



X-RAY CRYSTALLOGRAPHIC STUDIES OF SNAP190RcRd (SMALL NUCLEAR
RNA ACTIVATING PROTEIN) COMPLEX AND *E. COLI* GLYCOGEN SYNTHASE

By

Fang Sheng

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

2008

ABSTRACT

X-RAY CRYSTALLOGRAPHIC STUDIES OF SNAP190RcRd (SMALL NUCLEAR RNA ACTIVATING PROTEIN) COMPLEX AND *E. COLI* GLYCOGEN SYNTHASE

By

Fang Sheng

Small Nuclear RNA Activating Protein Complex (SNAPc) is a basal transcription factor that binds to the Proximal Sequence Element (PSE) of human Small Nuclear RNA (snRNA) gene promoter. The Myb-domain RcRd repeats of the largest SNAPc subunit SNAP190, SNAP190RcRd, was believed to be required and sufficient for SNAPc to bind to the PSE. Despite extensive crystallization trials, attempts to obtain the SNAP190RcRd-PSE complex structure in order to understand their specific interaction were unsuccessful. Our EMSA DNA-binding assay did not repeat the results reported in literature. Instead, it shows very limited binding of SNAP190RcRd to the PSE. The size-exclusion chromatogram of SNAP190RcRd suggests a high aggregation state. This result indicates a folding and aggregation problem with peptide SNAP190RcRd, which could hinder SNAP190RcRd binding to the PSE, causing crystallization failure.

Escherichia coli glycogen synthase (ECGS, EC2.4.1.21) is a retaining glycosyltransferase (GT) that transfers glucose from adenosine-diphosphate glucose (ADPGlc) to a glucan chain acceptor. Due to the acceptor analogue HEPPSO, we obtained the first structure of the catalytically active, closed conformation of two-domain protein GS, which hosts a deeply buried active site in the narrow interdomain cleft. Comparison of the ligand-bound *E.coli* GS structures to that of the apo-dmGS (C7S;

C408S) reveals a 15.2° overall domain-domain closure. The structure of the catalytically inactive mutant E377A complexed with oligosacchrides suggests that the glucan chain only binds to the GS N-terminal domain surface. This complies with the frequent open-close motions required by the undisturbed GS catalysis. The absence of a glucose moiety of ADPGlc in two of the E377A complex structures indicates that Glu377 plays an important role in positioning and stabilizing the glucose that is to be transferred. Our wild type *E.coli* GS (wtGS) structures demonstrate that residues Arg300 and Lys305 are close to the ADP phosphate and probably act as Lewis acids, working electrostatically to make the phosphate a better leaving group. Asp137 is suggested to play an important role in positioning these acceptor nucleophile analogues. These results are in agreement with previous biochemical and mutagenesis studies. The observation of the glucose and a DGM-like species in four of the wtGS crystal structures strongly suggests a S_N1 mechanism for the GT-B retaining enzymes over the once-prevalent double replacement mechanism. The highly reactive S_N1 mechanism intermediate, DGM, is probably stabilized by the nearby phosphate moiety, the incoming nucleophile, and the His161 backbone in the GT-B retaining enzymes. GT-B retaining enzyme family boasts approximately a thousand important sugar-modifying enzymes that are believed to share a virtually identical active site. Finally, based on our *E.coli* GS structures and the chimeric studies of GBSS and SSII, we speculate that the region (α 13 - α 18) in starch synthase is involved in positioning the two domains and that interaction of this region of GBSS with amylopectin possibly orients the domains for catalysis.

To my teacher, Prof. Zhanru Liao, who introduced me into research and
her infectious passion for science inspired me along the road.

ACKNOWLEDGEMENTS

I thank my advisor Dr. James H. Geiger for his support, inspiration, encouragement and most of all, his never-ending faith in me. Without that, I couldn't have ordered and tried a wide range of rare and expensive chemicals and among them find the magic reagent HEPPSO for GS crystallization. I thank our collaborator, Dr. Jack Preiss, for giving me the opportunity to work on GS. Now the puzzle of the glycogen synthesis finds its last piece: GS. I also thank our collaborator, Dr. Bill Henry, for being too busy to nag me but never too busy to be available and helpful. Talking with you has always been instructive and delightful. I am in gratitude to my committee members Dr. James K. McCusker, Dr. Thomas J. Pinnavaia, and Dr. David P. Weliky.

I want to say thank you to Dr. Stacy Hovde who took time out from her tight schedule to show me the ropes in the beginning. I appreciate your company in those around-the-clock synchrotron shifts. Our 2 AM tricycle racing in APS cycle was fun and did cheer us up from hundreds of poor diffracting crystals. I thank Dr. Xiangshu Jin for her invaluable help with the structure refinement and Dr. Sara E. Cnudde for paving the path to the usage of CCP4. I want to thank Xiaofei Jia for his great help with NMR trials and helpful discussion. I wish you the best in your career. I also want to thank Joe Leukyam at GTSF facility for assistance with [1-¹³C]-ADPGlc purification and Dr. Dan Holmes for help with NMR experiments. Paul Reed, you are the best computer technician anyone can ever ask for, thank you for putting up with my numerous and silly questions. I also want to thank Lisa in the graduate office for having the answers to all things bureaucratic and accommodating me to teach my favorite classes.

My sincere gratitude extends to all Geiger lab members, Lei Feng, Suzan, Andy, Blanka, Dorothy, Aimee, Justin and Kathy. Knowing you and working with you in lab has been a pleasure. I also want to thank my friends Jess Gunn for helping me with preparing my seminar, and Christine Kalcic for introducing me to line dancing which I really enjoy.

I want to say that my previous advisor in China, Prof. Zhanru Liao, was one of the biggest inspirations in my life. Her passion for science encouraged me to fly overseas to the States to pursue my Ph.D. I thank my parents and my dear brother who believe in me no matter what and whose love sustain me. Last, but not least, I thank my husband, Jian Yang, whose love and support have accompanied me and made the pursuit of my Ph.D joyful.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vii
LIST OF TABLES.....	xi
LIST OF FIGURES.....	xii
LIST OF ABBREVIATIONS.....	xx
CHAPTER I	
INTRODUCTION.....	1
I.1. Small Nuclear RNA Activating Protein Complex (SNAPc) 190RcRd.....	1
I.1.1. Transcription Regulation and Transcription Factors.....	1
I.1.2. Small Nuclear RNA Promoters (PSE, TATA and DSE).....	2
I.1.3. Small Nuclear RNA Activating Protein Complex (SNAPc).	4
I.1.4. SNAP190RcRd and Myb Domain.....	7
I.1.5. Objectives.....	14
I.2. Glycogen Synthase.....	15
I.2.1. Glycogen.....	15
I.2.2. Starch.....	17
I.2.3. Glycosyltransferase.....	19
I.2.4. The Retaining Glycosyltransferase Mechanism Controversy.....	22
I.2.5. Previous Studies on Bacterial Glycogen Synthase.....	30
I.2.6. Significance of <i>E.coli</i> GS Structural Studies.....	32
I.3. References.....	35
CHAPTER II	
CRYSTALLIZATION AND PRELIMINARY X-RAY DIFFRACTION ANALYSIS OF SNAP190RcRd.....	45
II.1. Experimental Procedures	45
II.1.1. Transformation and Overexpression of SNAP190 RcRd.....	45
II.1.2. Purification of SNAP190 RcRd.....	47
II.1.3. Purification of DNA and Annealing.....	49
II.1.4. Crystallization of SNAP190RcRd, SNAP190RcRd/ DNA and SNAP190RcRd / TBP/ DNA Complexes.....	50
II.1.5. Electrophoretic Mobility Shift Assay (EMSA).....	52
II.2. Results and Discussion.....	53
II.2.1.Preparation of SNAP190RcRd.....	53

II.2.2. Crystallization Trials.....	55
II.2.3. EMSA DNA-binding assay.....	55
II.3. Conclusion.....	60
II.4. Appendix.....	61
II.5. References.....	62
CHAPTER III	
THREE DIMENSIONAL STRUCTURES OF <i>Escherichia Coli</i> GLYCOGEN SYNTHASE AND ITS COMPLEXES.....	63
III.1. Theory.....	63
III.1.1. Structure Determination from X-ray Diffraction Data.....	63
III.1.2. Molecular Replacement	64
III.1.3. Structure Refinement.....	66
III.2. Experimental Procedures.....	68
III.2.1. Protein Overexpression.....	68
III.2.2. Protein Purification.....	69
III.2.2.A. His-tagged GS protein.....	69
III.2.2.B. Untagged dmGS protein.....	70
III.2.3. GS and GS Complex Crystallization.....	71
III.2.4. Diffraction Data Collection and Processing.....	73
III.2.5. ¹³ C-and ¹ H- ¹³ C NMR Experiment.....	73
III.2.5.A. Preparation of D-[1- ¹³ C]-ADPGlc from α-D-[1- ¹³ C] glucopyranosyl 1-phosphate.....	73
III.2.5.B. NMR experiment setting.....	75
III.2.5.C. NMR sample preparation.....	76
III.2.5.D. NMR spectra	76
III.3. Model Building and Structure Refinement.....	81
III.3.1. apo dmGS (C7S: C428S).....	81
III.3.2. Wild - type GS in Complex with ADP and D-glucopyranosylium (DGM) (wtGSa).....	86
III.3.3. Wild - type GS in Complex with ADP and Glucose (wtGSb, wtGSb and GSd).....	88
III.3.4. GS Mutant E377A in Complex with ADP and Oligosaccharides	95
III.3.5. GS Mutant E377A in Complex with ADP.....	97
III.4. <i>Escherichia coli</i> GS Structures.....	99
III.4.1. apo dmGS (C7A;C428S) Structure.....	99

III.4.2. GS- ADP-Glucose-HEPPSO Complex Structures	102
III.4.2.A. ADP binding site.....	104
III.4.2.B. Glucose binding site.....	108
III.4.2.C. Acceptor analogue HEPPSO binding site	111
III.4.3. GS -ADP-DGM-HEPPSO Complex Structure.....	114
III.4.4. E377A-ADP-HEPPSO Complex Structure	118
III.4.5. E377A- ADP-Oligosaccharide Complex Structure	119
III.4.5.A. Interdomain cleft oligosaccharide-binding site.....	120
III.4.5.B. The N-terminal surface oligosaccharide-binding sites.....	126
III.4.5.C. The glucose unit configuration in oligosaccharide-bound E377A...	133
III.4.5.D. Binding mode analysis.....	134
III.4.5.E. Oligosaccharide-binding causes little protein conformation change	134
III.4.5.F. Oligosaccharide conformation analysis.....	135
III.5. <i>E.coli</i> GS Structure Analysis	139
III.5.1. Open-Close: GS Dynamic Motion.....	139
III.5.1.A.Comparison of ADP conformations in GS open and closed form...	142
III.5.1.B. Large displacement of KTGGL loop in GS open and closed form..	143
III.5.1.C. What causes the enzyme to close?	146
III.5.1.D. Domain-wise ligand binding facilitates GS open-close motion.....	148
III.5.1.E. Substrate binding order.....	149
III.5.2. Important Residues	150
III.5.2.A. Asp137 is the +1 sugar positioning residue	151
III.5.2.B. His161 participates in –1 sugar positioning and intermediate DGM stabilization.....	152
III.5.2.C. The catalytic Lewis acid Arg300 side-chain switches in and out of the GS active site in response to the presence of ADP.....	152
III.5.2.D. Catalytic Lewis acid residue Lys305.....	155
III.5.2.E. Glu377 is responsible for the –1 sugar positioning and helps Lys305 maintain positive charge.....	157
III.5.3. GS Complex Structures Support S _N 1 Mechanism Over the Double Displacement Mechanism.....	160
III.5.3.A. DGM intermediate	160
III.5.3.B. DGM stabilization in GS.....	161
III.5.3.C. Effort to verify the existence of carbocation DGM using NMR.....	162
III.5.3.D. The substrate phosphate group deprotonates the incoming acceptor nucleophile.....	169
III.5.3.E. Evidence against the double-displacement mechanism	169
III.5.4. The GS Catalytic Scenario Suggested by <i>E.coli</i> GS Structural Studies	171

III.6. Mechanism Implication for Starch Synthase	171
III.6.1. Plant Starch Synthases Share a Similar Active Site and Catalytic Mechanism with Bacterial GS	171
III.6.2. Glucan Chain Binding in Starch Synthase	174
III.6.3. Insights into GBSS and SS Specificities.....	175
III.7. Conclusions.....	177
III.8. References.....	179

LIST OF TABLES

Table I.1. The sequence alignment of SNAP190 Myb domain and human A-, B-, c-Myb domains. Important DNA-interacting residues are in bold and underlined. Three well-defined helices in the human c-Myb domain are illustrated as tandem connected cylinder (H1, H2, and H3).....	12
Table I.2. The sequences of the PSEs present in the wild type and mutant probes. Uppercase letters correspond to the PSE sequence from the mouse U6 promoter, with red-colored characters corresponding to mutations.....	13
Table I.3. Kinetic data of DGM mimics in GT-B retaining enzymes.....	29
Table II.1. The SNAP190RcRd sequence used in crystallization trials and EMSA assay.....	46
Table II.2. Oligonucleotides used in crystallization trials.....	51
Table II.3. DNA sequences used in Figure II.6.....	59
Table III.1. GS complex crystallization trials.....	72
Table III.2. Data collection and refinement statistics of <i>E.coli</i> GS mutant and their complexes.....	82
Table III.3. Data collection and refinement statistics of wild type <i>E.coli</i> GS complexes.....	92
Table III.4. Comparison of the environment of DGM and glucose.....	116
Table III.5. Conformation of oligosaccharides when bound to GS, MalP, and GP.....	137
Table III.6. Kinetic parameters of wild type glycogen synthase and mutants.....	151
Table III.7. Selective conserved residues in bacterial GS and plant starch synthases. Those residues critical to GS activity and their equivalents in SSs are colored pink. Multiple sequence alignment was conducted with DNASTAR. GS sequences used were: <i>Escherichia coli</i> GS (P0A6U8), <i>Agrobacterium tumefaciens</i> GS (AAD03474), <i>Pyrococcus abyssi</i> GS (NP_125769), Sequences of granule-bound starch synthases from various organisms were those of barley (AAL77109), maize (P04713), potato (CAA41359), and rice (P19395), Other starch synthases used were those of maize SSI (AAB99957), potato SSI (P93568), wheat SSI (Q43654), wheat SSIIa (BAE48798), maize SSIIa (AAS77569), potato SSII (CAA61241), potato SSIII (Q43846), wheat SSIII (AAF87999) and <i>Chlamydomonas reinhardtii</i> SS (AAC17971).....	173

LIST OF FIGURES

Images in this dissertation are presented in color

Figure I.1 Composition of pol II and pol III snRNA promoters in higher eukaryotes. Typical eukaryote RNA pol II snRNA genes are U1-U5 genes and RNA pol III snRNA genes are U6 and 7-SK genes.....	3
Figure I.2. Architecture of SNAPc complex.....	5
Figure I.3. The composition of SNAP190.....	6
Figure I.4. Ternary Myb protein-enhancer binding protein-DNA complex (1H88.pdb). The Myb protein (39-190) is shown as cartoon and three repeats R1, R2, and R3 are in green, yellow, and blue, respectively.....	9
Figure I.5. The close-up view of the interaction between human Myb protein R2, R3 and DNA in the ternary protein-DNA complex (1H88.pdb). The conserved tryptophan residues are shown as magenta sticks.....	10
Figure I.6. Top: Typical GT-A fold illustrated by protein SpsA. Bottom: Structure of the GT-B fold enzyme GtfB. A Mn^{2+} ion in GtfB active site is depicted as blue sphere. The fold name is after their initial observation in the SpsA (<i>Bacillus subtilis</i> glycosyl-transferase) and GtfB (DNA β -glucosyltransferase structures).....	21
Figure I.7. Cartoon presentation of <i>Agrobacterium tumefaciens</i> GS complexed with ADP (blue). The N-terminal peptide 15-21 containing the KTGGL loop and Asp21 and the C-terminal residue Arg299 and Lys304 are colored red because they were indicated by mutagenesis and kinetic studies to be important in GS catalysis.....	22
Figure I.8. Proposed double-displacement mechanisms for GT retaining enzymes. Two essential components of this mechanism, the catalytic nucleophile and glucosyl-enzyme intermediate, are circled and framed, respectively.....	24
Figure I.9. S_N1 -like mechanism D-glucopyranosylium ion (DGM) intermediate is framed, which is otherwise considered a transition-state in the S_{Ni} -like mechanism. In both cases, positively charged DGM is stabilized by leaving group AMP-phosphate and the incoming nucleophile 4-OH group of sugar.....	26
Figure I.10. DGM (A) and DGM mimics (B-G) that have been tested in kinetic studies of GT-B retaining enzymes. The related kinetic data are listed in Table I.3.....	28
Figure I.11. Reactions catalyzed by GT-B retaining enzyme OtsA, MalP, and GP.....	31

Figure II.1. Predicted secondary structure of SNAP190RcRd (390-518) by method PSIPRED. Helix and coil are depicted as green cylinder and black straight lines. The confidence of predication (Conf) is indicated by the height of cyan columns. Pred and AA stands for the predicted secondary structure and target sequence, respectively.....46

Figure II.2. SDS-PAGE gel showing the GST-tagged RcRd peptide (lane 3 and 4) and the RcRd peptide alone (lane 6 and 7). The GST-tag was eluted by 100 mM reduced glutathione solution afterwards (lane 9 and 10).....48

Figure II.3. Diagram of SNAP190RcRd purification through Source-Q column.....49

Figure II.4. SDS-PAGE gel of SNAP190RcRd from Source-Q column. Lane 2, 3, and 4 corresponds to the fraction 27, 28, and 29, respectively.....49

Figure II.5. Size-exclusion chromatogram results from Sephacryl S-300 HiPrep 16/60 column. The blue line is Bio-Rad molecular weight standards (158 kDa bovine gamma-globulin, 44 kDa chicken ovalbumin, 17 kDa equine myoglobin, and 1.3 kDa vitamin B12. The red line is from SNAP190RcRd. Experiment was performed at 1mL /min flow rate and the fractions were 5 mL.....54

Figure II.6. An EMSA was performed with a probe containing B-MPSE (lane 1-6), S-MPSE (lane 7-12), or D-MPSE (lane 13-18). B-MPSE, S-MPSE, and D-MPSE are blunt ends-, single overhang-, and double overhang- mouse U6-21 PSE, respectively and their sequences are listed in Table II.2. In lane 1, 7, and 13, no SNAP190RcRd or recombinant SNAPc (rSNAPc) were added to the probes. Lane 2-5, 8-11, and 14-17 contain increasing amount of SNAP190RcRd (0.1, 0.3, 1, 3 μ g). Lane 6, 12, and 18 contain 3 μ g mSNAPc.....56

Figure II.7. An EMSA was performed with varied length of probes. mSNAPc was added to the * labeled lanes. Lane 1-2, 3-4, 11-12, 23-24, and 29-30, 35-36 contain increasing amount of mSNAPc (1 and 3 μ g). All probe sequences are listed in Table II.3. Lane 7-10, 13-16, 19-22, 25-28, 31-34, and 37-40 contain increasing amount of SNAP190RcRd (0.1, 0.3, 1, 3 μ g).....58

Figure III.1. Representative His-tagged *E.coli* GS SDS-PAGE gel. The most left lane is the molecular weight standard. Each elution lane accounts for the collection of a 5 mL elution buffer fraction.....69

Figure III.2. *E.coli* dmGS SDS-PAGE. The second left lane is the molecular weight standard and all other lanes are elution from Source-Q column with increasing concentration of KCl.....70

Figure III.3. HPLC spectrum from the DEAE-5PW column. Flow rate: 0.6 mL /min, Gradient: 2 min: 0% buffer B; 40 min: 25 % buffer B; 60 min: 100 % buffer B. Fraction ADPGlc elutes in the range of 14 –18 % buffer B.....	75
Figure III.4. NMR spectra of the starting material [1- ¹³ C]-Glc-1-P and batch I [1- ¹³ C]-ADPGlc.....	77
Figure III.5. NMR spectra of batch II [1- ¹³ C]-ADPGlc.....	78
Figure III.6. NMR spectra of the starting material [UL- ¹³ C ₆]-Glc-1-P and batch I [UL- ¹³ C ₆]-ADPGlc.....	79
Figure III.7. A crystal of dmGS.....	81
Figure III.8. Ramachandran plot of <i>E.coli</i> dmGS. Phi (degrees) is x and Psi (degrees) is y	85
Figure III.9. Crystals of wtGSa. The center crystal is the one the diffraction data was collected from.....	86
Figure III.10. Ramachandran plot of <i>E.coli</i> GS /ADP/ DGM /HEPPSO wtGSa complex structure. Phi (degrees) is x and Psi (degrees) is y	87
Figure III. 11. Crystals of wtGSb.....	88
Figure III.12. Ramachandran plot of <i>E.coli</i> GS / ADP / glucose / HEPPSO wtGSb complex structure. Phi (degrees) is x and Psi (degrees) is y.....	89
Figure III.13. Ramachandran plot of <i>E.coli</i> GS /ADP /glucose /HEPPSO wtGSb complex structure. Phi (degrees) is x and Psi (degrees) is y.....	90
Figure III.14. Ramachandran plot of <i>E.coli</i> GS /ADP /glucose /HEPPSO wtGSd complex structure. Phi (degrees) is x and Psi (degrees) is y	91
Figure III.15. Crystals of ADP, oligosacchride-bound E377A.....	95
Figure III.16. Ramachandran plot of E377A /ADP /oligosaccharides complex structure. Phi (degrees) is x and Psi (degrees) is y	96
Figure III.17. Crystals of ADP-bound E377A.....	97
Figure III.18. Ramachandran plot of E377A-ADP-HEPPSO complex structure. Phi (degrees) is x and Psi (degrees) is y	98

Figure III.19. Overall structure of *E.coli* dmGS. The N-terminal domain (1-241) and C-terminal domain (250-457) are colored yellow and blue, respectively. The interdomain peptide 242-254 and domain-spanning helix α 18 (458-476) are colored magenta and orange, respectively. The N-terminal loop 212-215 and helix α 7 (215-220) are in green whereas the C-terminal helix α 14 (398-403) is in light blue. The mutation sites of dmGS (C7S: C408S) are shown in red sticks.....100

Figure III.20. Interdomain interactions between Glu218 and Gly399; Thr214 and Gly399, and Val248 and Phe460 are shown as dotted lines in *E.coli* dmGS. Loop 212-215 and helix α 7 (215-220) are in green whereas the C-terminal helix α 14 (398-403) is in light - blue. The linker peptide 242-254 and domain-spanning helix α 18 (458-476) are colored magenta and orange, respectively.....101

Figure III.21. Overall structure of wtGSb complex. The N-terminal domain is in pink and C-terminal domain is in blue. The linker peptide 242-254 and domain-spanning helix α 18 (458-476) are colored magenta and orange, respectively.....103

Figure III.22. Top: ADP (shown in yellow and in atom colors) bound in the active site of wild type *E.coli* GS. Hydrogen bonds between ADP and protein are shown as broken lines. Residues from the N-terminal domain and the C-terminal domain are colored pink and blue, respectively. Bottom: 1.0 σ contoured 2Fo-Fc electron density map of the bound ADP in active site.....104

Figure III.23. Extensive interactions between ADP and residues Arg300 and Lys305 in the wild type *E.coli* GS active site. Water molecules are presented as red spheres.....106

Figure III.24. Helix α 1(17-32) (cyan) and α 13(382-389) (purple) point to the ADP phosphate group. The NH-ends of both helices are presented as yellow ribbons.....107

Figure III.25. Top: glucose (shown in yellow and in atom colors) bound in the active site of wild type *E.coli* GS. The nearby HEPPSO and ADP are shown as black and blue sticks. Hydrogen bonds between glucose and protein are shown as broken lines. Residues from the N-terminal domain and the C-terminal domain are colored pink and blue, respectively. Bottom: 1.0 σ contoured 2Fo-Fc electron density map of the bound glucose.....109

Figure III.26. Interactions between glucose, ADP and HEPPSO in the wtGSb active site.....110

Figure III.27. Diagram of HEPPSO and its atom numbering.....111

Figure III.28. Top: Interactions between HEPPSO and N-terminal residues, ADP, and glucose. Bottom: 1.0 σ contoured 2Fo-Fc electron density map of the bound HEPPSO112

Figure III.29. 2.5 σ Fo-Fc electron density map of wtGSa (2.83 Å resolution) with DGM (left), D-arabino-hex-1-enitol (middle) and wtGSb (2.22 Å resolution) with glucose (right). The ADP molecule is shown as lines for clarity.....114

Figure III.30. DGM in wtGSa is in a superimposable position of glucose in wtGSb. Ligands in wtGSb are in yellow whereas those in wtGSa are atom-wise colored.....115

Figure III.31. Positively charged C1 and O5 of DGM are stabilized by ADP, HEPPSO, and His161. Interactions starting from the DGM C1 are shown as red dotted lines.116

Figure III.32. Overlay of the active site of *E.coli* wtGSb and E377A. The residues and ligand of E377A structure are colored yellow whereas residues in wtGSb are in green and ligands are in magenta. Water W1, W2, and W3 are located between Arg300 and the ADP distal phosphate oxygen in the E377A structure.....118

Figure III.33. Overall structure of *E.coli* GS E377A in complex with ADP and oligosaccharides. ADP and oligosaccharides are shown as sticks. The secondary structure elements which host residues interacting with the surface-bound oligosaccharide at Gc and Gd sites are labeled.120

Figure III.34. 2Fo-Fc electron density map contoured at 1 σ with maltotriose at the Ga site. The ADP molecule is shown for reference.....120

Figure III.35. Schematic representation showing the interactions between *E.coli* GS E377A residues and the bound maltotriose at Ga site. Observed hydrogen bonding and aromatic stacking are depicted with regular and wide dashed lines, respectively. Van der Waals interaction is marked with the wavy line.....121

Figure III.36. The interaction between enzyme GS and the bound maltotriose at the Ga site.....122

Figure III.37. Superimposition of the oligosaccharide- binding site in E377A-ADP-oligosaccharide (green), wtGS-ADP-glucose-HEPPSO (yellow), and MalP-PLP-G5 (PDB: 1L6I) (blue) complexes. The HEPPSO that occupies the comparable position of the oligosaccharide was omitted for clarity. GS residues are labeled in black while MalP residues are in blue.....125

Figure III.38. Top: 2Fo-Fc electron density map contoured at 1 σ with maltose at the Gb site. Bottom: Interaction between protein GS and the bound maltose at the Gb site.127

Figure III.39. Top: 2Fo-Fc electron density map contoured at 1σ with maltopentaose at the Gc site. Bottom: 2Fo-Fc electron density map contoured at 1σ with maltohexaose at the Gd site.....	128
Figure III.40. Schematic diagram of interactions between protein and the bound maltopentaose at the Gb site. Observed hydrogen bonding and Van der Waals interaction are depicted with dashed lines. The ionic interaction is marked with arrow.....	129
Figure III.41. Interaction between protein GS and the bound maltopentaose at the Gb site.	130
Figure III. 42. Schematic diagram of interaction between protein and the bound maltohexaose at the Gc site. Observed hydrogen bonding and ionic interactions are depicted with dashed lines and arrows.....	132
Figure III.43. Interactions between protein and the bound maltohexaose at the Gc site	133
Figure III.44. Overlay of oligosaccharides at the Ga (red, 1-3), Gb (cyan, 4-5), Gc (green, 6-10), and Gd (yellow, 11-16) sites.....	136
Figure III.45. Comparison of the ideal glycogen helix model, glycogen phosphorylase – bound maltopentaose (blue, 1P2B.pdb), oligosaccharides at GS Ga (red), Gb (cyan, 4-5), Gc (green, 6-10), and Gd (yellow, 11-16) sites.....	138
Figure III.46. Superposition of the C-terminal domains from dmGS (cyan) and wtGSb (yellow). Ligands bound in wtGSb (HEPPSO, ADP and glucose) are shown as black sticks. Structural elements essential for catalysis are colored red and their counterparts in the open form dmGS are colored blue. Residues critical for glucosyl transfer (His161, Arg300, and Lys305) are shown as red sticks.....	141
Figure III.47. Structural comparison of ADP molecule in open form <i>A. tumefaciens</i> GS (yellow, 1rzu.pdb) and closed from <i>E.coli</i> GS (blue). The residues from <i>A. tumefaciens</i> GS are labeled in red and their carbons are colored yellow while those from <i>E.coli</i> GS are labeled in black and their carbons are colored blue. The interaction between ADP and protein are shown as broken lines.....	142
Figure III.48. Structural comparison of (bottom) loop 13-20 in apo-dmGS (yellow) and wtGSb complex (red) and (top) the equivalent loop14-24 in the UDP/ imidazole /Glc-6-P bound OtsA (green, 1gz5.pdb) and UDPGlc-bound OtsA (blue, 1uqu.pdb). Gly22 in OtsA and Gly18 in GS interacting with the phosphate oxygen are shown as sticks. Ligands other than UDP (OtsA) and ADP (GS) are omitted for clarity.....	144

Figure III.49. Schematic diagram showing the cross-domain network between ligand HEPPSO, glucose and ADP, which chiefly interacts with the N-terminal (red) and C-terminal (blue) residues, respectively.147

Figure III.50. In apo *E.coli* dmGS (cyan, atom-wise colored), apo AtGS (1rvz.pdb, ruby), apo PaGS (2bis.pdb, orange) and E377A-ADP-HEPPSO- complex (magenta). The side-chain of Arg300 and its equivalents are out of the GS active site and are in contact with Ala329 and Gly330 (Lys in PaGS). In ADP, Glc-bound wtGS (2qzs.pdb, green, atom-wise colored), oligosaccharide-bound E377A (blue), and ADP-bound AtGS (1rvz.pdb, yellow), Arg300 and its equivalent side-chain are close to the ADP phosphate and their interaction are shown as dotted lines (black lines for wtGS and red lines for AtGS complex). Hydrogen bonds between the Arg300 side-chain and Gly330 in apo-dmGS, and the Arg300 side-chain and maltotriose in oligosaccharide-bound E377A complex are also shown as dotted lines. Residues are labeled according to the *E.coli* GS sequence.....154

Figure III.51. Structural comparison of ADP, glucose binding site and active site residues among *E.coli* GS (in yellow and atom colored, wtGSb), MalP (blue, 2asv.pdb), rabbit R-GP (pink, 1gpa.pdb), OtsA (green, 1uqu.pdb) and AGT (cyan, 1y6f.pdb). MalP Ligands, PO_4^{3-} , ASO (1,5-anhydrosorbitol) and maltopentaose, occupy the equivalent positions of the ADP distal phosphate group, glucose and HEPPSO in the wtGSb structure, respectively. Maltotriose in E377A complex was also shown (red sticks). Residues are labeled according to the *E.coli* GS sequence. Top: the front view. Bottom: the back view.....156

Figure III.52 Top: The Glu377 environment in the wild-type GS structure. Arrows indicated the hydrogen bond proton donating direction. Bottom: The interaction between the modeled Asp377 and neighboring residues.....159

Figure III.53. NMR spectra of $[1-^{13}C]$ -ADPGlc -GS complex and $[1-^{13}C]$ -ADPGlc -GS-HEPPSO complex.....166

Figure III.54. NMR spectra of batch I $[1-^{13}C]$ -ADPGlc and its complex at different time points.....167

Figure III.55. NMR spectra of buffers in production of batch I and II $[1-^{13}C]$ -ADPGlc. The Tris and ammonium formate spectra are adopted from Spectral Database for organic compounds (SDBS).....168

Figure III.56. Modeled potato GBSS structure based on *E.coli* GS structure. The region ranging from helix $\alpha 13$ to $\alpha 18$ is colored red which is the determinant of most specific

properties of GBSS and may play a role in tuning the relative orientation of the two domains.....	176
---	-----

LIST OF ABBREVIATIONS

ADPGlc – adenosine diphosphate glucose
AGT - α -glucosyltransferase
Ala – alanine
AMP – adenosine monophosphate
Arg - arginine
ASO - 1,5-anhydrosorbitol
AtGS - *Agrobacterium tumefaciens* glycogen synthase
ATP – adenosine triphosphate
APS - advanced photon source

C – cysteine
C α - the alpha carbon in the peptide
CAZY - carbohydrate active enzymes
C-terminal - carboxyl terminal
CC - correlation coefficient
CCD - charged-coupled device
Cys - cysteine

dmGS – *E.coli* double mutant (C7S;C408S) glycogen synthase
DGM - D-glucopyranosylum
DEAE - diethylaminoethyl cellulose
DTT – dithiothreitol

E – glutamic acid
EMSA - electrophoretic mobility shift assay

F_{hkl} – structural factor

GBSS - granule-bound starch synthase
GH - glycoside hydrolase
Glc - glucose
Glc-1-P - glucose 1-phosphate
Glc-6-P - glucose 6-phosphate
Glu – glutamic acid
Gln - glutamine
Gly - glycine
GP - glycogen phosphorylase
GS - glycogen synthase
GST - glutathione S-transferase
GT - glycosyltransferase
GTA - N-acetylgalactosaminyltransferase

GTB - galactosyltransferase

HEPES - (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HEPPSO - 4-(2-Hydroxyethyl) piperazine-1- (2-hydroxypropane) sulfonic acid

His - histidine

ID - identification

IMD - imidazole

IPTG - isopropyl thiogalactopyranoside

LB – luria broth

Leu - leucine

Lys – lysine

MalP - maltodextrin phosphorylase

mm – millimeter

M.W.- molecular weight

N-terminal – the amino terminal of protein

ND – not determined

NMR – nuclear magnetic resonance

OtsA - trehalose-6-phosphate synthase

PaGS - *Pyrococcus abyssi* glycogen synthase

PDB - protein data bank

PEG – polyethylene glycol

Phe - phenylalanine

PLP - pyridoxal phosphate

RMSD - root mean square deviation

PMSF – phenylmethanesulphonylfluoride

Pro - proline

PSE – proximal sequence element

S – serine

SDS-PAGE - sodium dodecyl sulfate- poly acrylamide gel electrophoresis

Ser - serine

SNAPc – small nuclear RNA activating protein complex

snRNA – small nuclear RNA

SS - starch synthase

TAF - TBP-associated factor

TB – Terrific Broth

TBP - TATA-box binding protein

TDB - thrombin digestion buffer

TEA - triethanolamine
Thr - threonine
Tris - 2-amino-2- (hydroxymethyl)-1,3-propanediol
Trp – tryptophan
Tyr - tyrosine

UDP - uridine diphosphate
UDPGlc - uridine diphosphate glucose
 μg – microgram
 μl – micro liter

Val - valine

wat – water
wk – week
wtGS – wild type *E.coli* glycogen synthase
w/v – weight/volume

CHAPTER I

INTRODUCTION

I.1 Small Nuclear RNA Activating Protein Complex (SNAPc)

190 RcRd

I.1.1 Transcription Regulation and Transcription Factors

Gene expression is a fundamental process in which a gene DNA sequence is converted into a functional protein. It basically involves two steps, transcription and translation. RNA polymerase (pol) is the primary enzyme of the first step in which a specific RNA is made from the sequence information in the DNA genome. The resulting RNAs participate in the second step in a variety of ways to synthesize functional proteins and ensure the DNA genome information is faithfully and efficiently expressed(1).

Eukaryotes use three distinct RNA pols to catalyze gene transcription. RNA pol I transcribes the ribosomal RNA (rRNA) genes; RNA pol II is responsible for transcribing the messenger-RNA (mRNA) genes and the majority of the small nuclear RNA (snRNA) genes. RNA pol III catalyzes the transcription of transfer-RNA (tRNA) genes and some other snRNA genes (2,3).

To date the only genes that are known to be transcribed by both RNA pol II and pol III are snRNA genes, which encode essential RNA components of small nuclear ribonucleoprotein particles (snRNPs) involved in mRNA processing. Many of the snRNAs are very abundant in cells, ranging from 100,000 to 1 million molecules per nucleus, therefore, transcription from snRNA-type promoters represents a significant portion of the total amount of transcription initiated by both RNA pol II and pol III (4).

Although RNA pols are the primary enzymes that catalyze transcription, they are not able to recognize their target DNA sequences and initiate transcription on their own. A series of Transcription Factors (TFs) are required to assemble on the promoter region and form a preinitiation complex to recruit the specific RNA pol. The preinitiation TF complex is constructed and maintained through protein-DNA (TF-DNA) and protein-protein interactions (TF-TF). Some TFs bind to the promoter DNA directly. Some TFs bind to those TFs that are in direct contact with DNA and modulate their activities (2,3).

Most eukaryotic transcription initiation machinery is complicated in terms of promoter structure and TFs. The mRNA-type promoters are located upstream from transcription start sites (toward the 5' end of the transcription initiation site) whereas tRNA-type promoters reside both downstream and upstream from the transcription start site(5).

Unlike RNA pol II mRNA-type genes whose promoter structures and transcription factors are complex, eukaryote RNA pol II and pol III snRNA genes have a similar promoter upstream of the transcription start site and similar TFs binding to the promoter and directing transcription(5-7). Such distinct properties make eukaryote snRNA genes, including those of humans, a unique model to understand principles of polymerase selection and the mechanism of RNA pol II and pol III transcription machineries.

I.1.2 Small Nuclear RNA Promoters (PSE, TATA and DSE)

Like the promoters of most genes, snRNA promoters can be divided into two distinct functional regions: the core promoter region and the regulatory region. Core promoter regions contain the binding sites for the basal transcription factors that promote the assembly of preinitiation complexes, and therefore are sufficient to direct low levels of

transcription *in vitro*. Regulatory regions are responsible for the recruitment of activator or repressor proteins, which modulate the level of transcription (6,8,9).

The pol II and pol III core promoters in human snRNA genes contain a Proximal Sequence Element (PSE) which is centered ~55 base pairs (bp) upstream of the transcription start site (or denoted -55 considering zero for the transcription start site) and is recognized by the same Small Nuclear RNA Activating Protein Complex (SNAPc) (Figure I.1)(6,10). Both pol II and pol III regulatory regions contain a Distal Sequence Element (DSE) ~220 bp upstream of the transcription start site (-220) where a specific protein, Oct-1, binds and enhances transcription from the core promoter(8). In addition, the pol III snRNA gene core promoter contains a TATA box (TATAAAAG) at position 25 bp downstream of PSE(6). The presence of the TATA box in the human snRNA gene promoter is the determinant of RNA pol specificity because addition or deletion of the TATA box switches RNA pol specificity from II to III or vice versa, while exchanging PSEs or DSEs between human RNA pol II and III snRNA gene promoters does not have the same effect(6).

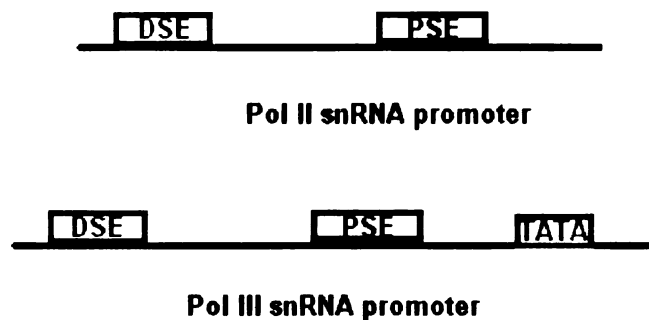


Figure I.1 Composition of pol II and pol III snRNA promoters in higher eukaryotes. Typical eukaryote RNA pol II snRNA genes are U1-U5 genes and RNA pol III snRNA genes are U6 and 7-SK genes.

The TATA box is bound by TATA-box Binding Protein (TBP). Interestingly, TBP is not only required for RNA pol III snRNA gene transcription from a promoter which has the TATA box, but also RNA pol II snRNA gene transcription from a promoter that does not have the TATA box. On the RNA pol III snRNA promoter the PSE is located at a fixed distance from the TATA box. Mutating either PSE or TATA box switches transcription to RNA polymerase II(11,12). Thus PSE or PSE-binding factor is likely to interact with TBP and to be involved in the RNA pol selection for human snRNA promoters(6).

The DSE is located approximately 160 bp upstream from PSE along the DNA chain, but in vivo the DSE and the PSE are brought into close proximity in space by a positioned nucleosome that resides between them and wraps about 140 bp DNA(5,9). The DSE contains a conserved octamer sequence (ATXXXXAT) (X is a residue not conserved) and the protein specifically binding to DSE through that octamer motif is called Oct-1. Oct-1 and the PSE binding factor SNAPc interact with each other and lead to cooperative binding (5,9,13).

1.1.3 Small Nuclear RNA Activating Protein Complex (SNAPc)

In 1991, Human snRNA PSE binding factor, SNAPc, was discovered in a nuclear HeLa cell extract. HeLa is a human cancer cell line commonly used to obtain biomaterials of human origin for its extremely rapid proliferation ability. SNAPc is a five subunit protein complex consisting of SNAP190 (190 kDa), SNAP50 (50 kDa), SNAP45 (45 kDa), SNAP43 (43 kDa), and SNAP19 (19 kDa) (14). The same SNAPc binds to the PSE of both human U1 and U6- type snRNA gene promoters and is involved in human snRNA transcription by pol II and pol III. Deletion of any subunit results in

reduced transcriptional activity, suggesting that each subunit plays a part in pol II and pol III snRNA gene transcription(15-18).

The architecture of SNAPc is shown in Figure I.2. SNAP190 is the largest subunit, serving as the backbone of the SNAPc complex. SNAP19 and SNAP45 associate toward the N- and C- terminus of SNAP190, respectively. SNAP43 associates with the same region of SNAP190 as SNAP19. SNAP50 joins the complex by associating with SNAP43 (6,9,19).

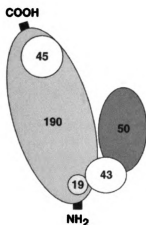


Figure I.2 Architecture of SNAPc complex (adopted from Hernandez et al(19)).

