Insights into the Structure and Mechanism of Anhydromuramic Acid Kinase (AnmK): A Novel Peptidoglycan Recycling Enzyme with Dual Hydrolase and Kinase Functionality

by

Catherine Leigh Allen

Department of Chemistry
Duke University

Date: ______________________

Approved:

___________________________
Professor Dewey G. McCafferty, Supervisor

___________________________
Professor Eric J. Toone

___________________________
Professor Jiyong Hong

___________________________
Professor Pei Zhou

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the Graduate School of Duke University

2011
ABSTRACT

Insights into the Structure and Mechanism of Anhydromuramic Acid Kinase (AnmK): A Novel Peptidoglycan Recycling Enzyme with Dual Hydrolase and Kinase Functionality

by

Catherine Leigh Allen

Department of Chemistry
Duke University

Date: ________________
Approved:

___________________________
Professor Dewey G. McCafferty, Supervisor

___________________________
Professor Eric J. Toone

___________________________
Professor Jiyong Hong

___________________________
Professor Pei Zhou

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctorate of Philosophy in the Department of Chemistry in the Graduate School of Duke University

2011
Abstract

Bacteria recycle pre-existing peptidoglycan in order to minimize the de novo synthesis of peptidoglycan precursors. The recycling pathway is under study for its chemotherapeutic target potential. Anhydromuramidic acid kinase (AnmK) is part of this recycling pathway and catalyzes the dual hydrolysis/phosphorylation of anhMurNAc to yield MurNAc-6-P. This enzyme has been previously discovered, but only minimally characterized. Therefore, the overarching goal of this work was to clone, express and purify AnmK to homogeneity; perform further kinetic characterization; solve the open, closed and transition state mimic-bound conformations of AnmK by x-ray crystallography; and develop a mechanism based on the accumulated research findings and 18O-labelling studies.

The anmK gene was successfully cloned as a hexahistidine fusion protein and overexpression was optimized to achieve homogeneity in three chromatographic steps. Additionally, the synthesis of both anhMurNAc and a pseudosubstrate (anhGlcNAc) were carried out to allow for both kinetic characterization and structural studies. To this end, the structure of de novo AnmK was solved using SAD and high-resolution (1.9 Å) data. Also, an ATP analog (ANP) and anhMurNAc substrate-bound, closed conformation structure (1.95 Å) was solved. These structures elucidated an 11° domain
closure of the enzyme upon substrate binding and also revealed the active site geometry to be used to determine potential molecular determinants of specificity.

Insights into the kinetic mechanism of AnmK were then gathered using multiple techniques. First, the structure of AnmK was solved with a known transition state analog, the MgADP-vanadate complex to 2.5 Å resolution. Following this structure, which sheds light on the potential importance of a residue (Glu330) other than the catalytic base (Asp187), isotopic labeling was performed with H$_2^{18}$O. Using NMR and MS, the regiochemical selectivity of AnmK hydrolysis to impart the solvent-derived oxygen at C1 was established. Additionally, this was carried out with stereochemical preference to initially create the α-anomer of the carbohydrate product. This regiochemistry and stereospecificity drove the design of our proposed concomitant hydrolysis/phosphorylation mechanism.

This research can be applied to the immediate goal of understanding the function of a single, novel enzyme with unique chemistry. The clarification of the AnmK mechanism will facilitate future investigation into enzymes with dual hydrolase/kinase functionality. Furthermore, this research contributes to the understanding of the complex bacterial peptidoglycan layer in order to harness new ideas for developing antibiotics.
Dedication

This thesis is dedicated to my family and to the memory of my grandfather,
Joseph Lindsay Albright, Sr. (1927 –2009).

“There is no doubt that it is around the family and the home that all the greatest virtues, the most dominating virtues of human society, are created, strengthened and maintained” – Winston Churchill
Contents

Abstract ................................................................................................................................. iv
List of Abbreviations ........................................................................................................... xv
List of Tables ......................................................................................................................... xxii
List of Figures ....................................................................................................................... xxiii
List of Schemes .................................................................................................................... xxvii
List of Mechanisms ........................................................................................................... xxviii
Acknowledgements ........................................................................................................... xxix
Chapter 1: General Introduction ......................................................................................... 1
  1.1 Bacterial Infection and Antibiotic Resistance .............................................................. 1
  1.2 Targets for Therapeutic Action .................................................................................... 2
  1.3 Components and Organization of the Bacterial Cell Wall ......................................... 3
    1.3.1 Peptidoglycan structure and linkage ................................................................... 7
    1.3.2 The peptidoglycan sacculus ............................................................................... 10
  1.4 Cell Wall Metabolism .................................................................................................. 10
    1.4.1 Peptidoglycan biosynthesis .............................................................................. 10
    1.4.2 Peptidoglycan turnover .................................................................................... 13
  1.5 Cell Wall Recycling .................................................................................................... 15
  1.6 Cell Wall Recycling Enzymes and Anhydromuramic Acid Kinase (AnmK) ............. 19
  1.7 Examination of AnmK Literature .............................................................................. 25
  1.8 Thesis Goals ............................................................................................................... 27
Chapter 2: Protein Biochemistry of AnmK .................................................................29

2.1 Introduction .................................................................................................................29

2.2 Methods for the Cloning, Expression and Purification of E. coli anmK .................30

2.2.1 Methods for cloning E. coli annK ........................................................................30

2.2.2 Methods for overexpression of AnmK ..................................................................33

2.2.2.1 IPTG-induced expression of AnmK .................................................................33

2.2.2.2 Se-Met enriched autoinduction expression .........................................................34

2.2.2.3 Autoinduction for expression of E. coli AnmK ..................................................35

2.2.3 Methods for purification of E. coli AnmK .................................................................37

2.2.3.1 General cell lysis, ultracentrifugation, ammonium sulfate precipitation, protease inhibitor, SDS-PAGE and desalting procedures ...........................................37

2.2.3.3 Glutathione-S-transferase (GST) column chromatography .................................38

2.2.3.4 Nickel immobilized metal affinity chromatography (Ni-IMAC) ............................39

2.2.3.5 Ion exchange chromatography (IEX) ..................................................................39

2.2.3.6 Phenyl Sepharose™ .........................................................................................40

2.2.3.7 Hydroxyapatite (HA) hydrophobic interaction column .......................................40

2.2.3.8 ATP-Sepharose ..................................................................................................40

2.2.3.9 S-200 Size exclusion chromatography ...............................................................41

2.2.3.10 TEV protease cleavage .....................................................................................41

2.2.3.11 Purification of AnmK following TEV cleavage ...................................................42

2.2.4 The final, optimized E. coli AnmK purification scheme ...........................................42

2.2.5 Methods for size exclusion chromatography, analytical ultracentrifugation, MALDI/MS and LC-ESI-MS of E. coli AnmK .................................................................43
2.2.5.1 Sephadex™ G-75 ................................................................. 43
2.2.5.2 Sephacryl S-100 HR ........................................................... 44
2.2.5.3 High resolution S-200 column ............................................. 44
2.2.5.4 Sedimentation equilibrium analytical ultracentrifugation (AUC) ........ 44
2.2.5.5 Calculations to discern molecular weight from AUC data .............. 45
2.2.5.6 Matrix-assisted laser desorption/ionization mass spectrometry (MALDI/MS) ........................................................................ 46
2.2.5.7 Liquid chromatography electrospray ionization mass spectrometry (LC-ESI-MS) ................................................................. 46
2.2.6 Methods of the determination of AnmK kinetic measurements ............ 47
2.2.6.1 Design of the AnmK activity assay ........................................ 47
2.2.6.2 AnmK activity assay with anhGlc (levoglucosan) .......................... 48
2.2.6.3 AnmK activity assay anhGlcNAc, GlcNAc and anhMurNAc ............. 48
2.2.6.4 Kᵢ and IC₅₀ Determination for ADP and ANP ............................. 49
2.2.6.5 pH optimum of AnmK ............................................................. 49
2.2.6.6 Kᵢ determination for ATP and the analog, ANP .......................... 49
2.2.6.7 Metal ion dependence ................................................................ 49
2.3 Results and Discussion .................................................................. 50
2.3.1 E. coli anmK cloning ................................................................. 50
2.3.2 AnmK expression ....................................................................... 54
2.3.3 AnmK purification ...................................................................... 59
2.3.4 Biophysical characterization ......................................................... 69
2.3.5 AnmK preliminary kinetics ......................................................... 71
Chapter 5: Structure of anhMurNAc-ANP-AnmK Ternary Complex .........................138

5.1 Introduction ........................................................................................................138

5.2 Experimental Procedures ..................................................................................139

5.2.1 Crystallization trials and substrate binding ..................................................139

5.2.2 Data collection, refinement, validation and deposition ....................................140

5.3 Results and Discussion for the Substrate-Bound Structure ................................142

5.3.1 Attaining a different crystal morphology .......................................................142

5.3.2 Phasing and refinement .................................................................................145

5.3.3 General structural notes ................................................................................148

5.3.4 Motifs within the structure ............................................................................151

5.3.5 Substrate binding sites ....................................................................................155

5.3.5.1 AnhMurNAc binding site .......................................................................155

5.3.5.2 ANP binding site .....................................................................................157

5.3.6 An interfacial zinc binding site .....................................................................161

5.4 Conclusions ........................................................................................................163

Chapter 6: Structure of MgADP•VO₃-bound AnmK..............................................169

6.1 Introduction .........................................................................................................169

6.2. Methods for MgADP•VO₃-bound AnmK Structure ........................................170

6.3 Results and Discussion of Mechanistic Insights ..............................................172

6.3.1 The structure of ADP•VO₃-bound AnmK .......................................................172

6.3.2 The ADP•VO₃ binding site in ADP•VO₃-bound AnmK .................................177

6.3.3 The anhMurNAc binding site in ADP•VO₃-bound AnmK ............................180
6.4 Conclusions .................................................................................................................. 182

Chapter 7: Probing the Mechanism of AnmK ................................................................. 186

7.1 Introduction .................................................................................................................. 186
   Probing the origin of the C1 oxygen via substitution with H\textsubscript{2}\textsuperscript{18}O .......... 186

7.2 Experimental Procedures .......................................................................................... 190
   7.2.1 Enzymatic synthesis and \textsuperscript{13}C NMR analysis of \textsuperscript{18}O-labeled GlcNAc-6-P .... 190
   7.2.2 Determination of \textsuperscript{18}O incorporation by mass spectrometry ......................... 190

7.3 Results and Discussion of Mechanistic Insights ....................................................... 191
   7.3.1 Mass spectrometry of \textsuperscript{18}O-labeled GlcNAc-6-P ........................................... 191
   7.3.2 Determination of \textsuperscript{18}O incorporation by \textsuperscript{13}C NMR ............................... 193

7.4 Conclusions ................................................................................................................. 196
   7.4.1 Comparison to known enzyme mechanisms ......................................................... 196
   7.4.2 Plausible Mechanism for AnmK Kinase/Hydrolase Activity ................................. 199

Chapter 8: Conclusions and Future Directions ............................................................... 205

8.1 General Conclusions .................................................................................................. 205

8.2 Future Directions ....................................................................................................... 208
   8.2.1 Investigation of a phosphoenzyme intermediate ................................................. 208
   8.2.2 Site-directed mutation studies .............................................................................. 208
   8.2.3 Conditional essentiality of AnmK ....................................................................... 209
   8.2.4 Inquiry into related kinases .................................................................................. 209
   8.2.5 Substrate mimetics as potential inhibitors .......................................................... 210

Appendix A – Supplemental Figures .............................................................................. 212
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Act</td>
<td>Actin</td>
</tr>
<tr>
<td>ADIT</td>
<td>Auto Deposition Input Tool</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine-5'-diphosphate</td>
</tr>
<tr>
<td>ADSC</td>
<td>Area Detector Systems Corporation</td>
</tr>
<tr>
<td>AFM</td>
<td>Atomic force microscopy</td>
</tr>
<tr>
<td>AIDS</td>
<td>Acquired immune deficiency syndrome</td>
</tr>
<tr>
<td>Ala</td>
<td>Alanine</td>
</tr>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>AmpD</td>
<td>1,6-anhydro-N-acetylmuramyl-L-alanine amidase</td>
</tr>
<tr>
<td>AmpG</td>
<td>Peptidoglycan permease</td>
</tr>
<tr>
<td>anhGlcNAc</td>
<td>2-Acetamido-1,6-anhydro-2-deoxy-D-glucose</td>
</tr>
<tr>
<td>anhMurNAc</td>
<td>2-Acetamido-1,6-anhydro-2-deoxy-D-muramic acid</td>
</tr>
<tr>
<td>anmK/AnmK</td>
<td>Anhydromuramic acid kinase</td>
</tr>
<tr>
<td>ANL</td>
<td>Argonne National Laboratory</td>
</tr>
<tr>
<td>ANP</td>
<td>Adenosine 5'-(β,γ-imido)triphosphate</td>
</tr>
<tr>
<td>APS</td>
<td>Advanced Photon Source</td>
</tr>
<tr>
<td>APT</td>
<td>Attached proton test</td>
</tr>
<tr>
<td>ARP/wARP</td>
<td>Automated Refinement for Protein crystallography/weighted ARP</td>
</tr>
</tbody>
</table>
ASKHA  Acetate and Sugar Kinases/Heat shock cognate/Actin family

ASU  Asymmetric unit

ATP  Adenosine-5'-triphosphate

*B. subtilis*  *Bacillus subtilis*

BNL  Brookhaven National Lab

bp  Base pair

CBASS  Crystallography at Brookhaven Acquisition Software System

DALI  Distance matrix Alignment

*mDAP*  meso-Diaminopimelic acid

DBU  1,8-diazabicyclo[5.4.0]undec-7-ene

DNA  Deoxyribonucleic acid

DpnI  *Streptococcus pneumoniae* G41 restriction endonuclease I

*E. coli*  *Escherichia coli*

EDTA  Ethylenediaminetetraacetic acid

EM  Electron microscopy

ESI-MS  Electrospray ionization mass spectrometry

EtOH  Ethanol

FPLC  Fast protein liquid chromatography

GlcNAc  *N*-Acetyl-D-glucosamine

GlcNAc-6-P  *N*-Acetyl-D-glucosamine-6-phosphate
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu</td>
<td>Glutamate (or glutamic acid)</td>
</tr>
<tr>
<td>GK</td>
<td>Glycerol kinase</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione-S-transferase</td>
</tr>
<tr>
<td>h</td>
<td>Hour(s)</td>
</tr>
<tr>
<td>H. pylori</td>
<td>Helicobacter pylori</td>
</tr>
<tr>
<td>HA</td>
<td>Hydroxyapatite</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>Hsc</td>
<td>Heat shock cognate</td>
</tr>
<tr>
<td>IDCP</td>
<td>Iodonium di-sym-collidine perchlorate</td>
</tr>
<tr>
<td>IEX</td>
<td>Ion exchange chromatography</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilized metal affinity chromatography</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>Kan</td>
<td>Kanamycin</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase</td>
</tr>
<tr>
<td>kD</td>
<td>Kilodalton</td>
</tr>
<tr>
<td>kg</td>
<td>Kilogram</td>
</tr>
<tr>
<td>KOH</td>
<td>Potassium hydroxide</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>Leu</td>
<td>Leucine</td>
</tr>
</tbody>
</table>
NagK  N-Acetylglucosamine kinase
NagZ  β-N-Acetylglucosaminidase
NaH  Sodium hydride
NEB  New England Biolabs
nm  Nanometer
nM  Nanomolar
NMR  Nuclear magnetic resonance
NSLS  National Synchrotron Light Source
NTA  Nitrilotriacetic acid
OD  Optical density
PCR  Polymerase chain reaction
PDB  Protein Data Bank
PEG  Polyethylene glycol
PEI  Polyethylenimine
PG  Peptidoglycan
PHENIX  Python-based Hierarchical ENvironment for Integrated Xtallography
pM  Picomolar
PMSF  Phenylmethylsulfonyl fluoride
ppm  Parts per million
psi  Pounds per square inch
PstI  *Providencia stuartii* restriction endonuclease I

PvuI  *Proteus vulgaris* restriction endonuclease I

PXRR  Protein Crystallography Research Resource

RCSB  Research Collaboratory for Structural Bioinformatics

RMSD  Root means square deviation (or difference)

ROK  Repressors, Open reading frames and Kinases family

rpm  Rotations per minute

rt  Room temperature

Rxn  Reaction

*S. aureus*  *Staphylococcus aureus*

*S. oneidensis*  *Shewanella oneidensis*

SAD  Single-wavelength anomalous dispersion (or diffraction)

SDS-PAGE  Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SeMet  Selenomethionine

SER-CAT  Southeastern Regional Collaborative Access Team

SHARP  Statistical Heavy Atom Refinement And Phasing

S-SAD  Sulfur SAD

Taq  *Thermophilus aquaticus*

TBAF  Tetra-N-butylammonium fluoride

TBS  *t*-Butyldimethylsilyl
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEM</td>
<td>Transmission electron microscopy</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco etch virus</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin layer chromatography</td>
</tr>
<tr>
<td>TLS</td>
<td>Translation/Libration/Screw</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris-hydroxymethylaminomethane hydrochloride</td>
</tr>
<tr>
<td>TsCl</td>
<td>Tosyl chloride</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>Volts</td>
</tr>
<tr>
<td>Å</td>
<td>Ångstrom (1 Å = 10^{-10} m)</td>
</tr>
<tr>
<td>µg</td>
<td>Microgram</td>
</tr>
<tr>
<td>µL</td>
<td>Microliters</td>
</tr>
<tr>
<td>µm</td>
<td>Micrometer (or micron)</td>
</tr>
<tr>
<td>µM</td>
<td>Micromolar</td>
</tr>
</tbody>
</table>
List of Tables

Table 1 - Reagents and Thermal Cycle Used For Pfx Amplification of anmK ..................32
Table 2 - Reagents and Thermal Cycle Used for Taq Verification of anmK ......................32
Table 3 - Final Reagent Concentrations in Autoinduction Media ................................36
Table 4 - Recipes for 50x5052, ZY and 50xM ..........................................................36
Table 5 - Components of the 1000x Trace Metals Mix (100mL) .................................36
Table 6 - Apparent Kinetic Values for the Carbohydrate Substrates .............................75
Table 7 - Data Collection Statistics for the Initial Home Source AnmK Data ...............122
Table 8 - Data Collection and Refinement Statistics for de novo AnmK Structure ........135
Table 9 - Data Collection and Refinement Statistics for anhMurNAc-ANP-AnmK Structure (3SFA) ..........................................................147
Table 10 - Data Collection and Refinement Statistics for ADP\(\cdot\)VO\(_3\)\-bound AnmK ..176
List of Figures

Figure 1 - Gram-Positive and Gram-Negative Bacterial Cell Membrane...............................6
Figure 2 – *E. coli* Peptidoglycan on a Molecular Level and in a Cross-Linked Complex....9
Figure 3 - Peptidoglycan Biosynthesis Starting from the First Committed Step, GlmS....11
Figure 4 - The Enzymes Responsible for Peptidoglycan Breakdown ................................14
Figure 5 - Interaction Between Peptidoglycan Synthesis, Breakdown and Recycling ......16
Figure 6 - Targets for Antibiotic Chemotherapy Including Cell Wall Recycling ..............18
Figure 7 - Cell Wall Recycling Following Entrance Through the Inner Membrane. ..........24
Figure 8 – An Additional Anhydrosugar Utilized in this Work and the Substrate of Levoglucosan Kinase...............................................................................................................26
Figure 9 - PCR Amplification of *anmK* Showing Appropriate Size (1,100 bp) ..............51
Figure 10 - *PvuI* Restriction Endonuclease Digestion of *anmK/TEV/pENTR* Vector to Verify Gene Insertion........................................................................................................52
Figure 11 - *PstI* Restriction Endonuclease Digestion of pDEST17/TEV *anmK* Vector to Verify Gene Insertion.................................................................53
Figure 12 - SDS-PAGE Characterization of Initial pDEST Vector Expression..................56
Figure 13 – A Representative FPLC UV Trace Comparing Auto- and IPTG-Induced AnmK After Ni-NTA Chromatography..........................................................58
Figure 14 - The UV Trace of Elution from the Glutathione Column...............................61
Figure 15 - SDS-PAGE experiment with AnmK post-Ni-NTA (1), post-Q-Sepharose (2) and post-S-200 (3).................................................................................................64
Figure 16 - Comparative SDS-PAGE of Pellet (a) and Supernatant (b) in (NH₄)₂SO₄ Precipitation Experiments. ..................................................................................67
Figure 17 - SDS-PAGE Comparison of 10% and 20% (NH₄)₂SO₄ Cut Purity.................68
Figure 18 - Radiometric Analysis of the AnmK pH Optimum ........................................73

Figure 19 - Radiometric Assay Showing AnmK Enzymatic Activity with AnhGlcNAc and Inhibition By 10 mM ADP .................................................................................73

Figure 20 - Representative PEI-Cellulose TLC KM Determination Using AnmK (10 nM), ATP (100 mM) and varying concentrations of anhGlcNAc (100 µM – 10 mM)) ................................76

Figure 21 – INTELLI-PLATE™ and Common Setup with Recipe Volumes ................113

Figure 22 - Typical Growth Pattern of AnmK Crystals in Screening Plates ............118

Figure 23 - Pictures of Different AnmK Crystal Forms ...........................................119

Figure 24 - Anomalous Difference Density for Mse 193 and Mse 195 .........................126

Figure 25 - The Magic Triangle (I3C) for Phasing .........................................................127

Figure 26 - Depiction of the RNase H-like Fold .........................................................130

Figure 27 - The Biological and Asymmetric Unit of AnmK .........................................131

Figure 28 - The Overall Topology of the Open Conformation of AnmK ..................134

Figure 29 - Hanging Drop With Two Crystal Morphologies .................................144

Figure 30 - The Asymmetric Unit of AnmK Divided into Four Biological Dimers ........146

Figure 31 - The Asymmetric Unit of AnmK Consisting of Eight Monomers .............149

Figure 32 - The Complete Overall Topology Diagram of AnmK with Residue Numbering for Secondary Structural Elements .................................................................150

Figure 33 - Active Site of AnmK with Electron Density Around Substrates ............152

Figure 34 - Structural Motifs in AnmK ......................................................................154

Figure 35 - Binding Site Interactions with AnhMurNAc .............................................156

Figure 36 - Binding Site Interactions with ANP .........................................................159
Figure 37 - Surface and Cartoon View of the AnmK (purple) Monomer with ANP (pink) Bound in the Active Site Showing Solvent Accessibility .................................................................160

Figure 38 – The Interfacial Zinc Binding Site at the Interface of Chain H (green) and Chain D (purple) ..............................................................................................................162

Figure 39 - Structural Overlay of the Open and Closed Conformations of AnmK............164

Figure 40 - Electron Density Mesh Around the AnhMurNAc Residues in the Structure ..............................................................................................................................165

Figure 41 - Sequence Homology Between Three AnmK Species..................................167

Figure 42 - Sequence Homology Colored by Conservation by ConSurf [157-159] ..........168

Figure 43 - Crystals in Hanging Drop Vapor Diffusion Trials Containing ADP•VO₃ ....175

Figure 44 - Important Binding Site Interactions with ADP, Mg⁺ and VO₃ .....................178

Figure 45 - Active Site Electron Density for ADP•VO₃ and Relevant VO₃ Bond Lengths ...........................................................................................................................179

Figure 46 - Important Binding Site Interactions with AnhMurNAc............................181

Figure 47 - Structural Overlay of AnmK and NagK Illustrating Active Site Geometry .184

Figure 48 - AnhGlcNAc Reaction Being Probed By ¹⁸O-Labeling Experiments ..........186

Figure 49 - Proposed ¹⁸O Labeling at C1 from Bulk Solvent ........................................187

Figure 50 - Mechanism of Mutorotation and Predicted Retention of Label (black).......189

Figure 51 - MS/MS Data of 301.9 m/z Parent Ion Peak in AnmK Reaction in 50% H₂¹⁸O 192

Figure 52 - The ¹³C NMR of the AnmK Control Reaction Mixture...............................194

Figure 53 - ¹³C NMR Spectra Comparing H₂¹⁶O with H₂¹⁸O Experiment with Peak Shift of the MurNAc-6-P C1 α Anomer.................................................................195

Figure 54 - General Mechanism of β-Glycosidase Enzymes ......................................196

Figure 55 - Comparison of AnmK to Two Other Kinases and A Hydrolase ..............198
Figure 56 - Electron Density of Important Active Site Residues and a Cartoon Simplification of the Proposed Mechanism .................................................................204

Figure 57 – Proposed AnhMurNAc Substrate Mimic Molecules .................................................211
List of Schemes

Scheme 1 - The ATP- and Magnesium Ion-Dependent AnmK Catalyzed Reaction........23
Scheme 2 - Paulsen, Himpkamp and Peters Synthesis of AnhMurNAc (1986)........82
Scheme 3 - Tailler, Jacquinet, Noirot and Beau Synthesis of 1,6-Anhydro-2-Azido-4-O- Benzyl-2-Deoxy-d-Glucose (1992).................................................................82
Scheme 4 - Kubasch and Schmidt Synthesis of anhMurNAc (2002)......................83
Scheme 5 - Kinzy and Schmidt (1987)................................................................83
Scheme 6 - Zottola, Alonso, Vite and Fraser-Reid Synthesis of Levoglucosan (1989)......84
Scheme 7 - Thomson and von Itzstein Synthesis of anhManNAc (1995).....................84
Scheme 8 - Retrosynthetic analysis of anhMurNAc .............................................85
Scheme 9 - Retrosynthetic analysis of anhGlcNAc.............................................87
Scheme 10 - The Overall Proposed Synthetic Scheme for AnhMurNAc.................101
Scheme 11 - Completed Steps in the AnhMurNAc Synthesis Starting from D-Gluca1...102
..............................................................................................................................103
Scheme 13 - The Completed Synthesis of AnhMurNAc.........................................104
Scheme 14 - The Final Synthetic Routes for AnhGlcNAc.......................................106
List of Mechanisms

Mechanism 1 - Proposed Lytic Transglycosylase Mechanism for the Formation of AnhMurNAc by Scheurwater, Reid and Clarke [67], and Blackburn and Clarke [65]. ....20

Mechanism 2 - Concomitant Hydrolysis and Phosphorylation with Asp187 as the Catalytic Base Glu330 in Close Proximity with A Potential Role in the Mechanism. .....200

Mechanism 3 - The Proposed Reaction Mechanism of AnmK.........................................................202
Acknowledgements

First and foremost, I would like to thank my thesis advisor, Dewey McCafferty, who has supported me throughout my graduate career. My time at Duke will never be forgotten as I have learned so much and grown immensely under his guidance. I will be forever grateful that he granted me the room to work in my own way by giving me rope, but trying not to let me hang myself with it.

I would also like to thank Nate Nicely and Charlie Pemble of the Duke X-Ray Facility. Nate generously let me be a guinea pig for the facility and invaluably taught me tips and tricks about a technique that has always fascinated me. Charlie was my PyMol, Keynote and graphics guru, and general sympathetic listener. I am also grateful to Karen Allen and her lab members who were instrumental in introducing me to the synchrotron. This thesis would not have been possible without the help and support from my fellow lab members. Although scientific assistance was valuable, it was perhaps Annual Puffy Vest Day, the Lab Calendar, Sporcle-ing and such that really kept me going. On a more scientific note, I would like to thank Brad Hover and Jake Elvey who worked with me on AnmK to provide countless moments of pure entertainment.

Finally, I’d like to thank the Durham dance community including ShaLeigh Comerford, my company, Blank Slate Dance Company, and Ninth Street Dance for embracing me as I sought a necessary artistic and healthy release from graduate school.
Chapter 1: General Introduction

1.1 Bacterial Infection and Antibiotic Resistance

The increased use of broad-spectrum antibiotics has lead to bacterial populations acquiring specialized drug resistance mechanisms [1-3]. Treatment failure and its related costs are on the rise [4] and unfortunately, antibiotic resistance has been witnessed for all of the drugs currently on the market [5] including the “drug of last resort,” – vancomycin [6]. The incidence of deaths in the United States attributed to antibiotic resistance has surpassed deaths from HIV/AIDS [7]. In fact, an estimated two million Americans (10% of all hospitalized patients) will acquire a nosocomial infection and more than 70% of these hospital-acquired infections will be resistant to at least one routinely prescribed antibiotic [8]. This amounts to one infected individual dying roughly every six minutes from a nosocomial infection, which leads to a total of 90,000 deaths annually. Unfortunately, this incidence of mortality is not only becoming more prevalent in hospital settings, but is on the rise in local communities as well [8, 9].

Furthermore, Clatworthy et al. [5] published a report that compared the deployment year of antibiotics with the year in which resistance was observed. This review illustrated how antibiotic resistant bacteria have been isolated for every current antibiotic on the market and in some cases, the time between drug release and observed resistance was ominously brief. Walsh direly states, in reference to the emergence of antibiotic resistant bacteria, that it is “not a matter of if, but a matter of when” [10].
Therefore, advancements in antimicrobial chemotherapy to reduce or circumvent antibiotic resistance are critical to completely address the looming public health specter of bacterial resistance.

1.2 Targets for Therapeutic Action

The ability to prevent the adverse activity of bacteria in humans requires an understanding of how they may be targeted. Currently, there are four established targets of antibiotics: (1) bacterial DNA/RNA replication and repair; (2) bacterial protein synthesis; (3) folate coenzyme biosynthesis; and (4) bacterial cell-wall biosynthesis [10, 11]. DNA/RNA replication and repair is the target of the quinolones, most notably ciprofloxacin, which inhibits DNA replication, whereas rifampicin interferes with RNA replication. Bacterial protein biosynthesis is inhibited by multiple classes of antimicrobial compounds including the macrolides, aminoglycosides and most recently, the oxazolidinones. The two sulfa drugs, sulfamethoxazole and trimethoprim, are responsible for interfering with folate metabolism and coenzyme biosynthesis. Finally, bacterial cell wall biosynthesis can be interrupted by a number of antibiotics such as the penicillins, fosfomycin, ramoplanin, moenomycin and bacitracin.

Recently, cell wall modification has been introduced as potential target of new antimicrobials [10] and a target for abrogating bacterial virulence [12]. Modification can occur through the action of sortase enzymes involved in covalently linking cell-surface proteins to the cell wall peptidoglycan and through the cellular recycling of existing
peptidoglycan, which will be the topic discussed in this work following an introduction to the constituents and construction of the bacterial cell wall.

