{2+} binding. But, it cannot support that these dynamic changes induce the formation of SufC dimer, since most peptides proposed to be located at the dimer interface do not show significant dynamic changes (Figure 4.14). The Walker A and ABC signature located at the dimer interface are supposed to provide binding site for nucleotide, but neither of them show changes in dynamics even though they have significant decreases in solvent accessibility due to ADP/Mg^{2+} binding. Therefore, the HDX-MS data cannot provide strong evidence for the formation of this type SufC dimer within $SufBC_2D$ complex in the presence of ADP and Mg^{2+}. It is possible that ADP cannot induce the dimer formation as a product, since ATP and ADP have different structures and energies.

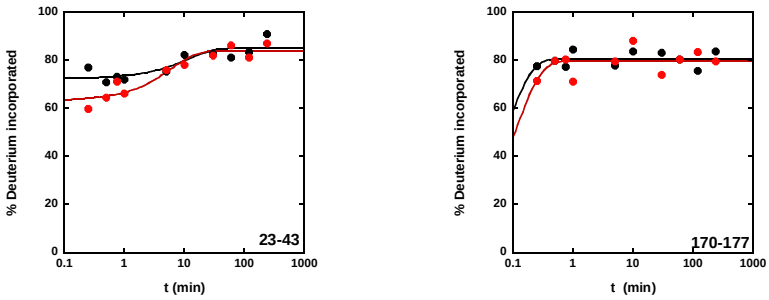


Figure 4.14. Deuterium incorporation curves of peptides from SufC dimer interface. The black and red curves represent percent of deuterium incorporation of SufD in the absence and presence of ADP/Mg^{2+}, respectively. Made with KaleidaGraph.

4.4.2 Insight into Dynamic Changes of SufB/SufD Protomer

In 2015, Hirabayashi *et al.* detected that SufC with ATP/Mg^{2+} induced exposure of hydrophobic regions of SufB/SufD subunits by using fluorescent labeling experiments. [28] The Cys405 of SufB, which is located at the heterodimer interface between SufB and SufD subunit, is a potential Fe-S cluster assembly site. His360 of SufD is another candidate for the cluster

coordination residue. [28] They reported that the residue Cys405 was exposed on ATP binding, which was buried inside the dimer interface. Moreover, His360 of SufD could also be exposed due to SufC dimerization. [28] However, the HDX-MS data cannot provide evidence for this point. The residues Cys405 of SufB and His360 of SufD are contained in peptides 401-411 of SufB and 358-368 of SufD, respectively. Both of these peptides do not show any significant changes in solvent accessibility and dynamics. Based on the HDX data, these two residues are not supported to be potential sites for Fe-S cluster assembly and exposed due to the SufC conformational changes. However, the HDX data shows the dynamic changes happening on other peptides that are located at the interface between SufC and SufB/SufD protomer and involved in β-sheet core domains of SufB/SufD protomer (Figure 4.10). The changes of these peptides indicate that the SufB/SufD protomer has conformational changes due to ADP/Mg^{2+} binding, which may provide an active conformation for iron acquisition and Fe-S cluster assembly. The peptide 276-283 from SufD contains Cys282 that is able to interact with Fe-S cluster. The dynamic change of the peptide 276-283 points out the possibility that the residue Cys282 may be a potential Fe-S cluster assembly site. The peptides 63-77, 87-91, 106-112, and 184-189 from SufD show decreasing dynamics, indicating the SufD has a regional conformational change. This change may contribute to the iron acquisition and cluster assembly under oxidative stress.

4.5 References

[1] Mihara, H., Kurihara, T., Yoshimura, T., Soda, K., and Esaki, N. Cysteine sulfinate desulfinase, a NIFS-like protein of *Escherichia coli* with selenocysteine lyase and cysteine desulfurase activities. Gene cloning, purification, and characterization of a novel pyridoxal enzyme. *J Biol Chem.* 1999;274:22417-22424.

[2] Mihara, H., Maeda, M., Fujii, T., Kurihara, T., Hata, Y., and Esaki, N. A nifS-like gene, csdB, encodes an *Escherichia coli* counterpart of mammalian selenocysteine lyase. Gene cloning, purification, characterization and preliminary x-ray crystallographic studies. *J Biol Chem.* 1999;274:14768-14772.