UV cross-linking experiments combined with immunoprecipitation indicate that within SNAPc, SNAP50 (15) and SNAP190 (17) are both in close proximity to the DNA. However, SNAP50 alone does not bind to PSE-containing probes(15), suggesting that SNAP190 is indispensable for DNA-binding of SNAPc.

SNAP190 bears 1469 residues and there is a Myb-DNA DNA-binding domain on its N-terminus (263-503). The rSNAPc that lacks the last two-thirds of SNAPc is still able to bind DNA and initiate transcription (9). In contrast, the rSNAPc without the N-terminus of SNAP190, where the Myb domain is located, completely loses the DNA-binding ability (18). Thus, the N-terminal Myb domain appears to be responsible for the DNA binding of SNAPc.

SNAP190 interacts with other SNAPc subunits at its two ends; its N-terminus interacts with SNAP19, SNAP43, and SNAP50 whereas its C-terminus interacts with SNAP45(19). The area between these two interacting regions does not mediate any subunit-subunit interactions and is solely used to interact with DNA (9) and other transcription factors, such as TBP(20) and Oct-1(18) (Figure I.3).

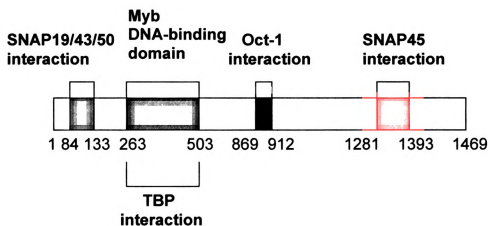


Figure I.3 The composition of SNAP190.

The interaction of SNAPc with TBP was first indicated by the observation that a sub-stoichiometric but detectable level of TBP copurified with SNAPc from human HeLa cells, which suggests their association *in vivo* (14). Electromobility shift assay (EMSA) experiments show that TBP enforces the DNA binding of SNAPc and *vice versa* (20). Such cooperative binding between TBP and SNAPc not only facilitates the SNAPc binding to PSE, but also, more importantly, ensures that TBP binds specifically to the promoter region. The DNA-binding of TBP is not as specific as its name might indicate and its dissociation from DNA is slow(21). The protein-protein interaction between SNAPc and TBP prevents TBP from binding irrelevant A/T rich sequences. TBP consists

of a highly conserved C-terminal DNA binding domain and an unconserved N-terminal regulation domain(22). SNAP190 directly interacts with the DNA-binding domain of TBP. Truncated SNAP190 containing only the Myb DNA binding domain is sufficient for TBP box in human U6 snRNA promoters(20).

The initiation of snRNA pol II transcription requires the assembly of transcriptional factors TFIIA, -B, -D, -E, -F and -H on the core promoter(2,23,24). In the most general case the orderly process of transcription initiation begins with TFIID recognizing and binding tightly to the TATA box(2). The TFIID-TATA box complex directs accretion of the remaining transcription factors to the human snRNA pol II promoter. TBP and TBP-associated factors, Bdp1 and Brf2, constitute the TFIIB complex and nucleated preinitiation complex assembly on the human snRNA pol III genes(25-27). Mutating either TATA box or PSE which is located at a fixed distance from the TATA box in the RNA pol III snRNA promoter switches transcription to RNA polymerase II(11,12). This suggests that the TATA box, together with the PSE, determines the assembly of a pol III-specific pre-initiation complex. Considering the cooperative binding between TBP and the PSE-binding factor, SNAPc, the interaction between the TBP C-terminal DNA binding domain and the SNAP190 Myb DNA binding domain may play a determinant role in assembling an RNA polymerase type-specific preinitiation complex.

I.1.4 SNAP190RcRd and Myb Domain

The SNAP190 Myb DNA binding domain is located at the N-terminus of SNAP190 and is important for SNAPc DNA binding(9). The recombinant SNAPc (rSNAPc) lacking the N-terminal SNAP190 cannot bind the PSE(9). The SNAP190 Myb DNA-binding domain also interacts with the DNA-binding domain of TBP and

stimulates its recruitment to the neighboring TATA box present in human U6 snRNA promoters(20). The preinitiation complex made of SNAPc and TBP then anchors RNA pol III to the promoters to initiate transcription(20).

The Myb domain was originally identified and defined in MYB proteins(28). The SNAP190 Myb domain is an unusual Myb domain in that it is not in a MYB protein and it has four and half repeats whereas the typical plant and yeast Myb domains have two repeats and typical animal Myb domains have three repeats (28,29).

MYB proteins are a group of transcription factors encoded by various *myb* genes and exist widely in organisms from fungi, yeast, plants, insects and humans(29). Three types of MYB protein, c-Myb, A-Myb, and B-Myb, are expressed in higher vertebrates and they have completely different biological roles as suggested by microarray experiments(30,31). In human, the c-Myb protein is a DNA binding transcriptional activator that induces transcription of a group of target genes and regulates both proliferation and apoptosis of hematopoietic cells(30). c-Myb is the most studied MYB protein because it is widely involved in the regulation of secondary metabolism, cellular morphogenesis, cell cycle, development, signal transduction, and disease resistance (32,33). The DNA binding domain of the c-Myb protein is located at the N-terminus and specifically binds the DNA sequence 5'-AACNG-3'. This DNA binding domain is well conserved among Myb families (c-Myb, A-Myb, and B-Myb) and between species from *Drosophila* to human(28). This conserved DNA binding domain is a characteristic of the Myb protein and so is called the "Myb domain".

The animal c-Myb domain is a stretch of 155 amino acids containing three

imperfect tandem repeats, R1, R2, and R3, each consisting of 51-52 amino acids) (34). The function of R1 is unknown. R2 and R3 together are responsible for the recognition of the specific DNA sequence but neither can interact with DNA separately (Figure I.4) (35). Most plant and yeast c-Myb domains do not have R1 repeats and only have R2 and R3 repeats (32,33). Nuclear magnetic resonance (NMR) studies of R2R3 repeats complexed with DNA revealed that both R2 and R3 contain three helices and the third helix of each is the recognition helix (36). The later *X-ray* structural studies of R1R2R3 repeats complexed with DNA and the enhancer binding protein is consistent with that NMR result and clearly shows that R2 and R3 are closely packed and intercalate in the major groove of DNA (Figure I.4)(37). The two recognition helices contact each other directly and bind to the specific base sequence cooperatively (37). In particular, the conserved Myb residues such as the K128 in the R2 repeat, and the N179, K182, and N183 in the R3 repeat, interact with the start AAC and the fifth G residue and determine the base pair recognition (Figure I.5).

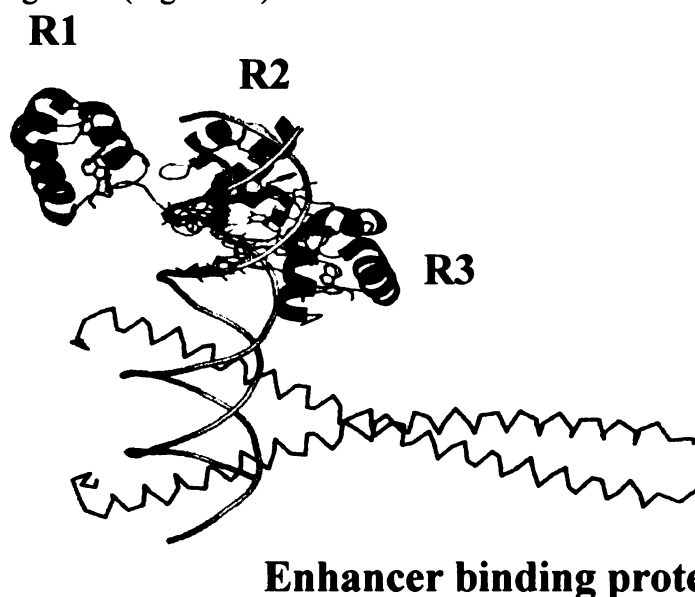


Figure I.4 Ternary human c-Myb protein-enhancer binding protein-DNA complex (1H88.pdb). The Myb protein (39-190) is shown as cartoon and three repeats R1, R2, and R3 are in green, yellow, and blue, respectively.

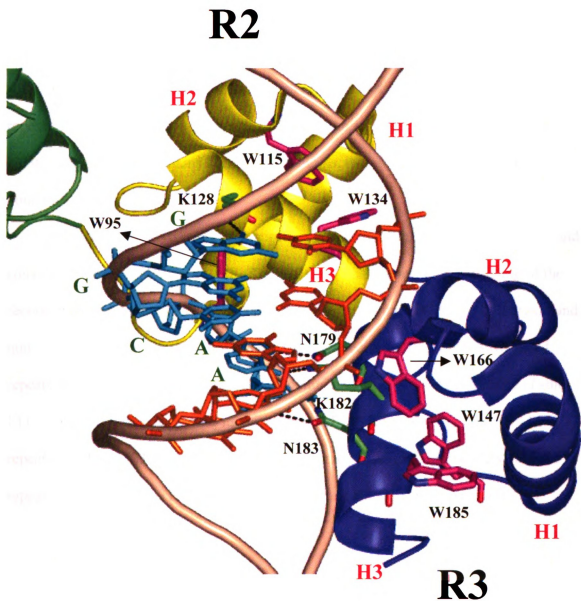


Figure I.5 The close-up view of the interaction between human c-Myb protein repeats R2, R3 (each constituting three helices H1, H2, and H3) and DNA (containing characteristic Myb-binding sequence 5'-AACNG-3') in the ternary protein-DNA complex (1H88.pdb). The conserved tryptophan residues are shown as magenta sticks.

All Myb -domain repeats contain three tryptophan residues that are highly conserved and regularly spaced 18 or 19 amino acids apart (29,38-40). They participate in the hydrophobic core that stabilizes the three-helix structure of Myb domain repeats(29). Based on these characteristic tryptophan residues, sequence analysis has identified a few Myb domains in non-myb proteins, including the SNAP190 Myb domain(9).

The SNAP190 Myb region encompasses 241 amino acids (Trp263 to Gln503) and constitutes four complete repeats (RaRbRcRd) and a half repeat (Rh, consisting of the second and third helices)(18). The sequence alignment of the SNAP190 Myb region and human c-Myb repeats shows considerable similarity between SNAP190 RaRbRcRd repeats and the R2R3 repeats in the c-Myb domain, especially SNAP190 RcRd (Table I.1). SNAP190 Rc repeat is 42% identical to SNAP190 Rd repeat(41). The SNAP190 Rc repeat is 38 % and 23% identical to the Myb R2 and R3 repeats, respectively. The Rd repeat is 38 % and 30 % identical to the Myb R2 and R3 repeats, respectively (18).

SNAP190 Rh	263		WEKISNINFEGRSAEEIRKFWQNSEHPS-	291
SNAP190 Ra	292		INKQ WS REEEERLQAI--AAAHGLE-- WQ KIAEELGTS-RSAFQCLQ KFQ QHNSKA--	343
SNAP190 Rb	344		LKR KEWT -EEEDRMLTQLVQEMRVGSHIPYRRIVVYMEG--RDSMQLIY WT TKSLDPG-	398
SNAP190 Rc	399		LKKGY W APEEDAKLLQA--VAKYGEQD-- WFK IRIEEVPG--RSDAQCRDR YL RLHFS-	450
SNAP190 Rd	451		LKKGR WN LKEEEQLIEL--IEKYGVGH-- WAK IASELPH--RSGSQCLSK WK IMMGKKQ	503
A-Myb	R1	33	WNRV KW TRDEDDKLKKL--VEQHGTD-- WT LIAASHLQN--RSD FQ CQHR WQ KVLNPE-	84
B-Myb	R1	29	KCKV KW THEEDEQLRAL--VRQFGQD-- WK FLASHFPN--RTDQCCQY WL RLVLPD-	80
c-Myb	R1	38	LGKTR W TREEDEKLKKL--VEQNGTDD-- WK VIANYLPN--RTDVQCCQHR WQ KVLNPE-	89
A-Myb	R2	85	LIKGP WT KKEEDQRVIEL--VQYGPGR-- WS LIAKHLKG--RIG KQ CRER WH NHLNPE-	136
B-Myb	R2	81	LVKGP WT KKEEDQKVIEL--VKKYGTKQ-- WT LIAAKHLKG--RLG KQ CRER WH NHLNPE-	132
c-Myb	R2	90	LIKGP WT KKEEDQRVIEL--VQYGPGR-- WS VIAKHLKG--RIG KQ CRER WH NHLNPE-	141
A-Myb	R3	137	VKK SSWT -EEEDRIIYE--AHKRLGNR-- WAE IAKLLPG--RTD NSI KNH WNSTMRKVV	188
B-Myb	R3	133	VKK SCWT -EEEDRIICE--AHKVLGNR-- WAE IAKMLPG--RTD NAV KNH WNSTIKRKV	184
c-Myb	R3	142	VKK TSWT -EEEDRIIYQ--AHKRLGNR-- WAE IAKLLPG--RTD NAI KNH WNSTMRKVV	193
				
Consensus			L.K.. WT .EED.R.I.L--V.K.G...-- W ..IA..LPG--R.D.QC.. RW ...L.P.-	

Table I.1 The sequence alignment of SNAP190 Myb domain and human A-, B-, c-Myb domains.. Important DNA-interacting residues are in bold and underlined. Three well-defined helices in the human c-Myb domain are illustrated as tandem connected cylinders (H1, H2, and H3).

Previous results have suggested that the Myb domain of SNAP190 is vital for PSE binding by SNAP190 and the entire SNAPc complex(18). EMSA results reported by Mee Wa Wong et al in 1998 show that a truncated SNAP190 peptide containing only the RcRd repeats (SNAP190RcRd) can bind to the wild-type but not to the mutant PSE (sequences shown in Table I.2), indicating that the SNAP190 Myb domain Rc and Rd repeats are required and sufficient for PSE binding(18). In 1999, The same lab found that the binding of SNAPc containing SNAP190 Δ RcRd (lacking the RcRd) was reduced >90 fold as compared with the wild-type SNAPc whereas the SNAPc containing SNAP190 Δ RhRaRb (lacking the RhRaRb) bound nearly as efficiently to the PSE as wild-type SNAPc (9). Taken together, SNAP190 RcRd was suggested to be the minimal and critical DNA binding element that enables efficient and specific binding of SNAPc to the core promoter PSE.

Table I.2 The sequences of the PSEs present in the wild-type and mutant probes. Uppercase letters correspond to the PSE sequence from the mouse U6 promoter, with red-colored characters corresponding to mutations.

wild-type mouse U6	ggatccgaaacTCACCTAACTGTAAAGTaattgtgtttctt
mutant mouse U6	ggatccgaaacTCCCACTACCGGTCCAGTaattgtgtttctt

The specific DNA binding ability and the sequence similarity (especially the conserved Trp residues) between SNAP190 RcRd and the c-Myb domain functional R2R3 repeats suggest that they share similar tri-helical structures and a similar DNA recognition manner. Indeed, the consensus AACNG sequence recognized by the c-Myb domains is in agreement with the AACTG sequence of the mouse U6 PSE, one of the

highest affinity PSEs for SNAPc. However, the residues critical for DNA sequence recognition of the c-Myb domain are not present in the SNAP190 RcRd (in red in Table I.1), suggesting that repeats Rc and Rd may recognize a different specific sequence and employ a different DNA-identification mechanism. If that is the case, the match between c-Myb consensus AACNG and the mouse U6 PSE sequence AACTG would be incidental. However, this hypothesis needs to be verified by an NMR or X-ray structure of the SNAP190cd /DNA complex.

I.1.5 Objectives

The specific binding of SNAPc to the PSE is the primary step in the assembly of the preinitiation transcription complex on human snRNA promoters, which is critical for RNA polymerase recruitment. Previous studies suggested that the SNAP190 Myb domain, particularly the RcRd repeats, is responsible for PSE binding of SNAPc.

SNAPc and TBP interact with each other and cooperatively bind to their target DNA sequence PSE and TATA box. SNAP190Myb domain RcRd repeats mediate the SNAPc interaction with TBP and are sufficient for TBP recruitment to the TATA box. TBP bends DNA when it binds to DNA. It is of interest to know how SNAP190RcRd interacts with TBP and whether a conformational change would result from their interaction. Perhaps it is relevant to the specific RNA pol recruitment.

Single crystal X-ray diffraction was used in the hope to obtain structural information of the SNAP190 RcRd peptide alone, various SNAP190RcRd /PSE complexes, and SNAP190 RcRd/ TBP/ DNA complexes. The purpose of structural studies of SNAP190 RcRd and its PSE complexes was to reveal the DNA binding and recognition pattern of SNAP190RcRd, which has been suggested to account for the

specific PSE binding of SNAPc. The goal of structural studies of the SNAP190RcRd/TBP/DNA complex was to map out the protein-protein interactions (SNAP190RcRd-TBP) and the protein-DNA interactions (SNAP190RcRd-PSE, TBP-TATA box) in the preinitiation complex of human RNA-pol III type snRNA genes, which is sufficient to recruit RNA pol III onto promoters.

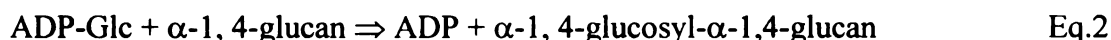
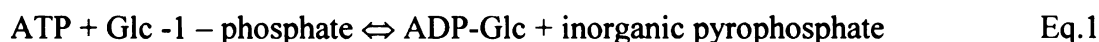
I.2 Glycogen Synthase

I.2.1 Glycogen

Glycogen is the main carbon source and energy storage molecule in bacteria and eukaryotic organisms. Glycogen molecules on average weigh 10^6 - 10^7 daltons, accounting for 5,550-50,000 glucose residues (42). Ninety percent of glucosyl units are linked by α -1, 4 glycosidic bonds and 10 % are linked by α -1, 6 glycosidic bonds. This huge glucose polymer constitutes an efficient way to store energy, glucose, by greatly reducing the large intracellular osmotic pressure that would result from its storage in monomeric form. In bacteria, glycogen is consumed rapidly when organisms are exposed to unfavorable conditions, and those rich in glycogen usually live longer than those without this energy reserve (43). The glycogen stored in the human body is enough to provide sugar to blood for 24-36 hours during fasting.

Glycogen metabolism is a major topic in general biochemistry. While catabolism of these polymers is much better known from a structure/function point of view (the 3D-structure of glycogen phosphorylase is one of the major milestones in protein crystallography) (44-46), biosynthesis has lagged behind.

Glycogen phosphorylases catalyze the breakdown of glycogen to glucose-1-phosphate, which enters glycolysis to fulfill the energetic requirements of the organism(47,48). For a long time, glycogen biosynthesis was considered a simple reversed phosphorylase degradation reaction. In the 1950s Leloir and coworkers (49) discovered that glycogen biosynthesis is an independent process and nucleoside diphosphate sugars are the starting material rather than glucose-1-phosphate. Later, three individual enzymes were revealed to participate in bacterial glycogen biosynthesis. In *E.coli*, they are ADPGlc –pyrophosphorylase EC 2.7.7.27, glycogen synthase EC.2.4.1.21, and branching enzyme EC 2.4.1.18(50,51). ADPGlc –pyrophosphorylase produces the “activated sugar” donor ADPGlc from ATP and glucose-1-phosphate (eq.1). Glycogen synthase then transfers the glucosyl unit from ADPGlc to elongate the α -1,4 glucan chain (Eq.2). In the third step, the branching enzyme (BE) creates branching points through 1,6-linkage for every 8-12 α -1,4 linked glucose residues (Eq.3). Although *in vitro* the first reaction is freely reversible, *in vivo* the hydrolysis of inorganic pyrophosphate by inorganic pyrophosphatase and the use of ADPGlc in the second step makes the first step reaction practically irreversible in the direction of ADP-Glc synthesis (50-52).



Studies on bacterial mutants, either with a deficit or overproduction of glycogen, have demonstrated that the levels of the polyglucan in these organisms correlates with

their levels of ADPGlc pyrophosphorylase and glycogen synthase activity(51). Recently, the structure of ADPGlc –pyrophosphorylase from potato tuber (53) and structures of glycogen synthase from *Agrobacterium tumefaciens* (54)and *Pyrococcus abyssi* (55) were solved. However, the reported GS structures, which either have no ligand or only ADP bound in the active site, display the catalytically inactive, open form. The lack of a GS structure of the active form, especially one with the entire substrate/ product present in the active site constitutes a major obstacle to fully understanding glycogen metabolism, particularly the structure/function relationship of GS and the catalytic mechanism GS employs.

I.2.2 Starch

While animal and bacterial cells store glucose in the form of glycogen, plants use starch, which is chemically very similar but contains fewer branches. Starch constitutes most of the dry matter obtained from crop plants, which is the primary source of the caloric intake in the diet of humans. Starch is also becoming increasingly important as a renewable industrial biomaterial in today's environmentally aware society (56,57).

Unlike glycogen, essentially homologous and water-soluble, starch is an insoluble semicrystalline material made of two distinct polysaccharide fractions: amylopectin and amylose. Normal wheat starch usually consists of 20-30 % amylose and 70-80 % amylopectin(58). The major fraction amylopectin is composed of α -1,4-linked chains that are clustered together by 5 % α -1, 6 linkages. Amylose is composed of longer chains with less than 1 % α -1, 6 linkage(59,60).

The physical and chemical properties of starch are chiefly determined by the relative amounts of amylose and amylopectin, which is directly related to the specific

functionality and has great impact on their industrial applications(61). For example, high amylose starches are more transparent, flexible, of higher tensile strength and tend to gelatinize more quickly than regular starches, so it is widely used as gelling agents in food processes and photographic film production(61). High amylose starches help create crispness and hamper the penetration of cooking oils in fried foods, which leads to a decrease in fat intake by consumers. Starches with high amylopectin levels are broadly used in the paper and adhesive industries because of their binding and bonding properties(56). Since the physicochemical properties of starch are essentially based on the degree of branching and /or polymerization, the modification of starch biosynthetic enzymes may enable us to adjust the ratio of amylopectin and amylose as well as to produce novel starches with improved functionality. The desirable products would be polymers with features that are intermediate between current amylopectin and amylose, or more highly branched than amylopectin or have a higher molecular weight than amylose(56,57,61).

Like glycogen in bacteria, plant starch synthesis also proceeds via these three steps. 1) to produce ADPGlucose, 2) to transfer glucose from ADPGlucose to a glucan chain, 3) to create branches(50,51). The second step, the elongation of the glucan chain, is catalyzed by starch synthase. To date, at least four isoforms of starch synthase (SS) have been found in plants, granule-bound SS I (GBSSI), SSI, SSII, and SSIII (50). While the precise role of soluble isoforms SSI, SSII and SSIII is less clear, GBSSI has been shown to be required for the production of the amylose component of starch(62,63).

Bacterial glycogen synthase (~ 450 aa) and plant starch synthase (~ 770-1100 aa) share an overall 29 -34 % sequence identity. Both of them belong to the GT-35 family

(see section I.2.3) based on thread analysis while a similar three-dimensional structural organization is suggested. As the residues that were identified to be critical to SS activity by site-directed mutagenesis and kinetic studies were also found important for GS activity, a very similar catalytic mechanism is believed to govern both GS and SS.

I.2.3 Glycosyltransferase

A huge amount of sugar-attached biomolecules are present in almost all biological kingdoms and play roles ranging from energy storage, cell-wall construction, cell-cell interaction, signaling, to host –pathogen interactions. The production of glycosylated molecules generally relies on the sugar-attaching glycosyltransferases (GTs). As of 2007, over 7200 sequences have been identified to be GT related (64) and this number is expected to grow as more and more genomes are being sequenced (65). Based on sequence homology, GTs are grouped into 87 families according to the carbohydrate-active enzyme server CAZy (64) and bacterial GS and plant SS belong to the GT35 family.

The large number of GTs in nature is not so surprising considering the great diversity of sugar donor and acceptor involved in glycosyltranslation. GTs can transfer a wide variety of sugars, such as glucose, galactose, N-acetylgalactose, N-acetylglucosamine, etc., to sugar, protein, nucleotide or lipid. The sugar donors usually are the lipid –phosphate sugars or nucleotide-sugars where the phosphate group is a good leaving - group and can drive the reaction in favor of synthesis, considering that most reactions take place in aqueous solution and hydrolytic processes are usually favored.

In spite of the great number of GTs, only two canonical folds (GT-A and GT-B) have been demonstrated from the known GT three-dimensional structures. The seeming

“dullness” in GT protein organization indicates that all GTs may have evolved from a small number of progenitor sequences (66). A typical GT-A fold consists of a single domain in which parallel β -strands are flanked on either side by α -helices. In contrast, the GT-B fold contains two Rossmann-fold-like domains that are separated by a deep cleft (Figure I.6). Based on thread analysis, bacterial glycogen synthase is grouped into GT 35 with plant starch synthase and will probably adopt the same fold. Crystallographic structures of GS from *Agrobacterium tumefaciens* (54) and *Pyrococcus abyssi* (55) reported in 2004 and 2006 explicitly reveal that GS is a GT-B fold enzyme (Figure I.7).



Figure I.6 Top: Typical GT-A fold illustrated by protein SpsA. Bottom: Structure of the GT-B fold enzyme GtfB. A Mn^{2+} ion in GtfB active site is depicted as blue sphere. The fold name is after their initial observation in the SpsA (*Bacillus subtilis* glycosyltransferase(67)) and GtfB (DNA β -glucosyltransferase) structures (19).

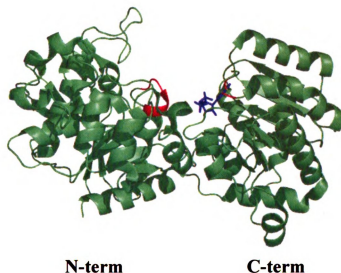


Figure I.7 Cartoon presentation of the structure of GS from *Agrobacterium tumefaciens* in complex with ADP (blue). The N-terminal peptide 15-21 containing the KTGL loop and Asp21 and the C-terminal residue Arg299 and Lys304 are colored red because they were indicated by mutagenesis and kinetic studies to be important in GS catalysis.

I.2.4 The Retaining Glycosyltransferase Mechanism Controversy

Glycosylation (the GT catalyzed sugar-attachment) proceeds with a high degree of stereo- and regio-selectivity. Errors during glycosylation could cause serious malfunction of glycosylated products, which is often linked to diseases such as neurodegeneration diseases, diabetes, and cancer (68). Depending on whether the product anomeric configuration is retained ($\alpha \rightarrow \alpha$) or inverted ($\alpha \rightarrow \beta$) with respect to the sugar donor, GTs are classified as retaining and inverting. It is noteworthy that the GTs' retaining/inversion feature is not correlated to their folds as the retaining GTs have been found in both the GT-A and GT-B fold, as have the inverting GTs.

Glycogen synthase is a retaining GT enzyme. Unlike inverting GT enzymes whose mechanism was well recognized to proceed via an S_N2 -like transition state (66,69), it is unclear which mechanism the retaining enzymes employ. Both double-displacement and S_Ni mechanisms (internal return) have been proposed for retaining GT enzymes in the past, but neither of them possesses solid and persuasive evidence over the other.

The double-displacement mechanism was first suggested by direct comparison to the retaining glycoside hydrolase (GHs), which includes branching enzyme, isoamylase, and cyclomaltodextrin glucanotransferase (70). GHs catalyze similar steps to GTs, the sugar dissociation and nucleophilic addition, but in GH the nucleophile is simply water, not a sophisticated acceptor nucleophile from sugar, protein, etc. (71). The double-displacement mechanism involves a catalytic nucleophilic attack from the β -face of the nucleotide-sugar donor and the formation of a covalently bound glucosyl – enzyme intermediate. Afterwards, the activated acceptor undertakes a nucleophile attack from the α - face of sugar and the retention of the anomeric configuration is achieved (Figure I.8).

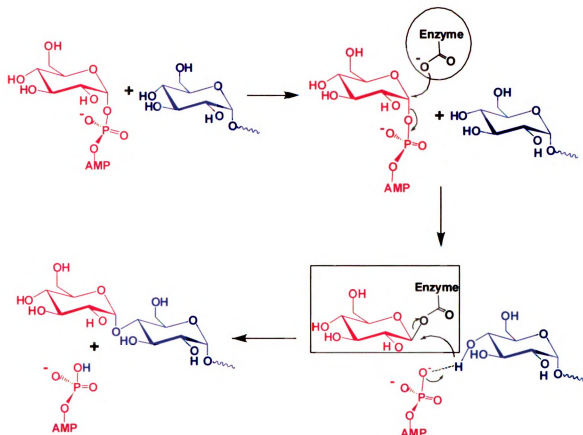


Figure 1.8 Proposed double-displacement mechanisms for GT retaining enzymes. Two essential components of this mechanism, the catalytic nucleophile and glucosyl-enzyme intermediate, are circled and framed, respectively.

In the case of the well-studied retaining glycosyl hydrolase (GH) enzymes, the evidence of double-displacement mechanism, a covalent bound glucosyl enzyme intermediate, was trapped and identified by both mass spectrometric and *X*-ray crystallographic characterization (66,72). Many attempts have been made to prove the existence of such an intermediate in the GT enzyme case in the last twenty years, but have never been successful. Recently it was reported that a galactosyl moiety covalently bound at Asp190 was observed by electrospray mass spectrometry of retaining GT-A enzyme LgtC Q189E (73). This piece of evidence initially appeared convincing and

attracted wide attention, but soon the author found that the Asp190, is 8.9 Å away from the anomeric center in either wild -type LgtC or its Q189E mutant, which is too far to function as a catalytic nucleophile if there were not a large geometric rearrangement at the active site. Moreover, in those GT-retaining enzymes whose structures have been determined, including glycogen phosphorylase(71,74), the first structure determined with a GT-B fold, no convincing potential catalytic nucleophile, a prerequisite for the double-placement mechanism, is found within an effective distance to the reaction center. In short, conclusive evidence for the classical double-displacement mechanism in retaining GTs has not been found.

The remaining option for the retaining GT mechanism is an S_N1 mechanism with some steric barrier at the β-side of the sugar. But such a mechanism involves an oxocarbenium cation intermediate (75). A series of solution experiments have indicated such electron deficient species have approximately a 10⁻¹² s half life which is not long enough for the sugar acceptor to undertake a directed attack (76-78). Almost out of despair, an S_Ni mechanism was proposed, in particular for GP, in response to the apparent lack of an appropriately positioned nucleophile (74,75). The S_Ni mechanism features not an oxocarbenium intermediate, but a transition state with significant oxocarbenium ion character (Figure I.9). In this mechanism, attack of the nucleophile occurs on the same face as the departure of the leaving group, and at nearly the same time, thus avoiding a discrete cation intermediate.

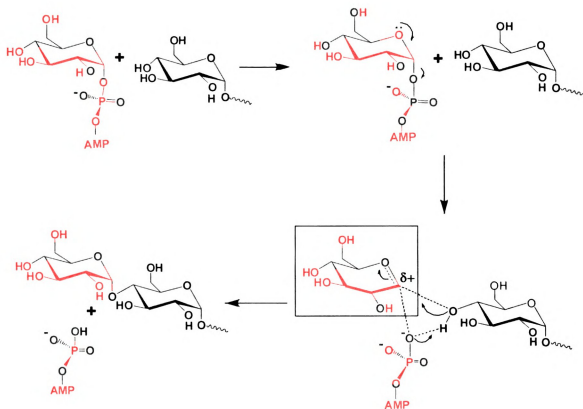


Figure I.9 S_N1 -like mechanism D-glucopyranosylium ion (DGM) intermediate is framed, which is otherwise considered a transition-state in the S_{Ni} -like mechanism. In both cases, the positively charged DGM is stabilized by the leaving group AMP- phosphate and the incoming nucleophile 4-OH group of SHPRT

The characteristic feature of the proposed S_{Ni}/S_N1 mechanism is the transition state/intermediate D-glucopyranosylium (DGM), and the stabilization of the oxocarbenium cation from the sugar donor phosphate and the attacking anionic nucleophile. Information derived from inhibitor studies has indicated the presence of these S_{Ni}/S_N1 mechanistic features for GT-retaining enzymes. A series of DGM analogues that either resembles the sp^2 -hybridized anomeric center and the consequential partly planar conformation of the glucopyranosyl ring or mimics the build-up and distribution of positive charge of the oxocarbenium ion have shown varied inhibiting

effects to GT-retaining enzymes (Figure I.10). Moreover, the adduct of the DGM analogue and phosphate has also exhibited a strong inhibiting effect on GT-retaining enzymes, such as D-gluconic acid 1,5-lactone and phosphate group in *Corynebacterium callunae* starch phosphorylase (79), nojirimycin tetrazole and phosphate in rabbit muscle glycogen phosphorylase (80), orthovanadate (phosphate analogue) and the nucleophile/leaving group in *Schizophyllum commune* trehalose phosphorylase (81). This indicates the involvement of phosphate in the DGM intermediate / transition state in these GT-retaining enzymes. In addition, the participation of an incoming acceptor nucleophile in the stabilization of the DGM intermediate /transition state has been implicated in the striking inhibition of validoxylamine A to *Schizophyllum commune* trehalose phosphorylase (K_i : 1.7 μ M, Figure I.10 G). Validoxylamine A not only mimics the positive charge build-up in DGM, but also mimics the critical ternary complex upon binding to the enzyme-phosphate complex (82).

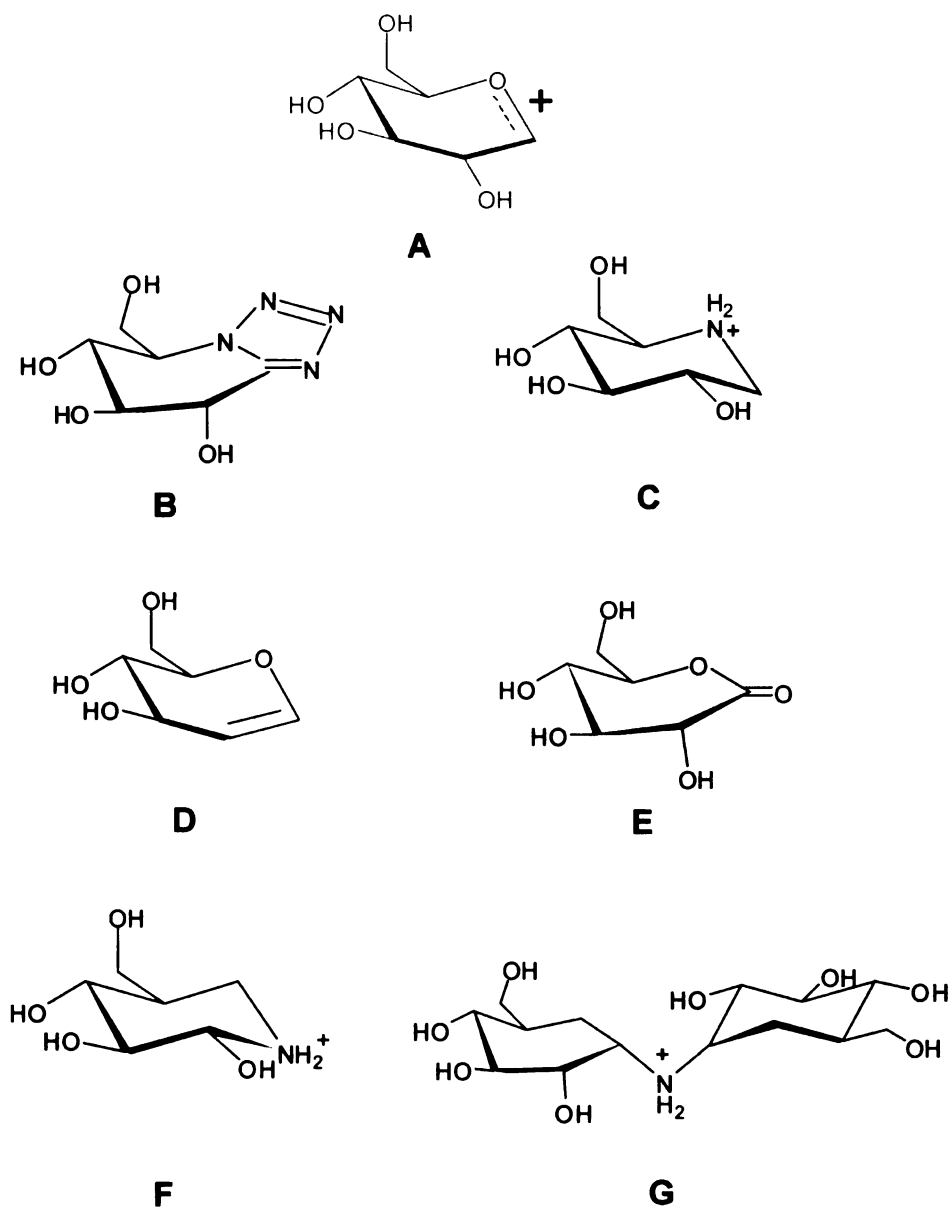


Figure I.10 DGM (A) and DGM mimics (B-G) that have been tested in kinetic studies of GT-B retaining enzymes. The related kinetic data are listed in Table I.3.

Table I.3 Kinetic data of DGM mimics in GT-B retaining enzymes

DGM and its mimics	Label in Figure I.10	K_i (μM)	GT-B retaining enzymes
D -glucopyranosylium (DGM)	A		
Nojirimycin tetrazole	B	700	Rabbit muscle GP(80)
1-deoxynorjirimycin	C	1200	<i>Schizophyllum commune</i> Trehalose phosphorylase(81)
D-glucal	D	320	<i>Schizophyllum commune</i> Trehalose phosphorylase(83)
D-gluconic acid 1, 5 -lactone	E	90	<i>Corynebacterium callunae</i> Starch phosphorylase(79)
		380	<i>E.coli</i> GS(84)
		920	Rabbit muscle GP(85)
Isofagomine	F	56	<i>Schizophyllum commune</i> Trehalose phosphorylase(81)
Validoxylamine	G	1.7	<i>Schizophyllum commune</i> Trehalose phosphorylase(82)

Although the above kinetic evidence apparently supports an S_{Ni} / S_{N1} mechanism, there is no definitive and direct experimental evidence to verify it, especially the oxocarbenium cation transition state /intermediate. As a result, the S_{Ni} / S_{N1} mechanism has not been firmly established for retaining GTs yet.

I.2.5 Previous Studies on Bacterial Glycogen Synthase

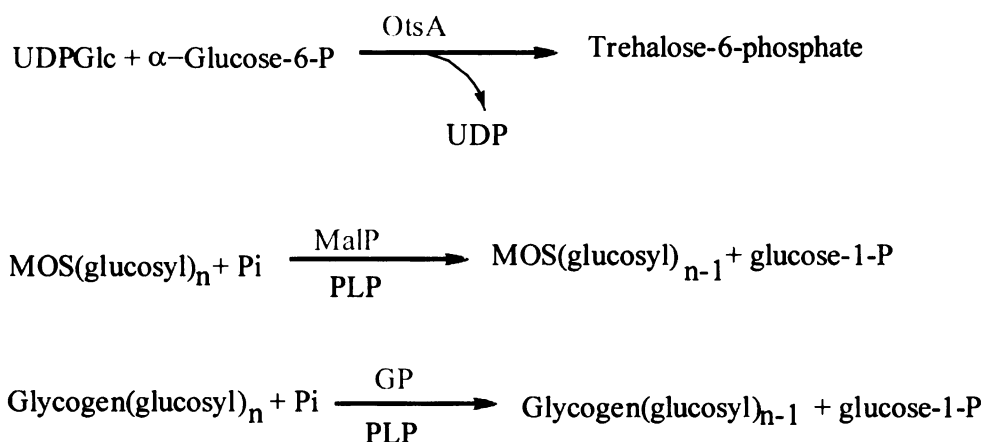
E.coli glycogen synthase (GS) is a 52.8 kDa protein containing 477 amino acid residues. The first successful GS separation from cell extract was done by Jack Preiss and his coworkers with *E.coli* (84). The cloning and expression of GS became available and widely used after the nucleotide sequence encoding *E.coli* GS, the *glgA* gene, was identified in the 1980s (86-88).

In vitro, *E.coli* GS is able to transfer glucose from ADPGlc to maltose, higher oligosaccharides, and rabbit liver glycogen with increasing efficiency, but not to glucose, even at a concentration as high as 1.4 M, indicating that a long glucan chain is preferred by glycogen synthase, which distinguishes GS from short-glucan chain elongation enzymes, such as MalP and OtsA.

Extensive chemical, mutagenesis, and kinetic studies have been performed on *E.coli* GS to identify the architecture of the reaction center in Preiss and Fukui's lab. Fukui first found that the ϵ -amino group of Lys15 is involved in ADPGlc binding and has an impact on GS activity(89). Lys15 is the first residue of the Lys-X-Gly- Gly motif, which is conserved in all glycogen synthases including human GS (X is a residue not conserved). The neighboring glycine residues were later found not only to affect substrate binding through interaction with Lys15, but also to participate in catalysis themselves(90).

E.coli GS has three cysteine residues, which were suggested to play two distinct roles: Cys379 is involved in substrate -binding and catalysis, whereas Cys7 and Cys408 are associated with glycogen binding (91,92). During their investigation with Cys379, Preiss's group found that the residue Glu377 is a critical catalytic residue because mutating it to Ala causes a 10, 000-fold decrease in enzyme activity(92). The impact of

Glu377 to *E.coli* GS stems mainly from its negative charge as a much smaller effect (57 fold decrease in enzyme activity) was observed when Glu377 is mutated to a similar, negatively charged residue aspartate. Mutagenesis and kinetic studies have also found Arg300, Lys305, His161 and Asp137 very important for GS activity (93). Interestingly, all these catalysis-related residues are highly conserved in GSs, SSs, and other GT-B retaining enzymes, such as OtsA, MalP and GP, indicating they all share a similar active site (Figure I.11).