### 1.3 Components and Organization of the Bacterial Cell Wall

Despite the new advances in elucidating alternative targets for antibiotics, the bacterial cell wall may still warrant further analysis. In the 1950s, transmission electron microscopy (TEM) allowed the first high-resolution visualization of the cell wall [13] and these images corroborated the distinction already elucidated by Gram [14] – the differences between Gram-positive and Gram-negative cell wall. While the Gram-negative cell wall contains both an inner and outer membrane surrounding the thin layer of peptidoglycan (also known as murein) in the periplasmic space, Gram-positive bacteria lack the outer membrane and are only surrounded by a thick layer of peptidoglycan around the cytoplasmic membrane (Figure 1).

The cell wall comprises approximately 10% of the complete dry cell mass in Gram-negative bacteria [15]. In *E. coli*, 75 to 80% of the peptidoglycan is a monolayer [16] that is covalently attached to lipoprotein (Lpp; Braun’s lipoprotein) inside the periplasmic space, which is then attached to the outer membrane (Figure 1). In contrast to Gram-negative bacteria, the peptidoglycan layer covering Gram-positive bacteria is between 10- and 20-fold thicker and peptidoglycan accounts for approximately 50 to 70% of the cell dry mass [15].
In addition to peptidoglycan, the cell wall of Gram-positive bacteria contains peptidoglycan- and membrane-anchored anionic polymers such as teichoic and lipoteichoic acids [17, 18]. The main chain of teichoic acid contains approximately 40 repeating phosphate units linked by a polyhydroxylated alkyl chain. This polyphosphate chain is connected to the peptidoglycan via several glycerol-phosphate (GroP) residues and the ManNAc-β-1,4-GlcNAc-1-phosphate disaccharide that connects to the C6 hydroxyl of a peptidoglycan MurNAc acid residue [19-22]. Gram-positive bacterial cell walls also contain porins that can sort molecules by their molecular weight to allow the passage of essential hydrophilic nutrients through the outer membrane [23].

Despite the architectural differences, there is one molecule common to both Gram-positive and Gram-negative bacteria – peptidoglycan. This molecule can be described as the stabilizing mesh present in all bacteria with the rare exception of classes such as mycoplasma and planctomycetes [24]. Peptidoglycan is a single polymeric macromolecule that jackets the entire cell and its primary function is to provide a discrete barrier for the bacterium against bursting due to elevated osmotic pressure that is part of the intrinsic turgor experienced by the cell [25, 26]. This exoskeleton-like covering insures that the acquired internal pressure (2 to 5 atm in Gram-negative bacteria and up to 50 atm in Gram-positive [26]) does not destroy the cell and discharge its contents. The maintenance of bacterial life relies on this barrier between the cellular contents and the ever-changing external environment. Without the cell wall
peptidoglycan, bacteria would certainly be unable to survive. Therefore, targeting the cell wall is a logical choice for targeting of bacterial death.

The following sections summarize the primary structure of peptidoglycan units, the intermolecular linkages that make up the secondary structure and finally, the tertiary structure of the peptidoglycan sacculus. These summaries will segue into the description of how peptidoglycan is metabolized in the bacterial cell and lead to a discussion of bacterial cell wall modification by recycling.
Figure 1 - Gram-Positive and Gram-Negative Bacterial Cell Membrane

(A) Gram-negative bacteria contain both the inner (IM) and outer (OM) membranes (both blue with green lipid tail) in addition to lipoprotein (Lpp; orange) attached to the peptidoglycan layer(s) (purple) (B) Gram-positive bacteria lack the outer membrane (blue with teal porins) instead they are covered with 10 to 20 layers of peptidoglycan (purple). Both contain integral membrane proteins (red).
1.3.1 Peptidoglycan structure and linkage

Peptidoglycan is comprised of a glycan backbone with alternating units of \(N\)-acetylglucosamine (GlcNAc) and its 3-O-lactyl ether derivative, \(N\)-acetylmuramic acid (MurNAc) that are linked by \(\beta\)-1,4-glycosidic bonds (Figure 2). In \(E.\ coli\), these glycan strands terminate with 1,6-anhydro-\(N\)-acetylmuramic acid (anhMurNAc) at the formally reducing end and GlcNAc at the non-reducing end \([27, 28]\). Generally, the lengths of the glycan strands vary \([29, 30]\), but \(E.\ coli\) peptidoglycan is approximately 21 disaccharide units on average \([31]\).

The carboxy terminus of the 3-O-lactyl ether of the MurNAc residues is covalently attached by a \(D, L\)-amide bond to the peptidoglycan pentapeptide, which can vary in amino acid composition between bacterial species. The pentapeptide contains both unusual non-codogenic \(D\)-amino acids (\(D\)-Ala, \(D\)-Glu) and a normal \(L\)-amino acid (\(L\)-Ala). \(E.\ coli\) also contain an \(\epsilon\)-carboxy derivative of lysine, \(meso\)-diaminopimelic acid (\(mDAP\)) in the pendant \(L\)-alanine-\(\gamma\)-\(D\)-glutamate-(\(L\))-\(mDAP\)-\(D\)-alanine-(\(D\)-alanine) (\(L\)-Ala-\(\gamma\)-\(D\)-Glu-\(mDAP\)-\(D\)-Ala-(\(D\)-Ala)) pentapeptide. The fifth amino acid is only present if it has not yet been utilized in the transpeptidation crosslinking reaction between two separate pendant pentapeptide chains.

The composition of these pendant (or stem) pentapeptides varies among bacterial species. The di-basic amino acid at the third position is the most notable variant. \(mDAP\), an intermediate in the biosynthesis of lysine \([20, 31, 32]\) is present in
Gram-negative bacteria and certain Bacilli, but most Gram-positive bacteria contain L-lysine (L-Lys) at this position. In a small number of cases, the third position is occupied by non-natural amino acids such as L-ornithine (L-Orn), L-5-hydroxylysine and L,L-diaminopimelate (LL-DAP) [33]. In fact, the only position without variation among bacterial species is the fourth (D-Ala) position [33].

The mesh-like network of bonds that joins the glycan backbone strands together is formed by the peptide moieties of peptidoglycan. This occurs between the di-basic amino acid on one chain (mDAP in E. coli) and a peptide carboxyl group on another chain. The E. coli D,D-type crosslinks are formed between the ε-amino group of mDAP and the amide group of D-Ala (in the fourth position) of a neighboring glycan strand [34, 35]. In Gram-positve species such as S. aureus, the bridge between the two pendant pentapeptides is not direct; instead, it is mediated by a pentaglycine linker peptide [36]. The strain and growth conditions cause variation in peptidoglycan crosslinking [34, 37], but approximately 40 to 60% of the E. coli pendant peptides are involved in crosslinks that form the peptidoglycan layer [34].
Figure 2 – E. coli Peptidoglycan on a Molecular Level and in a Cross-Linked Complex
The encircled figure (left) is the E. coli peptidoglycan building block – GlcNAc-MurNAc-pentapeptide (L-Ala-γ-D-Glu-m-DAP-D-Ala-D-Ala). GlcNAc (orange) and MurNAc (red) are connected by a β-1,4-glycosidic bond and the peptide is connected via a D,L-amide bond to the D-lactic acid ether of MurNAc. The right-hand illustration is colored to match the molecular structure and serves to show the crosslinking between the disaccharide subunits. For Gram-negative bacteria, the glycan strands are connected via a D,D-peptide crosslink between the terminal amino group of mDAP and the amide group of the amino acid at the fourth position (D-Ala).
1.3.2 The peptidoglycan sacculus

The sacculus is a mesh-like cellular peptidoglycan envelope that can reversibly expand and contract to its original size in response to the turgor pressure experienced by the cell [38]. In some cases, this elastic expansion can be as large as three-fold the resting state [39]. This enlargement is facilitated by the alternating D-L-D-L-peptide sequence starting with the d-lactate attached to MurNAc, which allows for flexibility [40]. In contrast to the peptide, the glycan backbone is rather inelastic [41]. Multiple mutually exclusive models have been introduced to aid in the understanding of the architecture and connectivity of peptidoglycan. Thus, the following section aims to summarize both the catabolic and anabolic steps in peptidoglycan biosynthesis.

1.4 Cell Wall Metabolism
1.4.1 Peptidoglycan biosynthesis

For both Gram-positive and Gram-negative bacteria, peptidoglycan biosynthesis is nearly identical occurring in three different cellular compartments (cytoplasm, periplasm and peptidoglycan) in a three step process illustrated in Figure 3. First, UDP-MurNAc-peptide, also known as “Park’s nucleotide [42],” and UDP-GlcNAc are synthesized in the cytoplasm. Then, lipid-linked intermediates (lipid I and lipid II) are synthesized on the inner side of the cytoplasmic membrane and lipid II is translocated across the membrane. The process terminates in the periplasm where the disaccharide-peptide monomer units of lipid II are inserted into the growing peptidoglycan sacculus.
Figure 3 - Peptidoglycan Biosynthesis Starting from the First Committed Step, GlmS
The assembly of the UDP-GlcNAc precursor starts with fructose-6-P and L-glutamine and utilizes three enzymes in four steps: the aminotransferase, GlmS [43]; the mutase, GlmM [44, 45]; and GlmU, a bifunctional transferase [46, 47]. Then, UDP-MurNAc is formed by two enzymes, MurA and MurB [48]. Hereafter, the production of UDP-MurNAc-pentapeptide proceeds iteratively by the addition of the peptides onto the D-lactyl ether group of UDP-MurNAc with concomitant ATP hydrolysis. This process is catalyzed by the ligases MurC, MurD, and MurE [35, 48]. Lastly, the D-Ala-D-Ala moiety is attached by the MurF ligase to the mDAP residues to provide amino acids important in the transpeptidase crosslinking reaction [49].

The first step in the second compartment is carried out by the MraY transferase on the inside of the cytoplasmic membrane. MraY transfers the phospho-moiety of UDP-MurNAc-pentapeptide to the membrane-associated undecaprenyl-phosphate to yield MurNAc-pentapeptide-pyrophosphoryl undecaprenol, or lipid I [48, 50]. The MurG transferase then adds the GlcNAc of UDP-GlcNAc to lipid I to produce undecaprenyl-pyrophosphoryl-MurNAc-pentapeptide-GlcNAc (lipid II) [51], which is transported across the membrane by an essential “flippase” recently annotated through bioinformatic investigation as MurJ [52, 53].

The final stage in the biosynthesis of peptidoglycan involves the formation of lipid II polymers and subsequent incorporation into the peptidoglycan sacculus. These processes occur outside the cytoplasmic membrane in the periplasmic space. In E. coli, a
glycosyl transferase, MtgA [54] polymerizes the glycan strands that are then transferred to transpeptidases for linkage of the peptide moieties [18]. This is a two-step reaction in \textit{E. coli} starting with the hydrolysis of D-Ala-D-Ala to form an enzyme-substrate intermediate and release the terminal D-Ala. Then, the resultant tetrapeptide is transferred to the acceptor dibasic amino acid (\textit{mDAP}) in the second pendant peptide [18] to form the covalent linkages discussed earlier (section 1.3.1).

\subsection*{1.4.2 Peptidoglycan turnover}

Likewise, there are specific hydrolases to target nearly every covalent linkage made in the biosynthesis of peptidoglycan. These enzymes are classified in five categories in reference to the specific bond that they cleave (Figure 4). The glucosaminidases cleave the GlcNAc residues from non-reducing ends of the peptidoglycan strands. The \textit{N}-acetylmuramyl-L-alanine amidases remove the polypeptide from MurNAc and the carboxypeptidases target either the DAP-Ala L,D-peptide linkage or the Ala-Ala D,D linkage. The endopeptidases catalyze the cleavage of the peptide bond crosslinking the pendant pentapeptide strands between two peptidoglycan strands. Finally, the lytic transglycosylases catalyze the cleavage of the GlcNAc-MurNAc glycosidic linkages and are the first step in peptidoglycan recycling.
Figure 4 - The Enzymes Responsible for Peptidoglycan Breakdown
1.5 Cell Wall Recycling

Bacteria continuously recycle pre-existing peptidoglycan in order to minimize the \textit{de novo} synthesis of peptidoglycan precursors. Cell wall recycling was first postulated and experimentally verified in the 1980s [55, 56]. This process captures peptidoglycan catabolites for conversion into cellular building blocks that feed into cell wall synthesis and minimize the energy and resources required for the cell to synthesize new cell wall. Instead of targeting \textit{de novo} bacterial cell wall synthesis like the penicillins, targeting bacterial cell wall recycling of pre-existing murein could retard or arrest cell growth. Peptidoglycan recycling is an essential process because bacteria such as \textit{E. coli}, for example, can break down approximately 60\% of their cell wall peptidoglycan during every replication [55, 57]. Various studies have confirmed that glycopeptides arising from recycling are harnessed to regenerate 25\% of the growing cell wall [27, 58, 59]. Figure 5 illustrates where peptidoglycan recycling fits into the cycle of peptidoglycan breakdown and synthesis. Figure 6 depicts the addition of recycling into an illustration of three of the four aforementioned classic targets of antimicrobial chemotherapy.
Figure 5 - Interaction Between Peptidoglycan Synthesis, Breakdown and Recycling
If blocking the peptidoglycan recycling pathway by itself is not effective, it may play an integral role in antibiotic combination therapy. This combinatorial approach to medical treatment is currently being employed in the treatment of tuberculosis [11, 60], HIV/AIDS [61] and bacterial sepsis [62], and often proves to be more useful than monotherapy [63, 64].
Figure 6 - Targets for Antibiotic Chemotherapy Including Cell Wall Recycling
Adapted from [11] to show the three main targets for antibiotic chemotherapy and the additional pathway occurring in bacteria, peptidoglycan recycling.
1.6 Cell Wall Recycling Enzymes and Anhydromuramic Acid Kinase (AnmK)

The cell wall recycling pathway begins with the initial peptidoglycan breakdown by lytic transglycosylases (LTGases) [65-69] that catalyze peptidoglycan cleavage between MurNAc and GlcNAc residues in the backbone to result in the formation of an anhMurNAc residue. In the reaction mechanism (Mechanism 1), proposed by Scheurwater, Reid and Clarke [67] and Blackburn and Clarke [65], a catalytic glutamic acid acts as a general acid to protonate the glycosidic linkage. By neighboring group participation, this protonation leads to the formation of a muramoyl oxazolinium ion intermediate. Glutamate then acts as a general base to abstract the C6-hydroxyl proton from the oxazolinium ion, which consequently promotes the collapse of the transition state and the concomitant formation of anhMurNAc.
Mechanism 1 - Proposed Lytic Transglycosylase Mechanism for the Formation of AnhMurNAc by Scheurwater, Reid and Clarke [67], and Blackburn and Clarke [65]. R and R'' are the adjacent GlcNAc molecules in the peptidoglycan chain and R' is the pendant pentapeptide.
This anhydromuropeptide (GlcNAc-anhMurNAc-pentapeptide) is then imported by the cell permease, AmpG (Figure 7) [52, 70-75], which is a membrane spanning substrate transport protein [76, 77]. It has been established that virtually all of the peptidoglycan-derived mDAP undergoes turnover and recycling passes through AmpG [59]. AmpG is specific for the GlcNAc-anhMurNAc disaccharide and the GlcNAc-anhMurNAc disaccharide with the pendant peptide [70]. AmpG does not transport anhMurNAc-tripeptide. Therefore, AmpG specifically transports intact GlcNAc-anhMurNAc with or without the pendant polypeptide into the cytoplasm [70].

Following entry into the cytoplasm, NagZ cleaves the GlcNAc-anhMurNAc glycosidic bond (Figure 7). NagZ from E. coli was first characterized by Yem and Wu [78] in 1976. This enzyme is present in the cytoplasm and a deletion mutant of nagZ, lacking β-N-acetylglucosaminidase activity, results in accumulation of large amounts of the GlcNAc-anhMurNAc in the cytoplasm [79]. The NagZ enzyme is active against muropeptides in addition to anhydromuropeptides. The substrate specificity of NagZ was further investigated by Vötsch and Templin who discovered that NagZ is inhibited by one of its reaction products, GlcNAc [80].

To avoid inhibition by GlcNAc, the cell can convert GlcNAc into GlcNAc-6-P (3) using N-acetylglcosamine kinase (NagK) (Figure 7). The discovery of this enzyme revealed the key C6 phosphorylation step in the conversion of free GlcNAc sugars into substrates UDP-GlcNAc-6-P, which can be shunted into the peptidoglycan biosynthesis
pathway [58]. The initial discovery of NagK activity occurred in 1966 [81] and the gene was identified through N-terminal protein sequencing years later [82]. NagK is highly specific for GlcNAc, but at extremely high concentrations, glucose may also become phosphorylated [82]. Sequence homology reveals that NagK is a member of the ROK family [82]. The ROK family includes repressors, open reading frames, and kinases, which includes members that phosphorylate a vast range of structurally distinct hexoses including D-glucose, several D-glucose epimers and various acetylated hexosamines [83].

Following the GlcNAc-anhMurNAc glycosidic bond cleavage by NagZ, AmpD is responsible for the scission of the anhMurNAc-pentapeptide bond (Figure 7). An E. coli strain with an ampD knockout contains large amounts of anhMurNAc-tripeptide in the cytoplasm [59]. The anhMurNAc-tripeptide is a substrate for AmpD, which selectively cleaves the anhMurNAc-L-Ala bond with high specificity [84]. When the cleavage activity of purified AmpD was probed with MurNAc-peptides and UDP-MurNAc-pentapeptide, the rate was more than 10,000 times slower than that of the anhMurNAc-tripeptide [84]. Therefore, AmpD in the bacterial cytoplasm cleaves the desired bond without destroying UDP-MurNAc-pentapeptide present in the cytoplasm for peptidoglycan synthesis. Additionally, Höltje et al. demonstrated the anhydrosugar specificity of AmpD, which only cleaves the anhMurNAc-L-Ala bond and not the MurNAc-L-Ala bonds contained in the glycan strands [85].
After the final processing by AmpD and NagZ, anhMurNAc is liberated from the imported disaccharide-polypeptide (GlcNAc-anhMurNAc-pentapeptide) in a 1C\(_{4}\) chair confirmation (Figure 7, inside central blue pentagon). In 2005, Uehara et al. [57] described a new enzyme, AnmK, which was isolated from soluble E. coli cell extract based on the activity it displayed to convert [\(^3\)H]anhMurNAc into MurNAc-6-P. AnmK is involved in the essential process [58] of cell wall recycling and utilizes ATP and 2-acetamido-1,6-anhydro-2-deoxy-D-muramic acid (anhMurNAc, 1) to achieve hydrolysis of the intermolecular sugar acetal and phosphorylation of the liberated C6-hydroxylate functionality. This results in the formal addition of H\(_3\)PO\(_4\) to form 2-acetamido-2-deoxy-6-phospho-D-muramic acid (MurNAc-6-P, 2) in a 4C\(_{1}\) chair confirmation (Scheme 1 and Figure 7). AnmK and the recently discovered lactyl etherase, MurQ collectively convert anhMurNAc to 2-acetamido-2-deoxy-6-O-phospho-D-glucose (GlcNAc-6-P, 3), which can then be converted into UDP-GlcNAc-6-P for use in peptidoglycan biosynthesis [56].

![Scheme 1 - The ATP- and Magnesium Ion-Dependent AnmK Catalyzed Reaction](image-url)
Figure 7 - Cell Wall Recycling Following Entrance Through the Inner Membrane.
A lytic transglycosylase (LTGase) is responsible for the modification of peptidoglycan to terminate with an anhMurNAc residue. AmpG imports the GlcNAc-anhMurNAc-peptide into the cytoplasm where the GlcNAc-anhMurNAc glycosidic bond as cleaved by NagZ and the peptide is cleaved off of anhMurNAc by AmpD. anhMurNAc (1) is thus liberated in its unique, strained chair conformation, which is concurrently hydrolyzed and phosphorylated by AnmK. The product of this reaction, MurNAc-6-P (2), is metabolized by MurQ, a lactyl etherase that cleaves the C3-lactyl ether from MurNAc to yield GlcNAc-6-P (3). Meanwhile, NagK phosphorylates the GlcNAc that was the other product of NagZ. The resultant GlcNAc-6-P carries on into additional recycling process that yields reformed peptidoglycan.
1.7 Examination of AnmK Literature

To identify the enzyme responsible for the metabolism of anhMurNAc, Park and coworkers first inhibited peptidoglycan recycling by a nagZ knockout that caused the accumulation of GlcNAc-anhMurNAc in the cytoplasm. Similarly, a nagEBACD [the only other known source (NagE) and breakdown (NagA) enzymes for GlcNAc-6-P] knockout resulted in high levels of GlcNAc and GlcNAc-phosphate in the cytoplasm and a complete lack of anhMurNAc in the cytoplasm [56]. AnhMurNAc was also absent from the growth medium. When E. coli soluble cell extract was incubated with [3H]anhMurNAc, only trace amounts of product were isolable by high-performance liquid chromatography (HPLC) or thin-layer chromatography (TLC) from anhMurNAc [58]. However, when Mg\(^2+\) and ATP were added to the incubation mixture, the reaction proceeded to convert all [3H]anhMurNAc to two products, GlcNAc-P and MurNAc-P [57]. Mass spectrometry analysis following trypsin digestion of the partially purified enzyme mixture that converted anhMurNAc to MurNAc-6-P was utilized to obtain sequence fragments. One unknown open reading frame, ydhH, was identified and renamed AnmK [57]. AnmK homologs have been identified in most species of Gram-negative and some Gram-positive bacteria [56].

Furthermore, the unique activity of AnmK has fueled the investigation for enzymes with similar chemistry. The fungi Aspergillus niger [86] and Lipomyces starkeyi [87], and other microorganisms [88, 89] utilize levoglucosan kinase (LgK) to convert 1,6-
anhydroglucose (levoglucosan or anhGlc, 4, Figure 8) to glucose-6-phosphate. Both enzymes have exhibited substrate specificity for their respective carbohydrates. For example, LgK does not phosphorylate 1,6-anhydromannose (mannosan), the C2-epimer of levoglucosan as well as levoglucosan itself [90] and AnmK cannot phosphorylate MurNAc [57], the hydrolyzed form of anhMurNAc (1). Both of these enzymes exhibit kinase activity in the presence of ATP/Mg\(^{2+}\), but bear little sequence homology to any of the known families of kinases. Instead, they are a part of an uncharacterized protein family UPF0075. Initial alignments show close relation to each other and distant homology to enzymes from the ASKHA (Acetate and Sugar Kinases/Heat shock cognate/Actin family) subfamily of the ROK sugar kinases. These particular kinases are U-shaped and contain a $\beta_3\beta_2\alpha_1\beta_4\alpha_5\alpha$ characteristic ASKHA RNase H-like fold. This shape allows for the domain closure upon sugar binding in a deep cleft between the N- and C-termini [91]. Further research into these enzymes may elucidate an entirely new family of kinases that work like hexose kinases, but act on anhydrohexoses.

Figure 8 – An Additional Anhydrosugar Utilized in this Work and the Substrate of Levoglucosan Kinase
1.8 Thesis Goals

Pioneered by JT Park [56-58, 75], anhydromuramic acid kinase (AnmK) has been identified and is a recent addition to the peptidoglycan recycling pathway in bacteria. This enzyme is responsible for the concomitant hydrolysis and phosphorylation of anhMurNAc by an unknown mechanism. Although it is a potential target for new antimicrobial compounds because of its function in cell wall metabolism and remodeling, AnmK literature precedents report minimal biochemical, biophysical and structural characterization.

Towards this end, the first goal of this thesis is to create an expression and purification system to obtain soluble, active AnmK and to synthesize its substrate, anhMurNAc, which will enable all future studies. Through cloning, expression and purification of hexahistidine-labeled E. coli AnmK, the homogenous protein is produced in high yield. In addition, the anhMurNAc substrate is synthesized from a commercially available intermediate and the pseudosubstrate, anhGlcNAc is synthesized from N-acetylglucosamine.

Together, the AnmK protein and its substrate facilitate the second goal of this thesis, the elucidation of the substrate requirements of enzyme turnover and preliminary kinetics. Using a PEI-cellulose TLC assay designed to separate ATP from the reaction product, ADP, steady state parameters validate those set forth by Park [57]. The enzymatic activity of AnmK can be investigated with a series of four carbohydrate
substrates (anhMurNAc, anhGlcNAc, anhGlc and GlcNAc) and three nucleotide substrates (ATP, ADP and ANP).

The third and major aim of this thesis is to better understand the macromolecular structure of AnmK, its relationships to other kinases and molecular determinants of substrate specificity within the active site. Through x-ray crystallography, the de novo structure solution of AnmK was solved in an unbound conformation. With this information in hand, the closed conformation was solved in complex with anhMurNAc and ANP, a nonhydrolyzable ATP analog and with ADP and vanadate, a putative mimic of the transition state. From these structures, AnmK was confirmed as an ASKHA kinase family member and the binding site motifs were identified for both anhMurNAc and the ATP analog, ANP. Additionally, important residues that may play a role in catalysis for were discovered and a putative mechanism was designed.

Finally, the fourth goal is to perform an informed mechanistic study for further insight into the concomitant substrate hydrolysis and phosphorylation. This was accomplished by nuclear magnetic resonance (NMR) and mass spectrometric (MS) analyses of the incorporation of an $^{18}$O label into anhGlcNAc by AnmK. This information led the to support of our proposed mechanism for the concomitant hydrolysis and phosphorylation of the anhMurNAc carbohydrate substrate by AnmK.
Chapter 2: Protein Biochemistry of AnmK

2.1 Introduction

AnmK catalyzes the concomitant hydrolysis and phosphorylation of the anhydrohexose substrate, anhMurNAc (1) to yield MurNAc-6-P (2). This reaction is significant because of the unprecedented dual kinase/hydrolase functionality of AnmK and also because of the uniqueness of the covalent structure of its substrate. The production and purification of this enzyme would allow for further investigation of this novel chemistry. Initial work by Uehara et al. describes the expression, partial purification and preliminary activity characterization of AnmK [57].

Although AnmK has been expressed recombinantly by Uehara et al. [57], it has not been purified to homogeneity and fully characterized biochemically, biophysically or by structure. To facilitate these studies, we developed an expression and purification protocol to produce homogenous AnmK (10 mg/L cells) in quantities suitable for mechanistic studies and crystallographic analyses. The final AnmK purification scheme is an improvement on prior AnmK precedents [57] in that it entails fewer chromatographic steps and requires less than 36 hours to achieve full purification for both hexahistidine-tagged and selenomethionine-labeled AnmK.

In addition to the cloning, expression, chromatographic trials and final purification of AnmK, this chapter includes various biophysical and kinetic experiments utilized to characterize AnmK. This includes a radiometric thin-layer chromatography
(TLC)-based activity assay that separates ATP from ADP (one of the reaction products), sedimentation rate equilibrium analytical ultracentrifugation and matrix assisted laser desorption ionization (MALDI) mass spectrometry for the molecular weight of hexahistidine-tagged and selenomethionine-labeled AnmK monomers.

2.2 Methods for the Cloning, Expression and Purification of E. coli anmK

2.2.1 Methods for cloning E. coli anmK

The anmK gene was amplified from wild-type E. coli MG1655 genomic DNA through polymerase chain reaction (PCR) [92]. The reaction mixture contained Pfx DNA polymerase (Stratagene), wild type E. coli (MG1655) genomic DNA, and the primers, 5′-CAC CAT GAA ATC GGG CCG CTT TAT TGG-3′ and 5′-TCA CGG GTT TGC GGG GAA AAT AGC-3′. The reaction components and PCR thermal cycle parameters are listed in Table 1.

The full-length, 1110 base pair (bp) blunt-end anmK PCR product was visualized by agarose gel (1%), extracted and purified using a QIAquick Gel Extraction kit (QIAGEN). The isolated PCR product was then inserted into a pENTR™/D-TOPO® vector using the pENTR™ directional TOPO cloning kit (Invitrogen). After transformation into DH5α cells, plasmids from colonies demonstrating resistance to kanamycin (50 µg/mL) were isolated using a plasmid Mini-Prep Kit (QIAGEN). The sequence fidelity was confirmed by DNA sequencing (Duke DNA Analysis Facility), the use of restriction digest (PvuI; New England Biolabs (NEB)) and Taq (Stratagene) PCR
with the standard M13 forward and reverse primers (Invitrogen). For the Taq insert verification PCR, the following reagents and PCR thermal cycle parameters were employed (Table 2)
Table 1 - Reagents and Thermal Cycle Used For Pfx Amplification of anmK

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
<th>Temperature</th>
<th>Time</th>
<th>Repetition</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x Pfx buffer</td>
<td>2.5 µL</td>
<td>94 °C</td>
<td>2 min</td>
<td>Once</td>
</tr>
<tr>
<td>10 mM dNTPs</td>
<td>1 µL</td>
<td>94 °C</td>
<td>15 sec</td>
<td>35 times</td>
</tr>
<tr>
<td>MgSO₄</td>
<td>0.75 µL</td>
<td>52 – 63* °C</td>
<td>30 sec</td>
<td></td>
</tr>
<tr>
<td>20 µM Forward primer</td>
<td>0.5 µL</td>
<td>68 °C</td>
<td>2 min</td>
<td></td>
</tr>
<tr>
<td>20 µM Reverse primer</td>
<td>0.5 µL</td>
<td>4 °C</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>Genomic DNA (5.4ng/µL)</td>
<td>0.5 µL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pfx</td>
<td>0.5 µL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H₂O</td>
<td>18.75 µL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The annealing temperature did not affect yield.