[3] Mihara, H., Kurihara, T., Yoshimura, T., and Esaki, N. Kinetic and mutational studies of three NifS homologs from *Escherichia coli*: mechanistic difference between L-cysteine desulfurase and L- selenocysteine lyase reactions. *J Biochem.* (Tokyo) 2000;127:559-567.

[4] Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M., and Barras F. Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase. *J Biol Chem.* 2003;278:38352-38359.

[5] Fontecave, M., de Choudens, S. O., and Barras, B. Py. F. Mechanism of iron-sulfur cluster assembly: the SUF machinery. *J Biol Inorg Chem.* 2005;10:713-721.

[6] Outten, F. W., Wood, M. J., Munoz, F. M., and Storz, G. The SufE protein and SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in *Escherichia coli*. *J Biol Chem.* 2003;278:45713-45719.

[7] Ollagnier-de-Choudens, S., Lascoux, D., Loiseau, L., Barras, F., Forest, E., and Fontecave, M. Mechanistic studies of the SufS-SufE cysteine desulfurase: evidence for sulfur transfer from SufS to SufE. *FEBS Lett.* 2003;555:263-267.

[8] Singh, H., Dai, Y., Outten, F. W., and Busenlehner, L. S. *Escherichia coli* SufE sulfur transfer protein modulates the SufS cysteine desulfurase through allosteric conformational dynamics. *J Biol Chem.* 2013;288:36189-36200.

[9] Wollers, S., Layer, G., Garcia-Serres, R., Signor, L., Clemancey, M., Latour, J. M., Fontecave, M., and Ollagnier-de-Choudens, S. Iron-sulfur (Fe-S) cluster assembly: the SufBCD complex is a new type of Fe-S scaffold with a flavin redox cofactor. *J Biol Chem.* 2010;285:23331-23341.

[10] Chahal, H. K., Dai, Y., Saini, A., Ayala-Castro, C., and Outten, F. W. The SufBCD Fe-S scaffold complex interacts with SufA for Fe-S cluster transfer. *Biochemistry.* 2009;48:10644-10653.

[11] Layer, G., Gaddam, S. A., Ayala-Castro, C. N., Ollagnier-de-Choudens, S., Lascoux, D., Fontecave, M., and Outten, F. W. SufE transfers sulfur from SufS to SufB for iron-sulfur cluster assembly. *J Biol Chem.* 2007;282:13342-13350.

[12] Outten, F. W., Djaman, O., and Storz, G. A *suf* operon requirement for Fe-S cluster assembly during iron starvation in *Escherichia coli*. *Mol Microbiol.* 2004;25:861-872.

[13] Tokumoto, U., Kitamura, S., Fukuyama, K., and Takahashi, Y. Interchangeability and distinct properties of bacteria Fe-S cluster assembly systems: functional replacement of the *isc* and *suf* operons in *Escherichia coli* with the *nifSU*-like operon from *Helicobacter pylori*. *J BioChem.* 2004;136:199-209.

[14] Outten, F. W., Gjaman, O., and Storz, G. A *suf* operon requirement for Fe-S cluster assembly during iron starvation in *Escherichia coli*. *Mol Microbiol.* 2004;52:861-872.
[15] Outten, W. O. Recent advances in the Suf Fe-S cluster biogenesis pathway: beyond the proteobacteria. *Biochim Biophys Acta.* 2014;1853:1464-1469.

[16] Hirano, T. SMC proteins and chromosome mechanics: from bacteria to humans. *Philos Trans R Soc Lond B Biol Sci.* 2005;360:507-514.

[17] Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P., and Tainer, J. A. Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell.* 2000;101:789-800.

[18] Graumann, P. L. SMC proteins in bacteria: condensation motors for chromosome segregation. *Biochimie.* 2001;83:53-59.

[19] Nachin, L., Loiseau, L., Expert, D., and Barras, F. SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. *Embo J.* 2003;22:427-437.

[20] Rangachari, K., Davis, C. T., Eccleston, J. F., Saldanha, J. W., Strath, M., and Wilson, R. J. SufC hydrolyzes ATP and interacts with SufB from *Thermotoga maritime*. *FEBS Lett.* 2002;514:225-228.