OtsA: Trehalose-6-phosphate synthase

MOS: Malto-oligosaccharides

MalP: Maltodextrin phosphorylase

GP: Glycogen phosphorylase

PLP: Pyridoxal phosphate

Pi: Inorganic phosphate

Figure I.11 Reactions catalyzed by GT-B retaining enzyme OtsA, MalP, and GP.

Although several residues have been implicated in the substrate binding and GS catalytic performance, a full understanding of substrate specificity and the catalytic mechanism requires knowledge of the three-dimensional structure of the enzyme-substrate complex, especially the structure of the active form. In 2004, the first structure of GS from *Agrobacterium tumefaciens* was reported (54). It displays a typical GT-B fold and a wide cleft between two Rossmann fold domains. Since all available GT-B retaining enzyme structures show a strikingly identical active site, structural superimposition was performed with the AtGS structure and structures of OtsA, MalP, and GP. It reveals most of the critical catalytic residues are located in the interdomain cleft, but a huge displacement exists between those residues on the unaligned C-terminal domain. Therefore, a domain-domain closure movement from the observed AtGS organization was suggested to bring catalysis residues together and form a competent catalytic center in the interdomain cleft. Two years later, an apo- *Pyrococcus abyssi* GS structure demonstrated a similar open, inactive form of GS by which again only limited details regarding the active site were provided (55). Apparently, the essential features of the GS active site have not been revealed by currently available structures. The active, closed form GS structure is in great demand to elucidate the detailed GS active site architecture and its catalytic mechanism.

1.2.6 Significance of *E.coli* GS Structural Studies

Glycogen and starch are the major carbon and energy storage compounds in nearly all-living organisms. The central step in glycogen/starch biosynthesis involves adding thousands of glucose units from substrate ADPGlc to build long α -1, 4-glycogen/starch

backbones. Structural homologous enzymes glycogen synthase and starch synthase are responsible for this process in bacteria, animals, and plants, respectively.

Previous biochemical studies have identified a few residues related to GS/SS activity. Overall GS organization revealed by the recent AtGS and PaGS structures demonstrates a typical GT-B fold with twin-Rossmann fold domains. The ADP-bound AtGS structure represents the first GS structure with ligand in the active site, but the structural information regarding the sugar to be transferred (the glucose moiety of ADPGlc) and glucan acceptor is missing. Moreover, in comparison to other GT-B retaining enzymes, there is a substantial displacement between their active site residues and AtGS's, indicating the open form revealed in apo-PaGS, apo-AtGS and ADP-bound AtGS is not a catalytically active form.

I am interested in crystallizing the *bona fide* active form of GS. The atomic resolution of the GS structure, together with ample experimental data from biochemical studies on *E.coli* GS, will reveal the real active site architecture of the enzyme and elucidate the structural basis for substrate and acceptor binding. The anticipated active form of the GS structure will be compared to the previously reported GS structures to confirm whether a significant domain-domain closure occurs to transform GS from an inactive to an active conformation. The hypothesis that all GT-B retaining enzymes share a strikingly identical active site will therefore be examined.

Crystallization trials of a number of *E.coli* GS complexes with various combinations of substrate, substrate analogue and inhibitor will be initiated. The anticipated multiple ligand-bound GS structure will provide a revealing picture of the GS catalysis process, one step closer to nature's sweet secrets of synthesis. With luck, the

structures may even offer three-dimensional insights into the process of glycosidic bond formation of GS, or even the wider, GT-B retaining family. The most positive scenario would be some solid piece of evidence from our GS structures that would put an end to the ongoing mechanism debates and advance our understanding of the biosynthesis of glycosylated molecules with retention of configuration.

Both glycogen synthase and starch synthase (SS) belong to the GT-35 family and almost certainly adopt the same GT-B fold. GS and SS share significant sequence identity and a number of residues conserved between GS and SS are found to be critical to both enzyme activities. This fact means GS resembles SS a great deal, especially in terms of their active site core structure. An improved understanding of SS is expected from an active form GS structure and ultimately benefits the lucrative starch bioengineering field.

In animals, there is an evident link between glycogen metabolism and the physiopathological events. In particular, important diseases affecting man like Glycogen Storage Diseases and Type II diabetes mellitus are directly or indirectly linked to carbohydrate metabolism. Human and yeast GS are comprised of approximately 700 residues. The additional 220 residues compared to *E.coli* GS (477 residues) were suggested to form a regulating domain to control GS activity in response to metabolites(94). Most critical catalytic residues found in *E.coli* GS are also conserved in human GS, indicating that the basic catalytic center is shared between *E.coli* GS and human GS. The results from our structural studies of *E.coli* GS will therefore shed light on the active site of human GS, which would facilitate understanding human GS.

References

1. Voet, D., Voet, J.G. and Pratt, C.W. (1998) Fundamentals of biochemistry, 25, *John Wiley & Sons, Inc.*, Chichester, UK
2. Patikoglou, G. and Burley, S.K. (1997) Eukaryotic transcription factor-DNA complexes *Annual Review of Biophysics and Biomolecular Structure*, 26, 289-325
3. Wolberger, C. (1999) Multiprotein-DNA complexes in transcriptional regulation *Annual Review of Biophysics and Biomolecular Structure*, 28, 29-56
4. Dahlberg, J.E. and Lund, E. (1998) Structure and function of major and minor small nuclear ribonucleoprotein particles, 38-70
5. Schramm, L. and Hernandez, N. (2002) Recruitment of RNA polymerase III to its target promoters *Genes & Development*, 16, 2593-2620
6. Hernandez, N. (2001) Small nuclear RNA genes: A model system to study fundamental mechanism of transcription *J. Biol. Chem.*, 276, 26733-26736
7. Henry, R.W., Ford, E., Mital, R., Mittal, V. and Hernandez, N. (1998) in *Cold spring harbor symposia on quantitative biology*, Vol. LXIII. 111-120
8. Mittal, V., Cleary, M.A., Henry, W. and Hernandez, N. (1996) The Oct-1 POU-specific domain can stimulate small nuclear RNA gene transcription by stabilizing the basal transcription complex SNAPc *Molecular and Cellular Biology*, 16, 1955-1965
9. Mittal, V., Ma, B. and Hernandez, N. (1999) SNAPc: A core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain *Genes & Development*, 13, 1807-1821
10. Hernandez, N. (1992) in *Transcriptional regulation* (McKnight and Yamamoto), Vol. 11. 290, *Cold spring harbor laboratory press*, Plainview
11. Das, G., Hinkley, C.S. and Henry, W. (1995) Basal promoter elements as a selective determinant of transcriptional activator function. *Nature*, 374, 657-660

12. Lobo, S.M. and Hernandez, N. (1989) A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. *Cell*, 58, 55-67
13. Zhao, X., Pendergrast, P.S. and Hernandez, N. (2001) A positioned nucleosome on the human U6 promoter allows recruitment of SNAPc by the Oct-1 POU domain. *Molecular Cell*, 7, 539-549
14. Henry, R.W., Sadowski, C.L., Kobayashi, R. and Hernandez, N. (1995) A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerases II and III *Nature*, 374, 653-657
15. Henry, R.W., Ma, B., Sadowski, C.L., Kobayashi, R. and Hernandez, N. (1996) Cloning and characterization of SNAP50, a subunit of the snRNA-activating protein complex SNAPc. *EMBO J.*, 15, 7129-7136
16. Sadowski, C.L., Henry, R.W., Kobayashi, R. and Hernandez, N. (1996) The SNAP45 subunit of the small nuclear RNA (snRNA) activating protein complex is required for RNA polymerase II and III snRNA gene transcription and interacts with the TATA box binding protein *Proc. Natl. Acad. Sci. U. S. A.*, 93, 4289-4293
17. Yoon, J.B., Murphy, S., Bai, L., Wang, Z. and Roeder, R.G. (1995) Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes *Molecular and Cellular Biology*, 15, 2019-2027
18. Wong, M.W., Henry, R.W., Ma, B., Kobayashi, R., Klages, N., Matthias, P., Strubin, M. and Hernandez, N. (1998) The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with Oct-1 *Molecular and Cellular Biology*, 18, 368-377
19. Ma, B. and Hernandez, N. (2001) A map of protein-protein contacts within the small nuclear RNA-activating protein complex SNAPc *J. Biol. Chem.*, 276, 5027-5035
20. Hinkley, C.S., Hirsch, H.A., Gu, L., Lamere, B. and Henry, R.W. (2003) The small nuclear RNA-activating protein 190 Myb DNA binding domain stimulates TATA box-binding protein-TATA box recognition *J. Biol. Chem.*, 278, 18649-18657

21. Coleman, R.A. and Pugh, B.F. (1995) Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA *Journal of Biological Chemistry*, 270, 13850-13859
22. Davidson, I. (2003) The genetics of TBP and TBP-related factors *Trends in Biochemical Sciences*, 28, 391-398
23. Zawel, L. and Reinberg, D. (1995) Common themes in assembly and function of eukaryotic transcription complexes *Annual review of biochemistry*, 64, 533-561
24. Kuhlman, T.C., Cho, H., Reinberg, D. and Hernandez, N. (1999) The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter. *Molecular and cellular biology*, 19, 2130-2141
25. Cabart, P. and Murphy, S. (2001) BRFU, a TFIIB-like factor, is directly recruited to the TATA-box of polymerase III small nuclear RNA gene promoters through its interaction with TATA-binding protein. *The Journal of biological chemistry*, 276, 43056-43064
26. Mcculloch, V., Hardin, P., Peng, W., Ruppert, J.M. and Lobo-Ruppert, S.M. (2000) Alternatively spliced hBRF variants function at different RNA polymerase III promoters. *EMBO J.*, 19, 4134-4143
27. Willis, I.M. (2002) A universal nomenclature for subunits of the RNA polymerase III transcription initiation factor TFIIB *Genes & development*, 16, 1337-1338
28. Lipsick, J.S. (1996) One billion years of Myb *Oncogene*, 13, 223-235
29. Ogata, K., Tahirov, T.H. and Ishii, S. (2004) in *Myb transcription factors: Their role in growth, differentiation and disease* (Frampton, ed), Vol. 11. 223-238, *Kluwer Academic Publisher*
30. Ness, S.A. (2003) Myb protein specificity: Evidence of a context-specific transcription factor code *Blood Cells, Molecules, & Diseases*, 31, 192-200
31. Rushton, J.J., Davis, L.M., Lei, W., Mo, X., Leutz, A. and Ness, S.A. (2003) Distinct changes in gene expression induced by A-Myb, B-Myb and c-Myb proteins *Oncogene*, 22, 308-313

32. Martin, C. and Paz-Ares, J. (1997) Myb transcription factors in plants. *Trends in Genetics*, 13, 67-73
33. Jin, H. and Martin, C. (1999) Multifunctionality and diversity within the plant myb-gene family. *Plant Molecular Biology*, 41, 577-585
34. Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S. and Nishimura, Y. (1992) Solution structure of a DNA-binding unit of Myb: A helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proc. Natl. Acad. Sci. U. S. A.* , 89, 6428-6432
35. Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T.J. and Ishii, S. (1989) Delineation of three functional domains of the transcriptional activator encoded by the c-myb protooncogene *Proc. Natl. Acad. Sci. U. S. A.* , 86, 5758-5762
36. Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*, 79, 639-648
37. Tahirov, T.H., Sato, K., Ichikawa-Iwata, E., Sasaki, M., Inoue-Bungo, T., Shiina, M., Kimura, K., Takata, S., Fujikawa, A., Morii, H., Kumasaka, T., Yamamoto, M., Ishii, S. and Ogata, K. (2002) Mechanism of c-Myb-c/ebp beta cooperation from separated sites on a promoter. *Cell*, 108, 57-70
38. Majello, B., Kenyon, L.C. and Dalla-Favera, R. (1986) Human c-myb protooncogene: Nucleotide sequence of cDNA and organization of the genomic locus *Proc. Natl. Acad. Sci. U. S. A.* , 83, 9636-9640
39. Slamon, D.J., Boone, T.C., Murdock, D.C., Keith, D.E., Press, M.F., Larson, R.A. and Souza, L.M. (1986) Studies of the human c-myb gene and its product in human acute leukemias *Science*, 233, 347-351
40. Nomura, N., Takahashi, M., Matsui, M., Ishii, S., Date, T., Sasamoto, S. and Ishizaki, R. (1988) Isolation of human cDNA clones of myb-related genes, a-myb and b-myb *Nucleic Acids Res.* , 16, 11075-11089
41. Tatusova, T.A. and Madden, T.L. (1999) Blast 2 sequences - a new tool for comparing protein and nucleotide sequences *FEMS Microbiol Lett.*, 174, 247-250

42. Geddes, R., Harvey, J.D. and Wills, P.R. (1977) The molecular size and shape of liver glycogen *Biochem. J.* , 163, 201–209
43. Strange, R.E. (1968) Bacterial "glycogen" and survival *Nature*, 220, 606 - 607
44. Sprang, S.R., Goldsmith, E.J., Fletterick, R.J., Withers, S.G. and Madsen, N.B. (1982) Catalytic site of glycogen phosphorylase: Structure of the T state and specificity for α -D-glucose. *Biochemistry*, 21, 5364-5371
45. Withers, S.G., Madsen, N.B., Sprang, S.R. and Fletterick, R.J. (1982) Catalytic site of glycogen phosphorylase: Structural changes during activation and mechanistic implications. *Biochemistry*, 21, 5372-5382
46. Mclaughlin, P.J., Stuart, D.I., Klein, H.W., Oikonomakos, N.G. and Johnson, L.N. (1984) Substrate-cofactor interactions for glycogen phosphorylase b: A binding study in the crystal with heptenitol and heptulose 2-phosphate. *Biochemistry*, 23, 5862-5873
47. Johnson, L.N. (1992) Glycogen phosphorylase: Control by phosphorylation and allosteric effectors *FASEB J.*, 6, 2274-2282
48. Johnson, L.N., Hu, S.H. and Barford, D. (1992) Catalytic mechanism of glycogen phosphorylase *Faraday Discussions*, 93, 131-142
49. Leloir, L.F. (1971) Two decades of research on the biosynthesis of saccharides *Science*, 172, 1299-1303
50. Preiss, J. and Sivak, M. (1999) in *Comprehensive natural products chemistry* (Pinto, edition), Vol. 3. 441-495, *Elsevier Science B.V., Amsterdam, Neth*
51. Iglesias, A.A. and Preiss, J. (1992) Bacterial glycogen and plant starch biosynthesis *Biochemical Education*, 20, 196-203
52. Ballicora, M.A., Iglesias, A.A. and Preiss, J. (2003) ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis. *Microbiology and Molecular Biology Reviews*, 67, 213-225

53. Jin, X., Ballicora, M.A., Preiss, J. and Geiger, J.H. (2005) Crystal structure of potato tuber ADP-glucose pyrophosphorylase *EMBO J.*, 24, 694-704
54. Buschiazzo, A., Ugalde, J.E., Guerin, M.E., Shepard, W., Ugalde, R.A. and Alzari, P.M. (2004) Crystal structure of glycogen synthase: Homologous enzymes catalyze glycogen synthesis and degradation *EMBO J.*, 23, 3196-3205
55. Cristina, H., Guinovart, J.J., Fita, I. and Ferrer, J.C. (2006) Crystal structure of an archaeal glycogen synthase: Insights into oligomerization and substrate binding of eukaryotic glycogen synthases *J. Biol. Chem.*, 281, 2923-2931
56. Jobling, S. (2004) Improving starch for food and industrial applications *Curr. Opin. Plant Biol.*, 7, 210-218
57. Tharanathan, R.N. (2005) Starch — value addition by modification *Critical Reviews in Food Science and Nutrition*, 45, 371-384
58. Akaoka, M., Watanabe, S., Sassa, H., Yamamori, M., Nakamura, T., Sasakuma, T. and Hirano, H. (1997) Structural characterization of high molecular weight starch granule-bound proteins in wheat (*Triticum aestivum* L.) *Journal of Agricultural and Food Chemistry*, 45, 2929-2934
59. Buleon, A., Colonna, P., Planchot, V. and Ball, S. (1998) Starch granules: Structure and biosynthesis *International Journal of Biological Macromolecules*, 23, 85-112
60. Manners, D.J. (1989) Recent developments in our understanding of amylopectin structure *Carbohydrate Polymers*, 11, 87-112
61. Slattery, J.C., Kavakli, I.H. and Okita, W.T. (2000) Engineering starch for increased quantity and quality *Trends Plant Sci.*, 291-298
62. Mason-Gamer, R.J., Weil, C.F. and Kellogg, E.A. (1998) Granule-bound starch synthase: Structure, function, and phylogenetic utility *Molecular Biology and Evolution*, 15, 1658-1673
63. Ball, S.G. and Morell, M.K. (2003) From bacterial glycogen to starch: Understanding the biogenesis of the plant starch granule *Annual Review of Plant Biology*, 54, 207-233

64. [Http://Afmdb.Cnrs-Mrs.Fr/Cazy/Fam/Acc_GT.Html](http://Afmdb.Cnrs-Mrs.Fr/Cazy/Fam/Acc_GT.Html).
65. Coutinho, P.M., Deleury, E., Davies, G.J. and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases *J. Mol. Biol.*, 328, 307-317
66. Lairson, L.L. and Withers, S.G. (2004) Mechanistic analogies amongst carbohydrate modifying enzymes *Chem. Commun. (Cambridge, U. K.)*, 20, 2243-2248
67. Charnock, S.J. and Davies, G.J. (1999) Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms *Biochemistry*, 38, 6380-6385
68. Patenaude, S.I., Seto, N.O.L., Borisova, S.N., Szpacenko, A., Marcus, S.L., Palcic, M.M. and Evans, S.V. (2002) The structural basis for specificity in human ABO(H) blood group biosynthesis *Nature Structural Biology*, 9, 685 - 690
69. Ünligil, U.M. and Rini, J.M. (2000) Glycosyltransferase structure and mechanism *Curr. Opin. Struct. Biol.* , 10, 510-517
70. Coutinho, P.M. and Henrissat, B. (1999) in *Recent advances in carbohydrate bioengineering* (Gilbert, Davies, Henrissat and Svensson). 3-12, *The Royal Society of Chemistry*, Cambridge
71. Sinnott, M.L. (1990) Catalytic mechanisms of enzymatic glycosyl transfer *Chem. Rev. (Washington, DC, U. S.)*, 90, 1171-1202
72. Vocadlo, D.J., Davies, G.J., Laine, R. and Withers, S.G. (2001) Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate *Nature*, 412, 835-838
73. Lairson, L.L., Chiu, C.P.C., Ly, H.D., He, S., Wakarchuk, W.W., Strynadka, N.C.J. and Withers, S.G. (2004) Intermediate trapping on a mutant retaining a-galactosyltransferase identifies an unexpected aspartate residue *J. Biol. Chem.*, 279, 28339-28344
74. Klein, H.W., Im, M.J. and Palm, D. (1986) Mechanism of the phosphorylase reaction. Utilization of D-gluco-hept-1-enitol in the absence of primer. *European journal of biochemistry / FEBS*, 157, 107-114

75. Sinnott, M.L. and Jencks, W.P. (1980) Solvolysis of D-glucopyranosyl derivatives in mixtures of ethanol and 2, 2, 2-trifluoroethanol *J. Am. Chem. Soc.*, 102, 2026 - 2032
76. Banait, N.S. and Jencks, W.P. (1991) Reaction of anionic nucleophiles with α -D-glucopyranosyl fluoride in aqueous solution through a concerted, $A_ND_N(S_N2)$ mechanism *J. Am. Chem. Soc.*, 113, 7951-7958
77. Banait, N.S. and Jencks, W.P. (1991) General-acid and general -base catalysis of the cleavage of α -D-glucopyranosyl fluoride *J. Am. Chem. Soc.*, 113, 7958-7963
78. Chiappe, C., Moro, G.L. and Munforte, P. (1997) Lifetime of the glucosyl oxocarbenium ion and stereoselectivity in the glycosidation of phenols with 1, 2-anhydro-3, 4, 6-tri-O-methyl- α -D-glucopyranose *Tetrahedron*, 53, 10471-10478
79. Schwarz, A., Pierfederici, F.M. and Nidetzky, B. (2005) Catalytic mechanism of α -retaining glucosyl transfer by corynebacterium callunae starch phosphorylase: The role of histidine-334 examined through kinetic characterization of site-directed mutants *Biochem. J.*, 387, 437-445
80. Mitchell, E.P., Withers, S.G., Ermert, P., Vasella, A.T., Garman, E.F., Oikonomakos, N.G. and Johnson, L.N. (1996) Ternary complex crystal structures of glycogen phosphorylase with the transition state analog nojirimycin tetrazole and phosphate in the T and R states. *Biochemistry*, 35, 7341-7355
81. Nidetzky, B. and Eis, C. (2001) A-retaining glucosyl transfer catalysed by trehalose phosphorylase from *schizophyllum commune*: Mechanistic evidence obtained from steady-state kinetic studies with substrate analogues and inhibitors *Biochem. J.*, 360, 727-736
82. Goedl, C., Griessler, R., Schwarz, A. and Nidetzky, B. (2006) Structure–function relationships for schizophyllum commune trehalose phosphorylase and their implications for the catalytic mechanism of family GT-4 glycosyltransferases *Biochem. J.* , 397, 491–500
83. Eis, C., Watkins, M., Prohaska, T. and Nidetzky, B. (2001) Fungal trehalose phosphorylase: Kinetic mechanism, pH-dependence of the reaction and some structural properties of the enzyme from schizophyllum commune *Biochem. J.* , 356, 757–767

84. Fox, J., Kawaguchi, K., Greenberg, E. and Preiss, J. (1976) Biosynthesis of bacterial glycogen. Purification and properties of the *Escherichia coli* b ADPglucose:1, 4- α -D-glucan 4- α -glucosyltransferase *Biochemistry*, 15, 849-847
85. Papageorgiou, A.C., Oikonomakos, N.G., Leonidas, D.D., Bernet, B., Beer, D. and Vasella, A. (1991) The binding of D-gluconohydroximo-1, 5-lactone to glycogen phosphorylase. Kinetic, ultracentrifugation and crystallographic studies. *Biochem. J.*, 274
86. Kumar, A., Larsen, C.E. and Preiss, J. (1986) Biosynthesis of bacterial glycogen, primary structure of *Escherichia coli* ADP-glucose: α -1, 4-glucan, 4-glucosyltransferase as deduced from the nucleotide sequence of the *glgA* gene *J. Biol. Chem.*, 261, 16256-16259
87. Okita, T.W., Rodriguez, R.L. and Preiss, J. (1981) Biosynthesis of bacterial glycogen. Cloning of the glycogen biosynthetic enzyme structural genes of *Escherichia coli*. *J. Biol. Chem.*, 256, 6944-6952
88. Okita, T.W., Rodriguez, R.L. and Preiss, J. (1982) Isolation of *Escherichia coli* structural genes coding for the glycogen biosynthetic enzymes *Methods in Enzymology*, 83, 549-556
89. Furukawa, K., Tagaya, M., Inouye, M., Preiss, J. and Fukui, T. (1990) Identification of lysine 15 at the active site in *Escherichia coli* glycogen synthase, conservation of a Lys-X-Gly-Gly sequence in the bacterial and mammalian enzymes *J. Biol. Chem.*, 265, 2086-2090
90. Furukawa, K., Tagaya, M., Tanizawa, K. and Fukui, T. (1993) Role of the conserved Lys-X-Gly-Gly sequence at the ADP-glucose-binding site in *Escherichia coli* glycogen synthase *J. Biol. Chem.*, 268, 23837-23842
91. Holmes, E. and Preiss, J. (1982) Detection of two essential sulfhydryl residues in *Escherichia coli* b glycogen synthase. *Arch. Biochem. Biophys.*, 216, 736-740
92. Yep, A., Ballicora, M.A., Sivak, M.N. and Preiss, J. (2004) Identification and characterization of a critical region in the glycogen synthase from *Escherichia coli* *J. Biol. Chem.*, 279, 8359-8367

93. Yep, A., Ballicora, M.A. and Preiss, J. (2004) The active site of the *Escherichia coli* glycogen synthase is similar to the active site of retaining GT-b glycosyltransferases *Biochem. Biophys. Res. Commun.*, 316, 960-966
94. Cid, E., Gomis, R.R., Geremia, R.A., Guinovart, J. and Ferrer, J.C. (2000) Identification of two essential glutamic acid residues in glycogen synthase *J. Biol. Chem.*, 275, 33614-33621

Chapter II

CRYSTALLIZATION AND PRELIMINARY X-RAY DIFFRACTION ANALYSIS OF SNAP190RcRd

II.1 Experimental Procedures

II.1.1 Transformation and Overexpression of SNAP190RcRd

Human SNAP190RcRd (390-518) was overexpressed in *E.coli* cells as a GST-tag fusion peptide. The peptide sequence and the predicted secondary structure are shown in Table II.1 and Figure II.1, respectively. The plasmid was provided by Dr. William Henry (department of Biochemistry and Molecular Biology, Michigan State University) (1) and transformed into *E.coli* BL21-*Condon* Plus competent cell (Cat. No. 230245, Stratagene, La Jolla, CA.). The plasmid-bearing cells were then plated on ampicillin and chloramphenicol containing, agar Luria Broth (LB)-filled petri dishes and grown for 12 - 16 h at 37 °C. Colonies were picked from the plate and transferred into 50 mL Terrific Broth (TB) flasks containing 0.1 mg /mL ampicillin and 0.025 mg /mL chloramphenicol. The cells were grown at 37 °C at 250 rpm shaking for 8-12 h before being transferred to 1 L TB flasks containing the same concentration of ampicillin and chloramphenicol for further growth. When the cell broth optical density (O.D.) absorbance at 600 nm reached 0.6-0.8, the inducing agent IPTG was added to achieve a final concentration of 1 mM and the cells were grown for an additional 3 h at 16 °C before being collected by centrifugation at 5000 rpm.

II.1.2 Purification of SNAP190 RcRd

Cell pellets were resuspended in HEMGT 250 (15 mL/L cell culture)(composition see appendix). A protease inhibitor cocktail tablet (containing chymotrypsin, thermolysin, papain, pancreatic extract, and trypsin) (Cat. No.11-836-170-001, Roche Applied Science, Mannheim, Germany) was added to the mixture. After sonication, the lysate was spun down for 30 min and the presence of protein in the supernatant was confirmed by SDS-PAGE. The glutathione sepharose column (Cat. No. G4510, Sigma, St. Louis, MO) was equilibrated with 50 mL of HEMGT 250 buffer. The supernatant was then loaded into the column and gently shaken overnight for a thorough binding. The column was washed thoroughly with HEMGT 250 followed by 30 mL HEMGT 100, 20 mL TDB, and 10 mL TDB-DTT. A small sample of glutathione sepharose resin (approximate total volume 5-8 μ L) was used to run SDS-PAGE in order to check for the bound protein. Twenty units of thrombin (Sigma) were added to the resin to cleave the GST tag and release the protein SNAP190RcRd. After 2.5 hours of shaking, PMSF solution was added to achieve a final concentration of 0.5 mM in the resin to inhibit the thrombin activity. The protein without the GST tag was eluted with TDB-DTT. The SDS-PAGE gel is shown in Figure I.2.

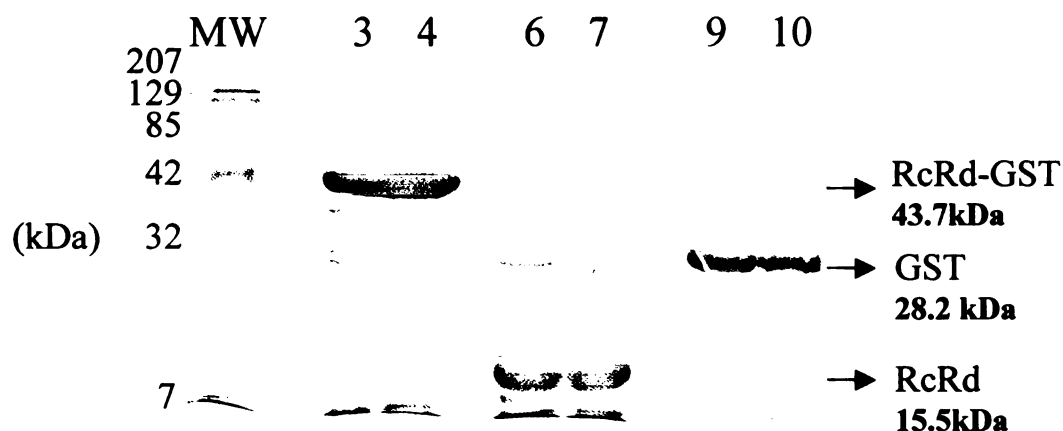


Figure II.2 SDS-PAGE gel showing the GST-tagged RcRd peptide (lane 3 and 4) and the RcRd peptide alone (lane 6 and 7). The GST-tag was eluted by 100 mM reduced glutathione solution afterwards (lane 9 and 10).

The peptide obtained from the previous GST purification step was buffer-exchanged to QA buffer (20 mM Tris-HCl pH 7.5, 5 % glycerol, 0.2 mM EDTA, and 1 mM DTT) and loaded onto a Source Q ion-exchange column (Cat. No.17-0947-01, GE, Piscataway, NJ). Pure SNAP190RcRd was eluted within the range of 50 % - 63 % mixture of QA and QB buffer (1M NaCl, 20 mM Tris-HCl pH 7.5, 5 % glycerol, 0.2 mM EDTA, and 1 mM DTT) (Figure II.3). The collected protein fractions were concentrated to ~ 4.5 mg/mL, pooled, and quick-frozen in liquid N₂ for storage at -80 °C. The SDS-PAGE gel of SNAP190RcRd eluted from the Source-Q column is shown in Figure II.4.

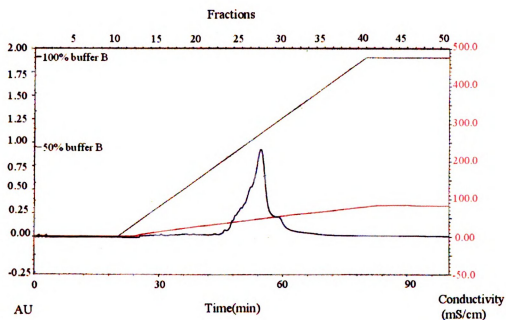


Figure II.3 Diagram of SNAP190RcRd purification through Source-Q column.

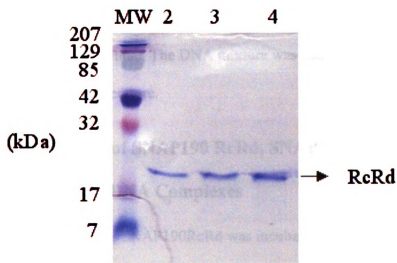


Figure II.4 SDS-PAGE gel of SNAP190RcRd from Source-Q column. Lane 2, 3, and 4 corresponds to the fraction 27, 28, and 29, respectively.

II.1.3 Purification of DNA and Annealing

Oligonucleotides were ordered from W. M. Keck Facility (Yale university) and were purified in an anion exchange column (Source Q-5, Pharmacia) on a Perkin Elmer

HPLC system with UV detection at 260 nm and 280 nm. One micromole of DNA was loaded onto the column with buffer A (10 mM NaOH and 0.2 M NaCl) and eluted with a shallow gradient of buffer B (10 mM NaOH and 1.0 M NaCl). Collected fractions were approximately 50-80 mL in total. After dilution with 4 column volumes of 10 mM Tris, the DNA solution was loaded onto a 1 mL DEAE cellulose column and the liquid and impurity was allowed to drip off the column completely. The purified DNA was then eluted with 2-3 mL of 1 M NaCl. Several cycles of concentration and dilution were utilized to lower the DNA salt concentration to 50 mM for the next step, annealing. Concentration of DNA was performed using a Centricon-3 Concentrator (Millipore Corporate, Billerica, MA).

Equimolar amounts of complementary DNA strands were added to the annealing buffer (50 mM NaCl and 10 mM Tris pH 7.0) and heated to their melting point (90 - 95 °C) in a water-bath for one minute. The DNA mixture was then left in the water bath and allowed to cool to room temperature.

II.1.4 Crystallization of SNAP190 RcRd, SNAP190RcRd / DNA and SNAP190RcRd / TBP / DNA Complexes

Prior to crystallization, SNAP190RcRd was incubated in a 1:1.2 molar ratio with double-stranded DNA and in a 1:1:1.2 molar ratio with yeast TBP and double-stranded DNA. Table II.2 lists all of the oligonucleotides utilized in the crystallization trials. Yeast TBP was provided by Dr. Bill Henry (Department of Biochemistry and Molecular Biology, Michigan State University). Crystallization was attempted using the hanging drop vapor diffusion method at both room temperature and 4 °C. The reservoir in the

crystallizing plates was about 1 mL and the primary screening solution kits (Crystal Screen™, Crystal Screen 2™) were purchased from Hampton Research (Aliso Viejo, CA). The hanging drop consisted of 2 µL SNAP190RcRd or SNAP190RcRd complexes and 2 µL reservoir solution.

Table II.2 Oligonucleotides used in crystallization trials

U1-18-blunt ends	TGACCGTGTGTGTAAAGA ACTGGCACACACATTTCT
U1-18single overhang	aTGACCGTGTGTGTAAAGA ACTGGCACACACATTTCTt
U1-18-double overhang	caTGACCGTGTGTGTAAAGA ACTGGCACACACATTTCTgt
U2-18-blunt ends	TCACCGCGACTTGAATGT AGTGGCGCTGAACTTACA
U2-18-single overhang	aTCACCGCGACTTGAATGT AGTGGCGCTGAACTTACAt
U2-18-double overhang	caTCACCGCGACTTGAATGT AGTGGCGCTGAACTTACAg
U2-23-double overhang	cacTCACCGCGACTTGAATGTgg gAGTGGCGCTGAACTTACAccgt
U4B-18-blunt ends	TCACCTTTGCGAAATAGG AGTGGAACGCTTTATCC
U4B-18-single overhang	aTCACCTTTGCGAAATAGG AGTGGAACGCTTTATCCt
U4B-18-double overhang	caTCACCTTTGCGAAATAGG AGTGGAACGCTTTATCCgt
U6-18-blunt ends	TTACCGTAACTTGAAAGT AATGGCATTGAACTTTCA
U6-18-single overhang	aTTACCGTAACTTGAAAGT AATGGCATTGAACTTTCA
U6-18-double overhang	caTTACCGTAACTTGAAAGT AATGGCATTGAACTTTCAg
U6-23-double overhang	cacTTACCGTAACTTGAAAGTat gAATGGCATTGAACTTTCAtagt

Table II.2 (cont'd)

H1-18-blunt ends	TCACCATAAACGTGAAAT AGTGGTATTTGCACTTTA
H1-18-single overhang	aTCACCATAAACGTGAAAT AGTGGTATTTGCACTTTA
H1-18-double overhang	caTCACCATAAACGTGAAAT AGTGGTATTTGCACTTTAgt
Mouse U6-21-blunt ends	ACTCACCCCTAACTGTAAAG
B-MPSE	TGAGTGGGATTGACATTTTC
Mouse U6-21-single overhang	aACTCACCCCTAACTGTAAAG
S-MPSE	TGAGTGGGATTGACATTTTCt
Mouse U6-21-double overhang	caACTCACCCCTAACTGTAAAG
D-MPSE	TGAGTGGGATTGACATTTTCgt
Mouse U6-22-double overhang	caCTCACCCCTAACTGTAAAGTa GAGTGGGATTGACATTTTCatgt
Mouse U6-27-double overhang A	caCTCACCCCTAACTGTAAAGTat ttcg GAGTGGGATTGACATTTTCataaagcgt
Mouse U6-27-double overhang B	caatatgCTCACCCCTAACTGTAAAGTa tatacGAGTGGGATTGACATTTTCatgt
Mouse U6-32-double overhang with human U6 flanker	caatatgCTCACCCCTAACTGTAAAGTat ttcg tatacGAGTGGGATTGACATTTTCataaagcgt
Mouse U6-32-double overhang with mouse U6 flanker	caggaaaCTCACCCCTAACTGTAAAGTaattgt cctttGAGTGGGATTGACATTTTCattaacagt

II.1.5 Electrophoretic Mobility Shift Assay (EMSA)

To radiolabel DNA probes, 250 ng single-stranded DNA was mixed with 1 μ L 10x T4 polynucleotide kinase buffer, 2 μ L [γ - 32 P] ATP, and 1 μ L T4 polynucleotide kinase (Amersham Pharmacia Biotech, Piscataway, NJ). The total volume was 10 μ L. The mixture was incubated at 37 °C for 1 h. The resulting γ - 32 P labeled DNA was purified on 5 % TGE gel and annealed with its complementary single strand DNA.

The DNA binding reactions were performed in a 20 μ l total volume. DNA binding reaction using only rSNAP_c or SNAP190RcRd were performed in a buffer containing 60

mM KCl, 20 mM HEPES pH 7.9, 5 mM MgCl₂, 0.2 mM EDTA, 10 % glycerol, 0.5 µg of poly (dI-dC), and 0.5 µg of pUC119 plasmid. The reactions were incubated for 30 min at room temperature after which 25,000 cpm of ³²P-labeled DNA probe was added and the reactions were then incubated an additional 30 min. The samples were loaded onto a 5 % nondenaturing polyacrylamide gel (acrylamide: bisacrylamide ratio 39:1). The EMSA (Electrophoretic Mobility Shift Assay) was then run in TGE running buffer (50 mM Tris, 380 mM glycine, 2 mM EDTA) at 150 V for 3-5 h at room temperature. Reactions containing both SNAP_c/ SNAP190RcRd and TBP were run in a buffer containing 100 mM KCl, 20 mM HEPES pH 7.9, 5 mM MgCl₂, 0.2 mM EDTA, 10 % glycerol, 1 mM dithiothreitol, 0.07 % Tween 20, 0.2 µg of poly (dG-dC), and 0.2 µg of pUC119 plasmid. Reactions with SNAP190 RcRd and TBP also contained 0.5 µL of fetal bovine serum as a protein carrier and 50 mM NaF but lacked KCl. The samples were loaded on a 5 % nondenaturing polyacrylamide gel (39:1) and EMSA was run in TGEM running buffer (50 mM Tris, 380 mM glycine, 2 mM EDTA, 5 mM MgCl₂) at 150 V for 3-5 h at room temperature. The gels were dried and autoradiographed.

II.2. Results and Discussion

II.2.1. Preparation of SNAP190RcRd

The human SNAP190RcRd plasmid was cloned into a T7-based vector for expression in *E.coli*. At first, the expression was tried on ordinary *E.coli* BL21 (DE3) competent cells, but was not successful. Forced high-level expression of a gene containing condons rarely expressed in *E.coli* can deplete the *E.coli* 's internal tRNA pools. Therefore *E.coli* BL21-*Condon* Plus competent cells (Stratagene, La Jolla, CA),

which contain supplemented codons, were specifically chosen to overcome the codon bias, which resulted in an efficient expression (800 colonies/ μg DNA).

In an effort to obtain a large amount of pure protein for crystallization screens, two different cell growth media were tested and terrific broth (TB) was chosen over luria broth (LB) as TB gains 4.5 mg of protein per liter of cell culture, triple the amount from LB. A detailed recipe for TB and LB are listed in the appendix. Compared to LB, growth media TB contains a higher amount of yeast extract, extra glycerol and potassium phosphate salts instead of NaCl. The yeast extract provided vitamins and trace elements for cell growth. The higher amount of yeast extract in TB allows higher protein yield per volume. Glycerol was added to the TB as an additional carbon and energy source. Potassium phosphate buffer were used instead of NaCl in LB to provide K^+ for cellular systems while maintaining an optimal pH that prevents cell death due to a drop in pH.

The size-exclusion chromatography of peptide SNAP190RcRd suggests that the molecular weight of SNAP190RcRd in solution is about 158kDa, if not higher (Figure II.5). The considerable aggregation state of SNAP190RcRd is likely to lead to a large discrepancy between the estimated and experimental pI values and may also hinder the peptide from binding to DNA.

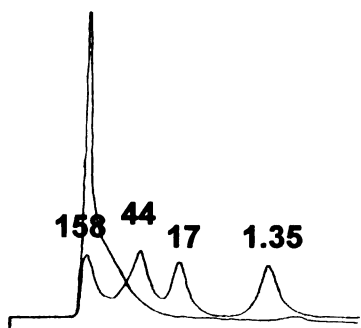


Figure II.5 Size-exclusion chromatogram results from Sephacryl S-300 HiPrep 16/60 column. The blue line is Bio-Rad molecular weight standards (158 kDa bovinegamma-globulin, 44 kDa chicken ovalbumin, 17 kDa equine myoglobin, and 1.3 kDa Vitamin B12). The red line is from SNAP190RcRd. Experiment was performed at 1mL/min flow rate and the fractions were 5 mL.

II.2.2. Crystallization Trials

Our crystallization trials include the SNAP190RcRd peptide alone as well as various SNAP190RcRd –PSE and SNAP190RcRd –PSE-TBP complexes. Unfortunately, none of these trials has produced diffraction-quality peptide or peptide complex crystals to date. Although some crystals were found in our crystallization trays from time to time, later the dye test (using Izit Crystal Dye, Hampton Research, Aliso Viejo, CA) and diffraction examination prove that they were all salt crystals from the crystallization buffer.

II.2.3. EMSA DNA-binding Assay

The failure to obtain SNAP190RcRd –PSE complex raised doubt about the binding ability of SNAP190RcRd to the PSE that was claimed in Wong's paper (3). Therefore, a series of EMSA SNAP190RcRd-PSE binding assays were carried out with the mouse U6 PSE, which has the highest affinity to the SNAP complex and was used in Wong's EMSA experiments. Unfortunately none of our assays show that SNAP190RcRd can bind to the mouse PSE, nor selectively bind wild type over mutant mouse U6 PSE. One of our EMSA gels is shown in Figure II.6.