Table 2 - Reagents and Thermal Cycle Used for Taq Verification of anmK

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
<th>Temperature</th>
<th>Time</th>
<th>Repetition</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x Taq buffer</td>
<td>2.5 µL</td>
<td>94 °C</td>
<td>2 min</td>
<td>Once</td>
</tr>
<tr>
<td>10 mM dNTPs</td>
<td>1 µL</td>
<td>94 °C</td>
<td>30 sec</td>
<td>35 times</td>
</tr>
<tr>
<td>50 mM MgCl₂</td>
<td>0.75 µL</td>
<td>62 °C</td>
<td>30 sec</td>
<td></td>
</tr>
<tr>
<td>20 µM Forward primer</td>
<td>0.5 µL</td>
<td>68 °C</td>
<td>2 min</td>
<td></td>
</tr>
<tr>
<td>20 µM Reverse primer</td>
<td>0.5 µL</td>
<td>72 °C</td>
<td>2 min</td>
<td>Once</td>
</tr>
<tr>
<td>Mini-Prep DNA</td>
<td>0.5 µL</td>
<td>4 °C</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>Taq</td>
<td>0.1 µL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H₂O</td>
<td>19.15 µL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Following isolation of the entry clone, multiple expression vectors were constructed using an LR Recombination Reaction kit (Invitrogen) and the supplier’s protocols. \textit{anmK} was cloned into the pDEST14\textsuperscript{TM}, pDEST15\textsuperscript{TM} and pDEST17\textsuperscript{TM} Gateway\textsuperscript{®} destination vectors. Plasmids from colonies with resistance to ampicillin (100 \( \mu \text{g/mL} \)), indicating insertion into the various pDEST vectors, were screened by \textit{PstI} digestion and transformed into BL21(DE3) cells to create cell lines for expression of the native, glutathione-S-transferase (GST) \textit{N}-terminal fusion and hexahistidine \textit{N}-terminal fusion proteins, respectively.

The same process was repeated to incorporate a Tobacco Etch Virus (TEV) cleavage site for future affinity tag removal. The initial \textit{anmK} PCR product was cloned into a \textit{pENTR}/TEV/D-TOPO\textsuperscript{®} vector using the \textit{pENTR} directional TOPO cloning kit (Invitrogen). The pDEST-based expression constructs were thus recreated as above with the \textit{N}-terminal incorporation of the TEV cleavage site.

\textbf{2.2.2 Methods for overexpression of AnmK}

\textbf{2.2.2.1 IPTG-induced expression of AnmK}

The five aforementioned pDEST\textsuperscript{TM} vectors containing \textit{anmK} were transformed into BL21(DE3) cells for expression. A bacterial colony from each of the Luria-Bertani (LB) agar and ampicillin (100 \( \mu \text{g/mL} \)) (Amp\textsubscript{100}) plates was used to inoculate a 250 mL flask containing 50 mL of LB broth supplemented with Amp\textsubscript{100}. These overnight preculture growths were incubated overnight with rotary shaking (180 rpm) at 37 \(^\circ\text{C} \).
The resultant bacterial growth was distributed among 8, 4 L flasks containing 1 L of LB media supplemented with ampicillin (100 µg/mL) and placed back in the rotary shaker at 37 °C. The growth progress was carefully monitored and then induced by IPTG (Sigma) when the optical density (OD<sub>600</sub>) of the growth reached 0.6. The IPTG-induced overexpression of the AnmK recombinant proteins then proceeded for 3 hours. Following overexpression, the cell pellet was harvested by centrifugation (2500 × g; 10 min; 4 °C). The supernatant liquid was decanted and the pellet was stored at -20 °C.

2.2.2.2 Se-Met enriched autoinduction expression

For the production of the Se-Met substituted AnmK, a non-induced, log phase starter culture was prepared from a single colony in LB broth supplemented with Amp<sub>100</sub> to serve as an inoculum. This culture was incubated with rotary shaking at 180 rpm until the OD<sub>600</sub> was between 0.5 and 1.0. The Overnight Express Autoinduction System 2 (Novagen) medium was prepared aseptically. Selenomethionine (125 mg/L), vitamin B12 (final concentration 100 nM) and the appropriate antibiotic (Amp<sub>100</sub>) for the host strain and plasmid were added. The media was inoculated with the starter culture and grown at 37 °C with rotary shaking at 180 rpm until saturated. Saturation was reached after approximately 16 h. The cells were harvested by centrifugation (2500 × g; 10 min; 4 °C) for the saturated, stationary phase culture for expression analysis and protein purification.
2.2.2.3 Autoinduction for expression of *E. coli* AnmK

Based on the work of Studier [93], autoinduction media (ZYM-5052) was prepared for use in place of IPTG-induced protein expression. This media was created using 5 different stock solutions: 50x5052, ZY, 1M MgSO₄, 1000x metals and 50xM (Table 4 and Table 5). To make ZYM-5052 (1000 mL), ZY (957 mL), 1M MgSO₄ (2 mL), 1000x metals (200 µL), 50x5052 (20 mL) and 50xM (20 mL) were added to two, 2.8 mL baffled flasks so that the final concentrations were as follows (Table 3).

As with the production of Se-Met substituted AnmK, a non-induced, log phase starter culture was prepared from a single colony in LB broth supplemented with Amp100 to serve as an inoculum. This culture was incubated overnight and diluted 1:100 into fresh media in the baffled flasks used for overexpression. The media was inoculated with the starter culture and grown at 37 °C with shaking at 180 rpm until saturated (approximately 16 h). The cells were harvested by centrifugation (2500 × g; 10 min; 4 °C) and stored at -20 °C.
Table 3 - Final Reagent Concentrations in Autoinduction Media

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>1 %</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>0.5%</td>
</tr>
<tr>
<td>Na₂HPO₄</td>
<td>25 mM</td>
</tr>
<tr>
<td>KH₂PO₄</td>
<td>25 mM</td>
</tr>
<tr>
<td>NH₄Cl</td>
<td>50 mM</td>
</tr>
<tr>
<td>Na₂SO₄</td>
<td>5 mM</td>
</tr>
<tr>
<td>MgSO₄</td>
<td>2 mM</td>
</tr>
<tr>
<td>0.2x metals</td>
<td></td>
</tr>
<tr>
<td>Glycerol</td>
<td>0.5% (54 mM)</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.05% (2.8 mM)</td>
</tr>
<tr>
<td>α-Lactose</td>
<td>0.2% (5.6 mM)</td>
</tr>
</tbody>
</table>

Table 4 - Recipes for 50x5052, ZY and 50xM

<table>
<thead>
<tr>
<th>50x5052 (100 mL)</th>
<th>ZY (1000 mL)</th>
<th>50xM (100 mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 g glycerol</td>
<td>10 g tryptone</td>
<td>80 mL H₂O</td>
</tr>
<tr>
<td>73 ml H₂O</td>
<td>5 g yeast extract</td>
<td>17.75 g Na₂HPO₄</td>
</tr>
<tr>
<td>2.5 g glucose</td>
<td>1 L H₂O</td>
<td>17.0 g KH₂PO₄</td>
</tr>
<tr>
<td>10 g α-lactose monohydrate</td>
<td></td>
<td>13.4 g NH₄Cl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.55 g Na₂SO₄</td>
</tr>
</tbody>
</table>

Table 5 - Components of the 1000x Trace Metals Mix (100mL)

<table>
<thead>
<tr>
<th>Volume</th>
<th>Concentration</th>
<th>Metal</th>
<th>MW</th>
<th>1 x Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 mL</td>
<td>0.1 M</td>
<td>FeCl₃ in ~0.12 M HCl</td>
<td>270.30</td>
<td>50 µM Fe</td>
</tr>
<tr>
<td>2 mL</td>
<td>1.0 M</td>
<td>CaCl₂</td>
<td>110.99</td>
<td>20 µM Ca</td>
</tr>
<tr>
<td>1 mL</td>
<td>1.0 M</td>
<td>MnCl₂·4H₂O</td>
<td>197.91</td>
<td>10 µM Mn</td>
</tr>
<tr>
<td>1 mL</td>
<td>1.0 M</td>
<td>ZnSO₄·7H₂O</td>
<td>287.56</td>
<td>10 µM Zn</td>
</tr>
<tr>
<td>1 mL</td>
<td>0.2 M</td>
<td>CoCl₂·6H₂O</td>
<td>237.95</td>
<td>2 µM Co</td>
</tr>
<tr>
<td>2 mL</td>
<td>0.1 M</td>
<td>CuCl₂·2H₂O</td>
<td>170.49</td>
<td>2 µM Cu</td>
</tr>
<tr>
<td>1 mL</td>
<td>0.2 M</td>
<td>NiCl₂·6H₂O</td>
<td>237.72</td>
<td>2 µM Ni</td>
</tr>
<tr>
<td>2 mL</td>
<td>0.1 M</td>
<td>Na₂MoO₇·2H₂O</td>
<td>241.98</td>
<td>2 µM Mo</td>
</tr>
<tr>
<td>2 mL</td>
<td>0.1 M</td>
<td>Na₂SeO₅·5H₂O</td>
<td>263.03</td>
<td>2 µM Se</td>
</tr>
<tr>
<td>2 mL</td>
<td>0.1 M</td>
<td>H₃BO₃</td>
<td>61.83</td>
<td>2 µM B</td>
</tr>
</tbody>
</table>

Note: 0.1 M FeCl₃ in ~0.12 M HCl should sterile filtered because autoclaving produces a precipitate. Once metals are combined, 36 mL H₂O is added.
2.2.3 Methods for purification of *E. coli* AnmK

2.2.3.1 General cell lysis, ultracentrifugation, ammonium sulfate precipitation, protease inhibitor, SDS-PAGE and desalting procedures

For the lysis of small volumes (up to 5 mL) of resuspended AnmK pellet, sonication was used. The cells were emulsified with a hand homogenizer then sonicated on ice in replicate pulses of 5 s with a 5 sec resting period at 20% amplitude for 10 minutes. For the lysis of larger volumes, the resuspended pellet was again emulsified with a hand homogenizer. This solution was transferred to an Emulsiflex C-5 (Avestin) high-pressure, liquid nitrogen-powered homogenizer for three cycles of cell lysis reaching a pressure range of 10,000 to 15,000 psi.

The resultant crude lysate from either sonication or high-pressure homogenization was then subjected to ultracentrifugation (Optima LE-80K, Beckman Coulter) at 40,000 rpm (~300,000 × g) for 40 minutes at 4 °C using either a Type 70 Ti or 45 Ti rotor. Pulverized ammonium sulfate ((NH₄)₂SO₄) was slowly added to the clarified lysate to obtain a final concentration 10% (w/v). The solution was stirred for 1 h at 4 °C and the pellet of precipitated protein was collected by centrifugation (3,000 × g; 10 min; 4 °C). The clear supernatant liquid was decanted and the pellet resuspended in the applicable buffer and carried forth into the first step of purification chromatography.

Generally, the first chromatographic step is Ni-NTA column chromatography. Following the Ni-NTA column, phenylmethylsulfonyl fluoride (PMSF) was added to the
5 mL fractions eluted from the FPLC. A 5x solution was used to obtain a final concentration of 1 mM.

Then, denaturing SDS-PAGE was used to visualize the protein in the chosen fractions. The gel used for all AnmK SDS-PAGE characterizations was a Laemmli gel consisting of a stacking (15% acrylamide) and a running (12% acrylamide) layer. (See Appendix B for how the gels were mixed, poured and chromatographed.)

After the appropriate fractions were identified, the pooled protein fractions were concentrated to 3 mL using an Amicon® Centicon® 30,000 kD MWCO centrifugal filter unit (Millipore). This volume was applied to either a pre-equilibrated PD-10 disposable benchtop (GE Healthcare) or an Econo-Pac® 10DG (BioRad) desalting column and a desalted 4 mL of protein was eluted isocratically.

2.2.3.3 Glutathione-S-transferase (GST) column chromatography

The cleared lysate was suspended in GST buffer A (PBS pH 7.3: 140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄ and 1.8 mM KH₂PO₄). The lysate was recirculated over a GSTPrep™ fast flow 16/10 glutathione Sepharose™ column (GE Healthcare) for 6 hours before being chromatographed on the Äkta FPLC. The column was washed with 20 mL GST buffer A and then the protein was eluted with 200 mL of 100% GST buffer B (50 mM Tris pH 8.0 and 10 mM reduced glutathione). Fractions containing protein were characterized by SDS-PAGE (12% Laemmli gel) and the AnmK-containing fractions were identified, pooled and desalted.
2.2.3.4 Nickel immobilized metal affinity chromatography (Ni-IMAC)

The cleared lysate was suspended in either Ni-NTA phosphate buffer A1 (50 mM NaPO₄ pH 8.0, 300 mM NaCl and 10 mM imidazole) or Tris buffer A1 (50 mM Tris pH 8.0, 300 mM NaCl and 10 mM imidazole) dependent upon the buffer system used for purification. The buffered lysate was then applied to a pre-equilibrated nickel-nitriloacetic acid (Ni-NTA) Superflow Sepharose® resin-packed column (Qiagen). The column was subjected to an initial wash of 30 mL in Buffer A1 and then a 70 mL linear gradient raised the concentration of buffer B (50 mM NaPO₄ pH 8.0, 300 mM NaCl and 750 mM imidazole) to 12%. The column was then washed with 80 mL buffer A2 (50 mM NaPO₄ pH 8.0, 300 mM NaCl, 100 mM imidazole and 0.05% Triton X-100) and another linear gradient increased the buffer B concentration to 25% (~200 mM imidazole). Buffer B was held at 25% for 40 mL and a final linear gradient increased the buffer B concentration to 75% over 150 mL. SDS-PAGE (12% Laemmli gel) was used to visualize the appropriate fractions from the purification and the AnmK-containing fractions were pooled and desalted.

2.2.3.5 Ion exchange chromatography (IEX)

Pooled desalted fractions from the IMAC column in 25 mM Tris pH 8.0 were applied to an IEX column. The Vivapure IEX starter kit was used and the strong anion, Q column was employed according the manufacturer’s protocol. For the larger scale Äkta purification, a Q-Sepharose® fast flow column was employed that was washed.
with of 25 mM Tris pH 8.0 (200 mL) before a 300 mL gradient up to 25 mM Tris pH 8.0 and 500 mM NaCl began. SDS-PAGE (12% Laemml gel) was used to identify AnmK-containing fractions that were subsequently desalted.

2.2.3.6 Phenyl Sepharose™

Partially purified AnmK was suspended in phenyl Sepharose™ buffer A (50 mM NaPO₄ pH 8.0, 1.0 M (NH₄)₂SO₄ and loaded onto a pre-equilibrated fast flow phenyl Sepharose™ resin-packed column. A 300 mL gradient was executed to end with 100% phenyl Sepharose™ buffer B (50 mM NaPO₄ pH 8.0) and was held until the baseline conductivity stabilized.

2.2.3.7 Hydroxyapatite (HA) hydrophobic interaction column

The HA-Ultrogel® hydroxyapatite column was packed according to the supplier’s protocol (Sigma). Partially purified AnmK was suspended in HA buffer A (10 mM sodium phosphate, pH 6.8) and loaded onto a pre-equilibrated HA column. The column was washed with HA buffer A (75 mL) and then a linear gradient to HA buffer B (0.4 M sodium phosphate, pH 6.8) was executed over 225 mL.

2.2.3.8 ATP-Sepharose

Dr. Timothy Haystead at in the Duke University Department of Pharmacology generously provided ATP-Sepharose™ resin, which was incubated in ATP buffer A (25 mM Tris, pH 8.0, 300 mM NaCl, 60 mM MgCl₂) with desalted AnmK and then packed into a disposable benchtop, gravity-flow column. Flow through was collected and then
a step gradient was executed to reach 100% ATP buffer B (25 mM Tris, pH 8.0, 300 mM NaCl, 60 mM MgCl₂, 100 mM ATP). SDS-PAGE was used to locate the fractions containing AnmK.

2.2.3.9 S-200 Size exclusion chromatography

A HiPrep 26/60 Sephacryl S-200 HR column (GE Healthcare) was equilibrated in 25 mM Tris, pH 8 and 250 mM NaCl. A 2 mL aliquot of AnmK was applied to the column flowing at a rate of 0.3 mL/min. The sample was eluted using 1 column volume of buffer and the column was re-equilibrated with 1 column volume of buffer or until the A₂₈₀ was stable.

2.2.3.10 TEV protease cleavage

Removal of the affinity tag on overexpressed recombinant AnmK was achieved by incubation with TEV in Ni-NTA buffer A1. This protease was obtained in three different ways. Louis Metzger from the Raetz Lab in the Duke University Department of Biochemistry graciously donated aliquots of 1 mg/mL TEV. He also supplied a freezer stock of the His₆-TEV fusion protein expression system that we used to obtain TEV. TEV is also commercially available and was purchased as both AcTEV™ (Invitrogen) and ProTEV (Promega). All of these TEV constructs were applied in the amount instructed by applicable protocol to a desalted solution of AnmK supplemented with EDTA (final concentration 1 mM). This solution was incubated at 4 °C for 16 hours (typically overnight) to achieve full affinity tag cleavage.
2.2.3.11 Purification of AnmK following TEV cleavage

The solution of AnmK and TEV was applied to a pre-equilibrated (buffer A1) benchtop Ni-NTA gravity flow column. The column was washed extensively with buffer A1 (at least 8 column volumes) and then a step gradient was applied to elute any uncleaved protein. This was achieved using 2 column volumes of 25% buffer B, 2 column volumes of 50% buffer B, 1 column volume of 75% buffer B and 1 column volume of 100% buffer B. SDS-PAGE analysis was performed and the cleaved AnmK was pooled, concentrated and typically buffer exchanged into crystallization buffer (10 mM Tris pH 8, 50 mM NaCl).

2.2.4 The final, optimized E. coli AnmK purification scheme

E. coli AnmK was expressed as a hexahistidine fusion protein in E. coli strain BL21(DE3). The cells were cultured in ZYM-5052 auto-induction media at 37°C for 18 hours. The cells were then harvested by centrifugation (SLA-3000 Rotor, Sorvall; 5,000 rpm/2500 × g; 10 min; 4 °C). The harvested cell pellet was resuspended in Ni-NTA buffer A1 (50 mM Tris pH 8.0, 300 mM NaCl and 10 mM imidazole) and mechanically lysed using the Emulsiflex C-5 high pressure liquid nitrogen-powered cell homogenizer. The lysate was clarified by ultracentrifugation (40,000 rpm/300,000 × g, 40 min, 4 °C; Type 70 Ti Beckman rotor) and then 10% (w/v) (NH₄)₂SO₄ was slowly combined with the supernatant at 4 °C with stirring.
After 1 hour the precipitated protein pellet was by obtained by centrifugation (3,000 \times g; 10 min; 4 °C), the supernatant liquid was decanted and the pellet was resuspended in Ni-NTA buffer A1. This protein was loaded onto a pre-equilibrated Ni-NTA Sepharose™ affinity column and chromatographed on an Äkta FPLC machine using an optimized protocol that included a wash step buffer A2 (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole and 0.05% Triton X-100). Aliquots of 5 mM PMSF (final concentration 1 mM) were added to the protein fractions as they eluted into the FPLC tubes. The appropriate fractions were pooled, concentrated to 3 mL and desalted on a PD-10 benchtop column. The eluted 4 mL were combined with a 1 mL aliquot of 1 mg/mL His<sub>6</sub>-TEV protease and supplemented with EDTA (final concentration 1 mM). The cleavage reaction proceeded overnight (16 h) and was subjected to crude Ni-NTA chromatography with a pre-equilibrated benchtop column. Purified AnmK eluted in the buffer A1 wash fractions and was verified by SDS-PAGE. Relevant fractions were combined, concentrated and buffer exchanged into crystallization buffer (50 mM NaCl, 10 mM Tris, pH 8).

### 2.2.5 Methods for size exclusion chromatography, analytical ultracentrifugation, MALDI/MS and LC-ESI-MS of E. coli AnmK

#### 2.2.5.1 Sephadex™ G-75

A Sephadex™ G-75 column (GE Healthcare) was equilibrated in 50 mM Tris, pH 7.5 and 250 mM NaCl. A 500 μL aliquot of AnmK (10 mg/mL) was applied to the column flowing at a rate of 0.5 mL/min. The sample was eluted using 1 column volume
of buffer and the column was re-equilibrated with 1 column volume of buffer or until the A$_{280}$ achieved baseline levels to indicate that there is no protein in the eluent.

### 2.2.5.2 Sephacryl S-100 HR

A HiPrep 16/60 Sephacryl S-100 HR column (GE Healthcare) was equilibrated in 25 mM Tris, pH 8 and 250 mM NaCl. Molecular weights were determined through comparison to a low molecular weight calibration kit (GE Healthcare). A 250 µL aliquot of AnmK was applied to the column flowing at a rate of 0.3 mL/min. The sample was eluted using 1 column volume of buffer and the column was re-equilibrated with 1 column volume of buffer or until the A$_{280}$ was stable.

### 2.2.5.3 High resolution S-200 column

A HiLoad 10/300 Superdex S-200 column was pre-equilibrated with gel filtration buffer (20 mM Tris, pH 7, 150 mM NaCl and 0.1 % NaN$_3$). Following a calibration run with blue dextran, aldolase (4 mg/mL; 158 kD) and ribonuclease A (3 mg/mL; 13.7 kD), a 100-µL aliquot of AnmK was injected onto the column flowing at a rate of 0.3 mL/min. The sample was eluted using 1 column volume of buffer and the column was re-equilibrated with 1 column volume of buffer or until the A$_{280}$ was stable.

### 2.2.5.4 Sedimentation equilibrium analytical ultracentrifugation (AUC)

The molecular weights of both His$_6$-tagged AnmK and AnmK after His$_6$ tag cleavage were determined by analytical ultracentrifugation in crystallization buffer (10 mM Tris, pH 8 and 50 mM NaCl) and Ni-NTA buffer A1 (25 mM Tris, pH 8, 10 mM
imidazole and 300 mM NaCl, respectively. These studies were carried out at 25 °C on a Beckman Model XL-A using sedimentation equilibrium at 9,000 rpm for 24 hours. In addition to an undiluted aliquot, 1:2 and 1:4 dilutions of each sample were injected into the three wells of an ultracentrifugation cell. The cells were scanned to confirm that the A280 was in the acceptable target range. A scan was then collected between 200 and 300 nm to determine the optimum wavelength for data monitoring. Readings were set to take place at 0, 6, 12, 18 and 24 h with 5 replicates at each data collection time.

2.2.5.5 Calculations to discern molecular weight from AUC data

The molecular weight (MW) was calculated after curve fitting adjustment using the ideal I program as supplied by Beckman with a partial specific volume (\( \bar{\nu} \)) of 0.7314 for the protein and a density (\( \rho \)) of 1.006 g/mL for crystallization buffer and 1.012 g/mL for Ni-NTA buffer A1. In equations 2.1 and 2.2, R is the gas constant, T is the temperature, \( \omega \) is the rotor speed in radians per second, A is the absorbance and r is the radius.

\[
MW = \frac{2RT}{\omega^2 (1 - \bar{\nu} \rho)} \frac{d \ln A}{dr^2}
\]  

(2.1)

A graph of ln(A) versus \( r^2 \) enables the use of equation 2.2.

\[
Slope = \frac{MW (\omega^2)(1 - \bar{\nu} \rho)}{2RT}
\]  

(2.2)

These parameters (\( \bar{\nu}, \rho \)) were determined using the SEDNTERP program [94].
Additionally, the data can be used to extrapolate an approximate calculation of the \( K_D \) for dissociation of the AnmK monomers from a simple calculation of \( K_A \) as follows:

\[
2A \rightleftharpoons A_2 \quad \text{and} \quad K_A = \frac{[A_2]}{[A]^2} = \frac{\epsilon}{\frac{\epsilon}{\epsilon_2}} = \left( \frac{\epsilon}{\epsilon_2} \right) \frac{\epsilon}{\epsilon_2} \quad \text{where} \quad \epsilon_{292} \approx 20,000.
\] (2.3)

### 2.2.5.6 Matrix-assisted laser desorption/ionization mass spectrometry (MALDI/MS)

For both the native AnmK and the SeMet-labeled protein, the samples were absorbed onto a sinapinic acid matrix using isopropanol:water:formic acid (2:3:1) as the solvent. MALDI/MS analysis was carried out on utilizing at Applied Biosystems DE-Pro instrument (Applied Biosystems, Foster City, CA). The percent incorporation of selenium can be calculated by the following equation (2.4).

\[
\% \text{ Se incorporation} = \frac{\text{Mass of SeMet AnmK} - \text{Mass of AnmK}}{(\text{Mass of Se} - \text{Mass of S}) \times \text{Met residues in AnmK}} \times 100
\] (2.4)

### 2.2.5.7 Liquid chromatography electrospray ionization mass spectrometry (LC-ESI-MS)

AnmK and the SeMet-labeled protein samples were submitted to the Duke University Mass Spectrometry facility. LC/MS analysis was carried out on an Agilent LC/MSD Trap instrument (Agilent Technologies, Santa Clara, CA). The percent incorporation of selenium can be calculated using equation 2.4.
2.2.6 Methods of the determination of AnmK kinetic measurements

2.2.6.1 Design of the AnmK activity assay

Silica gel with UV-254 indicator and PEI cellulose TLC plates were used to obtain the optimal conditions for ATP and ADP separation and a guanidine concentration range of 0.1 to 2.0 M was tested. The optimized guanidine concentration (1.1 M for ADP \( R_f = 0.6 \) and ATP \( R_f = 0.35 \)) was obtained for use on the PEI cellulose plates used in the AnmK activity assay. This concentration afforded complete baseline separation and allowed for a facile discontinuous radioactive assay. The radioactive assay was preferable for this application because of the high sensitivity of the phosphor screen in concert with the low limit of detection of the phosphorimaging system that allowed for the detection of \( [\alpha^{-32}\text{P}]\text{ATP} \) at concentrations greater than 100 pM.

The data obtained from the phosphoimages was quantified by comparison to standards of known specific activity located on the TLC plate and converted to concentrations. These data were then fit to the following equations for pseudo-first order kinetics using GraFit (v. 9, Erithacus Software) with an enzyme concentration of 10 nM.

\[
rate = \frac{V_{\text{max}}[S]}{K_M + [S]} \quad (2.5)
\]

\[
k_{\text{cat}} = \frac{V_{\text{max}}}{[E]} \quad (2.6)
\]
2.2.6.2 AnmK activity assay with anhGlc (levoglucosan)

To assay the activity of AnmK, a modified protocol from Uehara, *et al.* [57], was used. Aliquots (4 µL) of purified enzyme were incubated at 37 °C for 45 min with 8 µL of a mixture containing 10 mM MgCl₂, 10 mM ATP with 0.2% v/v 11.8 MBq [α-³²P]ATP, 10 mM CAPS Buffer (pH 10.5), and varying concentrations of levoglucosan (from 0 to 10 mM). After the 45 min incubation, samples were inactivated by incubation at 98 °C for 3 min and loaded onto PEI-cellulose for thin-layer chromatography and the product was eluted with 1.1 M guanidine HCl. The plate was air-dried after separation, then wrapped in plastic Saran Wrap and exposed to a storage phosphor screen (GE Healthcare) overnight (14 - 16 hours). A Typhoon 9400 Phosphorimager (GE Healthcare) was used to capture the phosphoimages and quantification was completed using ImageJ (ImageJ v.1.29, NIH). This technique analyzed the formation of [α-³²P]ADP relative to that of [α-³²P]ATP and was quantified based on the known amount of [α-³²P]ATP (0.2% v/v 11.8 MBq) added to the known concentration of non-radioactive ATP (10 mM). Kinetic parameters (K_M and k_cat) were determined by fitting to equations 2.5 and 2.6 using GraFit (v. 7, Erithacus).

2.2.6.3 AnmK activity assay anhGlcNAc, GlcNAc and anhMurNAc

The assay was performed and the data was processed as above (section 2.2.6.2), but using 1.0 M formic acid and 0.7 M LiCl as the mobile phase with a shorter, 15 minute incubation period for the assay.
2.2.6.4 $K_i$ and $IC_{50}$ Determination for ADP and ANP

The assay was performed and the data was processed as above (section 2.2.6.3), but with a constant 10 mM anhGlcNAc and incremental concentrations of ADP (50 mM, 30 mM, 10 mM, 1 mM, 500 µM, 100 µM, 50 µM, 10 µM, 1 µM, 0.5 µM)

2.2.6.5 pH optimum of AnmK

The assay was performed and the data was processed as above (section 2.2.6.2) with a constant concentration of anhGlcNAc (10 mM), but using a pH range (4.5, 6.0, 7.0, 8.0, 9.0 and 10.5) of buffers.

2.2.6.6 $K_m$ determination for ATP and the analog, ANP

The assay was performed as above (section 2.2.6.3), but with constant anhGlcNAc (10 mM) and varying concentrations of ATP and ANP (0 to 10 mM).

2.2.6.7 Metal ion dependence

The assay was performed as above (section 2.2.6.3) without varying the carbohydrate concentration. AnhGlcNAc was used at a concentration of 10 mM. Instead, various mono and divalent cations (MgCl₂, MgSO₄, ZnSO₄, CaCl₂, MnSO₄ and NaCl) were used (10 mM) both with and without the presence of magnesium (10 mM).
2.3 Results and Discussion

2.3.1 E. coli anmK cloning

The anmK PCR reaction with E. coli genomic DNA successfully amplified the full-length anmK gene, which was 1,100 base pairs (bp) in size. This PCR product was visualized and purified by gel electrophoresis (Figure 9) and extracted from the agarose gel (1%). Both entry clone constructs (anmK/pENTR and anmK/TEV/pENTR) were generated using a 5'-CACC-3' Kozak sequence [95] to start to the forward primer, which is necessary for the Gateway® system and enables directional cloning [96].

In addition to verification by sequence analysis, the presence of anmK insert was verified using Taq PCR amplification followed by PvuI restriction digest analysis. Standard M13 primers annealed to the pENTR™/D-TOPO® vector regions flanking the multiple cloning site and were used for PCR. The PvuI restriction endonuclease was chosen because it cut once in both the vector and the anmK insert; therefore, only vector with the insert would produce two fragments (1540 and 2180 bp) upon PvuI digest (Figure 10). DNA sequencing was used to verify sequence fidelity and directionality of the anmK insert.

Subsequent cloning produced five different anmK constructs: anmK/pDEST17, anmK/pDEST15, anmK/pDEST14, anmK/TEV/pDEST15 and anmK/TEV/pDEST17. Again, Taq PCR and restriction digest was utilized to confirm and verify the integrity of the anmK sequence in the pDEST™ vectors. When the PstI restriction enzyme was used (37
°C, 6 h) with anmK/TEV/pDEST17, the expected fragments (2162 and 3713 bp) were obtained (Figure 11).

