Using Whole Genome Sequencing to Accurately Predict β-lactam Antimicrobial Susceptibility Profiles of Invasive *Streptococcus pneumoniae* Isolated in Canada

By

Rachel Kathleen Hink

A Thesis Submitted to the Faculty of Graduate Studies

of the University of Manitoba

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Department of Medical Microbiology and Infectious Diseases

Rady Faculty of Health Sciences

University of Manitoba

Winnipeg, Manitoba, Canada

Copyright © March 16th 2020 by Rachel Hink

ABSTRACT

Streptococcus pneumoniae is a common etiologic agent of invasive bacterial disease worldwide. A proportion of invasive *S. pneumoniae* isolates demonstrate resistance to β-lactam antibiotics through the mechanism of target site modification. β-lactams are frequently used to treat invasive pneumococcal disease (IPD), therefore necessitating prompt detection of resistance. Whole-genome sequencing (WGS) potentially offers reduced time to results compared to phenotypic antimicrobial susceptibility testing (AST) methods. The purpose of this thesis was to develop three WGS-based AST models predicting *S. pneumoniae* β-lactam MICs. We hypothesized that WGS cannot be used exclusively in the determination of *S. pneumoniae* β-lactam MIC values.

Invasive *S. pneumoniae* isolates were selected from the SAVE study for WGS. Sequencing was performed through the SAVE partnership between the Canadian Antimicrobial Resistance Alliance and the National Microbiology Laboratory. Genetic analyses included the characterization of PBP amino acid sequences, phylogeny and amino acid conservation analysis. AST of study isolates was performed in triplicate for three β-lactams of interest. Multiple linear regression models were developed and validated for the three β-lactams, generating predicted MICs that were compared to phenotypic values.

Of the six pneumococcal PBPs, PBP2x, PBP2b and PBP1a were the primary focus of this study. PBP amino acid sequences were shown to be highly mutated in β-lactam-nonsusceptible isolates, whereby the level of sequence variation ranged from 0-13.8%, 0-7.1% and 0-12.4% for PBP2x, PBP2b and PBP1a, respectively. Sequence variability was correlated with serotype and the level of resistance; PBP sequence clusters were comprised of predominant serotype/MIC combinations in each of the three phylogenies. Distinct regions of variability in the 3D

structures of the PBPs were observed by conservation analysis, commonly within the transpeptidase domain active site.

Several known resistance-conferring substitutions were selected as predictors of MIC. Variables were weighted using forward regression, generating models for penicillin, ceftriaxone and meropenem with adjusted R^2 values of 0.964, 0.936, and 0.971, respectively. All three models demonstrated EA (≥95%) and CA (≥90%) rates within CLSI acceptable performance specifications. Model rates of error were also at acceptable levels, except for VME rates (3.6%) for penicillin by CLSI IV meningitis breakpoints and mE rates (10.5%) for ceftriaxone.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors, Dr. George Zhanel and Dr. Heather Adam, for agreeing to take me on as their graduate student and for providing me with countless opportunities that have greatly contributed to both my personal and professional growth. Your expertise, guidance and encouragement have been instrumental over the course of my studies, and I have benefitted immensely from being a member of the Canadian Antimicrobial Resistance Alliance. I would also like to thank the members of my advisory committee, Dr. Michael Mulvey, Dr. Denice Bay and Dr. Sheryl Zelenitsky, for your advice and support over the years.

Thank you to both past and present members of the Canadian Antimicrobial Resistance Alliance, Nancy Laing, Melanie Baxter, Kim Nichol, Karen Wake, Ravinder Vashisht, Sali Biju, Barb Weshnoweski, Mary Tarka, Dr. Daryl Hoban and Dr. James Karlowsky, for your continuous support and encouragement, you have all truly made my time in the laboratory a wonderful experience. Special thanks go to Dr. Alyssa Golden for mentoring me throughout my time as a graduate student and for providing me with invaluable training and expertise. I would also like to thank members of the Streptococcus and STI Unit, particularly Irene Martin, Walter Demczuk and Shelley Peterson, for the training that I received at the NML and for being so welcoming during my time in the laboratory.

Thank you to my wonderful parents, Chris and Jacqueline, for your unwavering support, patience and encouragement, and for teaching me the importance of hard work in achieving my goals. To Brad, thank you for providing me with daily support, positivity and laughs and for being a shoulder to lean on throughout this long journey. Lastly, thank you to Sarah and Lisa for your life-long friendship.

TABLE OF CONTENTS

ABSTRACT ..... **i**

ACKNOWLEDGMENTS ..... **iii**

TABLE OF CONTENTS ..... **iv**

LIST OF TABLES ..... **xi**

LIST OF FIGURES ..... **xiii**

LIST OF ABBREVIATIONS ..... **xv**

1. INTRODUCTION **1**

1.1 *Streptococcus pneumoniae* **1**

1.1.1 Laboratory Isolation, Identification and Serotyping 1

1.1.2 Colonization and Transmission 2

1.1.3 Clonal Evolution, Genetic Variation and the Genome of *S. pneumoniae* 3

1.1.4 Pathogenesis and Invasive Pneumococcal Disease 5

1.2 Treatment of IPD **7**

1.3 β-lactam Resistance in *S. pneumoniae* **8**

1.3.1 β-lactam Antibiotics and Penicillin Binding Proteins (PBPs) 8

1.3.2 Pneumococcal PBPs 10

1.3.3 Mechanism of β-lactam Resistance in *S. pneumoniae* 11

1.3.4 Genetic Exchange of β-lactam Resistance Determinants 12

1.4 Detection of β-lactam Resistance in *S. pneumoniae* 13

1.4.1 Phenotypic Antimicrobial Susceptibility Testing (AST) Methods 13

1.4.2 Whole Genome Sequencing (WGS)-based AST 15

1.5 The SAVE Study 17

1.5.1 Background on PCV use in Canada 17

1.5.2 Premise and Study Design 18

2. RATIONALE AND OBJECTIVES 20

3. HYPOTHESIS 22

4. MATERIALS AND METHODS 23

4.1 AST 23

4.1.1 Preparation of Antimicrobial Solutions 23

4.1.2 Broth Microdilution 23

4.1.3 MIC Interpretive Criteria 24

4.2 Isolate Selection 24

4.3 AST of β-lactam-Susceptible Isolates 24

4.4 WGS 29

4.4.1 DNA Extraction and Quantification 29

4.4.2 DNA Library Preparation 30

4.4.3 Sequencing, Assembly and Annotation 32

4.4.4 Analysis of PBP2x, PBP2b and PBP1a DNA and Protein Sequences 33

4.4.4.1 Sequence Extraction and Alignment 33

4.4.4.2 Identification of Mutations 35

4.4.4.3 Amino Acid Conservation Analysis and Mutation Mapping 35

4.4.4.4 Phylogenetic Analysis and Cluster Identification 36

4.4.4.5 Sequence Comparison 37

4.5 Identification and Classification of β-lactam-Resistance Conferring Mutations in *S. pneumoniae* PBP2x, PBP2b and PBP1a Sequences 37

4.6 Statistical Analysis 41

4.6.1 Principal Components Analysis (PCA) 41

4.6.2 Multiple Linear Regression 42

4.6.3 β-lactam MIC Model Validation 43

5. RESULTS 45

Part A: Molecular Characterization of *S. pneumoniae* PBP2x, PBP2b and PBP1a Amino Acid Sequences 45

5.1 Characterization of PBP2x 45

5.1.1 Conservation Analysis and Mapping of PBP2x Mutations 45

5.1.2 PBP2x Phylogenetic Analysis and Cluster Identification 50

5.1.3 PBP2x Sequence Comparison 55

5.1.4 Pattern Analysis of PBP2x Mutations 60

5.2 Characterization of PBP2b 65

5.2.1 Conservation Analysis and Mapping of PBP2b Mutations 65

5.2.2 PBP2b Phylogenetic Analysis and Cluster Identification 70

5.2.3 PBP2b Sequence Comparison 75

5.2.4 Pattern Analysis of PBP2b Mutations 80

5.3 Characterization of PBP1a 84

5.3.1 Conservation Analysis and Mapping of PBP1a Mutations 84

5.3.2 PBP1a Phylogenetic Analysis and Cluster Identification 89

5.3.3 PBP1a Sequence Comparison 94

5.3.4 Pattern Analysis of PBP1a Mutations 99

Part B: Tracking the Development of β-lactam Resistance in Different *S. pneumoniae* Serotypes 103

5.4 Comparison of PBP2x, PBP2b and PBP1a Mutation Profiles by β-lactam MIC uniquely associated with different *S. pneumoniae* serotypes 103

5.4.1 Serotype 19A 103

5.4.2 Serotype 6AC 107

5.4.3 Serotype 15AB 109

5.4.4 Serotype 35B 111

Part C: Development and Validation of Statistical Models to Predict β-lactam MICs in *S. pneumoniae* 113

5.5 Variable Reduction by PCA 113

5.5.1 PCA of PBP2x-Associated Variables 113

5.5.2 PCA of PBP2b-Associated Variables 117

5.5.3 PCA of PBP1a-Associated Variables 120

5.6 Multiple Linear Regression Models Predicting *S. pneumoniae* β-lactam MICs 123

5.6.1 Penicillin MIC Model Development 123

5.6.2 Penicillin MIC Model Validation 127

5.6.3 Ceftriaxone MIC Model Development 130

5.6.4 Ceftriaxone MIC Model Validation 134

5.6.5 Meropenem MIC Model Development 138

5.6.6 Meropenem MIC Model Validation 142

6. DISCUSSION 146

6.1 Putting it all Together: Characterization of Pneumococcal PBPs 146

6.1.1 Associating Sequence Variability with Serotype and β-lactam Resistance 146

6.1.2 Associating Regions of Variability with Mechanisms of β-lactam Resistance 148

6.1.2.1 Variability within the TP Domain Active Site 149

6.1.2.2 Regions of Variability in Other Structural Domains 149

6.2 PBP Mutational Patterns and Serotype-specific Mechanisms of β-lactam Resistance 151

6.2.1 PBP2b: Not Just for Low-level β-lactam Resistance? 152

6.2.2 PBP1a: Not Just for High-level β-lactam Resistance? 153

6.3 Evaluation of WGS-based Statistical Models Predicting *S. pneumoniae* β-lactam MICs 154

6.3.1 Assessment of β-lactam MIC Model Performance 154

6.3.2 Model Performance Compared to Other WGS-based AST Models Predicting β-lactam MICs 162

6.3.2.1 A Different WGS-based AST Model Applied to *S. pneumoniae* 162

6.3.2.2 A Similar WGS-based AST Model Applied to *N. gonorrhoeae* 164

6.3.3 Clinical Applications and Future Model Improvements 165

6.4 Study Limitations 168

7. CONCLUSIONS 170

8. FUTURE DIRECTIONS 173

9. REFERENCES 175

10. APPENDIX A 184

11. APPENDIX B 192

LIST OF TABLES

TABLE 4.1: *S. pneumoniae* gene and protein sequences utilized as references for PBP2x, PBP2b and PBP1a sequence analysis. 34

TABLE 4.2: Previously reported amino acid substitutions in PBP2x, PBP2b and PBP1a conferring reduced β-lactam susceptibility in *S. pneumoniae*. 39

TABLE 5.1: Characterization of the fifteen *S. pneumoniae* PBP2x sequence clusters highlighted in Figure 5.2. 53

TABLE 5.2: Analysis of the eleven previously reported amino acid substitutions in PBP2x associated with reduced β-lactam susceptibility in *S. pneumoniae* 62

TABLE 5.3: Frequency distribution of the eleven previously reported PBP2x substitutions analyzed collectively by penicillin MIC 63

TABLE 5.4: Characterization of the nine *S. pneumoniae* PBP2b sequence clusters highlighted in Figure 5.5 73

TABLE 5.5: Analysis of the nine previously reported amino acid substitutions in PBP2b associated with reduced β-lactam susceptibility in *S. pneumoniae* 82

TABLE 5.6: Frequency distribution of the nine previously reported PBP2b substitutions analyzed collectively by penicillin MIC 83

TABLE 5.7: Characterization of the five *S. pneumoniae* PBP1a sequence clusters and two sub-clusters highlighted in Figure 5.8 92

TABLE 5.8: Analysis of the eleven previously reported amino acid substitutions in PBP1a associated with reduced β-lactam susceptibility in *S. pneumoniae* 101

TABLE 5.9: Frequency distribution of the eleven previously reported PBP1a substitutions analyzed collectively by penicillin MIC 102

TABLE 5.10: PCA of the eleven PBP2x-associated variables displaying the loading values for each variable in a rotated component matrix 116

TABLE 5.11: PCA of the nine PBP2b-associated variables displaying the loading values for each variable in a rotated component matrix 119

TABLE 5.12: PCA of the eight PBP1a-associated variables displaying the loading values for each variable in a rotated component matrix 122

TABLE 5.13: Forward multiple linear regression analysis summary predicting *S. pneumoniae* penicillin MIC values 125

TABLE 5.14: Unstandardized and standardized coefficients of the eight independent variables included in the final linear regression model predicting *S. pneumoniae* penicillin MICs 126

TABLE 5.15: Summary of model predicted penicillin MIC values generated by leave-one-out cross-validation 128

TABLE 5.16: Forward multiple linear regression analysis summary predicting *S. pneumoniae* ceftriaxone MICs 132

TABLE 5.17: Unstandardized and standardized coefficients of the seven independent variables included in the final linear regression model predicting *S. pneumoniae* ceftriaxone MICs 133

TABLE 5.18: Summary of model predicted ceftriaxone MIC values generated by leave-one-out cross-validation 136

TABLE 5.19: Forward multiple linear regression analysis summary predicting *S. pneumoniae* meropenem MICs..... 140

TABLE 5.20: Unstandardized and standardized coefficients of the six independent variables included in the final linear regression model predicting *S. pneumoniae* meropenem MICs 141

TABLE 5.21: Summary of model predicted meropenem MIC values generated by leave-one-out cross-validation 144

LIST OF FIGURES

FIGURE 4.1: Expanded penicillin MIC distributions of 196 invasive *S. pneumoniae* isolates following repeat AST 26

FIGURE 4.2: Expanded ceftriaxone MIC distributions of 196 invasive *S. pneumoniae* isolates following repeat AST 27

FIGURE 4.3: Expanded meropenem MIC distributions of 196 invasive *S. pneumoniae* isolates following repeat AST 28

FIGURE 5.1: Amino acid conservation analysis of PBP2x from *S. pneumoniae* 47

FIGURE 5.2: Maximum likelihood phylogenetic tree of 190 *S. pneumoniae* PBP2x amino acid sequences 52

FIGURE 5.3: Structure-based multiple sequence alignment of select *S. pneumoniae* PBP2x amino acid sequences 57

FIGURE 5.4: Amino acid conservation analysis of *S. pneumoniae* PBP2b 67

FIGURE 5.5: Maximum likelihood phylogenetic tree of 190 *S. pneumoniae* PBP2b amino acid sequences 72

FIGURE 5.6: Structure-based multiple sequence alignment of select *S. pneumoniae* PBP2b amino acid sequences 77

FIGURE 5.7: Amino acid conservation analysis of the *S. pneumoniae* PBP1a TP domain 86

FIGURE 5.8: Maximum likelihood phylogenetic tree of 190 *S. pneumoniae* PBP1a amino acid sequences 91

FIGURE 5.9: Structure-based multiple sequence alignment of select *S. pneumoniae* PBP1a amino acid sequences 96

FIGURE 5.10: Combined PBP2x, PBP2b and PBP1a mutation profiles of six selected *S. pneumoniae* serotype 19A isolates with equivalent penicillin and ceftriaxone MICs 105

FIGURE 5.11: Combined PBP2x, PBP2b and PBP1a mutation profiles of six selected *S. pneumoniae* serotype 19A isolates with non-equivalent penicillin and ceftriaxone MICs 106

FIGURE 5.12: Combined PBP2x, PBP2b and PBP1a mutation profiles of six selected *S. pneumoniae* serotype 15AB isolates displayed by increasing penicillin MIC values 108

FIGURE 5.13: Combined PBP2x, PBP2b and PBP1a mutation profiles of six selected *S. pneumoniae* serotype 6C isolates displayed by increasing penicillin MIC values 110

FIGURE 5.14: Combined PBP2x, PBP2b and PBP1a mutation profiles of four selected *S. pneumoniae* serotype 35B isolates displayed by increasing penicillin MIC values 112

FIGURE 5.15: Accuracy of model predicted penicillin MICs compared to measured penicillin MICs for *S. pneumoniae* 129

FIGURE 5.16: Accuracy of model predicted ceftriaxone MICs compared to measured ceftriaxone MICs for *S. pneumoniae* 137

FIGURE 5.17: Accuracy of model predicted meropenem MICs compared to measured meropenem MICs for *S. pneumoniae* 145

LIST OF ABBREVIATIONS

3D	Three-dimensional
5%-SBA	Trypticase soy agar with 5% sheep blood
aa	Amino acid
ABCs	Active Bacterial Core Surveillance
AIC	Akaike information criteria
AMR	Antimicrobial resistance
AOM	Acute otitis media
AST	Antimicrobial susceptibility testing
ATCC	American Type Culture Collection
bp	Basepair
CA	Categorical agreement
CARA	Canadian Antimicrobial Resistance Alliance
CDC	Centers for Disease Control
CFU	Colony-forming unit
CLSI	Clinical and Laboratory Standards Institute
contig	Contiguous sequence
CPHLN	Canadian Public Health Laboratory Network
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
EA	Essential agreement
EtOH	Ethanol
EUCAST	European Committee on Antimicrobial Susceptibility Testing
g	Gram
GT	Glycosyltransferase
Hib	*Haemophilus influenzae* type b
HMM	High molecular mass
IPD	Invasive pneumococcal disease
IS	Insertion sequence
IV	Intravenous
JTT	Jones-Taylor-Thornton
LMM	Low molecular mass
MDR	Multidrug resistant
mE	Minor error
ME	Major error
mg	Milligram
MIC	Minimum inhibitory concentration
mL	Millilitre
ML	Machine learning
MLST	Multi-locus sequence typing
mM	Millimolar
NAAT	Nucleic acid amplification test
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
ng	Nanogram

nm	Nanometer
nt	Nucleotide
PBP	Penicillin-binding protein
PCA	Principal components analysis
PCR	Polymerase chain reaction
PCV	Pneumococcal conjugate vaccine
PDB	Protein Data Bank
PHAC-NML	Public Health Agency of Canada – National Microbiology Laboratory
RNA	Ribonucleic acid
RPM	Revolution per minute
SAVE	*Streptococcus pneumoniae* Serotyping and Antimicrobial Susceptibility: Assessment for Vaccine Efficacy in Canada
SNV	Single nucleotide variants
TAT	Turnaround time
VME	Very major error
WGS	Whole genome sequencing
WHO	World Health Organization
µg	Microgram
µL	Microlitre

1. INTRODUCTION

1.1 *Streptococcus pneumoniae*

1.1.1 Laboratory Isolation, Identification and Serotyping

Bacteria belonging to the genus *Streptococcus* are facultative anaerobic, catalase-negative, Gram-positive cocci that produce lactic acid as their major metabolic end product[1,2]. Current taxonomic considerations place *S. pneumoniae* in the *S. mitis* group of the larger viridans group streptococci[1,2]. Laboratory culture of *S. pneumoniae* requires an enriched media such as trypticase soy agar with 5% sheep blood (5%-SBA) and incubation at 35-37ºC in approximately 5% CO_2[1]. Colonies of *S. pneumoniae* typically appear small, grey and α-hemolytic due to partial degradation of the blood cells in the media by the pneumococcal toxin pneumolysin[1]. Following isolation, traditional phenotypic identification is based on solubility in bile (sodium deoxycholate) and sensitivity to optochin, thereby differentiating *S. pneumoniae* from other viridans group streptococci[1,2]. Some *S. pneumoniae* isolates demonstrate a mucoid colonial morphology as a result of increased capsular polysaccharide production[1]. When grown in liquid media, *S. pneumoniae* characteristically forms lancet-shaped pairs of diplococci.

Typing of *S. pneumoniae* is most commonly performed by research laboratories to detect and monitor changing serotype distributions within a geographic region. The gold standard phenotypic method for the typing of *S. pneumoniae* is the Quellung (Neufeld) reaction[1]. This reaction is based on the ability of the pneumococcal polysaccharide capsule to elicit serotype-specific immune responses[3]. Visible capsular swelling by phase-contrast microscopy following the addition of pooled, group-, type- or factor-specific antisera indicates a positive reaction[1]. To date, 97 distinct *S. pneumoniae* serotypes have been identified by this method[3].

Due to the significant cost burden associated with Quellung reagents, molecular typing systems such as multilocus sequence typing (MLST) have been established for *S. pneumoniae*[1].

1.1.2 Colonization and Transmission

S. pneumoniae is a common component of the normal bacterial flora within the human nasopharynx, as this is its major ecological niche[4]. Colonization of this region by *S. pneumoniae* occurs asymptomatically in healthy individuals following an acquisition event by the host[4-6]. Persistent colonization of the pneumococcus, termed carriage, is more prevalent in children than adults; the highest rates of carriage are demonstrated in children between 1-2 years of age and steadily declines to <10% carriage in children and adults >10 years of age[4,7,8]. In addition to age, carriage rates are influenced by various socioeconomic determinants, geography, host immune status, as well as environmental factors such as crowding and family size[4]. The duration of a carriage episode is also dependent upon a multitude of host- and organism-specific factors and can range anywhere from weeks to months[6].

Once *S. pneumoniae* enters the nasal cavity, it establishes itself on the surface of the mucous layer overlying the nasopharyngeal epithelium[5,6]. It then penetrates this layer and attaches to the epithelial surface by binding to host cell receptors[4-6]. This evasion of mucus-mediated clearance by *S. pneumoniae* is aided by phase variation, the molecular mechanism by which capsular polysaccharide production is regulated in response to changing host environments[5,6,8]. Phase variants with thinner capsules ("transparent variants") are selected for during the adherence stage of colonization[4-6,8]. The downregulation of capsular polysaccharide production promotes cell surface attachment by exposing pneumococcal cell surface proteins[4,6]. Conversely, phase variants with thicker capsules ("opaque variants") are selected for during the

transition through the mucous layer and tissue invasion[5,6,8]. These variants are more resistant to immune-mediated clearance and are generally more virulent[5,6,8].

Successful colonization of the nasopharynx is an essential pre-requisite for the horizontal transmission of pneumococcal strains[4,6,7]. *S. pneumoniae* is transmitted by direct contact with respiratory droplet secretions from carriers or symptomatic individuals[4,6,7]. Young children are recognized as the organism's major reservoir for spread due to the high carriage rates within this population[4]. The primary risk factor for increased transmission of *S. pneumoniae* is overcrowding; outbreaks of pneumococcal disease often occur in areas with increased close contacts such as day-care centers, hospitals, military bases and penitentiaries[4,9].

1.1.3 Clonal Evolution, Genetic Variation and the Genome of *S. pneumoniae*

The human nasopharynx is a highly diverse and dynamic niche that is home to a vast number of commensal microbial species[5]. Colonizing strains of *S. pneumoniae* are subjected to numerous selective pressures within this environment, such as the host's innate and adaptive immune responses, competitive interactions with other co-colonizing species and frequently, sub-lethal concentrations of antibiotics[5,10]. Factors such as these contribute to the selection and propagation of clonal pneumococcal lineages[10,11]. However, the primary driving force behind clonal evolution in *S. pneumoniae* is genetic recombination[12,11]. *S. pneumoniae* frequently exchanges its genetic material with other co-colonizing pneumococci and closely related streptococci species by way of natural transformation[10,11]. Natural transformation is the horizontal gene transfer mechanism by which competent bacteria uptake novel dsDNA from the external environment and integrate it into their genome through homologous recombination[10,11]. This mechanism significantly increases the genetic diversity within pneumococcal populations, thereby allowing for the selection of successful clones from the diverse background of genotypes

when faced with certain environmental pressures[11]. Clonal evolution in *S. pneumoniae* is further mediated by the availability of exogenous DNA for recombination[10]. Competent *S. pneumoniae* cells utilize the mechanism of fratricide to release chromosomal DNA from non-competent cells by allolysis, thereby increasing the gene pool within the nasopharynx[10,13].

The first whole genomes of *S. pneumoniae* were published in the early 2000s, revealing a circular chromosome of approximately 2.1 million base pairs (bp) of DNA with a guanosine-cytosine content of approximately 40%[11,14,15]. Overall, the pneumococcal genome contains upwards of 2000 gene coding regions depending on strain virulence[14-16]. *S. pneumoniae* uniquely demonstrates an overabundance of DNA repeat elements within its genome, including both insertion sequence (IS) elements and non-IS elements (BOX and RUP elements)[10,14,15]. Simple IS elements are present at high frequencies, accounting for approximately 5% of the genome[10,15]; however, they are often highly mutated and contain non-functional transposase enzymes[14]. Despite this, these elements serve an important function in *S. pneumoniae* by providing sites of homology for intra- and interspecies recombination[10,15]. Mobile genetic elements such as composite transposons that carry one or more antibiotic resistance genes are also commonly identified in the genomes of select pneumococcal strains[15].

Recent advancements and wide-spread use of next-generation sequencing technologies have led to an increased number of publically available pneumococcal whole genomes[11]. Comparative genomic analysis has shown that *S. pneumoniae* displays extensive genomic diversity and possesses a large and open pan-genome[10-12]. The core genome of *S. pneumoniae* accounts for approximately 75% of the total gene repertoire and is comprised of essential genes involved in the basic biological functioning of the species[11]. The accessory (dispensable) genome makes up the remaining portion of the pneumococcal pan-genome concept and

encompasses all non-essential genes[11,12]. Despite their non-essentiality, these genes often confer certain selective advantages such as antimicrobial resistance and increased niche adaptation[11]. *S. pneumoniae* frequently alters the contents of its accessory genome through horizontal gene transfer; therefore, genomic diversity is primarily due to variation in the accessory genome[11,12]. Pan-genome analysis of multiple *S. pneumoniae* strains revealed that closely related *S. mitis* was the major contributor to the increased genetic variability within the pneumococcal accessory genome[12].

1.1.4 Pathogenesis and Invasive Pneumococcal Disease

S. pneumoniae is capable of causing a variety of clinical syndromes, including upper and lower respiratory tract infections, as well as disseminated invasive infections[5,9,17]. Progression from nasopharyngeal colonization to pneumococcal disease occurs following bacterial penetration of the host's respiratory tissues[6]. The pneumococcal capsule is recognized as the critical determinant in the progression to disease[5]. The inherent invasive capabilities of different *S. pneumoniae* strains is a serotype-dependent trait, where strain virulence and capacity to cause disease are directly associated with capsular thickness[5]. Capsule production is an essential immune defense mechanism employed by the pneumococcus that inhibits opsonophagocytosis by preventing interactions with phagocytic cell receptors[5,6,17]. Tissue invasion by *S. pneumoniae* is facilitated further by numerous other serotype-independent factors that aid in host immune evasion, induce inflammatory immune responses or directly damage host tissues[6,9,17].

S. pneumoniae is a common etiologic agent of non-invasive mucosal infections such as acute otitis media (AOM), sinusitis and pneumonia[9,17]. The milder forms of disease (AOM and sinusitis) develop following contiguous spread by colonizing strains in the nasopharynx[9,17,18]. It

is estimated that *S. pneumoniae* is responsible for up to 50% of all cases of AOM globally[18]. Incidences of AOM caused by *S. pneumoniae* most commonly occur in early childhood, corresponding to peak rates of carriage[18]. Conversely, pneumonia is the most common manifestation of pneumococcal disease in adults, causing up to 30% of all community-acquired pneumonia cases[9,17,19]. Pneumococcal pneumonia develops following the extension of colonizing strains in the nasopharynx into the lower respiratory tract through aspiration[9,17]. The World Health Organization (WHO) states that *S. pneumoniae* is the most frequent cause of severe pneumonia and is responsible for the most pneumonia-related deaths world-wide[20].

Invasive pneumococcal disease (IPD) is the most severe and life-threatening manifestation of disease and is defined as the isolation of *S. pneumoniae* from a normally sterile body site[17]. Despite current vaccination efforts, IPD remains a significant cause of global morbidity and mortality, particularly in children <2 and adults ≥65 years of age[17,19,21,22]. In addition to age, other host-specific risk factors for IPD include patient comorbidities and immune deficiencies, as well as smoking and increased contact with children in immunocompetent adults[17]. The most important organism-specific factor for the development of IPD is serotype; of the greater than 90 pneumococcal serotypes identified, a small subset of serotypes demonstrate increased invasive potential and are associated with the majority of IPD cases[9,17].

The route to IPD development is a complex process requiring several steps that ultimately result in the presence of *S. pneumoniae* in the bloodstream[6]. Pneumococcal bacteremia is the most prevalent type of IPD in young children and often develops without an identifiable focus of infection[9,19]. Conversely, bacteremia is a common sequela of pneumococcal pneumonia in adults, occurring in up to 30% of cases[17,19]. Another critical form of IPD is meningitis; pneumococcal meningitis occurs following haematogenous dissemination and

invasion of the meninges by *S. pneumoniae*[9,6]. A significant factor in the development of meningitis is prolonged pneumococcal bacteremia with a high bacterial load[9]. *S. pneumoniae* is a leading cause of bacterial meningitis in both children and adults, with case-fatality rates of approximately ≤10% and up to 25% in the respective populations[9,19]. Importantly, *S. pneumoniae* is now the primary cause of bacterial meningitis in children in the United States due to the success of the *Haemophilus influenzae* type b (Hib) vaccine[19]. Additionally, other sterile sites such as the pleura, pericardium, peritoneum and joints may be infected by *S. pneumoniae* following haematogenous spread[17].

1.2 Treatment of IPD

Prompt initiation of therapy with the appropriate antimicrobial agent is of critical importance in cases of IPD due to the severity of the illness[23]. For patients presenting to the hospital with suspected IPD, combination empiric therapy is initiated, usually using two antimicrobial agents with known anti-pneumococcal activity, such as a 3rd generation cephalosporin plus a macrolide[23]. Due to the lack of available rapid diagnostic tools for bacterial identification and antimicrobial susceptibility testing (AST), empiric therapy should be selected based on knowledge of regional antimicrobial resistance trends and should cover all potential etiologic agents[23,24]. Once a positive direct Gram-stain confirms an invasive bacterial infection, therapy should be adjusted according to the morphology reported by the laboratory. Next, antimicrobial therapy should again be adjusted based on the *in vitro* identification and AST results of the cultured pathogen[25]. In cases of confirmed pneumococcal bacteremia and meningitis, standard directed therapy is based on the organism's susceptibility to penicillin[25,15]. For penicillin-susceptible IPD isolates, the standard directed therapy is penicillin G or ampicillin with the discontinuation of other empirically prescribed antimicrobial agents[25,15]. Conversely,

for IPD isolates that are both penicillin- and cephalosporin-resistant by CLSI IV meningitis breakpoints, the continuation of empiric therapy is recommended[25,15]. Alternative therapies such as fluoroquinolones and carbapenems may be considered in cases of IPD caused by penicillin-resistant strains[25].

1.3 β-lactam Resistance in *S. pneumoniae*

1.3.1 β-lactam Antibiotics and Penicillin Binding Proteins (PBPs)

β-lactams are a family of commonly prescribed bactericidal agents that can be used to treat a variety of bacterial infections[26]. Members of this diverse family are characterized by a β-lactam ring in their core molecular structure[26,27]. The bactericidal activity of β-lactams is a time-dependent characteristic; contingent on the class of β-lactam selected for treatment, the free drug serum concentration should be maintained at a level above the minimum inhibitory concentration (MIC) for approximately 40-70% of the dosing interval to ensure maximal killing[28]. Their common mode of action is inhibition of cell wall synthesis, which ultimately results in cell lysis[24,29].

Treatment of pneumococcal infections has long been dependent upon β-lactams, starting in the early 1940s with the discovery of penicillin[9,24]. The molecular structure of penicillin is relatively simple, consisting of a 4-membered β-lactam ring fused to 5-membered thiazolidine ring, forming the penam core[27]. In the decades following discovery, penicillin was regarded as the treatment of choice for pneumococcal infections due to the organism's uniform susceptibility[9,24]. Subsequent to the large-scale production of penicillin, other novel β-lactam classes were developed that were also highly effective in treating pneumococcal infections[24]. One such class was the cephalosporins; their core molecular structure is comprised of a 6-

membered dihydrothiazine ring fused to the β-lactam ring[27]. Members of this class are commonly in use today for the empiric and directed treatment of IPD, particularly 3rd-generation cephalosporins such as ceftriaxone[23,25]. Another class of β-lactam less commonly used for the treatment of pneumococcal infections are the carbapenems. Carbapenems demonstrate the broadest spectrum of activity of all β-lactams owing to their extensively modified penam structure[27]. These agents are selected for the treatment of pneumococcal infections caused by penicillin-resistant strains[26].

The natural cell wall target of the β-lactam antibiotics is the appropriately named PBP[29,30,31]. PBPs are a diverse family of membrane-bound enzymes with essential roles in peptidoglycan biosynthesis[30,31]. These enzymes are initially classified on the basis of molecular weight into high molecular mass (HMM) PBPs and low molecular mass (LMM) PBPs[30,31]. All HMM PBPs catalyze the transpeptidation (TP) reaction between two nascent glycan stands, yielding cross-linked peptidoglycan[29-31]. Their tertiary structure broadly consists of joined N- and C-terminal domains that are anchored to the cytoplasmic membrane by small transmembrane and cytoplasmic segments[31]. HMM PBPs are further subdivided into either class A or B according to the function of their N-terminal domain[30,31]. Class A HMM PBPs are bifunctional enzymes that demonstrate both glycosyltransferase (GT) and TP activity in their N- and C-terminal domains, respectively[30,31]. The N-terminal GT domain of class A HMM PBPs functions in earlier stages of peptidoglycan synthesis and is responsible for the polymerization of important lipid intermediates into uncross-linked glycan strands[30]. Conversely, the N-terminal domain of class B HMM PBPs is believed to be involved in other cellular functions and is generally not associated with penicillin binding[30,31]. Comparative genomic analysis has revealed

that all PBPs contain three conserved sequence motifs within their TP domain active site: SXXK, containing the catalytic serine residue, (S/Y)XN and (K/A)(S/T)G[29-32].