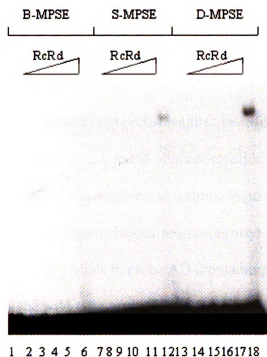


Figure II.6 An EMSA was performed with a probe containing B-MPSE (lane 1-6), S-MPSE (lane 7-12), or D-MPSE (lane 13-18). B-MPSE, S-MPSE, and D-MPSE are blunt ended, single overhang-, and double overhang- mouse U6-21 PSE, respectively, and their sequences are listed in Table II.2. In lane 1, 7, and 13, no SNAP190RcRd or recombinant SNAPc (rSNAPc) were added to the probes. Lane 2-5, 8-11, and 14-17 contain increasing amount of SNAP190RcRd (0.1, 0.3, 1, 3 μ g). Lane 6, 12, and 18 contain 3 μ g mSNAPc.

Mini SNAPc complex (mSNAPc) is composed of SNAP190 (1-505), SNAP50, SNAP43, and SNAP19. As Figure II.5 shows, the mSNAPc binds to the double-overhang mouse U6 PSE (D-MPSE) better than the single-overhang mouse U6 PSE (S-MPSE). The blunt-overhang mouse U6 PSE (B-MPSE) displays little binding. This result suggests that the neighboring base pair of the PSE is involved in mSNAPc binding. Therefore, we thought that adding flanking base pairs to the mouse U6 PSE may enhance its affinity to the peptide SNAP190RcRd. A series of mouse U6 PSE derivatives that

contain varied lengths of flanking base pairs at the 3'-, the 5'- or at both ends were used to perform EMSA assays to identify the optimized length of DNA for SNAP190RcRd binding.

The EMSA gel does not show the expected results. Instead, almost identical binding behavior was displayed for each tested oligonucleotide at high concentrations of SNAP190RcRd, regardless of the oligonucleotide lengths. Figure II.7 shows the EMSA gel and Table II.3 lists all of the oligonucleotide sequences used in the EMSA experiment. mSNAPc specifically binds to probe AD (containing the human U6 promoter whose PSE is replaced by the wild-type mouse U6 PSE and a mutated TATA box) but not to probe BD (containing the human U6 promoter whose PSE is replaced by a mutant mouse U6 PSE and a mutated TATA box). However, no apparent binding of mSNAPc to the oligonucleotides was observed, except for the smear underneath the loading position. Such a smearing phenomenon was also observed for the lower concentrations of SNAP190RcRd and probably is due to the relatively low ratio of labeled oligonucleotide to unlabeled ones. As the oligonucleotides we tested in the EMSA do not include a faithful flanker from the human U6 sequence (Table II.3), the possibility exists that the oligonucleotide that contains the mouse U6 PSE and an accurate flanker from the human U6 sequence may bind to SNAP190RcRd.

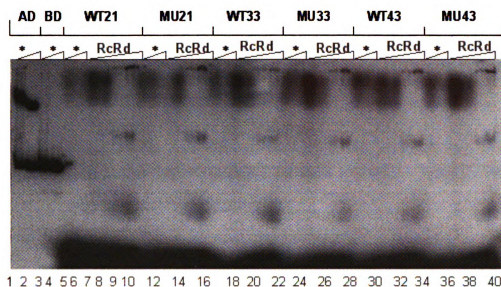


Figure II.7 An EMSA was performed with varied length of probe. mSNAPc was added to the * labeled lanes. Lane 1-2, 3-4, 11-12, 23-24, and 29-30,35-36 contain increasing amount of mSNAPc (1 and 3 μ g). All probe sequences are listed in Table II.3. Lane 7-10, 13-16, 19-22, 25-28,31-34, and 37-40 contain increasing amount of SNAP190RcRd (0.1, 0.3, 1, 3 μ g).

Table II.3 DNA sequences used in Figure II.6.

probe AD	agcggataacaatttcacacaggaaacagctatgacatgattacgaattcgagctcgg taccgggggatccgaaaCTCACCCTAACTGTAAAGTaattgtgtttcttggtTCTCGAGCCtt gggaagctggcactggcctcgtaaattg	
probe BD	agcggataacaatttcacacaggaaacagctatgacatgattacgaattcgagctcgg taccgggggatccgaaaCTCCCACTACCGGTCCAGTaattgtgtttcttggtTCTCGAGCCtt gaggaagctggcactggcctcgtaacaattg	
WT21	wild type mouse U6 double overhang	caCTCACCCTAACTGTAAAGT GAGTGGGATTGACATTTCagt
MU21	mutant mouse U6 double overhang	caCTCCCACTACCGGTCCAGT GAGGGTGATGGCCAGGTCagt
WT33	wild-type mouse U6 double overhang 12-base flanker	caatatgCTCACCCTAACTGTAAAGTgattcga tatacGAGTGGGATTGACATTTCActaagctgt
MU33	mutant mouse U6 double overhang 12 -base flanker	caatatgCTCCCACTACCGGTCCAGTgattcga tatacGAGGGTGATGGCCAGGTCActaagctgt
WT43	wild-type mouse U6 double overhang 22-base flanker	cactatcatatgCTCACCCTAACTGTAAAGTgattcgatttct gatagtatacGAGTGGGATTGACATTTCActaagctaaagagt
MU43	mutant mouse U6 double overhang 22-base flanker	cactatcatatgCTCCCACTACCGGTCCAGTgattcgatttct gatagtatacGAGGGTGATGGCCAGGTCActaagctaaagagt
human U6 PSE NCBI [gi:577b71]	251	gactatcatatgCTTACCGTAACTTGAAAGTatttcgatttct

Wild-type PSE in probe AD and mutant PSE in probe BD are in underscored uppercase.
The mutant TATA box in both probes are in italic uppercase. In other tested DNAs
as well as human U6 sequence, PSE is in uppercase and mutated bases in PSE are underlined.

II.3. Conclusion

The RcRd repeats of the N-terminal Myb domain of SNAP190 are required for the SNAPc complex to bind to the PSE, as the binding of SNAPc containing SNAP190 Δ RcRd (lacking the RcRd) was reduced >90 fold as compared to that of the wild-type SNAPc(4). Data reported by Wong et al. in 1998 further suggests that a truncated SNAP190 peptide containing only the RcRd repeats (SNAP190RcRd) can bind to the wild-type but not to the mutant PSE, indicating that the SNAP190 Myb domain Rc and Rd repeats are required and sufficient for PSE binding(3).

To investigate the interaction between SNAP190RcRd and the PSE, we performed crystallization trials with a combination of peptide 390-518 (encompassing the RcRd repeats which was used in the literature of Wong et al, 1998(3)) and 25 different PSEs. None of them has yielded any hits to date. A DNA-binding assay carried out on peptide 390-518 and the mouse U6 PSE, which has a high affinity to the SNAPc complex, revealed an opposite result to that which was reported in the literature(3). SNAP190RcRd does not favorably bind to the PSE, although the binding is slightly enhanced by the PSE flanker base pair. Size-exclusion chromatography suggests SNAP190RcRd is present in solution minimally as a decamer. We ascribe the observed low PSE affinity of SNAP190RcRd to its high aggregation state, which may bury the DNA-interacting elements into the interior and hinder SNAP190RcRd DNA-binding. Thus, efforts to investigate the interaction between SNAP190RcRd and the PSE in the future may need not only the RcRd repeats and the PSE, but also structural elements that prevent SNAP190RcRd from aggregating.

Appendix

Luria Broth (LB): to make 1 L of LB media, 15 g of tryptone, 5 g of yeast extract, and 5 g of NaCl were mixed and dissolved in 1 L of water. The solution was autoclaved. Before use appropriate antibiotics were added.

Terrific Broth (TB): to make 1 L of TB media, 12 g of tryptone, 24 g of yeast extract, and 4 ml glycerol were mixed and dissolved in 900mL of water and autoclaved. Before use 100 mL of salt solution containing 2.31 g KH_2PO_4 , and 12.54 g K_2HPO_4 were added to the broth.

HEMGT 250: 25 mM Hepes pH 7.9, 2 mM EDTA, 12.5 mM MgCl_2 , 10 % glycerol, 0.1 % Tween-20, 250 mM KCl, 3 mM DTT

HEMGT 100: 25 mM Hepes pH 7.9, 2 mM EDTA, 12.5 mM MgCl_2 , 10 % glycerol, 0.1 % Tween-20, 100 mM KCl, 3 mM DTT

TDB (thrombin digestion buffer) (10x): 200 mM Tris-HCl pH 8.4, 1.5 M NaCl, 25 mM CaCl_2

TDB –DTT (10x): 200 mM Tris-HCl pH 8.4, 1.5 M NaCl, 25 mM CaCl_2 , 3 mM DTT

QA: 20 mM Tris-HCl pH 7.5, 5 % glycerol, 0.2 mM EDTA, 1 mM DTT

QB: 1 M NaCl, 20 mM Tris-HCl pH 7.5, 5 % glycerol, 0.2 mM EDTA, 1 mM DTT

5 % TGE gel (10 mL): 1.25 mL Acrylamide (39:1), 2 mL 5x TGE, 0.5 mL 50 % glycerol, 8 μL Tetramethylethylenediamine, 80 μL 10 % ammonium persulfate and 6.25mL H_2O .

5 x TGE buffer: 500 mM Tris, 3.80 M glycine, 20 mM EDTA

References

1. Hinkley, C.S., Hirsch, H.A., Gu, L., Lamere, B. and Henry, R.W. (2003) The small nuclear RNA-activating protein 190 Myb DNA binding domain stimulates TATA box-binding protein-TATA box recognition *J. Biol. Chem.*, 278, 18649-18657
2. Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices *J. Mol. Biol.*, 292, 195-202
3. Wong, M.W., Henry, R.W., Ma, B., Kobayashi, R., Klages, N., Matthias, P., Strubin, M. and Hernandez, N. (1998) The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with Oct-1 *Molecular and Cellular Biology*, 18, 368-377
4. Mittal, V., Ma, B. and Hernandez, N. (1999) SNAPc: A core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain *Genes & Development*, 13, 1807-1821

Chapter III

THREE DIMENSIONAL STRUCTURES OF *Escherichia coli* GLYCOGEN SYNTHASE AND ITS COMPLEXES

III.1. Theory

III.1.1. Structure Determination from X-ray Diffraction Data

A crystal arranges huge numbers of molecules in a highly ordered and repetitive pattern. When an *X*-ray beam strikes the electron clouds(1) of these molecules, the scattered waves emitted from electrons can add up in phase and generate a measurable signal. This phenomenon is popularly termed as *X*-ray diffraction where the orderly arrayed molecules form imaginary reflecting planes and in-phase scattered waves are considered as the product of diffractions that comply with Bragg's law

$$2d \sin\theta = n\lambda$$

where *d* is the distance between reflecting planes. θ is the angle between the incident beam and the reflected beam, λ is the wavelength of incident radiation. The most commonly used *X*-ray wavelength for biological macromolecule structure determination is about 1 Å, compatible for atomic bond lengths (1.5-1.7 Å) and atoms themselves.

The *X*-ray diffraction pattern reflects the spatial distribution of the electrons and a three-dimensional picture of the electron density of a biological macromolecule can be inversely derived from the collected *X*-ray diffraction data with the aid of Fourier transformation. The Fourier transformation is a mathematical operation that can link the diffraction pattern and the object diffracting the waves. Electron density at every position *xyz* is presented by the following equation

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

The structure factor $F(hkl)$ is a complete description of all the atoms that contribute to the reflection labeled *hkl*. $F(hkl)$ is a wave function with frequency, amplitude $|F(hkl)|$ and phase $\alpha(hkl)$. The frequency of $F(hkl)$ is identical to the *X*-ray source. The amplitude is proportional to the square root of the intensity of the reflection, which can be measured from the recorded *X*-ray diffraction pattern. The phase cannot be directly derived from the diffraction pattern and constitutes the well-known Phase Problem in crystallography.

III.1.2. Molecular Replacement

Molecular replacement is one of the methods used to solve the “Phase Problem” in biological macromolecular crystallography. It takes advantage of the structure of a known protein to calculate the electron density map of a biological entity whose *X*-ray diffraction pattern has been recorded. Usually, a protein that shares >25 % sequence identity with the target protein can serve as a decent initial search model and stands a good chance of success in molecular replacement. In our case, the AtGS sequence is

43% identical to that of *E. coli* GS (2) and therefore was chosen as the initial model to solve the *E. coli* GS structure.

The search for the correct molecule position relative to the counterpart in the known structure involves three rotational and three translational parameters. In order to reduce otherwise enormous computation time, the search for these six parameters is carried out in two separate steps. The first is the rotation search, which uses the Patterson function to determine the correct orientation of the search model. The Patterson function is the Fourier transform using $F(hkl)^2$ and contains sets of peaks representing intra-atomic vectors (3). As Patterson functions can be calculated from the amplitudes only, the orientation of the search molecule can be found without the aid of phase information from the unknown structure. In the correct orientation, a maximal overlap with the target structure should be reached. The rotation search in the case of *E. coli* GS is all performed by the automated Patterson search routine of the program MolRep (4).

Once the correct orientation of the model (α, β, γ) has been found, the correct position can be determined through the translation function. The translation vector for each atom can be split into two vectors, one (t) in the rotation axis direction and the other (s) perpendicular to the rotation axis direction. The vector t is the same for all atoms while the self-vector depends on the distance of each atom to the rotation axis. Superimposing the self-vector set onto the intermolecular vectors (cross vectors) gives a number of positions where some agreement between the vector peaks is obtained. The structure factor calculated (F_{cal}) based on the solution of rotational search and translational function will be compared to the observed ones from data (F_{obs}). A properly

positioned model should give a relatively high correlation coefficient (*C.C.*) and low *R*-factor.

$$C.C. = \frac{\sum_{hkl} \left(|F_{obs}|^2 - \overline{|F_{obs}|^2} \right) \left(|F_{cal}|^2 - \overline{|F_{cal}|^2} \right)}{\left[\sum_{hkl} \left(|F_{obs}|^2 - \overline{|F_{obs}|^2} \right)^2 \sum_{hkl} \left(|F_{cal}|^2 - \overline{|F_{cal}|^2} \right)^2 \right]^{1/2}}$$

$$R = \frac{\sum_{hkl} \left| |F_{obs}| - |F_{cal}| \right|}{\sum_{hkl} |F_{obs}|} \times 100\%$$

III.1.3. Structure Refinement

After applying the initial phase from the known structure to the diffraction data of the protein of interest, an approximate model can be obtained, but most of time, that model is rather coarse and needs to be refined. In fact, an *R*-factor of 40-50 % is common for the model fresh from molecular replacement. After thorough refinements it can be lowered to 8-25 %, depending on the data quality.

Due to thermal vibrations and static disorder, the atoms in the crystal are not strictly fixed and therefore their atomic scattering factors deviate somewhat from the standard value. Temperature factors (*B*) reflect the extent of the dynamic disorders and these are refined in structure determination as well as three positional parameters (*x*, *y*, *z*). However, the number of independent *X*-ray reflections of the protein alone generally is not sufficient for solving all those parameters. Therefore, many additional “observations” are incorporated, such as geometric restraints: bond lengths, bond angles,

torsion angles and van der Waals contacts. The refinement takes advantage of these additional “observations” in a restrained refinement where the stereochemical parameters are only allowed to vary around a standard value (from small molecule geometry bank). On the other hand, the number of parameters to be refined can be substantially reduced if the entire protein or domain is taken as rigid. This refinement strategy is termed “rigid refinement”. In the *E.coli* GS case, a rigid-body refinement was first applied to find a best overall- fit position. Then a great number of restrained refinements along with manual refitting of the model to the electron density maps were carried out. The interactive graphics program Turbo Frodo(5) was used for manual fitting and Refmac(6) in the CCP4 program suite was used for both types of refinements. The R-factor of the final models of *E.coli* GS and its complexes are in the range of 17-21 %, substantially lower than the R-factor straight from the molecular replacement (31-51 %). Another widely used parameter to assess the closeness between the atomic model and the observed data is the free R value (R-free) (7). R-free is computed with a small set of randomly chosen intensities (“test set”, usually 5-10 % of the data), which are set aside from the beginning and not used during refinements. R-free measures how well the atomic model predicts a subset of the measured intensities that were not included in the refinement ($|F_{test}|$), whereas R-factor measures how well the model predicts the entire data set that produced the model ($|F_{obs}|$). The difference between R value and R-free value of an accurate model should not be large, usually less than 7-10 %(7,8). All determined *E. coli* GS structures display a nice agreement between R factor and free R factor (Table III.1).

Besides R and R-free values, a Ramachandran plot is another means commonly used to evaluate the model quality in biological macromolecular crystallography. Due to

steric hindrance, the mainchain of a polypeptide usually assumes preferred, energetically favorable conformations(9). Typically for each residue, these conformations are characterized by torsion angles, ϕ (Phi) $C_{i-1}-N_i-C_{\alpha i}-C_i$ and ψ (Psi) $N_i-C_{\alpha i}-C_i-N_{i+1}$. The distribution of ϕ and ψ is called the Ramachandran plot which shows how well the ϕ and ψ angles of protein residues cluster. As ϕ and ψ angles are not usually restrained during *X-ray* refinement, a Ramachandran plot provides a simple, sensitive, and independent means for assessing the quality of a protein model (10). The Ramachandran plot of *E.coli* GS structures and the related discussion is described in their individual sections.

III.2. Experimental Procedures

III.2.1. Protein Overexpression

E.coli GS (both His-tagged and untagged) was overexpressed in *E.coli* BL21 (DE3) cells. Fifty milliliters of LB medium containing 50 $\mu\text{g/mL}$ kanamycin was inoculated with a single colony of *E.coli* strain BL21 (DE3) bacteria containing the GS expression plasmid (pAY3 for His-tagged GS, pAY1 for untagged GS)(11). Cells were grown at 37 °C with shaking at 250 rpm for 3-4 h or until saturated. One liter of LB / kanamycin was inoculated with 50 mL of saturated culture and growth was maintained at 37 °C with 250 rpm shaking until the O.D.₆₀₀ reached 0.6-0.8. The cell culture was then cooled down to 4 °C before induced by 1 mM IPTG. Growth continued for 12 h at room temperature (20-22 °C) before the cells were harvested by centrifugation at 5000 rpm.

III.2.2. Protein purification

III.2.2.A. His-tagged GS protein

Cells were resuspended in the buffer (50 mM potassium phosphate at pH 8.0, 300 mM NaCl, 10 mM imidazole) and lysed by sonication. Cell debris was spun down by centrifugation at 5000 rpm for 30 minutes. The supernatant was mixed with Ni-NTA resin (Qiagen, Valencia, CA) in volume ratio 4:1. The mixture was stirred slowly at 4 °C for 1 h to allow thorough binding before loaded onto an empty column. The impurity was removed by the wash buffer (50 mM potassium phosphate at pH 8.0, 300 mM NaCl, 50 mM imidazole). The His-tagged GS protein was eluted with the elution buffer (50 mM potassium phosphate at pH 8.0, 300 mM NaCl, 250 mM imidazole).

Fractions were analyzed by SDS-PAGE for homogeneity (Figure III.1), and were thoroughly buffer- exchanged into the storage buffer (20 mM TEA/HCl pH 8.0 and 5 mM DTT) using desalting columns (Bio-Rad Econo-Pac 10 DG) and concentrated to 6-8 mg/ mL for crystallization.

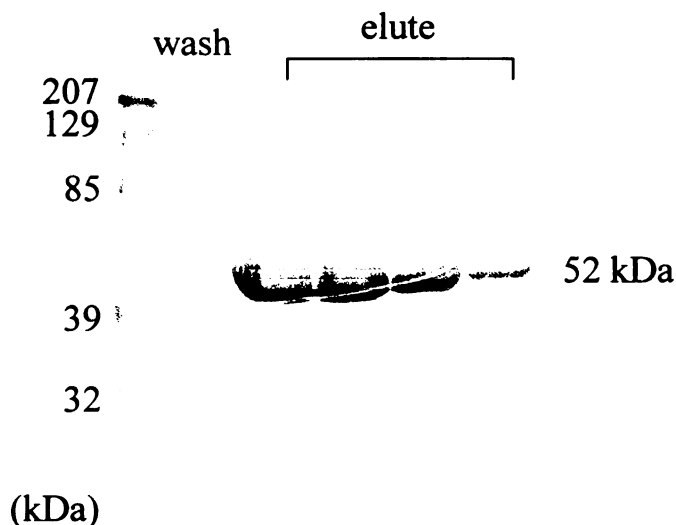


Figure III.1 Representative His-tagged *E.coli* GS SDS-PAGE gel. The most left lane is the molecular weight standard. Each elution lane accounts for the collection of a 5 mL elution buffer fraction.

III.2.2.B. Untagged dmGS protein

Cells were resuspended in the buffer (50 mM potassium phosphate at pH 6.8, 100 mM NaCl, 5 mM DTT) and lysed by sonication. Cell debris was spun down by centrifugation at 10,000 rpm for 30 minutes. The supernatant was then brought to the 25% saturation of ammonium sulfate and the mixture was stirred 30 minutes and centrifuged at 10,000 rpm. Most of the impurities stay in the supernatant and was discarded. The pellet was dissolved in buffer A (20 mM TEA/ HCl pH 7.5 and 5 mM DTT) and was desalted by desalting column or dialysis. Protein was filtered through 0.22 μ m filter and the solution was loaded onto a Source -Q column. GS protein eluted between 250 -350 mM KCl, 20 mM TEA/ HCl pH 7.5 and 5 mM DTT (Figure III.2). The purified GS protein was poured onto a desalting column three times to remove 99 % salt and concentrated to 6-8 mg/mL before crystallization.

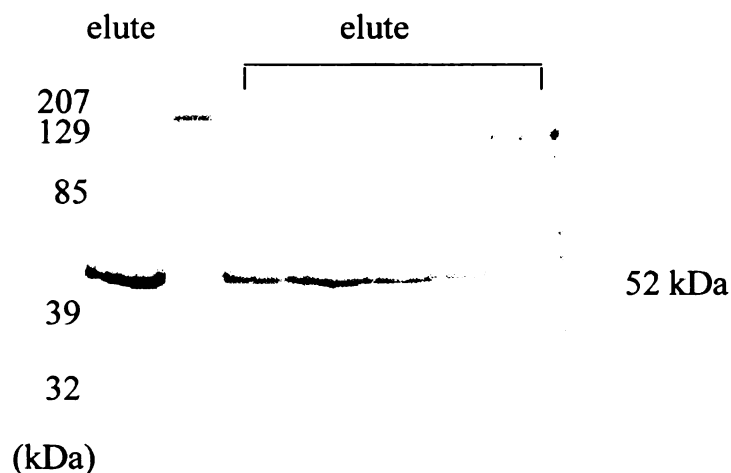


Figure III.2 *E.coli* dmGS SDS-PAGE. The second from left lane is the molecular weight standard and all other lanes are elution from Source -Q column with increasing concentration of KCl.

III.2.3. GS and GS Complex Crystallization

GS protein was incubated with various substrates and inhibitors for 30 minutes to generate GS complexes. A full list of substrates and inhibitors that have been used in crystallization trials are shown in Table III.1. The resulting GS complex solution was then subjected to spin-filtration using a spin filter with 0.45 μm cut-off membrane (Microcon® 0.45 μm , Amicon bioseparations) to remove precipitates that would cause excessive nucleation in crystallization.

Crystallization was performed using both sitting drop and hanging drop vapor diffusion methods at room temperature and 4 °C. The reservoir in the crystallizing plates was about 1 mL and the primary screening solution kits (Crystal Screen, Crystal Screen 2) were purchased from Hampton Research (Aliso Viejo, CA).

Once the primary crystallization condition was found, further screening of that condition was conducted by altering buffer, precipitate, additive, and temperature to obtain diffraction-quality crystals. Other attempts to improve crystal quality include the micro-dialysis, micro- and macro-seeding (12). When co-crystallization of GS with ligand did not produce suitable crystals for data collection, the apo-GS crystal was soaked in the stabilizer solution with a high concentration of ligand (50-500 times K_m or K_i). Usually, the stabilizer solution, which is able to prevent the crystal from dissolving, is the original reservoir solution with 5 –10 % higher concentration of precipitate (PEG 4000). For the ligand- bound GS crystals, multiple data sets were collected and the one with the best quality was used for the structure reported in the later sections except for the wild-type GS-ADP-HEPPSO complexes.

Table III.1 GS complex crystallization trials

Ligand	Crystal	Condition	Note
Substrate			
ADPGlc(<i>K_m</i> 18 μ M)	YES	40% PEG4K + 0.1M Tris pH 8.5	room temperature, space group C222
	YES	40% PEG4K + 0.1M HEPPSO pH 7.5	E377A
	YES	40% PEG4K + 0.1M HEPPSO pH 8.1 +0.2 M NaAc	wtGSa
		40% PEG4K + 0.1M HEPPSO pH 7.7 +0.2 M Natartrate	wtGSb,wtGSd
		40% PEG4K + 0.1M HEPPSO pH 7.6 +0.2 M Natartrate	wtGSc
	YES	36% PEG4K + 0.1M Tris pH 8.5 + 0.2 M NaAc	4 $\text{\AA}$
Maltose(<i>K_m</i> 800 mM)	NO		
Maltotriose	NO		
Maltopentaose	NO		
Maltohexaose (<i>K_m</i> >150 mM)	NO		
inhibitor			
ADP (<i>K_i</i> 45 μ M)	YES	36% PEG4K +0.1M Tris pH 8.1 +0.2M NaAc+0.03M Betaine	room temperature, large mosiacity, 5 $\text{\AA}$
AMP	NO		
Glucose-1-phosphate (<i>K_i</i> 1.15 mM)	YES	30% PEG4K + 0.1M Tris pH 8.5 + 0.2 M MgCl ₂	weak diffraction
		40% PEG4K + 0.1M Tris pH 8.5 + 75 mM NaAc	large mosiacity
	YES	38% PEG4K + 0.1M Tris pH 8.5 + 0.05 M NaAc	weak diffraction
		40% PEG4K + 0.1M Tris pH 8.1+0.2M Natartrate	5 $\text{\AA}$, E377A
Combination			
ADPGlc+maltose	YES	40% PEG4K + 0.1M HEPPSO pH 7.5	3 $\text{\AA}$
ADPGlc+maltotriose	YES	40% PEG4K + 0.1M HEPPSO pH 7.5 +0.1 M glycine	E377A
	YES	40% PEG4K + 0.1M HEPPSO pH 7.5 +0.1 M taurine	R300A
ADPGlc+maltopentaose	YES	40% PEG4K + 0.1M HEPPSO pH 7.5	E377A
	YES	40% PEG4K + 0.1M HEPPSO pH 7.5 +0.1 M taurine	R300A
	YES	40% PEG4K + 0.1M HEPPSO pH 7.5	E377A
ADPGlc+acarbose+maltohexaose	NO		
ADP +maltose	NO		
ADP+maltotriose	YES	40% PEG4K + 0.1M HEPPSO pH 7.5 +0.1 M glycine	weak diffraction
ADP+maltopentaose	NO		
ADP+maltohexaose	NO		
ADP+ δ -gluconolactone	NO		

if not noted, complex is made with wild-type *E. coli* GS at 4 $^{\circ}\text{C}$

III.2.4. Diffraction Data Collection and Processing

All crystals were obtained at 4 °C and were harvested into the mother liquid with 15 % glycerol as cryoprotectant and flash-frozen before the data collection. Data were collected at the Advance Photon Source of Argonne National Laboratories (APS) (Argonne, IL) at various beamlines (details in Table III.2 and III.3). Data reduction and scaling were performed using Denzo and Scalepack (13) either at home or at the beamline (HKL 2000). Complete data collection and process information is shown in the table of each individual section.

III.2.5 ¹³C- and ¹H-¹³C NMR Experiment

III.2.5.A Preparation of D- [1-¹³C]-ADPGlc from α-D- [1-¹³C]

glucopyranosyl 1-phosphate

α-D- [1-¹³C] glucopyranosyl 1-phosphate (dicyclohexylammonium salt, monohydrate) (MW 477.51, Cat GLC-015) and α-D- [UL-¹³C₆] glucopyranosyl 1-phosphate (dicyclohexylammonium salt, monohydrate, uniform labeled) (MW477.51, Cat. GLC-074) were purchased from Omicron Biochemicals, Inc. (South Bend, IN). The TSK-GEL DEAE-5PW column was purchased from Tosoh Biosciences LLC (Montgomery Ville, PA).

The reaction mixtures contain 100 mM Bicine or Hepes buffer (pH 8.0), 0.2 mg/mL bovine serum album, 7 mM MgCl₂, 1.5 mM ATP, 2 mM fructose 1,6-bisphosphate, 1.5 unit /mL inorganic pyrophosphatase, 0.5 mM glucose-1-phosphate(14). Reaction was initiated by addition of the enzyme *E. coli* ADPGlc pyrophosphorylase and the reaction mixture was incubated for 2 hrs at 37 °C before it was terminated by EDTA.

Paper chromatography was used to check the presence of product ADPGlc. Five μL of 20 mM ATP, ADP, and ADPGlc and 30 μL reaction mix were streaked on Whatman No.1 paper and chromatographed (descending) for 24 hrs in the ethanol – ammonium acetate pH 7.5 solvent system (95 % ethanol /1 M ammonium acetate, pH 7.5 (5:2)). The chromatogram was dried in the hood and the ADPGlc spot was located with UV-light. Descending paper chromatography revealed that there are two main UV absorbing spots, which correspond to the product ADPGlc and the unreacted ATP. To separate ^{13}C -labeled ADPGlc from the reaction mix, particularly from the unreacted ATP, the reaction mix was injected into the anion exchange DEAE-5PW column after filtered through a 0.45 μm filter. The separation was done on a DIONEX P680 pump system with gradient between buffer A (25 mM ammonium formate, pH 7.2) and buffer B (25 mM ammonium formate and 0.5 M NaCl, pH 7.2) (batch I) or buffer A (25 mM Tris, pH 7.2) and buffer B (25 mM Tris pH 7.2 and 0.5 M NaCl, pH 7.2) (batch II). Comparison of the HPLC spectrum of the reaction mix and the ATP, ADPGlc, ADP and AMP standards revealed that, besides the expected ADPGlc and ATP, there was a small amount of AMP and ADP resulting from the degradation of ATP/ADPGlc (Figure III.3). The second peak was identified as ADPGlc and the fraction at that position was collected and lyophilized to give a white solid.

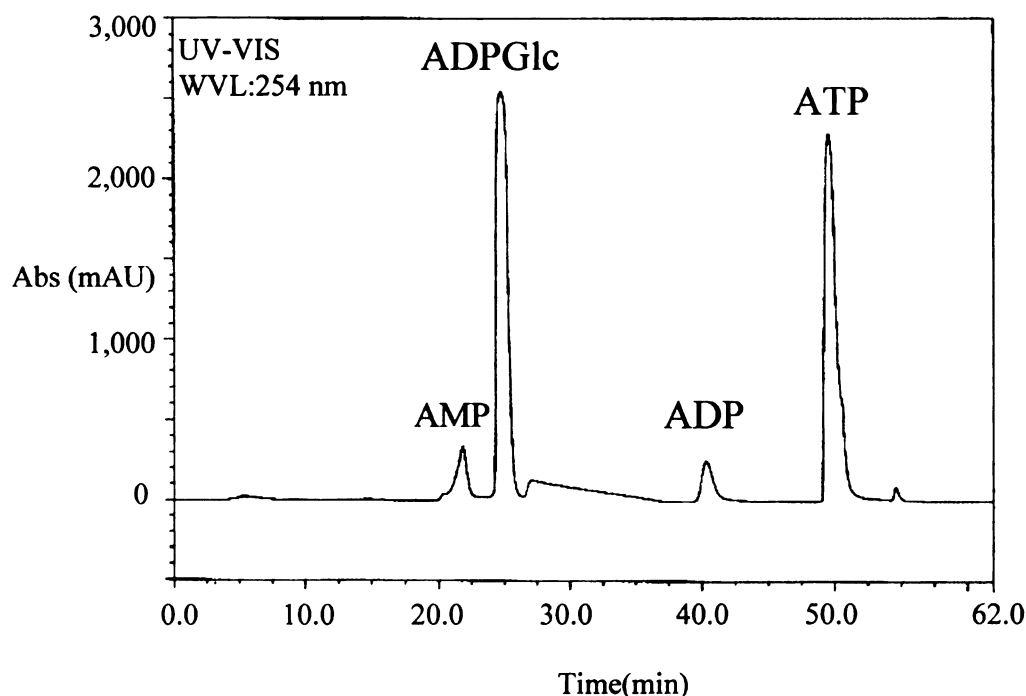


Figure III.3 HPLC spectrum from the DEAE-5PW column. Flow rate: 0.6 mL/min, Gradient: 2 min: 0% buffer B; 40 min: 25 % buffer B; 60 min: 100 % buffer B. Fraction ADPGlc elutes in the range of 14 –18 % buffer B.

A typical adenosine derivative displays an absorption maximum at 259 nm (A_{289} : $A_{260} = \sim 0.15$ and $A_{250}:A_{260} = \sim 0.80$). The spectral analysis of the $[1-^{13}\text{C}]$ and $[\text{UL}-^{13}\text{C}_6]$ -ADPGlc at pH 7.0 is in agreement with that. The quantity of ADPGlc was estimated based on the equation “g · per liter = ($\text{Abs}_{259 \text{ nm}} \cdot \text{Molecular Weight}$) / α_M , average molar absorbance index $\alpha_M = 15.4 \times 10^3 \text{ L/mol}$ ” (15). The yield of $[1-^{13}\text{C}]$ -ADPGlc and $[\text{UL}-^{13}\text{C}_6]$ -ADPGlc are 66 % and 87 %, respectively.

III.2.5.B. NMR experiment setting

^1H NMR, ^{13}C NMR and phase sensitive gHMQC two-dimensional spectra were recorded at 500 MHz (Varian Unity Plus 500) or at 600 MHz (Varian Inova 600). The

NMR facility temperature was calibrated with pure methanol prior to experiment and all data were collected at 4 -5 °C. Solvent D₂O with reference TSP (sodium salt of 3-(trimethylsilyl) propionic -2, 2, 3, 3- d₄ acid) was purchased from Sigma-Aldrich (Catalogue # 293040) and was used without further purification. All NMR data were processed with the Varian machine software and the spectra images were made with software Mestrec23.

III.2.5.C. NMR sample preparation

Wild-type *E.coli* GS (7-9 mg /mL) was thawed from – 80 °C and diluted to ~ 2-3 mg/mL with D₂O. ¹³C –labeled ADPGlc and HEPPSO was added in a 1:10:100 ratio to GS, ADPGlc and HEPPSO. The GS-ADPGlc-HEPPSO adduct was then concentrated by centrifugation until reaching its maximum concentration, approximately 9 mg /mL. The total complex volume was ~ 0.7 mL, a minimum volume to obtain decent NMR data.

III.2.5.D. NMR spectra

The starting material [1-¹³C]-Glc-1-P displays a major doublet α -¹³C₁ peak at 96.50 and 96.44 ppm and a minor doublet β -¹³C₁ at 99.89 and 99.85 ppm (approximately 1.8 % of the height of the 96.50 ppm peak) (Figure III.4). The first batch of [1-¹³C]-ADPGlc exhibits a large doublet α -¹³C₁ peak at 98.50 and 98.55 ppm and a minor doublet β -¹³C₁ peak at 99.95 and 100.02 ppm (approximately 3.0 % of the height of the 98.50 ppm peak) (Figure III.4).

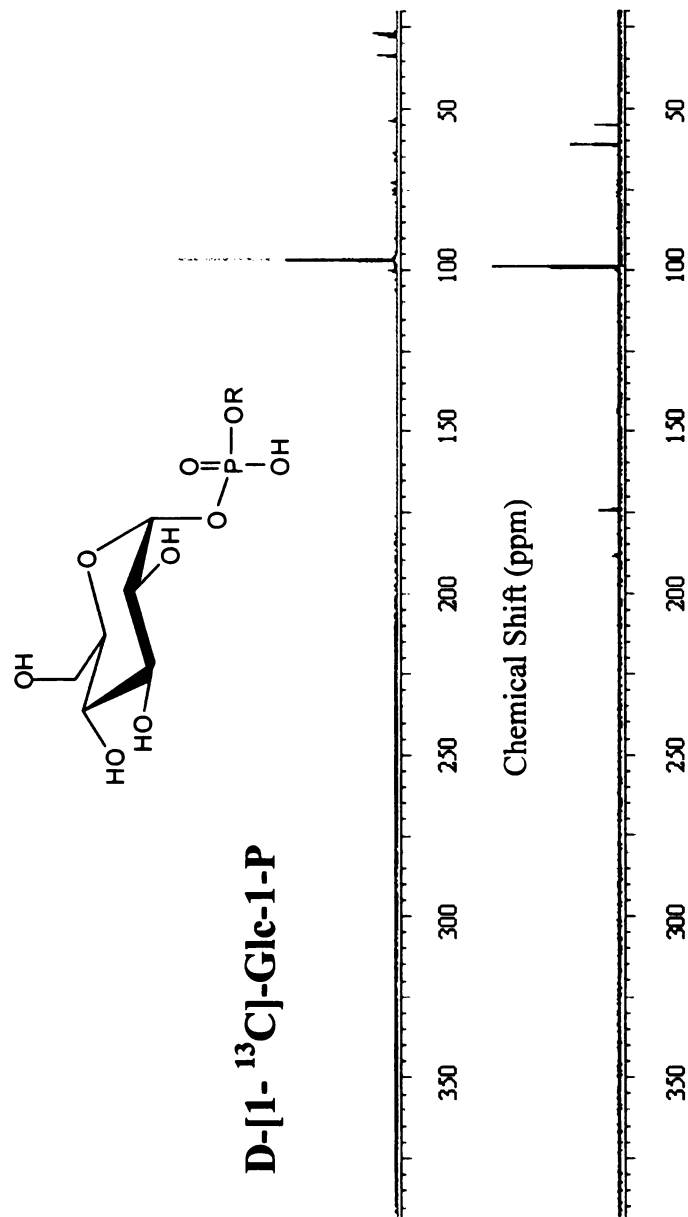


Figure III.4 NMR spectra of the starting material [1- ^{13}C]-Glc-1-P and batch I [1- ^{13}C]-ADPGlc.

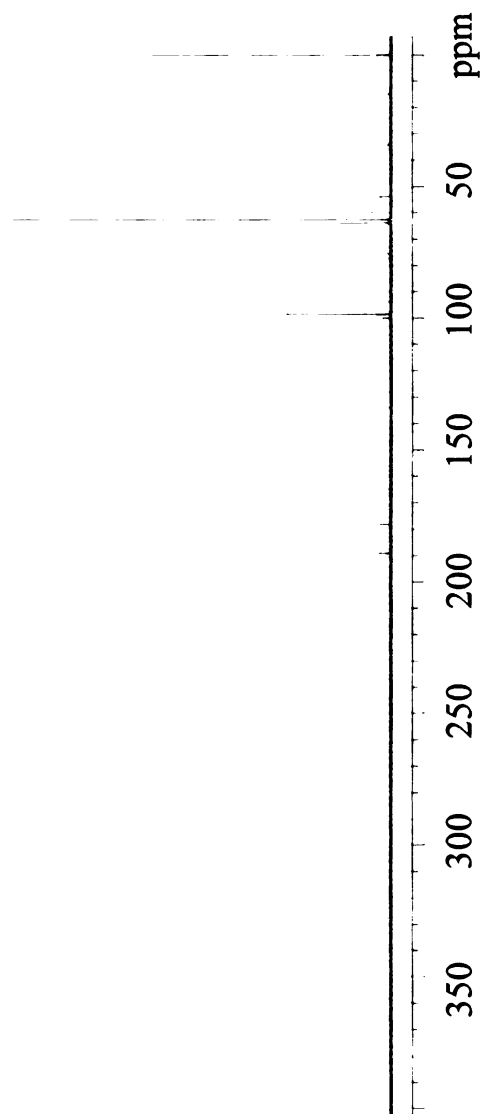
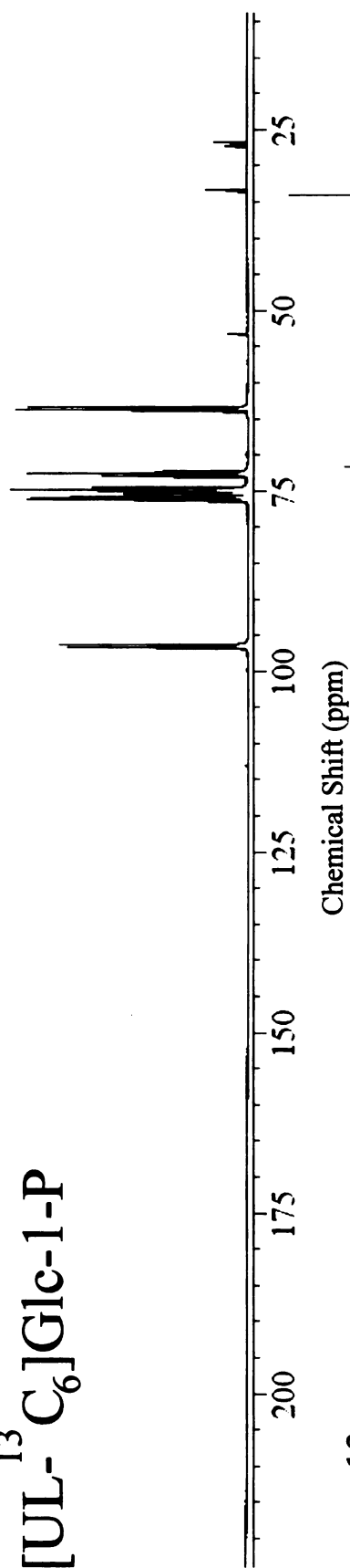


Figure III.5 NMR spectra of batch II [1- ^{13}C]-ADPGlc.