Figure 9 - PCR Amplification of anmK Showing Appropriate Size (1,100 bp)
Gel electrophoresis in 1% agarose gel for 60 min at 80 V.
Figure 10 - *PvuI* Restriction Endonuclease Digestion of *annK/TEV/pENTR* Vector to Verify Gene Insertion.
Figure 11 - PstI Restriction Endonuclease Digestion of pDEST17/TEV \textit{anmK} Vector to Verify Gene Insertion.
The use of the Gateway® system for cloning allowed shuttling of the DNA insert between multiple vectors to rapidly prepare expression vectors to facilitate native, GST-labeled and hexahistidine tagged protein overexpression. Furthermore, the incorporation of the TEV cleavage site enables the removal of the affinity tags following protein purification with only three additional amino acid residues (SFT) on the protein N-terminus. Therefore, the recombinant protein most closely mimics the native protein, but still has the affinity tags necessary to expedite the protein purification process, which is discussed in detail in section 2.2.3 (Purification of *E. coli* AnmK).

### 2.3.2 AnmK expression

The three pDEST™ vectors (14, 15 and 17) containing *anmK* and TEV, when applicable, were expressed in BL21(DE3) cells. Aliquots were taken from the growth media at various stages of overexpression both before and after IPTG-induction. Figure 12 shows the sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of representative samples (lysed by sonication) for pDEST14, pDEST15 and pDEST17 expression at various induction times where 0 h indicates the uninduced growth media.

All three expression systems produced sufficient native AnmK protein (39 kD) by overexpression to be detected by SDS-PAGE (Figure 12). For pDEST15, the GST-label increases the molecular weight of AnmK by 27.7 kD and for pDEST17, the hexahistidine-tag added 2.6 kD. For the native protein, there was a significant band at approximately 30 kD, which is likely a protein degradation product. The GST-labeled pDEST15
expression vector exhibited the most abundant protein overexpression at 37 °C by SDS-PAGE analysis. However, both \textit{anmK/TEV/pDEST15} and \textit{anmK/TEV/pDEST17} were carried onto purification trials to be investigated in parallel effort.

To facilitate later \textit{de novo} phasing x-ray diffraction experiments, AnmK was expressed to include selenomethionine (SeMet). Several methods are available for the incorporation of SeMet into recombinant proteins [97]. The first method involves the use of a methionine auxotrophic strain of \textit{E. coli} (B834) that cannot naturally synthesize methionine [98]. Although this strain can be grown in nutrient-rich LB media, the isolated protein yield is often poor and the growth of the strain is not as robust as a non-auxotrophic strain. Therefore, the approach chosen in this work was developed to use altered media instead of an altered bacterial strain. This methodology, using the Overnight Expression Autoinduction System II, relies on the bacterial deprivation of methionine and its replacement through supplementation with SeMet [93]. This method is advantageous because it allows for bacterial growth and expression in a more cost effective laboratory strain. AnmK was successfully overexpressed to incorporate SeMet in place of the methionine residues and its percent incorporation is discussed later (section 2.3.4).
Figure 12 - SDS-PAGE Characterization of Initial pDEST Vector Expression. Arrows (blue) indicate the band that corresponds to AnmK.
Because of the success and ease experienced with the autoinduction system for the production of SeMet-AnmK, this protocol was adopted for use with the overexpression of AnmK (Table 4 and Table 5). Autoinduction resulted in a greater quantity of protein as measured by UV absorbance (mAU) in comparison to IPTG-induction. Figure 13 shows a comparison of IPTG induced and autoinduced AnmK where the growth volume was halved for autoinduction, but the crude protein obtained after initial Ni-NTA purification was greater. Note that the peaks are slightly offset due to different nickel affinity column dimensions, but both columns contained similar resin volumes.
Figure 13 – A Representative FPLC UV Trace Comparing Auto- and IPTG-Induced AnmK After Ni-NTA Chromatography.
2.3.3 AnmK purification

The Park group [57] used an *E. coli* BL21(DE3)/pTanmK expression system that lacked affinity tags and carried the *anmK* gene under the control of the T7 promoter of pGEM-T. Their purification required the use of four different columns (DEAE-Sepacel, hydroxlapatite and two MonoQ). In order to simplify this process and improve upon the overall protein yield, a new purification scheme was devised to obtain AnmK purified to homogeneity. As seen in Figure 12, there were three destination vectors available for use and based on initial expression testing, the two to be investigated were *anmK*/TEV/pDEST15 and *anmK*/TEV/pDEST17. After initial lysis by either sonication or high-pressure homogenization, the *anmK*/TEV/pDEST15 protein product was applied to a glutathione Sepharose column and the *anmK*/TEV/pDEST17 protein product was applied to a Ni-NTA column.

Because the pDEST15 sample in Figure 12 appeared to contain the largest amount of soluble, overproduced protein, this construct was chosen for initial large-scale purification. However, upon glutathione column chromatography, the FPLC UV trace (Figure 14) exhibited a surprisingly small peak of AnmK protein at 15 minutes (60 mL) following the elution of initial non-specific binding proteins.

The SDS-PAGE characterization revealed that the eluted GST-tagged protein was not homogeneous and was present in a very low yield. Therefore, the chromatographic purification efforts were concentrated on Ni-NTA chromatography. An initial, simple,
linear gradient to 66.6% buffer B (300 mM NaCl, 500 mM imidazole and 100 mM Tris pH 8) revealed that AnmK bound the Ni-NTA resin with high affinity, eluting in 350 mM imidazole. Therefore, discrete wash steps at two concentrations (100 mM and 200 mM) were employed to clean superfluous, non-specifically bound proteins from the column prior to the linear AnmK elution gradient. This method was adopted for permanent use as the initial chromatographic step in AnmK purification.
Figure 14 - The UV Trace of Elution from the Glutathione Column
Despite rigorous initial Ni-NTA optimization, AnmK was still not isolable as a homogenous protein solution. Therefore, other methods were attempted in an effort to further purify AnmK. The first of these was IEX, which was scaled up for FPLC purification using Q Sepharose® resin after it was tested on a small scale. IEX revealed an initial fraction of protein eluted in the wash and then two peaks following later in the elution gradient. SDS-PAGE revealed AnmK in the second peaks, but in very low yield and with the same impurities seen after Ni-NTA.

The next two chromatography resins employed were both variations of hydrophobic interaction column (HIC) chromatography designed to capitalize on the inherent but differential hydrophobic nature of proteins. Unfortunately, the AnmK protein never eluted after being applied to the phenyl sepharose column except for a small amount that eluted in the flow through. The second HIC resin, hydroxyapatite, was also unsuccessful. In this instance, the problem was equal but opposite as approximately 95% of the AnmK protein flowed directly through the column. A small portion became bound to the column, but eluted very early in the gradient to afford insufficient purification. Therefore, HIC was abandoned for size exclusion chromatography.

An S-200 column was chosen for size exclusion chromatography and AnmK was applied, but there was not a resultant increase in isolated protein purity. Figure 15 depicts a summary SDS-PAGE analysis of protein samples post-Ni-NTA, post-Q-
Sepharose and post-S-200. This reveals essentially no difference in purity from the initial Ni-NTA column chromatography, and thus no progress in the effort toward the purification to of AnmK to homogeneity.
**Figure 15** - SDS-PAGE experiment with AnmK post-Ni-NTA (1), post-Q-Sepharose (2) and post-S-200 (3).
ATP-Sepharose resin, called KinaseBind™, is a γ-phosphate-linked ATP resin designed to bind proteins with affinity for ATP. The resin is also designed with a long hydrophilic spacer to separate the protein from the agarose beads in order to minimize non-specific hydrophobic interactions. Fortunately, the ATP-agarose resin is not sensitive to divalent metal cations, so it is suitable for use after Ni-NTA chromatography where nickel (Ni²⁺) may still be in solution. The resin is not sensitive to protease inhibitors, detergents or reducing agents; thus, it is also suitable for use with AnmK after desalting the pooled and concentrated fractions from Ni-NTA chromatography. Unfortunately, the utilization of this resin did not further improve upon the purity of AnmK. However, it did bind AnmK that was selectively eluted with ATP, which was a promising indication that AnmK was in the correctly folded.

With the rising popularity of affinity column chromatography, ammonium sulfate ((NH₄)₂SO₄) precipitation as a purification method may be becoming underutilized or forgotten [99]. In this work, (NH₄)₂SO₄ vastly improved the purity of AnmK to the point where a single Ni-NTA column was sufficient for purification to homogeneity. SDS-PAGE was used to visualize the results of an ammonium sulfate screen, which was designed to ascertain the appropriate concentration of (NH₄)₂SO₄ to isolate AnmK by altering its solubility to result in precipitation. Figure 16 shows the SDS-PAGE characterization of the pellet (Figure 16a) and the supernatant (Figure 16b) fractions from the (NH₄)₂SO₄ small-scale precipitation experiments.
AnmK was selectively precipitated by 10% (w/v) (NH₄)₂SO₄. However, the 20% (w/v) pellet fraction also contained AnmK with enhanced purity. Out of a range of 10 to 80% (NH₄)₂SO₄ cuts, the protein most cleanly precipitated at 10%, which is illustrated in Figure 17 as a comparison to the 20% (w/v) fraction following Ni-NTA chromatography. Hence, the 10% (w/v) (NH₄)₂SO₄ precipitation was added to the AnmK protein purification scheme prior to Ni-NTA chromatography and was likely the main source of ease in subsequently rapidly obtaining homogenous protein.

Following the attainment of this protein, a hexahistidine-tagged TEV protease was utilized to cleave the N-terminal tag from hexahistidine-tagged AnmK. This was achieved using commercially available protease stocks and proceeded for 12 hours at 4 °C without the introduction of degradation products and with complete cleavage of the hexahistidine tag.
Figure 16 - Comparative SDS-PAGE of Pellet (a) and Supernatant (b) in (NH₄)₂SO₄ Precipitation Experiments.
Figure 17 - SDS-PAGE Comparison of 10% and 20% (NH$_4$)$_2$SO$_4$ Cut Purity. Red arrow points out intervening band present in 20% but not 10% cut.
Ultracentrifugation, a Triton X-100 wash in Ni-NTA chromatography and the use of protease inhibitors also enabled improvement in AnmK purity. The clarification of the crude lysate was initially carried out by routine centrifugation (28,000 × g; 45 min; 4 °C), but when ultracentrifugation (300,000 × g, 40 minutes, 4 °C) was employed, the purity of the resultant AnmK was improved. Also, the addition of the nonionic Triton X-100 detergent (0.05%) as a part of the 100 mM imidazole column elution program removed a fraction of presumably hydrophobically bound protein impurities. Finally, post-Ni-NTA PMSF supplementation (1 mM) was used to decrease proteolytic breakdown products of AnmK. Therefore, ultracentrifugation, Triton X-100 and PMSF were incorporated into the final, optimized purification scheme as well (section 2.2.4).

2.3.4 Biophysical characterization

With homogeneous protein by SDS-PAGE analysis, neither a G-75 preparative grade gel filtration column nor a Sephacryl S-100 column led to the calculation of AnmK and SeMet AnmK molecular weights. Therefore, a high-resolution Superdex S-200 column was used in the Duke University X-Ray Crystallography Shared Resource facility that resulted in an AnmK molecular weight of 100 kD. With the inconsistent results from gel filtration chromatography, a more exacting method of mass determination was utilized.

Dr. Harvey Sage, PhD, at the Duke University Analytical Ultracentrifugation Shared Resource assisted in the determination of the dimerization state of AnmK for
both hexahistidine and tag-free AnmK. Using the appropriate parameters for equation 2.1, AnmK was determined to be a dimer with a molecular weight of 78 kD following hexahistidine tag cleavage by the TEV protease. Interestingly, the sample with the hexahistidine tag still intact was a trimer in solution (MW = 119 kD). The hexahistidine tag is proximal to the putative ATP binding loop that had been predicted by homology modeling and that was later confirmed by inspection of the structure. Therefore, when left intact, the tag may interact with important regions of the protein to induce unusual polymers of AnmK. This foreshadowed the difficulty and ultimate failure of the hexahistidine-tagged AnmK to crystallize and also indicated that the hexahistidine-tagged AnmK should not be used in place of the affinity tag-free AnmK in kinetic experiments, as its dimerization state is different.

Mass spectrometry was the final technique used to biophysically characterize AnmK and SeMet AnmK. The purpose was to obtain a more accurate molecular weight of the AnmK monomer and to allow the calculation of the percent selenium incorporation in SeMet-AnmK. MALDI/MS revealed a crude experimental mass for AnmK (39,831.0 amu) and SeMet-AnmK (39,987.7 amu). Therefore, these data indicated that the SeMet percent incorporation was approximately 50%. Following this technique, LC-ESI-MS was also employed to obtain higher resolution masses of AnmK and SeMet-AnmK, which were 39,918.1 ± 0.72 and 40,342.9 ± 0.80, respectively. For these values, equation 2.4 revealed the following incorporation:
This is clearly the optimal selenium incorporation and is sufficient for anomalous scattering by the selenium atoms to determine the obligatory crystallographic phases.

### 2.3.5 AnmK preliminary kinetics

The radiometric assay for AnmK activity was designed to utilize PEI-cellulose TLC to separate ATP from ADP, the reaction product. Measurements used densitometric analysis of \([\alpha^{32}\text{P}]\text{ATP}\) from phosphor screens and could be performed with high sensitivity (pM). The data were adjusted for background noise and baseline ATP hydrolysis using control samples. Then, densitometric analysis was performed to quantify the phosphorimage. This was converted to concentrations by comparison with internal standards and fit to the Michaelis-Menten equations (2.5 and 2.6) using GraFit.

Using anhGlcNAc (10 mM), the kinase activity was confirmed to have a dependence on both ATP \((K_m\) value of 0.7 mM) and \(\text{Mg}^{2+}\). This work also demonstrated the activity of AnmK with \(\text{Mn}^{2+}\) in place of \(\text{Mg}^{2+}\). No other cations tested (zinc \((\text{ZnSO}_4)\), calcium \((\text{CaCl}_2)\) and sodium \((\text{NaCl})\)) displayed detectable enzyme turnover in the absence of \(\text{Mg}^{2+}\) over baseline ATP hydrolysis.

Using similar reaction conditions, AnmK was found to qualitatively exhibit little activity at acidic pH values (4.5 and 6.0) followed by increased activity at neutral and basic pH values (7.0 through pH 10.5) (Figure 18). However, it is not clear that at high pH values such as pH 10.5, activity measurement would give an accurate enzyme-
catalyzed rate. This pH value could cause a high enough concentration of hydroxide in solution to chemically catalyze any base-driven mechanism.

Further kinetic analysis of AnmK was also performed using ADP and the non-hydrolyzable ATP analog, ANP. Comparing the inhibitory properties of ADP and ANP at 10 mM ATP, the $K_i$ and $IC_{50}$ values of ADP were 1 mM and 15 mM, respectively, while ANP had a $K_i$ and $IC_{50}$ values of 6 mM and 90 mM. Additionally, using ADP and GlcNAc-6-P (not pictured), only ADP, not the phosphorylated sugar, strongly inhibits the enzyme (Figure 19).
Figure 18 - Radiometric Analysis of the AnmK pH Optimum. The assay was performed with 10 nM AnmK, 10 mM MgCl₂ 10 mM anhGlcNAc and 10 mM ATP and varying pH values as indicated.

![Radiometric Analysis of the AnmK pH Optimum](image)

Figure 19 - Radiometric Assay Showing AnmK Enzymatic Activity with AnhGlcNAc and Inhibition By 10 mM ADP. The lane contents are as follows: 1 & 2 – 100 mM ATP; 3 & 4 – 100 mM ATP and 10 nM AnmK; 5 & 6 – 10 nM AnmK and 10 mM anhGlcNAc; 7 & 8 - 10 nM AnmK, 100 mM ATP and 10 mM anhGlcNAc (Reaction); 9 & 10 – Reaction with 10mM ADP; 15-18 – [α-³²P]ATP standards (100 pM, 1 nM, 10 nM and 100 nM)

![Radiometric Assay Showing AnmK Enzymatic Activity with AnhGlcNAc](image)
To determine the apparent $K_M$ for levoglucosan, anhGlcNAc, anhMurNAc and GlcNAc, the carbohydrate substrate was combined with [$\alpha$-32P]ATP and the substrate concentration was varied for the reaction while holding the ATP concentration constant (10 mM). While AnmK had no detectable activity with GlcNAc, the enzyme did demonstrate weak utilization of levoglucosan with a $K_M$ of 4.6 mM, and $V_{max}$ and $k_{cat}$ values of 0.4 $\mu$M/min and 40 min$^{-1}$. $K_M$ values of anhMurNAc and anhGlcNAc were 0.33 mM and 1.4 mM, respectively. The $V_{max}$ and $k_{cat}$ values for anhMurNAc were 9.5 $\mu$M/s and 950 min$^{-1}$, while anhGlcNAc had $V_{max}$ and $k_{cat}$ values of 3.4 $\mu$M/s and 340 min$^{-1}$.

These steady state parameters (apparent $K_M$, $V_{max}$, $k_{cat}$ and $k_{cat}/K_M$) determined through densitometric analysis from the PEI-cellulose TLC assay (Figure 20) are summarized in Table 6 for AnmK with four carbohydrate substrates.

Because a TLC-based activity assay using [$\alpha$-32P]ATP only evaluates the ability to hydrolyze ATP, the hydrolysis of the carbohydrate substrate from anhGlcNAc to GlcNAc-6-P was confirmed through MS. Through MS, the correct hydrolyzed and phosphorylated product ([M-H] 299.9 m/z) peaks were identified.
Table 6 - Apparent Kinetic Values for the Carbohydrate Substrates

<table>
<thead>
<tr>
<th>Apparent Kinetic Parameters</th>
<th>AnhMurNAc</th>
<th>AnhGlcNAc</th>
<th>AnhGlc</th>
<th>GlcNAc</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{M}^{app}$</td>
<td>0.33 mM</td>
<td>1.4 mM</td>
<td>4.6 mM</td>
<td>ND*</td>
</tr>
<tr>
<td>$V_{max}$</td>
<td>0.0095 mM min$^{-1}$</td>
<td>0.0034 mM min$^{-1}$</td>
<td>0.0004 mM min$^{-1}$</td>
<td>ND*</td>
</tr>
<tr>
<td>$k_{cat}$</td>
<td>950 min$^{-1}$</td>
<td>340 min$^{-1}$</td>
<td>40 min$^{-1}$</td>
<td>ND*</td>
</tr>
<tr>
<td>$k_{cat}/K_{M}$</td>
<td>2900 mM$^{-1}$ min$^{-1}$</td>
<td>240 mM$^{-1}$ min$^{-1}$</td>
<td>8.7 mM$^{-1}$ min$^{-1}$</td>
<td>ND*</td>
</tr>
</tbody>
</table>

* ND - No Activity Detected
Figure 20 - Representative PEI-Cellulose TLC $K_M$ Determination Using AnmK (10 nM), ATP (100 mM) and varying concentrations of anhGlcNAc (100 $\mu$M – 10 mM). The lane contents are as follows: 1 & 2 – 100 mM ATP; 3 & 4 – 100 mM ATP and 10 nM AnmK; 5 & 6 – 100 mM, 10 nM AnmK and 10 mM anhGlcNAc (Rxn); 7 & 8 – Rxn with 5 mM anhGlcNAc; 9 & 10 – Rxn with 1 mM anhGlcNAc; 11 & 12 – Rxn with 500 $\mu$M anhGlcNAc; 13 & 14 – Rxn with 100 $\mu$M anhGlcNAc.
2.4 Conclusions

From *E. coli* genomic DNA, AnmK was cloned, expressed and purified to homogeneity in high yield. The cloning technique strategically integrated the ability to manipulate the *anmK* insert to incorporate multiple different N- or C- terminal affinity tags. An N-terminal hexahistidine fusion proved to be the optimal choice for AnmK expression and purification. The purification was not trivial, but was ultimately condensed into a short but effective protocol (section 2.2.4) that reduced a previous four step chromatographic process into two steps and yielded approximately 10 mg/L homogeneous AnmK on average. Furthermore, affinity tag removal is a prerequisite for either kinetic or crystallographic experiments because analytical ultracentrifugation revealed the presence of a trimer (119 kD) when the tag remained uncleaved from the protein N-terminus. Using TEV protease to remove the hexahistidine tag results in AnmK that has three superfluous residues on the N-terminus (SFT) and exists as a dimer (78 kD).

In preparation for x-ray crystallography, AnmK was expressed using auto-induction to incorporate selenomethionine for use in crystallographic phasing experiments. As confirmed by LC-ESI-MS, SeMet was incorporated in the place of Met residues with 100% efficiency. Because of the increased protein expression achieved using autoinduction for SeMet-labeling, this methodology was adopted for the
expression of AnmK. Homogenous AnmK and SeMet-AnmK were taken forth into the kinetic and crystallographic experiments that follow.

The purified *E. coli* AnmK protein was shown to be catalytically active using levoglucosan and synthetic anhMurNAc and anhGlcNAc. A kinetic assay was successfully designed that separated ATP and ADP on PEI-cellulose to afford the apparent $K_M$, $V_{max}$, $k_{cat}$ and $k_{cat}/K_M$ for four different carbohydrate substrates of AnmK and also for its other substrate, ATP. These studies revealed that the preferred substrate for AnmK was anhMurNAc, its natural substrate, with a $K_M$ of 0.4 mM. This is comparable to the 0.2 mM $K_M$ determined by Bacik *et al.* [100] with *P. aeruginosa* AnmK in a continuous, coupled enzyme assay, but less than the preliminary $K_M$ determined by Uehara *et al.* [57] using $[^3]$HanhMurNAc. As hypothesized, anhGlcNAc was a suitable pseudosubstrate with a $K_M$ of 1.5 mM. The experimentally determined pH optimum range of AnmK verified earlier findings by Uehara, *et al.* [57] that AnmK has little activity at pH 4.5 and 6.0 followed an increase in activity to pH 10.5; however, this apparent activity may be imparted by base-mediated catalysis from the hydroxide ions present in solution.

In summary, ATP and Mg$^{2+}$/Mn$^{2+}$ are requisite for catalysis, which occurs best under basic pH conditions and a product of this reaction, ADP, is an inhibitor of the reaction. AnhMurNAc and anhGlcNAc are the preferred substrates, but the hydrolysis and phosphorylation of anhGlc occurs to a lesser extent; furthermore, AnmK will not
utilize GlcNAc (or MurNAc [56, 57]) as a substrate. Thus, we began x-ray
crystallographic structural studies to examine the active site geometry in order to
determine how this enzyme might confer specificity to anhydro carbohydrates.
Chapter 3: Substrate Synthesis

3.1 Introduction

For the desired biochemical, biophysical and structural studies with AnmK, a suitable carbohydrate substrate for the enzyme is crucial. We determined (Table 6) that AnmK will not utilize GlcNAc to form the desired, GlcNAc-6-P product and that anhMurNAc isolation from a natural source is not efficient (40 μmol per 114 g of E. coli cell paste). Therefore, synthetic methodology was designed to generate the natural substrate for AnmK along with a simplified substrate to probe the molecular determinants of AnmK specificity – anhMurNAc and anhGlcNAc, respectively.

The distinctive carbohydrate scaffold of the anhydropyranoses has been examined by many carbohydrate chemists investigating the synthesis of both antibiotics like sibirosamine [101] and the production of 1,6-anhydromuramylpeptides [102]. The first anhydrohexoses were produced by pyrolysis and included 1,6-anhydroglucose (levoglucosan), for example, which can be produced by the pyrolysis of starch on a kilogram scale [103]. As the Cerný pyrolysis review in 1977 [103] notes, the yield for these types of reactions is often low (8%) and not always reproducible; therefore, extensive work has been done in the field to synthesize anhydropyranoses in a stereochemically controlled, high-yielding manner [104].

The basic strategy for the production of the anhydropyranose skeleton can follow one of five different routes. When the C2, C3 and C4 positions are adequately protected,
a leaving group at C1 can participate in cyclization under either acidic or basic conditions. The acidic route proceeds through an acid-catalyzed oxocarbenium ion intermediate, and conversely the basic route leads to the alkoxide attack on C1 and expulsion of the leaving group. If the reverse situation is generated and the leaving group lies on the C6 position, then 1,6-cyclization is straightforward under basic conditions.

Alternatively, an intramolecular Ferrier rearrangement (IFR) of D-glucal can lead to 1,6-anhydrohexopyranoses when both unprotected and 3,4-di-O-substituted. This process involves Lewis acids such as cupric sulfate or claysil, a compound especially developed for these reactions. Finally, the fifth route is the intramolecular halocyclization of glycals such as D-glucal. In 3,4-di-O-substituted compounds, iodocyclization can occur using iodonium di-sym-collidine perchlorate (IDCP) in moderate yield. Conversely, tin chemistry can be employed with unprotected D-glucal. The tin increases the nucleophilicity of the C6 oxygen atom to attack the cyclic iodonium ion formed in situ between C2 and C3.

Tin chemistry has been employed most frequently in the synthesis of anhMurNAc to date. Alternatively, the tosylation route most efficiently affords anhGlcNAc. The synthetic precedents that were published at the inception of this work along with the retrosynthetic analysis of the two desired compounds follow.
3.1.1 Current literature regarding target compounds

3.1.1.1 AnhMurNAc

In 1986, the Peters group at the University of Hamburg, Germany published the synthesis of anhMurNAc as an intermediate synthon for the production of 1,6-anhydromuramylpeptides [102]. Their synthetic design begins with 1,6-anhydro-2-azido-4-O-benzyl-2-deoxy-β-D-glucose (5) and proceeds through a five step synthesis (Scheme 2) to afford anhMurNAc (1).

Scheme 2 - Paulsen, Himpkamp and Peters Synthesis of AnhMurNAc (1986)

In 1992, Tailler et al. [105] published an improved synthesis of a compound similar to the starting material for the Paulsen et al. [102] synthetic scheme. This four-step procedure started with D-glucal (8) and afforded 1,6-anhydro-2-azido-3-O-benzyl-2-deoxy-D-glucose (12) (Scheme 3) in a stereochemically-controlled manner.

Scheme 3 - Tailler, Jacquinet, Noirot and Beau Synthesis of 1,6-Anhydro-2-Azido-4-O-Benzyl-2-Deoxy-D-Glucose (1992)
Then, in 2002, Kubasch and Schmidt published a more extended synthesis of anhMurNAc (Scheme 4) [106] where they synthesized the starting material (13) by the selective 3,6-O-TBS silanyl protection of D-glucal by Kinzy and Schmidt (Scheme 5) [107].

**Scheme 4 - Kubasch and Schmidt Synthesis of anhMurNAc (2002)**

![Scheme 4](image)

**Scheme 5 - Kinzy and Schmidt (1987)**

![Scheme 5](image)

With these syntheses in mind, we embarked upon the synthesis of anhMurNAc using an amalgamation of these literature precedents.
3.1.1.2 AnhGlcNAc

In 1989, the Fraser-Reid lab at Duke University revealed their syntheses of 1,6-anhydrohexopyranoses on a large-scale starting from inexpensive and readily available starting materials such as D-glucose and D-mannose. Starting with D-glucose does not yield the desired product, however, it establishes the framework of base-mediated hydroxylate attack to expel the tosylate leaving group in the formation of the 1,6-anhydroglucose ring (Scheme 6).

Scheme 6 - Zottola, Alonso, Vite and Fraser-Reid Synthesis of Levoglucosan (1989)

Later, in 1995, Thomson and von Itzstein [108] developed a route to the cyclization of N-acetylmannopyranose using the chemistry above and utilizing 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) as the base (Scheme 7). This intermediate was important for their synthesis of sialic acid compounds, which are important in various biological processes from immune response to cell recognition.

Scheme 7 - Thomson and von Itzstein Synthesis of anhManNAc (1995)
With the literature precedent that the tosylate cyclization reaction will occur with an acetamide group at the C2 position, these two methods can be combined to afford anhGlcNAc.

### 3.1.2 Retrosynthetic analysis of desired carbohydrates

#### 3.1.2.1 AnhMurNAc

AnhMurNAc is a moderately functionalized carbohydrate molecule and though it is only a C_{11} molecule, it is comprised of two rings and six stereocenters. The six-membered central ring poses the largest challenge with all five carbons being stereocenters. The strategy to employ in this case is to use a starting material with the maximum number of stereocenters set with the correct chirality. To derive the appropriate starting material, there are four logical disconnections in the retrosynthetic analysis of the product (Scheme 8).

**Scheme 8 - Retrosynthetic analysis of anhMurNAc**

The disconnections begin with the lactyl ether moiety that differentiates MurNAc from GlcNAc. This lactate group can be installed using the coupling of the selectively
deprotonated C3-hydroxyl of benzyl anhGlcNAc (24) with a lactyl triflate or 2-chloropropionic acid, which are both commercially available [106, 109, 110].

The second disconnect involves the functional group manipulation at C2. Ultimately, the acetamide can be readily produced from an azide, which would come from the iodide (14) via a 1,6:2,3-dianhydro intermediate [102, 105, 106, 111]. The third, and likely most challenging disconnect occurs between C1 and C6 to form the intermolecular acetal. This process changes the 4C1 chair conformation of the protected glucal (13) into the less favorable 1C1 conformation and requires the selective formation of the [3.2.1] bicyclic compound (14) [102, 103, 108, 111-113].

Finally, fourth disconnect, the installation of the benzyl protecting group at C4, allows the selectivity needed for the installation of the lactyl ether moiety at C3. This group must be established early in the synthetic scheme (24) to carry forward until the final steps. A careful 3,6 protection of D-glucal would allow for the selective protection of the C4 hydroxyl group [107]. The Zemplén deprotection of 3,4,6-tri-O-acetyl-D-glucal yields 8 [114, 115].

3.1.2.2 AnhGlcNAc

AnhGlcNAc only has one less functional group than anhMurNAc – the lactyl ether moiety. Thus, this C₈ carbohydrate molecule is also comprised of two rings, but has one less stereocenter (five). Again, the central ring poses the largest challenge with all five carbons being stereocenters; however, retrosynthetic analysis reveals that none of
the stereocenters are altered in the formation of the intermolecular acetal to make anhGlcNAc (Scheme 9). This is the one main disconnect in the analysis. Proceeding from the tosylated glucosamine derivative (22) to anhGlcNAc (21) would proceed through base catalyzed dehydration via the formation of an intermolecular acetal.

Finally, further simplification of the tosylate reveals that it requires a simple selective protection of $N$-acetylglucosamine with a tosylation reagent such as $p$-toluenesulfonic acid chloride (TsCl).