β-lactams are mimetic molecules that inhibit bacterial cell wall synthesis by directly binding to the TP domain active site of PBPs[29-33]. Their overall structure resembles that of the C-terminal D-Ala-D-Ala dipeptide substrate of precursor glycan strands, thereby enabling direct antibiotic access to the active site region[29-33]. Following antibiotic entry, the β-lactam ring is subject to nucleophilic attack by the catalytic serine residue buried deep within the active site, resulting in the formation of a biologically inactive acyl-enzyme complex[29-33]. Complex formation results in irreversible inhibition of PBP TP activity, thus, preventing further peptidoglycan cross-linking[30-33]. The kinetics of the reaction is dependent on both PBP and β-lactam type, respectively[29,32].

1.3.2 Pneumococcal PBPs

Both the overall number and the types of PBPs that a bacterium possesses is highly variable amongst different species[31]. To date, a total of six unique PBPs have been identified in *S. pneumoniae*, with representatives from each of the aforementioned PBP classes: three class A HMM PBPs (PBP1a, PBP1b and PBP2a); two class B HMM PBPs (PBP2x and PBP2b); and a single class C LMM PBP (PBP3)[29,31- 33]. Gene inactivation studies revealed that both PBP2x and PBP2b (class B HMM) are required for *in vitro* growth[24,29,32,34]. Conversely, the three class A PBPs were shown to be individually non-essential under laboratory conditions and can each be inactivated without notable changes to growth or morphology[29,32,35]. However, either PBP1a or PBP2a, in combination with the class B HMM PBPs, are required for growth *in vitro* since *pbp1a/pbp2a* double-disruption mutants are nonviable[29,32,35]. Despite their differences in

essentiality, both class A and B pneumococcal PBPs are associated with various important cell division and growth processes[29,31].

1.3.3 Mechanism of β-lactam Resistance in *S. pneumoniae*

β-lactam-resistance in *S. pneumoniae* was first noted in the early 1970s, subsequent to widespread β-lactam use[9,24]. Incidences of pneumococcal disease caused by β-lactam-resistant strains steadily increased thereafter due to the global dissemination of a number of successful β-lactam-resistant clones[24]. Unlike many Gram-negative pathogens which produce β-lactamases as their major mechanism of resistance, β-lactam resistance in *S. pneumoniae* is exclusively based on target site modification, resulting in the expression of low-affinity PBPs[24,29,32]. The amino acid sequences of these low-affinity PBP variants display numerous substitutions, mainly concentrated within the TP domain active site[24,29,32]. This accumulation of mutations leads to overall changes in the physicochemical properties of the enzyme, such as its charge, polarity and flexibility, thereby affecting antibiotic active site access and binding[24].

Of the six previously described pneumococcal PBPs, three are shown to be consistently altered in clinically obtained isolates of *S. pneumoniae* with reduced susceptibility to β-lactams, namely PBP2x, PBP2b and PBP1a[24,29,32]. The two class B PBPs (PBP2x and PBP2b) are essential to the resistance process and are commonly implicated together in causing lower levels of β-lactam resistance[24,29,32]. Despite being of the same class, PBP2x and PBP2b differ in their affinities for select β-lactam classes; it has been demonstrated that PBP2x is capable of binding most β-lactams *in vitro*, whereas PBP2b preferentially binds penicillins over cephalosporins[24,29,32,36-39]. Conversely, higher levels of β-lactam resistance are associated with highly altered PBP1a proteins and are therefore considered the most clinically important *S.*

pneumoniae PBP type[24,29,32,36,39]. PBP1a mediated β-lactam resistance requires one or both of the class B PBPs to be altered in combination with altered PBP1a[24,29,32].

1.3.4 Genetic Exchange of β-lactam Resistance Determinants

Numerous combinations of mutations within PBP2x, PBP2b and PBP1a have been identified in β-lactam-non-susceptible clinical isolates of *S. pneumoniae* and are associated with wide-ranging MIC values[24]. Sequencing has revealed that low-affinity variants of the three PBPs are encoded by mosaic genes containing large, highly variable blocks of sequence[24,29,32]. These divergent sequence blocks are often not of pneumococcal origin and are acquired through multiple recombination events within the nasopharynx[24,29,32]. During colonization, closely related commensal streptococcal species develop resistance-conferring point mutations as a result of β-lactam selective pressure[24,32]. *S. pneumoniae* then freely exchanges its genetic material with these β-lactam-non-susceptible donors through homologous recombination, resulting in the incorporation of these highly variable PBP gene fragments into the pneumococcal genome[24,32]. Through genetic recombination analysis, it has been determined that most *S. pneumoniae* interspecies PBP-associated DNA exchange events involve *S. mitis* and *S. oralis*[24,40,41]. Mosaic *pbp2x*, *pbp2b* and *pbp1a* sequences can differ from non-mosaic alleles of wild-type susceptible pneumococcal strains by up to 20%, depending on the level of β-lactam non-susceptibility of a particular isolate[29].

In addition to the closely related *Streptococcus mitis* group, other pneumococci can serve as DNA donors for the genetic exchange of β-lactam resistance-conferring loci[24,29,32]. Of particular importance are the intraspecies recombination events involving the *cps* (capsular polysaccharide) locus[24,29]. The *cps* locus in *S. pneumoniae* encodes multiple genes involved in capsular biosynthesis and its components differ according to serotype[5]. Intraspecies exchange of

the *cps* locus occurs in response to selective vaccine pressure and can result in a serotype switch event[24]. Large fragments of variant PBP1a and PBP2x genes can be transferred in a single capsular switch event between a β-lactam-nonsusceptible donor and a β-lactam-susceptible recipient due to the proximity of these genes to the *cps* locus[24,29]. By this mechanism, the resultant progeny strains are of a different serotype and demonstrate a unique β-lactam resistance profile[24,29]. Following the introduction of the first pneumococcal conjugate vaccine (PCV), PCV-7, incidences of IPD caused by penicillin-resistant non-PCV-7 serotypes increased due to the emergence and global dissemination of a number of these vaccine escape recombinants[24].

1.4 Detection of β-lactam Resistance in *S. pneumoniae*

1.4.1 Phenotypic Antimicrobial Susceptibility Testing (AST) Methods

Phenotypic AST is routinely performed in clinical microbiology laboratories to determine the *in vitro* susceptibility of a bacterial pathogen following challenge with an antimicrobial agent[42]. Multiple culture-based AST methods have been developed that are based on the principles of either antimicrobial diffusion or dilution using solid and liquid media[43]. Kirby-Bauer disc diffusion and gradient diffusion are commonly performed diffusion methods involving the placement of an antibiotic-containing disc (or strip), respectively, on the surface of solid agar media following the inoculation of a standardized bacterial suspension[43,44]. The disc diffusion method produces a zone of bacterial growth inhibition that is measured and interpreted qualitatively as susceptible, intermediate and resistant using standardized breakpoint criteria[43,44].

Another commonly performed phenotypic AST method is broth microdilution, involving the inoculation of a standardized bacterial suspension into microtitre wells containing doubling dilutions of select antimicrobial agents in liquid media[44]. Unlike the disc diffusion method,

quantitative results in the form of MIC values are produced by broth microdilution; MICs are defined as the lowest concentration of the tested antimicrobial that completely inhibits visible growth of the organism[42-44]. Broth microdilution is recognized as the gold standard AST method by the Clinical and Laboratory Standards Institute (CLSI) and is the suggested reference method for the evaluation of new antimicrobial susceptibility tests[45].

Antimicrobial resistance is detected in clinically obtained isolates of *S. pneumoniae* by the previously described phenotypic AST methods. Results from phenotypic AST are reported to physicians by the clinical microbiology laboratory to guide clinical decision making for the treatment of pneumococcal infections[24]. Phenotypic AST is recognized as the gold standard for the determination of antimicrobial susceptibility due to the ease of method performance and the availability of published standards developed by professional organizations such as CLSI and the European Committee on AST (EUCAST)[24,42,46,47]. These methods are commonplace in public health laboratories for the detection and monitoring of emerging antimicrobial resistance trends in bacterial pathogens such as *S. pneumoniae*[42,24].

Despite widespread implementation, the major limitation of phenotypic AST methods is their culture-dependence, resulting in prolonged turn-around times (TATs)[24,46]. The requirement of a pure, freshly grown bacterial culture prohibits AST of currently nonculturable bacteria by these methods[42]. For culturable organisms, TATs for phenotypic AST may extend beyond the normal 48 hours due to mixed or insufficient bacterial growth following the direct culture of clinical specimens or if additional AST is required for multidrug-resistant organisms[46,47]. Phenotypic AST of *S. pneumoniae* is particularly challenging due to the organism's autolytic tendencies in culture[24]. Such delays in testing can adversely affect patient treatment outcomes, leading to extended hospitalizations and increased health-care-associated costs[47]. Additionally,

delayed reporting of AST results encourages the extended administration of empirically prescribed broad-spectrum antimicrobial agents. This unnecessary use of antimicrobials is a major contributing factor to the growing global antimicrobial resistance crisis[46]. Therefore, the development of rapid diagnostic tools for bacterial AST with improved time-to-results is of increasing clinical importance[47].

1.4.2 Whole Genome Sequencing (WGS)-based AST

Recent advancements in next-generation sequencing technologies, plus the overall reduction in sequencing costs have driven the development of genotypic AST methods[42,48]. At present, most genotypic methods are based on the detection and amplification of a few select antibiotic resistance genes[46,48]. These nucleic acid amplification test (NAAT)-based methods are limited to a certain number of target genes per assay and require *a priori* knowledge of the genetic mechanisms of resistance, thereby increasing the likelihood of false-negatives[46,48]. In recent years, WGS has emerged as a viable alternative approach to AST with the potential to overcome the major limitations associated with other phenotypic and genotypic AST methods[46,48,49]. The sequencing of an organism's whole genome not only allows access to its entire repertoire of resistance genes (resistome) but also provides important information relating to its epidemiology and virulence[46].

Currently, two approaches are commonly described in published literature relating to WGS-based AST: (1) rules-based approaches and (2) model-based approaches[46,48]. Both aim to accurately predict antimicrobial resistance (AMR) phenotypes from genomic data. Rules-based approaches attempt to predict phenotypic AST results based on the presence or absence of known AMR genes or resistance-conferring mutations that are detected in genetic sequences using either multispecies or species-specific AMR databases[46,48]. By this approach, AMR genes

are most commonly correlated with categorical interpretations of susceptibility[48]. Conversely, model-based approaches employ machine learning (ML) or statistical models to determine which genetic features in the dataset are significantly associated with the resistant phenotype[46,48]. By this approach, the model classifier is trained using a designated set of genomes with previously determined phenotypic AST results[46,48]. Multiple classification algorithms can be used for model training purposes depending on the type of genomic input data available. Additionally, both WGS-based AST approaches can be combined into a single "rules-based model", whereby the model classifier is trained on a set of known resistance-conferring genetic determinants[48].

WGS-based AST has recently been the subject of intense research and has been well-studied in several bacterial species[46,48,49]. To date, the majority of these studies have been performed using second-generation (short-read) sequencing technologies, demonstrating varying degrees of success for pathogens such as *Mycobacterium tuberculosis, Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa, Salmonella* spp., *Staphylococcus aureus, Neisseria gonorrhoeae* and *Enterococcus* spp[46,48,49]. Recently, different WGS-based models have been applied to *S. pneumoniae* for the prediction of resistance to agents commonly used in the treatment of pneumococcal infections[46,48]. Most notably was the successful application of a model-based PBP typing scheme developed by Metcalf *et al*. for the prediction of β-lactam resistance in *S. pneumoniae*[50,51].

The major advantage of second-generation sequencing technologies such as the Illumina platform is their high accuracy and low error rates, thereby enabling SNP level analysis; however, they are limited in their detection of DNA repeat regions, often resulting in fragmented genome assemblies[46]. This is particularly problematic for WGS-based AST since DNA repeat regions are commonly found flanking chromosomal and plasmid-associated AMR genes[46,49]. As

a result, third-generation sequencing technologies that produce reads of longer lengths are increasingly being used for AMR gene detection[46]. Importantly, however, resistance-conferring point mutations cannot be reliably detected by these newer technologies since they are more error-prone than short-read sequencing[46]. In addition to the inherent limitations associated with the different sequencing technologies, several other challenges of WGS-based AST have been identified. In a recent review on the subject, it was concluded that at the time of publication, WGS-based AST should not replace traditional phenotypic methods due to the lack of published evidence for most bacterial species[49]. Other noteworthy limitations of WGS-based AST include insufficient knowledge of the mode of action and genetic basis of resistance for several antibiotics, inability to detect novel mechanisms of resistance, lack of standardized quality control metrics and bioinformatics pipelines, as well as the requirement of sample batching to reduce costs[46,48,49]. These factors together limit the clinical utility of WGS as a tool to infer the antimicrobial susceptibility of important bacterial pathogens such as *S. pneumoniae*[49].

1.5 The SAVE Study

1.5.1 Background on PCV use in Canada

Vaccination with PCVs has proven to be an effective strategy for reducing IPD caused by vaccine serotypes both directly and indirectly in pediatric and adult populations, respectively[52-54]. In Canada, routine infant vaccination with PCVs began in the early 2000's with PCV-7, a seven-valent formulation that offers protection against common pediatric IPD serotypes 4, 6B, 9V, 14, 18C, 19F and 23F[52]. Introduction of PCV-7 significantly reduced the overall burden of IPD; however, it was later replaced by the higher-valent formulation PCV-13 (Prevnar®13) in response to the subsequent increase in IPD caused by non-vaccine serotypes[52,53]. In 2010, PCV-13 was licensed for use in Canada and successively replaced PCV-7 in routine infant vaccination

programs across all provinces and territories by mid-2011[55]. PCV-13 was recommended for use in Canada as it was the PCV formulation with the maximum serotype coverage available, offering extended coverage against PCV-7 serotypes, plus 1, 3, 5, 6A, 7F and importantly 19A[56]. Following the introduction of PCV-13 in Canada, the incidence of IPD in all age groups decreased from 9.8 to 9.0 cases per 100,000 population between 2009 and 2015[57].

1.5.2 Premise and Study Design

The SAVE (*Streptococcus pneumoniae* Serotyping and Antimicrobial Susceptibility: Assessment for Vaccine Efficacy in Canada) study is an ongoing national surveillance program that was initiated in 2011 following the introduction of PCV-13 in Canada. This study is carried out through a collaboration between the Canadian Antimicrobial Resistance Alliance (CARA) and Public Health Agency of Canada – National Microbiology Laboratory (PHAC-NML). Invasive isolates of *S. pneumoniae* were submitted to the Streptococcus and STI Unit at the PHAC-NML for serotyping by provincial public health and hospital laboratories who are part of the Canadian Public Health Laboratory Network (CPHLN). As part of the collaboration, IPD isolates from eight participating CPHLN sites (Saskatchewan, Manitoba, Ontario, Quebec, Nova Scotia, Prince Edward Island, Newfoundland and Labrador, and a portion of isolates from New Brunswick) were forwarded by PHAC-NML to CARA for antimicrobial susceptibility testing. Prior to sending, PHAC-NML confirmed the identification of the submitted isolates as *S. pneumoniae* using bile solubility and optochin tests, as well as performed serotyping by the Quellung reaction using commercially prepared pool, group, type and factor specific antisera (Statens Serum Institute, Copenhagen, Denmark). Molecular testing was performed using *rpoB* gene sequencing on isolates that were phenotypically nontypeable. In total, 7622 invasive *S. pneumoniae* isolates collected as part of the SAVE study between 2011 and 2016 were

forwarded to CARA for antimicrobial susceptibility testing. The annual isolate totals were 1379 from 2011, 1285 from 2012, 1138 from 2013, 1274 from 2014, 1262 from 2015 and 1284 from 2016.

.

2. RATIONALE AND OBJECTIVES

S. pneumoniae is a clinically important bacterium responsible for significant global morbidity and mortality despite vaccine availability. This organism is commonly isolated from the nasopharynx of asymptomatic human hosts as this is its primary ecological niche; however, virulent strains can disseminate beyond this region, causing severe respiratory and invasive infections. The severity of illness necessitates prompt initiation of therapy for IPD patients with effective antimicrobial agents. Treatment of IPD has long been reliant on penicillin and other β-lactams due to the organisms' uniform susceptibility; however, resistant *S. pneumoniae* strains emerged subsequent to widespread β-lactam use and disseminated globally thereafter.

Resistance to β-lactams in *S. pneumoniae* is attributable to highly modified PBP target proteins that demonstrate decreased affinities for these agents, namely PBP2x, PBP2b and PBP1a. Detection of β-lactam resistance in clinical *S. pneumoniae* isolates routinely involves gold standard phenotypic AST methods. These methods are limited by their culture-dependence, resulting in prolonged TATs that can delay the time to effective therapy for IPD patients. In recent years, WGS as a diagnostic tool for bacterial identification and AST has been the subject of intense research, offering decreased time to results compared to standard AST methods. Future use of WGS to rapidly and accurately predict *S.pneumoniae* β-lactam MICs could improve treatment outcomes for IPD patients through the optimization of pharmacotherapy.

The purpose of this thesis was to determine if WGS-based AST models can be used to accurately predict *S. pneumoniae* β-lactam MIC values. The primary objective of this work was to develop and validate multiple linear regression models predicting MICs of three β-lactams commonly used for the empiric and directed treatment of IPD, including penicillin, ceftriaxone and meropenem. To address this objective, a cohort of invasive *S. pneumoniae* isolates collected

as part of the SAVE study were selected for WGS. Previously reported PBP-associated genetic determinants conferring reduced β-lactam susceptibility in *S. pneumoniae* were identified as potential predictors of β-lactam MIC and were detected in isolate PBP2x, PBP2b and PBP1a amino acid sequences. Predictive models were generated by the forward multiple linear regression method[98] and were subsequently validated using leave-one-out cross-validation[99]. Predicted β-lactam MICs were compared to phenotypic AST values measured by broth microdilution to evaluate model performance. A secondary objective of this study was to determine if additional phenotypic AST would be required for the implementation of our best-determined β-lactam MIC models in clinical microbiology laboratories. This objective was addressed in our evaluation of model performance while taking into account differences in time to results between WGS-based AST and phenotypic AST methods.

3. HYPOTHESIS

The hypothesis of this study is that WGS, as a tool to infer antimicrobial susceptibilities, cannot be used exclusively in the determination of *S. pneumoniae* β-lactam MIC values routinely used to guide clinical decision making for the treatment of IPD.

4. MATERIALS AND METHODS

4.1 AST

4.1.1 Preparation of Antimicrobial Solutions

Concentrated antimicrobial stock solutions were prepared in accordance with CLSI methods using laboratory grade powders[58]. The CLSI broth macrodilution method was performed to confirm the activity of the prepared antimicrobial solutions; dilute stock solutions were tested against *S. pneumoniae* ATCC® 49619, *Staphylococcus aureus* ATCC® 29213, *Enterococcus faecalis* ATCC® 29212, *Escherichia coli* ATCC® 25922 and *Pseudomonas aeruginosa* ATCC® 27853 quality control strains.

4.1.2 Broth Microdilution

AST was performed on all invasive *S. pneumoniae* isolates collected as part of the SAVE study by the broth microdilution method as described by CLSI M07[58]. Custom-designed, 96-well microtitre panels were prepared in-house and contained doubling dilutions of the following antimicrobial agents in 100 µL of cation-adjusted MHB supplemented with 4% lysed horse blood: penicillin, ceftriaxone, cefuroxime, clarithromycin, clindamycin, telithromycin, levofloxacin, moxifloxacin, linezolid, trimethoprim-sulfamethoxazole, doxycycline, tigecycline, chloramphenicol and vancomycin.

Standardized bacterial inoculums were prepared in sterile water using the CLSI colony suspension method following two consecutive subcultures from frozen stock on 5%-SBA. The inoculum suspensions were adjusted to a 0.5 McFarland turbidity standard (1-2 x 10^8 CFU/mL) and diluted in sterile water to achieve a final concentration of approximately 5 x 10^5 CFU/mL in each microtiter well. Purity checks of the inoculum suspensions were performed for all isolates.

4.1.3 MIC Interpretive Criteria

Following 20-24 hours of incubation at 35˚C in 5% CO_2, MICs were determined for the *S. pneumoniae* isolates as the lowest concentration of the antimicrobial agent that completely inhibits visible organism growth. MICs were interpreted as susceptible, intermediate or resistant using CLSI M100-S28 (2018) breakpoint criteria[59].

4.2 Isolate Selection

A total of 7622 invasive *S. pneumoniae* isolates were collected as part of the SAVE 2011-2016 study. From the total, a subset of 196 isolates was selected for this study based on their penicillin MIC: 10% of all isolates for each penicillin MIC value, ranging from 0.06 to 4 µg/mL, were selected using a random number generator. The original isolate totals per penicillin MIC were: 36 from 0.06 µg/mL, 26 from 0.12 µg/mL, 18 from 0.25 µg/mL, 8 from 0.5 µg/mL, 9 from 1 µg/mL, 11 from 2 µg/mL and 9 from 4 µg/mL. All SAVE isolates with a penicillin MIC of 8 µg/mL were included (3 total). To achieve sufficient representation of wild-type susceptibility, the remaining isolates were randomly selected from isolates with penicillin MICs ≤0.03 µg/mL (76 total).

AST was repeated in duplicate for all *S. pneumoniae* study isolates with penicillin MICs ≥0.06 µg/mL, ceftriaxone MICs ≥0.25 µg/mL and meropenem MICs ≥0.12 µg/mL. Following repeat testing, the new isolate totals per penicillin MIC value were: 78 from ≤0.03 µg/mL, 34 from 0.06 µg/mL, 28 from 0.12 µg/mL, 11 from 0.25 µg/mL, 10 from 0.5 µg/mL, 9 from 1 µg/mL, 9 from 2 µg/mL, 16 from 4 µg/mL and 2 from 8 µg/mL.

4.3 AST of β-lactam-Susceptible Isolates

AST was performed by the broth microdilution method as previously described (outlined in Section 4.1) on all β-lactam-susceptible *S. pneumoniae* isolates with MICs below previously determined quantification limits to establish true MIC values for all study isolates. MICs were determined in triplicate for isolates with penicillin MICs ≤0.03 µg/mL, ceftriaxone MICs ≤0.12 µg/mL and meropenem MICs ≤0.06 µg/mL. Custom-designed, 96-well microtitre panels containing doubling dilutions of penicillin (0.001-1 µg/mL), ceftriaxone (0.001-0.25 µg/mL) or meropenem (0.001-0.25 µg/mL) were prepared in accordance with CLSI M07. Quality control testing was performed using *S. pneumoniae* ATCC® 49619 strain. The updated and expanded isolate MIC distributions for penicillin, ceftriaxone and meropenem following all repeat broth microdilution testing are shown in Figures 4.1, 4.2 and 4.3, respectively.

FIGURE 4.1: Expanded penicillin MIC distributions of 196 invasive *S. pneumoniae* isolates following triplicate AST by broth microdilution

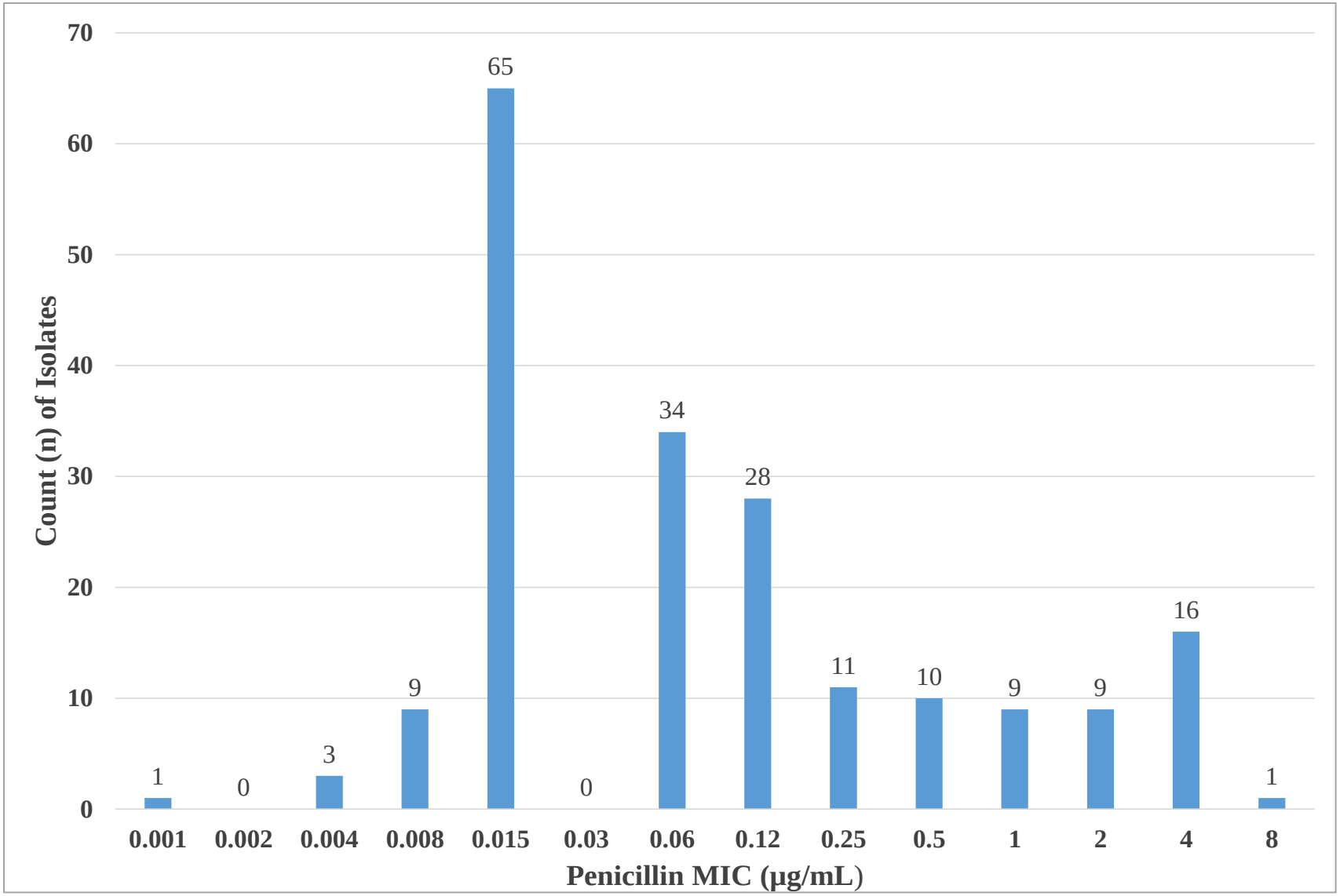


FIGURE 4.2: Expanded ceftriaxone MIC distributions of 196 invasive *S. pneumoniae* isolates following triplicate AST by broth microdilution

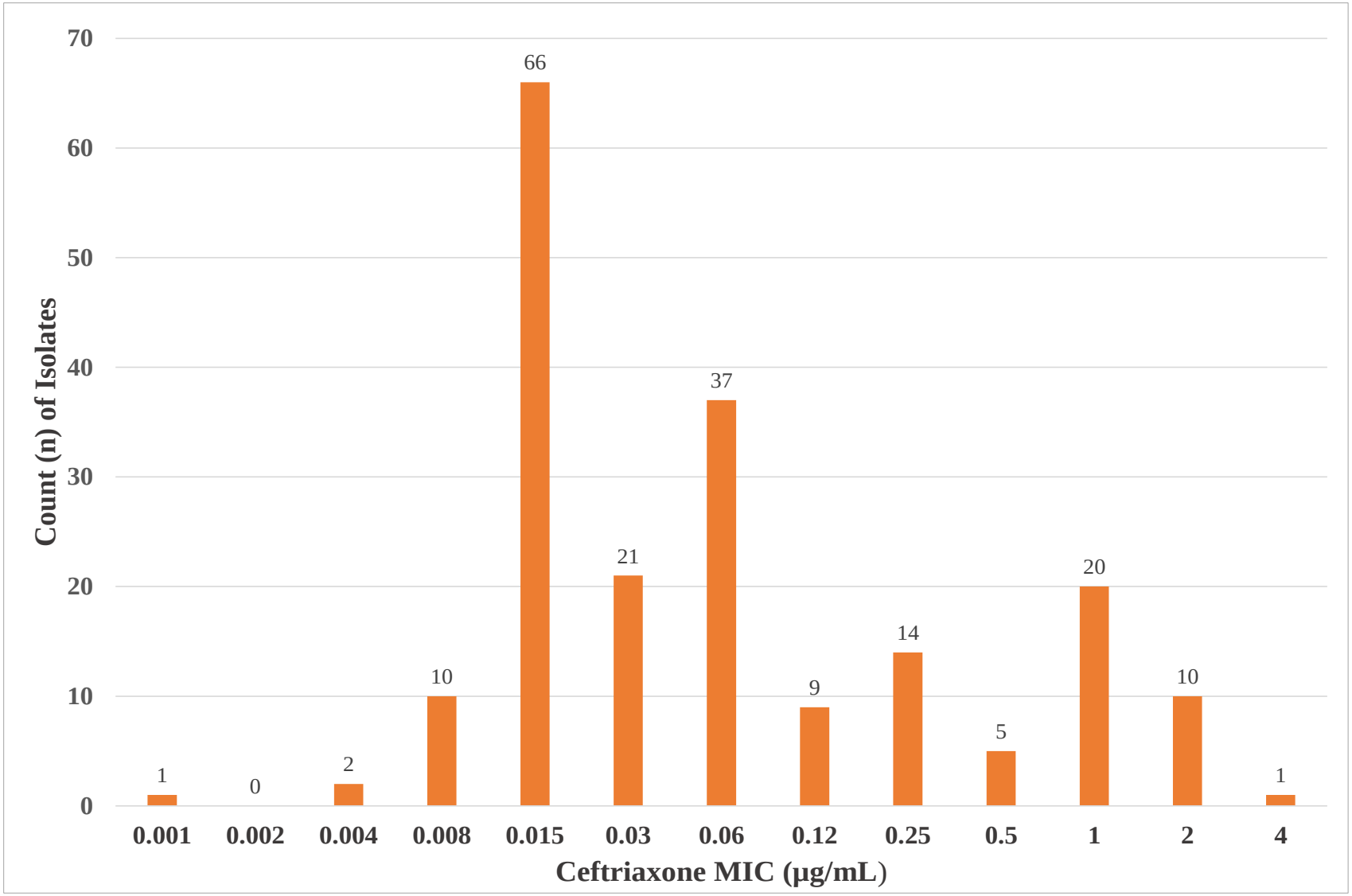
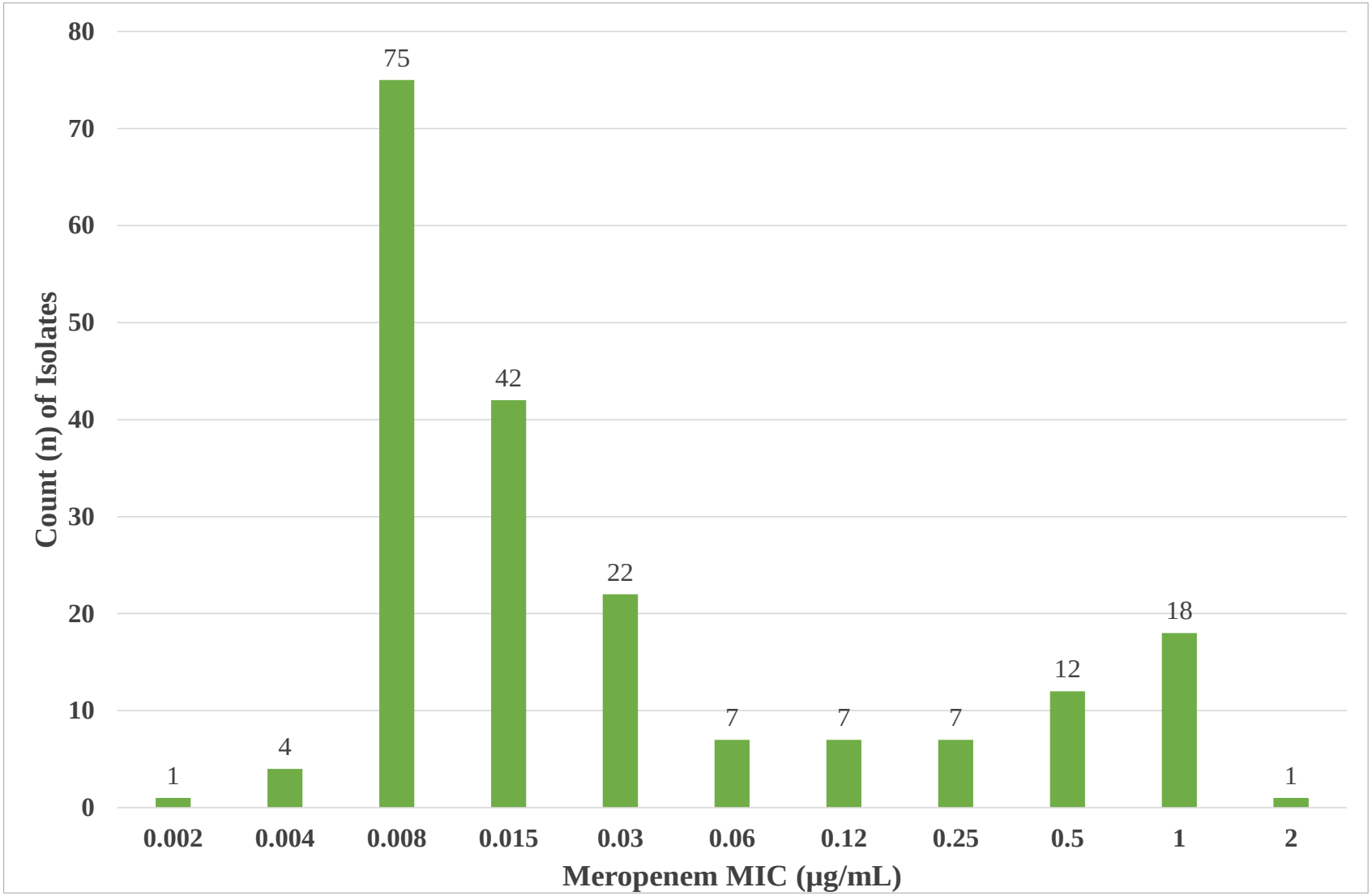


FIGURE 4.3: Expanded meropenem MIC distributions of 196 invasive *S. pneumoniae* isolates following triplicate AST by broth microdilution



4.4 WGS

Of note, four of the 196 *S. pneumoniae* isolates selected for this study were previously sequenced by the Streptococcus & STI Unit at PHAC-NML. The raw sequence data for these four isolates was forwarded by our colleagues following WGS of the remaining 192 study isolates. The methods described below are relevant for all 196 study isolates.