$[UL-^{13}C_6]Glc-1-P$



$[UL-^{13}C_6]ADPGlc$

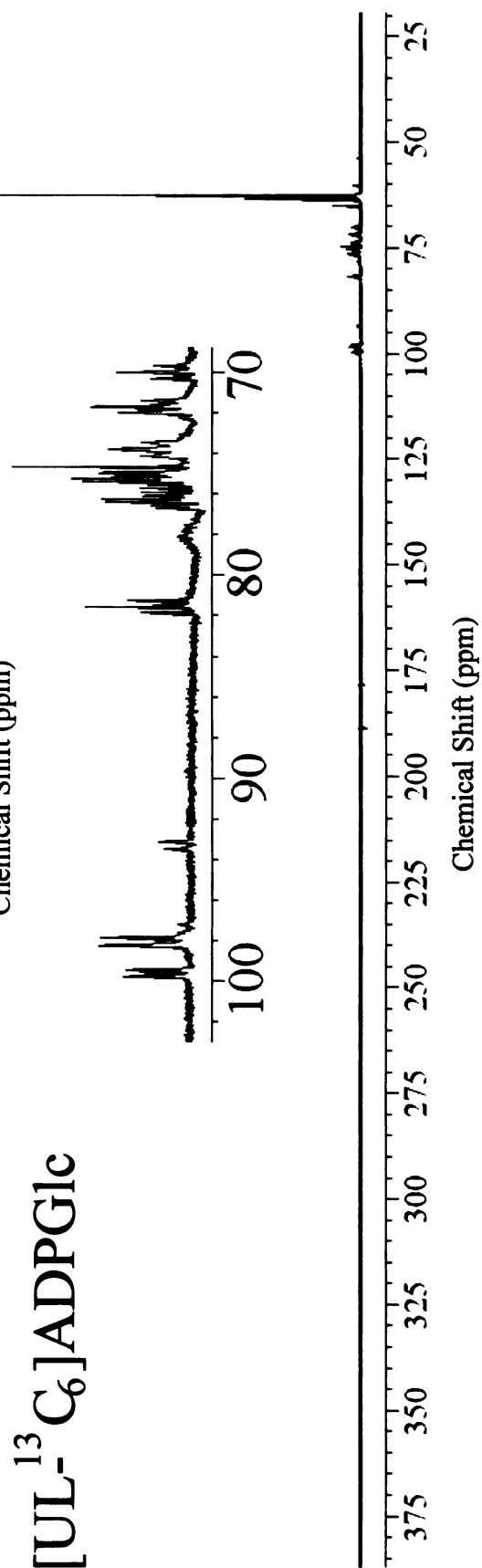


Figure III.6 NMR spectra of the starting material $[UL-^{13}C_6]Glc-1-P$ and $[UL-^{13}C_6]ADPGlc$.

In the carbocation DGM, positively charged C1 is deshielded and probably its peak will appear at lower field than the ADPGlc C1 peak (98.5 ppm). However, in the batch 1 [1-¹³C]-ADPGlc NMR spectrum, we see two peaks (173.8 and 188.0 ppm) in the low field region where the DGM C1 peak may appear. To vacant the low field region for a clear interpretation of the C1 peak of DGM in GS-ADPGlc-HEPPSO complex, the enzymatic synthesis buffer bicine and HPLC buffer ammonium formate were purposely replaced by HEPES and Tris, respectively, in the production of batch 2 [1-¹³C]-ADPGlc and [UL-¹³C₆] -ADPGlc. However, the 173.8 and 188.0 ppm peak persisted, indicating they were from some other source than buffer bicine or ammonium formate (Figure III.5 and Figure III.6). On the other hand, unlike ammonium formate which has low vapor pressure and most of it was removed in the lypholization process, the replacement buffer, Tris, was condensed during lypholization and display a large peak at 53.9 and 57.2 ppm in batch 2 [1-¹³C]-ADPGlc and [UL-¹³C₆] –ADPGlc (Figure III.5 and Figure III.6).

NMR spectra of uniformly labeled Glc-1-P ([UL-¹³C₆] –Glc-1-P) and ADPGlc ([UL-¹³C₆] –ADPGlc) are shown in Figure III.6. Compared to the regular ADPGlc ¹³C NMR spectrum (C1 : 96.24 /95.93 ppm; C2 :73.53 /73.39 ppm; C3 :73.39 ppm; C4 :69.92 ppm; C5 :72.08 ppm and C6 : 61.07 ppm) (16), [UL-¹³C₆] –ADPGlc displayed a complicated pattern due to the coupling between neighboring sugar ¹³C-labeled carbons. The central chemical shift of [UL-¹³C₆] –ADPGlc sugar carbon peaks are as followed: C1 : 98.27/97.86; C2 :76.20 ppm; C3 :75.34 ppm; C4 :71.66 ppm; C5 :73.76 ppm and C6 : 63.05 ppm (Figure III.6). Moreover, the appreciable doublet peak at 99.41 and 99.77 ppm corresponds to the C1 of β-[UL-¹³C₆] –ADPGlc, indicating that there is a considerable amount of the β anomer in [UL-¹³C₆] –ADPGlc.

III.3. Model Building and Structure Refinement

III.3.1. apo dmGS (C7S: C428S)

E.coli dmGS crystals were obtained from the complex of dmGS (2.5 mg/mL) and ADPGlc (0.54 mM) in a buffer of 40 % (w/v) PEG 4000, 0.1 M Tris (pH 7.5) and 0.2 M Na tartrate. The crystal used for data collection incubated for 16 wk before reaching its maximum size 0.2 × 0.2 × 0.3 mm (Figure III.7).

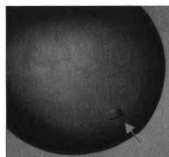


Figure III.7 A crystal of dmGS

The structure of apo-dmGS was solved by the Molecular Replacement method with the program MOLREP (Collaborative Computational Project, CCP4)(17) using the *A. tumefaciens* GS structure (1RZV.pdb) as a search model, with non-conserved amino acid residues mutated to alanine, except for glycine. Initially, plausible rotation and translation functions for the N-terminus (1-240) were identified and the C-terminus (271-456) was subsequently incorporated to search for its counterpart. Placement of the model was optimized by rigid-body refinement. The *apo* dmGS as well as all other GS structures was obtained through REFMAC5 (CCP4) refinement and map calculations with model building using TURBO-FRODO. All the structure refinement statistics are listed in Table III.2. The final structure of apo-dmGS consists of the first 475 residues of the protein (total 477 in *E.coli* GS) and only one molecule in each asymmetric unit. Figure III.8 represents the Ramachandran plot of the *apo* dmGS structure.

Table III.2 Data collection and refinement statistics of *E.coli* GS mutant and their complexes

	dmGS(C7S;C408S)	E377A-ADP-HEPPSO	E377A-ADP-oligosaccharides
Space group	H 32	I 41	I 41
Unit cell			
a / b / c (Å)	233.0/ 233.0/ 85.5	126.10,/126.10/ 152.07	126.3/ 126.3/ 152.3
α / β / γ (°)	90.0/ 90.0/120.0	90.0/ 90.0/90.0	90.0/ 90.0/ 90.0
Data processing statistics			
Beam line	COMCAT-ID32	SBCCAT-BM19	DNDCA-BM5
Wavelength (Å)	1.01	1.03	0.99
^a Resolution (Å)	50.00 - 3.02 (3.15-3.02)	50.00 - 2.30 (2.38-2.30)	97.13-2.29 (2.32-2.29)
^a Total reflections	21871 (2146)	30260 (2994)	335515 (25704)
^a Multiplicity	9.4 (9.5)	3.5(3.6)	3.21 (2.47)
^a Completeness (%)	99.9 (100)	92.5 (88.7)	98.1(99.0)
^{a, b} Rmerge (%)	7.6 (34.6)	7.8 (42.0)	8.5 (52.84)
^a I / σ (I)	28.3 (1.94)	13.7 (2.0)	14.5 (1.06)
Refinement statistics			
^c R work(%)	20.92	17.41	18.60
^d R free(%)	26.54	19.70	22.10
r.m.s. deviation from ideal values			
Bond length (Å)	0.012	0.009	0.011
Bond angle (°)	1.48	1.25	1.43

Table III.2 (cont'd)

	dmGS(C7S;C408S)	E377A-ADP-HEPPSO	E377A-ADP-oligosaccharides
Ramachandran statistics (%)			
Most favoured regions	85.0	88.7	90.4
Additional allowed regions	14.8	11.1	9.3
Generally allowed regions	0.2	0.2	0.2
Disallowed regions	0.0	0.0	0.0
B factor			
Average B ($\text{\AA}^2$)	50.44	44.49	51.28
$^{\circ}$ Protein	50.49 (3721)	43.70 (3749)	49.32 (3751)
$^{\circ}$ ADP		39.02(27)	47.35 (27)
$^{\circ}$ HEPPSO		42.06 (17)	101.80(17)
$^{\circ}$ Glucose at surface		85.87 (12)	
$^{\circ}$ Oligosacchride at Ga site			53.48 (34)
$^{\circ}$ Oligosacchride at Gb site			92.80 (23)
$^{\circ}$ Oligosacchride at Gc site			68.98 (56)
$^{\circ}$ Oligosacchride at Gd site			65.50(67)
$^{\circ}$ polyethylene glycol chain			74.99 (107)
$^{\circ}$ water	41.97(21)	53.38 (298)	56.70 (201)
Solvent content (%)	70.7	78.4	78.5
Protein Data Bank code		3COP	3CX4

^a The parentheses denote those values for the last resolution shell.

$$^b R_{merge} = \frac{\sum ||I| - \langle I \rangle||}{\sum |I|} \quad \text{where } I \text{ is the observed intensity and } \langle I \rangle \text{ is}$$

the average intensity obtained from multiple observations of symmetry-related reflections.

$$^c R = \frac{\sum ||F_{obs}| - |F_{cal}||}{\sum |F_{obs}|}$$

$$^d R_{free} = \frac{\sum ||F_{obs}| - |F_{cal}||}{\sum |F_{obs}|} \quad \text{where reflections belong to a test set of 5 \%}$$

randomly selected data.

^e Values in parentheses are for the atom number of that species in category.

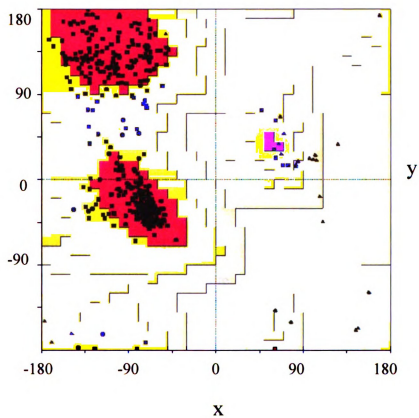


Figure III.8 Ramachandran plot of *E.coli* dmGS. Phi (degrees) is x and Psi (degrees) is y.

III.3.2. Wild - type GS in Complex with ADP and DGM (wtGSa)

The three- dimension structure of wild-type GS in complex with substrate revealing the authentic active site of the protein is of great importance to elucidate substrate- binding and understand the glucosyl transfer mechanism. CocrySTALLIZATION of wtGS and substrate ADPGlc produced several diffraction – quality crystals, among which wtGSa is the only one that was obtained from short (4 wk) incubation and contains ADP and DGM in its active site.

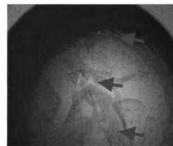


Figure III.9 Crystals of wtGSa. The center crystal is the one the diffraction data was collected from.

Single crystals of wtGSa of size ranging from $0.05 \times 0.1 \times 0.1$ mm to $0.1 \times 0.2 \times 0.2$ mm were obtained after 4 wk incubation in a solution of 40 % (w/v) PEG 4000, 0.2 M NaAc and 0.1 M HEPPSO (pH 8.1) (Figure III.9).

The structure of the first *E.coli* GS, *apo* dmGS, was used as the phasing model for molecular replacement performed with MOLREP (CCP4). There is only one molecule in each asymmetric unit. The final structure contains all residues except for the His-6 tag attached to the wtGS C-terminus. In addition, one buffer molecule HEPPSO, three PEG links and two acetate molecules are found in the structure. The structure also includes 239 water molecules. Figure III.10 represents the Ramachandran plot of the wtGSa structure, showing that none of the residues are in the disallowed region. The refinement statistics are listed in Table III.3.

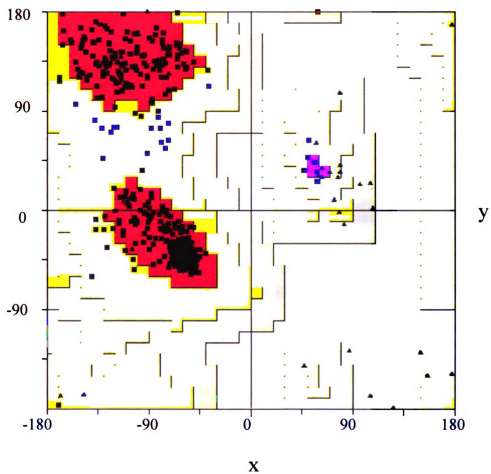


Figure III.10 Ramachandran plot of *E.coli* GS /ADP/ DGM /HEPPSO wtGSa complex structure. Phi (degree) is x and Psi (degree) is y.

III.3.3. Wild - type GS in Complex with ADP and Glucose (wtGSb, wtGS_c and GS_d)

The crystals of wtGS_b and wtGS_d grew in the buffer 40 % (w/v) PEG 4000, 0.2 M Na tartrate and 0.1 M HEPPSO (pH 7.7), and reached a maximum size $0.25 \times 0.25 \times 0.25$ mm and $0.2 \times 0.2 \times 0.2$ mm, respectively, within 12 wk (Figure III.11). The $0.2 \times 0.2 \times 0.2$ mm single crystal used for wtGS_c data collection was also obtained from a

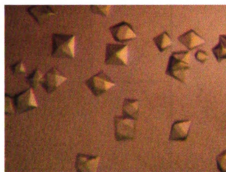


Figure III. 11 Crystals of wtGS_b

12 wk incubation under the same conditions except that the pH of HEPPSO was 7.6.

wtGS_b, wtGS_c and wtGS_d are from slightly different conditions (pH) and all contain ADP and glucose in their active sites. The structure of *E. coli* GS wtGS_a (PDB ID: 2QYY) was used as the phasing model for molecular replacement performed by MOLREP (CCP4). There is only one molecule in each asymmetric unit. The final structure includes all the protein residues, ADP, glucose, buffer molecule HEPPSO. In addition, wtGS_c and wtGS_d have one and four PEG links in their structures, respectively. The His-6 tag attached at the wtGS C-terminus was not traceable. The wtGS_b, wtGS_c and wtGS_d structures also include 246, 239, and 209 water molecules, respectively. Figure III.12, 13, and 14 represents the Ramachandran plots of wtGS_b, wtGS_c, and wtGS_d complex structure, respectively. There are no residues in the disallowed region. The refinement statistics are listed in Table III.3.

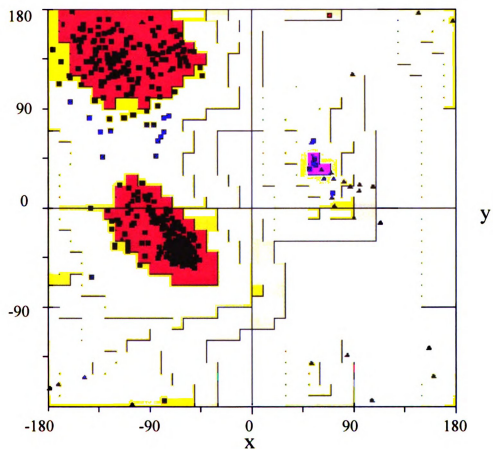


Figure III.12 Ramachandran plot of *E.coli* GS / ADP / glucose / HEPPSO wtGSb complex structure. Phi (degrees) is x and Psi (degrees) is y.

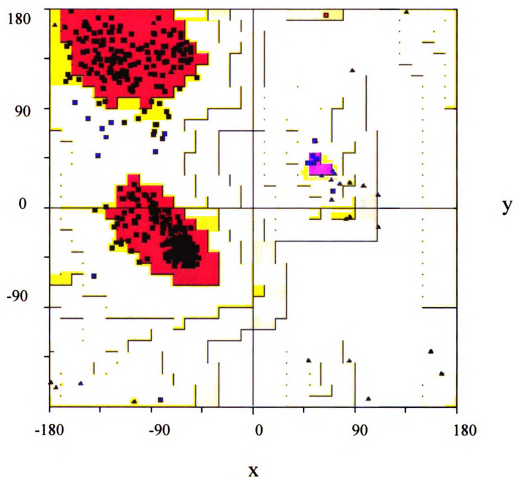


Figure III.13 Ramachandran plot of *E.coli* GS /ADP /glucose /HEPPSO wtGSc complex structure. Phi (degree) is x and Psi (degree) is y.

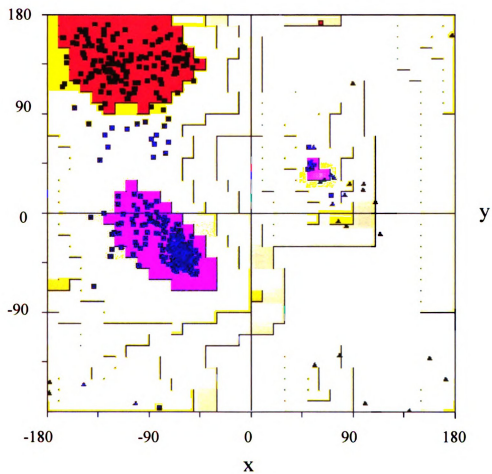


Figure III.14 Ramachandran plot of *E.coli* GS /ADP /glucose /HEPPSO wtGSd complex structure. Phi (degree) is x and Psi (degree) is y.

Table III.3 Data collection and refinement statistics of wild type *E.coli* GS complexes

	wtGSa	wtGSb	wtGSc	wtGSd
Space group	I 41	I 41	I 41	I 41
Unit cell				
a / b / c (Å)	125.84/125.84/153.26	126.49/126.49/151.88	127.10/127.10/152.00	126.20/126.20/151.60
α / β / γ (°)	90.0 / 90.0 / 90.0	90.0/90.0/90.0	90.0 / 90.0/90.0	90.0/90.0/90.0
Data processing statistics				
Beam line	BiOCARS-ID14	DND-CAT-BM5	DND-CAT-BM5	DND-CAT-BM5
Wavelength (Å)	1.01	0.99	0.99	0.99
^a Resolution (Å)	50 - 2.8 (2.87-2.80)	97.13 - 2.22 (2.43-2.22)	97.13 - 2.26 (2.38-2.26)	97.13 - 2.37 (2.53-2.37)
^a Total reflections	27528 (1765)	57263 (4243)	53217 (3870)	44756 (3333)
^a Multiplicity	7.5 (3.5)	3.15 (3.06)	2.49 (2.56)	2.68 (2.76)
^a Completeness (%)	98.5 (87.6)	99.8 (99.3)	99.2 (99.1)	97.7 (90.8)
^{a, b} R _{merge} (%)	10.9 (36.0)	5.7 (28.9)	4.21 (20.60)	6.12 (32.66)
^a I / σ (I)	18.07 (2.0)	11.29 (2.04)	10.61 (2.0)	10.34 (1.77)
Refinement statistics				
^c R _{work} (%)	17.5	20.3	16.3	18.2
^d R _{free} (%)	19.9	22.8	17.9	20.2
r.m.s. deviation from ideal values				
Bond length (Å)	0.009	0.014	0.008	0.011
Bond angle (°)	1.32	1.59	1.25	1.45

Table III.3 (cont'd)

	wtGSa	wtGSb	wtGSc	wtGSd
Ramachandran statistics (%)				
Most favoured regions	90.0	90.0	90.0	90.7
Additional allowed regio	9.8	9.8	9.8	9.1
Generally allowed regio	0.2	0.2	0.2	0.2
Disallowed regions	0.0	0.0	0.0	0.0
B factor				
Average B ($\text{\AA}^2$)	49.02	37.97	38.00	41.93
°Protein	48.12 (3735)	37.44 (3735)	37.08 (3738)	42.19 (3735)
°ADP	51.38 (27)	33.95 (27)	33.08 (27)	40.15 (27)
°Glucose		23.08 (12)	28.80 (12)	34.41 (12)
°DGM	24.43 (11)			
°HEPPSO	75.34 (17)	55.06 (17)	68.36 (17)	76.20 (17)
°Polyethylene glycol	79.77 (61)		75.18 (94)	81.88 (24)
°water	52.48 (209)	47.49 (208)	51.31 (335)	48.24 (199)
Solvent content (%)	77.9	78.3	77.9	78.1
Protein Data Bank code	2QYY	2QZS	2R4T	2R4U

^a The parentheses denote those values for the last resolution shell.

$$^b R_{merge} = \frac{\sum ||I| - \langle I \rangle||}{\sum |I|} \quad \text{where } I \text{ is the observed intensity and } \langle I \rangle \text{ is}$$

the average intensity obtained from multiple observations of symmetry-related reflections.

$$^c R = \frac{\sum ||F_{obs}| - |F_{cal}||}{\sum |F_{obs}|}$$

$$^d R_{free} = \frac{\sum ||F_{obs}| - |F_{cal}||}{\sum |F_{obs}|} \quad \text{where reflections belong to a test set of 5 \%}$$

randomly selected data.

^e Values in parentheses are for the atom number of that species in structure.

III.3.4. GS Mutant E377A in Complex with ADP and Oligosaccharides

The structure of E377A in complex with ADP and oligosaccharides was obtained from oligosaccharide cocrystallization and soaking. *E. coli* GS E377A (8.6 mg/mL) was first cocrystallized with 5 mM ADPGlc and 5 mM maltohexaose in buffer 40 % (w/v) PEG 4000 and 0.1 M HEPPSO (pH

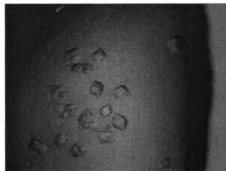


Figure III.15 Crystals of ADP, oligosaccharide-bound E377A.

7.5). Crystals of size ($0.05 \times 0.05 \times 0.05$ mm) were spotted after 4 wk and to that 3 μ l crystal-containing drop 2 μ l soaking solution was added which is comprised of equal volumes of well solution and 100 mM maltopentaose (Figure III.15).

The structure was solved by the Molecular Replacement method with MOLREP (Collaborative Computational Project, CCP4) (17) using the *E. coli* GS wtGSb (PDB ID: 2QZS) structure as a search model. Refinement and map calculations were carried out with the program REFMAC5 (CCP4)(17). All model building including manual adjustments and fitting of carbohydrate density was performed using TURBO-FRODO (18) running on a Silicon Graphics workstation. Water molecules were added using coot and inspected visually prior to deposition. Five percent of the observations were flagged as “free”(7) and used to monitor R_{free} . The final model displays good stereochemistry as evaluated with the program PROCHECK (CCP4)(17). Figure III.16 represents the Ramachandran plot of oligosaccharide-bound E377A structure. None of residues are in the disallowed region. The refinement statistics are listed in Table III.2.

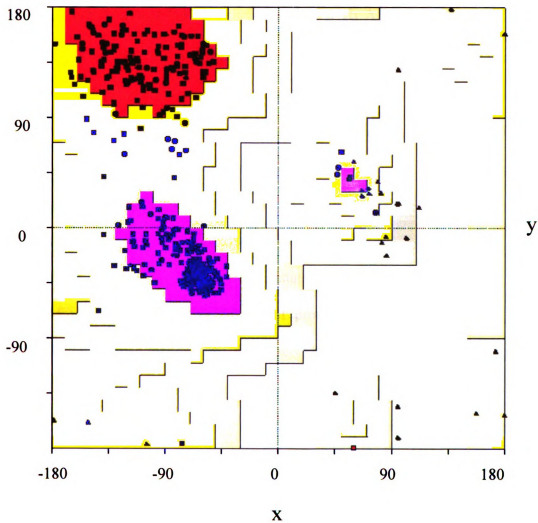


Figure III.16 Ramachandran plot of E377A /ADP /oligosaccharides complex structure. Phi (degrees) is x and Psi (degrees) is y.

III.3.5. GS Mutant E377A in Complex with ADP

Glu377 is a 100 % conserved residue throughout the whole glycogen synthase family across bacteria and animals. Mutating Glu377 to Ala diminishes the enzyme activity while the ability of the protein to bind substrate ADPGlc still remains (11). To better understand the role of Glu377 in the GS catalysis, the mutant E377A was co-crystallized with ADPGlc and crystals were obtained within 14 weeks.

E. coli GS E377A +ADP complex crystals were obtained from a buffer containing 7.8 mg/mL protein, 3 mM ADPGlc, 3 mM maltotriose, 40 % (w/v) PEG 4000 and 0.1 M HEPPSO (pH 7.5). After a 16 wk

incubation, the crystals grew to $0.2 \times 0.2 \times 0.2$ mm and diffracted to 2.3 Å (Figure III.17).

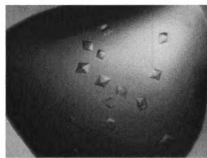


Figure III.17 Crystals of ADP-bound E377A.

The structure of *E. coli* GS wtGSa (PDB ID: 2QYY) was used as the phasing model for molecular replacement performed by MOLREP (CCP4). There is only one molecule in each asymmetric unit. The final structure includes residues 1-476, buffer molecule HEPPSO and an azide molecule. The last residue 477 and the His-6 tag attached to the E377A C-terminus were not traceable. The structure also includes 298 water molecules. Figure III.18 represents the Ramachandran plot for the E377A structure; none of the residues are in the disallowed region. The refinement statistics are listed in Table III.2.

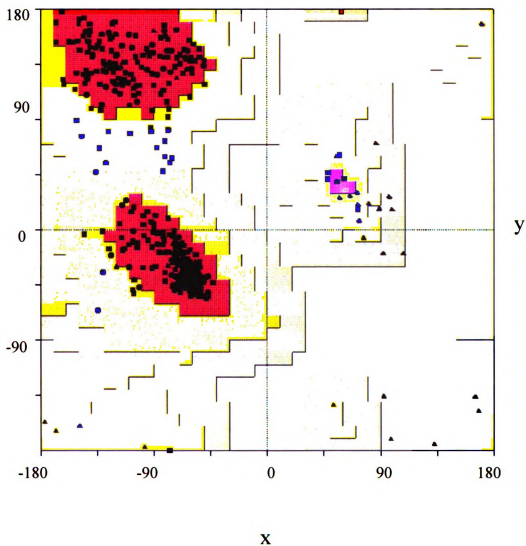


Figure III.18 Ramachandran plot of E377A-ADP-HEPPSO complex structure. Phi (degrees) is x and Psi (degrees) is y.

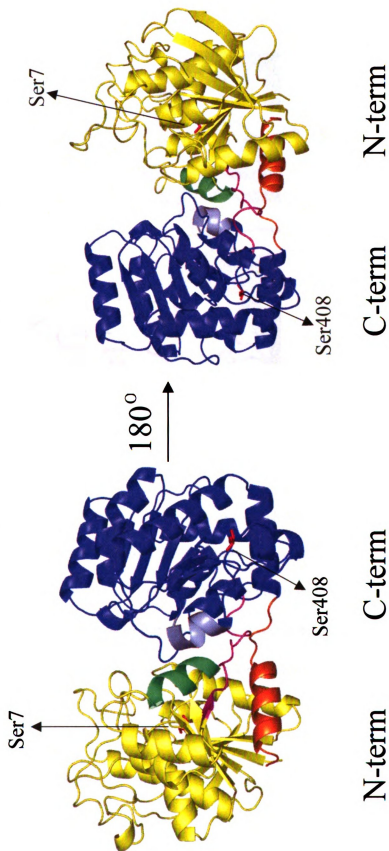
III.4. *Escherichia coli* GS Structures

III.4.1. Apo dmGS (C7S: C428S) Structure

E. coli GS is a 52.8 kDa protein with a total of 477 amino acid residues. Apo dmGS (C7S: C428S) exhibited a typical twin-Rossmann GT-B fold; the N- terminal domain (1-241) and C-terminal domain (250-457), each of which is composed of a “sandwich” of parallel β -sheets between α -helices, were separated by a deep cleft. The extended interdomain linker peptide (242-249) connects the N- and C-terminal halves. The helical tail, α 18 (458 - 476), crosses over from the C-terminal domain to pack against the N-terminal domain (Figure III.19). These results are consistent with the previously reported open form structure of PaGS (19) and AtGS (20). *E. coli* dmGS (C7S: C428S) exhibits specific activity and substrate ADPGlc affinity comparable to *E. coli* wild-type GS (11). The *apo* dmGS structure reveals that the mutant sites Ser7 and Ser408 are not close to the active site and are separately buried in the protein structure (Figure III.19), suggesting that this double mutation has little impact on the enzyme tertiary structure and the wild-type GS, when not bound with ligand, should adopt a similar open, inactive form as seen in apo dmGS.

As Figure III.19 shows, the two domains of apo dmGS are loosely connected. Most interdomain interactions occur in a defined region, in particular, between the linker peptide (242-249) and domain-spanning helix α 18 (458 - 476) (Val248 N-Phe460 O, 3.14 Å), and the N-terminal helix α 7 (212-220) and the C-terminal helix α 14 (398-403) (Thr214 OG-Gly399 O, 2.99 Å; Glu218 carboxyl oxygen-Gly399 N, 3.07, 3.09 Å) (Figure III.20).

Figure III.19 Overall structure of *E. coli* dmGS. The N-terminal domain (1-241) and C-terminal domain (250-457) are colored yellow and blue, respectively. The interdomain peptide 242-254 and domain-spanning helix $\alpha 18$ (458-476) are colored magenta and orange, respectively. The N-terminal loop 212-215 and helix $\alpha 7$ (215-220) are in green whereas the C-terminal helix $\alpha 14$ (398-403) is in light blue. The mutation sites of dmGS (C7S: C408S) are shown in red sticks.



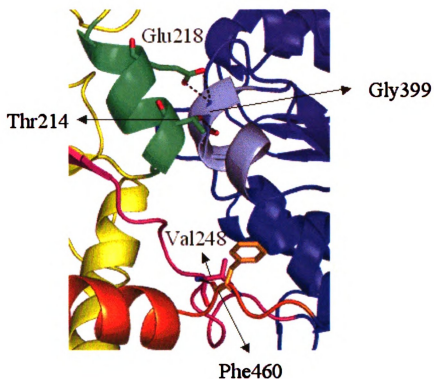


Figure III.20 Interdomain interactions between Glu218 and Gly399; Thr214 and Gly399, and Val248 and Phe460 are shown as dotted lines in *E. coli* dmGS. Loop 212-215 and helix $\alpha 7$ (215-220) are in green whereas the C-terminal helix $\alpha 14$ (398-403) is in light -blue. The linker peptide 242-254 and domain-spanning helix $\alpha 18$ (458-476) are colored magenta and orange, respectively.

III.4.2. GS - ADP - glucose - HEPPSO Complex Structures

GS-ADP-glucose-HEPPSO complex structures wtGSb, wtGS_c, and wtGS_d display a typical twin-Rossmann fold, which is similar to apo dmGS (C7S: C428S). But the two domains are very close together in wild-type GS complexes while the two domains are spread relatively far apart in apo dmGS. Several new interdomain interactions are observed between Asn162 and Gln304, and Lys15 and Glu357 in the wild-type GS complex narrow interdomain cleft. We designate the molecular organization seen in wild-type GS complexes with a interdomain narrow cleft as the closed form while that of apo dmGS with a wide cleft as the open form of GS. A detailed comparison of these two forms is described in section III.5.1.

Wild type GS complex structures wtGSb (2.20 Å), wtGS_c (2.26 Å), and wtGS_d (2.37 Å) are from three independent data sets, but exhibit almost identical structural organization as the structural displacement index RMSD is less than 0.1 Å between them. Except for small conformational difference of the HEPPSO ethyl end, ligands bound in the GS active site (ADP and glucose) in these three structures are in superimposable positions. Therefore we choose the wtGSb, the highest resolution structure, as a representative, to illustrate the interaction between ligand and protein.

The narrow cleft at the N- and C-domain interface serves as the ligand-binding pocket (Figure III.21). ADP binds mostly along the C-terminal wall and the glucose resides at the bottom of the cleft. HEPPSO is exclusively on the side of the N-terminal domain. The detailed description of ligand binding in GS follows.

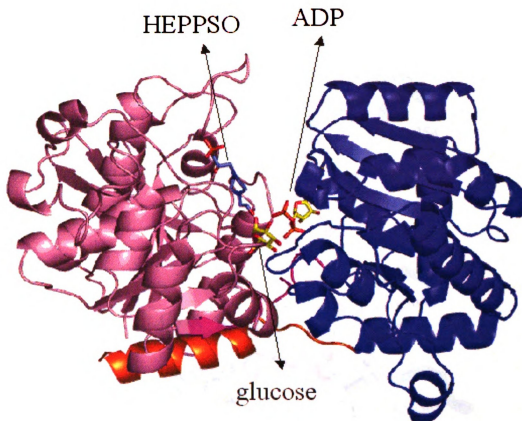


Figure III.21 Overall structure of the wtGSb complex. The N-terminal domain is in pink and C-terminal domain is in blue. The linker peptide 242-254 and domain-spanning helix $\alpha 18$ (458-476) are colored magenta and orange, respectively.

III.4.2.A. ADP binding site

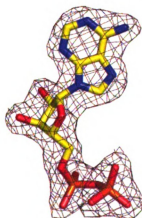
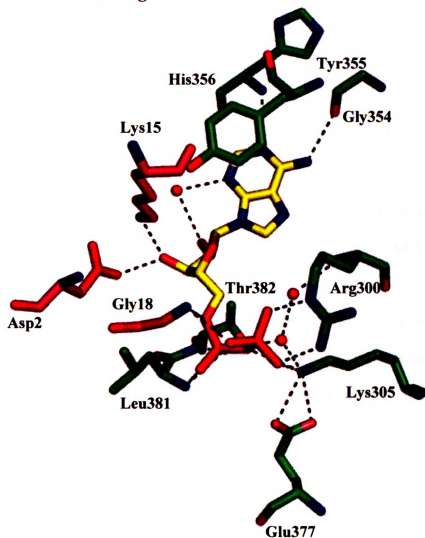


Figure III.22 Top: ADP (shown in yellow and in atom colors) bound in the active site of wild type *E. coli* GS. Hydrogen bonds between ADP and protein are shown as broken lines.

Residues from the N-terminal domain and the C-terminal domain are colored pink and blue, respectively. **Bottom:** 1.0 σ contoured 2Fo-Fc electron density map of the bound ADP in the active site.

The ADP heterocyclic adenine ring stacks on the conserved Tyr355 (Phe in animal GS and plant SS (Figure III.22). The adenine N1 atom donates a hydrogen bond to the backbone amide of His356 (3.08 Å). The adenine amide N6 is in contact with the Gly354 backbone amide (2.91 Å).

The ribose of ADP interacts with the Asp21 (2.84 Å) and Lys15 (2.74 Å) side-chain with its O3' hydroxyl group. The ribose moiety adopts a C3'-endo conformation relative to the adenine base (Figure III.22) and the water-mediated hydrogen bond between the ribose 2- hydroxyl (3.05 Å) and the adenine N3 atom (2.88 Å) presumably reinforces it.

The ADP phosphate groups are tucked back with respect to the adenine base so that O5* (connecting adenine and phosphate) and the distal phosphate O3B fall in the hydrogen bonding range of Gly18 (3.24 Å and 2.72 Å, respectively) (Figure III.22). The proximal phosphate is close to the invariant loop 374-381 and both phosphate O1A and O2A hydrogen bond to the main-chain amide of Leu381 (3.32 Å and 2.96 Å, respectively). The proximal phosphate O2A also hydrogen bonds to the Thr382 backbone amide (3.23 Å). The distal phosphate group is close to the loop 299-306 and extensively interacts with Arg300 and Lys305 through ionic interaction, direct hydrogen bonds, and water-mediated hydrogen bonds (Figure III.23).

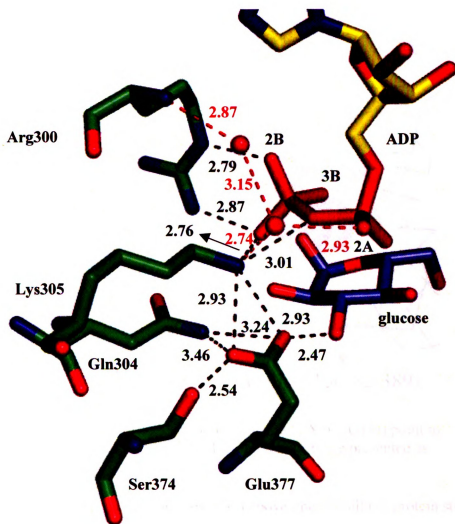


Figure III.23 Extensive interactions between ADP and residues Arg300, Lys305 and Glu377 in the wild type *E. coli* GS active site. Water molecules are presented as red spheres and water-mediated interactions are labeled as red.

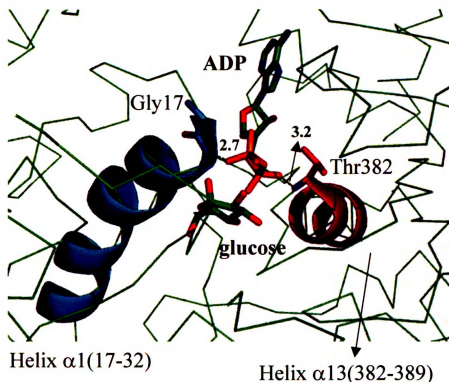


Figure III.24 Helix $\alpha 1(17-32)$ (cyan) and $\alpha 13(382-389)$ (purple) point to the ADP phosphate group. The NH-ends of both helices are presented as yellow ribbons.

Besides the local interactions described above, the overall GS protein structural organization also contributes to ADP binding, particularly the helix dipole moment of helix $\alpha 1(17-32)$ and helix $\alpha 13(382-389)$ whose NH-terminal ends point towards the ADP phosphate (Figure III.24). The helix dipole moment is caused by the aggregate effect of all the individual dipoles from the carbonyl groups of the peptide bonds pointing along the helix axis. Because the NH-terminal end of these two helices comprise neutral amino acid residues ($G_{17}GL$ and $T_{382}QL$, respectively), the NH-terminal positive charge from the overall helix dipole moment was not neutralized and directly interacts with the negative-charged ADP phosphates.

III.4.2.B. Glucose binding site

Wild-type GS complex wtGSb, wtGS_c, and wtGS_d were obtained from GS cocrystallization with substrate ADPGlc. However, the difference density maps of these wtGS structures all showed considerable gaps between the distal phosphate oxygen and the anomeric carbon of the glucose moiety of ADP, indicating the phosphate-glucosyl bond had been broken and an individual glucose is present.

The individual glucose derived from ADPGlc is exposed at the domain interface and extensively interacts with the C-terminal 373-380 loop and the N-terminal residues His161 and Asn162. The 2-hydroxyl group forms hydrogen bonds with the side chain of Asn162 (3.11 Å) and Gln304 (3.22 Å). The 3-hydroxyl group forms a 2.50 Å hydrogen bond with the carboxyl oxygen of Glu377 and makes additional hydrogen bonds with the backbone amide of Cys379 (2.98 Å) and Gly380 (2.89 Å) (Figure III.25). The 4-hydroxyl group makes a hydrogen bond with O1A of the proximal phosphate (2.41 Å), and with the Gly380 backbone amide (2.94 Å). The 6-hydroxyl group hydrogen bonds to the interdomain peptide residue Asn246's side chain carbonyl group (2.90 Å) and the His161 side chain (2.84 Å) (Figure III.25).

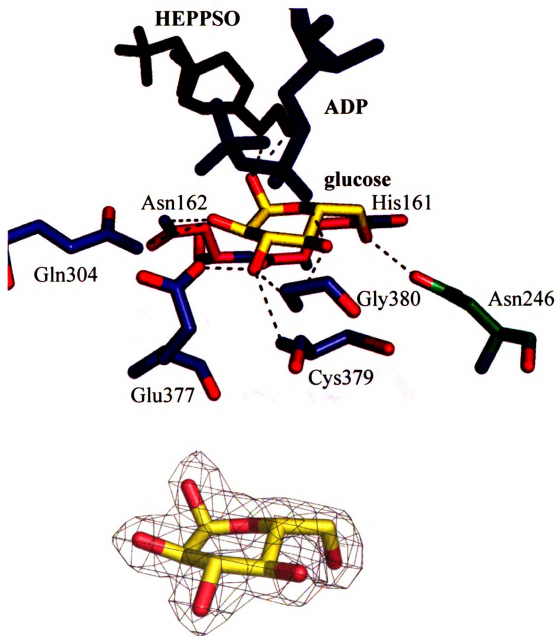


Figure III.25 Top: glucose (shown in yellow and in atom colors) bound in the active site of wild type *E. coli* GS. The nearby HEPPSO and ADP are shown as black and blue sticks. Hydrogen bonds between glucose and protein are shown as broken lines. Residues from the N-terminal domain and the C-terminal domain are colored pink and blue, respectively. Bottom: 1.0 σ contoured 2Fo-Fc electron density map of the bound glucose.

In all wtGS-ADP-glucose structures the anomeric C1 is 3.51 –3.65 Å from oxygen O3B of the distal phosphate, which is almost twice a regular carbon-oxygen bond length, clearly showing that the glucose phosphate bond is broken. On the other hand, the 2.50-2.99 Å distance between the glucose 1- hydroxyl and O3B of the distal phosphate suggests a hydrogen bond between the glucose and the leaving group phosphate (Figure III.26).

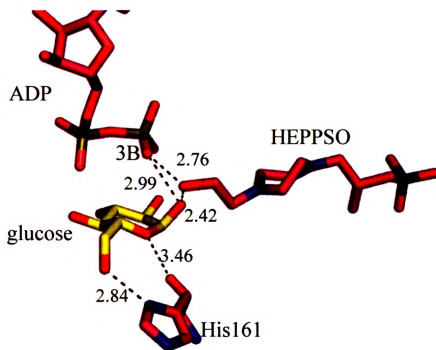


Figure III.26 Interactions between glucose, ADP and HEPPSO in the wtGSb active site.