[21] Saini, A., Mapolelo, D. T., Chahal, H. K., Johnson, M. K., and Outten, F. W. SufD and SufC ATPase activity are required for iron acquisition during in vivo Fe-S cluster formation on SufB. *Biochemistry.* 2010;49:9402-9412.

[22] Xu, X. M., and Moller, S. G. AtNAP7 is a plastidic SufC-like ATP-binding cassette/ATPase essential for Arabidopsis embryogenesis. *Proc Natl Acad Sci U. S. A.* 2004;101:9143-9148.

[23] Petrovic, A., Davis, C. T., Rangachari, K., Clough, B., Wilson, R. J., and Eccleston, J F. Hydrodynamic characterization of the SufBC and SufCD complexes and their interaction with fluorescent adenosine nucleotides. *Protein Sci.* 2008;17:1264-1274.

[24] Eccleston, J F., Petrovic, A., Davis, C. T., Rangachari, K., and Wilson, R. J. The kinetic mechanism of the SufC ATPase: the cleavage step is accelerated by SufB. *J Biol Chem.* 2006;281:8371-8378.

[25] Selbach, B. P., pradhan, P. K., and Dos Santos, P. C. Protected sulfur transfer reactions by the *Escherichia coli* Suf system. *Biochemistry.* 2013;52:4089-4096.

[26] Gupta, V., Sendra, M., Naik, S. G., Chahal, H. K., Huynh, B. H., Outten, F. W., Fontecave, M., and Ollagnier-de-Choudens, S. Native *Escherichia coli* SufA, coexpressed with SufBCDSE, purified as a [2Fe-2S] protein and acts as an Fe-S transporter to Fe-s target enzymes. *J Am Chem Soc.* 2009;131:6149-6153.

[27] Wada, K., Sumi, N., Nagai, R., Iwasaki, K., Sato, T., Suzuku, K., Hasegawa, Y., Kitaoka, S., Minami, Y., Outten, F. W., Takahashi, Y., and Fukuyama, K. Molecular dynamism of Fe-S cluster biosynthesis implicated by the structure of the $SufC_2$-$SufD_2$ complex. *J Mol Biol.* 2009;387:245-258.

[28] Hirabayashi, K., Yuda, E., Tanaka, N., Katayama, S., Iwasaki, K., Matsumoto, T., Kurisu, G., Outten, F. W., Fukuyama, K., Takahashi, Y., and Wada, K. Functional dynamics revealed by the structure of the SufBCD complex, a novel ATP-binding cassette (ABC) protein that serves as a scaffold for iron-sulfur cluster biogenesis. *J Biol Chem.* 2015;290:29717-29731.

[29] Shintaro, K., Wada, K., Hasegawa, Y., Minami, Y., Fukuyama, K., and Takahashi, Y. Crystal structure of *Escherichia coli* SufC, an ABC-type ATPase component of the SUF iron-sulfur cluster assembly machinery. *FEBS Letters.* 2006;580:137-143.

[30] Smith, P. C., Karpowich, N., Millen, L., Moody, J. E., Rosen, J., Thomas, P. J., and Hunt, J. F. ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell.* 2002;10:139-149.

[31] Zaitseva, J., Jenewein, S., Jumpertz, T., Holland, I. B., and Schmitt, L. H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J.* 2005;24:1901-1910.

[32] Davidson, A. L., and Chen, J. ATP-binding cassette transporters in bacteria. *Annu Rev Biochem.* 2004;73:241-268.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 Summary

Apart from the structure of a protein, dynamics also contribute to protein function. There exists an intimate relationship between protein dynamics and protein function, which has been widely accepted. [1] So, the study of protein dynamics is useful and necessary for understanding protein function. Several techniques have been developed to study protein dynamics, including NMR spectroscopy, EPR spectroscopy, and MS. [2-4] Compared with NMR and EPR spectroscopies, MS has less stringent sample requirements, including protein concentration and protein size. Moreover, the mass accuracy, sensitivity, and faster data analysis also contribute to the rapid growth of MS. [4] HDX-MS, a combination of HPLC and MS, has become a common and sensitive tool to probe protein structural flexibility and solution dynamics. [5-6] HDX-MS can be applied in study of protein structure and dynamics, studying protein-ligand and protein-protein interactions, protein folding, and protein therapeutics discovery and development. [7] In this dissertation, HDX-MS was applied to study dynamic and conformational changes of glycosyltransferases and $SufBC_2D$ complex upon substrate binding.