4.4.1 DNA Extraction and Quantification

The Epicenter MasterPure™ Complete DNA and RNA Purification Kit was used to extract high-quality DNA from *S. pneumoniae* isolates. First, a lysis solution was prepared for each isolate containing 300 µL of 2X T&C Lysis Solution, 2 µL of Proteinase K (50 µg/mL), 2 µL of ReadyLyse Lysozyme (20,000U/µL) and 2 µL of Mutanolysin (10U/µL, Sigma Aldrich). A loop full of fresh bacterial culture was suspended in the prepared solution and vortexed for 10 seconds. Bacterial suspensions were incubated in a dry bath incubator at 56ºC for 60 minutes with constant mixing at 1000 RPM. Lysed samples were cooled to 37ºC, mixed with 5 µL of RNase A (100 mg/mL), and incubated at 37ºC for 15 minutes with constant mixing at 600 RPM. Prior to protein precipitation, sample lysates were cooled at -80ºC for 1 minute. Cooled lysates were then mixed with 150 µL of MPC Protein Precipitation Reagent and vortexed vigorously for 5 seconds. Next, 300 µL of 100% EtOH was added to each tube and vortexed for 3 seconds to precipitate DNA in sample lysates.

The Qiagen QIAamp DNA Mini Kit was used to isolate and purify high-quality DNA from sample lysates. Sample lysates were applied to spin columns with attached collection tubes and centrifuged at 14,000 RPM for 1 minute. Spin columns containing sample DNA were transferred to clean collection tubes, washed with 500 µL of Buffer AW1, and centrifuged at

14,000 RPM for 1 minute. A second wash step was performed similarly using 500 µL of Buffer AW2. Column membranes were dried in clean collection tubes by centrifugation at 14,000 RPM for 1 minute. Next, 70 µL of heated (60ºC) TE Buffer was applied to the dried membranes, incubated at room temperature for 1 minute and centrifuged at 9,700 RPM for 1 minute to elute sample DNA from the columns.

Sample DNA was quantified using the Qubit® 4 Fluorometer and Qubit® dsDNA BR Assay kit (Invitrogen) to confirm that a minimum of 5 ng/µL of DNA was present for sequencing. For the quantification of multiple samples, a master mix was prepared by combining 199 µL of dsDNA BR Buffer with 1 µL of dsDNA BR Reagent per isolate. The master mix was immediately pipetted into Qubit® assay tubes at a volume of 199 µL and combined with 1 µL of sample DNA. Following incubation at room temperature for 2-minutes, sample assays were read by the fluorometer, yielding a DNA concentration value in ng/µL. Before sample reading, the fluorometer was calibrated using 10 µL of Qubit® dsDNA BR Standard #1 (0 ng/µL) and Standard #2 (100 ng/µL), combined separately with 190 µL of master mix. DNA extraction was repeated for isolates with a DNA concentration of <5 ng/µL. Sample DNA was stored at -20 ºC for sequencing.

4.4.2 DNA Library Preparation

Library preparation was performed at PHAC-NML Genomics Core Facility using Nextera® XT DNA Library Preparation Kits (Illumina, San Diego, California). Before library preparation, the submitted sample DNA was normalized to a concentration of 0.2 ng/µL. The first step of DNA library preparation was tagmentation of sample DNA, where 5 µl of Tagment DNA Buffer, 3 µl of H_2O, 1 µl of Amplicon Tagment Mix and 1 µl of 0.2 ng/µl sample DNA

was added to each well of a 96-well PCR plate. The sample plate was sealed and centrifuged at 280 x g for 1 minute and then immediately incubated at 55ºC for 5 minutes. Next, 2.5 µl of Neutralize Tagment Buffer was added to each sample well to neutralize the reaction. The sample plate was then centrifuged at 280 x g for 1 minute, followed by room temperature incubation for 5 minutes. For the amplification of tagmented sample DNA, 7.5 µL of Nextera PCR Master Mix, 2.5 µL of Nextera Index Primer 1, and 2.5 µL of Nextera Index Primer 2 were added to each well and gently mixed. The sample plate was centrifuged at 280 x g for 1 minute and placed in a thermocycler. Cycling parameters used for amplification were: 72ºC for 3 minutes, 95ºC for 30 seconds, followed by 12 cycles of 95ºC for 10 seconds (denaturation), 55ºC for 30 seconds (annealing) and 72ºC for 30 seconds (extension). The sample plate was held at 10ºC, following the final extension step (72ºC for 5 minutes).

For the cleaning of sample PCR products, 17.5 µl of Agencourt® AMPure XP (Beckman Coulter) was added to each sample well and gently mixed. The sample plate was incubated at room temperature for 5 minutes and then placed on a magnetic plate stand to pull down the beads bound with sample DNA. After approximately 2 minutes, the clear supernatant was carefully pipetted off and discarded. Bead pellets were washed with 200 µL of 80% EtOH and incubated at room temperature for 30 seconds. Following removal of the supernatant, the EtOH wash step was repeated similarly for a total of two washes. The bead pellets were air-dried at room temperature for 15 minutes or until all residual EtOH evaporated. Next, the dried bead pellets were resuspended in 25 µL of Resuspension Buffer and incubated at room temperature for 2 minutes to elute the sample DNA. The beads were separated from the eluate by placing the sample plate on a magnetic plate stand and incubated at room temperature until the supernatant cleared. The supernatant was carefully pipetted off and transferred to a new 96-well plate.

The DNA present in the individual sample libraries was quantitated using Picogreen ds DNA assay on the Biomek Automated Workstation (Beckman Coulter). Pooled libraries were prepared by combining predetermined volumes of the individual sample libraries in a single microcentrifuge tube (96 libraries per pool). The volume of each sample library required for the pool was determined by an in-house pooling calculator using the sample library concentration values as input. Pooled DNA libraries were concentrated using a MicroCon® Centrifugal Filter Unit and run on the BluePippin (Sage Science) DNA size selection system. The selected DNA fragment size was 600-1000bp. The DNA in the recovered pools was quantitated using the Qubit® 2.0 Fluorometer and Qubit® dsDNA HS Assay kit (Invitrogen) as previously described (outlined in Sections 4.4.1). The DNA pools were cleaned using Agencourt® AMPure XP, as previously described, at 0.6X the volume of the recovered pool. A final Qubit quantitation step was performed to determine the final concentration of DNA in each pooled library.

Prior to sequencing, pooled libraries were diluted using an in-house dilution calculator to achieve a final library concentration of 18 pM in a 600 µL volume. The diluted pools were denatured by adding 1 µL of 0.5N NaOH and incubated at room temperature for 5 minutes. Following denaturation, 583 µL of HT1 Hybridization Buffer and 6 µL of dilute PhiX Control (20 pM) was added to each diluted pool (10 µL) and gently mixed. The final samples were held on ice until sample loading.

4.4.3 Sequencing, Assembly and Annotation

Sequencing was performed at PHAC-NML Genomics Core Facility using the Illumina MiSeq platform and 600-cycle MiSeq Reagent Kits v3 (Illumina). The quality of the raw sequence data was assessed using the FastQC v.0.11.4 tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), yielding an average of 574,644

reads per genome with an average estimated genome coverage of 84x. Prior to genome assembly, reads with <20x genome coverage were removed (n=4).

To extend the length of the raw reads, FLASH v.1.2.11.1 was run with a minimum overlap = 20 and maximum overlap = 300[60]. The extended, high-quality reads were assembled *de novo* into contiguous sequences (contigs) using SPAdes v.3.11.1[61]. The quality of the assembled genomes was assessed using multiple assembly metrics; on average, 50 contigs were generated per genome with an average contig length of 54,510 bp. The average N50 contig length was 101,119 bp. Genome annotation was performed using Prokka v.1.13[62] on the collection of assembled contigs.

4.4.4 Analysis of PBP2x, PBP2b and PBP1a DNA and Protein Sequences

4.4.4.1 Sequence Extraction and Alignment

Nucleotide and amino acid sequences of PBP2x, PBP2b and PBP1a were extracted from Prokka output files and concatenated into separate FASTA files by PBP and data type, respectively. Extracted sequences of PBP2x, PBP2b and PBP1a were aligned to the corresponding reference sequence derived from β-lactam-susceptible reference strain *S. pneumoniae* R6 (Table 4.1) using ClustalW multiple sequence alignment program[63]. Sequence alignments were assessed for quality using AliView v.1.25 alignment viewer and editor program[64]. Following examination, three isolates displayed significantly higher single nucleotide variant (SNV) densities in their *pbp2x, pbp2b* and *pbp1a* sequences compared to other sequences in the alignment. These isolates were removed from subsequent analysis due to potential errors in genome assembly. In total, seven isolates were removed following WGS assembly, annotation, and sequence extraction (n = 189).

TABLE 4.1: *S. pneumoniae* gene and protein sequences utilized as references for PBP2x, PBP2b and PBP1a sequence analysis.

Genome Name	NCBI Accession Number	Assembly Level	Gene Name	NCBI Gene ID	Gene Length (nt)	Protein Name	Protein Length (aa)
R6	NC_0.003098.1	complete	*pbpx*	934744	2,253	PBP2x	750
			pbp2b	933948	2,058	PBP2b	680
			pbpA	934791	2,160	PBP1a	719

4.4.4.2 Identification of Mutations

Pair-wise distances between aligned PBP2x, PBP2b and PBP1a nucleotide and amino acid sequences and the corresponding *S. pneumoniae* R6 reference sequences were calculated using MEGA-X software[65] by the p-distance method. The resulting output was a distance matrix for each PBP and data type containing p-distance values for each isolate; p-distance represents the proportion of sites that are different between two sequences being compared[66]. The total number of SNVs present in sample *pbp2x*, *pbp2b* and *pbp1a* sequences were determined by multiplying the nucleotide p-distance values of each isolate (compared to the reference) by the total number of nucleotide sites available for the comparison. The above calculation was performed similarly to determine the total number of substitutions in PBP2x, PBP2b and PBP1a sample sequences using amino acid p-distance values. SNV and substitution counts were performed manually on a select number of sequences to confirm the correct calculation.

4.4.4.3 Amino Acid Conservation Analysis and Mutation Mapping

Amino acid conservation analysis was performed by the ConSurf server[67] using PBP2x, PBP2b and PBP1a multiple sequence alignments as input. Evolutionary rates, or conservation scores, were calculated for individual amino acids in alignment using the Rate4Site algorithm[68] by the Jones-Taylor-Thornton (JTT) evolutionary substitution model[69]. The JTT model was chosen as the best evolutionary model for the analysis by ConSurf based on the Akaike information criterion (AIC). Amino acid conservation scores were then transformed to discrete values, grouped into nine categories and assigned colour grades (1-9) according to their level of conservation; highly variable (score = 1) residues were coloured turquoise and strictly conserved (score = 9) residues were coloured maroon. The colour grades were projected onto the 3D structures of PBP2x, PBP2b and PBP1a from *S. pneumoniae* obtained from the Protein Data

Bank (PDB) using PDB identifiers 5OAU, 2WAF and 2C6W, respectively[70]. Of note, the complete structure of PBP1a from *S. pneumoniae* was not available as a single unit chain in the PDB; 2C6W chain B is the structural representation of the PBP1a TP domain only. For all three proteins, the score of the alignment between the sequence extracted from the PDB ATOM record and the query sequence from the multiple sequence alignment (PBP2x, PBP2b and PBP1a from reference strain *S. pneumoniae* R6) was ≥ 97% identity. The complete structures of PBP2x and PBP2b and the TP domain of PBP1a coloured with ConSurf scores were visualized and edited using UCSF-Chimera v.1.13.1[71]. Positions of important previously reported mutations in PBP2x, PBP2b and PBP1a were located using the Chimera label residue command and highlighted using PowerPoint. These mutations will be described in detail, beginning in Section 4.5.

4.4.4.4 Phylogenetic Analysis and Cluster Identification

Maximum likelihood phylogenetic trees were constructed for PBP2x, PBP2b and PBP1a using MEGA-X software by the JTT evolutionary substitution model and 500 bootstrap replicates. This analysis involved 189 aligned PBP2x, PBP2b and PBP1a sample amino acid sequences, plus the corresponding reference sequence from *S. pneumoniae* R6. The JTT model was selected as the best evolutionary model for the data based on the goodness of fit test conducted by MEGA-X prior to tree construction. The goodness of fit was tested for 56 different models of evolution and was determined using Bayesian information criterion and AIC. ClusterPicker v.1.2.3[72] was used to identify clusters of sequences in the phylogenies based on node support (bootstrap) values and pairwise genetic distances. Cluster analysis was carried out at an initial threshold of 0.9, main support threshold of 0.9 and a genetic distance threshold of 4.5 for PBP2b and PBP1a. The initial threshold and main support threshold values were decreased

to 0.8 for PBP2x to promote the clustering of a larger number of sequences. The ClusterPicker modified phylogenetic trees of PBP2x, PBP2b and PBP1a were visualized using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree) and edited using InkScape v.0.92.4.

4.4.4.5 Sequence Comparison

Structure-based sequence alignments were generated for PBP2x, PBP2b and PBP1a using the ESPript 3.0 server[73]. Comparative analysis was carried out on representative sequences from the most common PBP2x (n = 6 sequences), PBP2b (n = 7 sequences) and PBP1a (n = 7 sequences) clusters identified by cluster analysis (Section 4.4.4.4). Individual sequences were randomly selected for each PBP and cluster type, respectively, and aligned using ClustalW. Sequence similarity was determined by ESPript using the % equivalent method for calculation of similarity score[74]. In brief, for each column in the alignment, a percentage of equivalent residues was calculated based on amino acid physicochemical properties and then compared to the global score threshold of 0.7. Strictly conserved residues (score = 1) were displayed as white characters on a red background, whereas highly similar residues (score = $\geq$0.7, <1) were displayed as red characters on a white background. Residues in the column with a similarity score <0.7 were considered weakly similar; these were shown as black characters on a white background. Protein secondary structure elements derived from the 3D structure of PBP2x, PBP2b and PBP1a were retrieved from the PDB by ESPript using PDB identifiers 5OAU, 2WAF and 2C6W, respectively, and displayed above the corresponding alignments. The alignments were edited for presentation using PowerPoint.

4.5 Identification and Classification of β-lactam-Resistance Conferring Mutations in *S. pneumoniae* PBP2x, PBP2b and PBP1a sequences

Alterations within or adjacent to the three active site motifs of PBP2x, PBP2b and PBP1a were detected as previously described[75]: STMK, AHSSNV, and LKSGT of PBP2x; SVVK, SSNT, and KTGTA of PBP2b; and STMK, SRNVP and KTG of PBP1a. In addition, an extensive literature review was performed to identify other mutations outside of the well-described motif alterations in PBP2x, PBP2b and PBP1a with experimentally determined roles in β-lactam resistance in *S. pnuemoniae.* Following review, a total of eleven β-lactam-resistance conferring amino acid substitutions were identified for PBP2x, nine were identified for PBP2b and eleven were identified for PBP1a. The selected substitutions are presented in Table 4.2 and are described in detail in Tables 10.1, 10.2 and 10.3 of Appendix A. Each substitution was assigned a categorical ranking, ranging from low to very high, according to the level of experimental evidence available to support its role in β-lactam resistance in *S. pneumoniae*. The four levels of evidence are described as follows: low = structure-based evidence only; medium = motif alteration with or without structural evidence; high = biochemical evidence by way of site-directed mutagenesis studies, with or without structural evidence; very high = motif alteration with biochemical and structural evidence. Substitutions of interest were detected in aligned sample sequences using AliView v.1.25 and recorded within a database as present (+) or absent (-).

.

TABLE 4.2: Previously reported amino acid substitutions in PBP2x, PBP2b and PBP1a conferring reduced β-lactam susceptibility in *S. pneumoniae*.

Protein	Amino Acid Substitution			Level of Evidence[a]	Reference(s)
	Allele 1	Position	Allele 2		
PBP2x	T	338	A	very high	75,76,77,78,79,80,81,82,83
	L	364	F	high	84
	A	369	V	high	84
	I	371	T	high	81,84
	R	384	G	high	81,84
	S	389	L	low	82
	H	394	L	medium	75,76,79
	N	514	H	low	82
	L	546	V	medium	75,76,78,79,80
	Q	552	E	very high	85,86,87,88
	N	605	T	high	81
PBP2b	S	412	P	low	89,90
	N	422	Y	low	89,90
	T	426	K	low	89,90
	Q	427	L	low	89,90
	T	446	A	very high	75,78,79,80,90,91
	A	619	G	medium	75,78,80,89
	D	625	G	low	89, 90

	Q	628	E	low	89,90
	T	630	N	low	89,90
PBP1a	T	371	A/S	very high	75,77,78,79,80,92, 93,94
	P	432	T	medium	75,77,78,80,94
	N	546	G	low	95
	TSQF	574-577	NTGY	high	93,94,96
	L	606	I	low	95
	N	609	D	low	95
	L	611	F	low	95
	T	612	L	low	95

[a], low = structure-based evidence only; medium = motif alteration with or without structural evidence; high = biochemical evidence with or without structural evidence; very high = motif alteration with biochemical and structural evidence.

4.6 Statistical Analysis

An additional cohort of 63 IPD isolates with penicillin MICs ≥ 0.12 µg/mL were included in the statistical portion of this study to increase the predictive power of the three multiple linear regression models predicting *S. pneumoniae* β-lactam MICs. The genomes of these supplementary isolates were analyzed previously[97] and were submitted as part of NCBI BioProject PRJNA510562. AST was repeated in duplicate as previously described (Section 4.1.2) to confirm penicillin, ceftriaxone and meropenem MIC values for these isolates. Isolate PBP2x, PBP2b and PBP1a amino acid sequences were analyzed as previously described (Sections 4.4.4.2 and 4.5). Of the 63 supplementary isolates that were included in the penicillin MIC model (n = 252), 49 (ceftriaxone MICs >0.12 µg/mL) and 44 (meropenem MICs >0.06 µg/mL) isolates were included in the models for ceftriaxone (n = 238) and meropenem (n = 233), respectively.

4.6.1 Principal Components Analysis (PCA)

PCA was performed using SPSS Statistics v.26 (IBM Corp., Armonk, NY, USA) software to assess the relationships amongst the selected PBP2x-, PBP2b- and PBP1a-associated variables and to derive a smaller number of variables that capture approximately the same information as the larger set of variables. The analysis was conducted separately for each PBP type by the principal components method of extraction using the eleven, nine and eight previously reported substitutions in PBP2x, PBP2b and PBP1a, respectively, as input variables. Component extraction was based on Eigenvalues greater than 1 criterion and a maximum iteration for convergence of 25. A rotated component matrix was produced using the varimax method of rotation, displaying the component loading values for each variable, ranging from -1.0 to 1.0; variables with high loading values (positive or negative) in the same component were

considered highly related. Highly related variables were aggregated into composite variables if they fit conceptually for each PBP type. The Kaiser-Meyer-Olkin (KMO) and Bartlett's tests were performed to check the assumptions of normality, linearity and correlations amongst the variables required for PCA[98].

4.6.2 Multiple Linear Regression

Multiple linear regression models were generated by SPSS Statistics v.26 software to predict *S. pneumoniae* β-lactam MIC values from a combination of several previously described β-lactam resistance-conferring determinants used as independent (predictor) variables. Prior to regression analysis, measured MIC values for penicillin, ceftriaxone and meropenem were transformed to the $\log_2$ scale. The $\log_2$ (MIC) values for each β-lactam were used as dependent variables in separate regression models. A total of twelve independent variables were used as input for each model, including eight composite variables derived from PCA for PBP2x, PBP2b and PBP1a, as well as amino acid p-distance values for the three PBP types. The eight PCA-derived composite variables were included as dichotomous variables, whereas p-distance values were included as scale variables, rounded to two decimal places.

Multiple linear regression models for the three β-lactams were produced using the forward selection approach to ensure that all independent variables were contributing significantly to the final model[98]. Forward multiple regression involved the addition of the selected independent variables, one by one in successive rounds of regression, continuing until all variables with *P*-values ≤ 0.05 were added[98]. Of the multiple models generated using this approach, the best model for each β-lactam was determined as the model that demonstrated the highest adjusted (Δ) R^2 value. Also taken into account in the determination of the best model was the condition of multicollinearity; tests for multicollinearity were conducted using the SPSS

collinearity diagnostics command. Multicollinearity was considered a problem if tolerance values were < 1 - R^2 for the variables included in the final model, indicating high intercorrelations amongst the independent variables[98]. Following best-model determination, the unstandardized beta coefficients plus the model constant terms were used to generate linear regression equations for each of the final models using the formula $Y = \beta_o + B_i X_i + e$, where Y [$\log_2$ (β-lactam MIC)] is dependent on '*i'* number of independent variables.

4.6.3 β-lactam MIC Model Validation

The leave-one-out cross-validation method was performed using SPSS Statistics v.26 software to evaluate overall model performance and to determine the error rates for the three generated linear regression models predicting *S. pneumoniae* β-lactam MICs. This method was selected for model validation due to a relatively small sample size that prohibited the designation of separate datasets of equal size for model training and validation purposes. Briefly, leave-one-out cross-validation involved the splitting of the original dataset into two subsets for each model; subset 1 contained a single excluded sample (n) for model validation, while subset 2 contained the remaining n-1 samples for model fitting[99]. The process was repeated for all samples in the dataset, generating *DfFit* values that represent the change in the predicted MIC value resulting from the exclusion of that particular sample[100]. Predicted $\log_2$ (β-lactam MIC) values were calculated by subtracting the *DfFit* value from the unstandardized predicted value for each sample and were subsequently rounded to the nearest doubling dilution and converted back to the original measurement scale (µg/mL).

Model performance was evaluated by calculating the essential agreement (EA) and categorical agreement (CA) between the measured β-lactam MIC values by broth microdilution (reference method) and WGS-based predicted β-lactam MICs. The rates for CA and EA should

be ≥ 90% according to CLSI performance specifications[45]. Categorical discrepancies were further classified as minor errors (mE), major errors (ME) or very major errors (VME) using CLSI breakpoint criteria for penicillin, ceftriaxone and meropenem. The three error types were defined as follows: mE = isolate intermediate by the reference method but predicted susceptible or resistant by WGS, or vice versa; ME = isolate susceptible by the reference method but predicted resistant by WGS; VME = isolate resistant by the reference method but predicted susceptible by WGS. Acceptable error rates for this study were ≤10% for mE and <3% for both ME and VME[45].

5. RESULTS

Part A: Molecular Characterization of *S. pneumoniae* PBP2x, PBP2b, and PBP1a Amino Acid Sequences by WGS

5.1 Characterization of PBP2x

5.1.1 Conservation Analysis and Mapping of PBP2x Mutations

Evolutionary conservation analysis was performed using the ConSurf server to assess the positions of amino acid conservation or variation in the 3D structure of PBP2x from *S. pneumoniae.* The analysis was carried out on a multiple sequence alignment of 189 sample PBP2x sequences plus PBP2x from reference strain *S. pneumoniae* R6. Slowly evolving (conserved, maroon) and rapidly evolving (variable, turquoise) sites were mapped onto the 3D structure of PBP2x from *S. pneumoniae* (PDB ID: 5OAU[86]) and presented in Figures 5.1A&B. Positions of the eleven previously reported mutations in PBP2x associated with reduced β-lactam susceptibility in *S. pneumoniae* (Section 4.5) were labelled in the enlarged image of the PBP2x active site region (black box, Figure 5.1B).

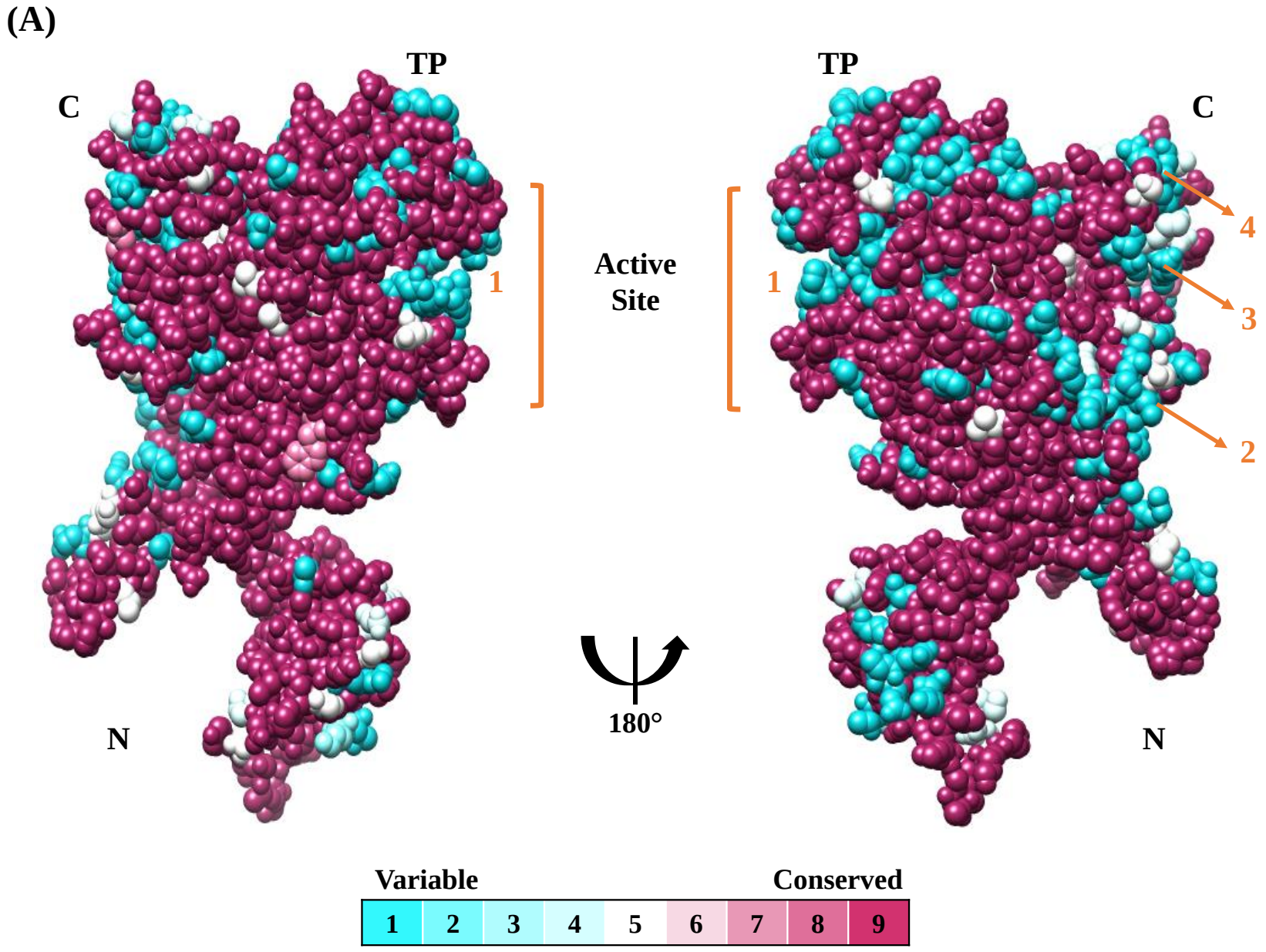
Overall, ConSurf analysis revealed that the structural core of PBP2x, comprised of N-terminal, TP and C-terminal domains, was mostly conserved. Despite this, clusters of highly variable residues forming distinct regions of variability were shown in both the space-filling (Figure 5.1A) and ribbon model (Figure 5.1B) representations of the protein. As expected, many highly variable sites were concentrated within the TP domain of PBP2x; however, the C-terminal domain [comprised of two PASTA domains (P1 and P2)] was also shown to be highly variable. Noted regions of variability on the surface of the TP and C-terminal domains of PBP2x were labelled from 1-4 in Figure 5.1A. Region 1 was recognized as a cluster of highly variable

residues within the TP domain surrounding the entry of the active site cleft, indicating that this region of PBP2x is rapidly evolving in *S. pneumoniae*.

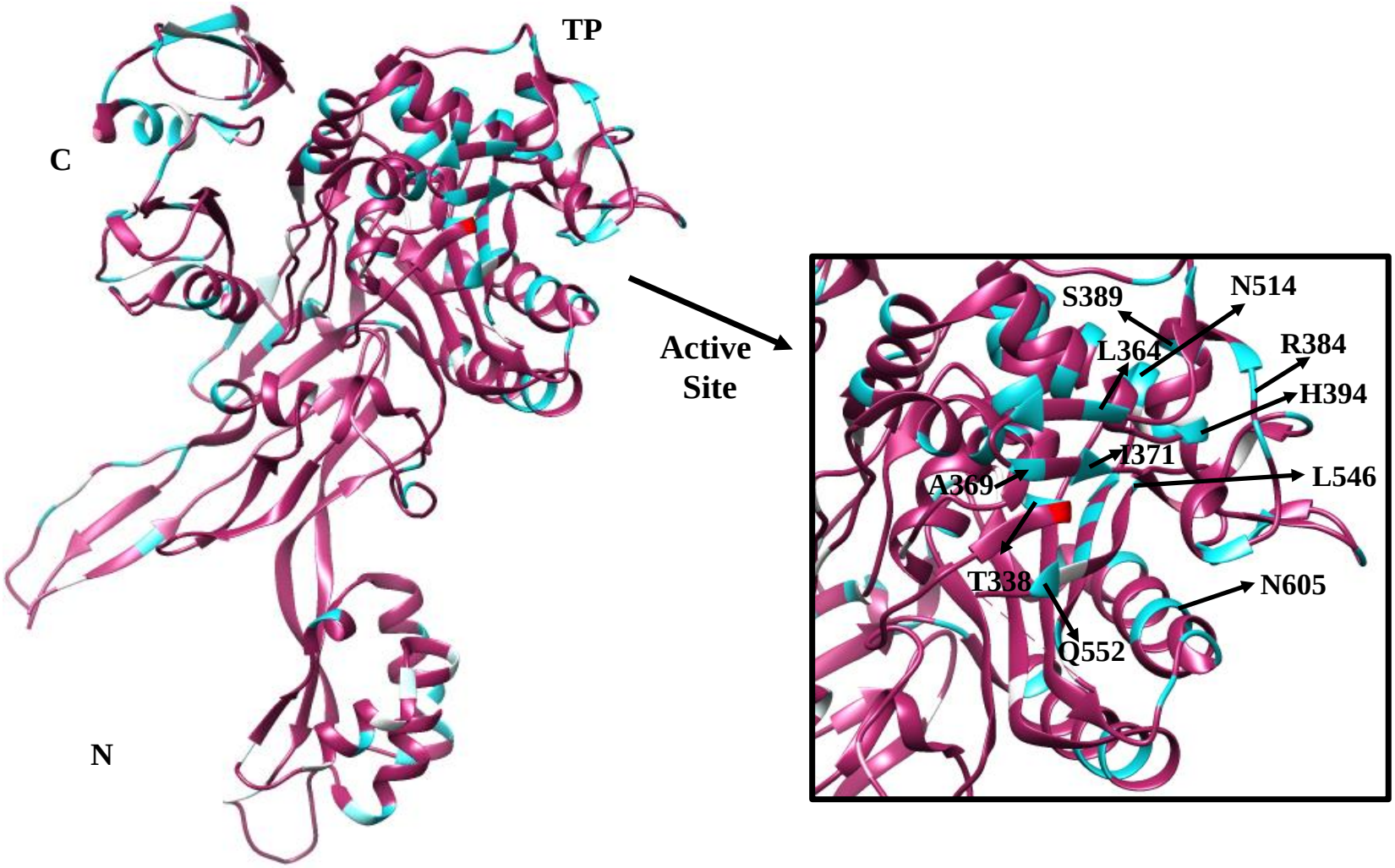
Region 2 was located opposite to the active site at the interface of the N-terminal, TP and C-terminal domains. This region was comprised of a small number of highly variable residues derived from the PBP2x linker region (residues 617-634), described further in Section 5.1.3. Highly variable regions 3 and 4 were both positioned in the C-terminal domain of PBP2x; region 3 was comprised of highly variable residues derived from the P1 domain (residues 635-691), whereas region 4 was comprised of highly variables residues from the P2 domain (residues 692-750). The C-terminal domain of PBP2x was analyzed further in Section 5.1.3. Overall, ConSurf analysis showed that additional sites outside of the PBP2x TP domain active site region are also rapidly evolving in *S. pneumoniae*. In general, the N-terminal domain was shown to be the most conserved of the three structural domains.

Shown in the ribbon model representation of PBP2x (Figure 5.1B) was an accumulation of highly variable residues deep within the TP domain active site cleft, many of which were located near the catalytic serine residue (S337, red) of the STMK active site motif. Importantly, positions of all eleven previously reported mutations in PBP2x were mapped within this region. In the enlarged image of the PBP2x TP domain active site, residues T338, L364, A369, I371, H394, L546 and Q552 were shown to be buried deep within the active site cleft. Conversely, residues R384, S389, N514 and N605 were positioned at various locations around the entry of the active site. Substitutions at these eleven positions were analyzed further, beginning in Section 5.1.4.

FIGURE 5.1: Amino acid conservation analysis of PBP2x from *S. pneumoniae*. A multiple sequence alignment of 189 sample PBP2x sequences plus PBP2x from *S. pneumoniae* R6, generated by ClustalW, was used as input for this analysis using the ConSurf server. Conservation scores were calculated for individual amino acids in the alignment by the JTT evolutionary substitution model. Residues were assigned colour grades based on their level of conservation (scale from 1-9): highly variable residues (score = 1) were coloured turquoise and strictly conserved residues (score = 9) were coloured maroon. Scores were mapped onto the 3D structure of PBP2x from *S. pneumoniae* (PDB ID: 5OIU). Figures were generated using UCSF Chimera v.1.13.1. The N-terminal, transpeptidase and C-terminal domains of PBP2x were labelled N, TP and C, respectively. (**A**) Space-filling representation of PBP2x highlighting sequence conservation on the surface of the protein. Highly variable regions were labelled from 1-4. (**B**) Ribbon model representation of PBP2x highlighting sequence conservation on the interior of the protein. The ribbon position of the catalytic serine residue (S337) of the STMK active site motif was coloured red. Positions of the eleven previously reported substitutions in PBP2x associated with reduced β-lactam susceptibility in *S. pneumoniae* were labelled within the enlarged image of the TP domain active site (black box).