**Scheme 9 - Retrosynthetic analysis of anhGlcNAc**
3.2 Experimental Procedures

3.2.1 Materials and methods

Unless otherwise noted, reactions were performed at ambient temperature (typically 20-25 °C) in oven-dried glassware under an argon or nitrogen atmosphere using dry, ACS grade solvents. Tosyl chloride, t-butyldimethylsilyl-chloride and sodium hydride were purchased from Aldrich and stored in a desiccator until use. Both N-acetylglucosamine and 1,6-anhydro-2-deoxy-2-azido-4-O-benzylglucopyranose were received and stored at -20 °C. All other commercially obtained reagents were used as received. Thin-layer chromatography (TLC) was performed using Merck silica gel 60 F254 precoated plates (0.25 mm) and visualized by UV fluorescence quenching, H2SO4, PMA or p-anisaldehyde. Agela Technologies (Newark, DE) silica gel (particle size 40-60 µm) was used for flash chromatography. 1H and 13C NMR were recorded on a Varian Mercury 300 (at 300 MHz and 75 MHz, respectively) and are reported relative to Me4Si (δ 0.0) or the applicable solvent peak for CD3OD (δ 3.31 and 49.00) or CDCl3 (δ 7.26 and 77.16). Data for 1H and 13C NMR spectra are reported as follows: chemical shift (δ ppm) (multiplicity, coupling constant (Hz), integration). Multiplicities are reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, sept. = septet, m = multiplet, comp. m = complex multiplet, app. = apparent, bs = broad singlet.
3.2.2 Preparation of compounds

**D-glucal (8).** 3,4,6-Tri-O-acetyl-D-glucal (13; 3.32 g; 12.19 mmol) was dissolved in ACS grade (> 99% purity) MeOH (25 mL). K$_2$CO$_3$ (337 mg; 2.44 mmol) was added while stirring the solution [116]. The reaction was allowed to proceed for 1 h and subsequently analyzed by thin-layer chromatography (TLC) with CH$_2$Cl$_2$/MeOH (9:1) as the eluent. After one hour, the starting material disappeared with the concurrent appearance of the unprotected sugar (8). There was a small impurity detected at $R_f$ = 0.41 in relation to the desired compound ($R_f$ = 0.35). Water was added (10 mL) to the solution, which was then subjected to rotary evaporation under reduced pressure until a thick syrup remained. The compound was purified by flash chromatography with silica gel using CH$_2$Cl$_2$/MeOH (85:15) as an eluent, yielding 1.78 g of compound 8 as an oil. $^1$H NMR (300 MHz, CD$_3$OD) $\delta$ 6.345 ($J_{AB}$ = 0.02, d, 1H, H-1), 4.67 (m, 1H, H-3), 4.11 ($J$ = 0.024, 0.006, dt, 1H, H-4), 3.90 – 3.70 (m, 3H, H-2 & H-6), 3.56 ($J$ = 0.006, t, 1H, H-5). $^{13}$C (75 MHz, CD$_3$OD) $\delta$ 143.74 (C1), 103.34 (C2), 79.15 (C5), 69.72 (C4), 69.36 (C3), 61.03 (C6).

**D-glucal (8).** 3,4,6-Tri-O-acetyl-D-glucal (13; 2.72 g; 10 mmol) was dissolved in MeOH:H$_2$O:EtOH (10:10:1, 60 mL:60 mL:6 mL) and stirred at room temperature for 5 hours as prescribed by the literature [113]. The reaction was then triturated using cycles
of EtOH solvation and rotary evaporation. The reaction mixture was stored over P₂O₅ for 10 hours to desiccate. TLC was performed with CH₂Cl₂:MeOH (9:1) as the eluent and then flash chromatographed with CH₂Cl₂:MeOH (85:15) as the eluent. A pure product (8) was again isolated in a quantitative yield of 1.46 g and exhibited the same NMR signals as shown above.

\[ \text{3,6-di-O-t-butyldimethylsilyl-2-deoxy-D-glucal (18). Deprotected D-glucal (8; 2.29 g; 15.67 mmol) was mixed with imidazole (4.27 g; 62.68 mmol) and tert-butyldimethylsilyl chloride (3.5 g; 34.5 mmol, 2 eq.) in DMF (25 mL) and allowed to react for 8 h at rt. The reaction was run 8 hours at room temperature. The organic layer was washed with 10 times its volume of water. The water was back-extracted into ether twice. The solvents in the organic layer were rotary evaporated to dryness under reduced pressure yielding an oil containing a white crystalline solid. TLC using petroleum ether/diethyl ether (6:1) provided the best separation of products. The desired bi-substituted product 18 (Rᵣ = 0.55) was purified by flash chromatography using hexane/ethyl acetate (10:1) in 25% yield.} \]

\[ \text{¹H NMR (300 MHz, CD₃OD) δ 0.09 – 0.14 (m, 12H, SiCH₃), 0.88 – 0.94 (m, 18H, Si-C(CH₃)₃), 3.30 (m, 1H, H-4), 3.80 (m, 1H, H-6a), 3.91 (m, 1H, H-6b), 4.54 (d, 1H, H-5), 4.84 (d, 1H, H-3), 5.10 (dd, J = 5.7, 1.5 Hz, 1H, H-2), 6.28 (d, J = 0.019, 1H, H-1).} \]
1,6-Anhydro-2-deoxy-2-iodo-β-D-glucose (9). D-Glucal (8; 0.86 g, 5.9 mmol) and activated molecular sieves (3Å) were placed in a three-necked, round-bottomed 500 mL flask, equipped with a condenser and stir bar. The complete apparatus was sealed. The system was subjected to vacuum followed by argon sparging for 4 cycles, and dry acetonitrile (35 mL) was then cannulated into the system. The resultant mixture was stirred with mild warming (45 °C) until all of the starting material had dissolved, after which bis(tributyltin) oxide (2.4 mL, 4.7 mmol) was added by injection and the colorless solution was heated under reflux with an argon atmosphere for 3 hours. Over the course of the reaction, the solution became opaque. When the reaction was presumed to be complete via TLC monitoring (hexane/ethyl acetate (9:1)), the mixture was cooled to room temperature and the solvent was removed by rotary evaporation under reduced pressure. The residual oil was dried under high vacuum (50 mTorr). Dry dichloromethane (35 mL) was used to dissolve residual oil under an argon blanket and the solution was cooled to 0 °C in an ice/water bath. Powdered iodine (1.79 g, 7.0 mmol) was added in one aliquot to the dichloromethane solution, which caused an immediate change in the solution color from clear to dark brown. The mixture stirred at 0 °C for 15 min and was then filtered through Celite to remove both the tin salts and the molecular
sieves. The Celite was washed with dichloromethane (2 × 25 mL) and the resulting filtrate was concentrated by rotary evaporation to a final volume of 5 mL. This solution was diluted with hexane (50 mL) and stirred with a 1 M aqueous sodium thiosulfate solution (50 mL) for 12 hours. The slightly yellow solution partitioned between water and ethyl acetate was extracted with ethyl acetate (7 × 100 mL) and then dried with anhydrous sodium sulfate (Na₂SO₄). The solvent was removed to give a white, oily solid. The crude product was dissolved in ethyl acetate, mixed with silica powder and concentrated to dryness by rotary evaporation under reduced pressure in preparation for chromatography. Hexane/acetone (3:2) was used as the mobile phase for flash chromatography and the desired product was concentrated by rotary evaporation to yield the product (9) a white solid [117]. ¹H (300 MHz, CD₃OD) δ 4.19 (s, 1H), 3.01 (d, 1H), 2.66 (t, 2H), 2.39 (s, 1H), 2.16 (m, 2H), 0.66 (m, 1H); ¹³C (75 MHz, CD₃OD) 105.04, 77.79, 76.47, 73.88, 66.94, 28.66; ¹³C-APT (75 MHz, CD₃OD; where up is 1° or 3° and down is 2° or 4°) 105.04 (up), 77.79 (up), 76.47 (up), 73.88 (up), 66.95 (down), 28.66 (up).
1,6-Anhydro-2-azido-4-O-benzyl-2-deoxy-3-O-[(1R)-1-carboxyethyl]-β-D-glucose (6). 1,6-Anhydro-2-azido-4-O-benzyl-2-deoxy-β-D-glucose (15; 0.28 g, 1 mmol) was dissolved in anhydrous dioxane (25 mL). NaH (0.168 g, 4.2 mmol; 60% suspension in mineral oil) was added slowly while stirring. The mixture was held constant at 45 °C for 10 min, and then the solution was then allowed to cool to room temperature. (S)-2-Chloropropionic acid (2 mL, 2.5 g, 2.4 mmol) was added very slowly to avoid rapid formation of solid precipitate in solution. The stirring was maintained for 3 h at 90 °C and the organic solvent was removed by rotary evaporation under reduced pressure until dryness. To decompose any leftover sodium hydride, a 30 mL aliquot of water was added at once. The resultant aqueous solution was extracted once with n-hexanes/EtOAc (1:1; 20 mL) and then the aqueous layer was filtered over a bed of charcoal. The solution was cooled to 0 °C over an ice water bath. This solution was acidified by the addition of 2.0 M HCl dropwise until the solution reached pH 3 and the resulting solution was quickly extracted with CH₂Cl₂ (3 × 30 mL). These extracts were combined, washed with water, dried over anhydrous sodium sulfate (NaSO₄), and rotary evaporated until dryness to yield an oil. Flash chromatography was performed
on a silica column using CHCl₃/acetone (3:1) to isolate the desired product (6; 0.229 g, 70% yield). ¹H NMR (300 MHz, CDCl₃) δ 1.38 (d, 3H), 3.32 (s, 1H, H-4), 3.45 (s, 1H, H-2), 3.59 (s, 1H, H-3), 3.71 (t, J = 6.5 Hz, 1H, H-6a), 3.93 - 4.00 (m, 2H, H-6b, Lac-α-H), 4.66 and 4.73 (AB, J = 12.0 Hz, 2H, OC₆H₄Ph), 4.65 (s, 1H, H-5), 5.48 (s, 1H, H-1), 7.26 - 7.38 (m, 5H), 9.16 (br. s., 1H, COOH); ¹³C NMR (75 MHz, CDCl₃) δ 18.59, 60.54 (C-2), 65.84 (C-6), 71.75 (OCH₃Ph), 74.30 (C-5), 74.47 (Lac-α-C), 75.66 (C-4), 77.22 (C-3), 100.57 (C-1), 128.09, 128.32, 128.74, 137.16, 176.96 (C=O).

2-Amino-1,6-anhydro-4-O-benzyl-2-deoxy-3-O-[1(R)-1-carboxyethyl]-β-D-glucose (29). 1,6-Anhydro-2-azido-4-O-benzyl-2-deoxy-3-O-[1(R)-1-carboxyethyl]-β-D-glucopyranose (6; 0.72 g, 2.1 mmol) was dissolved in MeOH (10 mL) containing 5% Pd/C (0.72 g). This reaction mixture was stirred vigorously under a room temperature hydrogen atmosphere for 1 h. The reaction mixture was filtered through Celite that was then washed thoroughly with MeOH. The filtrate was evaporated to dryness using rotary evaporation to yield the final amine product (29, 91% yield) that was used without further purification. ¹H NMR (300 MHz, CDCl₃) δ 1.36 (d, J = 6.0 Hz, 3H, Lac-β-CH₃), 3.31 (1H, H-2), 3.46 (1H, H-3), 3.57 (1H, H-4), 3.67 (t, J = 6 Hz, 1H, H-6a), 3.96 (q, J = 7.5 Hz, 1H, H-6b), 4.42 (q, J = 6 Hz, 1H, Lac-α-H), 4.62 (1H, H-5), 4.64 and 4.68 (2d, J = 12...
Hz, 2H, OCH2Ph), 5.47 (s, 1H, H-1), 7.28 - 7.38 (m, 5H); \(^{13}\)C NMR (75 MHz, CDCl\(_3\)) \(\delta\) 19.6, 55.2 (C-2), 67.2 (C-6), 71.9 (OCH2Ph), 75.8 (2s, C-3 and C-5), 77.8 (Lac-\(\alpha\)-C), 80.9 (C-4), 99.5 (C-1), 128.0, 128.3, 128.8, 137.2, 179.9 (C=O). \(^{13}\)C NMR APT (75 MHz, CDCl\(_3\)); where down is 1° or 3° and up is 2° or 4° \(\delta\) 19.6 (up), 55.3 (down), 67.2 (up), 72.0 (up), 75.8 (down), 77.8 (down), 80.9 (down), 99.5 (down), 128.0 (up), 128.3 (up), 128.7 (up), 137.2, 179.9 (up)

2-Acetamido-1,6-anhydro-4-O-benzyl-2-deoxy-3-O-[(1R)-1-carboxyethyl]-\(\beta\)-D-glucose (27). 2-Amino-1,6-anhydro-4-O-benzyl-2-deoxy-3-O-[(1R)-1-carboxyethyl]-\(\beta\)-D-glucopyranose (29; 0.70 g, 2.2 mmol) was dissolved in CH\(_2\)Cl\(_2\) (10 mL). Pyridine (5 mL) and acetic anhydride (7 mL) were added and allowed to stir at room temperature for 24 h. This solution was rotary evaporated to dryness and the resultant residue was dissolved in CH\(_2\)Cl\(_2\). This solution was washed twice with water, which was back-extracted with 10 mL of CH\(_2\)Cl\(_2\). The organic layers were combined, dried over anhydrous magnesium sulfate (MgSO\(_4\)), filtered and concentrated to dryness by rotary evaporation. The crude product was purified by column chromatography to afford the desired compound (27; 0.50 g, 63% yield). \(^1\)H NMR (300 MHz, CDCl\(_3\)) \(\delta\) 1.39 (d, \(J = 8\) Hz, 3H, Lac-\(\beta\)-CH\(_3\)), 1.96 (s, 3H), 3.42 and 3.46 (2s, 2H, H-3 and H-4), 3.66 (s, 1H, H-3), 3.76 (s,
1H, H-6a), 4.11 (2, J = 9.0 Hz, 1H, H-2), 4.12 (d, J = 6 Hz, 1H, H-6b), 4.22 (q, J = 6.9 Hz, 1H, Lac-α-H), 4.58 (d, J = 6 Hz, 1H, H-5), 4.64 (2H, OCH₂Ph), 5.36 (s, 1H, H-1), 6.36 (d, J = 8.0 Hz, 1H, NH), 7.32 - 7.40 (m, 5H); ¹³C NMR (75 MHz, CDCl₃) δ 18.3, 23.4, 52.4 (d, C-2), 65.6 (C-6), 71.6 (OCH₂Ph), 74.2 (C-5), 74.3 (d, Lac-α-C), 75.8 (C-3), 77.0 (C-4), 101.0 (C-1), 126.9, 128.0, 128.4, 169.9 (C=O), 173.4 (C=O); ¹³C NMR APT (75 MHz, CDCl₃; where down is 1° or 3° and up is 2° or 4°) δ 18.2 (down), 23.4 (down), 52.3 (down), 65.6 (up), 71.6 (up), 74.27 (down), 74.33 (down), 75.8 (down), 77.1 (down), 101.0 (down), 126.9 (down), 128.0 (down), 128.4 (down), 128.8 (down), 169.9 (up), 173.4 (up).

2-Acetamido-1,6-anhydro-2-deoxy-3-O-[(1R)-1-(methoxycarbonyl)ethyl]-β-D-glucose (28). 2-Acetamido-1,6-anhydro-4-O-benzyl-2-deoxy-3-O-[(1R)-1-carboxyethyl]-β-D-glucopyranose (27; 0.25 g, 0.69 mmol) was dissolved in MeOH (10 mL) and the solution was stirred in the presence of 10% Pd/C under an atmosphere of hydrogen at 50 psi for 4 h. The mixture was filtered through a layer of Celite, the Celite was washed with MeOH and the organics were combined and concentrated by rotary evaporation to provide the title compound as a methyl lactyl ether (28; 0.19 g, 0.68 mmol). ¹H NMR (300 MHz, CD₃OD) δ 1.37 (d, J = 8 Hz, 3H, Lac-β-CH₃), 1.98 (s, 3H), 3.46 (s, 1H, H-3), 3.66 (m, 2H, H-4 and H-6a), 3.98 (s, 1H, H-2), 4.13 (d, J = 6 Hz, 1H, H-6b), 4.29 (q, J = 8 Hz, 1H,
Lac-\(\alpha\)-H), 4.49 (d, \(J = 8\) Hz, 1H, H-5), 5.25 (s, 1H, H-1); \(^{13}\)C NMR (75 MHz, CD\(_3\)OD) \(\delta\) 18.8, 22.6, 52.1 (C-2), 66.4 (C-6), 70.6 (C-4), 75.4 (Lac-\(\alpha\)-C), 77.7 (C-5), 80.4 (C-3), 102.1 (C-1), 172.5 (C=O), 175.0 (C=O).

\[
\text{2-Acetamido-1,6-anhydro-2-deoxy-3-O-[(1R)-1-carboxyethyl]-\(\beta\)-D-glucose (1).}
\]

The saponification of the methyl lactate (28) was carried out in dioxane/MeOH (1:1) by the addition of an aqueous KOH solution at room temperature. The reaction was stirred and was complete in 2 h. This alkaline solution was neutralized by eluting it through an acidic ion exchange resin (Dowex 50 W) after which it was filtered. The solvent was evaporated to dryness by rotary evaporation under reduced pressure and the reaction proceeded to yield anhMurNAc (1) in quantitative yield (0.187 g; a total of 95% yield for two steps). \(^1\)H NMR (300 MHz, CD\(_3\)OD) \(\delta\) 1.37 (d, \(J = 8\) Hz, 3H, Lac-\(\beta\)-CH\(_3\)), 1.98 (s, 3H), 3.46 (s, 1H, H-3), 3.66 (m, 2H, H-4 and H-6a), 3.98 (s, 1H, H-2), 4.13 (d, \(J = 6\) Hz, 1H, H-6b), 4.29 (q, \(J = 8\) Hz, 1H, Lac-\(\alpha\)-H), 4.49 (d, \(J = 8\) Hz, 1H, H-5), 5.25 (s, 1H, H-1); \(^{13}\)C NMR (75 MHz, CD\(_3\)OD) \(\delta\) 18.8, 22.6, 52.1 (C-2), 66.4 (C-6), 70.6 (C-4), 75.4 (Lac-\(\alpha\)-C), 77.7 (C-5), 80.4 (C-3), 102.1 (C-1), 172.5 (C=O), 175.0 (C=O).
2-Acetamido-2-deoxy-6-O-p-toluenesulfonyl-β-D-glucose (22). A solution of p-toluenesulfonyl chloride in pyridine was added dropwise over the course of 30 minutes to a stirred solution of 2-acetamido-2-deoxy-β-D-glucose (GlcNAc, 25) dissolved in 25 mL anhydrous pyridine in a 100 mL round-bottom flask over a 0 °C ice-water bath. The reaction was allowed to stir for 4 hours, and was then quenched with 5 mL of MeOH. The solution was concentrated by rotary evaporation under reduced pressure, resolubilized in dichloromethane and washed with water (20 mL), NaCO₃ (15 mL) and water (20 mL). The organic layer was concentrated to yield 22 (60% yield). The authenticity of the compound was validated by comparison to literature values [118].

2-Acetamido-1,6-anhydro-2-deoxy-β-D-glucose (21). 2-Acetamido-2-deoxy-6-O-p-toluenesulfonyl-β-D-glucose (22) was dissolved in a 100 mL round-bottom flask containing ethanol (20 mL) and a stir bar. DBU was added at room temperature and the stirring was maintained until the starting material has disappeared (approximately 24 h). This reaction mixture was concentrated by rotary evaporation, then purified by flash
chromatography over silica gel with ethyl acetate-methanol (10:1) as the eluent. The desired final product (21) was isolated in 36% overall yield. \textit{H} (300 MHz, CDCl$_3$) $\delta$ 2.01 (s, 3H, CH$_3$), 3.48 (1H, H3), 3.67 (d, 1H, H4), 3.76 (d, $J$ = 3 Hz, 1H, H6a), 3.97 (d, $J$ = 12 Hz, 1H, H2), 4.27 (d, $J$ = 6 Hz, 1H, H6b), 4.55 (d, $J$ = 6 Hz, 1H, H5), 5.30 (d, $J$ = 3 Hz, 1H, H1), 6.64 (d, $J$ = 9 Hz, 1H, NH; \textit{C} (75 MHz, CDCl$_3$) 23.26 (CH$_3$), 51.87 (C2), 65.50 (C6), 70.86 (C4), 72.36 (C3), 76.36 (C5), 101.06 (C1), 170.83 (C=O).

2-Acetamido-1,6-anhydro-2-deoxy-\textit{\textbeta}-D-glucose (21). 2-Acetamido-2-deoxy-\textit{\textbeta}-D-glucose (GlcNAc, 25) was dissolved in 50 mL anhydrous pyridine in a 250 mL round-bottom flask equipped with a thermometer and a magnetic stir bar over a 0 °C ice-water bath. A solution of \textit{p}-toluenesulfonyl chloride was added dropwise to maintain an internal reaction temperature that did not exceed 22 °C. Following the addition of the \textit{p}-toluenesulfonyl chloride, the reaction was allowed to stir for an additional 2 hours. The pH of the reaction mixture (the \textalpha{} and \textbeta{} anomers) was then raised to pH 9 using 3 M sodium hydroxide. This pH change was accompanied by color change of the solution to fuchsia. After stirring for 90 minutes at pH 9, the solution was lowered to pH 7 using 1 M HCl. The solvent was then removed using rotary evaporation under reduced pressure that allowed for the azeotropic removal of pyridine with toluene (3 × 100 mL).
The product was trititated with absolute EtOH and passed over a pad of Florisil that was washed with multiple aliquots of EtOH (5 × 50 mL). The combined filtrates were freed of solvent by rotary evaporation and the resultant oil was dried under high vacuum to yield anhGlcNAc (21; 1.12 g; 46% yield). The $^1$H and $^{13}$C NMR peak values are listed in the experimental section above.
3.3 Results and Discussion

3.3.1 Synthetic trials and the synthesis of anhMurNAc

3.3.1.1 Synthetic trials originating with D-glucal

Based on the retrosynthetic analysis and the known literature at the inception of this project, the following synthetic scheme was proposed for the synthesis of anhMurNAc (Scheme 10).

Scheme 10 - The Overall Proposed Synthetic Scheme for AnhMurNAc
Based on the retrosynthesis alone, the first three important products were D-glucal (8), 4-O-benzyl-D-glucal (13) and 1,6-anhydro-2-azido-2-deoxy-D-glucose (6). The following scheme illustrates the steps attempted to obtain these products.

**Scheme 11 - Completed Steps in the AnhMurNAc Synthesis Starting from D-Glucal**

To prepare D-glucal, peracetylated 3,4,6-tri-O-acetyl-D-glucal was deprotected according to the method of Zemplén and Kunz [114, 115] with potassium carbonate, which proceeded in quantitative yield. Simultaneously, the selective 3,6-O-TBS protection and the halocyclization reaction were carried out in 25 and 18% yield, respectively. Then, this method was abandoned in favor of a different route to anhMurNAc. Ultimately, the reason for altering the synthetic course was a combination of experimenter sensitization to the tin compounds, difficulty in complete removal of the tin from the halocyclized products and the publication of new work in the field by the Mobashery group at Notre Dame [111].
In 2009, Hesek et al. [111] published their work on the total synthesis of N-acetylglucosamine-1,6-anhydro-N-acetylmuramylpentapeptide (GlcNAc-anhMurNAc-pentapeptide), which has already been discussed in this work as the product of peptidoglycan catabolism by lytic transglycosylases. In this synthesis, anhMurNAc is an intermediate that was obtained by the following route (Scheme 12).


Their methodology was the simplest, most direct route to the final product and it avoided the unpleasant use of H₂S utilized by Kubasch and Schmidt [106] and the necessary saponification carried out by Paulsen, Himpkamp and Peters [102], and replaced the use of sodium azide with a milder azidotrimethylsilane (Me₃SiN₃). Fortuitously, one of the intermediate compounds in many of the aforementioned synthesis became commercially available at this time. With the benefits of a pure,
commercially acquired compound in hand, the revised synthetic scheme (Scheme 13) commenced as follows from azide 15.

### 3.3.1.2 Synthesis of anhMurNAc

2-Acetamido-1,6-anhydro-2-deoxy-3-O-[(1R)-1-carboxyethyl]-β-D-glucose (anhMurNAc, 1) was prepared from 1,6-anhydro-2-azido-4-benzyl-D-glucal (15) as follows in Scheme 13 to produce 12 mg of product.

**Scheme 13 - The Completed Synthesis of AnhMurNAc**

In the conversion of azide (6) to amine (29), sodium hydride was used to deprotonate the C3-hydroxyl, which attacked the (S)-2-chloropropionic acid to expel the chloride leaving group in an S\(_2\) mechanism. The (S)-2-chloropropionic acid required slow, careful addition as this reaction was prone to precipitation, but otherwise proceeded smoothly in 59% yield.
The first of two hydrogenations was carried out with 5% Pd/C in near quantitative yield. These mild conditions assured that the benzyl protecting group was left intact. The hydrogenation product (29) was directly used for acetylation through the acetate coupling using acetic anhydride in pyridine to from acetamide 27. This reaction proceeded in 63% yield. The second hydrogenation using more vigorous conditions employed the catalytic hydrogenator (maintained at 50 psi) for 4 h and used 10% Pd/C. This reaction was carried out in near quantitative yield (97%) and produced the methyl ester of anhMurNAc. This compound was saponified by stirring with 1 M KOH for one hour and then worked up using Dowex 50 W acidic ion exchange resin to yield the final product (1) in an overall 35% yield for five steps.

### 3.3.2 Synthesis of anhGlcNAc

In addition to anhMurNAc, the native AnmK substrate, 2-acetamido-1,6-anhydro-2-deoxy-β-D-glucose (anhGlcNAc, 21) was prepared. Because anhGlcNAc lacks the lactyl ether moiety that is characteristic of muramic acid but still provides the same skeletal structure and acetamide moiety as anhMurNAc, it was synthesized as a pseudosubstrate. Its covalent connectivity is highly similar to anhMurNAc, so we expected this compound to also be metabolized by AnmK. Based on literature precedent, retrosynthetic analysis and experience with anhMurNAc, this pseudosubstrate was significantly easier to prepare. Therefore, these reactions were
carried out on a larger scale and afforded ample substrate (1.12 g) for crystallization and kinetic experiments.

Both methodologies presented in the review of literature regarding anhGlcNAc were adapted for the synthesis of anhGlcNAc (21) in this work (Scheme 14).

**Scheme 14 - The Final Synthetic Routes for AnhGlcNAc**

The first route (Scheme 14A) uses the chemistry set forth by Thomson and von Itzstein [108] and adapted for N-acetylglucosamine, a C2 epimer of their substrate, N-acetylmannosamine. This reaction was comprised of the tosylation of the carbohydrate substrate (25) to yield 22 in 60% yield. Next, DBU catalyzed a deprotonated C1-hydroxyl attack to liberate the tosylate leaving group and form the desired intermolecular acetal in 36% yield to afford an overall yield of 22%. The second route (Scheme 14B), involved the formation of the tosylate *in situ* and the subsequent addition of sodium hydroxide to elevate the pH to 9. This catalyzed the same intermolecular acetal formation and proceeded in an overall 77% yield (1.12 g).
3.4 Conclusions

Using established literature procedures, both anhMurNAc (1) and anhGlcNAc (21) were synthesized [101-103, 105-108, 111, 113-115, 119, 120]. Common to both these target molecules was the requisite formation of the intermolecular acetal. The 6-O-tosylation and base-assisted attack of the anomeric C1 carbon on C6 afforded anhGlcNAc in good yield (77%) and the reaction tolerated synthesis on a 1 g scale.

Significant progress was made in the initial steps of the synthesis of anhMurNAc from D-glucal. Ultimately, several difficulties in the ring-formation and subsequent work-up steps forced a re-evaluation of the retrosynthetic path. Coincidentally, this chronologically aligned with the publication of a revised synthetic route to obtain anhMurNAc and was also when a key intermediate became commercially available. The key ring formation was more difficult in the synthesis of anhMurNAc, but the obtained intermediate had the anhydrocarbohydrate skeleton intact. A five step synthesis afforded anhMurNAc in an overall yield of 35% to yield 12 mg.

Therefore, both desired compounds were obtained for use in the kinetic analysis and structural elucidation of AnmK.
Chapter 4: Structure of Unbound, Open Conformation of AnmK

4.1 Introduction

4.1.1 Structural Background of AnmK

At the inception of this work in 2008, there was no structural information available regarding AnmK. Because AnmK phosphorylates a carbohydrate it was hypothesized to be a member of the carbohydrate kinase family; however, bioinformatic investigation revealed poor CLUSTALW [121] sequence alignment with glucokinase and similar hexokinases. BLASTP [122] searches revealed that *E. coli* AnmK only aligned with other AnmK homologs and various uncharacterized and unknown proteins. Every BLASTP hit was a member of “uncharacterized protein family (UPF0075).” This family has no known function, but may be distantly related to HSP-70 metalloproteases. Hence, this work aims to eradicate this knowledge deficit by solving the crystal structure of *E. coli* AnmK. Structural elucidation will allow homology searches based on structure instead of sequence, which is often a better indication of an enzyme fold. More importantly, the crystal structure of AnmK will allow for a full characterization of the enzyme to aid in the design of mechanistic proposal for its dual hydrolysis and phosphorylation activities.
### 4.1.2 AnmK Phasing

Because a structure of AnmK would be novel, the phase problem must be addressed before the structural data can be rendered into models that help visualize the complicated inner workings of this enzyme. There are three main ways to obtain the phases for a novel structure. Molecular replacement (MR) is the most preferentially utilized method, however anomalous x-ray scattering and isomorphous replacement can be used as well. All three of these methods were attempted in order to obtain the structure for AnmK and ultimately, the structure was solved using single-wavelength anomalous dispersion (SAD) methods.

Anomalous dispersion employs the inherent anomalous scattering by the electron cloud around certain atoms (selenium in this case) that gets relatively large at or near the absorption edge. Data from these experiments help find the location of the scattering atom substructure of the molecule. With the rise in synchrotron flux and intensity, it is possible to obtain the necessary phases from one SAD dataset collected near the peak wavelength where the two anomalous scattering factors \((f'\text{ and } f'')\) are largely different. These factors play a role in the overall scattering of an atom \((f)\), which comes from the equation: \(f = f^\circ + f' + if''\) where \(f^\circ\) is the “normal” atomic scattering if the atoms are treated as free oscillators. \(f'\) and \(f''\) are the real and imaginary (respectively) terms to account for the fact that the atoms are, in fact, bound by their atomic orbitals and resonate around the absorption frequency of the electron shell. Therefore, one
dataset taken at the near peak of the Se K-edge will have the largest contribution to \( f \) by \( f' \) and \( f'' \) to allow for phasing. Moreover, this method reduces the crystal exposure to the beam, which, in turn, reduces radiation damage to the crystal in order to maximize data quality.