Chapter 2 focused on exploring the dynamic changes of *Mt*GpgS on ligand binding. *Mt*GpgS is the retaining GT-A glycosyltransferase catalyzing the first step in the MGLP biosynthesis. [8] The crystal structure of *Mt*GpgS and complex bound by substrates have been determined, however, little dynamic and conformational changes on substrate binding were detected. [9] In order to identify conformational changes of *Mt*GpgS on UDPG and 3PGA

binding, HDX-MS was performed and deuterium incorporation curves were compared between free *Mt*GpgS and substrate-bound complexes. Differences in solvent accessibility between *Mt*GpgS APO and *Mt*GpgS bound by UDPG or 3PGA revealed that both substrates UDPG and 3PGA could bind to *Mt*GpgS independently and 3PGA could bind to or allosterically affect the UDPG binding site. This conclusion was consistent with the substrate inhibition mechanism of *Mt*GpgS by 3PGA versus UDPG. [10] Moreover, the deuterium incorporation curves identified that *Mt*GpgS might provide a necessary conformation for UDPG binding. However, *Mt*GpgS showed dynamic changes in the substrate binding site and the dimer interface upon 3PGA binding.

Chapter 3 utilized HDX-MS to investigate conformational changes of *Cg*MshA on ligand binding. *Cg*MshA is a retaining GT-B glycosyltransferase catalyzing the initial step of MSH biosynthesis. [11] A large conformational change was observed in *Cg*MshA on nucleotide binding by superimposing APO structure of *Cg*MshA and complex structure with UDP. [12] HDX-MS was utilized to investigate conformational changes of *Cg*MshA on substrate binding on the aspect of dynamics, providing a complementary to static structures. Solvent accessibility changes of *Cg*MshA on substrate binding revealed that both substrates UDP-GlcNAc and I1P could bind to *Cg*MshA independently, disagreeing with the previous proposal that UDP-GlcNAc bound to *Cg*MshA first followed by I1P. Moreover, the HDX-MS data suggested that only the substrate UDP-GlcNAc could induce large conformational changes of *Cg*MshA even though the 3PGA binding could also lead to some dynamic changes in *Cg*MshA.

Chapter 2 and chapter 3 focused on investigation of dynamic changes of proteins due to ligand binding, so chapter 4 applied HDX-MS to study dynamic changes of the protein complex $SufBC_2D$ due to ligand binding and PPIs. $SufBC_2D$ complex is a scaffold for Fe-S clusters

assembly under oxidative stress. [13-14] Three proteins SufB, SufC, and SufD can interact with each other and are indispensable for Fe-S clusters assembly. So, it is important to understand the functions of each protein in the $SufBC_2D$ complex. HDX-MS was utilized to study conformational changes of SufC on ADP/Mg^{2+} binding and changes in SufB and SufD due to the interactions with SufC. The HDX data revealed that the *N*-terminal domain of SufC provided a binding site for ADP and SufC had conformational changes due to the ADP binding. Moreover, SufB and SufD were observed to have regional dynamic changes that might play a role in iron acquisition and Fe-S cluster assembly.

5.2 Impact of HDX-MS Study

The HDX-MS study in *Mt*GpgS gives a deeper insight into the substrate binding mechanism of GT-A glycosyltransferases. The solvent accessibility changes of *Mt*GpgS identified that both substrates can bind to *Mt*GpgS independently and 3PGA can bind to UDPG binding site and act as a competitive inhibitor versus UDPG. It disagrees with the previous proposal that *Mt*GpgS follows a sequential mechanism that UDPG binds to *Mt*GpgS first followed by 3PGA. [15] Most GT-A glycosyltransferases follow an ordered mechanism in which the divalent cation and nucleotide sugar donor bind first, prior to binding of the acceptor. [16-18] *Mt*GpgS is the first GT-A enzyme identified to not follow the sequential binding mechanism. Moreover, the HDX-MS data revealed that *Mt*GpgS has conformational changes upon 3PGA binding, disagreeing with the proposal that the catalytic site of *Mt*GpgS was preformed before donor and acceptor binding. So the investigation of *Mt*GpgS dynamics disagrees with some previous proposals and provides a good example for other members of GT-A family especially those having similar structures with *Mt*GpgS. Similarly, the dynamic investigation of *Cg*MshA also gives a deeper insight into substrate binding mechanism of GT-B glycosyltransferase. Initial