(A)
TP
C
1
Active Site
180°
N
TP
C
1
4
3
2
N
Variable
Conserved
1
2
3
4
5
6
7
8
9

(B)


TP
C
N
Active Site
S389
N514
L364
R384
H394
I371
A369
L546
T338
N605
Q552
Variable
Conserved
1
2
3
4
5
6
7
8
9

5.1.2 PBP2x Phylogenetic Analysis and Cluster Identification

To assess the genetic diversity of PBP2x in *S. pneumoniae,* a maximum likelihood phylogenetic tree of 189 sample PBP2x sequences plus PBP2x from reference strain *S. pneumoniae* R6 was produced and is presented in Figure 5.2. In total, ClusterPicker identified 15 clusters of PBP2x sequences, consisting of as few as two sequences to as many as 76. Detailed information on the 15 clusters of PBP2x sequences is provided in Table 5.1. The six most common PBP2x clusters containing >5 sequences were selected to be analyzed in detail, beginning in this section.

Cluster 1 was the largest cluster identified by cluster analysis, containing 76 sequences from 22 different serotypes, including PBP2x from *S. pneumoniae* R6. Sequences within cluster 1 were largely unaltered (wild-type), demonstrating a ≤0.5% amino acid difference compared to PBP2x of *S. pneumoniae* R6. On average, one amino acid substitution was identified per sequence. All cluster 1 sequences were derived from penicillin-susceptible isolates with penicillin MICs ≤0.015 µg/mL.

Overall, non-wild-type PBP2x sequences clustered mainly by serotype in isolates with penicillin MICs ≥ 0.06 µg/mL. Cluster 2 sequences demonstrated a moderately increased number of mutations compared to PBP2x of *S. pneumoniae* R6, averaging 19 substitutions per sequence with a percent amino acid difference of 2.5%. Sequences within cluster 2 (83.9%, 26/31) were predominantly from serotype 19A isolates with penicillin MICs of 0.06 µg/mL. In contrast, cluster 4 sequences were exclusively from serotype 15A isolates; these isolates were commonly penicillin-intermediate with penicillin MICs of 0.12 µg/mL (83.3%, 10/12). Sequences within cluster 4 differed from PBP2x of *S. pneumoniae* R6 by 3.6%, averaging 27

substitutions per sequence. Cluster 9 was comprised of sequences from serotypes 23A and 23B, representing 57.1% (4/7) and 42.9% (3/7) of sequences within the cluster, respectively. Similarly to cluster 4, cluster 9 sequences were exclusively from penicillin-intermediate isolates with penicillin MICs of 0.12 µg/mL. However, cluster 9 sequences demonstrated an increased number of mutations compared to cluster 4 sequences, averaging 54 substitutions with a percent amino acid difference of 7.2% (compared to PBP2x of *S. pneumoniae* R6).

A large proportion of PBP2x sequences from penicillin-resistant isolates were contained within clusters 10 and 12. Cluster 10 was comprised entirely of sequences from penicillin-resistant serotype 19A isolates with MICs of 4 µg/mL. Sequences within cluster 10 differed from PBP2x of *S. pneumoniae* R6 by 9.2%, averaging 69 substitutions per sequence. Of the six most common PBP2x sequence clusters, cluster 12 sequences were the most distantly related to PBP2x of *S. pneumoniae* R6; these sequences demonstrated the highest percent amino acid difference at 9.9%, averaging 74 substitutions per sequence. The majority of sequences within cluster 12 were from serotype 35B (71.4%, 5/7) isolates with penicillin MICs ranging from 1-2 µg/mL.

FIGURE 5.2: Maximum likelihood phylogenetic tree of 190 *S. pneumoniae* PBP2x amino acid sequences. Unrooted tree was generated by MEGA-X using the JTT substitution model and 500 bootstrap replicates. Clusters were identified using ClusterPicker v.1.2.3, highlighted using FigTree v.1.4.4 and labelled from 1-15. Clusters were labelled from lowest to highest percent amino acid difference compared to PBP2x of *S. pneumoniae* R6.

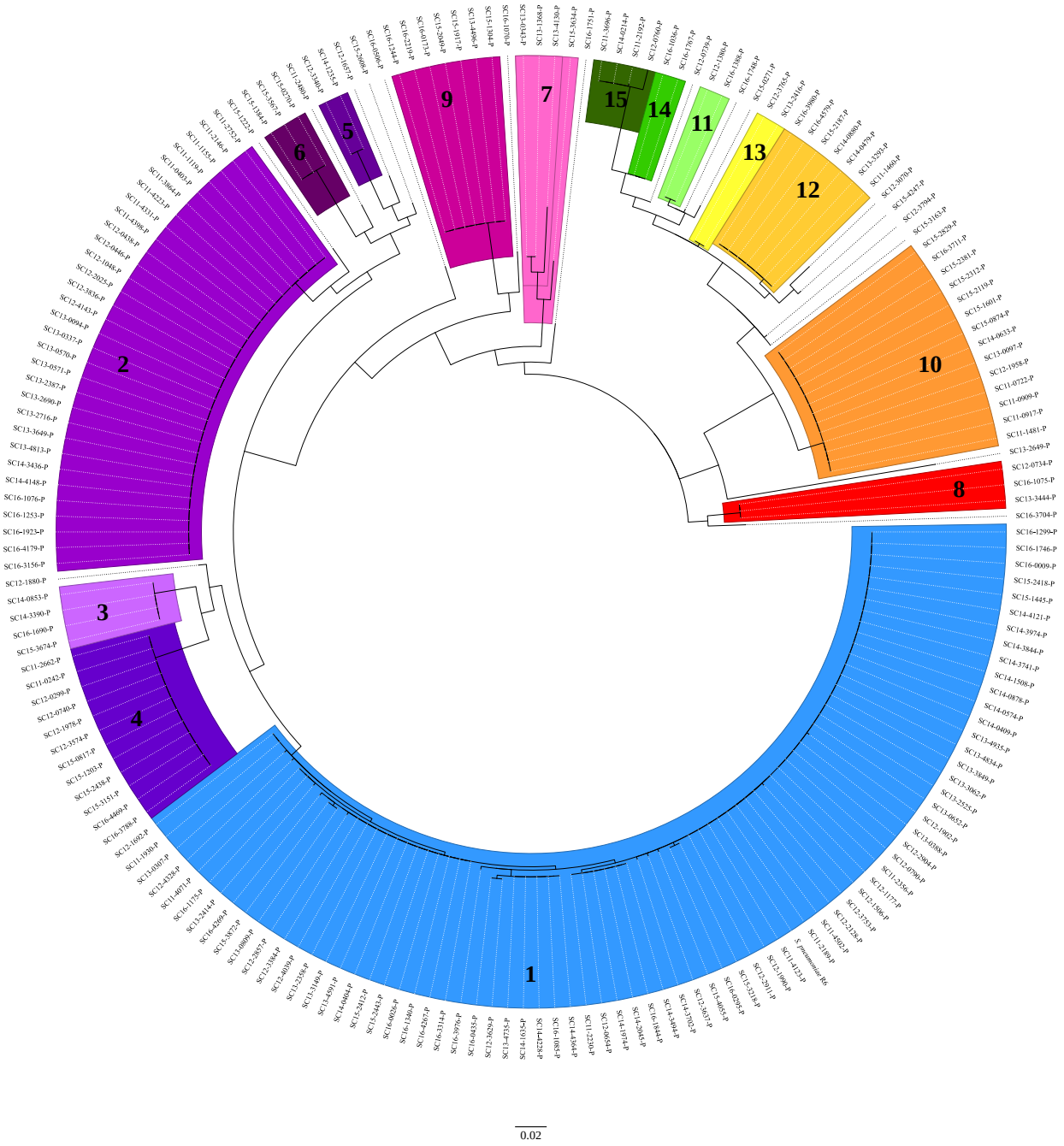


TABLE 5.1: Characterization of the fifteen *S. pneumoniae* PBP2x sequence clusters highlighted in Figure 5.2.

Cluster No.	Serotypes (n)	Amino acid p-distance[a] (%)	Average No. of Substitutions[b]	Penicillin MIC/range (µg/mL)	Ceftriaxone MIC/range (µg/mL)	Meropenem MIC/range (µg/mL)
1	22F (12), 3 (8), 12F (8), 8 (7), 7F (5), 11A (5), 16F (4), 35F (4), 6C (3), 9N (3), 34 (3), 15B (2), 20 (2) single STs: 7C, 10A, 19A, 23A, 25A, 33F, 35B, 37, 38	0-0.5	1	0.001-0.015	0.001-0.03	0.002-0.015
2	19A (29), 3 (1), 6A (1)	2.5	19	0.06-0.12	0.03-0.06	0.008-0.03
3	24F (4)	3.3	25	0.5-1	0.12-0.25	0.06-0.12
4	15A (12)	3.6	27	0.06-0.25	0.015-0.12	0.008-0.03
5	6C (2)	4.6	34	0.12-0.25	0.25	0.25
6	23A (3)	5.6	42	0.25	0.12-0.25	0.015-0.03
7	6A (2), 16F (1), 22F (1)	6.2	46	0.06	0.015-0.06	0.008-0.015

8	6C (2), 6A (1)	6.6	49	0.12-0.25	0.25	0.03-0.06
9	23A (4), 23B (3)	7.2	54	0.12	0.06	0.03
10	19A (14)	9.2	69	4	1-2	1-2
11	15A (2)	9.5	71	1-2	1	0.5
12	35B (5), 29 (2)	9.9	74	1-2	0.5-1	0.5-1
13	9V (1), 19A (1)	10.0	75	1	1	0.5
14	6C (2)	10.3	77	0.5-2	0.5-1	0.12-0.5
15	6C (1), 9V (1), 15A (1), 19A (1)	13.8	104	0.12-2	0.12-1	0.03-0.12
N/A	unclustered isolates[c]: 19A (4), 19F (3), 6C (2), 16F (2), 1 (1), 8 (1), 15B (1), 18 (1), 23F (1), 24F (1)	variable	variable	variable	variable	variable

[a], Percent amino acid difference compared to *S. pneumoniae* R6 PBP2x reference sequence. [b], Averaged from isolate total per cluster. [c], These isolates did not cluster using ClusterPicker at an initial threshold of 0.8, main support threshold of 0.8 and a genetic distance threshold of 4.5.

5.1.3 PBP2x Sequence Comparison

A structure-based multiple sequence alignment containing representative sequences from the six most common PBP2x sequences clusters identified in Section 5.1.2 was produced and is presented in Figure 5.3. The purpose of the alignment was to compare, by cluster, sequence variability within the three structural domains of PBP2x. Protein secondary structure elements derived from the crystal structure of PBP2x from *S. pneumoniae* (PDB ID: 5OAU[86]) were displayed above the alignment using the ESPript server. Coordinates of the three structural domains of PBP2x were labelled in Figure 5.3 as follows[82,85,88]: N-terminal domain (residues 49-265), TP domain (residues 266-616) and C-terminal domain (residues 635-750).

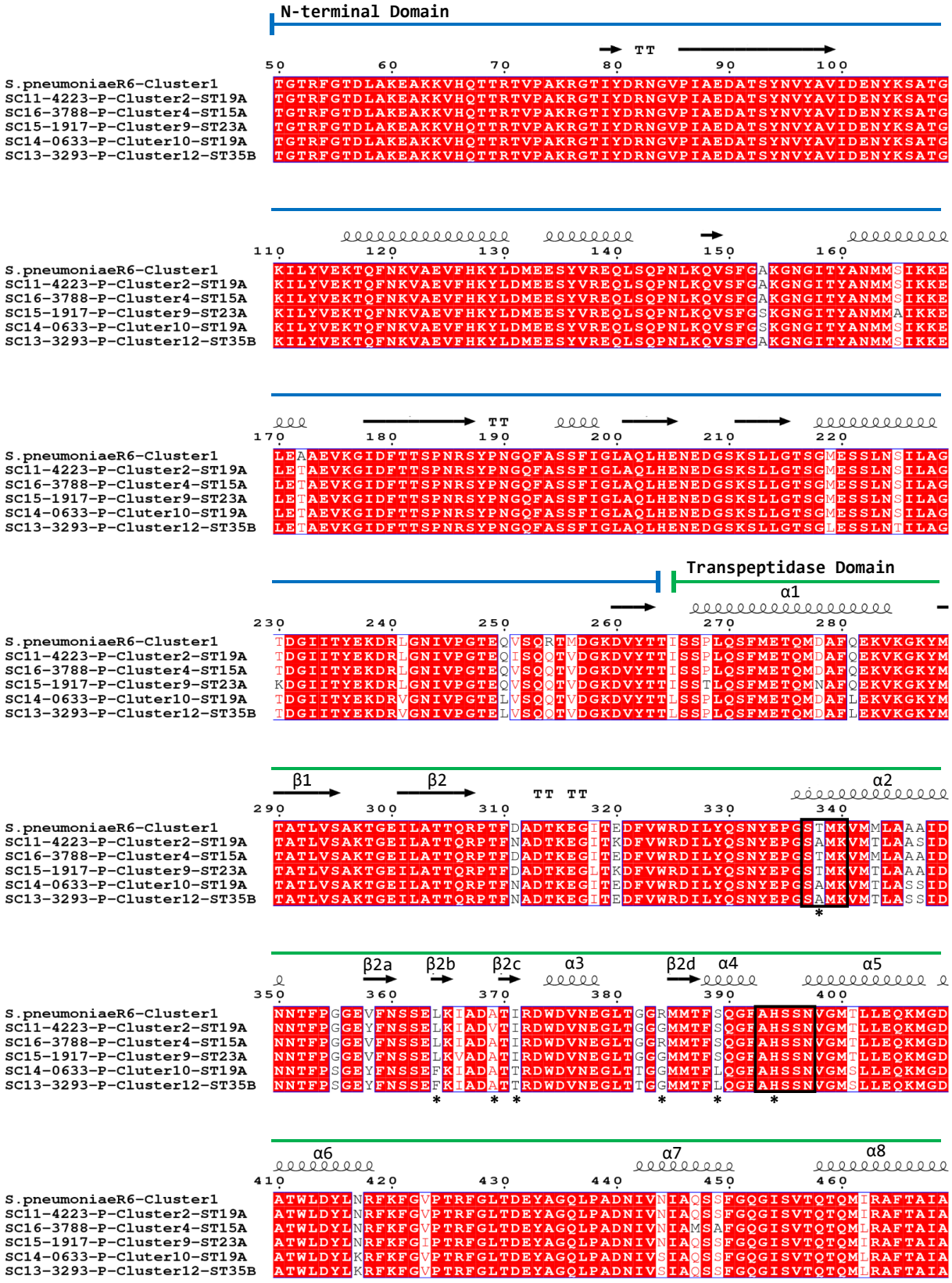
Comparison of the six representative sequences revealed that, overall, the N-terminal domain of PBP2x is largely unaltered; 94.5% (205/217) of residues within this region were strictly conserved. Conversely, the transpeptidase domain was shown to be increasingly variable in each of the five non-wild-type PBP2x sequences (cluster 2, 4, 9, 10, and 12 sequences). This is in agreement with ConSurf analysis of PBP2x (Section 5.1.1), where multiple distinct regions of variability were observed within the TP domain. Of the three structural domains of PBP2x analyzed, the majority of substitutions identified were concentrated within the TP domain in the five non-wild-type sequences; substitutions within the TP domain represented 47.4% (9/19), 81.5% (22/27), 48.2% (26/54), 52.2% (36/69), and 51.4% (38/74) of the total number of substitutions in cluster 2, 4, 9, 10 and 12 sequences, respectively. In comparison to corresponding region in PBP2x in *S. pneumoniae* R6 (cluster 1), the TP domain sequences of clusters 2, 4, 9, 10, and 12 differed from wild-type by 2.6% (9/351), 6.3% (22/351), 7.4% (26/351), 10.3% (36/351) and 10.8% (38/351), respectively. Interestingly, greater than half of the mutations within this region in the selected cluster 4 (54.5%), 9 (61.5%), 10 (66.7%) and 12

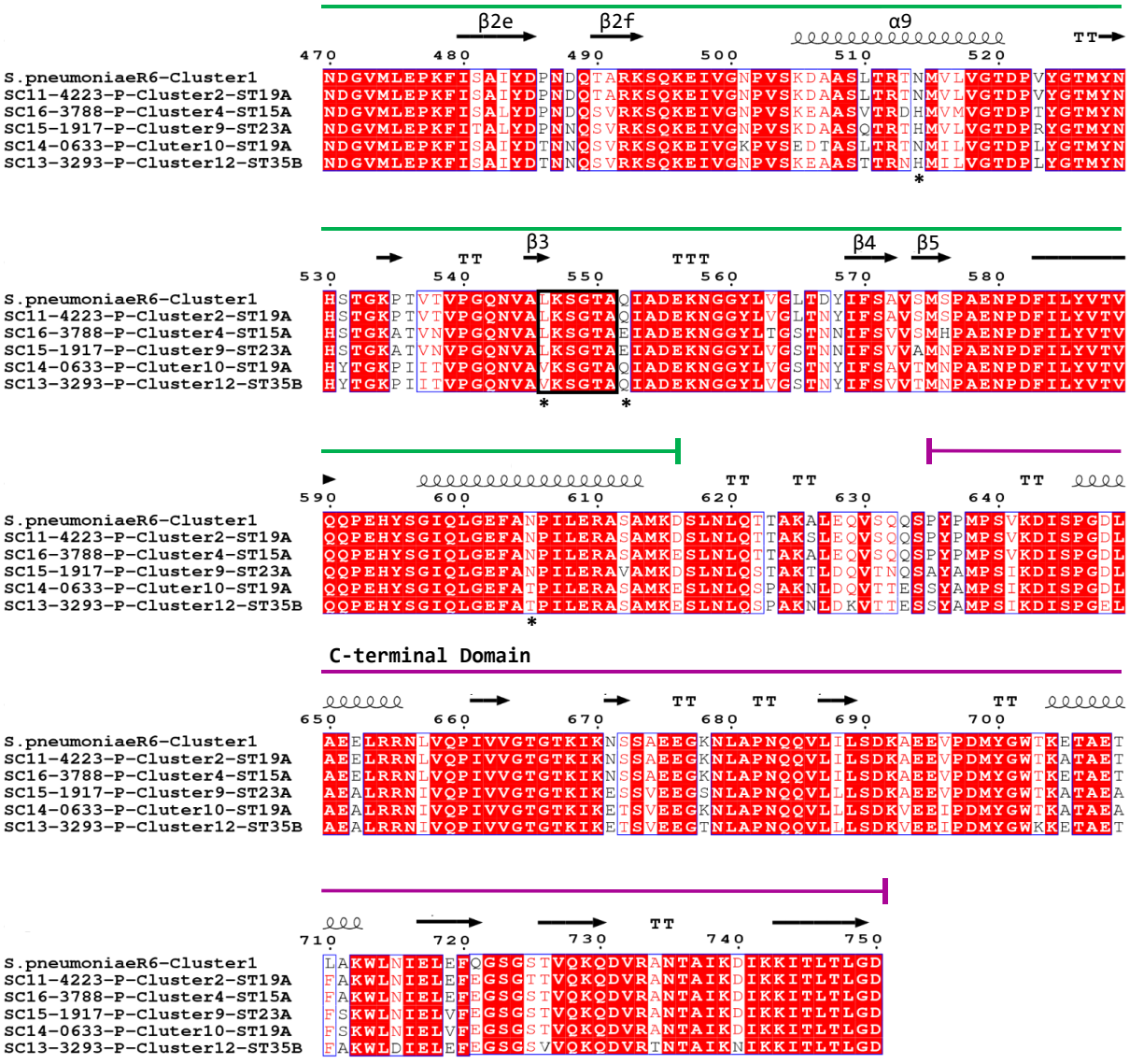
(60.5%) sequences were classified as non-conservative substitutions (shown as black characters on a white background) according to the ESPript % equivalent method for calculation of amino acid similarity type.

Similarly to the TP domain, the C-terminal domain of PBP2x was also shown to be highly variable by both ConSurf and ESPript analysis. Substitutions within this region represented 21.1% (4/19), 7.4% (2/27), 27.8% (15/54), 24.6% (17/69) and 27% (20/74) of the total number of substitutions in cluster 2, 4, 9, 10 and 12 sequences, respectively. Notably, 4 out of the 5 non-wild-type PBP2x sequences demonstrated a higher degree of variation in their C-terminal domain than their TP domain when compared to the corresponding regions of *S. pneumoniae* R6 PBP2b; the C-terminal domain sequences of clusters 2, 4, 9, 10 and 12 differed from wild-type by 3.5% (4/116), 1.7% (2/116), 12.9% (15/116), 14.7% (17/116) and 17.2% (20/116), respectively.

Noted in the alignment was a small, highly variable region between the TP and C-terminal domains of PBP2x, spanning residues 617-634. This group of residues related to a distinct hotspot of variability labelled region 2 in the structural model of PBP2x (Figure 5.1). Substitutions within this region of PBP2x represented 5.3% (1/19), 0% (0/27), 11.1% (6/54), 10.2% (7/69) and 10.8% (8/74) of the total number of substitutions in cluster 2, 4, 9, 10 and 12 sequences, respectively. Interestingly, the highest degree of sequence variation of all regions analyzed was demonstrated in this region in 4 out of the 5 non-wild-type sequences, differing from PBP2b of *S. pneumoniae* R6 by 5.5% (1/18), 0% (0/18), 33.3% (6/18), 38.9% (7/18) and 44.4% (8/18) in cluster 2, 4, 9, 10, and 12 sequences, respectively.

FIGURE 5.3: Structure-based multiple sequence alignment of select *S. pneumoniae* PBP2x amino acid sequences. Representative sequences from the six most common PBP2x sequence clusters (n >5 isolates) identified in Section 5.1.2 were aligned using ClustalW and displayed using ESPript 3.0 server. Protein secondary structure elements derived from the crystal structure of PBP2x (PDB ID: 5OAU[86]) were presented above the alignment: α-helices and β-strands were shown as medium squiggles and bold arrows, respectively. The coordinates of the three structural domains of PBP2x were highlighted with coloured bars above secondary structure elements. Secondary structure elements within the TP domain were labelled according to Dessen *et al*[82]. Individual amino acid residues were coloured according to the ESPript % equivalent colouring scheme: strictly conserved residues were shown as white characters on a red background; residues that displayed within-group similarity were shown as red characters on a white background; weakly similar residues were shown as black characters on a white background. The three conserved active site motifs of PBP2x were highlighted with black boxes. Positions of the eleven previously reported mutations in PBP2x associated with reduced β-lactam susceptibility in *S. pneumoniae* were indicated with asterisks.


N-terminal Domain
TT
50 60 70 80 90 100
S.pneumoniaeR6-Cluster1 TGTRFGTDLAKEAKKVHQTTRTVPAKRGTIYDRNGVPIAEDATSYNVYAVIDENYKSATG
SC11-4223-P-Cluster2-ST19A TGTRFGTDLAKEAKKVHQTTRTVPAKRGTIYDRNGVPIAEDATSYNVYAVIDENYKSATG
SC16-3788-P-Cluster4-ST15A TGTRFGTDLAKEAKKVHQTTRTVPAKRGTIYDRNGVPIAEDATSYNVYAVIDENYKSATG
SC15-1917-P-Cluster9-ST23A TGTRFGTDLAKEAKKVHQTTRTVPAKRGTIYDRNGVPIAEDATSYNVYAVIDENYKSATG
SC14-0633-P-Cluter10-ST19A TGTRFGTDLAKEAKKVHQTTRTVPAKRGTIYDRNGVPIAEDATSYNVYAVIDENYKSATG
SC13-3293-P-Cluster12-ST35B TGTRFGTDLAKEAKKVHQTTRTVPAKRGTIYDRNGVPIAEDATSYNVYAVIDENYKSATG
110 120 130 140 150 160
S.pneumoniaeR6-Cluster1 KILYVEKTQFNKVAEVFHKYLDMEESYVREQLSQPNLKQVSFGAKGNGITYANMMSIKKE
SC11-4223-P-Cluster2-ST19A KILYVEKTQFNKVAEVFHKYLDMEESYVREQLSQPNLKQVSFGAKGNGITYANMMSIKKE
SC16-3788-P-Cluster4-ST15A KILYVEKTQFNKVAEVFHKYLDMEESYVREQLSQPNLKQVSFGAKGNGITYANMMSIKKE
SC15-1917-P-Cluster9-ST23A KILYVEKTQFNKVAEVFHKYLDMEESYVREQLSQPNLKQVSFGSKGNGITYANMMAIKKE
SC14-0633-P-Cluter10-ST19A KILYVEKTQFNKVAEVFHKYLDMEESYVREQLSQPNLKQVSFGSKGNGITYANMMSIKKE
SC13-3293-P-Cluster12-ST35B KILYVEKTQFNKVAEVFHKYLDMEESYVREQLSQPNLKQVSFGAKGNGITYANMMSIKKE
TT
170 180 190 200 210 220
S.pneumoniaeR6-Cluster1 LEAAEVKGIDFTTSPNRSYPNGQFASSFIGLAQLHENEDGSKSLLGTSGMESSLNSILAG
SC11-4223-P-Cluster2-ST19A LETAEVKGIDFTTSPNRSYPNGQFASSFIGLAQLHENEDGSKSLLGTSGMESSLNSILAG
SC16-3788-P-Cluster4-ST15A LETAEVKGIDFTTSPNRSYPNGQFASSFIGLAQLHENEDGSKSLLGTSGMESSLNSILAG
SC15-1917-P-Cluster9-ST23A LETAEVKGIDFTTSPNRSYPNGQFASSFIGLAQLHENEDGSKSLLGTSGMESSLNSILAG
SC14-0633-P-Cluter10-ST19A LETAEVKGIDFTTSPNRSYPNGQFASSFIGLAQLHENEDGSKSLLGTSGMESSLNSILAG
SC13-3293-P-Cluster12-ST35B LETAEVKGIDFTTSPNRSYPNGQFASSFIGLAQLHENEDGSKSLLGTSGLESSLNTILAG
Transpeptidase Domain
α1
230 240 250 260 270 280
S.pneumoniaeR6-Cluster1 TDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGKYM
SC11-4223-P-Cluster2-ST19A TDGIITYEKDRLGNIVPGTEQISQQTVDGKDVYTTISSPLQSFMETQMDAFQEKVKGKYM
SC16-3788-P-Cluster4-ST15A TDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSPLQSFMETQMDAFQEKVKGKYM
SC15-1917-P-Cluster9-ST23A KDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSTLQSFMETQMNAFQEKVKGKYM
SC14-0633-P-Cluter10-ST19A TDGIITYEKDRVGNIVPGTELVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGKYM
SC13-3293-P-Cluster12-ST35B TDGIITYEKDRVGNIVPGTELVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGKYM
β1 β2 α2
TT TT
290 300 310 320 330 340
S.pneumoniaeR6-Cluster1 TATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLAAAID
SC11-4223-P-Cluster2-ST19A TATLVSAKTGEILATTQRPTFNADTKEGITKDFVWRDILYQSNYEPGSAMKVMTLAASID
SC16-3788-P-Cluster4-ST15A TATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLAAAID
SC15-1917-P-Cluster9-ST23A TATLVSAKTGEILATTQRPTFDADTKEGITKDFVWRDILYQSNYEPGSTMKVMTLAAAID
SC14-0633-P-Cluter10-ST19A TATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVMTLASSID
SC13-3293-P-Cluster12-ST35B TATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVMTLASSID
*
β2a β2b β2c α3 β2d α4 α5
350 360 370 380 390 400
S.pneumoniaeR6-Cluster1 NNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLEQKMGD
SC11-4223-P-Cluster2-ST19A NNTFPGGEYFNSSELKIADVTIRDWDVNEGLTGGMMTFSQGFAHSSNVGMTLLEQKMGD
SC16-3788-P-Cluster4-ST15A NNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLEQKMGD
SC15-1917-P-Cluster9-ST23A NNTFPGGEVFNSSELKVADATIRDWDVNEGLTGGMMTFSQGFAHSSNVGMTLLEQKMGD
SC14-0633-P-Cluter10-ST19A NNTFPSGEYFNSSEFKIADATTRDWDVNEGLTGGMMTFLQGFAHSSNVGMSLLEQKMGD
SC13-3293-P-Cluster12-ST35B NNTFPSGEYFNSSEFKIADATTRDWDVNEGLTGGMMTFLQGFAHSSNVGMSLLEQKMGD
* * * * * * *
α6 α7 α8
410 420 430 440 450 460
S.pneumoniaeR6-Cluster1 ATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIRAFTAIA
SC11-4223-P-Cluster2-ST19A ATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIRAFTAIA
SC16-3788-P-Cluster4-ST15A ATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAMSAFGQGISVTQTQMLRAFTAIA
SC15-1917-P-Cluster9-ST23A ATWLDYLNRFKFGIPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIRAFTAIA
SC14-0633-P-Cluter10-ST19A ATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRAFTAIA
SC13-3293-P-Cluster12-ST35B ATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRAFTAIA


β2e
β2f
α9
TT
470 480 490 500 510 520
S.pneumoniaeR6-Cluster1 NDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDPVYGTMYN
SC11-4223-P-Cluster2-ST19A NDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDPVYGTMYN
SC16-3788-P-Cluster4-ST15A NDGVMLEPKFISALYDPNDQSVRKSQKEIVGNPVSKEAASVTRDHMVMVGTDPTYGTMYN
SC15-1917-P-Cluster9-ST23A NDGVMLEPKFITALYDPNNQSVRKSQKEIVGNPVSKDAASQTRTHMVLVGTDPRYGTMYN
SC14-0633-P-Cluter10-ST19A NDGVMLEPKFISAIYDTNNQSVRKSQKEIVGKPVSEDTASLTRTNMILVGTDPLYGTMYN
SC13-3293-P-Cluster12-ST35B NDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKEAASTTRNHMILVGTDPLYGTMYN
β3
β4
β5
TT
TTT
530 540 550 560 570 580
S.pneumoniaeR6-Cluster1 HSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTDYIFSAVSMSPAENPDFILYVTV
SC11-4223-P-Cluster2-ST19A HSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTNYIFSAVSMSPAENPDFILYVTV
SC16-3788-P-Cluster4-ST15A HSTGKATVNVPGQNVALKSGTAEIADEKNGGYLTGSTNNIFSVVSMHPAENPDFILYVTV
SC15-1917-P-Cluster9-ST23A HSTGKATVNVPGQNVALKSGTAEIADEKNGGYLVGSTNNIFSVVAMNPAENPDFILYVTV
SC14-0633-P-Cluter10-ST19A HYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSAVTMNPAENPDFILYVTV
SC13-3293-P-Cluster12-ST35B HYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSVVTMNPAENPDFILYVTV
TT
TT
TT
590 600 610 620 630 640
S.pneumoniaeR6-Cluster1 QQPEHYSGIQLGEFANPILERASAMKDSLNLQTTAKALEQVSQQSPYPMPSVKDISPGDL
SC11-4223-P-Cluster2-ST19A QQPEHYSGIQLGEFANPILERASAMKDSLNLQTTAKSLEQVSQQSPYPMPSVKDISPGDL
SC16-3788-P-Cluster4-ST15A QQPEHYSGIQLGEFANPILERASAMKESLNLQTTAKALEQVSQQSPYPMPSVKDISPGDL
SC15-1917-P-Cluster9-ST23A QQPEHYSGIQLGEFANPILERAVAMKDSLNLQSTAKTLDQVTNQSAYAMPSIKDISPGDL
SC14-0633-P-Cluter10-ST19A QQPEHYSGIQLGEFATPILERASAMKESLNLQSPAKNLDQVTTESSYAMPSIKDISPGDL
SC13-3293-P-Cluster12-ST35B QQPEHYSGIQLGEFATPILERASAMKESLNLQSPAKNLDKVTTESSYAMPSIKDISPGEL
C-terminal Domain
TT
TT
TT
650 660 670 680 690 700
S.pneumoniaeR6-Cluster1 AEELRRNLVQPIVVGTGTKIKNSSAEEGKNLAPNQQVLILSDKAEEVPDMYGWTKETAET
SC11-4223-P-Cluster2-ST19A AEELRRNLVQPIVVGTGTKIKNSSAEEGKNLAPNQQVLILSDKAEEVPDMYGWTKATAET
SC16-3788-P-Cluster4-ST15A AEELRRNLVQPIVVGTGTKIKNSSAEEGKNLAPNQQVLILSDKAEEVPDMYGWTKETAET
SC15-1917-P-Cluster9-ST23A AEALRRNIVQPIVVGTGTKIKESSVEEGSNLAPNQQVLLLSDKAEEVPDMYGWTKATAEA
SC14-0633-P-Cluter10-ST19A AEALRRNIVQPIVVGTGTKIKETSVEEGKNLAPNQQVLLLSDKVEEIPDMYGWTKATAEA
SC13-3293-P-Cluster12-ST35B AEALRRNIVQPIVVGTGTKIKETSVEEGTNLAPNQQVLLLSDKVEEIPDMYGWKKETAET
TT
710 720 730 740 750
S.pneumoniaeR6-Cluster1 LAKWLNIELEFQGSGSTVQKQDVRANTAIKDIKKITLTLGD
SC11-4223-P-Cluster2-ST19A FAKWLNIELEFEGSGTTVQKQDVRANTAIKDIKKITLTLGD
SC16-3788-P-Cluster4-ST15A FAKWLNIELEFEGSGSTVQKQDVRANTAIKDIKKITLTLGD
SC15-1917-P-Cluster9-ST23A FSKWLNIELVFEGSGSTVQKQDVRANTAIKDIKKITLTLGD
SC14-0633-P-Cluter10-ST19A FSKWLNIELVFEGSGSTVQKQDVRANTAIKDIKKITLTLGD
SC13-3293-P-Cluster12-ST35B FAKWLDIELEFEGSGSVVQKQDVRTNTAIKNIKKITLTLGD

5.1.4 Pattern Analysis of PBP2x Mutations

A total of eleven previously reported substitutions in PBP2x were selected for analysis based on the availability of experimental evidence supporting their role in β-lactam resistance in *S. pneumoniae*. These substitutions were previously introduced in Section 4.5 and described further in Table 10.1 of Appendix A. Positions of the eleven residues were highlighted previously in the tertiary and secondary structures of PBP2x in Figure 5.1B (Section 5.1.1) and Figure 5.3 (Section 5.1.3), respectively. For the present analysis, the eleven substitutions were analyzed individually in Table 5.2 and collectively as an 11-residue mutation profile for PBP2x in Table 5.3.