The downside is that these experiments cannot be performed using a home source x-ray generator. They have to be performed at a synchrotron where the wavelength of the beam is tunable to the desired wavelength for the K-edge.

4.2 Experimental Procedures

4.2.1 Crystallization of AnmK and SeMet AnmK

4.2.1.1 General methods

Screening deep well blocks were obtained from Hampton Research (Aliso Viejo, CA), Qiagen (Valencia, CA) and Emerald BioSystems (Bainbridge Island, WA). INTELLI-PLATE™ 96-3 screening trays were obtained from Art Robbins Instruments (Sunnyvale, CA). The Linbro 24-well optimization plates and siliconized glass cover slides (22 mm diameter and 0.22 mm thickness) were obtained from Hampton Research. All chemicals used in the Alchemist BirdFeeder™ solvent reservoirs were ACS grade and sterile filtered prior to use.

4.2.1.2 Screening trays

Using the Phoenix liquid handling robot (Art Robbins Instruments, CA), numerous INTELLI-PLATE™ 96-3 trays were assembled using various crystal screens.
Crystals of AnmK and SeMetAnmK were grown initially using sitting drop vapor diffusion. Sitting drop screening trays contained three different ratios of protein (P; 9 mg/mL) to well solution (W) as follows (in P:W): 0.2:0.4 µL, 0.3:0.3 µL and 0.4:0.2 µL (vol/vol) in three subwells, and 70 µL of W in the large reservoir well (Figure 21). For the monitoring of crystal growth, the Minstrel™ HT crystal inspection system was used and the schedule was designed to photograph each of the 288 subwells twice on day one, once every other day starting day two for one week and then once at the end of weeks three and four.

4.2.1.3 Optimization trays

Following initial screening hits, hanging drop optimization Linbro-type plates were set up using equal volumes of protein (7 mg/mL in 10 mM Tris, pH 8 and 50 mM NaCl) and precipitant solution were mixed and equilibrated at 20 °C against well solutions (0.5 or 1 mL). The well solutions were prepared either by hand or with the Alchemist liquid handler (Rigaku). The well solution used for the structure of the open conformation of AnmK contained 0.5 M (NH₄)₂SO₄, 85 mM Tris (pH 8.5) and 15% glycerol. For data collection, crystals were extracted from the drop with a 0.2 – 0.3 micron loop (MiTeGen) and plunged into liquid nitrogen. They were then either stored for subsequent synchrotron data collection or directly transferred onto the crystal goniometer head with cryocooling to 100 K.
The optimization trays were also monitored using the Minstrel system and its companion CrystalTrak™ software, which has the ability to take photographs of the protein drop at different depths and choose the appropriate image through the detection of crystal edges. Most crystal growth occurred between day one and week two with variability due to the inherent differences in the crystal growth conditions.

4.2.1.4 Heavy atom soaks

Several optimized AnmK crystals were subjected to various soaks from the Heavy Atom Screen (Hampton Research) in an effort to isomorphously incorporate heavy atoms. For these trials, two methods were attempted. First, a crystal was looped and transferred to a new mother liquor enriched with a heavy atom solution in a 1:1 ratio of well to heavy atom solution. Secondly, 1 µL of heavy atom solution was added to the original hanging drop.

Additionally, a solution of 5-amino-2,4,6-triiodoisophthalic acid (I3C) was soaked into several AnmK crystals. I3C contains three covalently bound iodine atoms arranged as if forming the corners of equilateral triangle with 6 Å sides. For these experiments, stock solutions for soaks were obtained by dissolving I3C in water and adding double the equimolar amount of sodium hydroxide solution to fully deprotonate the carboxyl groups. Therefore, a 0.2 M sodium I3C solution was prepared and used at a final concentration of 10 mM.
Figure 21 – INTELLI-PLATE™ and Common Setup with Recipe Volumes
(Left) The Art Robbins INTELLI-PLATE™ crystallization screening plate. Rows A and B, and columns 1 and 2 are shown, but the plate has 8 rows and 12 columns total. (Right) Ratios in the subwells are given in protein (P) to well solution (W) in microliters.

Magnification of one well

2:4  3:3  4:2

70 µL well solution
4.2.2 Data collection

4.2.2.1 General methods

For data collection, the crystals were looped on an appropriately sized (generally 0.1-0.2 or 0.2-0.3 μm) Mounted Cryoloop™ (Hampton Research) that was secured to a CrystalCap™ Magnetic ALS style cap (Hampton Research) using nail enamel or ethyl-2-cyanoacrylate. When mounting the loop to the cap, the MicroTube was snapped at the second notch from the bottom as most data collection geometries prefer or require a 22 mm length between the base of the cap and the beam.

Once a crystal was mounted in the beam, test images were collected to determine the appropriate data collection strategy to minimize the number of images taken to maximize data completeness and redundancy. Generally, a step size of 0.5° and a sweep width for at least 90° was used for the open conformation of AnmK to ensure near 100% completeness of data overall and in the highest resolution shell.

4.2.2.2 In house (Cu Kα)

Home source (Cu Kα) data were collected at the Duke University X-Ray Crystallography Shared Resource under the direction of Dr. Nathan Nicely, PhD. X-ray diffraction data were collected in house using the Rigaku MicroMax™-007 HF microfocus X-ray generator and a sled-mounted RAXIS IV++ image plate detector. Data were indexed and scaled to 2.3-Å-resolution using the d*TREK software suite (Table 7).
4.2.2.3 Anomalous dispersion (BNL NSLS Beam X4C)

Anomalous diffraction data were collected at the National Synchrotron Light Source, Brookhaven National Laboratory (NSLS, BNL) on beam line X4C. This beam line has a double Si(111) crystal monochromator, a MAR 345mm CCD Detector and an Oxford cryojet. Data was at 100K collected for one peak (12666.465 eV) and one inflection point (12663.136 eV) as recommended by the pre-collection fluorescence scan and as suited for multi-wavelength anomalous dispersion (MAD). Data were indexed and scaled to 2.9-Å-resolution using HKL3000 (Table 8).

4.2.2.4 High resolution synchrotron (BNL NSLS Beam X29)

High resolution x-ray diffraction data (λ = 1.0809) were collected at the NSLS, BNL on beam line X29A. This beam line features an ADSC Quantum 315 detector and the detector is controlled by the Protein Crystallography Research Resource (PXRR) CBASS program for macromolecular crystallography. Data were indexed and scaled to 1.9-Å-resolution using the HKL2000 (Table 8).

4.2.3 De novo structure solution

The structure was determined with a single 2.90 Å SAD dataset from the peak wavelength (12.667 keV) using SHELX [123], SHARP [124] and ARP/wARP [125-127]. Model completion and refinement were performed with COOT [128, 129] and PHENIX (phenix.refine) [130]. Refinement statistics are summarized in Table 8. The final model includes two monomers comprised of residues 1–101, 113–369, the residues from the
expression tag (SFT, -2–0) and 103 water molecules in the asymmetric unit (ASU). No electron density was observed for residues 102–112.

4.2.4 High resolution (1.9 Å) structure

The high resolution structure was determined through the use of the structure solved from the SAD dataset as the search model for MR. The MR trials were carried out using PHENIX [130], then the initial model was constructed through manual rebuilding using COOT [128, 129]. Model completion and iterative structure refinement were performed with COOT and PHENIX (phenix.refine) [130]. This suite was also used for Cartesian coordinate-based simulated annealing and the incorporation of TLS refinement [131, 132].

Refinement statistics are summarized in Table 8. The final model includes two monomers comprised of residues 1–101, 113–369, 14 formic acid molecules (FMT), the residues from the expression tag (SFT, -2–0) and 290 water molecules in the asymmetric unit (ASU). No electron density was observed for residues 102–112.

4.2.5 Validation and deposition

Analysis of the stereochemical quality of the model was accomplished using AutoDepInputTool (ADIT) (http://deposit.pdb.org/adit/) and MolProbity [133]. Figures were prepared with PyMOL [134]. Atomic coordinates and experimental structure factors have been deposited in the Protein Data Bank (PDB) and are accessible under the codes 3SF9 for the high resolution structure.
4.3 Results and Discussion of de novo AnmK Structure

4.3.1 Crystallization of AnmK and SeMet AnmK

Crystals of AnmK were successfully prepared for both AnmK and SeMet AnmK using a combination of robotic screening followed by manually prepared optimization trays. Initial crystals grew within a day and reached full size (50 × 250 × 250 µm) in one week (Figure 22). The crystal morphology pictured in Figure 23e and informally referred to as the “football” was typical of AnmK, but AnmK exhibited several different crystal morphologies that are illustrated in Figure 23.

Fortuitously, many AnmK crystals grew in conditions that already offered some degree of cryoprotection – glycerol, ethylene glycol or small molecular weight polyethylene glycol (PEG). Therefore, they could be looped directly from the screening trays and immediately flash frozen in liquid nitrogen for initial data collection experiments. The conditions that yielded well-diffracting crystals (Figure 23e and f) were replicated and then optimized for better diffraction. Crystals that were not obvious single crystals did not diffract beyond 8 Å, if at all (Figure 23a, b, c, d and g).
After 6 hours  After 1 day  After 5 days

Figure 22 - Typical Growth Pattern of AnmK Crystals in Screening Plates
Figure 23 - Pictures of Different AnmK Crystal Forms
Crystals in (e) and (f) diffracted and (f) was used for data collection. All other morphologies did not diffract well (beyond 8 Å) if at all.
4.3.2 Modeling and initial molecular replacement

When initially constructing a preliminary model of *E. coli* AnmK using sequence
and structural homology databases, the resultant models were based on PDB ID 3CQY
[135], which was also later utilized by Bacik *et al.* [136]. *E. coli* AnmK has 49% sequence
identity with this homolog and was threaded onto the *S. oneidensis* structure using
SWISS-MODEL [137-141] to produce a putative model of *E. coli* AnmK. This monomer
model, fully intact, was used with preliminary home source data (Table 7) as a search
model for two copies (as indicated by the Matthews coefficient) in MR by PHASER.

Molecular replacement is the most commonly utilized method to obtain
crystallographic phases and one estimate revealed that 67.5% of the structures deposited
in the RCSB PDB in 2006 have been solved this manner [142]. As the number of
available structures grows, so does the capability of utilizing MR because this
methodology utilizes a known structure or structures to obtain a solution. MR uses
these structures as models to rotate and/or translate in order to fit the structure within
the electron density of the experimental data.

Unfortunately, phasing by MR was not successful with the model obtained
through threading. This is not entirely surprising as the carbohydrate/hexose kinases
often exhibit closure and domain movement. This property limits the feasibility of
solving *E. coli* AnmK with an intact, static model of AnmK from another organism
through the use of molecular replacement. Therefore, home source phasing was the next method used to attempt a solution of the inherent phase problem for this novel protein.
<table>
<thead>
<tr>
<th><strong>Data collection</strong></th>
<th><strong>AnmK</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>DHVI</td>
</tr>
<tr>
<td>Space group</td>
<td>$P4_212$</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td></td>
</tr>
<tr>
<td>$a, b, c$ (Å)</td>
<td>80.81, 80.81, 218.6</td>
</tr>
<tr>
<td>$\alpha, \beta, \gamma$ (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Wavelength</td>
<td>1.5418</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>37.9 – 2.3 (2.40 – 2.30)</td>
</tr>
<tr>
<td>$R_{sym}$ or $R_{merge}$</td>
<td>5.2 (25.8)</td>
</tr>
<tr>
<td>$I / \sigma I$</td>
<td>43.8 (7.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.3 (94.5)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>4.3 (4.2)</td>
</tr>
</tbody>
</table>
4.3.3 Home source phasing

In addition to the initial molecular replacement, we used S-SAD data collection on the home source in an attempt to phase the data [143]. Sulfur-SAD (S-SAD) data collection is a relatively new methodology, but has been investigated by many groups and has been shown to yield phases strong enough to lead to a structure solution [143-145]. The wavelength of the Cu Kα beam at the Duke University facility is not as suitable for this method as the Cr Kα wavelength (2.23 Å) emitted by other home source x-ray generators. Nonetheless, this wavelength was still been exploitable, so a dataset was collected for use with this technique.

Unfortunately, because of two major factors the data was not utilizable. First, this technique requires highly redundant data. This is not unobtainable, but practically impossible on the home source because of the long collection time required for AnmK crystals. However, the symmetry group obtained for AnmK crystals from indexing indicated tetragonal symmetry; thus, only approximately 90° of data needed to be collected to yield a complete dataset making redundancy not altogether unattainable. The longest unit cell edge of the AnmK ASU is approximately 220 Å, which necessitates a greater crystal-to-detector distance than the minimum (90 mm). In fact, any crystal-to-detector distance less than 220 mm yielded images where the spots corresponding to the long unit cell edge were too close together to be differentiated by any of the data.
reduction programs (HKL2000, iMosflm and d*TREK). Thus, this data was discarded by the program, which led to a vastly less complete or redundant dataset than expected.

Secondly, high mosaicity undermines the utilization of S-SAD phasing and as expected by the rounded shape, the AnmK crystals were highly mosaic. This is unfortunately an inherent quality of crystal forms and rarely varies unless the morphology is altered. Therefore, a different phasing technique was necessary to solve the initial crystal structure of *E. coli* AnmK

### 4.3.4 de novo structure solution

Home source S-SAD phasing and the *S. oneidensis* structure as an initial search model for phasing by molecular replacement were abandoned for *de novo* phasing enabled by the introduction of SeMet into the recombinant AnmK protein. In fact, this project was chosen to be an experimental case at the RapiData 2009 Data Collection and Structure Solving practical course held in the NSLS at BNL. The *de novo* crystal structure was solved on site with a single-wavelength anomalous dispersion (SAD) dataset that diffracted to 2.9 Å. The SeMet AnmK protein crystallized in the *P*4*3*2*1* space group with a unit cell of 81 x 81 x 219 Å. The data had a reasonable correlation coefficient and figure of merit, and yielded a heavy atom substructure of 18 selenium atoms in the ASU. Figure 24 shows two of the eighteen heavy atom sites (residues Met193 and Met195) that are conveniently separated by only one residue (Leu194). This allowed for an initial starting point for model building and the backbone of AnmK was built into the electron
density map generated by the SHARP density-modified SAD solution. This initial model was then used for the solution of the high-resolution AnmK structure.
Figure 24 - Anomalous Difference Density for Mse 193 and Mse 195
The anomalous difference density is shown (purple) around the two selenium atoms in the selenomethionine residues of the MLM sequence in AnmK.
4.3.5 Other phasing tools

The initial heavy atom soaks never produced a crystal that was not cracked; however, the I3C soaks did not damage the crystals. This compound is referred to as the “magic triangle” for phasing because of the three iodine atoms forming a 6 Å equilateral triangle [146, 147] (Figure 25)

![Image of Magic Triangle](image)

Figure 25 - The Magic Triangle (I3C) for Phasing
The iodine atoms depicted in purple surrounded by a mesh that illustrates the anticipated electron density.

Multiple datasets were collected following overnight soaks with I3C, but the anomalous signal was not sufficient for phasing. Presumably, the compound was in solution, but did not consistently bind AnmK at the same position in each monomer to yield strong signal at that location.
4.3.6 High resolution data collection and refinement

The hand-built structure from the de novo SAD solution was then used as the model in ARP/wARP with the data from the high resolution (1.95 Å) native dataset. The final structure yielded reasonable crystallographic and geometric indicators after multiple rounds of refinement including simulated annealing and TLS refinement [131, 132]. Data collection, model, and refinement statistics are summarized in Table 8. The structure has appropriate stereochemical merits, with a final R-values of $R_{work} = 18.9\%$ and $R_{free} = 23.9\%$ with 10% of the data withheld into the $R_{free}$ test set. The estimated solvent content of the ASU is 46.07%. The Ramachandran plot, produced by MolProbity [133], shows that 98.88\% of the residues are in favored regions and 0.14\% are outliers.

Overall, the appearance of the AnmK monomer is a U-shape containing N- and C-terminal RNAse H-like folds (Figure 26) to comprise the dyad characteristic of carbohydrate kinases. Specifically, this fold is characteristic of the ASKHA (Acetate and Sugar Kinases/Heat shock cognate/Actin) superfamily of proteins. The RNase H-like fold contains the $\beta_3\beta_2\beta_1\alpha\beta_4\alpha\beta_5\alpha$ architecture where $\beta_3$ is antiparallel to the others. In both monomers of the homodimer, contiguous density was not observed for the loop region encompassing Trp102 to Thr 112 due to presumed mobility of this loop region between $\beta_4$ and $\alpha_3$. AnmK forms stable homodimers confirmed by analytical ultracentrifugation. Thus, there are two molecules in both the asymmetric and biological unit.
Additionally, the biological dimer has an extended β-sheet bridging the monomers that is formed at the interface of the additional β-strand from the C-termini of the monomers (Figure 27).
Figure 26 - Depiction of the RNase H-like Fold. Also referred to as ββαβαβ fold. Strand order is 5-4-1-2-3 and β2 is antiparallel to the rest.
Figure 27 - The Biological and Asymmetric Unit of AnmK. The helices are shown in blue, the beta sheets in purple and the loops in grey. The C2 symmetry axis lies down the middle of the illustration.
This dimerization interface comes from the antiparallel interaction between the β12 sheets (Figure 27) of two different AnmK molecules. The surface area buried in the dimer interface is extensive (15701 Å²). There is a reciprocal hydrogen bond between the backbones of the Ile364 residues that forms the middle of the extended β-sheet architecture spanning the dimer interface. The α3 loop interaction between the monomers also bridges the interface. The major dimer salt bridges and hydrogen bonds (where the asterisk denotes chain B of the dimer) are as follows: Asp132 – Asn*119, Arg136 – Pro*366, Arg135 – Asp*66, Thr359 – Asn*368.

The AnmK protein has approximate dimensions of 81 × 81 × 218 Å³ and is comprised 14 α helices, 12 β sheets and 4 3₁₀ helices arranged into two domains (N-terminal and C-terminal). The N-terminal domain is comprised of residues -2–160 and 321–369, and the C-terminal domain contains residues 161–320. The N-terminus is ordered and quickly leads into a series of β sheets flanked by α and 3₁₀ helices. After the fourth β sheet (β4) there is a missing loop region that was not built due to aforementioned lack of contiguous density. The structure resumes with another β sheet, which leads to an α-helical region between the domains. Then, there is a repeat of the RNase H-like fold (5 β sheets flanked by various α and 3₁₀ helices), with a large helical insertion after the first β sheet (β10). The final 3₁₀ helix (η4) allows a quick turn to direct α helix (α13) back across toward the N-terminal domain. The final segment of AnmK
forms a small β sheet that aligns with the 5 initial β sheets forming the αβα core of the N-terminal domain.
Figure 28 - The Overall Topology of the Open Conformation of AnmA
Table 8 - Data Collection and Refinement Statistics for *de novo* AnmK Structure

<table>
<thead>
<tr>
<th>Data collection</th>
<th>SeMet-AnmK</th>
<th>AnmK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>NSLS X3A</td>
<td>NSLS X29</td>
</tr>
<tr>
<td>Space group</td>
<td>$P4_3212$</td>
<td>$P4_3212$</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>80.74, 80.74, 218.3</td>
<td>80.81, 80.81, 218.6</td>
</tr>
<tr>
<td>$\alpha$, $\beta$, $\gamma$ (°)</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Wavelength</td>
<td>0.9798</td>
<td>1.08</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50 – 2.73 (2.83 – 2.73)</td>
<td>37.9 – 1.9 (1.97 – 1.90)</td>
</tr>
<tr>
<td>$R_{sym}$ or $R_{merge}$</td>
<td>6.1 (21.7)</td>
<td>9.0 (48.2)</td>
</tr>
<tr>
<td>$I/\sigma I$</td>
<td>30.4 (3.4)</td>
<td>25.4 (3.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.2 (97.3)</td>
<td>99.5 (100)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td></td>
<td>17.0 (11.1)</td>
</tr>
<tr>
<td>Wilson B-factor (Å$^2$)</td>
<td></td>
<td>37.28</td>
</tr>
</tbody>
</table>

**Refinement**

| Resolution (Å) | 37.90 – 1.90 |
| No. reflections | 108,446 |
| $R_{work}/R_{free}$ (%) | 18.88/23.89 |
| No. atoms | 5792 |
| Protein | 5528 |
| Water | 264 |

| Ramachandran plot analysis |          |      |
| Favored | 98.88% |      |
| Outliers | 0.14% |      |
| Clashscore | 10.86 |      |
| Rotamer outliers | 0.88% |      |
| R.m.s. deviations |          |      |
| Bond lengths (Å) | 0.007 |      |
| Bond angles (°) | 1.10 |      |

Data were collected from a single crystal. Values in parentheses are for the highest-resolution shell.
4.4 Conclusions

This work elucidated the open, substrate-free x-ray crystallographic structure of *E. coli* AnmK and allowed the verification (through the DALI server) of its inclusion in the ASKHA family of proteins. The final structure has a 2.6 Å RMSD from the *S. oneidensis* structure and a 2.7 Å RMSD from the recently published *P. aeruginosa* substrate-bound structure. The average B factor (46.5) for this model was slightly elevated, however, the predicted mobility of the enzyme likely contributes to this anomaly. This may also explain the round appearance of the crystals that ultimately lead to the structure solution.

The Mse193-Leu194-Mse195 sequence present in AnmK greatly facilitated the ease of building the *de novo* structure by hand. Much like the easily distinguishable triangle of I3C, the MLM sequence offered a fingerprint region that when identified, led to relatively straightforward model building. After the construction of the structural model, the AnmK structure revealed two, U-shaped monomers that made an elegant dimer interface through an extended β-sheet architecture. The likely structural family for AnmK was verified, but there was no substrate present in the active site.

Because AnmK exhibits such novel chemistry related to the dual functionality carried out an a unique anhydro carbohydrate substrate, we commenced further structural characterization in hopes of obtaining a substrate-bound form to elucidate the
molecular basis of specificity for the substrates and also to investigate the active site geometry for an end goal of formulating a putative mechanism of AnmK.
Chapter 5: Structure of anhMurNAc-ANP-AnmK Ternary Complex

5.1 Introduction

Prior to this work and the very recent report by Bacik et al. [100], there was no knowledge of how the substrates for AnmK bind in the active site or what changes, if any, occur upon substrate binding. It was previously reported and we confirmed that AnmK must have anhMurNAc present in order to hydrolyze ATP [57, 100]. Therefore, we sought to obtain the ternary complex to gather insight into substrate binding when all of the requisite substrates were present. Thus, this chapter describes the elucidation of the high resolution (1.95 Å) anhMurNAc-ANP-AnmK crystal structure by x-ray crystallography and describes the changes that occur upon substrate binding and domain closure. In this chapter, we describe the structure of the liganded AnmK. Then this structure is compared to the recently reported structures of an AnmK homolog from \textit{P. aeruginosa} that contains either a single substrate (anhMurNAc) or the product (ADP), but not ATP, making it catalytically inactive. Lastly, this structure can be used to gain insight into the mechanism of AnmK by examining the sugar binding site for potential hydrolysis in the crystallize enzyme. AnhMurNAc in the active site would offer support of a concomitant hydrolysis and phosphorylation mechanism.
5.2 Experimental Procedures

5.2.1 Crystallization trials and substrate binding

In order to find the appropriate conditions for co-crystal formation of anhMurNAc, a non-hydrolyzable analog and AnmK, numerous crystallization trials were conducted using methods described in section 4.2.1. Various substrates (anhMurNAc, anhGlcNAc and anhGlc) and ATP-like molecules (ANP, ATPγS and ADP) were used. Prior to deposition into the crystallization drops, the protein (5 mg/mL) was preincubated for 15 minutes at 37 °C with 2-fold molar excess anhMurNAc and ANP in a solution containing 10 mM Tris pH 8 and 50 mM NaCl. In the screening trials, as before, the co-crystals of the complex with its substrates were grown using the sitting-drop vapor diffusion method in the INTELLI-PLATE™ trays. Initial hits were replicated and optimized using the hanging-drop vapor diffusion method with Linbro plates prepared with the Alchemist liquid handler. For the final AnmK anhMurNAc-ANP-complex, equal volumes of the AnmK-anhMurNAc-ANP mixture and precipitant solution containing 2.5 mM ammonium formate, 100 mM HEPES (pH 7.8), 10 mM MgCl₂ and a range of 5 to 15% (w/v) glycerol were mixed and allowed to equilibrate at 20 °C. Crystals appeared within a day and reached maximum size in approximately ten days. More specifically, the round “football”-shaped crystals formed quickly and the bound crystals grew in over the course of three to seven days in the same hanging drop.
5.2.2 Data collection, refinement, validation and deposition

The crystals produced from the mother liquor containing 2.5 M ammonium formate, 100 mM HEPES (pH 7.8), 10 mM MgCl₂ and 9% (w/v) glycerol diffracted to the highest resolution (1.95 Å) as determined by home source screening of crystal conditions. This high-resolution x-ray diffraction data (λ = 1.0000 Å) was collected at the Southeast Regional Collaborative Access Team (SER-CAT) bending magnet beamline (22-BM) at the Advanced Photon Source (APS) at Argonne National Laboratory (ANL). This beam line features a MAR MOSAIC 225 mm CCD detector and an Oxford instruments cryojet to facilitate data collection at 100 K. The SER-CAT sergui and marccd programs control the robotics and data collection at SER-CAT. The data collection protocol collected 180° of with a step width of 0.5° at a detector distance of 190 mm. Data were indexed and scaled to 1.95-Å-resolution using d*TREK (Table 9)[148].

The closed conformation, dual substrate (anhMurNAc and ANP) bound crystal structure of AnmK was solved to 1.95 Å-resolution. The anhMurNAc-ANP-AnmK protein crystallized in the P2₁2₁2₁ space group with a unit cell of 113 x 167 x 195 Å. Refinement proceeded smoothly and the final structure yielded reasonable crystallographic and geometric indicators after multiple iterative rounds of refinement including simulated annealing and TLS refinement [13, 14]. Data collection, model, and refinement statistics are summarized in Table 9. The structure has appropriate stereochemical merits, with a final R-values of R_work = 19.0% and R_free = 22.7% with 5% of
the data withheld into the $R_{\text{free}}$ test set. The Ramachandran plot, produced by MolProbity [9], shows that 97.5% of the residues are in favored regions and there are no outliers.

The structure was solved by molecular replacement using PHASER [149] with the native, unbound structure of AnmK as a search model. However, this model was divided into the C-terminal (residues 161 – 320) and N-terminal (residues 1 – 160 and 321 – 369) domains. The initial model was manually rebuilt using COOT [128, 129]. Iterative structure refinement was carried out using PHENIX [130], specifically phenix.refine with initial noncrystallographic symmetry (NCS) restraints for the refinement. Ligand molecular topologies for ANP were obtained using the HIC-Up server [150]. The molecular coordinates and topology of anhMurNAc was obtained using REEL that is a part the PHENIX suite [130], and through personal communication with Nigel Moriatry, PhD, the creator of REEL. The final refined model of the AnmK-ANP-anhMurNAc complex exhibited $R_{\text{work}}$ and $R_{\text{free}}$ values of 19.0 and 22.7%, respectively. The structure was validated using MOLPROBITY [133].

Analysis of the stereochemical quality of the model was accomplished using AutoDepInputTool (ADIT) (http://deposit.pdb.org/adit/) and MolProbity. Figures were prepared with PyMOL [134]. Atomic coordinates and experimental structure factors have been deposited in the Protein Data Bank (PDB) and are accessible under the code 3SFA.
5.3 Results and Discussion for the Substrate-Bound Structure

5.3.1 Attaining a different crystal morphology

Rounded edges for a crystal is often associated with disorder in the crystal lattice, and this was reflected in the mosaicity of the de novo structure; therefore, we hypothesized that substrate binding would produce a more ordered crystal morphology. The use of a preincubation step prior to crystal tray setup afforded a different, sharper crystal morphology, which indicated that the disorder associated with domain and loop mobility in the open conformation was eliminated. Prior to this discovery, screening and optimization trays constructed without prior incubation yielded the “football”-shaped morphology. Following a suggestion from the Hampton Research website (http://hamptonresearch.com/tips.aspx), the enzyme was preincubated with the substrates (ATP analog, ANP [151, 152], and anhMurNAc) for 15 minutes at 37 °C before being deposited in the screening conditions. Indeed, preliminary screening at the home source verified that AnmK now crystallized in the P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1} space group as opposed to the P4\textsubscript{2}12\textsubscript{1}2 group of the open conformation. When both crystal morphologies were present in a single drop (Figure 29), data collected on each type revealed only substrate bound in one. Henceforth, crystals that indexed in a tetragonal Laue group were solely the open conformation and data that indexed into an orthorhombic space group were substrate-bound. Because of the rapid formation of crystals exhibited by the AnmK protein in its open conformation, it was possible that a pre-incubation was
necessary to ensure that both of the substrates bound in the active site prior to the nucleation of the crystals.
Figure 29 - Hanging Drop With Two Crystal Morphologies
This photograph depicts the crystallization conditions where crystals of both unbound AnmK (center and bottom left) and anhMurNAc-ANP-AnmK (top right) were present.
5.3.2 Phasing and refinement

With the de novo solution in hand, phases could be obtained for this high-resolution closed conformation using molecular replacement. However, the model had to be divided into two parts, the constant (N-terminal; residues 1-160 and 321-369) and the variable domain (C-terminal; residues 161-320), and searched for in that order. Multiple rounds of PHASER [149] were required to locate both domains within the electron density so that a full monomer could be constructed. Once this was accomplished, then the dimer unit was located, and this dimer was used as a larger search model for the four requisite dimer copies in the ASU (Figure 30).