III.4.2.C. Acceptor analogue HEPPSO binding site

All wtGS complex crystals grew in the presence of 4-(2-Hydroxyethyl) piperazine-1- (2-hydroxypropane sulfonic acid) (HEPPSO) (Figure III.27). Clear electron density was consistently observed for HEPPSO in the interdomain cleft of wtGSb, wtGSc, and wtGSd structures. Comparison of wtGS –ADP-HEPPSO complexes (wtGSb, wtGSc, and wtGSd) with the MalP-PLP-maltopentaose complex structure reveals that HEPPSO occupies the first two glucose unit positions of the acceptor maltopentaose in the MalP complex, indicating HEPPSO probably acts as an oligosaccharide acceptor analogue for *E.coli* GS (21). This interpretation is further supported by our GS E377A-ADP-oligosaccharide structure where the oligosaccharide also binds at the HEPPSO position in the active site.

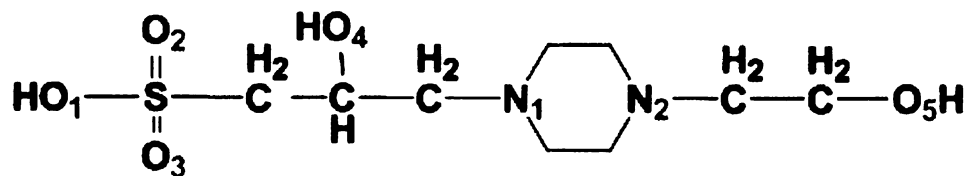


Figure III.27 Diagram of HEPPSO and its atom numbering.

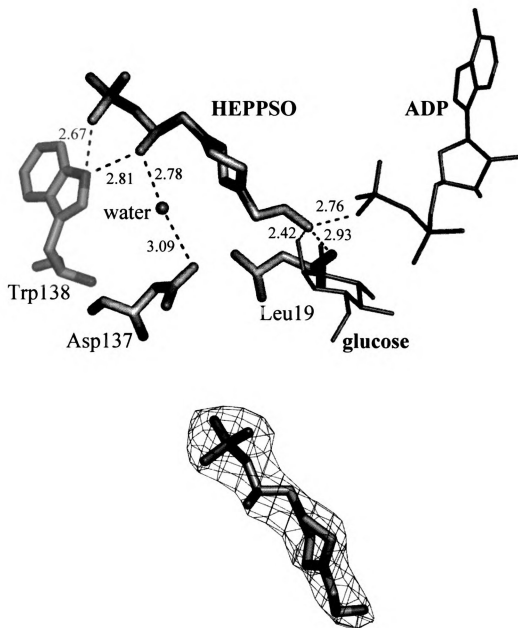


Figure III.28 Top: Interactions between HEPPSO and N-terminal residues, ADP, and glucose. **Bottom:** 1.0 σ contoured 2Fo-Fc electron density map of the bound HEPPSO.

HEPPSO lies along the N-terminal side of the crevice wall with its ethyl hydroxyl end near the glucose and ADP (Figure III.28). The ethyl hydroxyl end of HEPPSO is also close to the K₁₅TGGL motif loop and makes a 2.93 Å hydrogen bond with the Leu19 backbone. The sulfite end of HEPPSO points to the protein surface and interacts with the N-terminal residues along the cleft wall. The conserved Trp138 forms hydrogen bonds with both the branched hydroxyl group and the sulfite oxygen. The conserved residue Asp137 interacts with the branched hydroxyl group through a water-mediated hydrogen bond (Figure III.28).

Mutagenesis and modeling studies suggest that the conserved Asp137 and Trp138 are involved in recruiting and locking the glucan chain acceptor in place (22). The fact that HEPPSO cannot be substituted in crystallization experiments with HEPES, which has an ethane-sulfonic group instead of 2-hydroxypropanesulfonic group as seen in HEPPSO, indicates that the incorporation of HEPPSO into the protein requires the interactions of its branching hydroxyl O4 with Asp137 and Trp138.

III.4.3 GS-ADP-DGM-HEPPSO complex structure

While an individual glucose fits the density map of wtGSb (2.22 Å), wtGSc (2.26 Å) and wtGSd (2.37 Å) which was cocrystallized with ADPGlc for 12 weeks, the density map of wtGSa (2.83 Å), which was incubated with ADPGlc for four weeks, is not well-defined at the glucose site. Tentatively a D-glucopyranosylium cation (DGM) was assigned though the DGM deprotonation product D-arabino-hex-1-enitol fits just as well (Figure III.29).

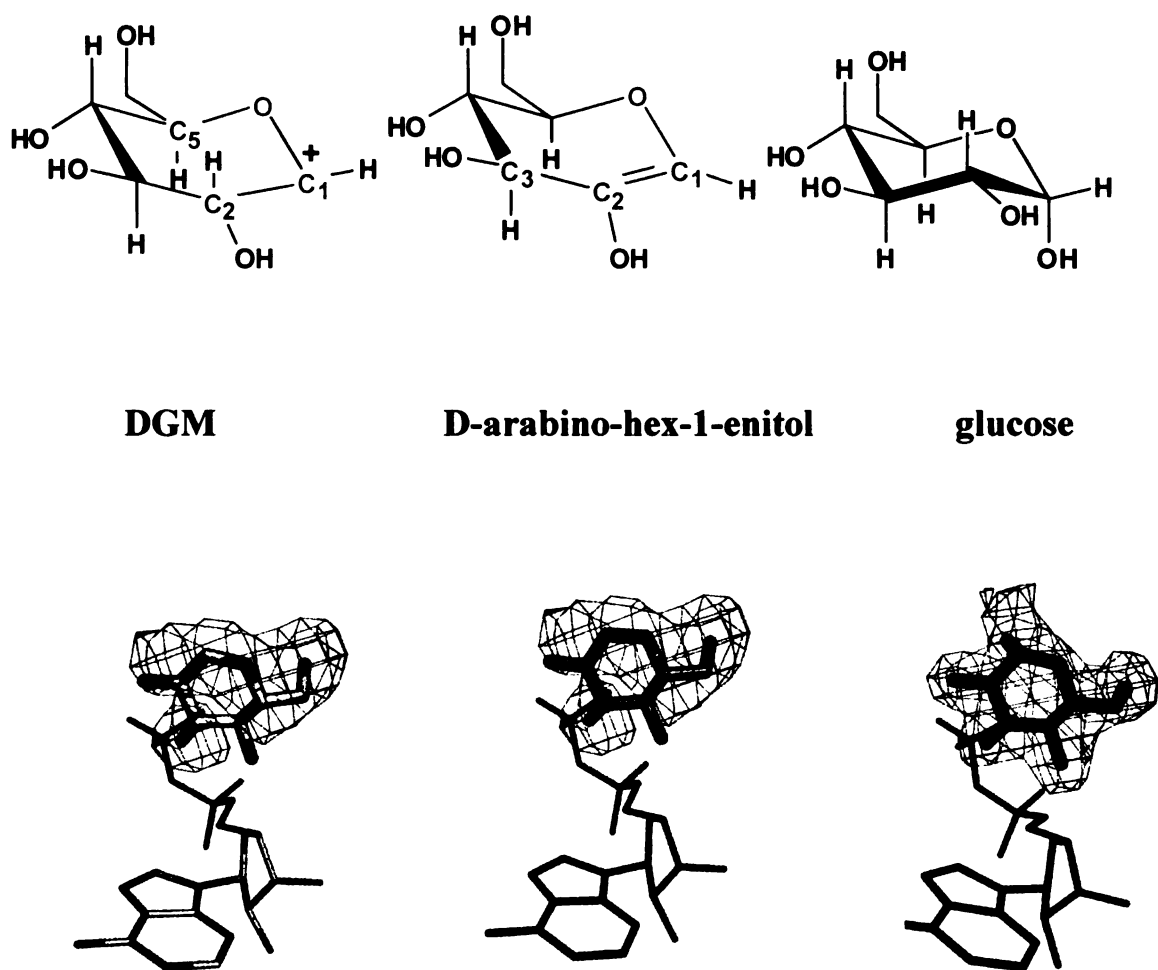


Figure III.29 2.5 σ Fo-Fc electron density map of wtGSa (2.83 Å resolution) with DGM (**left**), D-arabino-hex-1-enitol (**middle**) and wtGSb (2.22 Å resolution) with glucose (**right**). Co-planar atoms are in red. The ADP molecule is shown as lines for clarity.

The “D-glucopyranosylium cation (DGM)” occupies the equivalent position of glucose (Figure III.30) and most interactions between glucose hydroxyl groups and protein seen in wtGSb are retained in wtGSa for “DGM”. DGM differs from glucose in its positive charge on C1 and O5, the sp^2 hybridization of C1, and the subsequent partial planar conformation of the sugar ring. The distal phosphate O3B, HEPPSO ethyl end, and His161 backbone carbonyl group are close to the partially positively charged “DGM” C1 and O5, presumably providing electrostatic stabilization (Figure III.31). The detailed interacting distances in wtGSa and wtGSb are listed in Table III.4.

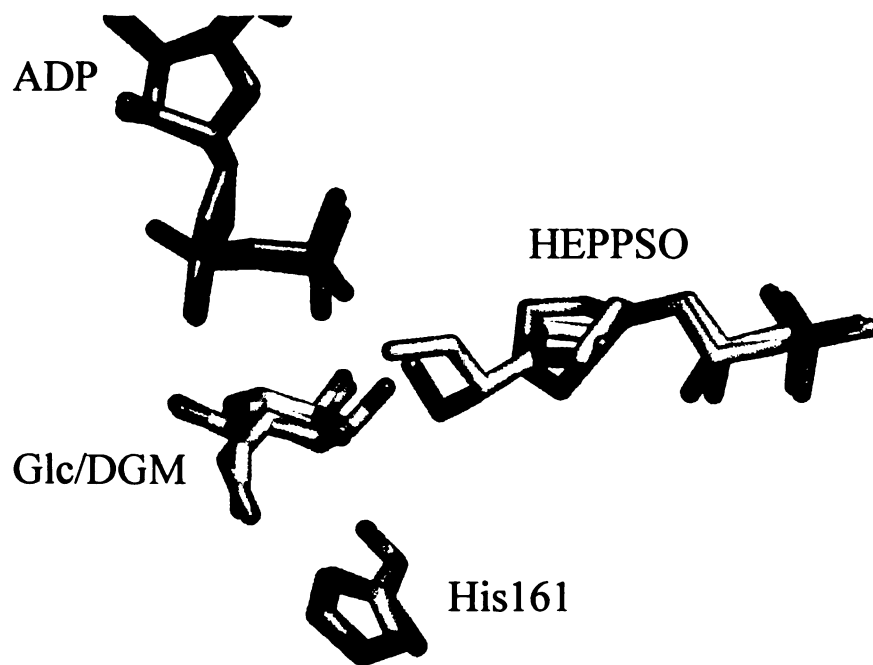


Figure III.30 DGM in wtGSa is in a superimposable position of glucose in wtGSb. Ligands in wtGSb are in yellow whereas those in wtGSa are atom-wise colored.

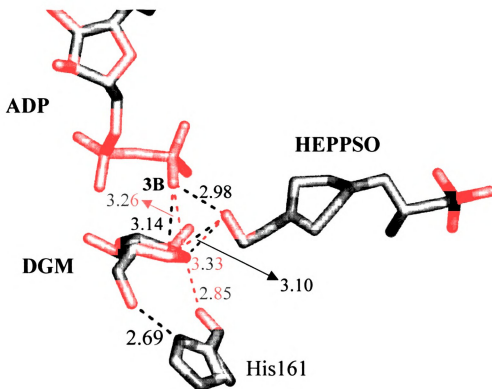


Figure III.31 Positively charged C1 and O5 of DGM are stabilized by ADP, HEPPSO, and His161. Interactions starting from the DGM C1 are shown as red dotted lines.

Table III.4 Comparison of the environment of DGM and glucose.

Atom A	Atom B	Distance A-B (Å) in wtGSa	Distance A-B (Å) in wtGSb
Glc/DGM C1	ADP O3B	3.26	3.65
	HEPPSO O5	3.33	3.31
	His161 O	2.85	2.90
Glc/DGM O5	ADP O3B	3.14	3.80
	HEPPSO O5	3.10	3.12
	His161 O	3.82	3.46
ADP O3B	HEPPSO O5	2.98	2.76
Glc 1-OH	ADP O3B	N/A	2.99
	HEPPSO O5	N/A	2.42
Glc/DGM 6-OH	His161 ND1	2.69	2.84

In spite of the limited resolution of the wtGSa (2.83 Å) crystal, the finding of DGM (or DGM deprotonation product D-arabino-hex-1-enitol) in the wtGSa complex active site is very significant. In particular, it provides strong evidence for an S_Ni/ S_N1 mechanism, as oppose to the once-prevailing double-displacement mechanism. The possibility that DGM is in the GS active site was reinenforced by the MalP-PLP- PO_4^{3-} maltopentaose structure (PDB: 2ASV, 1.95 Å) where the density map in its active site also suggests a glucose-like species lacking 1- hydroxyl group at the superimposable position. Strangely, the stable 1,5-anhydrosorbitol was built in, although DGM would also fulfill the density map there. Unfortunately, this MalP structure was only deposited in the PDB and is not officially published. As a result, the official quotation of DGM is missing and a firm piece of experimental evidence of DGM must be obtained before making the claim that DGM is the GS catalysis intermediate. To that end, we attempted to identify DGM in the GS active site using NMR and more details are described in section III.5.3. C.

III.4.4. E377A-ADP-HEPPSO Complex Structure

ADP-bound GS E377A displays an essentially identical structure to the wtGS complex (R.M.S.D of total 476 residues C α atoms = 0.53 Å). Unlike in the wtGS structure where ADPGlc was decomposed into ADP and individual Glc/DGM, only ADP is present in the active site of E377A (Figure III.32).

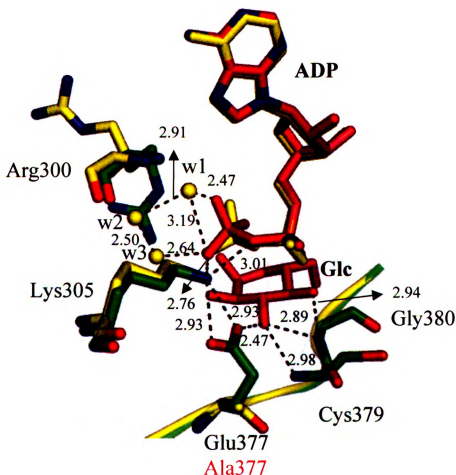


Figure III.32 Overlay of the active site of *E. coli* wtGSb and E377A. The residues and ligand of E377A structure are colored yellow whereas residues in wtGSb are in green and ligands are in magenta. Water W1, W2, and W3 are located between Arg300 and the ADP distal phosphate oxygen in the E377A structure.

The most significant difference between the active site of E377A and wtGSb is the loss of the glucose moiety of ADPGlc whose 3-hydroxyl group hydrogen bonded to the Glu377 side chain in wtGSb. Moreover, a drastic conformation change was observed at Arg300. Instead of being close to ADP and interacting with the ADP distal phosphate, Arg300 flipped its side chain out of the active site in the E377A-ADP complex structure.

III.4.5. E377A-ADP-Oligosaccharide Complex Structure

To understand the molecular basis of glycogen recognition to GS, we have determined the crystal structure of GS E377A in complex with ADP and oligosaccharide at 2.3 Å resolution. Four oligosaccharides were found in the interdomain cleft and on the enzyme's N-terminal surface (Figure III.33).

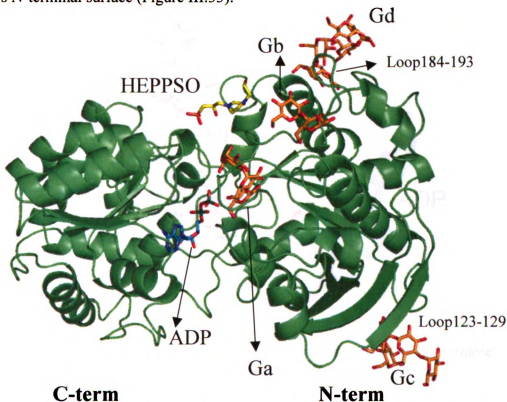


Figure III.33 Overall structure of *E. coli* GS E377A in complex with ADP and oligosaccharides. ADP and oligosaccharides are shown as sticks. The secondary structure elements which host residues interacting with the surface-bound oligosaccharide at Gc and Gd sites are labeled.

III.4.5.A. Interdomain cleft oligosaccharide binding

The 1σ -weighted difference Fourier map (2Fo-Fc) showed continuous density in the interdomain active site cleft (“Ga site”) (Figure III.34), corresponding to three well-defined glucose rings extending from the center toward the enzyme surface, which directly mimics the glycogen binding in the GS catalytic site. All protein residues that directly contact this oligosaccharide are well defined in the electron density map. The interaction scheme and the cartoon presentation are shown in Figure III.35 and III.36, respectively.

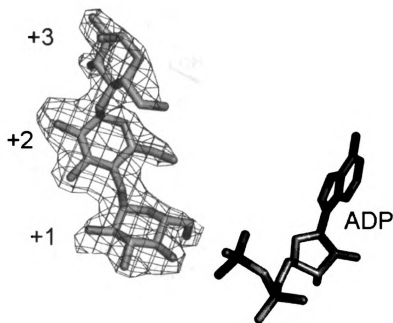


Figure III.34 2Fo-Fc electron density map contoured at 1σ with maltotriose at the Ga site. The ADP molecule is shown for reference.

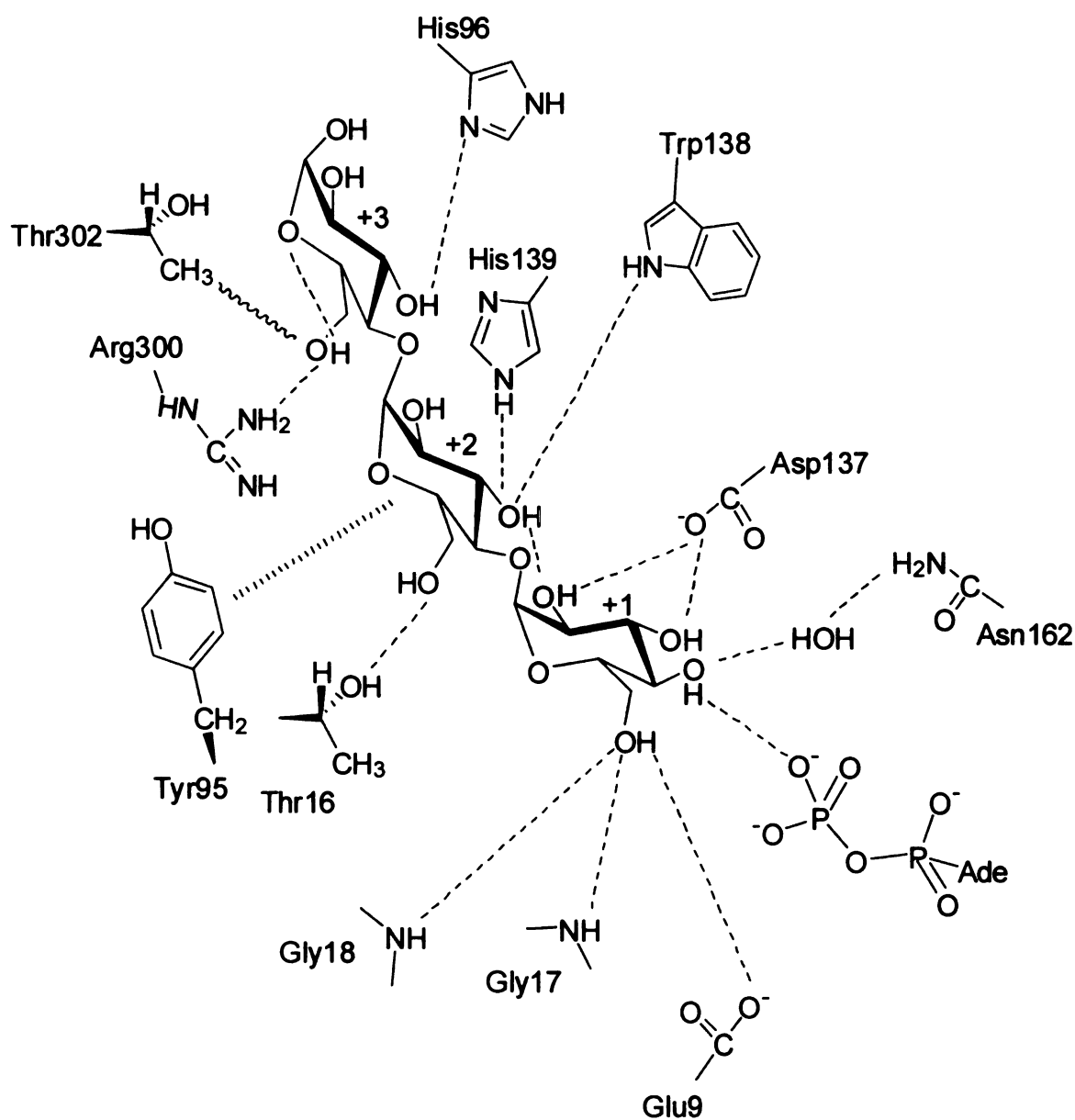


Figure III.35 Schematic representation showing the interactions between *E.coli* GS E377A residues and the bound maltotriose at Ga site. Observed hydrogen bonding and aromatic stacking are depicted with regular and wide dashed lines, respectively. Van der Waals interaction is marked with the wavy line.

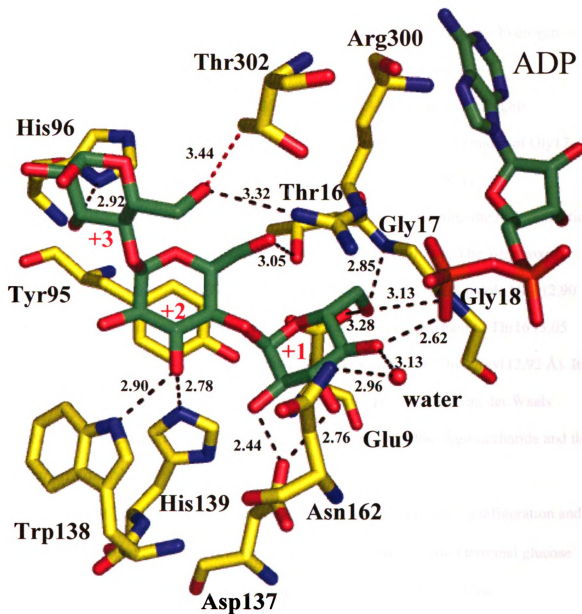


Figure III.36 The interaction between enzyme GS and the bound maltotriose at the Ga site.

The +1 sugar is buried close to ADP in the interdomain cleft (Figure III.35). Its 4-hydroxyl group is 2.62 Å from the ADP phosphate oxygen O3B. Bidentate hydrogen bonds are made between OH-2, OH-3 and Asp137 OD2 (2.44 Å and 2.76 Å, respectively). Its 6-hydroxyl group is close to the loop 9-19 that hosts the highly conserved K₁₅TGGL motif, forming hydrogen bonds to the backbone amides of Gly17 (2.85 Å) and Gly18 (3.13 Å) as well as the side-chain of Glu 9 (3.28 Å).

The sugar at subsite +2 stacks against Tyr95 (3.80 Å) presenting the only aromatic stacking between the bound oligosaccharide and GS (Figure III.36). The 3-hydroxyl makes interactions to the His139 imidazole NE2 (2.78 Å) and Trp138 indole NE1 (2.90 Å). The 6-hydroxyl is within hydrogen bond distance of the side chain of Thr16 (3.05 Å). The sugar at subsite +3 interacts with His96 ND1 through its 3-hydroxyl (2.92 Å). Its 4-hydroxyl group hydrogen bonds to the Arg300 mainchain and is in van der Waals contact with Thr302, presenting the only interactions between the oligosaccharide and the C-terminal domain.

The configuration of the reducing sugar in subsite +3 is in the α configuration and not the β configuration, which is the more favored configuration for a terminal glucose residue. We suggest that this hydroxyl group is not the reducing end of the oligosaccharide chain but there is at least one more α -1, 4 glycosidic link. The “invisibility” of the following sugars in the X-ray structure is probably due to disorder as there are few opportunities to make strong interactions with the protein on the surface.

The oligosaccharide at the Ga site in the GS interdomain cleft is in most intimate association with the active site and most interacting residues are highly conserved throughout GS, SS, MalP, and GP families. Mutating them, such as Asp137, Glu9,

His139 or Tyr95, decreased GS specific activity to a variable extent ((22) and Yep unpublished result). A close inspection of the overlaid G5- MalP complex (PDB ID: 1L6I)(21) and E377A-ADP-oligosacchride structure revealed that those conserved residues are lining the interdomain cleft and interact with oligosaccharide in a virtually identical manner (Figure III.37). This suggests that the glucan chain binding channels are constructed in a very similar way to guide substrates into or acceptors out of the deeply buried catalytic site in MalP and GS and this conclusion may apply to other polysaccharide processing enzymes like SS and GP, which were suggested to share a similar active site. It is noteworthy that mutating Tyr95 or His139 to Ala does lower the GS affinity for maltosaccharide, but not much to glycogen (Yep unpublished result). Such behavior is not difficult to understand as glycogen is a very long, branched glucan chain and probably has many more sugar binding sites on the GS enzyme than the short oligosaccharides have.

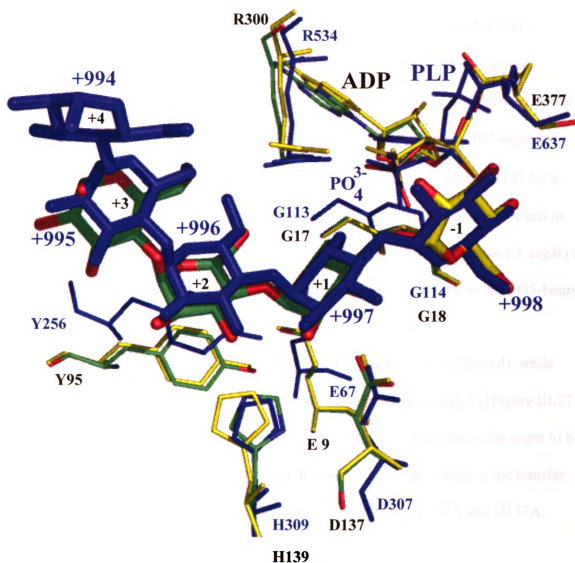


Figure III.37 Superimposition of the oligosaccharide- binding site in E377A-ADP-oligosaccharide (green), wtGS-ADP-glucose-HEPPSO (yellow), and MalP-PLP-G5 (PDB: 1L6I) (blue) complexes. The HEPPSO that occupies the comparable position of the oligosaccharide was omitted for clarity. GS residues are labeled in black while MalP residues are in blue.

As seen in Figure III.37, the +1 sugar in the E377A-ADP-oligosaccharide complex occupies almost an identical position as the second glucose (+997) of

maltopentaose in the G5-bound MalP structure and the rest of the oligosaccharide glucose units follow a similar orientation (21,23). The edge glucose +998 in the MalP-G5 complex overlaid well on top of the glucose molecule in the wtGS-ADP- glucose complex (PDB: 2QZS) and presumably reflects the conformation of transferred glucose (-1 sugar) in GS. The 174.73° glycosidic angle between the +998 and +997 sugar in MalP (-1 and +1 sugar in GS) is largely deviated from the average value of 114° for a regular polysaccharide (24). The neighboring residues presumably contribute much in the abrupt kink between the acceptor nucleophile (+1 sugar) and sugar donor (-1 sugar) as both sugars are involved in a number of hydrogen bonds to the protein, in both G5-bound MalP (23) and oligosaccharide-bound GS.

Glu377 makes a hydrogen bond to the -1 sugar (sugar to be transferred), while Asp137 forms a bidentate hydrogen bond to the +1 sugar (sugar acceptor) (Figure III.37). Both of them play an important role in precisely positioning and orienting the sugar to be added (-1 sugar) and the sugar acceptor, which is of extreme importance in the transfer efficiency. Actually, to date the most detrimental mutations are E377A and D137A, which cause a 10,000 and 8,400 fold decrease in GS activity, respectively.

III.4.5. B. The N-terminal surface oligosaccharide-binding sites

Three other oligosaccharide chains are observed at Gb, Gc and Gd sites on the N-terminal domain surface (Figure III.33). The Gb site is close to the interdomain cleft opening and the sugar +4 at the Gb site is approximately 6.5 Å from the +3 sugar at the Ga site. In contrast, the Gc and Gd sites are quite far away from the interdomain cleft and the closest distance between site Ga and Gc, and Ga and Gd are approximately 28.9 Å and 22.31 Å, respectively.

Two glucose rings (+4 and +5) are found at the Gb site and the electron density map is shown in Figure III.38. The +4 sugar OH-2 and OH-3 hydroxyls interact with the Gln194 side-chain. The +5 sugar OH-2 hydroxyl forms a close hydrogen bond to the His173 Ne. The +5 sugar 6-hydroxymethyl group is in a hydrogen bond distance to the Tyr103 mainchain oxygen. There are several van der Waals contacts between Tyr170 and sugars at the GB site (Figure III.38)

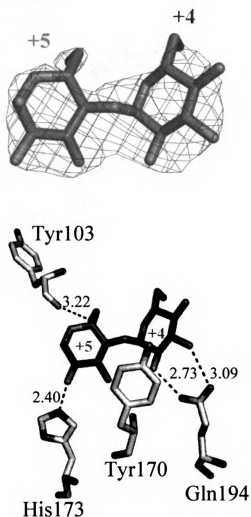


Figure III.38 Top: 2Fo-Fc electron density map contoured at 1σ with maltose at the Gb site. **Bottom:** Interaction between protein GS and the bound maltose at the Gb site.

The oligosaccharide at the Gc site consists of five glucose rings and is in proximity to β -sheet β 3 (55-58), β 4 (69-73) and loop 123-129. The oligosaccharide at the Gd site is comprised of six glucose units and is located between helix α 8 (229-237) and loop 184 - 93 (between α 5- β 9) (Figure III.33). Figure III.39 shows the electron density map for oligosaccharides at the Gc and Gd sites.

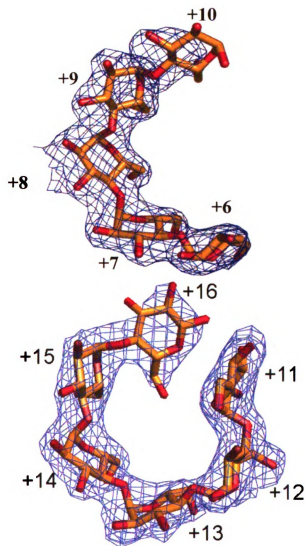


Figure III.39 **Top:** 2Fo-Fc electron density map contoured at 1 σ with maltopentaose at the Gc site. **Bottom:** 2Fo-Fc electron density map contoured at 1 σ with maltohexaose at the Gd site.

Due to the curled-up conformation, the surface-bound Gc and Gd exhibit few interactions with protein GS. The oligosaccharide at the Gc site interacts with GS only at subsites + 7 and + 8 (Figure III.40 and 41). The + 7 sugar 6-hydroxyl group is 3.23 Å from Gln55 NE2. +7 sugar 2- and 3- hydroxyls are in the van der Waals distance of Phe 127 CE. The Asp125 carboxyl group is in close contact with the adjacent 2-hydroxyl (2.52 Å) and 3-hydroxyl (3.24 Å) of the + 8 sugar and forms “double-handle” hydrogen bonds. In addition, the + 8 sugar 2-hydroxyl is in contact with the positively charged Arg38 side chain (3.01 Å).

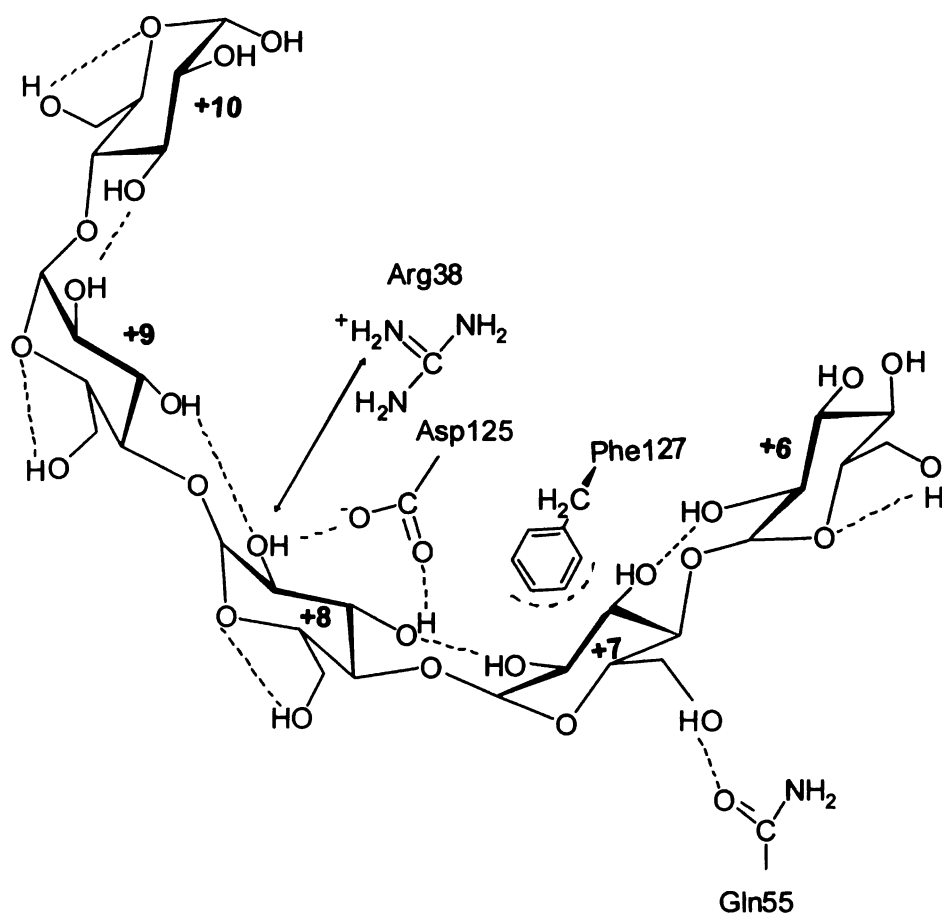


Figure III.40 Schematic diagram of interactions between protein and the bound maltopentaose at the Gc site. Observed hydrogen bonding and Van der Waals interaction are depicted with dashed lines. The ionic interaction is marked with arrow.

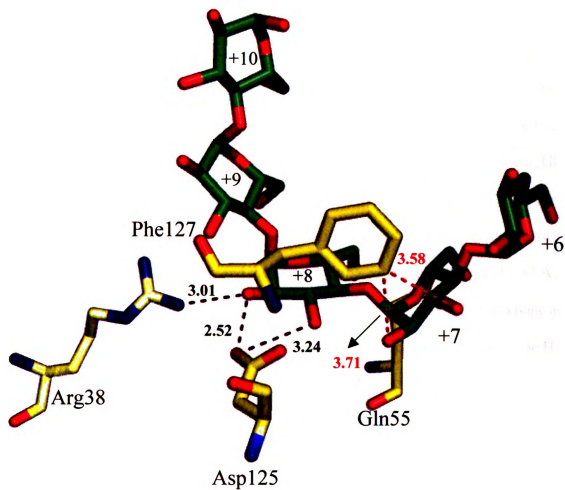


Figure III.41 Interaction between protein GS and the bound maltopentaose at the Gc site.

The oligosaccharide at the Gd site is the most circular oligosaccharide among the three bound to GS. The first sugar (+11 sugar) almost reaches the last glucose unit (+16 sugar), as the +11 sugar 3- hydroxyl is only 2.71 Å away from the +16 sugar 1-hydroxyl group (Figure III.42).

The interaction between the oligosaccharide at the Gd site and the protein mostly occurs at one end of the oligosaccharide, in particular at subsite +11, +12, and +13 (Figure III.43). The Ile186 backbone amide hydrogen bonds to the +11 sugar 6- hydroxyl (2.63 Å). The Lys199 amine is in close contact with adjacent hydroxyl groups OH-2 (2.93 Å) and OH-3 (3.20 Å) of the +12 sugar. On the other face of the +12 sugar, Gly227 and Gly230 are in hydrogen bond distance to the 6-hydroxymethyl group (2.62 Å and 2.97 Å, respectively). At subsite +13, “double-handle” hydrogen bonds occur between the adjacent 2- and 3-hydroxyls and the Glu190 carboxyl group (2.84 Å and 2.83 Å, respectively). The 3-hydroxyl group also hydrogen bonds to the Asn192 backbone amide (2.97 Å). The only interaction on the other end of oligosaccharide is between the His187 NE2, the +15 sugar 2-hydroxyl, and the +16 sugar O4.

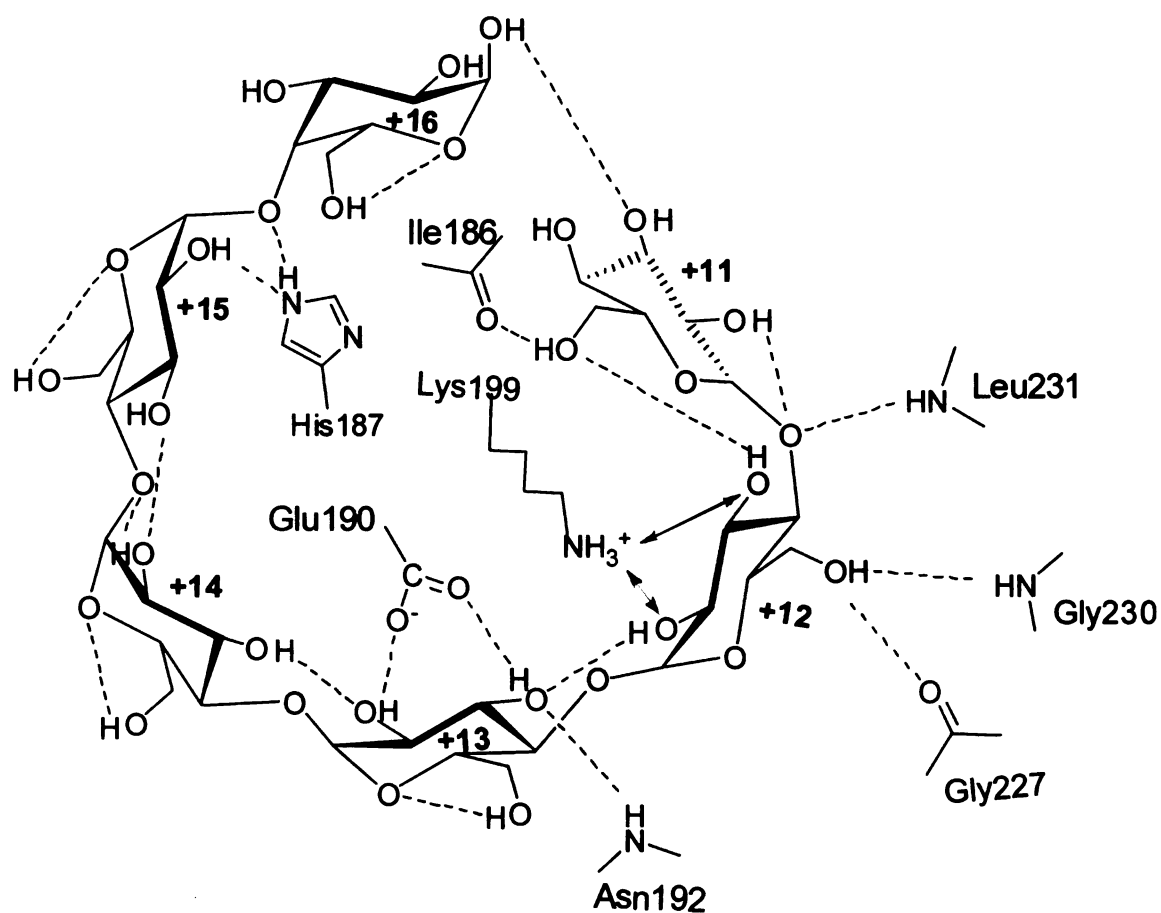


Figure III. 42 Schematic diagram of interaction between protein and the bound maltohexaose at the Gd site. Observed hydrogen bonding and ionic interactions are depicted with dashed lines and arrows.

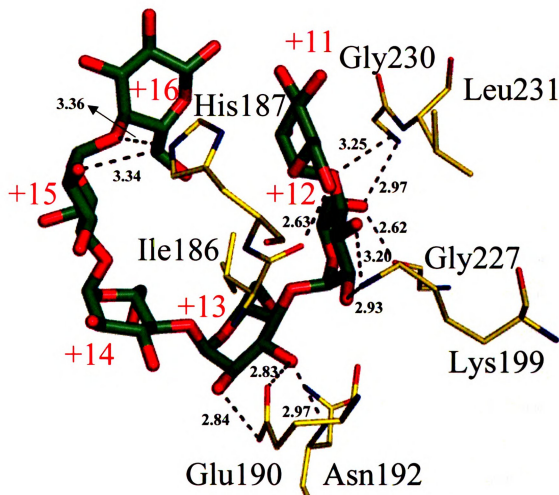


Figure III.43 Interactions between protein and the bound maltohexaose at the Gd site.