The frequencies of the eleven previously reported substitutions in PBP2x overall and by penicillin MIC are displayed in Table 5.2. The rank order of the frequency of the individual substitutions was: R384G > T338A > S389L > N605T > N514H > A369V > L546V > L364F > I371T > Q552E > H394L. Differences in the rank order of substitutions were demonstrated for each increasing penicillin MIC value. For penicillin-susceptible isolates with MICs of 0.06 µg/mL, the top three most common PBP2x substitutions were A369V (82.4%, 28/34), T338A (79.4%, 27/34) and R384G (79.4%, 27/34); the rank order of the top three most common PBP2x substitutions shifts to N514H (69.2%, 18/26) > Q552E (65.4%, 17/26) > R384G (57.7%, 15/26) in penicillin-intermediate isolates with MICs of 0.12 µg/mL. Beginning at a penicillin MIC value of 0.25 µg/mL, 10 out of 11 substitutions demonstrated apparent increasing or decreasing trends in proportion with each increasing penicillin MIC value; T338A, L364F, I371T, R384G, S389L, L546V and N605T substitutions increased from 0.25-4 µg/mL, whereas A369V, H394L and Q552E substitutions decreased over the same MIC range. The N514H substitution was

commonly present in isolates with penicillin MICs ranging from 0.06-2 µg/mL but was mostly absent in isolates with penicillin MICs of 4 µg/mL.

In total, 20 unique mutational patterns were identified when the eleven previously reported PBP2x substitutions were analyzed together. The top five most common PBP2x mutational patterns containing at least 2 substitutions were patterns 17 (n=32), 2 (n=21), 1 (n=16), 14 (n=13) and 13 (n=8). Pattern 17, characterized by the presence of the T338A, A369V, and R384G substitutions, was the PBP2x mutational profile displayed by all (31/31) cluster 2 sequences (Section 5.1.2); these sequences were commonly derived from serotype 19A isolates with penicillin MICs of 0.06 µg/mL. In contrast, pattern 14 was characterized by the presence of two substitutions, N514H and Q552E, in the more C-terminal region of the TP domain. Pattern 14 was commonly displayed by cluster 4 PBP2x sequences (serotype 15A, penicillin MIC = 0.06-0.25 µg/mL), accounting for 78.6% (11/14) sequences with this pattern. Pattern 13 was differentiated from pattern 14 by a single mutation, namely R384G. This pattern was the PBP2x mutational profile displayed by all cluster 9 sequences (serotype 23AB, penicillin MIC = 0.12 µg/mL), accounting for 87.5% (7/8) sequences with this pattern.

Patterns 1 and 2 were the predominant PBP2x mutational patterns demonstrated by penicillin-resistant *S. pneumoniae* isolates (MICs ≥2 µg/mL). These two mutational patterns were differentiated by a single mutation, namely the N514H substitution. Pattern 1 was demonstrated by all (14/14) cluster 10 PBP2x sequences (serotype 19A isolates, penicillin MIC = 4 µg/mL), plus two PBP2x sequences from unclustered serotype 19F isolates with penicillin MICs ≥ 4 µg/mL. Conversely, pattern 2 was the PBP2x mutational profile demonstrated by seven different *S. pneumoniae* serotypes (clusters 11-15); these sequences were derived from both penicillin-intermediate and -resistant isolates with penicillin MICs 1-2 µg/mL.

TABLE 5.2: Analysis of the eleven previously reported amino acid substitutions in PBP2x associated with reduced β-lactam susceptibility in *S. pneumoniae*

Amino Acid Substitution			Frequency[a]	Frequency of Substitution by Penicillin MIC[b] (µg/mL), % (n)							
Allele 1	**Pos.**	**Allele 2**	**% (n)**	**≤0.015**	**0.06**	**0.12**	**0.25**	**0.5**	**1**	**2**	**≥4**
T	**338***	**A**	38.1 (72)	1.3 (1)	79.4 (27)	19.2 (5)	11.1 (1)	55.6 (5)	77.8 (7)	100 (9)	100 (17)
L	**364**	**F**	20.1 (38)	0 (0)	0 (0)	3.9 (1)	0 (0)	44.4 (4)	77.8 (7)	100 (9)	100 (17)
A	**369**	**V**	23.8 (45)	1.3 (1)	82.4 (28)	23.1 (6)	77.8 (7)	33.3 (3)	0 (0)	0 (0)	0 (0)
I	**371**	**T**	19.6 (37)	0 (0)	0 (0)	3.9 (1)	0 (0)	33.3 (3)	77.8 (7)	100 (9)	100 (17)
R	**384**	**G**	45.5 (86)	1.3 (1)	79.4 (27)	57.7 (15)	11.1 (1)	77.8 (7)	100 (9)	100 (9)	100 (17)
S	**389**	**L**	28.6 (54)	0 (0)	2.9 (1)	19.2 (5)	55.6 (5)	88.8 (8)	100 (9)	100 (9)	100 (17)
H	**394***	**L**	5.8 (11)	0 (0)	2.9 (1)	7.7 (2)	66.7 (6)	22.2 (2)	0 (0)	0 (0)	0 (0)
N	**514**	**H**	25.9 (49)	0 (0)	17.6 (6)	69.2 (18)	22.2 (2)	66.7 (6)	77.8 (7)	100 (9)	5.9 (1)
L	**546***	**V**	22.2 (42)	0 (0)	0 (0)	3.9 (1)	44.4 (4)	44.4 (4)	77.8 (7)	100 (9)	100 (17)
Q	**552**	**E**	14.8 (28)	0 (0)	17.7 (6)	65.4 (17)	11.1 (1)	22.2 (2)	22.2 (2)	0 (0)	0 (0)
N	**605**	**T**	27 (51)	1.3 (1)	11.8 (4)	19.2 (5)	22.2 (2)	66.7 (6)	77.8 (7)	100 (9)	100 (17)

[a], Calculated from the total number of study isolates (n=189). [b], Calculated from the isolate total per MIC. *, Residues within or directly adjacent to conserved PBP2x active site motifs.

TABLE 5.3: Frequency distribution of the eleven previously reported PBP2x substitutions analyzed collectively by penicillin MIC

Pattern No.	PBP2x Amino Acid Substitutions											Frequency (n) by Penicillin MIC (µg/mL)							
	T 338 A	L 364 F	A 369 V	I 371 T	R 384 G	S 389 L	H 394 L	N 514 H	L 546 V	Q 552 E	N 605 T	≤0.015	0.06	0.12	0.25	0.5	1	2	≥4
1	+	+	-	+	+	+	-	-	+	-	+								16
2	+	+	-	+	+	+	-	+	+	-	+			1		3	7	9	1
3	+	+	-	-	+	+	-	+	+	-	+					1			
4	+	-	+	-	+	+	-	+	-	-	-					1			
5	-	-	+	-	-	-	+	+	-	-	+					1			
6	-	-	-	-	+	+	-	-	-	+	-					2	2		
7	-	-	+	-	-	-	+	-	+	-	-				2				
8	-	-	+	-	-	+	+	-	+	-	-				2				
9	-	-	+	-	-	+	+	-	-	-	-				1				
10	+	-	+	-	-	-	-	-	+	-	-				1				
11	-	-	+	-	-	+	+	-	-	-	+			2	1	1			
12	-	-	-	-	+	+	-	-	-	-	+			2	1				
13	-	-	-	-	+	-	-	+	-	+	-			8					

14	-	-	-	-	-	-	-	+	-	+	-		3	9	1
15	-	-	-	-	-	-	-	+	-	+	+		2		
16	-	-	-	-	-	-	-	-	-	+	+		1		
17	+	-	+	-	+	-	-	-	-	-	-	1	27	4	
18	-	-	+	-	-	+	+	+	-	-	+		1		
19	-	-	-	-	-	-	-	-	-	-	+	1			
20	-	-	-	-	-	-	-	-	-	-	-	74			

5.2 Characterization of PBP2b

5.2.1 Conservation Analysis and Mapping of PBP2b Mutations

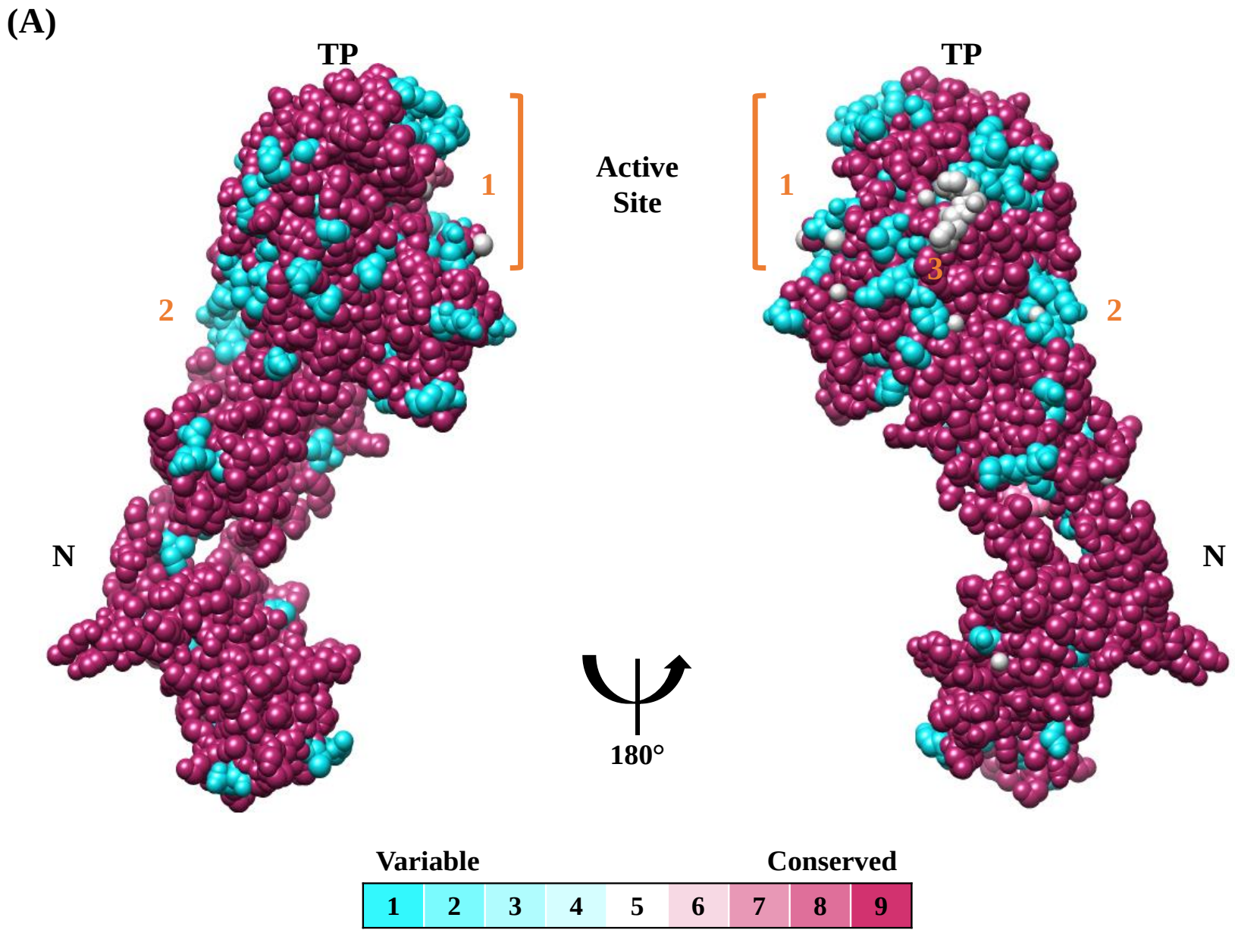
Evolutionary conservation analysis was performed to determine positions of amino acid conservation or variation in the 3D structure of PBP2b from *S. pneumoniae*. The analysis was carried out using the ConSurf server on a multiple sequence alignment of 189 sample PBP2b sequences, plus PBP2b from reference strain *S. pneumoniae* R6. The results from the ConSurf analysis of PBP2b are presented in Figures 5.4A and 5.4B. Conservation scores, ranging from 1-9, were calculated for individual amino acids in the alignment and the corresponding colour grades were projected onto the 3D structure of PBP2b from *S. pneumoniae* (PDB ID: 2WAF[89]); highly variable (rapidly evolving) residues were coloured turquoise and strictly conserved (slowly evolving) residues were coloured maroon.

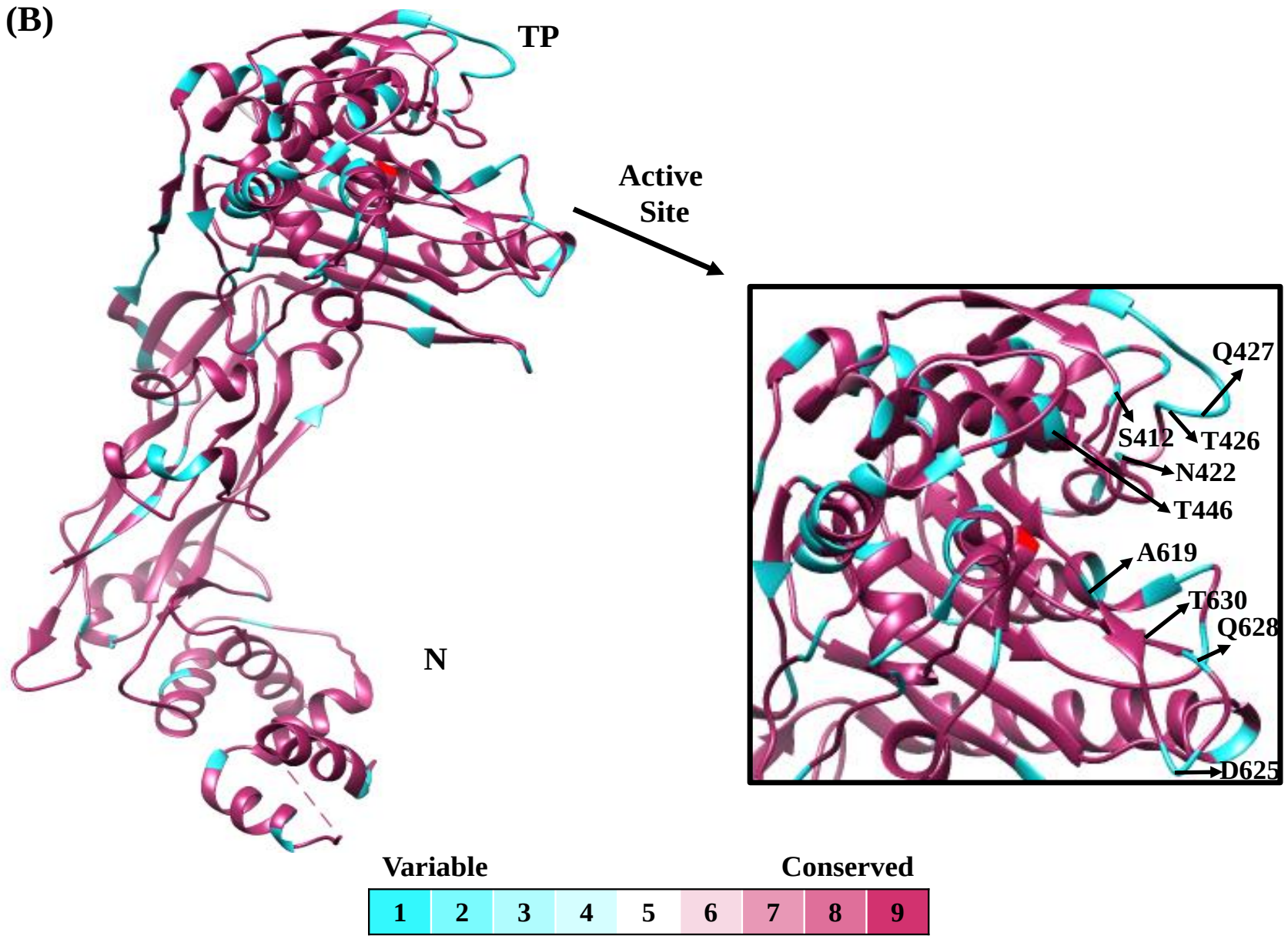
ConSurf analysis of PBP2b revealed that, overall, the core structure of the protein (comprised of N-terminal and TP domains) was mostly conserved. However, distinct regions of variability formed by the clustering of multiple highly variable residues were displayed on the surface of PBP2b in the space-filling model representation (Figure 5.4A); these highly variable regions were labelled from 1-3. Similarly to PBP2x, a notable region of variability, labelled region 1, was identified within the TP domain surrounding the entrance of the active site cleft. This shows that the TP active site region of PBP2b is rapidly evolving in *S. pneumoniae*. Highly variable region 2 was located distant to the active site at the interface of the N-terminal and TP-domains. This region was comprised of a small number of highly variable residues derived from the TP domain, spanning amino acid positions 565-569. Of note, a similar clustering of highly variable residues was demonstrated in ConSurf analysis of PBP2x (Figure 5.1A), labelled region 2. Also recognized was a clustering of highly variable residues across the mid-portion of the

PBP2b TP domain, labelled region 3. Region 3 was comprised of highly variable residues from the most C-terminal portion of the TP domain. All three highly variable regions of PBP2b were further described in Section 5.2.3.

Further shown in the ribbon model representation of PBP2b (Figure 5.4B) was the accumulation of highly variable residues within the TP domain active site cleft. Importantly, positions of all nine previously reported mutations in PBP2b (Section 4.5) were mapped within this region. Residues T446 and A619 were shown to be buried within the active site cleft near the catalytic serine residue (S386, red) of the SVVK active site motif, whereas the other seven residues were positioned at the entry of the active site. Residues S412, N422, T426 and Q427 were present on a single loop that formed the upper lip of the entry site. Conversely, residues D625, Q628 and T630, also present in a loop, decorated the lower region of the entry site. Substitutions at all nine positions were analyzed further, beginning in Section 5.2.4.

FIGURE 5.4: Amino acid conservation analysis of *S. pneumoniae* PBP2b. A multiple sequence alignment of 189 sample PBP2b sequences plus PBP2b from *S. pneumoniae* R6, generated by ClustalW, was used as input for this analysis using the ConSurf server. Conservation scores were calculated for individual amino acids in the alignment by the JTT evolutionary substitution model. Residues were assigned colour grades based on their level of conservation (scale from 1-9): highly variable residues (score = 1) were coloured turquoise and strictly conserved residues (score = 9) were coloured maroon. Scores were mapped onto the 3D structure of PBP2b from *S. pneumoniae* (PDB ID: 2WAF[89]). Figures were generated using UCSF Chimera v.1.13.1. The N-terminal and transpeptidase domains of PBP2b were labelled N and TP, respectively. (**A**) Space filling model representation of PBP2b highlighting sequence conservation on the surface of the protein. Highly variable regions were labelled 1-3. (**B**) Ribbon model representation of PBP2b highlighting sequence conservation on the interior of the protein. The ribbon position of the catalytic serine residue (S386) of the SVVK active site motif was coloured red. Positions of the nine previously reported mutations in PBP2b associated with reduced β-lactam susceptibility in *S. pneumoniae* were labelled in the enlarged image of the TP domain active site (black box).


(A)
TP
TP
1
2
3
Active
Site
N
N
180°
Variable
Conserved
1
2
3
4
5
6
7
8
9


(B)
TP
Active Site
N
Q427
S412
T426
N422
T446
A619
T630
Q628
D625
Variable
Conserved
1
2
3
4
5
6
7
8
9

5.2.2 PBP2b Phylogenetic Analysis and Cluster Identification

To assess the genetic diversity of PBP2b in *S. pneumoniae,* a maximum likelihood phylogenetic tree of 189 sample PBP2b sequences plus PBP2b from reference strain *S. pneumoniae* R6 was produced and is presented in Figure 5.5. ClusterPicker identified a total of nine PBP2b sequence clusters, containing as few as three sequences to as many as 98. Detailed information on the nine clusters of *S. pneumoniae* PBP2b sequences is provided in Table 5.4. Of the nine clusters, seven were comprised of >5 sequences; these clusters were selected for further analysis, beginning in this section.

The largest cluster identified by ClusterPicker was cluster 1, containing 98 PBP2b sequences from 23 different serotypes, including PBP2b from reference strain *S. pneumoniae* R6. Similarly to PBP2x, PBP2b cluster 1 sequences were largely unaltered (wild-type), demonstrating ≤0.6% amino acid difference compared to PBP2b of *S. pneumoniae* R6. The majority of these sequences originated from isolates with penicillin MICs ≤0.015, accounting for 78.6% (77/98) of sequences within the cluster. Cluster 1 also contained a large proportion of sequences from isolates with penicillin MICs of 0.06 µg/mL; these sequences made up the remaining 21.4% (21/98) of sequences within the cluster.

Non-wild-type PBP2b sequences clustered largely by serotype in isolates with penicillin MICs ≥0.06 µg/mL. Cluster 3 contained sequences from a subset of serotype 19A isolates with penicillin MICs of 0.06 µg/mL, representing 71.4% (10/14) of sequences within the cluster. Cluster 3 sequences demonstrated a moderately increased number of mutations compared to PBP2b of *S. pneumoniae* R6, averaging 13 substitutions per sequence with an amino acid difference of 1.9%.

Most PBP2b sequences from penicillin-intermediate isolates were contained within clusters 4, 5 and 8. Despite similarities in amino acid p-distance between clusters 4 and 5 (2.4% and 2.5%, respectively), clustering occurred separately according to serotype. The majority of cluster 4 sequences were from serotype 15A isolates with penicillin MICs of 0.12 µg/mL (68.8%, 11/16), whereas 83.3% (5/6) of cluster 5 sequences were from serotype 24F isolates. Similarly to cluster 4, cluster 8 was comprised mainly of sequences from isolates with penicillin MICs of 0.12 µg/mL, accounting for 69.2% (9/13) of sequences within the cluster. However, the predominant serotype of cluster 8 was serotype 23AB. Cluster 8 sequences demonstrated greater sequence variability compared to clusters 4 and 5, differing from PBP2b of *S. pneumoniae* R6 by 5.9%.

The majority of PBP2b sequences derived from penicillin-resistant isolates were contained within clusters 7 and 9. Cluster 7 sequences were predominantly from serotype 35B isolates with penicillin MICs ranging from 1-2 µg/mL (71.4%, 5/7). Interestingly, sequences within cluster 7 demonstrated a reduced number of mutations compared to cluster 8 sequences described above, averaging 21 substitutions per sequence. Of the seven most common PBP2b sequence clusters, cluster 9 sequences were the most distantly related to PBP2b of *S. pneumoniae* R6. Sequences within cluster 9 demonstrated an amino acid difference of 7.1% with an average of 48 substitutions per sequence. Cluster 9 was comprised exclusively of sequences from penicillin-resistant isolates (MICs 2-8 µg/mL); the predominant serotype was 19A.

FIGURE 5.5: Maximum likelihood phylogenetic tree of 190 *S. pneumoniae* PBP2b amino acid sequences. Unrooted tree was generated by MEGA-X using the JTT substitution model and 500 bootstrap replicates. Clusters were identified using ClusterPicker v.1.2.3, highlighted using FigTree v.1.4.4 and labelled from 1-9. Clusters were labelled from lowest to highest percent amino acid difference compared to PBP2b of *S. pneumoniae* R6.

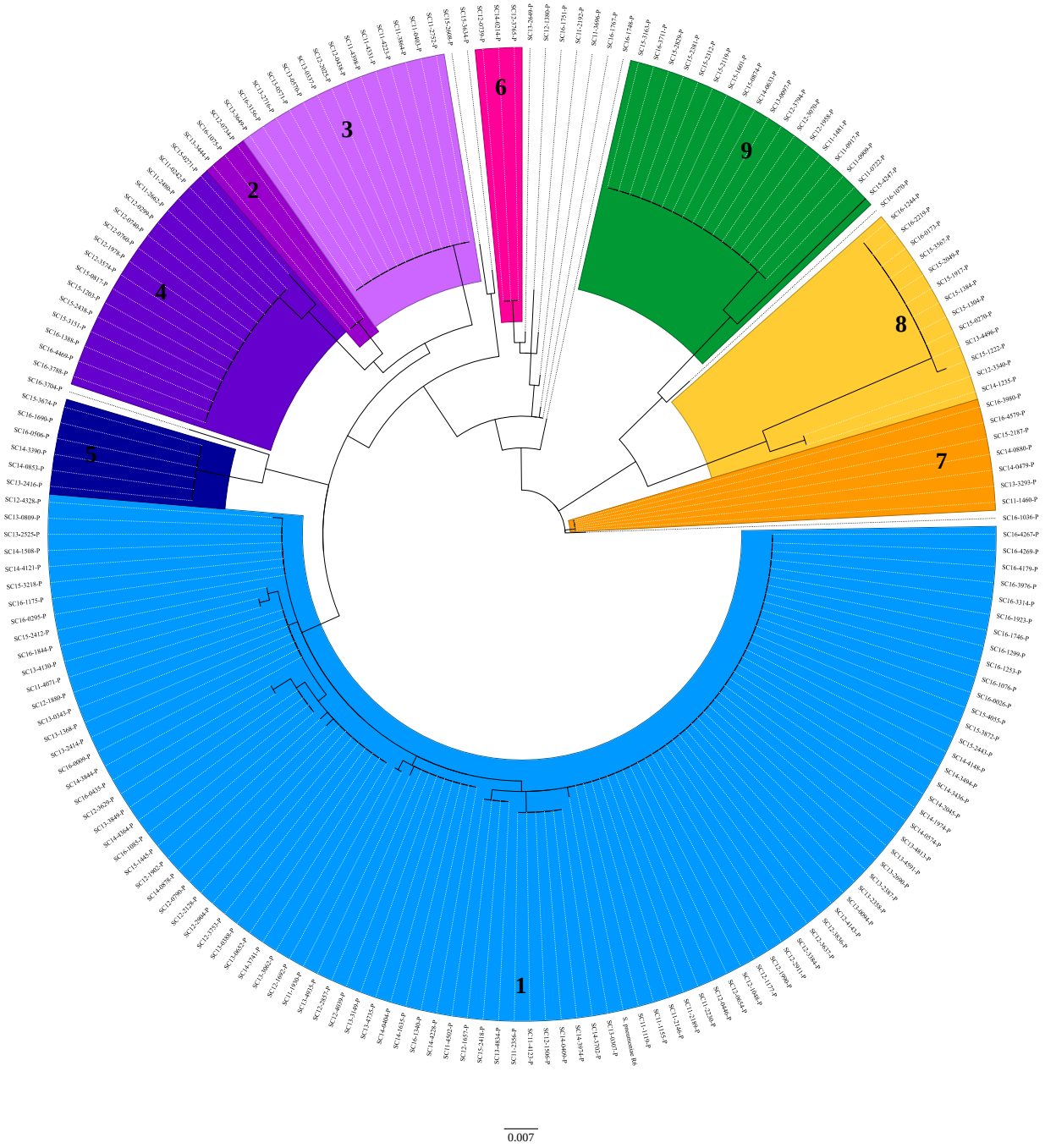


TABLE 5.4: Characterization of the nine *S. pneumoniae* PBP2b sequence clusters highlighted in Figure 5.5

Cluster No.	Serotypes (n)	Amino acid p-distance[a] (%)	Average No. of substitutions[b]	Penicillin MIC/range (µg/mL)	Ceftriaxone MIC/range (µg/mL)	Meropenem MIC/range (µg/mL)
1	19A (17), 22F (12), 3 (9), 12F (8), 8 (7), 16F (6), 7F (5), 11A (5), 6C (4), 35F (4), 9N (3), 34 (3), 6A (2), 15B (2), 20 (2) single STs: 7C, 10A, 23A, 25A, 33F, 35B, 37, 38	0-0.6	1	0.001-0.06	0.001-0.25	0.002-0.015
2	6C (2), 6A (1)	1.3	9	0.12-0.25	0.25	0.03-0.06
3	19A (13), 6A (1)	1.9	13	0.06-0.12	0.03-0.06	0.015-0.03
4	15A (14), 19A (1), 23F (1)	2.4	16	0.06-1	0.015-1	0.008-0.25
5	24F (5), 19A (1)	2.5	17	0.25-1	0.03-1	0.06-0.25
6	9V (2), 19A (1)	2.9	20	1-2	1	0.5
7	35B (5), 29 (2)	3.1	21	1-2	0.5-1	0.5-1

8	23A (7), 23B (3), 6C (2), 16F (1)	5.9	40	0.12-0.25	0.06-0.25	0.015-0.25
9	19A (16); 19F (2)	7.1	48	2-8	1-4	0.5-2
N/A	unclustered isolates[c]: 6C (4), 1 (1), 8 (1), 15A (1), 15B (1), 18C (1), 19A (1), 19F (1), 22F (1)	variable	variable	variable	variable	variable

[a], Percent amino acid difference compared to *S. pneumoniae* R6 PBP2b reference sequence. [b], Averaged from isolate total per cluster. [c], These isolates did not cluster using ClusterPicker at an initial threshold of 0.9, main support threshold of 0.9 and a genetic distance threshold of 4.5.

5.2.3 PBP2b Sequence Comparison

A structure-based multiple sequence alignment containing representative sequences from the seven most common PBP2b sequence clusters (containing >5 sequences) identified in Section 5.2.2 was prepared and is presented in Figure 5.6. Protein secondary structure elements derived from the crystal structure of PBP2b from *S. pneumoniae* (PDB ID: 2WAF[89]) were displayed above the alignment using the ESPript server. Coordinates of the two structural domains of PBP2b were labelled in Figure 5.6 as follows[89]: N-terminal domain (extracellular portion, residues 36-312) and the TP domain (residues 313-680).

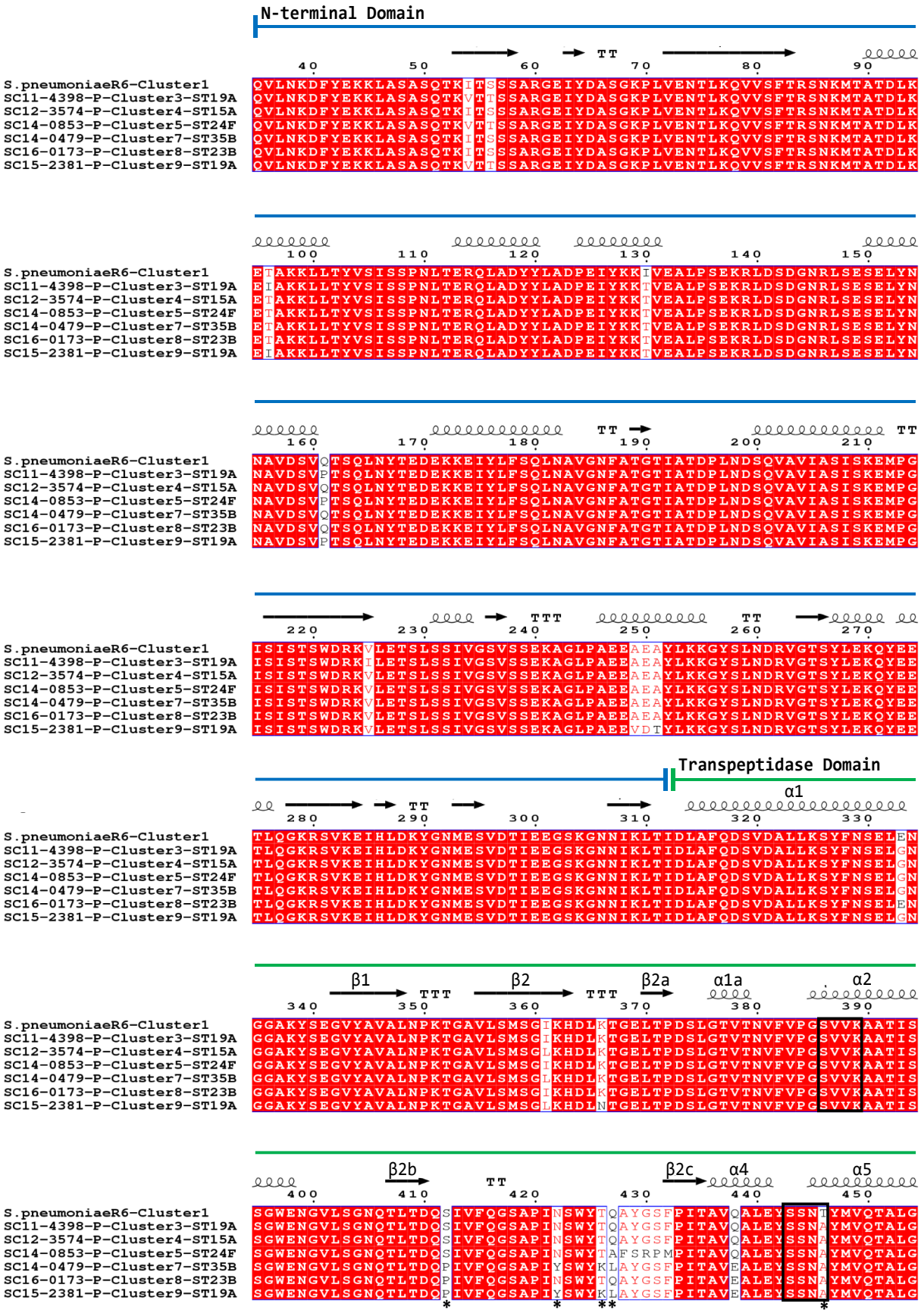
Similarly to PBP2x, comparative sequence analysis revealed that, overall, the N-terminal domain of PBP2b is largely unaltered; 96.8% (268/277) of amino acids within this region were strictly conserved amongst the seven sequences. This is in accordance with the ConSurf analysis of PBP2b (Section 5.2.1), where fewer highly variable residues were displayed within the N-terminal domain compared to the TP domain of the protein. Of the six representative non-wild-type PBP2b sequences (cluster 3, 4, 5, 7, 8 and 9 sequences) analyzed, cluster 9 sequences demonstrated the highest degree of sequence variation in the N-terminal domain, differing from the corresponding region in PBP2b of *S. pneumoniae* R6 by 2.9% (8/277).