The closed conformation, dual substrate (anhMurNAc and ANP) bound crystal structure of AnmK was solved to 1.95 Å-resolution. The anhMurNAc-ANP-AnmK protein crystallized in the P2_12_2_1 space group with a unit cell of 113 x 167 x 195 Å. Refinement proceeded smoothly and the final structure yielded reasonable crystallographic and geometric indicators after multiple iterative rounds of refinement including simulated annealing and TLS refinement [131, 132]. Data collection, model, and refinement statistics are summarized in Table 9. The structure has appropriate stereochemical merits, with a final R-values of R_{work} = 19.0% and R_{free} = 22.7% with 5% of the data withheld into the R_{free} test set. The Ramachandran plot, produced by MolProbity [133], shows that 97.5% of the residues are in favored regions and there are no outliers.
Figure 30 - The Asymmetric Unit of AnmK Divided into Four Biological Dimers
### Table 9 - Data Collection and Refinement Statistics for anhMurNAc-ANP-AnmK Structure (3SFA)

<table>
<thead>
<tr>
<th>Data collection</th>
<th>anhMurNAc-ANP-AnmK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location</strong></td>
<td>SER-CAT 22-BM</td>
</tr>
<tr>
<td><strong>Space group</strong></td>
<td>P2_1_2_1</td>
</tr>
<tr>
<td><strong>Unit cell dimensions</strong></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>113.3 168.2 194.0</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td><strong>Wavelength</strong></td>
<td>1.0000</td>
</tr>
<tr>
<td><strong>Resolution range (Å)</strong></td>
<td>32.49 – 1.94 (2.05 – 1.94)</td>
</tr>
<tr>
<td><strong>R_{sym} or R_{merge}</strong></td>
<td>11.2 (45.8)</td>
</tr>
<tr>
<td><strong>I / σI</strong></td>
<td>23.7 (1.9)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>99.92 (100)</td>
</tr>
<tr>
<td><strong>Multiplicity</strong></td>
<td>6.8 (6.4)</td>
</tr>
</tbody>
</table>

### Refinement

| Resolution (Å)                | 32.49 – 1.94        |
| **No. reflections**           | 270,247             |
| **R_{work}/R_{free} (%)**     | 19.01/22.69         |
| **No. atoms**                 | 23632               |
| Protein                       | 21550               |
| Solvent                       | 2082                |
| **Ramachandran plot analysis**|                     |
| Favored                       | 97.5%               |
| Outliers                      | 0.00%               |
| Clashscore                    | 16.11               |
| Rotamer outliers              | 0.70%               |
| **R.m.s. deviations**         |                     |
| Bond lengths (Å)              | 0.007               |
| Bond angles (°)               | 1.15                |

Data were collected from a single crystal. Values in parentheses are for the highest-resolution shell.
5.3.3 General structural notes

In addition to the change in space group from P4\textsuperscript{3}2\textsuperscript{1}2\textsuperscript{1}2\textsuperscript{1} to P2\textsuperscript{1}2\textsuperscript{1}2\textsuperscript{1} and the sharper edges of crystal morphology for this new structure, the loop (residues 102-112) that was unable to be built in the open conformation was ordered in anhMurNAc-ANP-AnmK. Therefore, the full set of molecular coordinates for all residues were determined. In the complete, contiguous structure, there are eight monomers comprised of residues 1–369, three HEPES solvent molecules (EPE), seven anhMurNAc molecules (AH0), seven ANP molecules, the residues from the expression tag (SFT, -2–0) and approximately 200 water molecules in the asymmetric unit (ASU). Figure 31 shows the full asymmetric unit of AnmK with each dimer as a different color. No contiguous electron density was observed for residues 160–320 for chain E of the model. However, this is not uncommon when there are numerous identical units in the ASU.

With this structure, because the previously missing loop was now present, an additional \(\beta\)-sheet was added to the overall molecular topology to afford a monomer model with 14 \(\alpha\)-helices (\(\alpha\)1-14), 13 \(\beta\)-sheets (\(\beta\)1-13) and 4 3\text{\_}10 helices (\(\eta\)1-4) as seen in Figure 32 with amino acid numbering. The same core RNase H-like fold (\(\beta\beta\alpha\beta\alpha\beta\alpha\) is present in both N and C-terminal domains.
Figure 31 - The Asymmetric Unit of AnmK Consisting of Eight Monomers
Figure 32 - The Complete Overall Topology Diagram of AnmK with Residue Numbering for Secondary Structural Elements.
5.3.4 Motifs within the structure

With both of the substrates bound, the binding motifs (PHOSPHATE1, PHOSPHATE2, ADENOSINE, CONNECT1 and CONNECT2) could be elucidated (Figure 33) and are named based on the Nicely et al. [91] and Bork et al. [153] publications regarding other ASKHA kinase family members. The PHOSPHATE1 motif is typically Gly-X-Asp-X-Gly, however, in AnmK it appears as Gly12-X-X-X-Asp16-Gly17 and is located at the loop between β1 and β2 (residues 11-18). The PHOSPHATE2 motif is simply the magnesium-coordinating Asn167 residue in the short β8-β9 (residues 168-171) loop, which is conserved in AnmK as part of a larger conserved residues NIGGI sequence from residues 167 – 171. In AnmK, however, Asp167 coordinates water (HOH 274 in chain A) instead of magnesium. The third 3_{10} helix (η3) contains the Gly295-Gly296-Gly297-X-X-X-Asn300 ADENOSINE motif, which along with Asp196, Lys209 and Asp210, coordinate the adenosine ring and sugar moiety of ATP. The construction of a morph model of the open conformation to the closed conformation revealed that the phosphate binding and ADENOSINE motifs are not in place for ATP- binding until the enzyme reaches its closed conformation. Figure 33 shows the PHOSPHATE1, PHOSPHATE2 and ADENOSINE binding motifs along with the electron density mesh surrounding the two substrates.
Figure 33 - Active Site of AnmK with Electron Density Around Substrates and Relevant Motifs Noted. Electron density is contoured at 1 σ.
There are also two other motifs (CONNECT1 and CONNECT2) that are conserved in the hinge region between the N- and C-terminal domains. The CONNECT1 motif involves the α5 helix, which is one of the two helices that cross at the base of the substrate binding cleft. There is no consensus sequence for this motif but there are generally multiple hydrophobic residues then a small residue and a polar residue, followed by another small residue. In the case of AnmK, this motif is L147-V148-X-X-A150-X-H152-H153-A154. The CONNECT2 motif involves the other helix of the cleft base cross (α13). It commences with a glycine (Gly326) that is the central point about which the two domains hinge. This motif is always a helix, but again, does not have a consensus sequence. It consists of largely hydrophobic residues with interspersed small residues. In AnmK, this helix spans residues 326 – 343 (GDDMEALAFAWLAWRTLA). Thus, based on the CONNECT motifs, residues A154 and G326 are the two residues central to the movement of the AnmK domains. The locations of these motifs are depicted in (Figure 34), which is a modified version of the AnmK topology diagram.
Figure 34 - Structural Motifs in AnmK.
The β-sheets in purple and α-helices in teal represent the core ββαβαβ fold. The two helices in green are the hinge region. All of the motifs are listed in red.
5.3.5 Substrate binding sites

5.3.5.1 AnhMurNAc binding site

The anhMurNAc binding site (Figure 35) lies close in proximity to the two helices that form the base of the enzyme cleft (α5 and α13). For anhMurNAc alignment in the active site, Arg134 hydrogen bonds the terminal carboxylate of the lactyl ether moiety off of C3, which is also coordinated by Leu147, Val148 and a water molecule. The C4-hydroxyl below the plane of the carbohydrate is hydrogen bonded to Asp187 and a water molecule, which is also interacting with C1. Additionally, there exists a hydrogen bond between Thr100 and the carbonyl of the acetamide group on C2 of anhMurNAc. The underside of the carbohydrate central ring is solvent accessible and the acetamide moiety projects out into solvent accessible surface area flanked by His97, Thr100, Gln103, Arg134, Ser240 and Glu244. Because of this, the water molecule coordinated between C1 and Asp187 likely enters the active site from just below the anhMurNAc molecule. All other backbone interactions are shown in Figure 35.

Taken together, these interactions in the anhMurNAc binding site indicate that both the lactyl ether and acetamide moieties are important molecular determinants of specificity. This corroborates the kinetic data as the change from anhMurNAc to anhGlcNAc (loss of lactly ether) and anhGlc (loss of lactyl ether and acetamide) is accompanied by a loss in catalytic efficiency of the enzyme with its substrate.
Figure 35 - Binding Site Interactions with AnhMurNAc
Bond lengths are shown in green (Å). Eyelash-like illustrations depict hydrophobic interactions with the backbone of various proximal amino acids.
5.3.5.2 ANP binding site

The ANP binding site is formed upon domain closure between the β1-β2 loop of the N-terminal domain and the β8-β9 loop of the C-terminal domain. Important residues are illustrated in Figure 36 and a thorough description follows that moves from the adenine ring to the phosphate terminus. With the domain closure, Asp196 moves into the appropriate proximity to hydrogen bond the 2'-OH and 3'-OH of ribose moiety of ANP. Also, Tyr207 aligns for the formation of pi stacking with the adenine moiety. Also on the adenine ring, the Asp210 side chain forms a hydrogen bond with the exocyclic nitrogen (N6) and the neighboring N1. Asn192 is hydrogen bonded to the 4'-OH of ribose, and there is a highly ordered water molecule coordinated between the ribose 2'-OH, the N3 of adenine, Nδ of Asn300, Oδ of Asn300 and another water molecule. The ribose 5'-OH and one alpha-phosphate oxygen is hydrogen bonded to Gly-296. A water molecule is highly coordinated to the same α-phosphate oxygen, a β-phosphate oxygen, a γ-phosphate oxygen, the Glu330 carboxylate, and Asn167. This highly coordinated water is where one might expect to see a magnesium ion, but the electron density map was not indicative of bound magnesium. However, magnesium not bound to the ATP analog has been previously noted in this family of enzymes [100]. The other α-phosphate oxygen is coordinated to two water molecules on the more solvent accessible side of the ATP binding pocket. Ser14 binds the oxygen bridging the α- and β-phosphates along with the other β-phosphate oxygen. The rest of the terminal
phosphate is coordinated by both the backbone and the alcohol moiety of Thr13, Glu330, Asn192, Ile171, a water molecule and the 6′-OH of anhMurNAc.

As seen in ASHKA kinases actin and human Hsp70, the N6 of ATP is oriented toward the solvent (Figure 37). This may not reflect the optimal conformation of the adenine ring as it does not allow for maximal hydrogen bonding interactions in the active site. However, this is a unique and indicative characteristic of kinases with the actin fold [154]. Also ATP is in an "extended" conformation [154], which allows the best interaction with the protein because it covers more surface area. The extended phosphate group also streamlines the molecule to fit into the narrow opening of the active site. The PHOSPHATE1 loop between β1 and β2 contains multiple glycine residues because any side chain would preclude the ATP from fitting into the narrow entrance to the binding site.
Figure 36 - Binding Site Interactions with ANP
Bond lengths are shown in green (Å). Eyelash-like illustrations depict interactions with the backbone of various proximal amino acid
Figure 37 - Surface and Cartoon View of the AnmK (purple) Monomer with ANP (pink) Bound in the Active Site Showing Solvent Accessibility
5.3.6 An interfacial zinc binding site

Upon examination of the $F_o - F_c$ difference density map, there was a large circular region with density unexplained by the current model. Examination of the residues surrounding this region revealed it contains a His-His-Asp triad comprised of Asp26 and His28 from one chain (H) and His257 of chain D with a water molecule as the fourth coordinating species (Figure 38). This is reminiscent of a catalytic zinc site [155], however the zinc binding site is not proximal to the active site to aid in catalysis.

Traditionally, “structural” zinc sites are tetrahedrally coordinated by four amino acid residues, but AnmK only has three ligands and the site is not between the monomers of the biological dimer to offer structural support [155]. Therefore, this zinc coordination is likely irrelevant to the mechanism and study of AnmK and is likely an artifact of the crystallization conditions as zinc is ubiquitous in buffers that are not specially prepared without metal ions. However, the added intermolecular coordination and the stability gained is likely a driving force for crystal nucleation when zinc is present in solution.
Figure 38 – The Interfacial Zinc Binding Site at the Interface of Chain H (green) and Chain D (purple)
5.4 Conclusions

This work revealed the structure of the ATP analog and substrate bound in a closed conformation, which mimics the components necessary for the hydrolysis and phosphorylation of anhMurNAc. The goal of this work was achieved as both substrates were simultaneously bound in the active site. This conformation of AnmK showed that there was an 11° closure upon substrate binding that was facilitated by hinge residues Ala154 and Gly326 in the two helices (α5 and α13) at the intersection of the N- and C-terminal domains (Figure 39). The CONNECT1 and CONNECT2 motifs contained the conserved residues in this region that enable this domain movement. Bacik et al. [100] noted that AnmK was unlike other carbohydrate kinases in that it did not exhibit domain motion; however, a enzyme closure upon substrate binding is a hallmark of ASKHA kinase family members [156]. Our findings contradict the previous claim that AnmK is static upon substrate binding, but these predictions were made using conformations of *P. aeruginosa* AnmK containing either ADP or anhMurNAc, not both.

Bacik et al. [100] also proposed a mechanism where the sugar is hydrolyzed in the active site prior to catalysis. However, examination of the electron density map for the sugar unambiguously revealed that anhMurNAc bound in the anhydro form and was not hydrolyzed even in the presence of an ATP analog (Figure 40). This key detail will be used as a basis for the design of the proposed mechanism that does not include an initial separate anhMurNAc hydrolysis step.
Figure 39 - Structural Overlay of the Open and Closed Conformations of AnmK
This illustrates the overall change in AnmK upon substrate binding where the dark blue monomer is the open conformation and the light blue structure is the closed conformation. The structural superposition was performed in COOT [128, 129] using the SSM algorithm.
Figure 40 - Electron Density Mesh Around the AnhMurNAc Residues in the Structure of AnhMurNAc-ANP-AnmK. Map is contoured at 1.2σ using a map generated in PHENIX [130].
Alignment of the two (open and closed conformations) structures obtained thus far in this work and the *P. aeruginosa* AnmK reveal that AnmK-ADP and AnmK-anhMurNAc forms from their work are at an intermediate stage of closure between our two forms. Furthermore the RMSD value for the closed conformation of AnmK in comparison to the published *P. aeruginosa* homolog is 0.9 and this value is 1.2 for the *S. oneidensis* homolog. These values are lower in the closed conformation of AnmK indicating that using these structures for molecular replacement may have been successful if the initial dataset for AnmK was in this conformation.

Moreover, the sequence alignment of *E. coli* (AnmK), *P. aeruginosa* (3QBX) and *S. oneidensis* (3CQY) homologs of AnmK reveal definite areas of sequence similarity (Figure 41). However, despite such low RMSD, they retain the same molecular topology though utilizing largely different sequences (47% conservation). In addition, an alignment of AnmK homologs in over 100 species was prepared and colored using a heat map of conservation with ConSurf (Figure 42). The blue regions represent the highly sequence conserved areas and each one of the motifs and active site resides identified for AnmK lie in these blue regions.

Taken together, the information gained from this structure allowed unequivocal location of the active site in AnmK and allowed for comparison with not only the closed conformation of AnmK but also the homologs from other species. Identification of the
active site geometry is crucial to predicting the mechanism of AnmK, and this mechanistic inquiry is the subject of the following chapter.

Figure 41 - Sequence Homology Between Three AnmK Species
This alignment was created using E. coli (3SF9, AnmK), P. aurginosa (3QBX) and S. oneidensis (3CQY) homologs in CLUSTALW [121].
Figure 42 - Sequence Homology Colored by Conservation by ConSurf [157-159]
Highly conserved residues are in blue and highly variable (less conserved) regions are in red.
Chapter 6: Structure of MgADP•VO₃-bound AnmK

6.1 Introduction

The use of vanadium-ADP as a transition state model

In the anhMurNAc-ANP-AnmK structure, the terminal phosphate of ANP was observed, but with only partial occupancy (50%). Various explanations involve radiation damage from the high-energy synchrotron radiation damage, chemical degradation in solution or enzyme-mediated chemical reaction. Therefore, investigation into the nature of the terminal phosphate moiety was necessary to determine more information about its role in catalysis. A phosphate mimic was the natural choice to substitute for the γ-phosphate. Traditionally, aluminum fluoride, beryllium fluoride and nitrate are used as phosphate mimics [160]. However, the metal oxoanion vanadate can also form transition state mimics of phosphoryl transfer reactions [160, 161].

The use of a terminal phosphate mimic with ADP is also a longstanding technique to probe the transition state of an enzyme that functions as an ATPase. The vanadate-ADP complexes have been proven as successful phosphate mimics in actin complexes, which are part of the same ASKHA kinase protein family as AnmK [160, 161]. Vanadate tends to form pentacovalent complexes with trigonal bipyramidal geometry, which makes this compound a close approximation of the transition state for phosphoryl transfer reactions [160]. Therefore, vanadium-ADP was investigated for use
as an appropriate phosphate mimic to visualize the terminal phosphate and also attempt the capture of AnmK in a transition state-like complex of phosphoryl transfer.

6.2. Methods for MgADP•VO$_3^-$-bound AnmK Structure

As with the AnmK-ANP-anhMurNAc complex (section 5.2.1), the enzyme was incubated at 37 °C with anhMurNAc and ADP•VO$_3^-$ prior to the addition of the well solution for crystallization. Using the Linbro-type hanging drop vapor diffusion trays, optimization experiments were carried out using the conditions that yielded crystals for the anhMurNAc-ANP-AnmK complex. Crystals formed under a variety of conditions, but the largest (500 × 200 × 100 µm) were grown in 10% PEG 400, 2.4 M ammonium formate and 100 mM HEPES (pH 7.8).

Home source (Cu Kα) data were collected at the Duke University X-Ray Crystallography Shared Resource. X-ray diffraction data were collected in house using the Rigaku MicroMax™-007 HF microfocus X-ray generator and a sled-mounted RAXIS IV++ image plate detector. Data were indexed and scaled to 2.50-Å-resolution using the d*TREK software suite (Table 10) where partially recorded reflections recorded on adjacent images were added to give full reflections. An oscillation range of 1° was used, and 180 frames were recorded with a crystal-to-image distance of 150 mm. A final unique, non-zero data set was obtained with an overall $R_{merge}$ of 0.072. The data set was 96.5% complete to 2.5-Å resolution (Table 10) with 99.4% completion in the highest resolution shell. Also, in this shell (2.59 – 2.50 Å) the average $I/\sigma$ was 3.7.
The anhMurNAc-ADP•VO₃⁻-AnmK protein crystallized in the $P2_12_12_1$ space group with a unit cell of 113 x 167 x 195 Å. Refinement proceeded smoothly and the final structure yielded reasonable crystallographic and geometric indicators after multiple iterative rounds of refinement including simulated annealing and TLS refinement [13, 14]. Data collection, model, and refinement statistics are summarized in Table 10. The structure has appropriate stereochemical merits, with a final R-values of R_work = 20.5% and R_free = 26.5% with 5% of the data withheld into the R_free test set. The Ramachandran plot, produced by MolProbity [9], shows that 95.6% of the residues are in favored regions and there are no outliers.

The structure was solved by molecular replacement using PHASER [149] iteratively solving using a dimer from the anhMurNAc-ANP-AnmK complex as a search model. The initial model was manually rebuilt using COOT [128, 129]. Iterative structure refinement was carried out using PHENIX [130], specifically phenix.refine with initial noncrystallographic symmetry (NCS) restraints for the refinement. Ligand molecular topologies for ADP and vandate (VN4) were obtained using the HIC-Up server [150]. The molecular coordinates and topology for anhMurNAc and VN4 were obtained and optimized using REEL that is a part the PHENIX suite [130], and through personal communication with Nigel Moriatry, PhD, the creator of REEL.

Analysis of the stereochemical quality of the model was accomplished using AutoDepInputTool (ADIT) (http://deposit.pdb.org/adit/) and MOLPROBITY [133].
Figures were prepared with PyMOL [134]. Atomic coordinates and experimental structure factors have been deposited in the Protein Data Bank (PDB) and are accessible under the code 3SFB.

6.3 Results and Discussion of Mechanistic Insights

6.3.1 The structure of ADP•VO₃-bound AnmK

The crystal structure of ADP•VO₃-bound AnmK was solved from home source data to a resolution of 2.5 Å. The ADP•VO₃-bound protein crystallized in the P2₁2₁2₁ space group with a unit cell of 113 x 169 x 195 Å. The crystal used for data collection had a mosaicity of 0.86, which was markedly lower than the anhMurNAc-ANP-AnmK complex, but similar to the unbound conformation of AnmK (0.79). The structure was solved using one of the biological dimers from the anhMurNAc-ANP-AnmK structure as the search model, because the use of the entire octamer was inefficient in silico. This structure contained eight monomers in the ASU with four sets of biological dimers. As seen before, the biological dimer has an extended β-sheet bridging the monomers that is formed at the interface of the additional β-strand from the C-termini of the monomers.

Each AnmK monomer is comprised of 14 α helices, 13 β sheets and 4 3₁₀ helices arranged into two domains (N-terminal and C-terminal). The N-terminal domain is comprised of residues 2–160 and 321–369, and the C-terminal domain contains residues 161–320. In the complete, contiguous structure, each monomer is comprised of residues 1–369, one anhMurNAc molecule (AH0), one ADP molecule, one magnesium, one
vanadate residue (VN4), the residues from the expression tag (SFT, -2–0) and approximately 100 water molecules. This structure also contained two interfacial zinc binding sites in the octameric ASU.

The final structure yielded reasonable crystallographic and geometric indicators after multiple rounds of refinement including simulated annealing and TLS refinement [131, 132]. Data collection, model, and refinement statistics are summarized in Table 10. The structure has appropriate stereochemical merits, with a final R-values of \( R_{\text{work}} = 20.6\% \) and \( R_{\text{free}} = 26.5\% \) with 5% of the data withheld into the \( R_{\text{free}} \) test set. The Ramachandran plot, produced by MolProbity, shows that 95.6% of the residues are in favored regions and only 0.4% are outliers. Again, most if these outliers lie within a monomer with weak density (chain E), in the terminal regions of the eight monomers or on solvent exposed loops. Any rotamer outliers are the large extended side chains of Lys and Arg residues stretched out into the bulk solvent. Due to the inherent mobility of these regions and these residues, the deviation is not unexpected.

Like the other two AnmK structures, the ADP•VO₃⁻-bound structure contains an N-terminal RNAse H-like fold and a similar C-terminal fold that comprises the dyad in each monomer. The disordered loop not present in the \textit{de novo} structure solution was also ordered in this structure like the anhMurNAc-ANP-AnmK structure. Here again, the disorder arose in chain E as one of the monomers lacked contiguous density for the entire sequence. This was observed in anhMurNAc-ANP-AnmK and is not uncommon.
in large polymeric ASU architectures. Additionally, the crystal morphology changed from the rounded appearance of unbound AnmK to the prism-like crystals seen in ANP-anhMurNAc-AnmK (Figure 43).
Figure 43 - Crystals in Hanging Drop Vapor Diffusion Trials Containing ADP•VO₃
### Table 10 - Data Collection and Refinement Statistics for ADP•VO₃-bound AnmK

<table>
<thead>
<tr>
<th>Data collection</th>
<th>ADP•VO₃ AnmK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>DHVI</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>113.24, 168.85, 194.85</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Wavelength</td>
<td>1.5418</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>42.68 – 2.50 (2.59 – 2.50)</td>
</tr>
<tr>
<td>R_sym or R_merge</td>
<td>7.2 (32.5)</td>
</tr>
<tr>
<td>I / σI</td>
<td>11.1 (3.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.5 (99.4)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.99 (6.29)</td>
</tr>
<tr>
<td>Wilson B (Å²)</td>
<td>45.16</td>
</tr>
</tbody>
</table>

**Refinement**

| Resolution (Å)                | 42.68 – 2.50 |
| No. reflections               | 124,775      |
| R_work/R_free (%)             | 20.58/26.48  |
| No. atoms                     | 22072        |
| Protein                       | 21962        |
| Water                         | 110          |
| Ramachandran plot analysis    |              |
| Favored                       | 95.6%        |
| Outliers                      | 0.4%         |
| Clashscore                    | 21.61        |
| Rotamer outliers              | 6.1%         |
| R.m.s. deviations             |              |
| Bond lengths (Å)              | 0.009        |
| Bond angles (°)               | 1.173        |

Data were collected from a single crystal. Values in parentheses are for the highest-resolution shell.
6.3.2 The ADP•VO$_3$ binding site in ADP•VO$_3$-bound AnmK

The ADP binding site is reminiscent of the ANP binding site in anhMurNAc-ANP-AnmK and it again formed upon domain closure between the β1-β2 loop of the N-terminal domain and the β8-β9 loop of the C-terminal domain. Important residues are illustrated in Figure 44. Starting from the adenosine ring and moving to the phosphate terminus, the key interactions will be enumerated. With the domain closure, Asp196 forms hydrogen bonds with the 2′-OH and 3′-OH of the ADP ribose moiety. Also, Tyr207 aligns for π-stacking with the adenosine. The Asp210 backbone forms a hydrogen bond with the exocyclic nitrogen (N6) and the neighboring N1 on the adenine ring. Asp16 is hydrogen bonded to an α-phosphate oxygen, which is also coordinated with a magnesium ion and a β-phosphate oxygen. One α-phosphate oxygen is hydrogen bonded to the Gly296 backbone. One of the remaining two β-phosphate oxygens is coordinated with the vanadate molecule. The Ser14 side chain hydroxyl is hydrogen bonded to the other β-phosphate oxygen, which is reciprocally bonded to the Ser14 backbone and Gly170.

The vanadate (VO$_3$) is coordinated to the terminal phosphate of ADP with the ADP β-phosphate oxygen acting as an apical ligand for the vanadium. The bond length is as expected – long (~1.95 Å) for the apical ligands and approximately 1.8 Å for the equatorial ligands [161]. Figure 45 shows these bond lengths in addition to the electron density around all of the appropriate ligands and important active site residues.
Figure 44 - Important Binding Site Interactions with ADP, Mg$^+$ and VO$_3^-$
Figure 45 - Active Site Electron Density for ADP•VO$_3^-$ and Relevant VO$_3^-$ Bond Lengths
ADP, anhMurNAc, Asp187 and Glu330 are shown in purple. The magnesium ion is green and 1 $\sigma$ electron density mesh is teal. Chain G (white) was used to generate this figure.
6.3.3 The anhMurNAc binding site in ADP•VO$_3$-bound AnmK

The anhMurNAc binding site lies in the same locale as compared to the anhMurNAc-ANP-AnmK structure, close to the base of the enzyme cleft (α5 and α13). For anhMurNAc alignment in the active site, Arg134 hydrogen bonds the terminal carboxylate of the lactyl ether moiety attached to C3, which is also coordinated by Leu147 and Val148. The C4-hydroxyl below the plane of the carbohydrate is hydrogen bonded to Asp187 and a water molecule (HOH502), which is also interacting with C1. Additionally, Thr100 is bound to the amide carbonyl of the acetamide group on C2 of anhMurNAc. From this perspective, the underside of the carbohydrate central ring is solvent accessible and the acetamide moiety projects out into solvent accessible surface area flanked by His103, Arg134, Ala145, Asn173 and Glu330, among others. Because of this, the water molecule coordinated between C1 and Asp187 likely enters the active site from just below the anhMurNAc molecule. All other backbone interactions are shown in Figure 46, which also shows hydrophobic interaction with the vanadate bound to the ADP terminus ("Vn43(g)").

The conservation of the same interactions as anhMurNAc-ANP-AnmK highlights their importance and further supports the necessity of conservation of the residues in the active site of AnmK in order to retain activity and specificity. Additionally, the position of vanadate in the active site indicates close proximity to the anhMurNAc carbohydrate substrate where a phosphate mimic would be expected.
Figure 46 - Important Binding Site Interactions with AnhMurNAc
6.4 Conclusions

As predicted, the vanadate ion bound the terminal $\beta$-phosphate of ADP to mimic the $\gamma$-phosphate of ATP. These studies led to the discovery of the terminal phosphate interaction with Glu330. This implicates Glu330 as an active site conformational stabilizer through the coordination of a protonated phosphate or $\text{Mg}^{2+}$. Glu330 may also be protonated and in turn, protonate the terminal phosphate to facilitate nucleophilic attack of the liberated carbohydrate alkoxide generated by hydrolytic ring opening of the carbohydrate substrate. Furthermore, it may act as a nucleophile to generate an enzyme phospho-intermediate.

Additionally, the enzymatic use of amino acid residues like Glu or Asp as general bases is not uncommon. For the enzyme NagK, a member of the peptidoglycan recycling pathway, Asp107 is used as a general base to deprotonate $N$-acetylglucosamine for phosphoryl transfer from ATP to result in $N$-acetylglucosamine-6-phosphate ($\text{GlcNAc-6-P}$). This may be applied to AnmK, however, it would require that anhMurNAc becomes protonated in the active site following hydrolysis so that Glu330 would have a proton to remove. Protonation of the anhMurNAc hydrolysis product would form MurNAc, which is not phosphorylated by AnmK [56, 57]. Therefore, it’s unlikely that Glu330 acts as a general base in the AnmK active site.
However, there are a few salient points regarding the comparison of AnmK and NagK. Figure 47 shows an overlay of AnmK with Glc- and GlcNAc-bound NagK aligned based on the RNase H-like fold of the two enzymes. The nucleotide phosphate is roughly in the same binding pocket of the enzyme. More notably, the catalytic base (Asp107) is in the same site as the potentially important AnmK residue (Glu330). The methylene difference between Asp and Glu may allow Glu330 to reach into approximately the same region of the carbohydrate substrate binding site. The proximity of ADP, vanadate, Glu330 and anhMurNAc noted in the vanadium-bound AnmK structure allow for the flexible side chain to shuttle a phosphate closer to the anhMurNAc as it is hydrolyzed.
Figure 47 - Structural Overlay of AnmK and NagK Illustrating Active Site Geometry
AnmK is shown in purple and NagK in green (GlcNAc) and red (Glc-bound).
AnmK also catalyzes the hydrolysis and phosphorylation of anhGlcNAc to GlcNAc-6-P. This may indicate that these enzymes (AnmK and NagK) have diverged from a common ancestor or conversely, have developed through convergent evolution to ultimately play a similar role. This sort of bioinformatics search was not in the scope of this work; however it is an attractive case for future investigation since the enzymes have minimal superimposable structural elements other than the RNase H-like fold, but very similar chemistry performed in the active site.