III.4.5.C. The glucose unit configuration in oligosaccharide-bound E377A

All the glucose units in the oligosaccharide-bound GSE377A complex adopt a typical 4C_1 chair conformation (25). Among 16 sugar units bound in the E377A-ADP-oligosaccharide complex, twelve of the glucose hydroxymethyl groups hold gauche-gauche (*gg*) and three have gauche-trans (*gt*) conformations. The OH-6 hydroxyl of the +2 sugar is in close contact with the Thr16 hydroxyl side-chain and the torsion angles of O5-C5-C6-O6 and C4-C5-C6-O6 are 129.13 ° and -106.44 °, respectively, moderately

beyond the definition of a gauche conformation (30° to 90° or -30° to -90°). The prevalent low-energy *gt* and *gg* conformation of hydroxymethyl groups are largely reinforced by their interaction with the intra ring oxygen O5 ($2.98 \pm 0.2 \text{ \AA}$) (Figure III.35, 40 and 42).

III.4.5.D. Binding mode analysis

Distinct bidentate hydrogen bonds from a protein carboxyl side chain to the adjacent sugar OH-2 and OH-3 hydroxyls are observed at each binding site (Asp137 at the G5a site, Asp125 at the G5c site, and Glu190 at the G5d site) and catch our attention. Such “bidentate” binding was also seen in the polar interaction between the Lys199 side-chain Nz and the +12 sugar. In contrast, the carbohydrate - aromatic interaction usually seen with sugar - processing proteins such as amylase (26,27), endoglucanase CelA(28), maltodextrin translocation protein maltoporin(29) are rare in our oligosaccharide-bound E377A structure (only between Tyr95 and the +2 sugar). It appears that GS predominantly relies on double - handle hydrogen bonding to recognize and orient glycogen.

III.4.5.E. Oligosaccharide-binding causes little protein conformation change

Both E377A-ADP-HEPPSO and E377A-ADP-oligosaccharide complexes belong to the I 41 space group with cell dimension $a=b=126.3 \text{ \AA}$, $c=152.3 \text{ \AA}$. The E377A-ADP-HEPPSO complex displays a closed conformation, partly due to the presence of HEPPSO in the interdomain cleft. Co-crystallization and further soaking with oligosaccharides brings three oligosaccharides to E377A, specifically in the interdomain cleft and on the N-terminal surface. The RMSD value of C α atoms between the E377A-ADP-HEPPSO and E377A-ADP-oligosaccharide complex is only $0.34 \text{ \AA}$ (30). It appears

that crystal packing and protein structural organization were not considerably affected by the action of the oligosaccharide replacing HEPPSO inside the interdomain channel or binding on the enzyme surface.

The few conformational changes that were observed were at the side-chain of His139 whose imidazole ring moved 1.03 Å to form a hydrogen bond to the +2 sugar and at Ile186 which rotated its side chain to avoid being too close to the +11 sugar OH-4 group (otherwise 1.81 Å). The side - chain of Arg300 adopts an identical conformation as in the wtGS- ADP- glucose -HEPPSO complex, but totally differs from that in the E377A-ADP-HEPPSO complex where the glucose moiety of ADPGlc is absent. More discussion regarding Arg300 conformational change in response to ligand binding and its function is in section III.5.3.B.

III.4.5.F. Oligosaccharide conformation analysis

To investigate the possible conformation alteration in oligosaccharide caused by protein binding, three GS-bound oligosaccharides at binding sites (Ga, Gc, and Gd) are compared to each other, to another GT-B enzyme GP-bound maltopentaose, and a proposed glycogen model by Sundaralingam (31) and Goldsmith(32).

The proposed helical conformation of the glycogen chain was referred to as a minimum energy conformation that has a characteristic intramolecular hydrogen bond between the 2-hydroxyl from one glucopyranose ring to the 3-hydroxyl of the adjacent glucopyranose. We construct an ideal glycogen helix to maintain such a hydrogen bond network as well as the guidelines derived from a typical maltose structure (24,33). They include: 1) the 1.4 Å α - (1,4) glycosidic linkage and the 114° glycosidic angle, 2) the

107° torsion angle of O5_n-C1_n-O4_{n+1}-C4_{n+1}, 3) the 122.7° torsion angle of C1_n-O4_n-C4_{n+1}-C3_{n+1}.

At the surface Gb, Gc and Gd sites, oligosaccharides maintain O3_n - O2_{n-1} hydrogen bonds between most glucosyl units. In contrast, the oligosaccharide at the Ga site that is held in the narrow interdomain cleft possesses a relatively rich interaction with protein, adopts a more extended conformation, and one O3_n - O2_{n-1} hydrogen bond out of two glycosidic linkages is lost. (Figure III.34). The distances between O3_n - O2_{n-1} in all three oligosaccharides are listed in Table III.5 and Figure III.44 shows the overlay of these oligosaccharides.

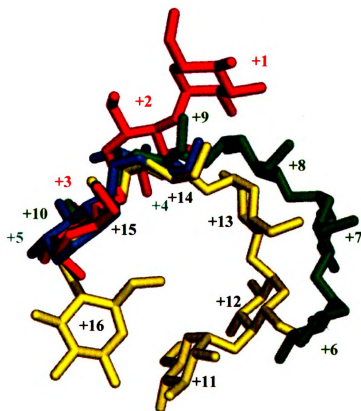


Figure III.44 Overlay of oligosaccharides at the Ga (red, 1-3), Gb (cyan, 4-5), Gc (green, 6-10), and Gd (yellow, 11-16) sites.

Table III.5 Conformation of oligosaccharides when bound to GS, MalP, and GP

binding sites	Phi(ϕ)	Psi(ψ)	τ	O _{2n} -O _{3n+1}
G5a				
(+1G)-(+2G)	99.48	120.22	114.33	3.05
(+2G)-(+3G)	66.15	85.06	106.00	3.76
G5b				
(+4G)-(+5G)	95.10	115.07	108.66	3.14
G5c				
(+6G)-(+7G)	101.96	100.76	113.00	2.85
(+7G)-(+8G)	102.50	125.86	114.16	2.92
(+8G)-(+9G)	107.60	124.62	111.48	2.83
(+9G)-(+10G)	101.58	114.42	109.21	2.96
G5d				
(+11G)-(+12G)	86.32	-51.41	114.64	5.53
(+12G)-(+13G)	101.52	119.05	115.51	3.18
(+13G)-(+14G)	120.20	123.68	116.15	2.78
(+14G)-(+15G)	114.47	128.36	114.49	2.84
(+15G)-(+16G)	47.01	104.43	110.60	4.95
Maltopentaose bound to MalP(1L6I.pdb)				
998-997 (equivalent to -1G to +1G)	71.96	59	-174.73	4.59
997-996 (equivalent to +1G to +2G)	102.93	109.82	112.92	2.66
996-995 (equivalent to +2G to +3G)	81.19	100.03	110.99	3.48
Maltopentaose bound to GP(1P2B.pdb)				
S7-S6	79.84	103.98	117	3.66
S6-S5	105.15	111.77	114.85	2.69
S5-S4	74.33	85.82	116.87	3.94
S4-S3	113.05	111.63	117.33	2.55

torsion angle: phi: O5(n)-C1(n)-O4(n+1)-C4(n+1)

torsion angle: psi: C1(n)-O4(n+1)-C4(n+1)-C3(n+1)

τ : glycosidic linkage angle: C1(n)-O4(n+1)-C4(n+1)

IUPCA-IUB 1983

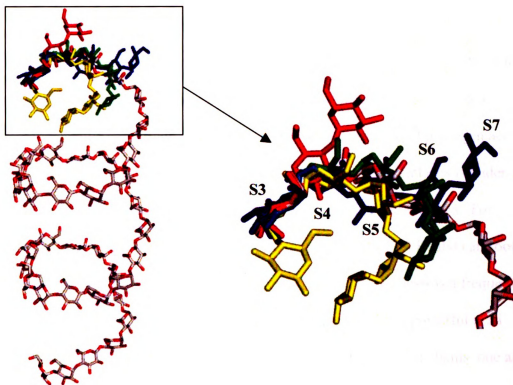


Figure III.45 Comparison of the ideal glycogen helix model, glycogen phosphorylase –bound maltopentaose (blue, 1P2B.pdb), oligosaccharides at GS Ga (red), Gb(cyan), Gc (green), and Gd (yellow) sites.

We also overlaid the GP-bound maltopentaose, glycogen helical model and GS-bound oligosaccharides. The sugar S3 in the GP-bound maltopentaose, sugar +3, +10, and +15 in GS are the glucosyl units with the least interaction with protein and were purposely chosen as the superimposition basis. As shown in Figure III.45, there is appreciable deviation in each sugar chain conformation compared to the ideal glycogen helical model, which are likely due to the protein binding. Actually, none of the protein bound oligosaccharides ($n \geq 5$) in the protein data bank has shown the faithful glycogen helix conformation to date.

III.5. *E.coli* GS Structure Analysis

Most bacteria transform external nutrients to glycogen for energy storage. Usually the amount of glycogen in bacteria is huge, because it is not only an energy depot for daily consumption, but also a reservoir that ensures the bacteria can survive any unforeseen negative conditions. In some bacterial species the mass of glycogen accounts for fifty percent of the dry cell weight(34). In humans, glycogen synthesis is under strict control of “secondary messengers”, but it still can be very fast and effective. For example, an average adult after a standard meal (approximately 700 Kcals) can soon produce up to 100 –120 g of glycogen. Apparently, glycogen synthesis is a frequent, fast and effective biological process. In that process, GS operates like a powerful mini machine that adds thousands of glucose units from ADPGlc to glucan chains, one at a time. The structural studies of *E.coli* GS described here address important questions like what type of molecular organization is designed to suit such a function? How do the substrates ADPGlc and the glucan chain bind to the enzyme? What does the active site look like? What is the mechanism of GS?

III.5.1. Open-Close: GS Dynamic Motion

All of our *E.coli* GS X-ray structures display a typical twin Rossmann fold with a cleft between two domains. However, the interdomain cleft of the *E.coli* GS complex is remarkably smaller than that in the *apo* dmGS, as well as the previously reported GS structures. Superposition of the C-terminal domain of the *apo* dmGS and the representative wtGSb structure revealed that the displacement on the apex of their N-terminal domains is as much as 9.3 Å and the overall domain rotation angle is approximately 15.2° according to the program DynDom(35) (Figure III.39). Since the

RMSD between the N-terminal domains (the α -carbon atoms of residues 1-241) and the C-terminal domains (α -carbon atoms of residues 251-375 and 382-477) between *apo* dmGS and the wtGS complex are only 0.679 Å and 0.605 Å, respectively, the domain-domain movement is basically rigid, and we designate the conformation observed in the wtGS complexes as the closed form of GS.

The *E.coli* GS active site is composed of residues on flexible loops between the β strand and α helix emanating from the domain boundaries (Figure III.46). Comparison of the open and closed forms of *E.coli* GS explains the role of such a motion in enzyme catalytic turnover: only after the domain–domain closure are these substrate-binding residues brought together to construct a competent active site.

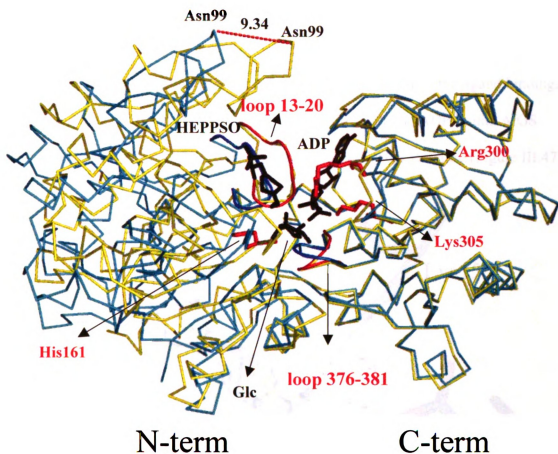


Figure III.46 Superposition of the C-terminal domains from dmGS (cyan) and wtGSb (yellow). Ligand bound in wtGSb (HEPPSO, ADP and glucose) are shown as black sticks. Structural elements essential for catalysis are colored red and their counterparts in the open form dmGS are colored blue. Residues critical for glucosyl transfer (His161, Arg300, and Lys305) are shown as red sticks.

III.5.1 A. Comparison of ADP conformations in GS open and closed form

In an effort to understand how the open and closed form affect the ligand binding, we compared the binding mode of ADP in the open form with the ADP-bound AtGS structure (20) and the closed form ADP-containing *E. coli* wtGS structure (Figure III.47).

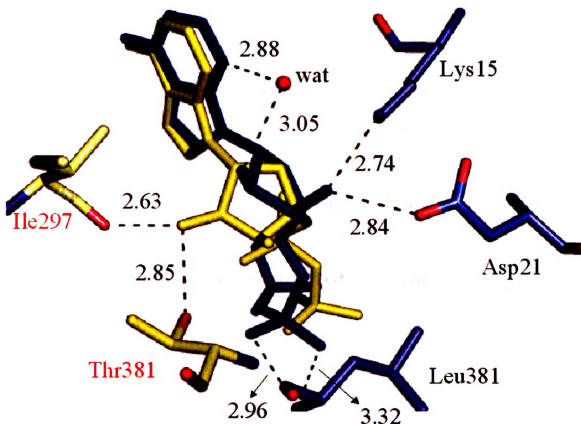


Figure III.47 Structural comparison of ADP molecule in open form *A. tumefaciens* GS (yellow, 1rzu.pdb) and closed form *E. coli* GS (blue). The residues from *A. tumefaciens* GS are labeled in red and their carbons are colored yellow while those from *E. coli* GS are labeled in black and their carbons are colored blue. The interaction between ADP and protein are shown as broken lines.

Compared to ADP in the AtGS open form, the adenosine diphosphate group in the closed form of *E.coli* wtGS exhibited a twisted conformation at the ribose ring and the proximal phosphate. The ribose ring in the closed form of *E.coli* wtGS rotated approximately 51.8° and the ribose 3-hydroxyl group then fell within hydrogen bonding range of the side chain of Lys15 and Asp21, which were approximately 8 Å away in the AtGS open form. The proximal phosphate was in contact with Leu381 and formed two hydrogen bonds to its backbone amide. The distal phosphate of ADP, which extensively interacts with C-terminal residue Arg300 and Lys305, did not show the same extent of twist as ribose and the proximate phosphate. It seems that the difference in conformation of ADP between open and closed forms of GS is initiated at the ribose where the N-terminal residues approach and interact, suggesting that closure of the enzyme is required for substrate ADPGlc to adopt a catalytically competent orientation in the protein.

III.5.1.B. Large displacement of the KTGGL loop in GS open and closed forms

The KTGGL motif is highly conserved among glycogen synthases and starch synthases and was suggested to be involved in substrate binding (36,37). The N-terminal KTGGL motif is located in the loop 13-20 at the N- and C-terminal domain interface, more than 5 Å away from the active site in the apo-dmGS open form (Figure III.48). The domain-domain closure brings this N-terminal loop into the vicinity of the C-terminal domain to interact with both substrate ADPGlc and acceptor. Lys15 and Gly18 both make hydrogen bonds with the ribose hydroxyl and β -phosphate oxygen, respectively (Figure III.22). Lys15 also participates in a domain-spanning hydrogen bonding network

with Glu357 and Tyr355, presumably further facilitating the cross- talk between the N- and C-terminal domains.

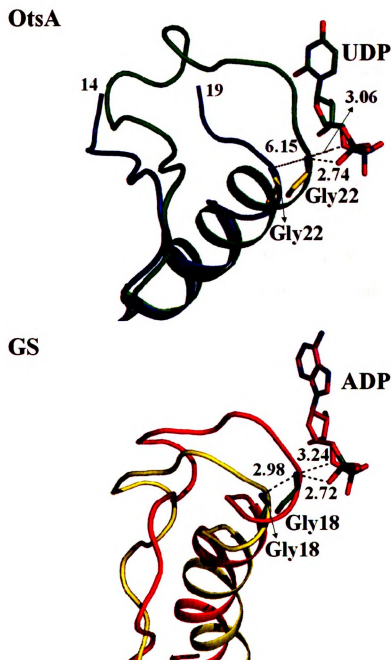


Figure III.48 Structural comparison of (**bottom**) loop 13-20 in apo-dmGS (yellow) and wtGSb complex (red) and (**top**) the equivalent loop 14-24 in the UDP/ imidazole /Glc-6-P bound OtsA (green, 1gz5.pdb) and UDPGlc-bound OtsA (blue, 1uqu.pdb). Gly22 in OtsA and Gly18 in GS interacting with the phosphate oxygen are shown as sticks. Ligands other than UDP (OtsA) and ADP (GS) are omitted for clarity.

The last three residues of the KTGGL motif are conserved in GT-B retaining enzymes, such as GP(38), MalP(21,23) and Trehalose –6-phosphate synthase (OtsA)(39,40). The GGL sequence at the beginning of helix IH1 in rabbit muscle glycogen phosphorylase has been shown to be in van der Waals contact with heptulose-2-phosphate located at the enzyme binding site of the natural substrate, glucose-1-phosphate(38).

OtsA transfers glucose from UDP-glucose to glucose-6-phosphate to form α,α -1,1-trehalose-6-phosphate(39,40). The OtsA GGL motif is also on the loop, 14-24, structurally equivalent to the *E.coli* GS loop 13-20 (Figure III.48).

In the substrate UDPGlc-bound OtsA structure, OtsA loop 14-24 is partially disordered (14-19) while the rest of the loop (20-24) is 7-9 Å away from the substrate UDPGlc (40), reminiscent of the comparable distance between the GS loop 13-20 and ADPGlc seen in the open apo-dmGS structure(Figure III.48). In the presence of the substrate analogue UDP/imidazole and acceptor Glc-6-P, the entire loop 14-24 of OtsA becomes structured and adjacent to the UDP molecule where the GGL motif interacts with the substrate donor and locks the substrate into position for the glucosyl transfer with its Gly22 (Gly18 equivalent) hydrogen- bonding O5* and β -phosphate oxygen of UDP (39) (Figure III.48).

In conclusion, OtsA manages to encapsulate the active site with the GGL motif via local conformational changes as evidenced by the small N-terminal domain displacement ($\sim$ 1-1.5 Å). In contrast, *E.coli* GS undergoes a large global domain-domain rotation (9.48 Å N-terminal displacement) to move the GGL motif into the interdomain catalytic center.

III.5.1.C. What causes the enzyme to close?

There are two major forms of *E.coli* GS, an open form, where the two domains are spread relatively far apart, and a closed form where the two domains are very close together. The apo-dmGS structure adopts the open form while all ligand-bound *E.coli* GS structures represent the catalytically active, closed form of the enzyme (Figure III.46).

Attempts to crystallize GS in its catalytically active, closed form have been made by several groups over the past five years before we first achieved this goal and obtained not only one, but several ligand-bound closed form GS structures. The one thing in common in those closed-form GS structures is the presence of an acceptor analogue HEPPSO or an oligosaccharide in addition to ADP, or ADP and glucose. In the absence of an acceptor or acceptor analogue, only the open form has been obtained both in our and other labs. For instance, incubation of AtGS with substrate ADPGlc alone resulted in the open form with only ADP visible, which binds to the C-terminal side of the interdomain cleft (20). Incubation of *E.coli* dmGS with ADPGlc results in a vacant interdomain cleft and wide-open conformation.

HEPPSO was added as a buffer in GS crystallization trials, but turns out to appear in the active sites of four closed wtGS structures, suggesting that HEPPSO is critical for capturing the closed form of *E.coli* GS. Although ADP, glucose and HEPPSO all bind in the GS interdomain cleft, HEPPSO exclusively interacts with the N-terminal domain whereas ADP and glucose predominantly interact with residues from the C-terminal domain. However, the interaction between HEPPSO, ADP, and glucose, directly and

through the K₁₅TGGL motif, establishes the cross-domain network which brings the two domains together (Figure III.49).

Almost equally important in catching the closed form of GS is the impotency of HEPPSO for glucosyl transfer; mostly due to the distance between the HEPPSO tip hydroxyl group and the anomeric carbon of glucose (3.3 Å). If HEPPSO were able to approach the glucose of ADPGlc closely enough and perform a nucleophilic attack, the GS machinery would start to run incessantly with frequent opening and closing, until all of the substrate ADPGlc is depleted. Then it would be extremely hard to maintain a closed form on the time scale of crystallization.

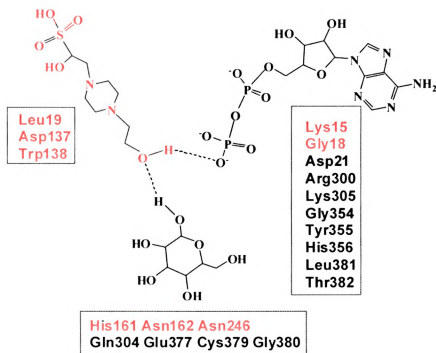


Figure III.49 Schematic diagram showing the cross-domain network between ligand HEPPSO, glucose and ADP, which chiefly interacts with the N-terminal (red) and C-terminal (blue) residues, respectively.

In the oligosaccharide-bound GS E377A structure, the oligosaccharide that replaces HEPPSO in the interdomain cleft also predominantly interacts with N-terminal residues (Figure III.36). The connection between this oligosaccharide and the C-terminal domain is mainly mediated through the ADP phosphate group and the KTGGL motif. It is noteworthy that there is no electron density for the glucose moiety of ADPGlc or an individual glucose molecule from ADPGlc. Due to such a lack of the sugar to be transferred, the glucose addition is not able to launch and the enzyme remains in the closed form.

In summary, our success in capturing the closed form of GS should be attributed to the simultaneous presence of ADP and oligosaccharide acceptor, or ADP, glucose and the acceptor analogue HEPPSO. The cross-domain network spanning the N-terminal residues, oligosaccharide acceptor/ HEPPSO, ADP / ADP+glucose, and the C-terminal residues provides internal bonds for GS to keep the two domains closed.

III.5.1. D. Domain -wise ligand binding facilitates GS open-close motion

The structural analysis of the open form apo-dmGS and closed form wtGS-ADP-glucose-HEPPSO complex as well as E377A-ADP-oligosaccharide complex suggests that domain-domain closure forms a competent active site in the center and domain-domain opening facilitates the product release. As a result, domain-domain opening and closing is believed to accompany GS glycogen chain production.

We notice that ligand ADP and glucose are predominantly bound by the C-terminal domain whereas acceptor oligosaccharide (and acceptor analogue HEPPSO) binds almost exclusively to the N-terminal domain, either in the interdomain cleft or on the surface. Such a domain-wise ligand binding facilitates the GS open-close motion because it

prevents the constraints that would result from a two-domain spanning bound glycogen polymer.

III.5.1.E. Substrate binding order

As ADP/Glc is deeply buried in the GS interdomain cleft, one would think that the binding of glucosyl donor ADPGlc must precede the binding of the acceptor glucan chain, which occupies the channel. However, the GS open form presents another possibility regarding substrate-binding order. Both ADPGlc and glucan chain could bind to the enzyme's open form without constraint on their binding order. Only when both bind does closure and formation of the active site happen. This scenario allows independent binding of substrates and does not require a specific binding order, therefore avoiding potential substrate inhibition (where the 2nd substrate binds first and must dissociate before the other binds).

Precedence exists for both the ordered bi-bi mechanism, in which ADPGlucose binds first, followed by the acceptor and the random kinetic mechanism, in which both donor and acceptor bind to enzyme without a specific order. Structural and calorimetric binding studies on GT-B retaining enzyme N-acetylgalactosaminyltransferase (GTA) and galactosyltransferase (GTB)(41) suggested that the binding of their donor facilitates the formation of the acceptor-binding site and therefore it occurs prior to acceptor binding. On the other hand, the extensive analysis of the GT-B retaining enzyme GP reveals that it employs a rapid equilibrium, random kinetic mechanism, in which both donor and acceptor substrates must be bound prior to the chemical event but without constraint on their binding order (reviewed in (42)).

Although there is no experimental evidence to rule out or support one over the other regarding the GS substrate binding order, the competitive inhibiting effect of acceptor analogue HEPPSO to ADPGlc ($K_i = 0.15 \text{ M}$) suggests that the acceptor binding somewhat weakens substrate ADPGlc binding to GS, indicating ADPGlc might bind to GS prior to the acceptor.

Interestingly, the interdomain-bound acceptor analogue HEPPSO shows substantially low inhibiting effect to glycogen ($K_i = 0.8 \text{ M}$), indicating the majority of glucan may bind outside the active site where they are not disturbed by the presence of HEPPSO. Such a hypothesis is consistent with the mutagenesis studies, which show that mutating residues that interact with the interdomain bound oligosaccharide such as Tyr95 or His139 to Ala does not lower the GS affinity to glycogen (Yep unpublished result).

III.5.2. Important Residues

Previous mutagenesis and kinetic studies identified several residues that are significant to GS activity (Table III.6) (11,22). Now, based on our GS-ligand complex structures, we are able to illustrate their particular roles in GS catalyzed glucan chain elongation.

Table III.6 Kinetic parameters of wild type glycogen synthase and mutants

	$V_{\max}$ (U/mg)	Decrease (-fold)	K _m (ADPGlc) μ M	Residue roles suggested by our GS structures
WT	500 $\pm$ 65	1	18 $\pm$ 3	
D137A	0.07 $\pm$ 0.01	8140	21 $\pm$ 4	holding +1 Glc in place
H161A	0.8 $\pm$ 0.1	710	200 $\pm$ 16	positioning -1 Glc and stabilizing intermediate DGM
R300A	0.22 $\pm$ 0.02	2590	150 $\pm$ 15	acid catalyst making ADP a better leaving group
K305A	0.46 $\pm$ 0.05	1240	32 $\pm$ 5	acid catalyst making ADP a better leaving group
E377A	0.05 $\pm$ 0.01	10,000	70 $\pm$ 30	
E377Q	0.020 $\pm$ 0.001	25,000	61 $\pm$ 12	holding -1 Glc in place for transfer
E377D	10 $\pm$ 1	57	1500 $\pm$ 300	

III.5.2.A. Asp137 is the +1 sugar - positioning residue

Mutating Asp137 to Ala is an almost lethal mutations of GS as the enzyme activity drops 8,140- fold (22). In the interdomain catalytic cleft of the oligosaccharide-bound E377A structure, Asp137 side chain OD2 makes bidentate hydrogen bonds to the +1 sugar OH-2 (2.44 Å) and OH-3 hydroxyl (2.76 Å) (Figure III.36). As the +1 sugar is the immediate acceptor of the substrate ADPGlc glucose, its orientation is critical to the transfer efficiency and the bidentate hydrogen bonds from Asp137 play a critical role in correctly positioning the +1 sugar.

III.5.2.B. His161 participates in –1 sugar positioning and intermediate DGM stabilization

His161 forms a 2.84 Å hydrogen bond between its side chain ND1 to the 6-hydroxyl of Glc in the wtGSb structure (Figure III.26) and a 2.69 Å hydrogen bond to DGM 6-OH in the wtGSa structure (Figure III.31). This hydrogen bond plays an important role in positioning the –1 Glc as mutating His to Ala at this position caused a 710- fold decrease in enzyme specific activity (22). The His161 backbone carbonyl group is only 2.85 Å from the DGM C1, presumably offering electro static stabilization to the positively charged intermediate DGM.

III.5.2.C. The catalytic Lewis acid Arg300 side-chain switches in and out of the GS active site in response to the presence of ADP

Mutating Arg300 to Ala decreases GS specific activity 2,600 fold (22). The ADP, glucose-bound wtGS structure reveals that Arg300 is close to the phosphate group and probably acts as a Lewis acid to donate a proton to the ADPGlc phosphate to facilitate its departure from the glucose moiety (Figure III.22).

A virtually identical conformation of Arg300 where its side chain is proximal to the ADP phosphate was seen in almost all GS structures with ADP or an ADP moiety bound, including the ADP-AtGS complex (20). In contrast, when there is no ligand in the active site as in the apo-dmGS, apo-AtGS, and apo-PaGS structures, the Arg300 (Arg299 in AtGS; Arg257 in PaGS) side-chain is flipped out of the active site. It rests on the loop 327-331 and is stabilized by the hydrogen bond from the Ala329 backbone amide and several Van der Waals contacts from Ala329 and Gly330 (Figure III.50).

Apparently, the Arg300 side chain switches in and out of the GS active site in response to the presence of the ligand ADP, particularly the distal phosphate. However, in the ADP-bound E377A structure several water molecules occupy the space between ADP phosphate and Arg300. As a result, the Arg300 side chain flipped away from the active site and the distance between Arg300 NH2 and phosphate O1B is approximately 10.4 Å (Figure III.32).

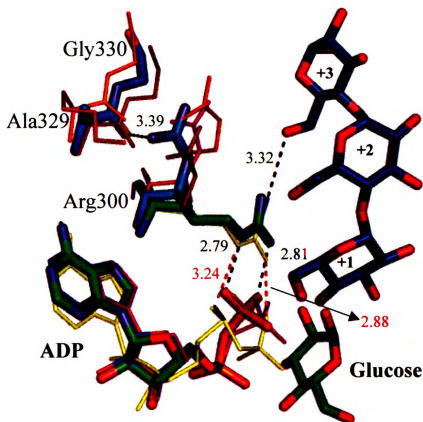


Figure III.50 In apo-*E. coli* dmGS (cyan, colored by atom), apo-AtGS (1rzv.pdb, ruby), apo- PaGS (2bis.pdb, orange) and E377A-ADP-HEPPSO- complex (magenta). The side-chain of Arg300 and its equivalents are out of the GS active site and are in contact with Ala329 and Gly330 (Lys in PaGS). In ADP, Glc-bound wtGS (2qzs.pdb, green, atom-wise colored), oligosaccharide-bound E377A (blue), and ADP-bound AtGS (1rzu.pdb, yellow), Arg300 and its equivalent side-chain are close to the ADP phosphate and their interaction are shown as dotted lines (black lines for wtGS and red lines for AtGS complex). Hydrogen bonds between the Arg300 side-chain and Gly330 in apo-dmGS, and the Arg300 side-chain and maltotriose in oligosaccharide-bound E377A complex are also shown as dotted lines. Residues are labeled according to the *E. coli* GS sequence.

III.5.2.D. Catalytic Lewis acid residue Lys305

Lys305 is located close to the distal phosphate and is a potential candidate for protonation of the phosphate to make it a better leaving group (Figure III.24). Consistent with this supposition is the fact that mutating Lys305 to Ala resulted in more than a thousand fold decrease in specific activity (22). Equivalent lysine 305 and arginine 300 residues occupy identical positions relative to the phosphate in all available GT-B retaining enzyme structures (GP (43), MalP(2asv.pdb) , OtsA (40), AGT (44)) (Figure III.51) that do not require a divalent Mn^{2+} ion for their enzymatic action. It therefore seems likely that nature either uses an exogenous divalent metal ion as for GT-A enzymes, or positively charged side chains in the majority of GT-B enzymes to make phosphate a better leaving group.

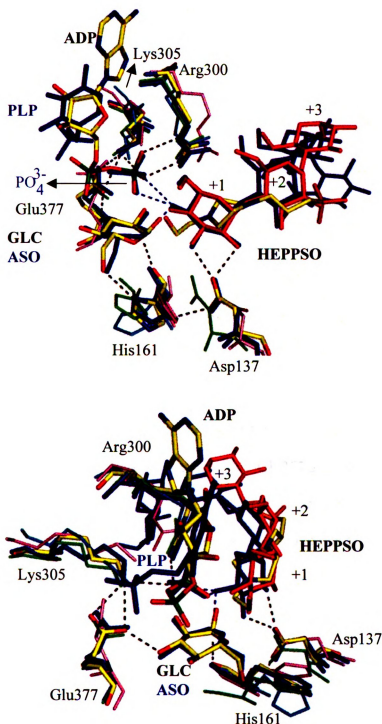


Figure III.51 Structural comparison of ADP, glucose binding site and active site residues between *E. coli* GS (in yellow and atom colored, wtGSb), MalP (blue, 2asv.pdb), rabbit R-GP (pink, 1gpa.pdb), OtsA (green, 1uqu.pdb) and AGT (cyan, 1y6f.pdb). MalP Ligand, PO_4^{3-} , ASO (1,5-anhydrosorbitol) and maltopentaose, occupies the equivalent positions of the ADP distal phosphate group, glucose and HEPPSO in the wtGSb structure, respectively. Maltotriose in E377A complex was also shown (red sticks). Residues are labeled according to the *E. coli* GS sequence. Top: the front view. Bottom: the back view.

III.5.2.E. Glu377 is responsible for the –1 sugar positioning and helps Lys305 maintain positive charge

Glu377 deserves special attention because mutation at this position caused the most deadly effect to GS specific activity (E377A: 10,000 fold decrease; E377Q: 25,000 fold decrease). Glu377 is widely conserved in MalP, GP, trehalose phosphorylase, N-acetylglucosaminyltransferase (GTA), galactosyltransferase (GTB) and OtsA (equivalent residue Asp) (Figure III.51) and is the first residue in the E-X7-E motif of some retaining glycosyltransferases, such as eukaryotic glycogen synthases (GT3), α -N-acetylglucosaminyltransferase (GT4), starch synthases, bacterial glycogen synthase (GT5) and α -mannosyltransferase (GT15). Substitution of *E.coli* GS Glu377 by Gln/Ala(11), MalP Glu637 by Gln(45), and GTB Glu303 by Ala (41) all result in a drastic reduction in enzyme activity. Mutating Glu510 (the E377 equivalent) completely inactivated the human muscle glycogen synthase (46). Glu377 and its equivalents were therefore thought to be either the catalytic nucleophile or the general acid/base catalyst for these glycosyltransferases (11,41,46-48), however we found that Glu377 is located on the α -face of the glucose to be transferred and is relatively far both from the C1 glucose position and from the ADP β -phosphate position in our ECGS structures, which is inconsistent with Glu377 playing a direct catalytic role in the reaction. Positively charged Arg300 and Lys305 interact with the substrate phosphate group (Figure III.22), and are therefore much more likely to play the role of a proton transfer agent than Glu377. In fact, the Glu377 carboxyl oxygen atoms are 2.93 Å from Lys305 (Figure III.52). Perhaps Glu377 is deprotonated and negatively charged which would help the neighboring Lys305 maintain its positively charged side chain. More importantly, Glu377

side-chain forms a short hydrogen bond to the glucose 3-hydroxyl and plays an important role in positioning and stabilizing the α -1 glucose (Figure III.52). Our E377A-ADP-HEPPSO and E377A-ADP-oligosaccharide structures both show that the glucose moiety is missing in their GS active sites. As shown in Figure III.52, an extensive hydrogen bond network is present between Glu377, catalytic residues Lys305 and Arg300, and glucose. Each of Glu377 carboxyl side-chain oxygen accepts two hydrogen bonds. Replacing either one of them to nitrogen which has one lone pair and can only accept one hydrogen bond, would seriously disrupt the whole network, which perfectly explains the dramatic enzyme activity decrease (25,000 fold) observed in mutant E377Q. To understand why the enzyme retained the appreciable enzyme activity when Glu377 was mutated to Asp, we modeled Asp at the position 377 in the conformation of the OtsA Asp361 (equivalent to GS Glu377) and find that Asp137 is still able to interact with glucose and hold it in place (Figure III.52). Due to the shorter side-chain of Asp, the interactions between the position 377, Gln304, and Lys305 observed in the wtGS structure are lost, which probably leads to the 57- fold activity decrease in the mutant E377D.

Taken together, Glu377 positions and stabilizes the α -1 glucose and is an extremely important residue to GS catalysis. Although Glu377 was proposed to play a direct catalytic role and our structure does show that Glu377 contributes in the maintaining the positive charge of catalytic residue Lys305 which makes ADP a better leaving group, its glucose locator role is dominating and has an overwhelming impact to the enzyme activity.

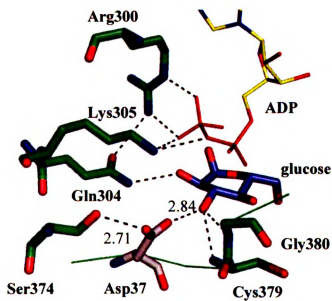
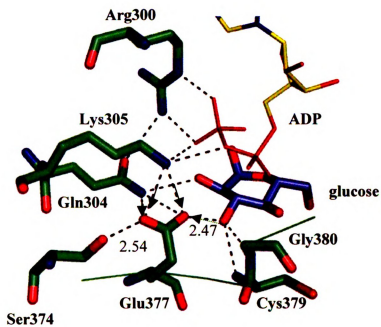


Figure III.52 Top: The Glu377 environment in the wild-type GS structure. Arrows indicated the hydrogen bond proton donating direction. **Bottom:** The interaction between the modeled Asp377 and neighboring residues.

III.5.3. GS Complex Structures Support S_N1 Mechanism Versus The Double Displacement Mechanism

III.5.3.A. DGM intermediate

Our three separate high-resolution wtGS structures (wtGSb, wtGS_c, wtGS_d) from 12 week incubations with ADPGlc have shown individual glucose at a superimposable position in the active site. In a crystal harvested in a shorter time, 4 weeks, we see density for most of the glucose moiety, but do not see density for the O4 glucose oxygen or for the O1 glucose oxygen. We therefore believe that we have captured some intermediate form that will subsequently be hydrolyzed to the glucose we see in the above-mentioned three structures. Interestingly, in the active site of MalP which is also a GT-B retaining enzyme and shares an almost identical active site as *E.coli* GS, there exists the density for a similar glucose-like species lacking a 1-hydroxyl group at the superimposable position of glucose in our wtGSb, wtGS_c, and wtGS_d. Although that density was assigned to 1,5-anhydrosorbitol (2ASV.pdb), the oxocarbenium ion DGM would fit the density just as well and hints that such a species is also formed in the wtGSa active site.

A DGM intermediate is one essential component of an S_N1 mechanism that was first suggested for the GT-retaining enzyme glycogen phosphorylase (GP) about twenty years ago(49). The observation of DGM and its hydrolyzation product glucose in the *E.coli* GS active site as well as DGM in the MalP active site strongly suggest that those GT-B retaining enzymes, which share active site structure, proceed through an S_N1 mechanism. Consistent with our hypothesis, several oxocarbenium cation mimics have been shown to strongly inhibit GT-B retaining enzymes such as *E.coli* GS (50), *Schizophyllum commune* trehalose phosphorylase (51-53), rabbit muscle glycogen

phosphorylase(52,54,55), and starch phosphorylase (26). Moreover, the kinetic behavior of various oxocarbenium cation mimics on *Schizophyllum commune* trehalose phosphorylase suggested that the positive charge of the oxocarbenium ion is focused on C1 rather than O-5, which is consistent with the requirement for GT glucose transfer (52).

III.5.3.B. DGM stabilization in GS

Like all carbocations, DGM is electron deficient and has long been considered as an unstable, transient species. The estimated half-life of DGM in open solvent is on the order of pico -seconds, slightly longer than bond vibration times(56,57). The extreme instability of DGM was the major obstacle for the scientific community to accept an S_N1 mechanism in the past. But today, our wtGS complex structures point out that there are multiple interactions to stabilize DGM and it is possible for DGM to live much longer when bound in the active site of GT-B retaining enzymes than alone when exposed to solvent.

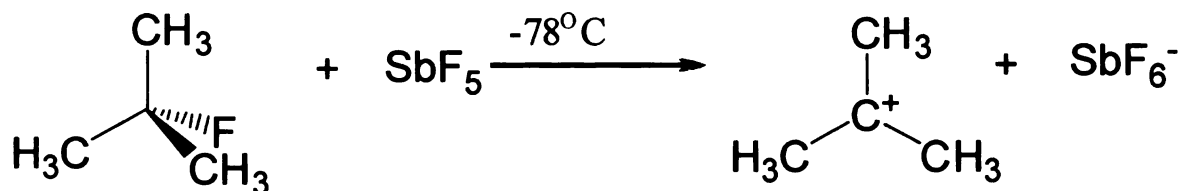
The most prominent electrostatic stabilization of DGM is provided by the substrate phosphate moiety whose oxygen O3B is 3.26 Å and 3.14 Å to glucose C1 and O5, respectively in the wtGSa structure (Figure III.32). In fact, a stronger inhibition and tighter binding was observed when oxocarbenium cation mimics were tested with phosphate, such as D-gluconic acid 1,5-lactone (GL) and phosphate (26) , and nojirimycin tetrazole and phosphate (55). This again suggests that the nearby leaving group phosphate is involved in stabilizing the DGM intermediate.

In addition to the substrate ADP moiety, the protein itself also contributes to stabilize the intermediate DGM, particularly through residue His161. The His161

backbone carbonyl group is only 2.85 Å from the positively charged DGM C1, presumably offering electrostatic stabilization to the intermediate DGM.

III.5.3.C. Effort to verify the existence of carbocation DGM using NMR

C. K. Ingold and E. H. Huphs first proposed that a carbocation is the key intermediate in unimolecular nucleophilic substitution (S_N1) and unimolecular elimination reactions (E1). Their detailed kinetic, stereochemical and product investigations suggest that formation of carbocations is the slow rate-determining step in those S_N1 and E1 reactions. Later, the carbocation concept was generalized to many other organic reactions and its significance in organic reaction mechanisms are well recognized. However, due to their electron deficient nature, carbocations are extremely reactive towards surrounding molecules e.g. solvent molecules or negative ions (nucleophile). The direct observation of a stable, long-lived carbocation was so much of a challenge that it did not become possible until the highly acidic superacid chemical system was developed. In 1962, George A. Olah used tertiary butyl fluoride to react with the superacid antimony pentafluoride (SbF_5) at $-78\text{ }^\circ\text{C}$ and obtained the first stable, long-lived tert-butyl cation. That tert-butyl cation can stay at $-78\text{ }^\circ\text{C}$ for many hours, long enough for both spectroscopic and chemical study.



Superacids are acids stronger than 100 % sulfuric acid. Superacid is the key to obtain stable, long-lived carbocations because their extremely low nucleophilicity suppresses the deprotonation of alkyl cations to olefins. If deprotonation took place, the carbocation (a strong acid) would immediately react with olefin (a good π base) and generate complex mixtures through a variety of reactions, such as alkylation, oligomerization, polymerization, and cyclization.