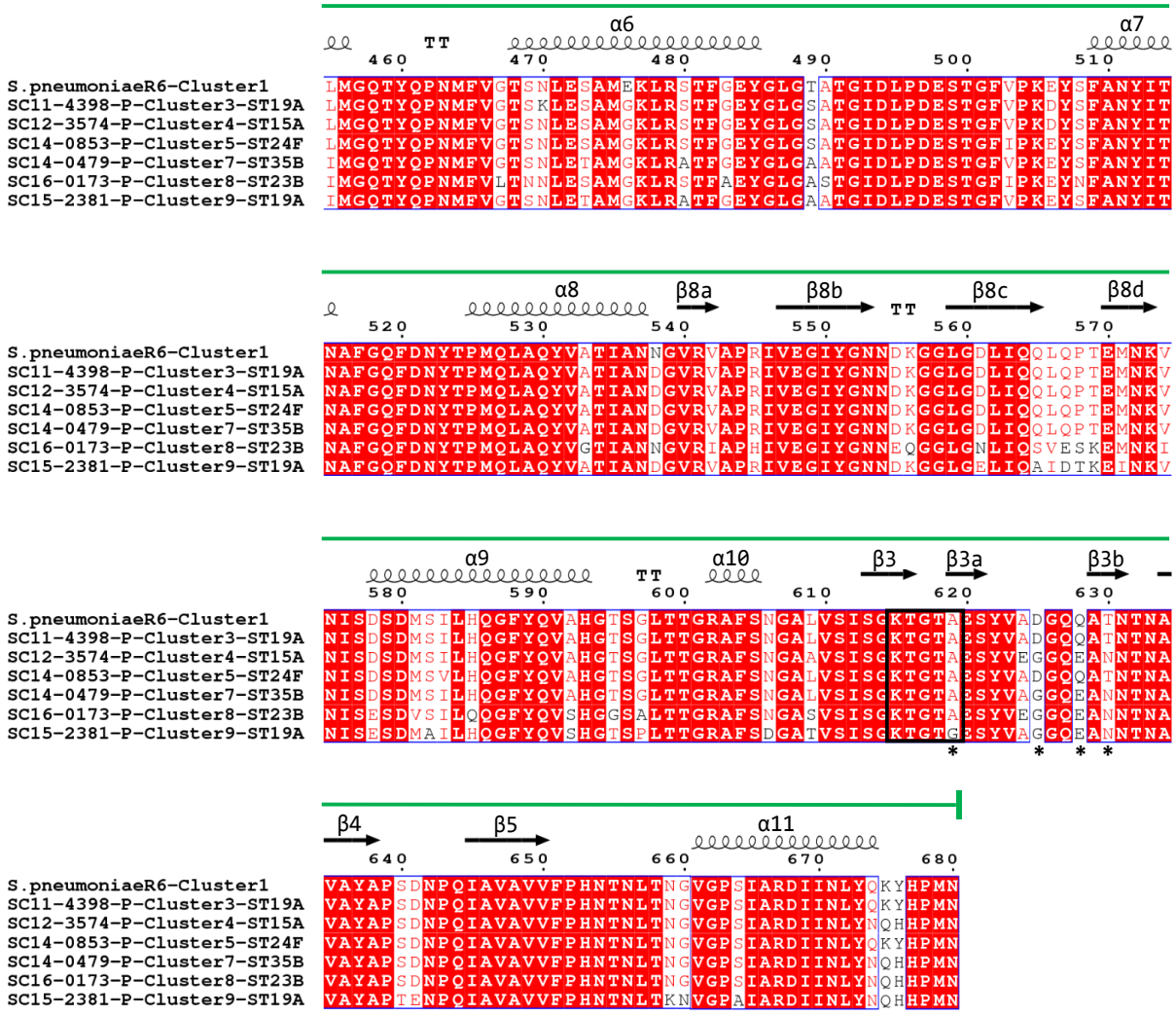
Conversely, the TP domain of PBP2b was shown to be increasingly variable in each of the six non-wild-type PBP2b sequences. Substitutions within the TP domain represented 53.9% (7/13), 93.8% (15/16), 76.5% (13/17), 95.2% (20/21), 95% (38/40) and 83.3% (40/48) of the total number of substitutions identified in cluster 3, 4, 5, 7, 8 and 9 sequences, respectively. The majority of substitutions (>60%) within this region were classified as non-conservative mutations (black characters on a white background) according to the ESPript % equivalent method for calculation of amino acid similarity type in all six non-wild-type PBP2b sequences.

Sequence variation within the TP domain compared to the PBP2b TP domain of *S. pneumoniae* R6 was 1.9% (7/368), 4.1% (15/368), 3.5% (13/368), 5.4% (20/368), 10.3% (38/368), and 10.9% (40/368) for cluster 3, 4, 5, 7, 8 and 9 sequences, respectively.

Most substitutions within the TP domain were concentrated in the mid-portion of the domain in cluster 3, 5 and 7 sequences, downstream of the SSNT active site motif. Comparatively, the majority of TP domain substitutions in cluster 4, 8 and 9 sequences were concentrated in the more C-terminal region of the TP domain, surrounding the KTGTA active site motif. This accumulation of substitutions is related to a highly variable region in the structural model of PBP2b (Figure 5.4) labelled region 3. Also noted in the alignment was a small, highly variable region in cluster 8 and 9 sequences between residues 565-569, corresponding to highly variable region 2 in Figure 5.4. Interestingly, this region was also related to the highly variable linker region of PBP2x (region 2, Figure 5.1).

FIGURE 5.6: Structure-based multiple sequence alignment of select *S. pneumoniae* PBP2b amino acid sequences. Representative sequences from the seven most common PBP2b sequence clusters (n > 5 isolates) identified in Section 5.2.2 were aligned using ClustalW and displayed using ESPript 3.0 server. Protein secondary structure elements derived from the crystal structure of PBP2b (PDB code: 2WAF[89]) were presented above the alignment: α-helices and β-strands were shown as medium squiggles and bold arrows, respectively. Protein domain coordinates were highlighted with coloured bars above the secondary structure elements. Individual amino acids were coloured according to the ESPript % equivalent colouring scheme: strictly conserved residues were shown as white characters on a red background; residues that displayed within-group similarity were shown as red characters on a white background; weakly similar residues were shown as black characters on a white background. The three conserved active site motifs of PBP2b were highlighted with black boxes. Positions of the nine previously reported mutations in PBP2b associated with reduced β-lactam susceptibility in *S. pneumoniae* were indicated with asterisks.


N-terminal Domain
Transpeptidase Domain
α1
β1
β2
β2a
α1a
α2
β2b
β2c
α4
α5
S.pneumoniaeR6-Cluster1
SC11-4398-P-Cluster3-ST19A
SC12-3574-P-Cluster4-ST15A
SC14-0853-P-Cluster5-ST24F
SC14-0479-P-Cluster7-ST35B
SC16-0173-P-Cluster8-ST23B
SC15-2381-P-Cluster9-ST19A


α6
TT
α7
460
470
480
490
500
510
S.pneumoniaeR6-Cluster1
SC11-4398-P-Cluster3-ST19A
SC12-3574-P-Cluster4-ST15A
SC14-0853-P-Cluster5-ST24F
SC14-0479-P-Cluster7-ST35B
SC16-0173-P-Cluster8-ST23B
SC15-2381-P-Cluster9-ST19A
α8
β8a
β8b
TT
β8c
β8d
520
530
540
550
560
570
α9
TT
α10
β3
β3a
β3b
580
590
600
610
620
630
β4
β5
α11
640
650
660
670
680

5.2.4 Pattern Analysis of PBP2b Mutations

A total of nine previously reported substitutions in PBP2b were selected for analysis based on the availability of experimental evidence supporting their role in β-lactam resistance in *S. pneumoniae*. These substitutions were previously introduced in Section 4.5 and described further in Table 10.2 of Appendix A. Positions of all nine residues in the tertiary and secondary structures of PBP2b were highlighted previously in Figure 5.4B (Section 5.2.1) and Figure 5.5 (Section 5.2.3), respectively. Each of the nine PBP2b substitutions were analyzed individually in Table 5.5 and collectively in Table 5.6, generating a 9-residue mutation profile for PBP2b.

The frequencies of the nine selected PBP2b substitutions overall, and by penicillin MIC are displayed in Table 5.5. The most common alteration was the T446A substitution within the SSNT active site motif; this substitution was displayed by 11.8% (13/110), 100% (53/53) and 100% (26/26) of penicillin-susceptible, -intermediate and -resistant isolates, respectively. Importantly, the T446A substitution was present in all non-wild-type patterns of substitutions identified in PBP2b sequences (patterns 1-8, Table 5.6). Pattern 8, containing T446A only, was the mutation profile displayed by all cluster 3 PBP2b sequences (Section 5.2.2); these sequences were commonly from serotype 19A isolates with penicillin MICs of 0.06 µg/mL.

The top five most common patterns of substitutions containing at least 2 of the 9 selected PBP2b substitutions were patterns 1 (n=18), 7 (n=15), 2 (n=13), 6 (n=13) and 3 (n=8). Patterns 6 and 7, characterized by the D625G, Q628E and T630N group of substitutions in the C-terminal region of the TP domain, were differentiated by the S412P substitution. Both patterns of substitutions were commonly demonstrated by penicillin-intermediate isolates. Pattern 6, containing the S412P substitution, was the common mutational profile displayed by serotype 23AB isolates with penicillin MICs of 0.12-0.25 µg/mL (cluster 8, Section 5.2.2), representing

76.9% (10/13) of sequences with this pattern. In comparison, pattern 7 was displayed exclusively by cluster 4 PBP2b sequences (Section 5.2.2); these sequences were predominantly from serotype 15A isolates with penicillin MICs of 0.12 µg/mL.

Similarly to PBP2x (Section 5.1.4), penicillin-resistant isolates commonly demonstrated two patterns of substitutions, namely patterns 1 and 2, in their PBP2b sequences. These patterns were differentiated by the A619G substitution adjacent to the KTG active site motif. Pattern 1, containing the A619G mutation, was the mutational profile displayed by all cluster 9 PBP2b sequences (Section 5.2.2) derived from penicillin-resistant serotype 19AF isolates with MICs ≥2 µg/mL. Conversely, pattern 2 was demonstrated by both penicillin-intermediate and -resistant isolates from 6 different serotypes (clusters 7 and 8, Section 5.2.2).

TABLE 5.5: Analysis of the nine previously reported amino acid substitutions in PBP2b associated with reduced β-lactam susceptibility in *S. pneumoniae*

Amino Acid Substitution			Frequency[a]	Frequency of Substitution by Penicillin MIC[b] (µg/mL), % (n)							
Allele 1	**Pos.**	**Allele 2**	**% (n)**	**≤0.015**	**0.06**	**0.12**	**0.25**	**0.5**	**1**	**2**	**≥4**
S	**412**	**P**	27.5 (52)	0 (0)	0 (0)	34.6 (9)	66.7 (6)	66.7 (6)	55.6 (5)	100 (9)	100 (17)
N	**422**	**Y**	20.6 (39)	0 (0)	0 (0)	3.9 (1)	22.2 (2)	55.6 (5)	55.6 (5)	100 (9)	100 (17)
T	**426**	**K**	20.6 (39)	0 (0)	0 (0)	3.9 (1)	22.2 (2)	55.6 (5)	55.6 (5)	100 (9)	100 (17)
Q	**427**	**L**	24.3 (46)	0 (0)	0 (0)	3.9 (1)	33.3 (3)	88.9 (8)	88.9 (8)	100 (9)	100 (17)
T	**446***	**A**	48.7 (92)	0 (0)	38.2 (13)	100 (26)	100 (9)	100 (9)	100 (9)	100 (9)	100 (17)
A	**619***	**G**	9.5 (18)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	11.1 (1)	100 (17)
D	**625**	**G**	31.2 (59)	0 (0)	2.9 (1)	80.8 (21)	66.7 (6)	33.3 (3)	44.4 (4)	77.8 (7)	100 (17)
Q	**628**	**E**	32.8 (62)	0 (0)	2.9 (1)	88.5 (23)	77.8 (7)	33.3 (3)	44.4 (4)	77.8 (7)	100 (17)
T	**630**	**N**	31.2 (59)	0 (0)	2.9 (1)	80.8 (21)	66.7 (6)	33.3 (3)	44.4 (4)	77.8 (7)	100 (17)

[a], Calculated from the total number of study isolates (n=189). [b], Calculated from the isolate total per MIC. *, Residues within or directly adjacent to conserved PBP2b active site motifs.

TABLE 5.6: Frequency distribution of the nine previously reported PBP2b substitutions analyzed collectively by penicillin MIC

Pattern No.	PBP2b Amino Acid Substitution									Frequency (n) by Penicillin MIC (µg/mL)							
	S 412 P	N 422 Y	T 426 K	Q 427 L	T 446 A	A 619 G	D 625 G	Q 628 E	T 630 N	≤0.015	0.06	0.12	0.25	0.5	1	2	≥4
1	+	+	+	+	+	+	+	+	+							1	17
2	+	+	+	+	+	-	+	+	+			1	1	2	3	6	
3	+	+	+	+	+	-	-	-	-				1	3	2	2	
4	-	-	-	+	+	-	-	-	-				1	3	3		
5	-	-	-	-	+	-	-	+	-			2	1				
6	+	-	-	-	+	-	+	+	+			8	4	1			
7	-	-	-	-	+	-	+	+	+		1	12	1		1		
8	-	-	-	-	+	-	-	-	-		12	3					
9	-	-	-	-	-	-	-	-	-	76	21						

5.3 Characterization of PBP1a

5.3.1 Conservation Analysis and Mapping of PBP1a Mutations

Evolutionary conservation analysis was performed to assess positions of amino acid conservation or variation in the 3D structure of PBP1a from *S. pneumoniae.* Conservation scores were determined for individual amino acids in a multiple sequence alignment of 189 sample PBP1a sequences, plus PBP1a from reference strain *S. pneumoniae* R6, using the ConSurf server. Scores were projected onto the 3D structure of the PBP1a TP domain from *S. pneumoniae* (PDB ID: 2C6W[95]) and presented in Figures 5.7A & 5.7B; highly variable (rapidly evolving) and strictly conserved (slowly evolving) residues were coloured turquoise and maroon, respectively.

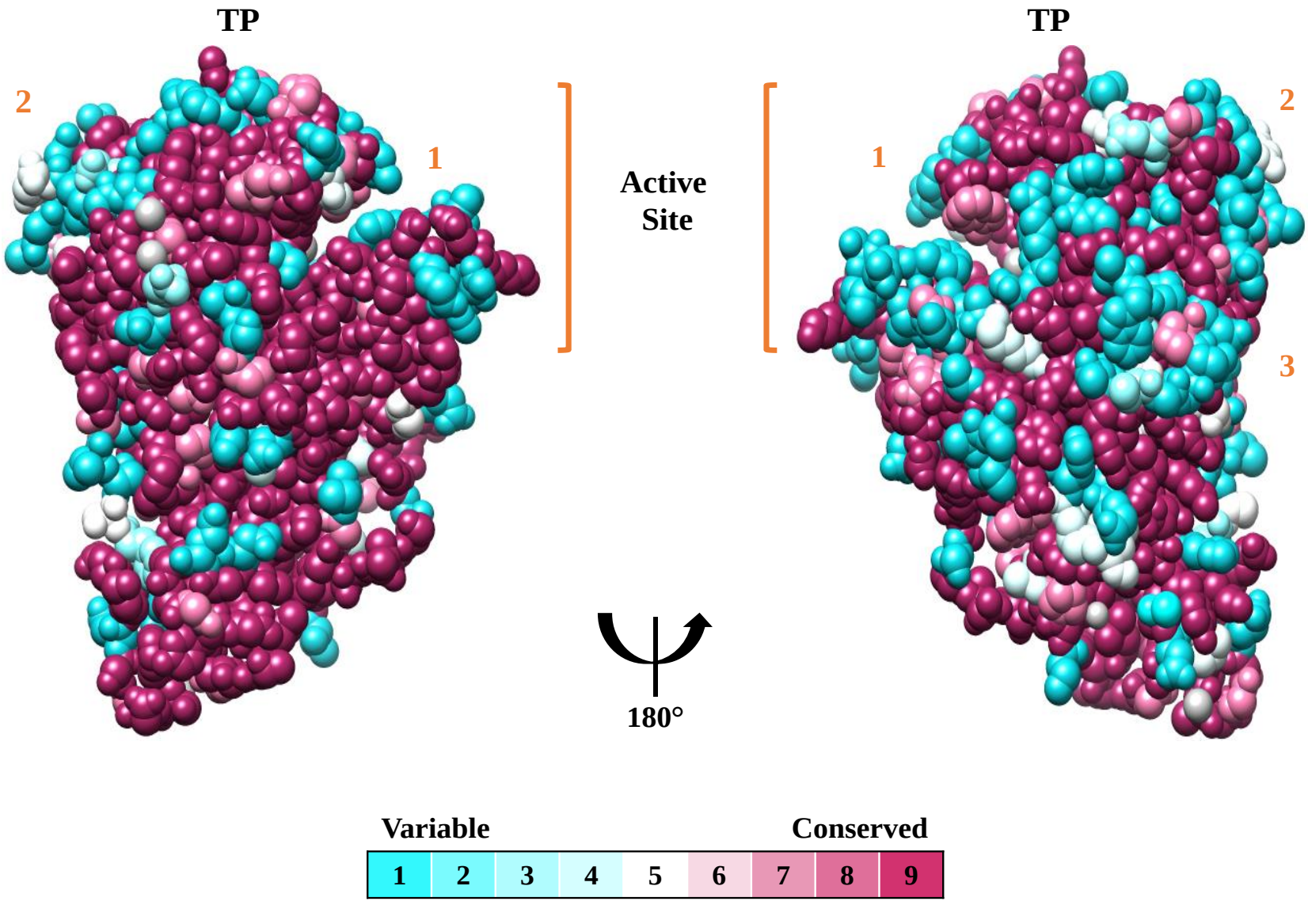
Similarly to what was observed in the ConSurf analysis of PBP2x and PBP2b (Sections 5.1.1 & 5.2.1), clusters of highly variable residues forming distinct regions of variability were displayed on the surface of PBP1a TP domain. These highly variable regions were labelled from 1-3 in the space-filling model representation of PBP1a (Figure 5.7A). Region 1 was characterized by a cluster of highly variable residues surrounding the entry of the TP domain active site. Similar modifications were shown in PBP2x and PBP2b, highlighting a common mutational mechanism amongst the three PBPs. Region 2, located opposite of the active site in the upper portion of the domain, was comprised of highly variable residues derived from a portion of the PBP1a C-terminal tail (approximately spanning residues 638-650) and a small number of residues derived from N-terminal region of the TP domain (approximately spanning residues 382-395). Also notable on the surface was a clustering of highly variable residues at the mid-point of the domain, labelled region 3. Region 3 was primarily comprised of variable

residues from the most C-terminal portion of the TP domain. These highly variable regions were described further in Section 5.3.3.

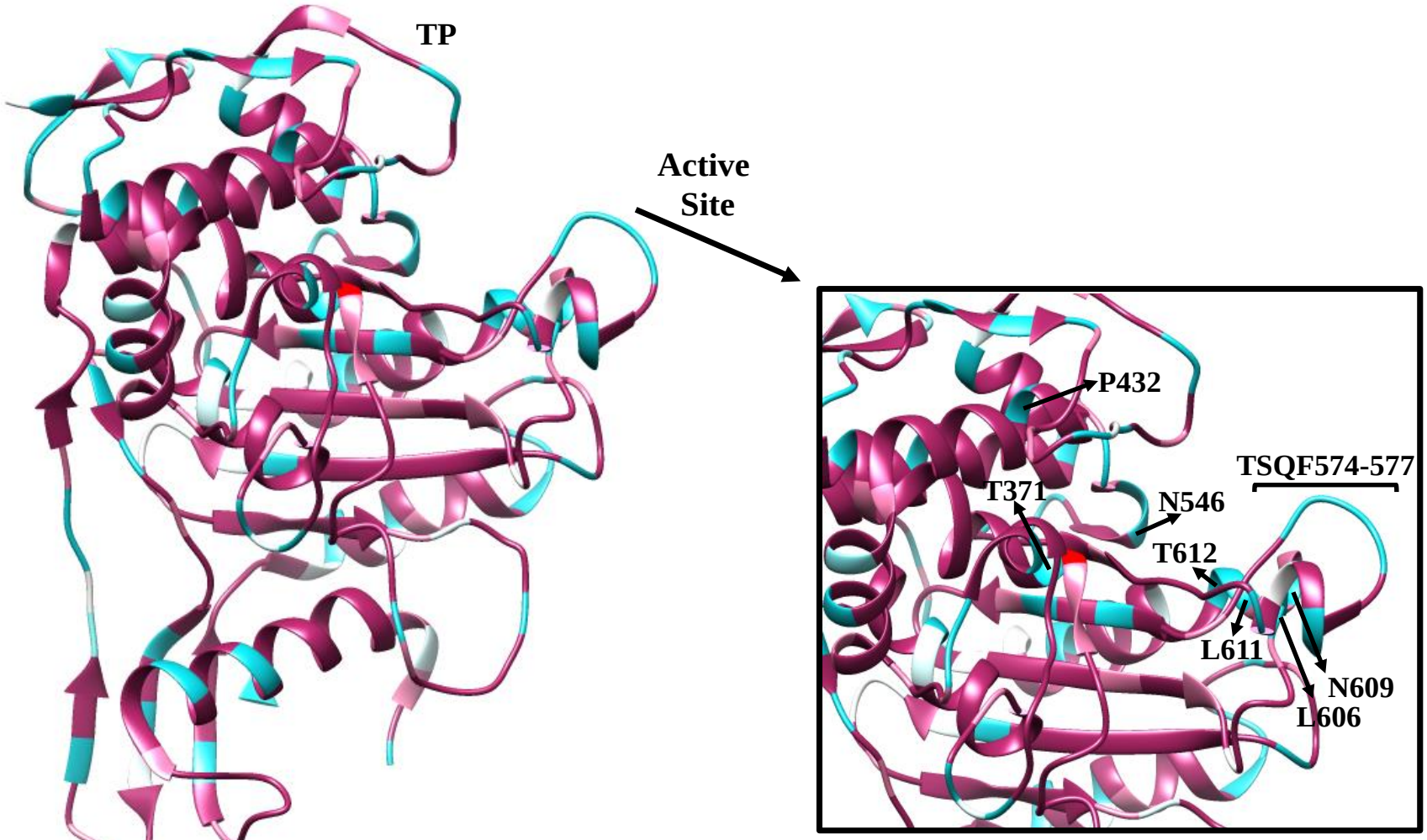
Examination of the ribbon model representation of the PBP1a TP domain (Figure 5.7B) revealed an accumulation of highly variable residues within the active site cleft. Many highly variable sites were located near the catalytic serine residue (S370, red) of the STMK active site motif. Notably, positions of all eleven previously reported substitutions in PBP1a associated with reduced β-lactam susceptibility in *S. pneumoniae* (Section 4.5) were mapped within this region (Figure 5.7B, black box). Residues T371 and P432 were buried deep within the active site cleft, whereas residues N546, TSQF574-577, L606, N609, L611 and T612 were positioned around the entry of the active site. Substitutions at all eleven positions were analyzed further in Section 5.3.4.

FIGURE 5.7: Amino acid conservation analysis of the *S. pneumoniae* PBP1a TP domain. A multiple sequence alignment of 189 sample PBP1a sequences plus PBP1a from *S. pneumoniae* R6 was generated using ClustalW and was used as input for ConSurf analysis. Conservation scores were calculated for individual amino acids by the JTT evolutionary substitution model were assigned colour grades based on their level of conservation (scale from 1-9): highly variable residues (score = 1) were coloured turquoise and strictly conserved residues (score = 9) were coloured maroon. Scores were projected onto the 3D structure of the PBP1a TP domain from *S. pneumoniae* (PDB ID: 2C6W[95]). Figures were generated using UCSF Chimera v.1.13.1. (**A**) Space-filling model representation of PBP1a TP domain highlighting sequence conservation/variation on the surface of the protein. Highly variable regions were labelled 1-3. (**B**) Ribbon model representation of the PBP1a TP domain highlighting sequence conservation on the interior. Positions of the eleven previously reported mutations in PBP1a associated with reduced β-lactam susceptibility in *S. pneumoniae* were labelled within the enlarged image of TP domain active site (black box).

(A)



(B)


TP
Active Site
P432
TSQF574-577
T371
N546
T612
L611
N609
L606
Variable
Conserved
1
2
3
4
5
6
7
8
9

5.3.2 PBP1a Phylogenetic Analysis and Cluster Identification

To assess the genetic diversity of PBP1a in *S. pneumoniae*, a maximum likelihood phylogenetic tree of 189 sample PBP1a sequences plus PBP2b from reference strain *S. pneumoniae* R6 was produced and is presented in Figure 5.8. In total, five clusters and two sub-clusters of PBP1a sequences were identified, containing as few as three sequences to as many as 111. Detailed information on the five clusters and two subclusters of *S. pneumoniae* PBP1a sequences is provided in Table 5.7.

Cluster 1 was considered a single cluster by ClusterPicker, containing a total of 136 PBP1a sequences from 25 different serotypes. However, by manual analysis, cluster 1 was divided into two sub-clusters based on differences in amino-acid p-distance values (compared to PBP1a of reference strain *S. pneumoniae* R6) between the two groups. Similarly to cluster 1 of PBP2x and PBP2b (Sections 5.1.2 & 5.2.2), cluster 1A of PBP1a was comprised of the largest number of sequences (111 total), including PBP1a from *S. pneumoniae* R6. Sequences within cluster 1A were largely unaltered (wild-type), demonstrating ≤0.8% amino acid difference compared to PBP1a of *S. pneumoniae* R6. In comparison to cluster 1 of PBP2x and PBP2b, PBP1a cluster 1A sequences originated from isolates demonstrating a wider range of penicillin MIC values. Cluster 1A contained a large proportion of sequences from isolates with penicillin MICs of 0.06 µg/mL (15.3%, 17/111) and 0.12 µg/mL (15.3%, 17/111), plus a single isolate with a penicillin MIC of 0.25 µg/mL. The remaining cluster 1A sequences were derived from wild-type β-lactam susceptible isolates with penicillin MICs ≤0.015 µg/mL (68.5%, 76/111).

PBP1a cluster 1B sequences were grouped separately from cluster 1A sequences due to a notable increase in the total number of mutations, averaging 19 substitutions per sequence. In

total, cluster 1B contained sequences from 5 different serotypes; the predominant serotype/penicillin MIC combination in cluster 1B was 19A and 0.06 µg/mL, respectively, accounting for 64% (16/25) of sequences within the sub-cluster. Interestingly, these sequences were derived from a different subgroup of penicillin-susceptible serotype 19A isolates (penicillin MICs of 0.06 µg/mL) than cluster 3 of PBP2b (Section 5.2.2).

Well-defined clustering by serotype was demonstrated in clusters 2 and 5. Both clusters were largely comprised of sequences from penicillin-intermediate isolates. Cluster 2 contained sequences exclusively from serotype 15A isolates that commonly demonstrated penicillin MICs of 0.12 µg/mL (80%, 8/10). These sequences differed from PBP1a of *S. pneumoniae* R6 by 5.4%, averaging 39 substitutions per sequence. Conversely, the predominant serotype of cluster 5 was serotype 24F. Cluster 5 sequences differed from PBP1a of *S. pneumoniae* R6 by 8.9%, averaging 64 substitutions per sequence.

All PBP1a sequences from penicillin-resistant isolates were contained within clusters 3, 4 and 6. Clusters 3 and 4 both contained a small number of sequences and demonstrated similar amino acid p-distance values, differing from PBP1a of *S. pneumoniae* R6 by 7.8% and 7.9%, respectively. Of the five clusters and two sub-clusters of PBP1a sequences, cluster 6 sequences were the most distantly related to PBP1a of *S. pneumoniae* R6, averaging 79 substitutions per sequence. The predominant serotype/penicillin MIC combination in cluster 6 was 19A and 4 µg/mL, respectively, accounting for 57.7% (15/26) of sequences within the cluster.

FIGURE 5.8: Maximum likelihood phylogenetic tree of 190 *S. pneumoniae* PBP1a amino acid sequences. Unrooted tree was generated by MEGA-X using the JTT substitution model and 500 bootstrap replicates. Clusters were identified using ClusterPicker v.1.2.3, highlighted using FigTree v.1.4.4 and labelled from 1AB-6. Clusters were labelled from lowest to highest percent amino acid difference compared to PBP1a of *S. pneumoniae* R6.

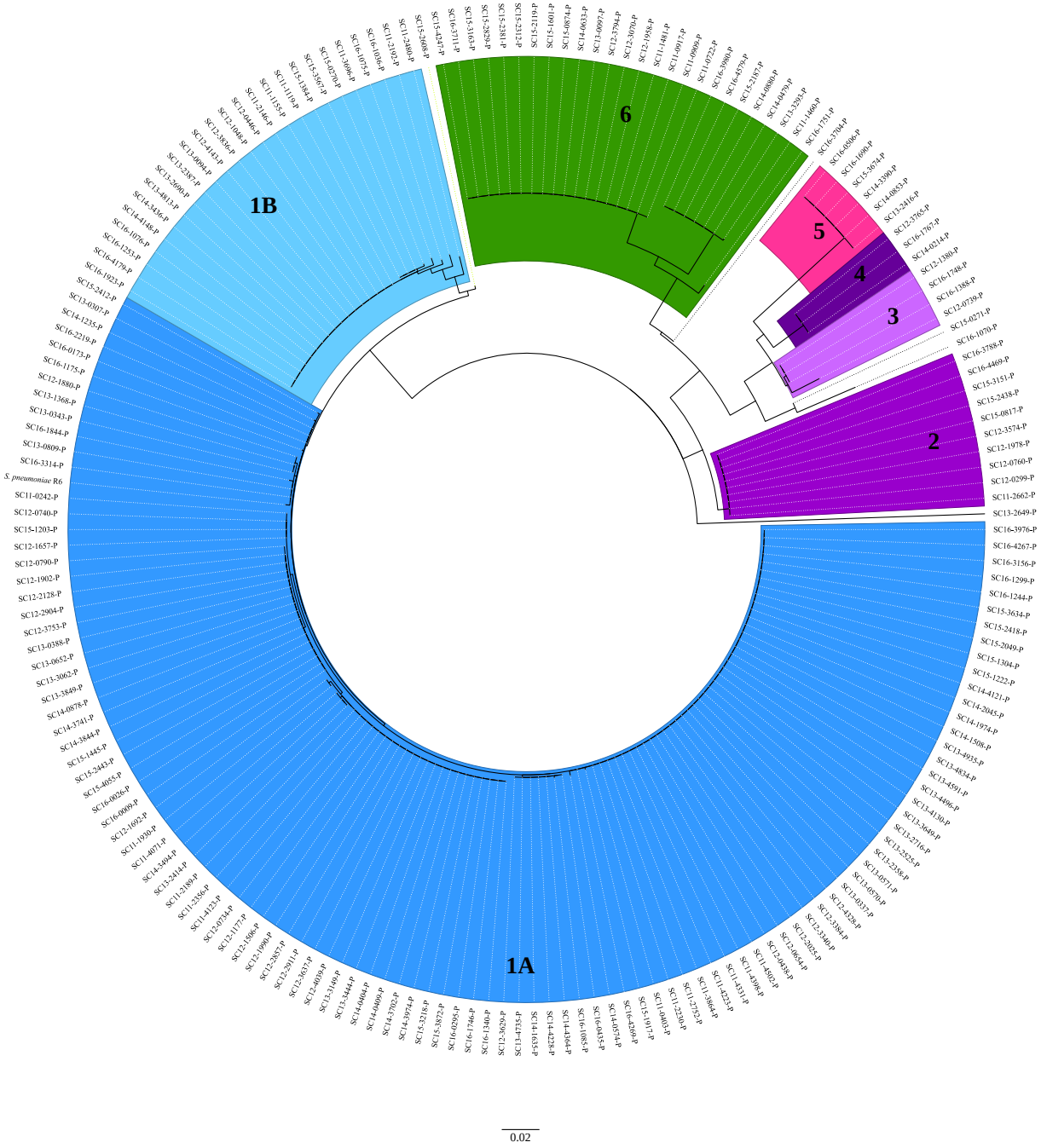


TABLE 5.7: Characterization of the five *S. pneumoniae* PBP1a sequence clusters and two sub-clusters highlighted in Figure 5.8

Cluster No.	Serotypes (n)	Amino acid p-distance[a] (%)	Average No. of Substitutions[b]	Penicillin MIC/range (µg/mL)	Ceftriaxone MIC/range (µg/mL)	Meropenem MIC/range (µg/mL)
1A[c]	19A (14), 22F (13), 3 (8), 6C (8), 12F (8), 8 (7), 16F (7), 7F (5), 11A (5), 23A (5), 35F (4), 6A (3), 9N (3), 15A (3), 23B (3), 34 (3), 15B (2), 20 (2) single STs: 7C, 10A, 25A, 33F, 35B, 37, 38	0-0.8	4	0.001-0.25	0.001-0.25	0.002-0.25
1B[c]	19A (18), 23A (3), 6C (2), 6A (1), 3 (1)	2.2-2.8	19	0.008-0.5	0.03-0.5	0.008-0.25
2	15A (10)	5.4	39	0.06-0.25	0.015-0.12	0.008-0.03
3	15A (2), 19A (1), 19F (1)	7.8	56	1-2	1	0.25-0.5

4	9V (2), 6C (1)	7.9	57	1-2	1	0.5
5	24F (5), 19A (1)	8.9	64	0.25-1	0.03-1	0.06-0.25
6	19A (16), 35B (5), 19F (2), 29 (2), 1 (1)	10.5-12.4	79	0.5-8	0.25-4	0.25-2
N/A	unclustered isolates[d]: 6C (1), 8 (1), 15B (1), 18C (1), 23F (1)	variable	variable	variable	variable	variable

[a], Percent amino acid difference compared to *S. pneumoniae* R6 PBP1a reference sequence. [b], Averaged from isolate total per cluster. [c], Cluster 1 was considered a single cluster by ClusterPicker; subgroups A and B were identified by manual analysis and were separated based on differences in amino acid p-distance from the reference. [d], These isolates did not cluster using ClusterPicker at an initial threshold of 0.9, main support threshold of 0.9 and a genetic distance threshold of 4.5.

5.3.3 PBP1a Sequence Comparison

A structure-based multiple sequence alignment containing representative sequences from the five PBP1a sequence clusters and two sub-clusters outlined in Section 5.3.2 was prepared using the ESPript 3.0 server and presented in Figure 5.9. Individual amino acids in the alignment were coloured according to the ESPript % equivalent colouring scheme and associated with protein secondary structure elements derived from the crystal structure of PBP1a from *S. pneumoniae* (PDB ID: 2C6W[95]). The coordinates of the two structural domains and the C-terminal region of PBP1a were labelled above the alignment as follows[95]: GT domain (residues 31-245), TP domain (residues 304-627) and C-terminal region (residues 628-719).