Finally and notably, this structure allows for the identification of the position of the magnesium counter ion commonly seen with ATP or ADP and requisite for catalysis as determined previously. The lack of observable magnesium in the anhMurNAc-ANP-AnmK structure might be due to a more planar nature of the phosphoramidite resonance contributors not present in the terminal phosphoester of ATP. The magnesium presence in related enzymes [162, 163] serves to stabilize the developing negative charge of the pentavalent transition state structure of the terminal ATP γ-phosphate moiety during transfer. The lack of magnesium counterions was also encountered by Bacik et al. [100] in *P. aeruginosa* and may be characteristic of the AnmK pre-catalytic structure, but further investigation would be necessary to support this hypothesis.
Chapter 7: Probing the Mechanism of AnmK

7.1 Introduction

Probing the origin of the C1 oxygen via substitution with $H_2^{18}O$

The use of isotopically labeled substrates is prevalent in the literature for the investigation of kinetic mechanism. The use of such labels imparts altered spectroscopic properties that allow for the differentiation from the naturally isotopically abundant atom. In previous kinetic studies of AnmK, Bacik et al. [100] utilized $^1H$ NMR to show initial formation of the $\alpha$ anomer of MurNAc-6-P. The anticipated mutarotation of the reducing sugar created the $\beta$ anomer as time elapsed. These studies, however, did not address the origin of the oxygen at C1, which can either come from the bulk solvent or bond breaking between C6 and the five-membered ring oxygen. The reaction being probed through the use of a pseudosubstrate (anhGlcNAc) follows in Figure 48 and an $H_2^{18}O$ wash-in experiment will help identify the resultant oxygen species at C1.

![Figure 48 - AnhGlcNAc Reaction Being Probed By $^{18}O$-Labeling Experiments](image)
Based on the data obtained from the crystal structures, the form of the carbohydrate in the active site in spite of the presence of an ATP analog is the anhydro carbohydrate that has not undergone hydrolysis. This evidence does not support the formation of MurNAc, or in this case, GlcNAc in the active site without another process working in tandem, because carbohydrate hydrolysis could have already occurred in the presence of the enzyme and appropriate substrates. The conserved water residue present in every AnmK monomer suggests that it plays a role in catalysis. Therefore, the hypothetical labeling reaction with H\textsubscript{2}\textsuperscript{18}O will occur as follows (Figure 49).

Figure 49 - Proposed \textsuperscript{18}O Labeling at C1 from Bulk Solvent
A combination of mass spectrometry and NMR analysis can elucidate the presence of this C1-adduct. For NMR, the resonance near 100 ppm is indicative of the anhydro carbohydrate C1 position. When the anhGlcNAc is cleaved by AnmK, this results in two peaks for the α and β anomers of GlcNAc-6-P. The mutarotation for the GlcNAc-6-P anomer is illustrated below (Figure 50) and is hypothesized to only be a minor contributor to the wash out of $^{18}$O label. When incubated with 50% $H_2^{18}$O, the isotopically labeled product would give rise to a second chemical shift for the atom to which it is attached. The shift is not expected to be large (0.02 – 0.05 ppm) [164], but should be distinguishable. In combination with ESI-MS techniques to confirm the appropriate change in $m/z$ and fragmentation by MS/MS, the incorporation of $^{18}$O at C1 by AnmK can be investigated in the reaction with anhGlcNAc.

Taken together the biophysical techniques described in this chapter will help characterize the product of the AnmK reaction with anhGlcNAc, a suitable pseudosubstrate, and aid in the design of a mechanism for simultaneous substrate hydrolysis and phosphorylation.
Figure 50 - Mechanism of Mutorotation and Predicted Retention of Label (black)
7.2 Experimental Procedures

7.2.1 Enzymatic synthesis and $^{13}$C NMR analysis of $^{18}$O-labeled GlcNAc-6-P

A 1 mL solution of 100 mM Tris pH 8, MgCl$_2$ (25 mM), and ATP (25 mM) was prepared and then anhGlcNAc (19.5 mg) and AnmK (10 µg) were dissolved therein. For the control reaction, an additional 1.0 mL of H$_2$O was added. The experimental sample contained an extra 1.0 mL aliquot of H$_2^{18}$O to obtain a final volume of 2 mL of 50% H$_2^{16}$O: 50%H$_2^{18}$O. Incubation of the mixture was performed at 4 °C for 16 h, whereupon the reaction was terminated by the filtration of AnmK out of solution using a 10K MWCO centrifugal filter unit (Millipore, Billerica, MA). For NMR analysis, 7 µL of D$_2$O was added to 70 µL of the reaction mixture to reach a final 10% D$_2$O requisite for shimming the sample.

7.2.2 Determination of $^{18}$O incorporation by mass spectrometry

The AnmK H$_2^{18}$O enzymatic reaction products were also analyzed by ESI-MS. For this analysis, 5 µL of each of the reaction mixtures above (section 6.2.2) was diluted to 1 mL. The [M-H] peak (299.9 m/z) was identified using the H$_2^{16}$O control reaction. Upon incubation of the reaction mixture in 50% H$_2^{16}$O: 50%H$_2^{18}$O, a peak at 301.9 m/z appeared in addition to the peak at 299.9 m/z. Further MS/MS fragmentation was performed on the signals at 299.9 m/z and 301.9 m/z.
7.3 Results and Discussion of Mechanistic Insights

7.3.1 Mass spectrometry of $^{18}$O-labeled GlcNAc-6-P

Because of the significantly improved sensitivity offered by mass spectrometry over NMR-based methods, investigation into the AnmK $H_2^{18}$O enzymatic reaction products was carried out using ESI-MS. This experiment confirms that the $^{18}$O label is attached to C1 of GlcNAc-6-P. From a control reaction using $H_2^{16}$O without $H_2^{18}$O, the [M-H] peak (299.9 m/z) was identified based on the ionization of an average MW of 301.18 g/mol for GlcNAc-6-P. Upon incubation of the reaction mixture in 50% $H_2^{16}$O: 50%$H_2^{18}$O, a peak at 301.9 m/z appeared in addition to the control peak at 299.9 m/z.

Furthermore, when MS/MS was performed in negative ion mode with 301.9 m/z as the parent ion, the $^{18}$O-label was located at the C1 position based on the fragmentation products (Figure 51) that follow. As expected, there is loss of water from the polyhydroxylated compound to yield a peak at 283.8 m/z. However, there is also the loss of $H_2^{18}$O that results in a peak at 281.8 m/z. The change in mass from the parent ion to the other abundant fragmentation product indicates that there is a loss of the HO-C1-C2-NHAc group to give rise to a signal at 198.7 m/z. This is the same signal exhibited by the MS/MS with 299.9 m/z as the parent ion for fragmentation (not pictured); therefore, the $^{18}$O label is contained in the fragment lost. Taken together, this information indicates that the loss of $H_2^{18}$O comes from the loss of the readily ionized C1-OH on the HO-C1-C2-NHAc fragment of anhGlcNAc.
Figure 51 - MS/MS Data of 301.9 m/z Parent Ion Peak in AnmK Reaction in 50% H$_{2}^{18}$O
7.3.2 Determination of $^{18}$O incorporation by $^{13}$C NMR

For this experiment, a 50% solution of $^{18}$O-labeled water was used to probe the origin of the oxygen on the C1 position of the sugar. The control reaction established the expected NMR peaks using H$_2^{16}$O. Upon enzyme hydrolysis and phosphorylation, if the hydroxyl moiety on C1 originates from the bulk solvent, then the C1 position should contain 50% $^{18}$O. Therefore, the aqueous reaction was doped with H$_2^{18}$O, which led to the formation of a second peak at the C1 position of the $\alpha$ anomer of GlcNAc-6-P (Figure 53).

Initial experiments were carried out to validate that concomitant sugar hydrolysis and phosphorylation would be detectable by $^{13}$C NMR using reasonable amounts of enzyme and substrates. Freshly purified AnmK led to the turnover of the pseudosubstrate, anhGlcNAc, in amounts detectable by NMR. These two sugars are easily distinguishable because of the peak shift from 100 ppm in anhGlcNAc to 95 ppm ($\beta$ anomer) and 91 ppm ($\alpha$ anomer) in GlcNAc-6-P (Figure 52). As with any reducing sugar, mutorotation is inevitable. Figure 50 illustrates the retention of the $^{18}$O label during mutorotation; however, hydrolysis of the aldehyde during the step labeled “rotation” is certainly a concern. For this reason, the reaction was performed at lower temperatures, which was also important for minimizing denaturation and loss of activity for the AnmK enzyme over time.
After the appropriate controls, the reaction containing H$_{18}$O was successfully performed to yield a $^{13}$C spectrum with the hypothesized apparent peak doubling for the $\alpha$ anomer only (Figure 53) indicating that this reaction is not only regiospecific, but also stereospecific.

Figure 52 - The $^{13}$C NMR of the AnmK Control Reaction Mixture
The $^{13}$C NMR peaks shown ($\delta$) are: anhMurNAc C1 (100.68), GlcNAc-6-P $\beta$ C1 (95.54), GlcNAc-6-P $\alpha$ C1 (91.40), ATP ribose signals (87.57 – 83.84)
Figure 53 - $^{13}$C NMR Spectra Comparing H$_2^{16}$O with H$_2^{18}$O Experiment with Peak Shift of the MurNAc-6-P C1 α Anomer
7.4 Conclusions

7.4.1 Comparison to known enzyme mechanisms

When the important active site residues of AnmK were compared to NagK, hexokinase and β-glycosidases, it allowed speculation about the dual function of AnmK. Regarding the hydrolytic functionality, AnmK is similar to the inverting β-glycosidases with the same base-mediated deprotonation. However, instead of C1 RO⁻ release by the β-glycosidases (Figure 54) to be protonated by the Asp (or any carboxyl-containing amino acid) above the sugar (as oriented in Figure 55), AnmK releases the C6-O⁻ of anhMurNAc (or anhGlcNAc) to attack proximal γ-phosphate of ATP.

Figure 54 - General Mechanism of β-Glycosidase Enzymes
When comparing AnmK to the other kinases, it does not have the conserved Asp residue, but instead contains the conserved Glu330, which can be utilized by $\beta$-glycosidases. NagK and hexokinase perform similar functions, but NagK lacks the Asp residue below the active site. This may mean that this Asp is important for coordinating the $N$-acetyl moiety present in GlcNAc (and anhMurNAc/anhGlcNAc) and not glucose.
For comparison to AnmK, NagK, hexokinase and the general form of the β-glycosidases were used. Below the PDB code (blue) is the function of the appropriate enzyme listed above the active site components (purple).

Figure 55 - Comparison of AnmK to Two Other Kinases and A Hydrolase
7.4.2 Plausible Mechanism for AnmK Kinase/Hydrolase Activity

All kinases perform a phosphoryl transfer reaction, however, the dual hydrolysis and phosphoryl transfer catalyzed by AnmK is unique and may be unprecedented enzyme chemistry. When considering the reaction performed by AnmK, there are many different mechanistic possibilities. Some of which have already been ruled out by the uncleaved anhMurNAc found in the active site.

The most plausible mechanism for the dual hydrolysis and phosphorylation by AnmK is outlined below. This involves a catalytic base-mediated deprotonation of water, which then attacks at the anti-bonding orbital of the C1 to O bond that makes up the intramolecular acetal. This breaks that bond to release the C6-hydroxylate for nucleophilic attack on the γ-phosphate of ATP and releases ADP and MurNAc-6-P (Mechanism 2). Glu330 is also pictured as it is proximal to the bound substrates and may protonate the terminal phosphate to assist in alkoxide nucleophilic attack. The associative or dissociative nature of the phosphoryl transfer will not be speculated. Further studies would have to be carried out in order to propose a mechanism that fully characterizes phosphoryl transfer as well.
Mechanism 2 - Concomitant Hydrolysis and Phosphorylation with Asp187 as the Catalytic Base Glu330 in Close Proximity with A Potential Role in the Mechanism.
As mentioned in previous sections, a distinct, two-step mechanism that forms MurNAc is not supported by our data. In addition, the hydrolysis of the anhydro ring between C6 and the ring-bound oxygen is also unlikely because of the $^{18}$O labeling present at C1 that was elucidated in this work. If the ring were to undergo nucleophilic attack at C6, the liberated hydroxyl would initially be in the β anomeric form, which is not supported by this work or Bacik et al. [100]. Therefore, the regio- and stereoselectivity confirmed by these biophysical methods lead to the more detailed mechanism below (Mechanism 3). This mechanism is entirely concerted, but a two-step-like mechanism that forms the C6 hydroxyl cannot be ruled out yet.
Mechanism 3 - The Proposed Reaction Mechanism of AnmK

The inset (b) illustrates the 1,3 diaxial interactions that are present in the starting material and are one of the driving forces in the carbohydrate ring opening.
The other potential mechanism for the dual hydrolysis and phosphorylation of AnmK involves a transient enzyme phosphointermediate as already discussed following the previous ADP-vanadate structure. Based on the data from the ADP-vanadate structure, the role of Glu330 may be important due to its proximity to the terminal phosphate; therefore, this mechanism cannot be ruled out as a possibility either. Active site density and the mechanism proposed (Mechanism 3) from this work is represented again in Figure 56, which represents the goodness of fit for the electron density and the proximity of the substrates. This arrangement of substrates in the active site allows for the base-catalyzed hydrolysis at C1 coupled with phosphoryl transfer via nucleophilic attack of the liberated hydroxyl moiety on the γ-phosphate terminus of ATP to enable phosphate transfer that yields ADP and MurNAc-6-P. The mechanism as shown indicates the formation of a pentavalent intermediate during phosphoryl transfer that has been cited for this family of enzymes [165, 166], but further investigation would be necessary to determine the nature of this phosphate transfer.
Figure 56 - Electron Density of Important Active Site Residues and a Cartoon Simplification of the Proposed Mechanism
Chapter 8: Conclusions and Future Directions

8.1 General Conclusions

This dissertation work has provided an enhancement in the information known about a previously incompletely characterized enzyme, AnmK. The *anmK* gene was isolated and recombinantly expressed to incorporate a hexahistidine affinity tag. Protein overexpression conditions were then determined to rapidly result in homogeneous protein in high yield and it was characterized as a homodimer by sedimentation rate analytical ultracentrifugation.

Also, in this work, anhMurNAc, the natural substrate, and anhGlcNAc, a pseudosubstrate, were synthesized for use in kinetic and crystallographic studies. Following the attainment of homogeneous protein and these two synthetic products, the activity of AnmK was assessed using a PEI-cellulose radioactive assay to monitor the formation of ADP from ATP indirectly indicating that phosphoryl transfer had occurred. These studies verified the literature reports for the *Kₐ* for both substrates and the magnesium and ATP dependence of the reaction [57, 100]. In concert with the TLC assay, the dual hydrolysis/phosphorylation of anhGlcNAc was also verified by mass spectrometry to confirm the GlcNAc-6-P product.

After verification of AnmK activity, the protein was investigated by x-ray crystallography. Using cryocrystallographic techniques and with the aid of tunable synchrotron radiation for SAD phasing, three novel crystal structures were obtained.
These included the open conformation, the closed, ternary complex and the ADP-vanadate-bound transition state-like structure. The open conformation revealed a homodimer in the asymmetric unit where each monomer was comprised of two subdomains connected by a deep active site cleft. The core fold in each subdomain identified AnmK as a member of the ASKHA family of kinases. These enzymes catalyze ATP hydrolysis or phosphate transfer paired with a large conformational change. This change was seen in the closed conformation of AnmK where an 11° change allowed the two domains to close around the nucleotide enabling catalysis.

Additionally, the origin of the hydroxyl group added in the hydrolysis of anhMurNAc was investigated by $^{18}$O-labeling experiments, which revealed that the C1-hydroxyl oxygen is derived from bulk solvent. Presumably, this is the ordered water molecule that is present in each monomer below the plane of the anhMurNAc ring coordinated to Asp187. The geometric arrangement of the binding site residues in concert with the $^{18}$O-labelling experiment and basic kinetic data allowed for the development of a proposed mechanism for AnmK. This mechanism involves the hydrolysis of anhMurNAc initiated by the deprotonation of a water molecule to facilitate attack at C1, forming the α-anomer of MurNAc and releasing the C6-hydroxyl to participate in an inline phosphoryl transfer to form MurNAc-6-P and ADP. As discussed, this mechanism best describes the amalgamation of our findings; however, we were unable to disprove the possibility of a transient phosphoenzyme intermediate.
or the role of the Glu330 residue. Continuation of this work would commence with this as the subject of further studies on AnmK.
8.2 Future Directions

8.2.1 Investigation of a phosphoenzyme intermediate

As mentioned previously, a mechanistic possibility that involves the formation of a phosphoenzyme intermediate for AnmK cannot be ruled out. A combination of burst kinetics, mass spectrometry and mutagenesis can be employed to attain support or evidence against this mechanism.

8.2.2 Site-directed mutation studies

Abrogation of AnmK activity has been shown for the alanine mutation of the catalytic Asp187. However, Glu330 has not been investigated through mutagenic analysis. An E330A mutation would help discern the essentiality of the Glu330 residue for catalysis. A reduction or loss in AnmK activity upon mutation to alanine would reveal that this residue plays a larger role in catalysis than previously believed. Currently, it is not factored into our molecular mechanism or the one put forth by Bacik et al. [100]. Furthermore, the E330D mutation would mutate AnmK to resemble NagK and investigation of this change may reveal that it could change the enzyme specificity to also become more like NagK. What would be most interesting is if this mutation could create a kinase with conditional hydrolase activity dependent upon which substrate is present like anhMurNAc or MurNAc, for example.

Additionally, the crystal structure of an activity abrogating mutation may be able to best represent the active site geometry if the substrates will still bind in the active site.
In this case, an ATP analog would not be necessary, and this structure would give an unbiased view of the active site geometry without the small change in chemical character and bond lengths associated with an analog molecule.

**8.2.3 Conditional essentiality of AnmK**

Park and Uehara [56] and Uehara et al. [57] have published that AnmK is not an essential enzyme; however, it is a part of an essential pathway in bacteria. Investigation of its essentiality under stressful cellular conditions like those of an infection is proposed. If AnmK is knocked out of a strain of pathogenic *E. coli*, this may reveal that when bacteria are rapidly dividing, the peptidoglycan recycling pathway plays a larger role and the lack of AnmK may then be detrimental.

**8.2.4 Inquiry into related kinases**

According to Miller and Raines [167], members of the ROK kinase family, the larger family that the ASKHA kinases belong to, can employ substrate promiscuity and increased enzyme production to evade shunts in their respective pathways. This is not surprising as AnmK can act on at least three anhydro carbohydrates including the substrate of another anhydro carbohydrate kinase, levoglucosan kinase (LgK). However, LgK has only been identified in fungi, which don’t contain MurNAc. Therefore, perhaps LgK is the fungal AnmK involved in the metabolism of their unique chitin cell wall.
Nonetheless, bioinformatic investigation of the phylogenetic relationship of these two enzymes may reveal a recent divergence due to the fact that LgK has a higher sequence similarity with *E. coli* AnmK (68%) than *E. coli* AnmK has with *S. oneidensis* AnmK (40%). It may appear that LgK is just a promiscuous homolog of AnmK; however, Dai *et al.* [168] published studies comparing substrate utilization and LgK does not metabolize anhMurNAc.

### 8.2.5 Substrate mimetics as potential inhibitors

For further study of the molecular determinants of specificity, it would be useful to synthesize a series of molecules based on the modification of the natural substrate (Figure 57). The novel [3.2.1] bicyclic geometry of the carbohydrate substrate should remain unchanged as this enzyme has displayed a strong predilection for sugars in the anhydro form.

The first and likely most facile compound recommended for synthesis and study is 1,6-anhydroglucosamines 30a and 30b. Using the same procedure that was used for the 6-O-tosyl-N-acetylglucosamine compound (22), the starting material would simply be switched to glucosamine. Presumably, this compound may be a poor substrate with a binding affinity and turnover comparable to levogluconsan (4), but the amine in place of the hydroxyl moiety may help characterize the importance of a nitrogen at the C2 position. Along the same lines, the deacetylation of the C2-NHAc group and the retention of the C3-lactyl ether group may actually be a better indication of the
importance of acetylation at C2. This proposed substrate is actually the penultimate compound (29) in the synthesis of anhMurNAc (1).

In addition to the anhMurNAc simplification products, molecules with nitrogen in place of the in-ring oxygens may prove useful. The oxygen shared by both the 5- and 6-membered ring (C5-OH of the sugar), when replaced by nitrogen, may create compounds (31a and 31b, 32a and 32b) that may stabilize the formation of a planar intermediate. Alternatively, the elimination of both ring oxygens to form an all carbon backbone (33a and 33b) could reveal information about the necessity of these atoms in the ring. Furthermore, the two related compounds with each one of the two ring oxygens missing could be prepared (34a and 34b, 35a and 35b). The use of any or all of these molecules may provide valuable insight into the molecular determinants of specificity for the carbohydrate substrate.

![Diagram of proposed anhMurNAc substrate mimic molecules](image)

**Figure 57 – Proposed AnhMurNAc Substrate Mimic Molecules**
Appendix A – Supplemental Figures

**Guide to Running a Gel**

**Agarose Gels**

Making a 1% Gel
1. Combine 0.5g agarose and 50 mL TAE buffer in a 250mL Erlenmeyer.
2. Heat 1.5-2 minutes in microwave
3. Cool until glass is comfortable to hold
4. Add 5µL of ethidium bromide (EtBr)
5. Pour into gel apparatus and insert comb
6. Let cool until uniformly cloudy (~1 hour)
7. Fill wells with TAE buffer until gel is submerged
8. (Optional) Put 5-10µL EtBr in buffer well near cathode

Loading an Agarose Gel
1. Assume loading dye is 6x
2. Combine dye with sample
3. Pipette into gel wells

Running an Agarose Gel
1. Make sure gel is set up for DNA to run toward cathode
2. For most gels, 80V – 100V works well

**Agarose Loading Dye**

<table>
<thead>
<tr>
<th>Component</th>
<th>10 mL</th>
<th>20 mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xylene Cyanol FF (0.25%)</td>
<td>25 mg</td>
<td>50 mg</td>
</tr>
<tr>
<td>Bromophenol blue (0.25%)</td>
<td>25 mg</td>
<td>50 mg</td>
</tr>
<tr>
<td>Glycerol (30%)</td>
<td>6 mL</td>
<td>6 mL</td>
</tr>
<tr>
<td>Water</td>
<td>To 10 mL</td>
<td>To 20 mL</td>
</tr>
</tbody>
</table>

**SDS-PAGE Gels**

Making a Gel of Specified % Acrylamide
1. Combine ingredients for running gel. Add TEMED last
2. Swirl and pour into gel cassettes until you reach the top horizontal bar (see dotted line in picture below)
3. Squirt/pour isopropanol over this layer
4. Let running gel set up
5. Combine ingredients for stacking gel adding TEMED last
6. Pour out isopropanol and rinse with H₂O
7. Pour in the stacking gel layer
8. Insert desired combs before it gels
9. Store at 4 °C wrapped in paper towels soaked in used running buffer

Loading an SDS-PAGE Gel
1. Combine 5µL dye with 20µL sample
2. Denature at 98°C for 10 min
3. Pipette into gel wells

Running an SDS-PAGE Gel
1. Make sure tape is removed from gel cassette.
2. For most gels, use 180V – 220V. Up to 300V can be used if resolution is less important

**SDS-PAGE Running Buffer**

<table>
<thead>
<tr>
<th>Component</th>
<th>100 mL</th>
<th>1L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stain 0.05% (w/v)</td>
<td>Brilliant Blue R-250</td>
<td>50 mg</td>
</tr>
<tr>
<td>40% (v/v) Ethanol (to dissolve dye)</td>
<td>40 mL</td>
<td>400 mL</td>
</tr>
<tr>
<td>10% (v/v) Glacial acetic acid</td>
<td>10 mL</td>
<td>100 mL</td>
</tr>
<tr>
<td>50% (v/v) Water</td>
<td>50 mL</td>
<td>500 mL</td>
</tr>
</tbody>
</table>

Destain: Exactly like stain, but without the blue dye

**SDS-PAGE Loading Dye**

<table>
<thead>
<tr>
<th>Component</th>
<th>1 mL</th>
<th>5 mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2M Tris, pH 6.8</td>
<td>50 µL</td>
<td>250 µL</td>
</tr>
<tr>
<td>1M DTT in 10mM NaAc, pH 5.2</td>
<td>200 µL</td>
<td>1 mL</td>
</tr>
<tr>
<td>10% SDS</td>
<td>400 µL</td>
<td>2 mL</td>
</tr>
<tr>
<td>Bromophenol blue</td>
<td>2 µg</td>
<td>10 µg</td>
</tr>
<tr>
<td>80% glycerol</td>
<td>250 µL</td>
<td>1.25 mL</td>
</tr>
<tr>
<td>H₂O</td>
<td>to 1 mL</td>
<td>to 5 mL</td>
</tr>
</tbody>
</table>

**SDS-PAGE Gel Visualization Solutions**

<table>
<thead>
<tr>
<th>Component</th>
<th>100mL</th>
<th>1L</th>
</tr>
</thead>
<tbody>
<tr>
<td>40% Acrylamide (30%)</td>
<td>375 µL</td>
<td>(500)</td>
</tr>
<tr>
<td>1.0 M Tris, pH 6.8</td>
<td>380 µL</td>
<td>2.28 mL</td>
</tr>
<tr>
<td>10% SDS (100mg/mL)</td>
<td>30 µL</td>
<td>180 µL</td>
</tr>
<tr>
<td>10% APS (100mg/mL)</td>
<td>30 µL</td>
<td>180 µL</td>
</tr>
<tr>
<td>TEMED</td>
<td>3 µL</td>
<td>18 µL</td>
</tr>
<tr>
<td>H₂O</td>
<td>2.18 mL (2.06)</td>
<td>13.08 mL (12.36)</td>
</tr>
</tbody>
</table>

**SDS-PAGE Gel Layers**

<table>
<thead>
<tr>
<th>Component</th>
<th>100mL</th>
<th>1L</th>
</tr>
</thead>
<tbody>
<tr>
<td>40% Acrylamide (30%)</td>
<td>375 µL</td>
<td>(500)</td>
</tr>
<tr>
<td>1.0 M Tris, pH 6.8</td>
<td>380 µL</td>
<td>2.28 mL</td>
</tr>
<tr>
<td>10% SDS (100mg/mL)</td>
<td>30 µL</td>
<td>180 µL</td>
</tr>
<tr>
<td>10% APS (100mg/mL)</td>
<td>30 µL</td>
<td>180 µL</td>
</tr>
<tr>
<td>TEMED</td>
<td>3 µL</td>
<td>18 µL</td>
</tr>
<tr>
<td>H₂O</td>
<td>2.18 mL (2.06)</td>
<td>13.08 mL (12.36)</td>
</tr>
</tbody>
</table>
References


14. Gram, H. (1884) Über die isolierte färbung der Schizomyceten in schnitt- und
trockenpräparaten, Fortschritte der Medizin 2, 185-189.

Strategies, Cabi.

of a "more-than-single-layered" peptidoglycan architecture of Escherichia coli W7:
a neutron small-angle scattering study, Journal of Bacteriology 173, 751.

glycopolymers in Gram-positive physiology and host interactions, Nature
Reviews Microbiology 6, 276-287.

bacteria and mechanisms of their targeting to the cell wall envelope, Microbiology
and Molecular Biology Reviews 63, 174-229.

CRC Critical Reviews in Microbiology 17, 121-135.

peptidoglycan, FEMS Microbiology Reviews 32, 287-306.

Staphylococcus aureus wall teichoic acid biosynthesis based on in vitro
reconstitution of the intracellular steps, Chemistry & Biology 15, 12-21.

of linkage unit intermediates in the biosynthesis of ribitol teichoic acids in
Staphylococcus aureus H and Bacillus subtilis W23, European Journal of Biochemistry
161, 479-489.

Annual Review of Microbiology 42, 359-393.


Biotechnology 11, 115-125.


217


bond cleavage and phosphoryl transfer by *Pseudomonas aeruginosa* 1,6-anhydro-N-acetyl-

efficient large-scale synthesis of 1,6-anhydrohexopyranoses, *Journal of Organic  
Chemistry* 54, 6123-6125.

oligosaccharides. 69. Synthesis of 1,6-anhydromuramyl peptides, *Liebig's Annalen  
Der Chemie*, 664-674.

103. Cerný, M., and Stanek Jr, J. (1977) 1,6-Anhydro derivatives of aldohexoses, in  
*Advances in Carbohydrate Chemistry and Biochemistry* (Tipson, R. S., and Derek, H.,  
Eds.), pp 23-177, Academic Press.

applications of D-and L-form 1, 6-anhydrohexopyranoses for the synthesis of  

and stereocontrolled preparation of 2-azido-2-deoxy-beta-D-glucopyranose  
derivatives from D-glucal, *Journal of the Chemical Society-Perkin Transactions 1*,  
3163-3164.


2,4-dideoxy-mannopyranoses using 1,6-anhydro sugar chemistry, *Carbohydrate  
Research* 274, 29-44.

trifluoromethanesulfonates and their reactions with alcohols, *Journal of Organic  
Chemistry* 38, 3673-3677.


Catherine Leigh Allen was born in Charlotte, North Carolina on May 3, 1984 to Donald Wayne Allen, Jr. and Sarah Albright Allen, now Sarah Albright Greene.

She attended Wake Forest University in 2002 and carried out undergraduate research in two chemistry department labs. Under the supervision of S. Bruce King, Ph.D., Leigh was included on a 2005 Tetrahedron Letters paper titled: “A convenient method for the synthesis of N-hydroxyureas.” After research in the King lab, she moved across the hall to work with Bernard A. Brown II, Ph.D. Here, she completed her Honors thesis titled: “Structural and Mutagenic Analyses of the M. jannaschii L7Ae Loop 9 RNA-Binding Region.” Leigh graduated in May of 2006 with an Honors B.S. in chemistry with a concentration in biochemistry and a minor in psychology.

From Winston-Salem, she moved to Durham and began work for Prof. Dewey McCafferty at Duke University. After her first year, Leigh got accepted into the Pharmacological Sciences Training Program, which was a merit-based two-year fellowship certificate program. Following that, she was awarded the Burroughs Wellcome Fellowship in 2009 for exceptional independent research ability and promising academic and professional development. In June 2011, Leigh successfully defended her Ph.D. dissertation in chemistry and also earned a pharmacology certificate. She is moving to Buffalo, NY to start work as a Postdoctoral Associate in Dr. Andrew Gulick’s Lab at Hauptman-Woodward Medical Research Institute.