velocity studies indicated an ordered mechanism of *Cg*MshA with UDP-GlcNAc binding first followed by I1P. [12] However, the HDX-MS data disagrees with this proposal and shows that both UDP-GlcNAc and I1P can bind to *Cg*MshA independently. Thus, the study of *Mt*GpgS and *Cg*MshA gives a better understanding of substrate binding mechanism of glycosyltransferases and provides dynamic evidences for studying substrate binding and conformational changes of other glycosyltransferases.

The dynamic investigation of $SufBC_2D$ complex provides evidence for understanding the roles of each protein SufB, SufC, and SufD playing in Suf pathway. Even though the HDX-MS data cannot support the SufC can form a dimer in the presence of ATP and Mg^{2+}, it displays that the SufC has conformational changes on ADP/Mg^{2+} binding within $SufBC_2D$ complex. Moreover, the dynamic changes of SufD points out the possibility that the SufD residue Cys282 may be a potential Fe-S cluster assembly site and the peptides 63-77, 87-91, 106-112, and 184-189 may contribute to the iron acquisition under oxidative stress. So, the clarification of structure and dynamics of $SufBC_2D$ complex helps to understand how Fe-S clusters are assembled and transferred when iron or sulfur mechanism is disrupted by iron starvation or oxidative stress.

5.3 Future Work

The HDX-MS data has revealed that *Mt*GpgS provides a necessary conformation for UDPG and has a global conformational change upon 3PGA binding site. But the conformational changes of *Mt*GpgS induced by 3PGA in the presence of UDPG are still not identified. So HDX-MS could be performed to study dynamic changes of *Mt*GpgS induced by 3PGA in the presence of UDPG. The determination of these changes would clarify the peptides contributing to 3PGA binding and the necessary conformation required by both substrates. The complementary of dynamic changes of *Mt*GpgS would benefit understanding catalytic mechanism of *Mt*GpgS and

being amodel in GT-A glycosyltransferases. Furthermore, dynamic changes of *Mt*GpgS could provide evidences for comparison between GT-A and GT-B glycosyltransferases.

In the Suf pathway, $SufBC_2D$ complex is responsible for Fe-S clusters assembly and has the ability to bind multiple cluster types. [19] Each of cluster types has its own geometrical constraints. It is very likely that the formation and stabilization of a particular cluster type is in part controlled by protein dynamics. [20] So, different conformations of $SufBC_2D$ may be important to support the intermediate cluster types along the Suf pathway. HDX-MS could be applied to define the structural and dynamic changes that occur between APO $SufBC_2D$ and $SufBC_2D$ with Fe-S clusters. It could also provide a mechanistic understanding of Fe-S clusters biogenesis and cluster stabilization.

5.4 References

[1] Engen, J. R. Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. *Anal Chem*. 2009;81:7870-7875.

[2] Kern, D., Eisnmesser, E. Z., and Wolf-watz, M. Enzyme dynamics during catalysis measured by NMR spectroscopy. *Method Enzymol*. 2005;394:507-524.

[3] Yang, H., Luo, G., Karnchanaphanurach, P., Louie, T. M., Rech, I., Cova, S., Xun, L., and Xie, X. S. Protein conformational dynamics probed by single-molecule electron transfer. *Science.* 2003;302:262-266.

[4] Ong, S. E., and Mann, M. Mass spectrometry-based proteomics turns quantitative. Nat. Chem. Biol. 2005;1:252-262.

[5] Wales, T. E., and Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev*. 2006;25:158-170.

[6] Smith, D. L., Deng, Y., and Zhang, Z. Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J. Mass Spectrom.* 1997;32:135-146.