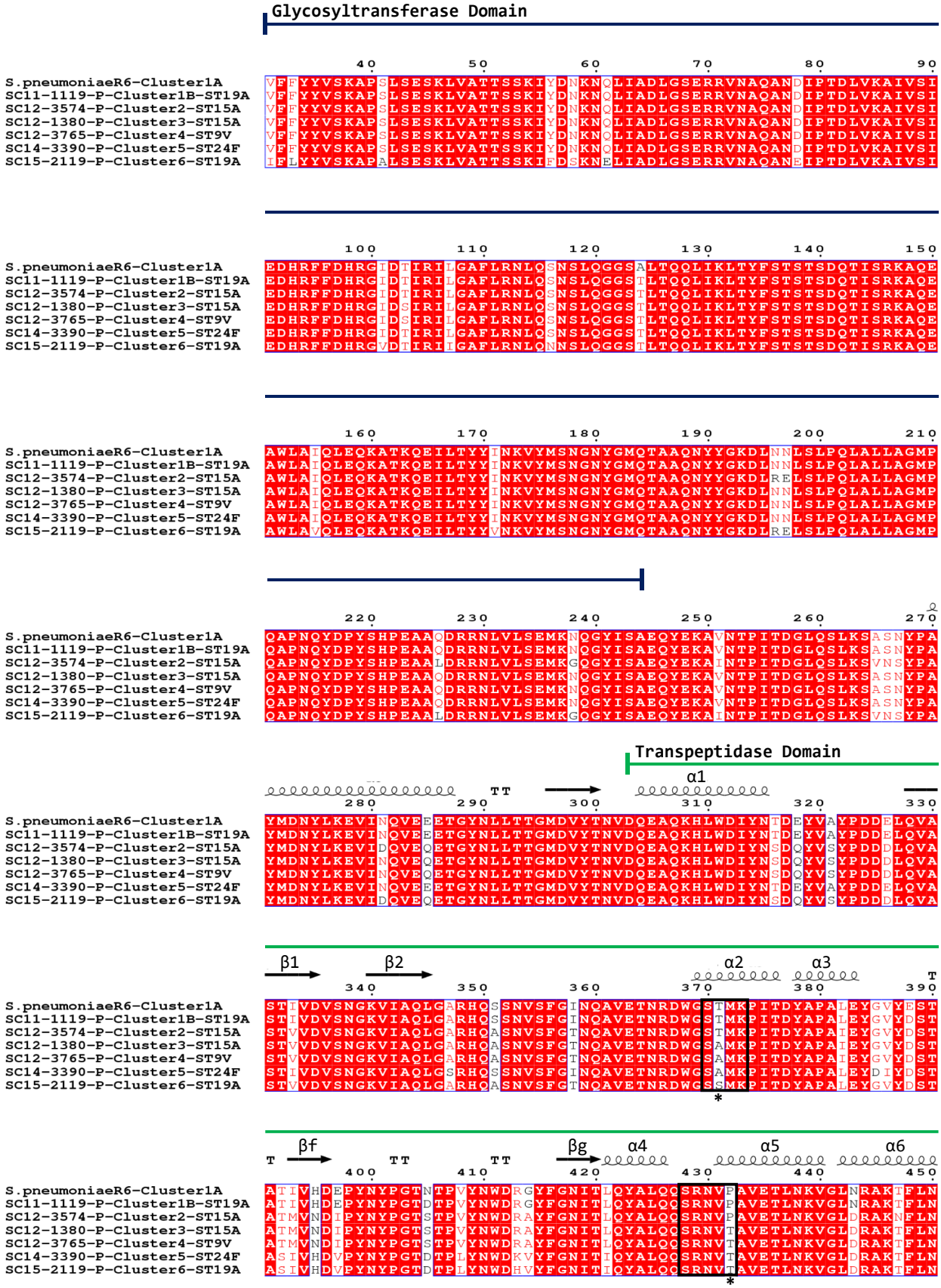
In comparison to PBP2x and PBP2b, PBP1a contains an additional functional domain, the GT domain, at the N-terminus of the protein. Based on ESPript analysis, the GT domain of PBP1a was shown to be mostly unaltered; 91.6% (197/215) of amino acids within this region were strictly conserved amongst the seven sequences. Of the six representative non-wild-type PBP1a sequences (clusters 1B, 2, 3, 4, 5 and 6), cluster 6 sequences demonstrated the highest degree of variability within the GT domain, differing from the corresponding region in PBP1a of *S. pneumoniae* R6 by 7.9% (17/215).

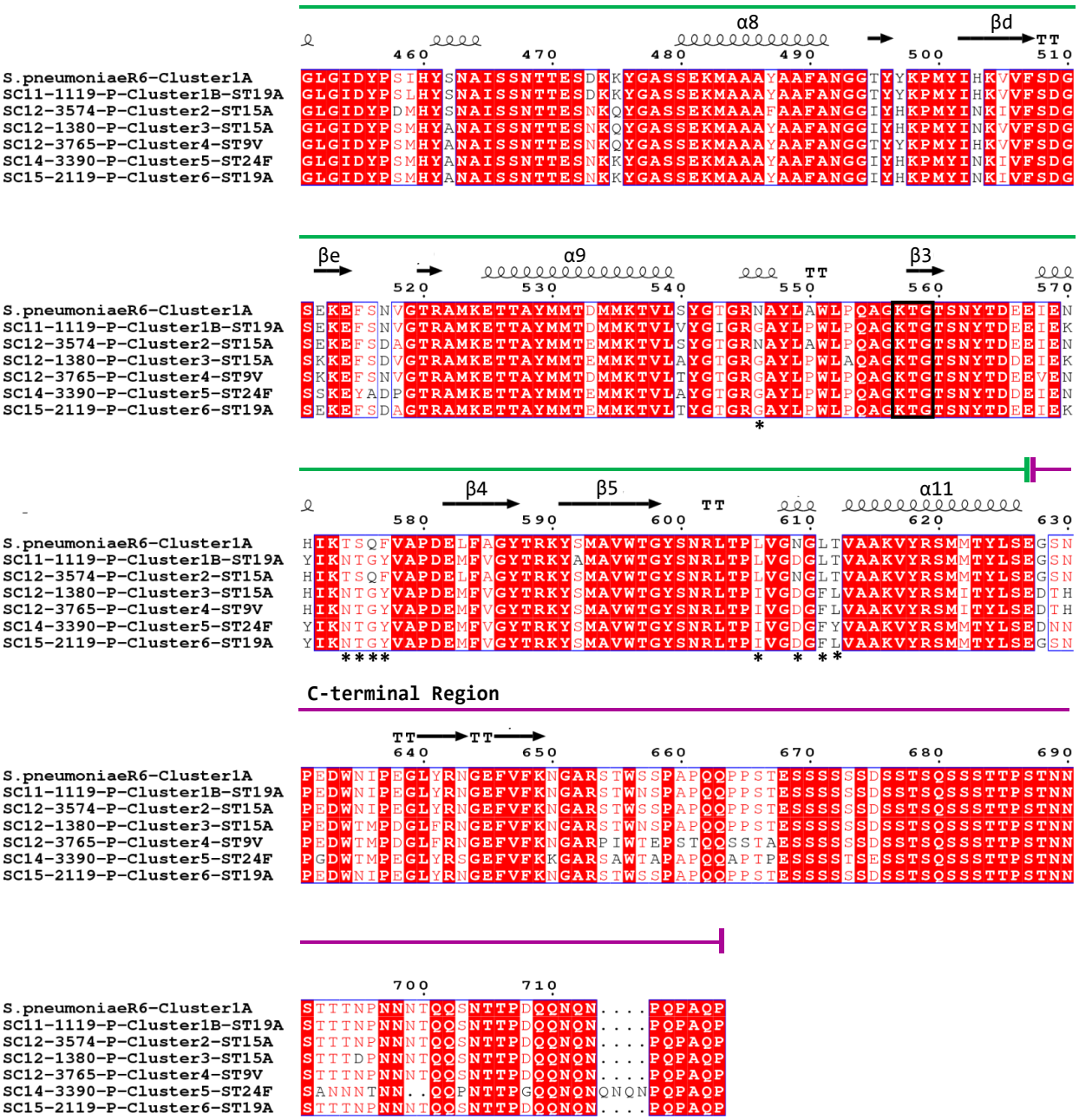
Comparatively, the TP domain of PBP1a was shown to be highly variable by both ConSurf (Section 5.3.1) and ESPript analysis. Overall, the majority of substitutions identified in the six non-wild-type PBP1a sequences were concentrated within the TP domain, representing 89.5% (17/19), 71.8% (28/39), 78.6% (44/56), 64.9% (37/57), 65.6% (42/64) and 57.3% (43/75) of the total number of substitutions in cluster 1B, 2, 3, 4, 5 and 6 sequences, respectively. Similarly to the ESPript analysis of PBP2x and PBP2b (Sections 5.1.3 & 5.2.3), over 50% of substitutions within this region were classified as non-conservative substitutions (black

characters on a white background) according to the ESPript % equivalent method for calculation of amino acid similarity type in each of the six non-wild-type PBP1a sequences. Sequence variability in the PBP1a TP domain compared to the PBP1a TP domain of *S. pneumoniae* R6 was 5.3% (17/324), 8.6% (28/324), 13.5% (44/324), 11.4% (37/324), 12.9% (42/324), and 13.3% (43/324) in cluster 1B, 2, 3, 4, 5 and 6 sequences, respectively. For cluster 1B sequences, substitutions within the TP domain were largely concentrated in the more C-terminal portion of the domain, surrounding the KTG active site motif. In comparison, substitutions within the TP domain of cluster 2 sequences were displayed in the more N-terminal to mid-regions of the domain, surrounding the STMK and SRNVP active site motifs. Substitutions within the TP domain of cluster 3, 4, 5 and 6 sequences were evenly dispersed throughout.

The C-terminal region of PBP1a was also shown to be highly variable in 3 out of 6 non-wild-type sequences; C-terminal sequences of clusters 3, 4, and 5 differed from the C-terminal region of *S. pneumoniae* R6 PBP1a by 8.7% (8/92), 17.4% (16/92) and 21.7% (20/92), respectively. This region corresponded to a distinct region of variability labelled region 2 in the structural model of PBP1a (Figure 5.7). Also noted was a highly variable region spanning residues 654-667 in cluster 4 and 6 sequences. Interestingly, cluster 5 sequences also uniquely demonstrated a 2-residue deletion and a 4-residue insertion at the most C-terminal portion of the PBP1a sequence.

FIGURE 5.9: Structure-based multiple sequence alignment of select *S. pneumoniae* PBP1a amino acid sequences. Representative sequences from the five clusters and two sub-clusters of PBP1a sequences identified in Figure 5.8 were aligned using ClustalW and displayed using ESPript 3.0 server. Secondary structure elements derived from the crystal structure of PBP1a (PDB ID: 2C6W[95]) were presented above the alignment and labelled according to Contreras-Martel *et al*.: α-helices and β-strands were shown as medium squiggles and bold arrows, respectively. The two structural domains and C-terminal region of PBP1a were highlighted with coloured bars above secondary structure elements. Individual amino acids were coloured according to the ESPript % equivalent colouring scheme: strictly conserved residues were shown as white characters on a red background; residues that displayed within-group similarity were shown as red characters on a white background; weakly similar residues were shown as black characters on a white background. The three conserved active site motifs of PBP1a were highlighted with black boxes. Positions of the eleven previously reported mutations in PBP1a associated with reduced β-lactam susceptibility in *S. pneumoniae* were indicated with asterisks.


Glycosyltransferase Domain
S.pneumoniaeR6-Cluster1A
SC11-1119-P-Cluster1B-ST19A
SC12-3574-P-Cluster2-ST15A
SC12-1380-P-Cluster3-ST15A
SC12-3765-P-Cluster4-ST9V
SC14-3390-P-Cluster5-ST24F
SC15-2119-P-Cluster6-ST19A
Transpeptidase Domain
α1
β1
β2
α2
α3
βf
βg
α4
α5
α6


α8
βd
TT
460
470
480
490
500
510
S.pneumoniaeR6-Cluster1A GLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAYAAFANGGTYYKPMYIHKVVFSDG
SC11-1119-P-Cluster1B-ST19A GLGIDYPSLHYSNAISSNTTESDKKYGASSEKMAAAYAAFANGGTYYKPMYIHKVVFSDG
SC12-3574-P-Cluster2-ST15A GLGIDYPDMHYSNAISSNTTESNKQYGASSEKMAAAFAAFANGGIYHKPMYINKIVFSDG
SC12-1380-P-Cluster3-ST15A GLGIDYPSMHYANAISSNTTESNKQYGASSEKMAAAYAAFANGGIYHKPMYINKVVFSDG
SC12-3765-P-Cluster4-ST9V GLGIDYPSMHYANAISSNTTESNKQYGASSEKMAAAYAAFANGGTYYKPMYIHKVVFSDG
SC14-3390-P-Cluster5-ST24F GLGIDYPSMHYANAISSNTTESNKKYGASSEKMAAAYAAFANGGIYHKPMYINKIVFSDG
SC15-2119-P-Cluster6-ST19A GLGIDYPSMHYANAISSNTTESNKKYGASSEKMAAAYAAFANGGIYHKPMYINKIVFSDG
βe
α9
β3
TT
520
530
540
550
560
570
S.pneumoniaeR6-Cluster1A SEKEFSNVGTRAMKETTAYMMTDMMKTVLSYGTGRNAYLAWLPQAGKTGTSNYTDEEIEN
SC11-1119-P-Cluster1B-ST19A SEKEFSNVGTRAMKETTAYMMTDMMKTVLVYGIGRGAYLPWLPQAGKTGTSNYTDEEIEK
SC12-3574-P-Cluster2-ST15A SEKEFSDAGTRAMKETTAYMMTEMMKTVLSYGTGRNAYLAWLPQAGKTGTSNYTDEEIEN
SC12-1380-P-Cluster3-ST15A SKKEFSDVGTRAMKETTAYMMTEMMKTVLAYGTGRGAYLPWLAQAGKTGTSNYTDDEIEK
SC12-3765-P-Cluster4-ST9V SKKEFSNVGTRAMKETTAYMMTDMMKTVLTYGTGRGAYLPWLPQAGKTGTSNYTDEEVEN
SC14-3390-P-Cluster5-ST24F SSKEYADPGTRAMKETTAYMMTEMMKTVLAYGTGRGAYLPWLPQAGKTGTSNYTDDEIEN
SC15-2119-P-Cluster6-ST19A SEKEFSDAGTRAMKETTAYMMTEMMKTVLTYGTGRGAYLPWLPQAGKTGTSNYTDEEIEK
*
β4
β5
TT
α11
580
590
600
610
620
630
S.pneumoniaeR6-Cluster1A HIKTSQFVAPDELFAGYTRKYSMAVWTGYSNRLTPLVGNGLTVAAKVYRSMMTYLSEGSN
SC11-1119-P-Cluster1B-ST19A YIKNTGYVAPDEMFVGYTRKYAMAVWTGYSNRLTPLVGDGLTVAAKVYRSMMTYLSEGSN
SC12-3574-P-Cluster2-ST15A HIKTSQFVAPDELFAGYTRKYSMAVWTGYSNRLTPLVGNGLTVAAKVYRSMMTYLSEGSN
SC12-1380-P-Cluster3-ST15A HIKNTGYVAPDEMFVGYTRKYSMAVWTGYSNRLTPIVGDGFLVAAKVYRSMITYLSEDTH
SC12-3765-P-Cluster4-ST9V HIKNTGYVAPDEMFVGYTRKYSMAVWTGYSNRLTPIVGDGFLVAAKVYRSMITYLSEDTH
SC14-3390-P-Cluster5-ST24F YIKNTGYVAPDEMFVGYTRKYSMAVWTGYSNRLTPIVGDGFYVAAKVYRSMMTYLSEDNN
SC15-2119-P-Cluster6-ST19A YIKNTGYVAPDEMFVGYTRKYSMAVWTGYSNRLTPIVGDGFLVAAKVYRSMMTYLSEGSN
****
* * **
C-terminal Region
TT
TT
640
650
660
670
680
690
S.pneumoniaeR6-Cluster1A PEDWNIPEGLYRNGEFVFKNGARSTWSSPAPQQPPSTESSSSSSDSSTSQSSSTTPSTNN
SC11-1119-P-Cluster1B-ST19A PEDWNIPEGLYRNGEFVFKNGARSTWNSPAPQQPPSTESSSSSSDSSTSQSSSTTPSTNN
SC12-3574-P-Cluster2-ST15A PEDWNIPEGLYRNGEFVFKNGARSTWSSPAPQQPPSTESSSSSSDSSTSQSSSTTPSTNN
SC12-1380-P-Cluster3-ST15A PEDWTMPDGLFRNGEFVFKNGARSTWNSPAPQQPPSTESSSSSSDSSTSQSSSTTPSTNN
SC12-3765-P-Cluster4-ST9V PEDWTMPDGLFRNGEFVFKNGARPIWTEPSTQQSSTAESSSSSSDSSTSQSSSTTPSTNN
SC14-3390-P-Cluster5-ST24F PGDWTMPEGLYRSGEFVFKKGARSAWTAPAPQQAPTPESSSSTSESSTSQSSSTTPSTNN
SC15-2119-P-Cluster6-ST19A PEDWNIPEGLYRNGEFVFKNGARSTWSSPAPQQPPSTESSSSSSDSSTSQSSSTTPSTNN
700
710
S.pneumoniaeR6-Cluster1A STTTNPNNNTQQSNTTPDQQNQN....PQPAQP
SC11-1119-P-Cluster1B-ST19A STTTNPNNNTQQSNTTPDQQNQN....PQPAQP
SC12-3574-P-Cluster2-ST15A STTTNPNNNTQQSNTTPDQQNQN....PQPAQP
SC12-1380-P-Cluster3-ST15A STTTDPNNNTQQSNTTPDQQNQN....PQPAQP
SC12-3765-P-Cluster4-ST9V STTTNPNNNTQQSNTTPDQQNQN....PQPAQP
SC14-3390-P-Cluster5-ST24F SANNNTNN..QQPNTTPGQQNQNQNQNPQPAQP
SC15-2119-P-Cluster6-ST19A STTTNPNNNTQQSNTTPDQQNQN....PQPAQP

5.3.4 Pattern Analysis of PBP1a Mutations

A total of eleven previously reported substitutions in PBP1a were selected for analysis based on the availability of experimental evidence supporting their role in β-lactam resistance in *S. pneumoniae*. These substitutions were previously introduced in Section 4.5 and described in detail in Table 10.3 of Appendix A. Positions of all eleven substitutions in the secondary and tertiary structures of PBP1a were highlighted previously in Figure 5.9 (Section 5.3.3) and Figure 5.7 (Section 5.3.1), respectively. Of note, the T574N, S575T, Q576G and F577Y substitutions were shown to be consistently present together in the analyzed PBP1a sequences and were subsequently described as a single 4-residue block of substitutions.

Each of the eleven selected substitutions in PBP1a was analyzed individually in Table 5.8 and collectively in Table 5.9, generating an 11-residue mutation profile for PBP1a. The rank order of the overall frequency of the individual substitutions was: TSQF574-577NTGY ≈ N609D > N546G > L606I > L611F ≈ T612L > T371A/S > P432T. Notably, the T371A/S, P432T, L606I, L611F and T612L substitutions demonstrated trends of increasing proportion, beginning at a penicillin MIC of 0.25 µg/mL. Conversely, the N546G, TSQF574-577NTGY and N609D substitutions were commonly present at lower penicillin MIC values (0.06 µg/mL).

In total, eight unique patterns of substitutions were identified in PBP1a sequences. Pattern 8, characterized by the absence of all eleven substitutions, was the most frequently identified pattern overall and was demonstrated by the majority of penicillin-susceptible isolates with (MICs ≤0.06 µg/mL, 84.5%, 93/110), as well as most penicillin-intermediate isolates with an MIC of 0.12 µg/mL (96.2%, 25/26). The top three most common patterns of substitutions containing at least 2 of the 11 selected PBP1a substitutions were patterns 1 (n=32), 5 (n=23) and 2 (n=7). Pattern 5, described as containing the TSQF574-577NTGY group of substitutions, plus

N546G and N609D, was the mutational profile demonstrated by the majority of sequences within PBP1a sub-cluster 1B (Section 5.3.2, 92%, 23/25). As described previously, these sequences largely originated from serotype 19A isolates with penicillin MICs of 0.06 µg/mL. Conversely, patterns 1 and 2 were the predominant PBP1a mutational profiles demonstrated by penicillin-intermediate and -resistant isolates with MICs ≥1 µg/mL. These patterns were differentiated by the P432T substitution in the SRN active site motif of PBP1a. Pattern 1, containing P432T, was displayed by sequences from four different sequence clusters and seven different serotypes; the majority of these sequences were from penicillin-resistant isolates with MICs ≥2 µg/mL (68.8%, 22/32). In comparison, pattern 2 was the mutational profile uniquely demonstrated by sequences within cluster 6 (Section 5.3.2, serotype 35B, penicillin MICs 1-2 µg/mL).

TABLE 5.8: Analysis of the eleven previously reported amino acid substitutions in PBP1a associated with reduced β-lactam susceptibility in *S. pneumoniae*

Amino Acid Substitution			Frequency[a]	Frequency of Substitution by Penicillin MIC[b] (µg/mL), % (n)							
Allele 1	**Pos.**	**Allele 2**	**% (n)**	**≤0.015**	**0.06**	**0.12**	**0.25**	**0.5**	**1**	**2**	**≥4**
T	371*	A/S	21.2 (40)	0 (0)	0 (0)	0 (0)	11.1 (1)	44.4 (4)	100 (9)	100 (9)	100 (17)
P	432*	T	18.5 (35)	0 (0)	0 (0)	0 (0)	11.1 (1)	66.7 (6)	66.7 (6)	55.6 (5)	100 (17)
N	546	G	34.9 (66)	1.3 (1)	47.1 (16)	0 (0)	55.6 (5)	100 (9)	100 (9)	100 (9)	100 (17)
TSQF	574-577	NTGY	36.5 (69)	1.3 (1)	47.1 (16)	3.9 (1)	77.8 (7)	100 (9)	100 (9)	100 (9)	100 (17)
L	606	I	23.3 (44)	0 (0)	0 (0)	3.9 (1)	22.2 (2)	66.7 (6)	100 (9)	100 (9)	100 (17)
N	609	D	36 (68)	1.3 (1)	47.1 (16)	3.9 (1)	77.8 (7)	88.9 (8)	100 (9)	100 (9)	100 (17)
L	611	F	22.2 (42)	0 (0)	0 (0)	0 (0)	11.1 (1)	66.7 (6)	100 (9)	100 (9)	100 (17)
T	612	L	22.2 (42)	0 (0)	0 (0)	0 (0)	11.1 (1)	66.7 (6)	100 (9)	100 (9)	100 (17)

[a], Calculated from the total number of study isolates (n=189). [b], Calculated from the isolate total per MIC. *, Residues within or directly adjacent to conserved PBP1a active site motifs.

TABLE 5.9: Frequency distribution of the eleven previously reported PBP1a substitutions analyzed collectively by penicillin MIC

Pattern No.	PBP1a Amino Acid Substitution								Frequency (n) by Penicillin MIC (µg/mL)							
	T 371 A/S	P 432 T	N 546 G	TSQF 574-577 NTGY	L 606 I	N 609 D	L 611 F	T 612 L	≤0.015	0.06	0.12	0.25	0.5	1	2	≥4
1	+	+	+	+	+	+	+	+				1	3	6	5	17
2	+	-	+	+	+	+	+	+						3	4	
3	-	+	+	+	+	+	+	+					3			
4	+	-	+	+	-	-	-	-					1			
5	-	-	+	+	-	+	-	-	1	16		4	2			
6	-	-	-	+	+	+	-	-			1	1				
7	-	-	-	+	-	+	-	-				1				
8	-	-	-	-	-	-	-	-	75	18	25	2				

Part B: Tracking the Development of β-lactam Resistance in Different *S. pneumoniae* Serotypes

5.4 Comparison of PBP2x, PBP2b and PBP1a Mutation Profiles uniquely associated with different *S. pneumoniae* serotypes by β-lactam MIC

As was shown in Part A, different *S. pneumoniae* serotypes commonly demonstrated serotype-specific patterns of mutations in their PBP2x, PBP2b and PBP1a sequences. To further demonstrate this, the previously described 11-residue, 9-residue and 11-residue mutation profiles generated for PBP2x, PBP2b and PBP1a, respectively, were analyzed collectively in four different *S. pneumoniae* serotypes, namely serotypes 19A, 15A, 6C and 35B. The serotypes that were selected for this analysis had at least one representative isolate from each of the CLSI oral penicillin V interpretive categories of susceptible (≤0.06 µg/mL), intermediate (0.12-1 µg/mL) and resistant (≥2 µg/mL) for *S. pneumoniae*. Importantly, the representative *S. pneumoniae* isolates that were used to demonstrate unique changes in the three PBPs were either randomly selected from a cohort of isolates of the same serotype/β-lactam MIC profile, or were single isolates of a specific serotype/β-lactam MIC profile.

5.4.1 Serotype 19A

The combined PBP2x, PBP2b and PBP1a mutation profiles of select *S. pneumoniae* serotype 19A isolates with equivalent and non-equivalent penicillin and ceftriaxone MICs are presented in Figures 5.10 & 5.11, respectively. In both subsets of isolates, the first recognized alterations were the T338A, A369V and R384G substitutions in PBP2x at a penicillin MIC of 0.06 µg/mL. Serotype 19A isolates with equivalent penicillin and ceftriaxone MICs of 0.06 µg/mL also demonstrated key alterations in PBP1a, namely the N546G, TSQF574-577NTGY,

and N609D substitutions. Conversely, serotype 19A isolates with non-equivalent penicillin and ceftriaxone MICs commonly displayed the PBP2b T446A substitution without alterations in PBP1a at penicillin MICs of 0.06-0.12 µg/mL.

At penicillin MICs of ≥ 0.5 µg/mL, unique changes in the mutational profiles of three PBPs were demonstrated in both subsets of serotype 19A isolates. For the subset of isolates with equivalent penicillin and ceftriaxone MICs, the maximum level of alteration in PBP2x (pattern 2, Section 5.1.4) was observed at a penicillin MIC of 0.5 µg/mL. This isolate also demonstrated the S412P, N422Y, T426K and Q427L group of substitutions in PBP2b that were previously shown to be located at the entrance of the TP domain active site (Section 5.2.1). Interestingly, the isolate with penicillin and ceftriaxone MICs values of 1 µg/mL demonstrated similar mutational patterns in PBP2x and PBP2b; however, PBP1a was completely altered (pattern 1, Section 5.3.4) in this isolate. Notably, the only additional alterations demonstrated by the two penicillin-resistant isolates with penicillin and ceftriaxone MICs ≥2 µg/mL were the PBP2b substitutions A619G, D625G, Q628E and T630N.

A similar step-wise accumulation of alterations in the three PBPs was demonstrated by the subset of serotype 19A isolates with non-equivalent penicillin and ceftriaxone MICs. PBP1a was highly altered in the three selected isolates of this subset with penicillin MIC ≥0.5 µg/mL. Additional alterations in both PBP2x and PBP2b were demonstrated at penicillin MICs ≥2 µg/mL. Similarly to what was demonstrated by serotype 19A isolates with equivalent penicillin and ceftriaxone MICs, the A619G, D625G, Q628E and T630N group of substitutions in PBP2b were the last notable alterations in these isolates (penicillin MIC = 4 µg/mL). Notably absent in this subset of serotype 19A isolates was the commonly identified N514H substitution in PBP2x.

FIGURE 5.10: Combined PBP2x, PBP2b and PBP1a mutation profiles of six selected *S. pneumoniae* serotype 19A isolates with equivalent penicillin and ceftriaxone MICs

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP2x Substitutions										
				T338A	L364F	A369V	I371T	R384G	S389L	H394L	N514H	L546V	Q552E	N605T
SC12-4328-P	0.015	0.015	0.008	-	-	-	-	-	-	-	-	-	-	-
SC12-4143-P	0.06	0.06	0.015	+	-	+	-	+	-	-	-	-	-	-
SC11-2192-P	0.5	0.5	0.12	+	+	-	+	+	+	-	+	+	-	+
SC12-0739-P	1	1	0.5	+	+	-	+	+	+	-	+	+	-	+
SC15-4247-P	2	2	0.5	+	+	-	+	+	+	-	+	+	-	+
SC12-3070-P	4	4	1	+	+	-	+	+	+	-	+	+	-	+

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP2b Substitutions								
				S412P	N422Y	T426K	Q427L	T446A	A619G	D625G	Q628E	T630N
SC12-4328-P	0.015	0.015	0.008	-	-	-	-	-	-	-	-	-
SC12-4143-P	0.06	0.06	0.015	-	-	-	-	-	-	-	-	-
SC11-2192-P	0.5	0.5	0.12	+	+	+	+	+	-	-	-	-
SC12-0739-P	1	1	0.5	+	+	+	+	+	-	-	-	-
SC15-4247-P	2	2	0.5	+	+	+	+	+	+	+	+	+
SC12-3070-P	4	4	1	+	+	+	+	+	+	+	+	+

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP1a Substitutions							
				T371A/S	P432T	N546G	TSQF574-577NTGY	L606I	N609D	L611F	T612L
SC12-4328-P	≤0.015	≤0.015	≤0.008	-	-	-	-	-	-	-	-
SC12-4143-P	0.06	0.06	≤0.015	-	-	+	+	-	+	-	-
SC11-2192-P	0.5	0.5	0.12	+	-	+	+	-	-	-	-
SC12-0739-P	1	1	0.5	+	+	+	+	+	+	+	+
SC15-4247-P	2	2	0.5	+	+	+	+	+	+	+	+
SC12-3070-P	4	4	1	+	+	+	+	+	+	+	+

FIGURE 5.11: Combined PBP2x, PBP2b and PBP1a mutation profiles of six selected *S. pneumoniae* serotype 19A isolates with non-equivalent penicillin and ceftriaxone MICs

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP2x Substitutions										
				T338A	L364F	A369V	I371T	R384G	S389L	H394L	N514H	L546V	Q552E	N605T
SC12-4328-P	0.015	0.03	0.008	-	-	-	-	-	-	-	-	-	-	-
SC11-2752-P	0.06	0.03	0.015	+	-	+	-	+	-	-	-	-	-	-
SC13-0571-P	0.12	0.06	0.03	+	-	+	-	+	-	-	-	-	-	-
SC11-4254-P	0.5	0.25	0.06	-	-	+	-	-	+	-	-	-	-	+
SC11-0902-P	2	1	0.5	+	+	-	+	+	+	-	-	+	-	+
SC11-0917-P	4	1	1	+	+	-	+	+	+	-	-	+	-	+

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP2b Substitutions								
				S412P	N422Y	T426K	Q427L	T446A	A619G	D625G	Q628E	T630N
SC12-4328-P	0.015	0.03	0.008	-	-	-	-	-	-	-	-	-
SC11-2752-P	0.06	0.03	0.015	-	-	-	-	+	-	-	-	-
SC13-0571-P	0.12	0.06	0.03	-	-	-	-	+	-	-	-	-
SC11-4254-P	0.5	0.25	0.06	+	-	-	-	+	-	-	-	-
SC11-0902-P	2	1	0.5	+	+	+	+	+	-	-	-	-
SC11-0917-P	4	1	1	+	+	+	+	+	+	+	+	+

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP1a Substitutions							
				T371A/S	P432T	N546G	TSQF574-577NTGY	L606I	N609D	L611F	T612L
SC12-4328-P	0.015	0.03	0.008	-	-	-	-	-	-	-	-
SC11-2752-P	0.06	0.03	0.015	-	-	-	-	-	-	-	-
SC13-0571-P	0.12	0.06	0.03	-	-	-	-	-	-	-	-
SC11-4254-P	0.5	0.25	0.06	-	+	+	+	+	+	+	+
SC11-0902-P	2	1	0.5	+	+	+	+	+	+	+	+
SC11-0917-P	4	1	1	+	+	+	+	+	+	+	+

5.4.2 Serotype 15AB

The combined PBP2x, PBP2b and PBP1a mutation profiles of six representative *S. pneumoniae* serotype 15AB isolates are displayed by increasing β-lactam MIC in Figure 5.12. Of note, the only serotype 15B isolate selected for this analysis was wild-type β-lactam susceptible isolate SC15-2318-P. The first recognized alterations were the N514H and Q552E substitutions in PBP2x; these two substitutions were commonly observed together in serotype 15A isolates with penicillin MICs ranging from 0.06-0.25 μg/mL (pattern 14, Section 5.1.4). Interestingly, serotype 15A isolates that displayed this pattern of mutations in PBP2x also uniquely demonstrated the T446A, D625G, Q628E and T630N substitutions in PBP2b (pattern 7, Section 5.2.4), with unaltered PBP1a proteins. These patterns of substitutions in PBP2x and PBP2b were different than what was displayed by serotype 19A isolates with similar penicillin MIC values (Section 5.4.1).

A total of two serotype 15A isolates were available with penicillin MICs ≥1 μg/mL. Both isolates demonstrated highly altered PBP2x and PBP1a proteins and differed only in their mutational profiles of PBP2b. The lone penicillin-resistant serotype 15A isolate (MIC = 2 μg/mL) additionally demonstrated the S412P, N422Y, T426K and Q427L group of substitutions in PBP2b. Importantly, this observed accumulation of a specific group of mutations in PBP2b after PBP2x and PBP1a reached their maximum number of substitutions was similarly demonstrated in serotype 19A isolates; however, the groups of substitutions in PBP2b were different between the two serotypes. Notably absent in serotype 15A isolates was the A619G substitution in PBP2b.