Since the remarkable role of superacid in stabilizing carbocations were revealed by George A. Olah, who later won the Nobel Prize in 1994 for his work on carbocations and superacids, thousands of carbocations have been investigated and characterized. The applied physical methods include conductivity, UV-vis spectroscopy, IR, laser Raman, X-ray diffraction, NMR (nuclear magnetic resonance), and ESCA (electron spectroscopy for chemical analysis). Because of the rehybridization from sp^3 to sp^2 and the effect of significant positive charge, the carbocation carbon nucleus is greatly de-shielded and

generally displays a large chemical shift. Therefore, ^{13}C NMR, which was used to characterize the first stable alkyl cation by George A. Olah in 1962, is still one of the most straightforward techniques to detect the carbocation in solution.

In an effort to verify the existence of DGM in GS, ^{13}C and ^1H - ^{13}C 2D NMR analyses were employed with GS- $[1-^{13}\text{C}]$ ADPGlc-HEPPSO complex. The GS- $[1-^{13}\text{C}]$ ADPGlc-HEPPSO complex displayed a peak at 177.98 ppm which was not seen in GS, $[1-^{13}\text{C}]$ ADPGlc, or HEPPSO (Figure III.52). Although the chemical shifts of most carbocation carbon center is in the range of 200 -320 ppm, the positive charge on DGM C1 is partially delocalized to the neighboring O4 which may render its chemical shift lower than regular carbocations. The 177.98 ppm peak was strong and actually was the only candidate for the DGM positively charged C1. In the $[1-^{13}\text{C}]$ ADPGlc titration experiment, the peak at a similar position, 177.74 ppm, rose as the collection time increased (Figure III.53). To prove our supposition, we synthesized uniformly ^{13}C -labeled ADPGlc to explicitly identify the source of the 177.98/177.74 ppm peak through tracing glucosyl unit carbons in a two-dimensional ^1H - ^{13}C spectrum. The NMR scanning range was purposely set between -10- 400 ppm, covering all possible peak regions. This choice, however, turns out not to be sensible, as even long collection time as 36 hours was not able to provide sufficient data and decent resolution. We were also not able to reproduce the 177.98 ppm peak with the second batch of $[1-^{13}\text{C}]$ ADPGlc. The second batch of $[1-^{13}\text{C}]$ ADPGlc was made following exactly the same procedure as the first batch $[1-^{13}\text{C}]$ ADPGlc, except that HEPES in the enzymatic synthesis and the Tris in the HPLC separation replaced bicine and ammonium formate, respectively. Bicine and ammonium formate exhibit peaks between 160-190 ppm whereas HEPES and Tris ^{13}C

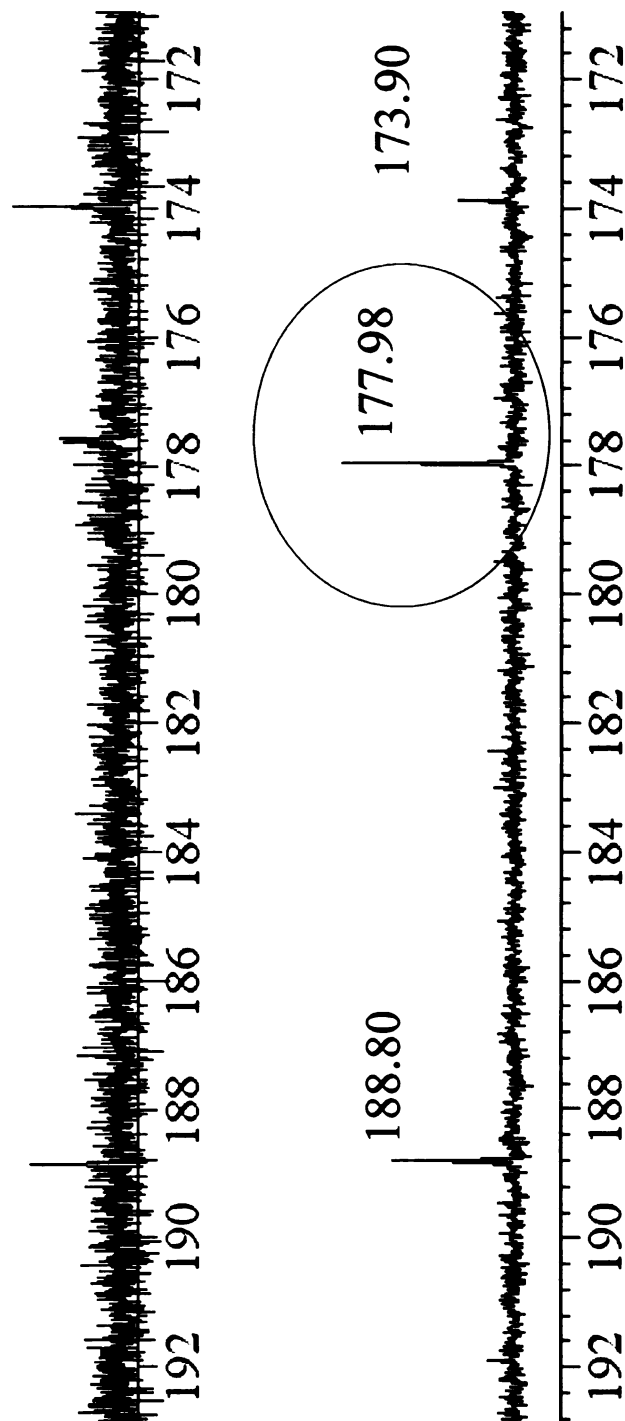
NMR peaks are all between 50 -60 ppm and won't interfere with the high chemical shift region where the DGM cation C1 peak may arise.

Although the NMR experiments we carried out did not offer evidence of long-lived DGM in the GS complex, the possibility to directly observe the long-lived carbocation DGM in the GS complex in the future still remains. Dramatically increasing GS solubility would allow more DGM in one NMR tube and using a more advanced NMR spectrometer and probe could be the solution.

Figure III.53 NMR spectra of $[1-^{13}\text{C}]\text{-ADPGlc-GS}$ complex and $[1-^{13}\text{C}]\text{-ADPGlc-GS-HEPPSO}$ complex

D-[1- ^{13}C]-ADPGlc+GS

3hr



D-[1- ^{13}C]-ADPGlc+GS+HEPPSO

10hr

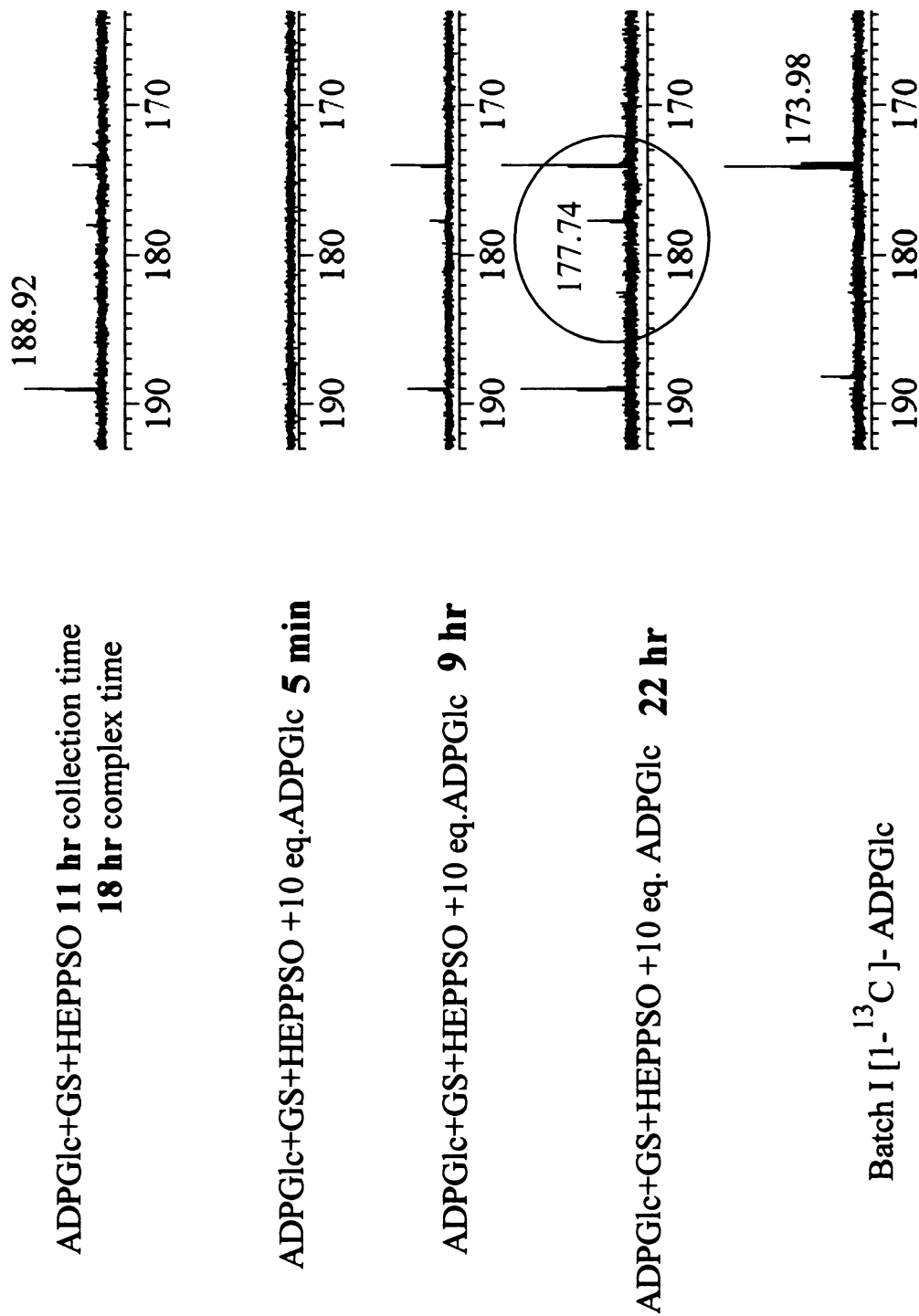


Figure III.54 NMR spectra of batch I [1-¹³C]-ADPGlc and its complex at different time points.

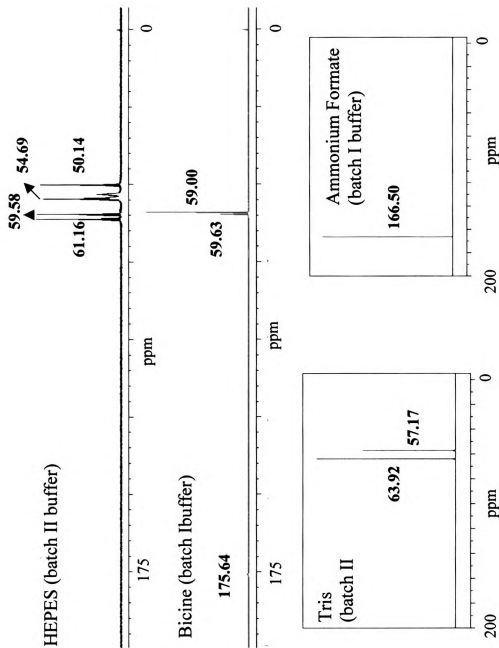


Figure III.55 NMR spectra of buffers in production of batch I and II [1- ^{13}C]-ADPGlc. The Tris and ammonium formate spectra are adapted from Spectral Database for organic compounds (SDBS).

III.5.3.D. The substrate phosphate group deprotonates the incoming acceptor nucleophile

The oligosaccharide-bound E377A structure shows that the β -phosphate O3B is 2.62 Å from the O4 of the glucan chain acceptor, ideally poised to deprotonate the incoming nucleophile 4-hydroxy group (Figure III.36). A similarly close distance was also spotted between O3B and the acceptor analogue HEPPSO hydroxyl-ethyl group O4 (2.76 Å) in the wtGSb structure (Figure III.26), also indicating that the phosphate group is involved in the deprotonation of the incoming acceptor nucleophile. This rationale explains why GS activity increases as the buffer pH goes up (from 6.8 to 7.5) because the high pH may promote the phosphate to deprotonate the acceptor hydroxyl group and facilitate the acceptor nucleophile attack (50). Similarly close distances between the incoming nucleophile hydroxyl group of the acceptor sugar and the phosphate oxygen of the leaving group is also seen in the GT-B retaining enzymes OtsA (39) and MalP(21).

In addition, the structures of GT-B retaining enzymes including *E.coli* GS all display a very similar overall disposition of the ADP/UDP and glucose in their active site. The phosphate moiety takes the unusual “tucked under” conformation relative to the ADP /UDP moiety which could provide some ground state destabilization as well as assisting acid catalysis of glycosyl transfer.

III.5.3.E. Evidence against the double-displacement

The double displacement mechanism was previously proposed for retaining GT enzymes with a direct reference to GH enzymes(58-61). It involves a nucleophilic attack from the β -face of the sugar donor and formation of a covalent glucosyl – enzyme intermediate (Figure I.12). Attempts to verify the double displacement mechanism for

GT-B enzymes, mainly to identify the catalytic nucleophile and to trap a covalent intermediate, have never been successful (62). Here, our wtGS complex structure presents some evidence against the double-displacement mechanism.

The catalytic nucleophile is an essential component of the double-displacement mechanism. For GS, Glu377 and His161 had long been proposed as potential candidates as the catalytic nucleophile in the double-displacement mechanism. However, in our wtGS complex structures Glu377 is located on the α -face of the glucose to be transferred and is relatively far both from the C1 glucose position and from the ADP -phosphate position, which is inconsistent with Glu377 playing a direct catalytic role in the reaction. Actually, our mutant E377A/ADP/HEPPSO structure reveals that the glucose moiety of ADPGlc is missing in the enzyme's active site, indicating that Glu377 is in fact critical for positioning the glucose properly in the active site (Figure III.52). Moreover, the Glu377 side chain two carboxyl oxygen are equally 2.93 Å to Lys305 and likely play a role in maintaining the positive charge on the catalytic residue Lys305.

The main chain of *E.coli* GS His161 and its structural equivalents in other retaining GT enzymes, such as His345 in maltodextrin phosphorylase (MalP), are located at the β -side of the sugar and close to the anomeric interaction center. Although amide groups are not ideal nucleophiles, such a function is not without precedence (63). To design an experiment to test whether it is the catalytic nucleophile is difficult because it is located in the main chain of the protein. Therefore, although the double replacement mechanism seems not plausible according to the available GT-B retaining enzyme structures, including our *E.coli* GS structures, and kinetic studies, it should not be totally ruled out.

III.5.4 The GS Catalytic Scenario Suggested by *E.coli* GS Structural Studies

Based on our structural studies of *E.coli* GS and its mutants, we propose the GS catalytic mechanism as follows. Firstly, GS adopts the open form with a wide-open interdomain cleft where the sugar donor ADPGlc and the sugar acceptor glucan chain binds to the C-terminal domain side and the N-terminal domain side, respectively. Then, the concurrence of sugar donor and acceptor stimulates the communication between the N- and C-domains and the enzyme closes. In the closed conformation of the enzyme, the basic residues Arg300 and Lys305 promote the phosphate-sugar bond break by protonating the ADP phosphate or stabilizing negative charge on the substrate ADP moiety. The His161 main chain (C=O) and the nearby substrate phosphate group stabilize the resulting DGM intermediate. The phosphate group also deprotonates the incoming acceptor nucleophile, which in turn attacks the partially positively charged DGM anomeric carbon C1 from the α -side. Finally, the substrate sugar is transferred to the acceptor and GS opens up to release the product ADP and prepare for the next cycle.

III.6 Mechanistic Implications for Starch Synthase

III.6.1 Plant Starch Synthases Share A Similar Active Site And Catalytic Mechanism With Bacterial GS

Like bacterial glycogen synthase, plant starch synthases also use ADPGlc to elongate α -1,4-linked glucan chains with retention of configuration. Both of them have a GT-B fold. Plant starch synthase (GBSS, SSII and SSIII) and bacterial GS share 29 -

34% sequence identity. We carried out a multiple sequence alignment on GSs, GBSS, SSII and SSIII from a variety of organisms and identified a number of conserved residues.

Almost all the residues in the *E.coli* GS active site have identical counterparts in SSs (Table III.7). They include the KTGGL motif, the glucose locator Glu377; the catalytic residues Lys305 and Arg300, DGM intermediate stabilizer His161, residue Tyr95, Asp137, and Trp138 that orient and bind to the glucan chain acceptor in the interdomain cleft. Moreover, Many of the buttressing residues of the active-site residues are also conserved in all SSs, such as Glu9 (interacting with KTGGLThr16 and Tyr95), and Gly400 (interacting with Glu377). Some of the conserved residues in SSs have been experimentally proven to participate in starch synthase catalysis, such as the KTGGL element Lys15(64), Asp21, Asp137 and Glu377(65). In addition, previous studies suggest that one arginine residue is involved in the ADP reaction as the arginine-specific reagent phenylglyoxal inhibited SSIIa but the repression can be recovered with a higher concentration of ADPGlc(66). Our sequence alignment points out that the arginine residue is probably the *E.coli* GS Arg300 equivalent: maize SSIIa Arg551. Taken together, GT-B retaining bacterial GS and plant starch synthase share a closely similar active site and are very likely to employ a similar mechanism to catalyze glucan chain elongation.

Table III.7 Selective conserved residues in bacterial GS and plant starch synthases.

Those residues critical to GS activity and their equivalents in SSs are colored pink.

Multiple sequence alignment was conducted with DNASTAR. GS sequences used were:

Escherichia. coli GS (P0A6U8), *Agrobacterium tumefaciens* GS (AAD03474), *Pyrococcus abyssi* GS (NP_125769), Sequences of granule-bound starch synthases from various organisms were those of barley (AAL77109), maize (P04713), potato (CAA41359), and rice (P19395), Other starch synthases used were those of maize SSI (AAB99957), potato SSI (P93568), wheat SSI (Q43654), wheat SSIIa (BAE48798), maize SSIIa (AAS77569), potato SSII (CAA61241), potato SSIII (Q43846), wheat SSIII (AAF87999) and *Chlamydomonas reinhardtii* SS (AAC17971).

Table III.7

9E	15	KT	GGL	ADV	95	Y	137	DW	H	161	H	N	246	N	G	<i>E.coli</i> GS
9E	15	KT	GGL	ADV	96	Y	138	DW	Q	162	H	N	246	N	G	AtGS
9E	14	KV	GGL	ADV	89	Y	128	DW	H	151	H	R	217	N	G	PaGS
84E	90	KT	GGL	GDV	176	Y	227	DW	H	257	H	N	346	N	G	barley GBSS
85E	91	KT	GGL	GDV	177	Y	228	DW	H	258	H	N	347	N	G	maize GBSS
89E	95	KT	GGL	GDV	181	Y	232	DW	H	262	H	N	351	N	G	potato GBSSI
91E	97	KT	GGL	GDV	183	Y	234	DW	H	264	H	N	353	N	G	rice GBSS
140E	146	KT	GGL	GDV	232	Y	277	DW	H	307	H	N	401	N	G	maize SSI
139E	145	KT	GGL	GDV	232	Y	277	DW	H	307	H	N	401	N	G	potato SSI
147E	153	KT	GGL	GDV	239	Y	284	DW	H	314	H	N	408	N	G	wheat SSI
316E	322	KT	GGL	GDV	403	Y	446	DW	H	476	H	N	562	N	G	wheat SSIIa
249E	255	KT	GGL	GDV	333	Y	376	DW	H	406	H	N	492	N	G	maize SSIIa
284E	290	KT	GGL	GDV	371	Y	414	DW	H	444	H	N	530	N	G	potato SSII
788E	794	KT	GGL	GDV	875	Y	909	DW	S	938	H	N	986	N	G	potato SSIII
1188E	1194	KT	GGL	GDV	1275	Y	1308	DW	S	1337	H	N	1385	N	G	wheat SSIII
156E	162	KT	GGL	GDV	259	Y	294	DW	H	323	H	N	416	N	G	<i>Chlamydomonas reinhardtii</i> SS

300	R	L	304Q	K	G	355	Y	373	PSRF	E	PCGLTQL	398	TG	G	L	<i>E.coli</i> GS
300	R	L	304Q	K	G	354	Y	372	PSRF	E	PCGLTQL	397	TG	G	L	AtGS
257	R	F	262Q	K	G	317	L	335	PSYF	E	PFGLTQL	360	VG	G	L	PaGS
401	R	L	405Q	K	G	457	F	475	TSRF	E	PCGL IQL	500	TG	G	L	barley GBSS
402	R	L	406Q	K	G	459	F	477	TSRF	E	PCGL IQL	502	TG	G	L	maize GBSS
406	R	L	410Q	K	G	461	F	480	PSRF	E	PCGL IQL	505	TG	G	L	potato GBSSI
408	R	L	412Q	K	G	463	F	481	PSRF	E	PCGL IQL	506	TG	G	L	rice GBSS
455	R	L	459Q	K	G	510	F	528	PSRF	E	PCGLNQL	553	TG	G	L	maize SSI
455	R	L	459Q	K	G	510	F	528	PSRF	E	PCGLNQL	553	TG	G	L	potato SSI
462	R	L	466Q	K	G	517	F	535	PSRF	E	PCGLNQL	561	TG	G	L	wheat SSI
621	R	L	625Q	K	G	676	F	694	PSRF	E	PCGLNQL	719	VG	G	L	wheat SSIIa
551	R	L	555Q	K	G	606	F	624	PSRF	E	PCGLNQL	649	VG	G	L	maize SSIIa
589	R	L	593Q	K	G	644	F	662	PSRF	E	PCGLNQL	687	VG	G	L	potato SSII
1040	R	L	1044Q	K	G	1100	Y	1118	PS IF	E	PCGLTQL	1143	TG	G	L	potato SSIII
1439	R	L	1443Q	K	G	1499	Y	1517	PS IF	E	PCGLTQL	1542	TG	G	L	wheat SSIII
470	R	L	474Q	K	G	526	Y	544	PSMF	E	PCGLTQL	569	TG	G	L	<i>Chlamydomonas reinhardtii</i> SS

III.6.2. Glucan Chain Binding In Starch Synthases

Multiple sequence alignment reveals that the glucan chain binding residues in the interdomain catalytic cleft are strictly conserved in SS, such as Tyr95, Asp137, and

Phe138. None of the surface binding residues especially in the Gb binding site is conserved in SSs. However five out of eight binding residues at the Gc site used their backbone amide group to lock the oligosaccharide. The binding residue, Lys199, was not conserved throughout the SSs, but was conserved in one SS subfamily, GBSS. Taken together, we believe that the glucan chain may not bind to SSs on the surface in exactly the same way as to GS.

III.6.3. Insights into GBSS and SS Specificities

Granule-bound starch synthase (GBSS) and soluble starch synthase SS (SSI, SSII and SSIII) are two major isoforms of starch synthase. GBSS preferably catalyzes transglycosylation reactions with long, linear amylose as the product, whereas the homologous soluble SS (SSI, SSII and SSIII) produces short, more branched amylopectin(67,68). One interesting feature of GBSS is that it can be activated by amylopectin but soluble SS cannot (69). In addition, GBSS and soluble SS exhibit quite a few different properties, such as affinities for ADPGlc and glucan substrate, thermosensitivity and the processivity of glucan chain extension.

In an effort to find out the determinants of GBSS and SS specificities, a series of potato GBSS and SSII chimeric and truncated proteins were tested(70). It turns out that the C-terminal region of GBSS confers most of the specific properties of GBSS. Based on our GS structure, this C-terminal region ($\alpha 13$ - $\alpha 18$) spans the two domains of the enzyme and defines a surface of the enzyme that is quite far from the active site (Figure III.56). As we mentioned before, SS and GS share a similar domain – spanning active site. This C-terminal region may bind to amylopectin and promote reorganization of two

domain relative position, in turn affecting catalysis. This assumption is now under examination in Dr. Geiger's lab.

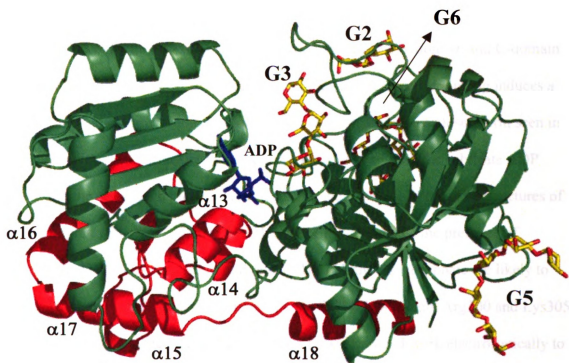


Figure III.56 Modeled potato GBSS structure based on *E.coli* GS structure. The region ranging from helix $\alpha 13$ to $\alpha 18$ is colored red which is the determinant of most specific properties of GBSS and may play a role in tuning the relative orientation of the two domains.

III.7. Conclusions

We have determined a series of *E.coli* GS and GS complex structures and they provide a comprehensive understanding of the structure and functional relationship of bacterial GS. All ligand-bound *E.coli* GS structures exhibit virtually identical molecular organization and a deeply buried active site in the narrow interdomain cleft. In contrast, the apo-dmGS exhibits a wide-open interdomain cleft and the N-domain and C-domain are spread relatively far apart from each other, indicating that ligand binding induces a considerably large domain-domain closure. The catalytically active closed form seen in the ligand – bound GS structure is due to the concurrence of both the substrate ADP moiety and the oligosaccharide acceptor /acceptor analogue HEPPSO. The structures of GS in complex with ADP, glucose /DGM, and HEPPSO indicate the presence of intermediate DGM in GS catalysis and reveals His161 and the phosphate are likely to participate in the stabilization of the intermediate DGM. In addition, Arg300 and Lys305 are close to the phosphate and probably act as a Lewis acid and work electrostatically to make the phosphate a better leaving group. The ADPGlc-incubated E377A crystal misses the glucose moiety of substrate in the active site, which evidently suggests that Glu377 plays a critical role in fixating the glucose to be transferred. In addition, Glu377 may help the catalytic residue Lys305 maintain its negative charge through two close hydrogen bonds. Our finding of Glu377's dual role in GS catalysis perfectly explains the observed dramatic decrease in enzyme activity upon mutation on that position. In our oligosaccharide-bound E377A and HEPPSO-bound wtGS structures, maltotriose and HEPPSO occupy a similar position in the interdomain cleft and Asp137 plays an important role in positioning these acceptor nucleophile analogues. Information gained

from our GS structures is in agreement with previous biochemical and mutagenesis studies. Taken together, our *E.coli* GS structural studies allows us to confidently propose an S_N1 mechanism over the double displacement mechanism for GS and other GT-B retaining enzymes such as SSs, MalP, OtsA and GP, which all share a virtually identical active site. In addition, the oligosaccharide-bound GS structure suggests that a long glucan chain binds only to the N-terminal domain allowing frequent opening and closing of the domains without dissociation of glucan on each cycle.

Finally, based on our *E.coli* GS structures and the previously reported chimeric studies of GBSS and SSII, we speculate that the region (α 13 - α 18) in starch synthase is involved in positioning two domains and that interaction of this region of GBSSI with amylopectin possibly orients the domains for catalysis.

References

1. Edwards, A., Borthakur, A., Bornemann, S., Venail, J., Denyer, K., Waite, D., Fulton, D., Smith, A.M. and Martin, C. (1999) Specificity of starch synthase isoforms from potato *European journal of biochemistry / FEBS*, 266, 724-736
2. Notredame, C., Higgins, D. and Heringa, J. (2000) T-coffee: A novel method for multiple sequence alignments *J. Mol. Biol.*, 302, 205-217
3. Drenth, J. (1994) Principles of protein X-ray crystallography (Springer, Ed.), New York
4. Vagin, A. and Teplyakov, A. (1997) Molrep: An automated program for molecular replacement *Journal of Applied Crystallography*, 30, 1022-1025
5. Jones, T.A. (1978) A graphics model building and refinement system for macromolecules. Frodo: A graphics fitting program for macromolecules. *Journal of Applied Crystallography*, 11, 268-272
6. Murshudov, G.N., A.Vagin, A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method *Acta Crystallogr., Sect. D: Biol. Crystallogr.* , 53, 240-255
7. Brünger, A.T. (1992) Free R value: A novel statistical quantity for assessing the accuracy of crystal structures *Nature*, 355, 472-475

8. Kleywegt, G.J. and Brunger, A.T. (1996) Checking your imagination: Applications of the free R value. *Structure*, 4, 897-904
9. Ramakrishnan, C. and Ramachandran, G.N. (1965) Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophysical Journal*, 5, 909-933
10. Kleywegt, G.J. and Jones, T.A. (1996) Phi/ psi-chology: Ramachandran revisited *Structure*, 15, 1395–1400
11. Yep, A., Ballicora, M.A., Sivak, M.N. and Preiss, J. (2004) Identification and characterization of a critical region in the glycogen synthase from *Escherichia coli* *J. Biol. Chem.*, 279, 8359-8367
12. Bergfors, T.M. (1999) Protein crystallization techniques, strategies, and tips, *International University Line*, La Jolla, CA
13. Otwinowski, Z.M.W. (1997) Scalepack *Methods Enzymol*, 276, 307-326
14. Yep, A., Bejar, C.M., Ballicora, M.A., Dubay, J.R., Iglesias, A.A. and Preiss, J. (2004) An assay for adenosine 5'-diphosphate (ADP)-glucose cpyrophosphorylase that measures the synthesis of radioactive ADP-glucose with glycogen synthase *Analytical Biochemistry*, 324, 52-59
15. Morell, S.A. and Bock, R.M. (1954) Ultraviolet absorption spectra of 5'-ribonucleotides in *American Chemical Society 126th meeting* (Chemistry). 44, New York
16. SDBSweb <http://riodb01.lbase.aist.go.jp/sdbs/> (national institute of advanced industrial science and technology, date of access)

17. Collaborative Computational Project, N. (1994) The CCP4 suite: Programs for protein crystallography *Acta Crystallogr., Sect. D: Biol. Crystallogr.* , 50, 760-763
18. McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, 16, 404-405
19. Cristina, H., Guinovart, J.J., Fita, I. and Ferrer, J.C. (2006) Crystal structure of an archaeal glycogen synthase: Insights into oligomerization and substrate binding of eukaryotic glycogen synthases *J. Biol. Chem.*, 281, 2923-2931
20. Buschiazzo, A., Ugalde, J.E., Guerin, M.E., Shepard, W., Ugalde, R.A. and Alzari, P.M. (2004) Crystal structure of glycogen synthase: Homologous enzymes catalyze glycogen synthesis and degradation *EMBO J.*, 23, 3196-3205
21. Watson, K.A., McCleverty, C., Geremia, S., Cottaz, S., Driguez, H. and Johnson, L.N. (1999) Phosphorylase recognition and phosphorolysis of its oligosaccharide substrate: Answer to a long outstanding question *EMBO J.*, 18, 4619-4632
22. Yep, A., Ballicora, M.A. and Preiss, J. (2004) The active site of the *Escherichia coli* glycogen synthase is similar to the active site of retaining GT-b glycosyltransferases *Biochem. Biophys. Res. Commun.*, 316, 960-966
23. Geremia, S., Campagnolo, M., Schinzel, R. and Johnson, L.N. (2002) Enzymatic catalysis in crystals of *Escherichia coli* maltodextrin phosphorylase *J. Mol. Biol.*, 322, 413-423
24. Rao, V.V.R., Qasba, P.K., Balaji, P.V. and Chandrasekaran, R. (1998) Conformation of carbohydrates, *Harwood Academic*, Australia

25. Appell, M., Strati, G., Willett, J.L. and Momany, F.A. (2004) B3LYP/6-311++G** study of alpha- and beta-D-glucopyranose and 1,5-anhydro-d-glucitol: 4C_1 and 1C_4 chairs, 3OB and B3O boats, and skew-boat conformations. *Carbohydrate Research*, 339, 537-551

26. Schwarz, A., Pierfederici, F.M. and Nidetzky, B. (2005) Catalytic mechanism of α -retaining glucosyl transfer by corynebacterium callunae starch phosphorylase: The role of histidine-334 examined through kinetic characterization of site-directed mutants *Biochem. J.* , 387, 437-445

27. Ryuta, K., Keiko, H., Toshihiko, A., Kunio, Y. and Kazuaki, H. (2006) Role of trp140 at subsite -6 on the maltohexaose production of maltohexaose-producing amylase from alkalophilic Bacillus sp.707 *Protein Sci.* , 15, 468-477

28. Alzari, P.M., Souchon, H.N. and Dominguez, R. (1996) The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from clostridium thermocellum *Structure*, 4, 265-275

29. Dutzler, R., Wang, Y.-F., Rizkallah, P.J., Rosenbusch, J.P. and Schirmer, T. (1996) Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway *Structure*, 4, 127-134

30. Claude, J.-B., Suhre, K., Notredame, C., Claverie, J.-M. and Abergel, C. (2004) Caspr: A web server for automated molecular replacement using homology modeling *Nucleic Acids Res.* , 32(Web Server), W606-W609

31. Sundaralingam, M. (1968) Some aspects of stereochemistry and hydrogen bonding of carbohydrates related to polysaccharide conformations. *Biopolymers*, 6, 189-213

32. Goldsmith, E., Sprang, S. and Fletterick, R. (1982) Structure of maltoheptaose by difference fourier methods and a model for glycogen. *J. Mol. Biol.*, 156, 411-427
33. Dowda, M.K., Zenga, J., Frenchb, A.D. and Reillya, P.J. (1992) Conformational analysis of the anomeric forms of kojibiose, nigerose, and maltose using MM3 *Carbohydrate Research*, 230, 223-244
34. Preiss, J. (1984) Bacterial glycogen synthase and its regulation *Annu. Rev. Microbiol.*, 38, 419-458
35. Qi, G., Lee, R. and Hayward, S. (2005) A comprehensive and non-redundant database of protein domain movements *Bioinformatics*, 21, 2832-2838
36. Furukawa, K., Tagaya, M., Inouye, M., Preiss, J. and Fukui, T. (1990) Identification of lysine 15 at the active site in *Escherichia coli* glycogen synthase, conservation of a Lys-X-Gly-Gly sequence in the bacterial and mammalian enzymes *J. Biol. Chem.*, 265, 2086-2090
37. Furukawa, K., Tagaya, M., Tanizawa, K. and Fukui, T. (1993) Role of the conserved Lys-X-Gly-Gly sequence at the ADP-glucose-binding site in *Escherichia coli* glycogen synthase *J. Biol. Chem.*, 268, 23837-23842
38. Johnson, L.N., Acharya, K.R., Jordan, M.D. and Mclaughlin, P.J. (1990) Refined crystal structure of the phosphorylase-heptulose 2-phosphate-oligosaccharide-AMP complex. *J. Mol. Biol.*, 211, 645-661
39. Gibson, R.P., Turkenburg, J.P., Charnock, S.J., Lloyd, R. and Davies, G.J. (2002) Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA *Chemistry & Biology*, 9, 1337-1346

40. Gibson, R.P., Tarling, C.A., Roberts, S., Withers, S.G. and Davies, G.J. (2004) The donor subsite of trehalose-6-phosphate synthase: Binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2a resolution *J. Biol. Chem.*, 279, 1950-1955
41. Patenaude, S.I., Seto, N.O.L., Borisova, S.N., Szpacenko, A., Marcus, S.L., Palcic, M.M. and Evans, S.V. (2002) The structural basis for specificity in human ABO(h) blood group biosynthesis *Nature Structural Biology*, 9, 685 - 690
42. Davies, G., Withers, S.G. and Sinnott, M.L. (1997) in *Comprehensive biological catalysis*. 119-209, *Academic Press, London*
43. Barford, D., Hu, S.H. and Johnson, L.N. (1991) Structural mechanism for glycogen phosphorylase control by phosphorylation and amp *J. Mol. Biol.*, 218, 233-260
44. Lariviere, L., Sommer, N. and Morera, S. (2005) Structural evidence of a passive base-flipping mechanism for AGT, an unusual GT-B glycosyltransferase *J. Mol. Biol.*, 352, 139-150
45. Schinzel, R. and Palm, D. (1990) *Escherichia coli* maltodextrin phosphorylase: Contribution of active site residues glutamate-637 and tyrosine-538 to the phosphorolytic cleavage of alpha-glucans *Biochemistry*, 29, 9956-9962
46. Cid, E., Gomis, R.R., Geremia, R.A., Guinovart, J. and Ferrer, J.C. (2000) Identification of two essential glutamic acid residues in glycogen synthase *J. Biol. Chem.*, 275, 33614-33621
47. Sinnott, M.L. (1990) Catalytic mechanisms of enzymatic glycosyl transfer *Chem. Rev. (Washington, DC, U. S.)*, 90, 1171-1202

48. Boix, E., Yingnan, Z., Swaminathan, G.J., Brew, K. and Acharya, K.R. (2002) Structure basis of ordered binding of donor and acceptor substrate to the retaining glycosyltransferase, α -1,3-galactosyltransferase *J. Biol. Chem.*, 277, 28310-28318
49. Sprang, S., Goldsmith, E. and Fletterick, R. (1987) Structure of the nucleotide activation switch in glycogen phosphorylase a. *Science*, 237, 1012-1019
50. Fox, J., Kawaguchi, K., Greenberg, E. and Preiss, J. (1976) Biosynthesis of bacterial glycogen. Purification and properties of the *Escherichia coli* b ADPglucose:1,4- α -D-glucan 4- α -glucosyltransferase *Biochemistry*, 15, 849-847
51. Goedl, C., Griessler, R., Schwarz, A. and Nidetzky, B. (2006) Structure–function relationships for schizophyllum commune trehalose phosphorylase and their implications for the catalytic mechanism of family GT-4 glycosyltransferases *Biochem. J.* , 397, 491–500
52. Nidetzky, B. and Eis, C. (2001) A-retaining glucosyl transfer catalysed by trehalose phosphorylase from *schizophyllum commune*: Mechanistic evidence obtained from steady-state kinetic studies with substrate analogues and inhibitors *Biochem. J.* , 360, 727-736
53. Eis, C., Watkins, M., Prohaska, T. and Nidetzky, B. (2001) Fungal trehalose phosphorylase: Kinetic mechanism, pH-dependence of the reaction and some structural properties of the enzyme from schizophyllum commune *Biochem. J.* , 356, 757–767
54. Papageorgiou, A.C., Oikonomakos, N.G., Leonidas, D.D., Bernet, B., Beer, D. and Vasella, A. (1991) The binding of D-gluconohydroximo-1, 5-lactone to glycogen phosphorylase. Kinetic, ultracentrifugation and crystallographic studies. *Biochem. J.* , 274

55. Mitchell, E.P., Withers, S.G., Ermert, P., Vasella, A.T., Garman, E.F., Oikonomakos, N.G. and Johnson, L.N. (1996) Ternary complex crystal structures of glycogen phosphorylase with the transition state analog nojirimycin tetrazole and phosphate in the T and R states. *Biochemistry*, 35, 7341-7355
56. Banait, N.S. and Jencks, W.P. (1991) Reaction of anionic nucleophiles with α -D-glucopyranosyl fluoride in aqueous solution through a concerted, $A_nD_n(S_n2)$ mechanism *J. Am. Chem. Soc.*, 113, 7951-7958
57. Banait, N.S. and Jencks, W.P. (1991) General-acid and general -base catalysis of the cleavage of α -D-glucopyranosyl fluoride *J. Am. Chem. Soc.*, 113, 7958-7963
58. Bowles, D., Isayenkova, J., Lim, E.-K. and Poppenberger, B. (2005) Glycosyltransferases: Managers of small molecules *Curr. Opin. Plant Biol.*, 8, 254-263
59. Breton, C., Najdrova, L., Jeanneau, C., Koa, J. and Imberty, A. (2006) Structures and mechanisms of glycosyltransferases *Glycobiology*, 16, 29-37
60. Davies, G.J. (2001) Sweet secrets of synthesis *Nature Structural Biology*, 8, 98-100
61. Ünligil, U.M. and Rini, J.M. (2000) Glycosyltransferase structure and mechanism *Curr. Opin. Struct. Biol.*, 10, 510-517
62. Lairson, L.L. and Withers, S.G. (2004) Mechanistic analogies amongst carbohydrate modifying enzymes *Chem. Commun. (Cambridge, U. K.)*, 20, 2243-2248

63. Persson, K., Ly, H.D., Dieckelmann, M., Wakarchuk, W.W., Withers, S.G. and Strynadka, N.C.J. (2001) Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs *Nature Structural Biology*, 8, 166-175
64. Gao, Z., Keeling, P., Shibles, R. and Guan, H. (2004) Involvement of lysine-193 of the conserved "K-T-G-G" motif in the catalysis of maize starch synthase IIa *Arch. Biochem. Biophys.* , 427, 1-7
65. Nichols, D.J., Keeling, P.L., Spalding, M. and Guan, H. (2000) Involvement of conserved aspartate and glutamate residues in the catalysis and substrate binding of maize starch synthase. *Biochemistry*, 39, 7820-7825
66. Imparl-Radosevich, J.M., Keeling, P.L. and Guan, H. (1999) Essential arginine residues in maize starch synthase IIa are involved in both ADP-glucose and primer binding *FEBS Lett.*, 457, 357-362
67. Manners, D.J. (1989) Recent developments in our understanding of amylopectin structure *Carbohydrate Polymers*, 11, 87-112
68. Slattery, J.C., Kavakli, I.H. and Okita, W.T. (2000) Engineering starch for increased quantity and quality *Trends Plant Sci.* , 291-298
69. Denyer, K., Waite, D., Edwards, A., Martin, C. and Smith, A.M. (1999) Interaction with amylopectin influences the ability of granule-bound starch synthase I to elongate malto-oligosaccharides *Biochem. J.* , 342, 647-653
70. Edwards, A., Borthakur, A., Bornemann, S., Venail, J., Denyer, K., Waite, D., Fulton, D., Smith, A.M. and Martin, C. (1999) Specificity of starch synthase isoforms from potato *European Journal of Biochemistry*, 266, 724-736

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02956 9211