[7] Lu, J., Witcher, D. R., White, M. A., Wang, X., Huang, L., Rathnachalam, R., Bealsand, J. M., and Kuhstoss, S. IL-1beta epitope mapping using site-directed mutagenesis and hydrogendeuterium exchange mass spectrometry analysis. *Biochemistry*, 2005;44:11106-14.

[8] Jackson, M., Brennan, P. Polymethylated polysaccharides from Mycobacterium species revisited. *J Biol Chem.* 2009;284:1949-1953.

[9] Pereira, P. J. B., Empadinhas, N., Albuquerque, L., Sa-Moura, B., Costa, M. S., and Macedo-Ribeiro, S. *Mycobacterium tuberculosis* glucosyl-3-phosphoglycerate synthase: structure of a key enzyme in methylglucose lipopolysaccharides biosynthesis. *Plos One.* 2008;3:e3748.

[10] Kumar, G., Guan, S. Q., and Frantom, P. A. Biochemical characterization of the retaining glycosyltransferase glucosyl-3-phosphoglycerate synthase from Mycobacterium tuberculosis. *Arch Biochem Biophys*. 2014;564:120-127.

[11] Newton, G. L., Ta, P., Bzymek, K. P., and Fahey, R. C. Biochemistry of the initial steps of mycothiol biosynthesis. *J Biol Chem.* 2006;281:33910-33920.

[12] Vetting, M. W., Frantom, P. A., and Blanchard, J. S. Structural and enzymatic analysis of MshA from *Corynebacterium glutamicum*. *J Biol Chem*. 2008;283:15834-15844.

[13] Wollers, S., Layer, G., Garcia-Serres, R., Signor, L., Clemancey, M., Latour, J. M., Fontecave, M., and Ollagnier-de-Choudens, S. Iron-sulfur (Fe-S) cluster assembly: the SufBCD complex is a new type of Fe-S scaffold with a flavin redox cofactor. *J Biol Chem.* 2010;285:23331-23341.

[14] Chahal, H. K., Dai, Y., Saini, A., Ayala-Castro, C., and Outten, F. W. The SufBCD Fe-S scaffold complex interacts with SufA for Fe-S cluster transfer. *Biochemistry.* 2009;48:10644-10653.

[15] Urresti, S., Albesa-Jove, D., Schaeffer, F., Pham, H. T., Kaur, D., Gest, P., van der Woerd, M. J., Carreras-Gonzalez, A., Lopez-Fernandez, S., Alzari, P. M., Brennan, P. J., Jackson, M., and Guerin, M. E . Mechanistic insights into the retaining glucosyl-3-phosphoglycerate synthase from mycobacteria. *J Biol Chem.* 2012;287:24649-24661.

[16] Brew, K., Tumbale, P., and Acharya, K. R. Family 6 glycosyltransferase in vertebrates and bacteria. Inactivation and horizontal gene transfer may enhance mutualism between vertebrates and bacteria. *J Biol Sci*. 2010;285:37121-37127.

[17] Qasba, P. K., Ramakrishnan, B., and Boeggeman, E. Substrate-induced conformational changes in glycosyltransferases. *Trends Biochem Sci*. 2005;30:53-62.

[18] Ly, H. D., Lougheed, B., Wakarchuk, W. W., and Withers, S. G. Mechanistic studies of a retaining α-galactosyltransferase from *Neisseria meningitidis*. *Biochemistry.* 2002;41:5075-5085.

[19] Moulis, J. M., Davasse, V., Golinelli, M. P., Meyer, J., and Quinkal, I. The coordination sphere of iron-sulfur clusters: lessons from site-directed mutagenesis experiments. *J Biol Chem*. 1996:1:2-14.

[20] Hirabayashi, K., Yuda, E., Tanaka, N., Katayama, S., Iwasaki, K., Matsumoto, T., Kurisu, G., Outten, F. W., Fukuyama, K., Takahashi, Y., and Wada, K. Functional dynamics revealed by the structure of the SufBCD complex, a novel ATP-binding cassette (ABC) protein that serves as a scaffold for iron-sulfur cluster biogenesis. *J Biol Chem.* 2015;290:29717-29731.