FIGURE 5.12: Combined PBP2x, PBP2b and PBP1a mutation profiles of six selected *S. pneumoniae* serotype 15AB isolates displayed by increasing penicillin MIC values

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP2x Substitutions										
				T338A	L364F	A369V	I371T	R384G	S389L	H394L	N514H	L546V	Q552E	N605T
SC15-2318-P	0.015	0.015	0.008	-	-	-	-	-	-	-	-	-	-	-
SC12-1978-P	0.06	0.03	0.015	-	-	-	-	-	-	-	+	-	+	-
SC12-0299-P	0.25	0.06	0.015	-	-	-	-	-	-	-	+	-	+	-
SC14-1380-P	0.25	0.5	0.06	+	+	-	+	+	+	-	+	+	-	+
SC16-1388-P	1	1	0.25	+	+	-	+	+	+	-	+	+	-	+
SC12-1380-P	2	1	0.5	+	+	-	+	+	+	-	+	+	-	+

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP2b Substitutions								
				S412P	N422Y	T426K	Q427L	T446A	A619G	D625G	Q628E	T630N
SC15-2318-P	0.015	0.015	0.008	-	-	-	-	-	-	-	-	-
SC12-1978-P	0.06	0.03	0.015	-	-	-	-	+	-	+	+	+
SC12-0299-P	0.25	0.06	0.015	-	-	-	-	+	-	+	+	+
SC14-1380-P	0.25	0.5	0.06	-	-	-	-	+	-	+	+	+
SC16-1388-P	1	1	0.25	-	-	-	-	+	-	+	+	+
SC12-1380-P	2	1	0.5	+	+	+	+	+	-	+	+	+

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP1a Substitutions							
				T371A/S	P432T	N546G	TSQF574-577NTGY	L606I	N609D	L611F	T612L
SC15-2318-P	0.015	0.015	0.008	-	-	-	-	-	-	-	-
SC12-1978-P	0.06	0.03	0.015	-	-	-	-	-	-	-	-
SC12-0299-P	0.25	0.06	0.015	-	-	-	-	-	-	-	-
SC14-1380-P	0.25	0.5	0.06	-	-	-	-	-	-	-	-
SC16-1388-P	1	1	0.25	+	+	+	+	+	+	+	+
SC12-1380-P	2	1	0.5	+	+	+	+	+	+	+	+

5.4.3 Serotype 6C

The combined PBP2x, PBP2b and PBP1a mutation profiles of six representative *S. pneumoniae* serotype 6C isolates are displayed by increasing β-lactam MICs in Figure 5.13. Similarly to what was demonstrated by both subsets of serotype 19A isolates (Section 5.4.1), substitutions in PBP2x were the first PBP alterations displayed by the lone serotype 6C isolate available with a penicillin MIC of 0.06 µg/mL. The pattern of substitutions in PBP2x shifts uniquely to R384G+ and S389+ at MICs of 0.12 µg/mL and 0.25 µg/mL and 0.03 µg/mL for penicillin, ceftriaxone and meropenem, respectively. Serotype 6C isolates with this specific β-lactam MIC profile also uniquely displayed the T446A and Q628E substitutions in PBP2b.

The three available serotype 6C isolates with penicillin MICs ≥0.5 µg/mL demonstrated additional alterations in PBP2x, PBP2b and PBP1a. PBP2x was highly altered (pattern 2, Section 5.1.4) in all three isolates, similarly to penicillin-resistant serotype 19A (with equivalent penicillin and ceftriaxone MICs) and 15A isolates. Interestingly, the PBP1a mutational profile demonstrated by the two serotype 6C isolates with penicillin MICs of 0.5 µg/mL (N546G+, TSQF574-577NTGY+, and N609D+) was also demonstrated by penicillin-susceptible serotype 19A isolates with penicillin and ceftriaxone MICs of 0.06 µg/mL. In contrast to serotype 19A and 15A isolates, the maximum level of alteration in PBP1a was demonstrated in the serotype 6C isolate with the highest penicillin MIC value (MIC = 2 µg/mL).

FIGURE 5.13: Combined PBP2x, PBP2b and PBP1a mutation profiles of six selected *S. pneumoniae* serotype 6C isolates displayed by increasing penicillin MIC values

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP2x Substitutions										
				T338A	L364F	A369V	I371T	R384G	S389L	H394L	N514H	L546V	Q552E	N605T
SC14-0878-P	0.015	0.015	0.008	-	-	-	-	-	-	-	-	-	-	-
SC12-1880-P	0.06	0.03	0.008	-	-	-	-	-	-	-	-	-	+	+
SC12-0734-P	0.12	0.25	0.03	-	-	-	-	+	+	-	-	-	-	-
SC11-3696-P	0.5	0.25	0.25	+	+	-	+	+	+	-	+	+	-	+
SC16-1036-P	0.5	0.5	0.12	+	+	-	+	+	+	-	+	+	-	+
SC16-1767-P	2	1	0.5	+	+	-	+	+	+	-	+	+	-	+

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP2b Substitutions								
				S412P	N422Y	T426K	Q427L	T446A	A619G	D625G	Q628E	T630N
SC14-0878-P	0.015	0.015	0.008	-	-	-	-	-	-	-	-	-
SC12-1880-P	0.06	0.03	0.008	-	-	-	-	-	-	-	-	-
SC12-0734-P	0.12	0.25	0.03	-	-	-	-	+	-	-	+	-
SC11-3696-P	0.5	0.25	0.25	+	+	+	+	+	-	-	-	-
SC16-1036-P	0.5	0.5	0.12	+	+	+	+	+	-	+	+	+
SC16-1767-P	2	1	0.5	+	+	+	+	+	-	-	-	-

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP1a Substitutions							
				T371A/S	P432T	N546G	TSQF574-577NTGY	L606I	N609D	L611F	T612L
SC14-0878-P	0.015	0.015	0.008	-	-	-	-	-	-	-	-
SC12-1880-P	0.06	0.03	0.008	-	-	-	-	-	-	-	-
SC12-0734-P	0.12	0.25	0.03	-	-	-	-	-	-	-	-
SC11-3696-P	0.5	0.25	0.25	-	-	+	+	-	+	-	-
SC16-1036-P	0.5	0.5	0.12	-	-	+	+	-	+	-	-
SC16-1767-P	2	1	0.5	+	+	+	+	+	+	+	+

5.4.4 Serotype 35B

The combined PBP2x, PBP2b and PBP1a mutation profiles of four representative *S. pneumoniae* serotype 35B isolates are displayed by increasing β-lactam MICs in Figure 5.14. Overall, the observed step-wise accumulation of alterations in all three PBPs of serotype 35B isolates was similar to what was displayed by serotype 15A isolates (Section 5.4.2). However, the first-step alterations, likely substitutions in PBP2x, were not demonstrated due to the unavailability of serotype 35B isolates with penicillin MICs ranging from 0.06-0.12 µg/mL. PBP2x and PBP1a were shown to be maximally altered in the lone serotype 35B isolate available with a penicillin MIC of 0.25 µg/mL; notably absent was the P432T substitution in PBP1a. This isolate also displayed the T446A, D625G, Q628E, and T630N group of substitutions in PBP2b (pattern 7, Section 5.2.4), similarly to penicillin-intermediate serotype 15A isolates. Again, the maximum number of alterations in PBP2b was demonstrated at higher penicillin MIC values in serotype 35B isolates compared to PBP2x and PBP1a; the S412P, N422Y, T426K and Q427L group of substitutions in PBP2b were displayed at penicillin MICs ≥0.5 µg/mL. The A619G was also notably absent in serotype 35B isolates.

FIGURE 5.14: Combined PBP2x, PBP2b and PBP1a mutation profiles of four selected *S. pneumoniae* serotype 35B isolates displayed by increasing penicillin MIC values

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP2x Substitutions										
				T338A	L364F	A369V	I371T	R384G	S389L	H394L	N514H	L546V	Q552E	N605T
SC13-0307-P	0.015	0.015	0.008	-	-	-	-	-	-	-	-	-	-	-
SC12-1751-P	0.25	1	0.12	+	+	-	+	+	+	-	+	+	-	+
SC13-3293-P	1	1	0.5	+	+	-	+	+	+	-	+	+	-	+
SC16-3980-P	2	1	1	+	+	-	+	+	+	-	+	+	-	+

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP2b Substitutions								
				S412P	N422Y	T426K	Q427L	T446A	A619G	D625G	Q628E	T630N
SC13-0307-P	0.015	0.015	0.008	-	-	-	-	-	-	-	-	-
SC12-1751-P	0.25	1	0.12	-	-	-	-	+	-	+	+	+
SC13-3293-P	1	1	0.5	+	+	+	+	+	-	+	+	+
SC16-3980-P	2	1	1	+	+	+	+	+	-	+	+	+

Isolate No.	PEN MIC (µg/mL)	CRO MIC (µg/mL)	MER MIC (µg/mL)	PBP1a Substitutions							
				T371A/S	P432T	N546G	TSQF574-577NTGY	L606I	N609D	L611F	T612L
SC13-0307-P	≤0.015	≤0.015	≤0.008	-	-	-	-	-	-	-	-
SC12-1751-P	0.25	1	0.12	+	-	+	+	+	+	+	+
SC13-3293-P	1	1	0.5	+	-	+	+	+	+	+	+
SC16-3980-P	2	1	1	+	-	+	+	+	+	+	+

Part C: Development and Validation of Statistical Models predicting *S. pneumoniae* β-lactam MICs

5.5 Variable Reduction by PCA

PCA was the statistical method used to mathematically reduce the relatively large number of PBP-associated independent variables prior to multiple linear regression analysis[98]. Variable reduction by PCA ensured that the most parsimonious model was obtained, as well as corrected for multicollinearity through the aggregation or removal of highly correlated variables. By using this approach, a condensed set of variables was produced that conveyed approximately the same information as the larger set of measured variables for each of the three PBPs.

5.5.1 PCA of PBP2x-Associated Variables

PCA was conducted to assess the relationships amongst the eleven selected variables for PBP2x and to determine which variables clustered together into discrete components. The results of the PCA are presented in Table 5.10, demonstrating a four-component solution using the eigenvalues >1 criterion and the varimax rotation method. Component loading values, ranging from -1 to 1, were displayed for each of the rotated variables; values $\geq \pm 0.6$ were considered strong loadings. The univariate correlations of the eleven variables are displayed in Table 11.1 of Appendix B.

Component 1 demonstrated very strong (>0.9) loadings for the variables L364F, I371T, S389L, L546V and N605T, indicating that these variables are highly related and should be considered for aggregation. Conceptually, aggregation of the five substitutions into a single composite variable was appropriate as they were commonly displayed together in the majority of PBP2x sequences from penicillin-intermediate and -resistant *S. pneumoniae* isolates with MICs ≥

0.5 µg/mL (Section 5.1.4). The singular composite variable derived from Component 1 was termed *PBP2x_comp1* and was coded dichotomously; *PBP2x_comp1* was coded 1 if two or more of the five listed substitutions were present and coded 0 if the substitutions were absent in the analyzed PBP2x sequences. This variable represented PBP2x proteins that were highly altered since the majority of sequences that were coded 1 (81/105) demonstrated all five of the listed substitutions.

Component 2 demonstrated strong loadings for the variables T338A/S, A369V and R384G, suggesting aggregation of this group of substitutions. These three substitutions were commonly demonstrated together in PBP2x sequences from penicillin-susceptible and -intermediate *S. pneumoniae* isolates with MICs ranging from 0.06-0.12 µg/mL (Section 5.1.4). The composite variable derived from Component 2 of PCA was termed *PBP2x_comp2*; PBP2x sequences were coded 1 if the A369V and R384G substitutions were present and coded 0 if the two substitutions were absent. Of note, the T338A/S substitution had a strong cross-loading on Component 1 and was therefore not included in the *PBP2x_comp2* variable.

Component 3 of PCA was defined by strong loadings for the variables Q552E and N514H. This combination of substitutions was commonly displayed in PBP2x sequences from penicillin-susceptible and -intermediate *S. pneumoniae* isolates with MICs ranging from 0.06-0.25 µg/mL (Section 5.1.4). However, the N514H substitution was not considered the defining variable in the variable derived from Component 3 (*PBP2x_Q552E)* since it was also commonly identified with Component 1 substitutions (cross-loading of 0.33). PBP2x sequences were coded 1 for *PBP2x_Q552E* if the substitution was present (by default, the L364F and A369V substitutions of Components 1 and 2, respectively, were absent) and coded 0 if it was absent. Similarly to Component 3, Component 4 was also defined by a single variable, namely the

H394L substitution. The variable derived from Component 4 was termed *PBP2x_H394L* and was coded dichotomously; coded 1 if the H394L was present (by default, the L364F and Q552E substitutions of Components 1 and 2, respectively, were absent) and coded 0 if it was absent in the analyzed PBP2x sequences.

TABLE 5.10: PCA of the eleven PBP2x-associated variables displaying the loading values for each variable in a rotated component matrix

PBP2x Variables	Component Loading Values[a]				Extraction Communality[b]
	1	2	3	4	
T338A/S	0.60	0.74	-0.13	-0.19	0.96
L364F	0.96	0.15	0.02	-0.20	0.98
A369V	-0.36	0.73	-0.21	0.52	0.98
I371T	0.95	0.15	0.02	-0.20	0.97
R384G	0.52	0.74	0.11	-0.23	0.87
S389L	0.91	0.11	0.02	0.16	0.87
H394L	0.04	-0.08	-0.06	0.99	0.99
N514H	0.33	0.05	0.85	-0.05	0.83
L546V	0.95	0.13	-0.01	-0.03	0.92
Q552E	-0.33	-0.14	0.86	-0.06	0.87
N605T	0.93	0.07	0.03	0.10	0.89
Eigenvalues	5.84	1.88	1.27	1.15	
% of variance[c]	53.1	17.04	11.52	10.41	

[a], Correlation coefficients (r) between the variables and the components. [b], Measure of the relationship between a single variable and all other input variables prior to varimax rotation. [c], Percentage of the total variance among variables explained by each component.

5.5.2 PCA of PBP2b-Associated Variables

The results from the second PCA, assessing the relationships amongst the nine selected PBP2b-associated variables are presented in Table 5.11. Component loading values, ranging from -1 to 1, were displayed for each of the rotated variables; values $\geq \pm 0.6$ were considered strong loadings. Results suggest a two-component solution for the variables using the eigenvalues >1 criterion and the varimax rotation method. The univariate correlations of the nine variables are displayed in Table 11.2 of Appendix B.

Component 1 demonstrated strong loadings for the variables S412P, N422Y, T426K and Q427L, originating from the more N-terminal region of the PBP2b TP domain. This group of substitutions was aggregated into a single composite variable as they were commonly displayed in combination with T446A in PBP2b sequences from penicillin-intermediate and -resistant *S. pneumoniae* isolates with MICs ranging from 0.25-2 µg/mL (patterns 3&4, Section 5.2.4). The dichotomous variable derived from Component 1 was termed *PBP2b_comp1* was coded 1 if the Q427L and T446A substitutions were present and coded 0 if they were absent in the analyzed PBP2b sequences. The Q427L substitution was selected as the representative variable of Component 1 since it was the most prevalent of the four highly related substitutions.

Conversely, Component 2 was defined by strong loadings for the variables D625G, Q628E and T630N, originating from the C-terminal region of the PBP2b TP domain. These three substitutions were aggregated into a single composite variable as they were frequently displayed in combination with T446A in PBP2b sequences from penicillin-intermediate *S. pneumoniae* isolates with MICs ranging from 0.12-1 µg/mL (patterns 5-7, Section 5.2.4). The dichotomous variable derived from Component 2 was termed *PBP2b_comp2* and was coded 1 if the Q628E and T446A substitutions were present and coded 0 if they were absent in the analyzed

sequences; Q628E was the most prevalent of the three highly related substitutions and was therefore selected as the representative variable for Component 2.

In addition to the two composite variables described above, a third variable was derived from the T446A substitution. Despite T446A cross-loading in both Components 1 and 2, this third variable was included since a subset of PBP2x sequences from penicillin-susceptible and -intermediate *S. pneumoniae* isolates with MICs ranging from 0.06-0.12 µg/mL displayed the T446A substitution in the absence of the other eight selected substitutions (pattern 8, Section 5.2.4). This variable was termed *PBP2b_T446A* and was coded 1 if the substitution was present and coded 0 if it was absent in the analyzed PBP2b sequences. Therefore, isolates displaying only the T446A substitution were coded 0 for *PBP2b_comp1*, 0 for *PBP2b_comp2* and 1 for *PBP2b_T446A*.

TABLE 5.11: PCA of the nine PBP2b-associated variables displaying the loading values for each variable in a rotated component matrix

PBP2b Variables	Component Loading Values[a]		Extraction Communality[b]
	1	2	
S412P	0.79	0.45	0.83
N422Y	0.96	0.25	0.97
T426K	0.96	0.25	0.97
Q427L	0.94	0.19	0.93
T446A	0.44	0.67	0.64
A619G	0.59	0.34	0.47
D625G	0.27	0.95	0.97
Q628E	0.23	0.95	0.96
T630N	0.27	0.95	0.97
Eigenvalues	6.09	1.62	
% of variance[c]	67.69	17.95	

[a], Correlation coefficients (r) between the variables and the components. [b], Measure of the relationship between a single variable and all other input variables prior to varimax rotation. [c], Percentage of the total variance among variables explained by each component.

5.5.3 PCA of PBP1a-Associated Variables

The results from the third PCA, assessing the relationships amongst the eight selected PBP1a-associated variables are presented in Table 5.12. Of note, the TSQF574-577NTGY block of substitutions was considered a single variable for the analysis. Component loading values, ranging from -1 to 1, were displayed for each of the rotated variables; values $\geq \pm 0.6$ were considered strong loadings. Results suggest a two-component solution for the variables using the eigenvalues >0.8 criteria and the varimax rotation method. The eigenvalue threshold was decreased for this analysis since only one component was extracted when the threshold value was set to >1. The univariate correlations of the eight variables are displayed in Table 11.3 of Appendix B.

Component 1 demonstrated strong loadings for the variables T371A/S, P432T, L606I, L611F and T612L, suggesting that this group of variables should be aggregated. Conceptually, aggregation is valid since these five substitutions were commonly demonstrated together in PBP1a sequences from penicillin-intermediate and -resistant *S. pneumoniae* isolates with MICs $\geq$0.5 µg/mL (patterns 1-3, Section 5.3.4). The composite variable derived from Component 1 was termed *PBP1a_comp1* and was coded dichotomously; *PBP1a_comp1* was coded 1 if minimally the L611F and T612L substitutions were present and coded 0 if the two substitutions were absent in the analyzed PBP1a sequences. The L611F and T612L substitutions were selected as the defining variables out of the five Component 1 variables as they were the most representative substitutions of highly altered PBP1a sequences (Section 5.3.4).

Component 2 was defined by strong loadings for the variables N546G, TSQF574-577NTGY and N609D. This group of substitutions were commonly identified together in the absence of the five other PBP1a substitutions in sequences from penicillin-susceptible and -

intermediate isolates with MICs ranging from 0.06-0.5 µg/mL (patterns 4-7, Section 5.3.4). The composite variable derived from Component 2 termed *PBP1a_comp2* was coded 1 if the listed substitutions were present and coded 0 if they were absent in the analyzed PBP1a sequences.

TABLE 5.12: PCA of the eight PBP1a-associated variables displaying the loading values for each variable in a rotated component matrix

PBP1a Variables	Component Loading Values[a]		Extraction Communality[b]
	1	2	
T371A/S	0.87	0.42	0.93
P432T	0.86	0.34	0.85
N546G	0.45	0.87	0.95
TSQF574-577NTGY	0.41	0.91	0.99
L606I	0.86	0.46	0.94
N609D	0.41	0.91	0.99
L611F	0.89	0.43	0.99
T612L	0.89	0.43	0.99
Eigenvalues	6.81	0.81	
% of variance[c]	85.09	10.13	

[a], Correlation coefficients (r) between the variables and the components. [b], Measure of the relationship between a single variable and all other input variables prior to varimax rotation. [c], Percentage of the total variance among variables explained by each component.

5.6 Multiple Linear Regression Models Predicting *S. pneumoniae* β-lactam MICs

5.6.1 Penicillin MIC Model Development

Multiple linear regression was conducted using the forward method to determine the most statistically significant combination of PBP-associated independent variables for predicting *S. pneumoniae* penicillin MICs. Specifically, forward multiple regression was computed using all previously described dichotomous variables determined by PCA for PBP2x (*PBP2x_comp1, PBP2x_comp2, PBP2x_Q552E* and *PBP2x_H394L),* PBP2b (*PBP2b_comp1, PBP2b_comp2* and *PBP2b_T446A*) and PBP1a (*PBP1a_comp1* and *PBP1a_comp2*), as well as p-distance values coded as scale variables for the three PBP types (*PBP2x_pdis, PBP2b_pdis* and *PBP1a_pdis*). In total, eight rounds of forward regression were required to generate a model where all independent variables were contributing significantly (*P*-values ≤0.05). The model summaries for the eight models generated by forward multiple regression are presented in Table 5.13.

Despite all eight models being statistically significant with *P*-values of <0.001, model 8 was deemed the best model [F (8, 252) = 836.47, *P*-value <0.001] since it demonstrated the highest adjusted R^2 value (0.964). This indicated that most of the variance (96.4%) in the dependent variable [$\log_2$ (penicillin MIC)] was predicted by model 8 that included 8 of 12 independent variables; variables *PBP2x_comp2, PBP2b_comp1, PBP2b_comp2* and *PBP1a_pdis* were not significant contributors (*P*-values >0.05) and were excluded from this model. The unstandardized beta coefficients of each variable included in model 8 are presented in Table 5.14 and were used to generate the following equation for the prediction of *S. pneumoniae* penicillin MIC values:

$\log_2$ (penicillin MIC) = -5.96 + 0.54(PBP2x_comp1) + 0.41(PBP2x_Q552E) – 0.70(PBP2x_H394L) + 11.17(PBP2x_pdis) + 1.29(PBP2b_T446A) + 23.42(PBP2b_pdis) + 1.65(PBP1a_comp1) + 1.35(PBP1a_comp2)

The standardized beta coefficients (β-weights) of the eight independent variables included in model 8 are also presented in Table 5.19. These values were included for the comparison of weights between independent variables in predicting the dependent variable since multiple measurement scales (dichotomous and scale variables) were used in the regression. Interestingly, the two PBP1a-associated variables (*PBP1a_comp1* and *PBP1a_comp2*) demonstrated the highest β-weights (0.27 and 0.23, respectively), indicating that they were the most significant contributors in predicting penicillin MICs, whereas the four PBP2x-associated variables contributed the least. Additionally, tolerance and VIF statistics for each predictor variable were displayed in Table 5.19 to check for multicollinearity within model 8; tolerance values greater than $1 - R^2$ were acceptable for this analysis. Since all independent variables in model 8 demonstrated tolerance values <0.035 (1 - 0.965 = 0.035), multicollinearity was not a significant issue for this model.

TABLE 5.13: Forward multiple linear regression analysis summary predicting *S. pneumoniae* penicillin MIC values

Model No.	Method[a]	Variables[b]		Model Summary[c]		ANOVA	
		Entered	Excluded	R^2	ΔR^2	F	*P*-value
1	Forward	PBP2x_pdis		0.811	0.810	1073.56	<0.001
2	Forward	+ PBP1a_comp1		0.899	0.898	1108.05	<0.001
3	Forward	+ PBP2b_T446A		0.934	0.933	1166.77	<0.001
4	Forward	+ PBP1a_comp2		0.948	0.947	1120.96	<0.001
5	Forward	+ PBP2b_pdis		0.960	0.960	1195.64	<0.001
6	Forward	+ PBP2x_S394L		0.963	0.962	1050.43	<0.001
7	Forward	+ PBP2x_Q552E		0.963	0.962	919.28	<0.001
8	Forward	+ PBP2x_comp1	- PBP2x_comp2, - PBP2b_comp1, - PBP2b_comp2, - PBP1a_pdis	0.965	0.964	836.47	<0.001

[a], Independent variables with the smallest *P*-values (≤0.05) were added one at a time using the forward multiple linear regression method. [b], Dependent variable: $\log_2$ (penicillin MIC). [c], Coefficient of determination [R^2 and adjusted (Δ) R^2] values range from 0 to 1.

TABLE 5.14: Unstandardized and standardized coefficients of the eight independent variables included in the final linear regression model predicting *S. pneumoniae* penicillin MICs

Model No.	Model Variables	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		***B*[a]**	***SEB*[b]**	**β**	**t**	***P*-value[c]**	**Tolerance[d]**	**VIF[e]**
8	(Constant)	-5.957	0.061		-97.280	<0.001		
	PBP2x_comp1	0.537	0.167	0.090	3.216	0.001	0.183	5.454
	PBP2x_Q552E	0.413	0.121	0.052	3.418	0.001	0.622	1.608
	PBP2x_H394L	-0.702	0.181	-0.053	-3.884	<0.001	0.776	1.288
	PBP2x_pdis	11.170	1.985	0.157	5.626	<0.001	0.186	5.365
	PBP2b_T446A	1.292	0.141	0.215	9.182	<0.001	0.262	3.812
	PBP2b_pdis	23.423	2.441	0.197	9.594	<0.001	0.340	2.938
	PBP1a_comp1	1.650	0.152	0.267	10.835	<0.001	0.238	4.209
	PBP1a_comp2	1.352	0.125	0.230	10.804	<0.001	0.317	3.154

[a], Unstandardized beta coefficients, or slope of the associated independent variable (x) in the formula $Y = \beta_0 + \beta_i x_i + e$. [b], Standard error of beta (*B*). [c], *P*-values were considered significant if ≤0.05. [d], Tolerance = $1 - R^2$; multicollinearity is a problem if tolerance values are $< 1 - R^2$. [e], Variance inflation factor (VIF) = 1/Tolerance.

5.6.2 Penicillin MIC Model Validation

The performance of the best-determined model (model 8, Section 5.6.1) predicting *S. pneumoniae* penicillin MICs was evaluated using leave-one-out cross-validation. By this method, predicted penicillin MIC values were generated for all isolates as previously described (Section 4.6.3) and were subsequently compared to phenotypic MIC values determined by broth microdilution. The WGS-based model predicted penicillin MIC values are presented in Table 5.15. Overall, 78.6% (198/252) of penicillin MICs predicted by model 8 were exact matches with measured MIC values. Furthermore, model 8 demonstrated an EA, where the model predicted penicillin MIC was within ±1 doubling dilution of the reference method, of 97.6% (246/252). Importantly, 66.7% (36/54) of model predicted MICs that were non-exact matches (≥ ±1 doubling dilution of the measured MIC) were predicted at a dilution higher than the measured value, indicating a slight tendency of this model to over-predict the level penicillin resistance.

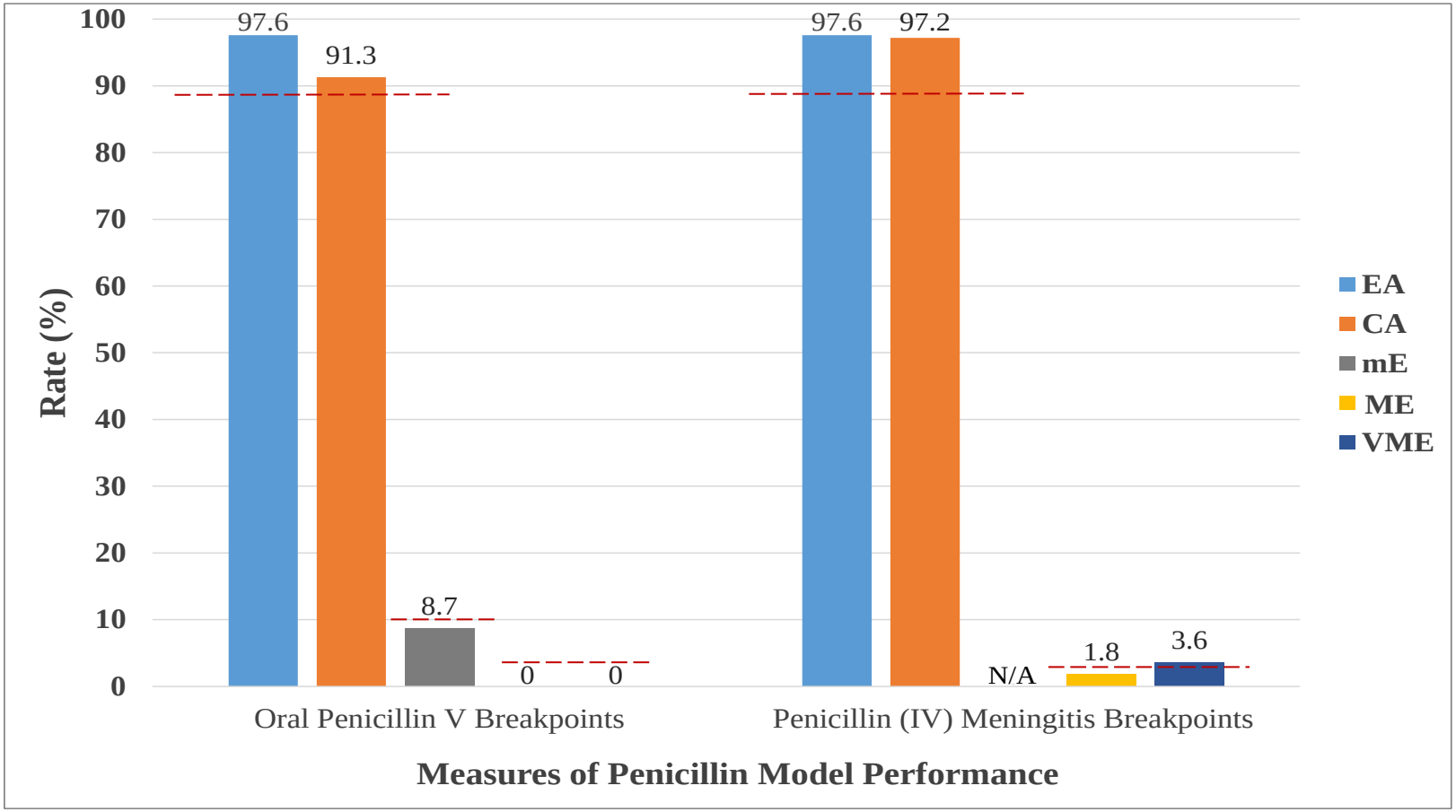
The overall error rates demonstrated by model 8 are shown in Figure 5.15 and were calculated as previously described (Section 4.6.3) according to both CLSI oral penicillin V (susceptible = ≤0.06 μg/mL, intermediate = 0.12-1 μg/mL and resistant = ≥2 μg/mL) and IV meningitis (susceptible = ≤0.06 μg/mL and resistant = ≥0.12 μg/mL) breakpoint criteria. The CA, mE, ME and VME rates demonstrated by this model were 91.3% (230/252), 8.7% (22/252), 0% (0/110) and 0% (0/63), respectively, when the oral penicillin V breakpoints were applied. The error rates of this model increased when the IV meningitis breakpoints were applied, demonstrating CA, ME and VME rates of 97.2% (245/252), 1.8% (2/113) and 3.6% (5/139), respectively. Overall, the error rates of this model were within acceptable performance specifications outlined by CLSI, except for the VME rate (>3%) when using the IV meningitis breakpoints.

TABLE 5.15: Summary of model predicted penicillin MIC values generated by leave-one-out cross-validation

		Predicted Penicillin MIC[a] (µg/mL)								
Measured Penicillin MIC[b] (µg/mL)		**≤0.015**	**0.06**	**0.12**	**0.25**	**0.5**	**1**	**2**	**4**	**Total Measured**
	≤0.015	**75**	1							76
	0.06	5	**27**	2						34
	0.12		5	**20**	10					35
	0.25			6	**9**	2	1			18
	0.5					**6**	3	3	1	13
	1					1	**2**	10		13
	2					1		**26**	3	30
	4								**33**	33
	Total Predicted	80	33	28	19	10	6	39	37	252

[a], WGS-based predicted penicillin MIC values generated by leave-one-out cross-validation of model 8. [b], Phenotypic penicillin MIC values determined by broth microdilution.

FIGURE 5.15: Accuracy of model predicted penicillin MICs compared to measured penicillin MICs for *S. pneumoniae*. The essential agreement (EA) and categorical agreement (CA), as well as minor errors (mE), major errors (ME) and very major errors (VME) between predicted and measured penicillin MICs were calculated according to CLSI oral penicillin V and IV meningitis breakpoint. The three error types were defined as follows: mE = phenotypically intermediate, WGS-prediction susceptible or resistant (and vice versa); ME = phenotypically susceptible, WGS-prediction resistant; VME = phenotypically resistant, WGS-prediction susceptible. Red dashed lines indicate threshold acceptable error rates: ≥90% for EA and CA, ≤10% for mE and <3% for ME and VME.



5.6.3 Ceftriaxone MIC Model Development

Similarly to penicillin MIC model development (Section 5.6.1), forward multiple linear regression was performed to determine the most statistically significant combination of PBP-associated independent variables for predicting *S. pneumoniae* ceftriaxone MICs. A total of twelve independent variables were selected as input for this analysis, including the nine previously described dichotomous variables determined by PCA for PBP2x, PBP2b and PBP1a (Sections 5.5.1-5.5.3), as well as three scale variables representing amino acid p-distance values for each PBP type. By the forward approach, independent variables were added individually in successive rounds of regression (7 total), starting with the variable that demonstrated the smallest *P*-value and continuing until all variables with *P*-values ≤0.05 were added. The model summaries for the seven forward regression models predicting *S. pneumoniae* ceftriaxone MICs are presented in Table 5.16.

In total, five of twelve input variables demonstrated *P*-values > 0.05 and were excluded from the final model. Interestingly, two of the five (*PBP2b_comp2* and *PBP1a_pdis*) excluded variables were also excluded in the best-determined model for penicillin (Section 5.6.1). Of the seven statistically significant models that were produced, model 7 was considered the best model as it predicted the most variance (93.6%) in the dependent variable [$\log_2$ (ceftriaxone MIC)], F (7, 238) = 411.44, P-value <0.001, adjusted R^2 = 0.936. The unstandardized coefficients of the seven predictor variables included in model 7, along with the model constant term, are presented in Table 5.17. The final linear regression equation derived from model 7 predicting *S. pneumoniae* ceftriaxone MIC values was:

$\log_2$ (ceftriaxone MIC) = -5.92 + 1.01(PBP2x_comp1) - 0.45(PBP2x_Q552E) + 28.53(PBP2x_pdis) - 0.54(PBP2b_comp1) + 15.88(PBP2b_pdis) + 1.21(PBP1a_comp1) + 0.78(PBP1a_comp2)

The standardized beta coefficients (β-weights) of the seven independent variables included in model 7 were also displayed in Table 5.21. These values were included for the comparison of weights between independent variables of different measurement scales in predicting the dependent variable. Overall, the top three most significant contributors in predicting ceftriaxone MICs were the variables *PBP2x_pdis*, *PBP1a_comp1* and *PBP2x_comp1*, demonstrating β-weights of 0.443, 0.222 and 0.191 (*P*-values <0.001), respectively. Interestingly, two of the top three variables were derived from PBP2x, highlighting the importance of PBP2x-associated genetic determinants in the prediction of *S. pneumoniae* ceftriaxone MICs. Notably, there were two significant negative predictors of ceftriaxone resistance; variables *PBP2b_comp1* and *PBP2x_Q552E* demonstrated standardized beta coefficients of -0.101 and -0.056, respectively. Additionally, tolerance and VIF statistics were computed for each of the independent variables included in the final model to check for high intercorrelations amongst variables causing multicollinearity. These values are also presented in Table 5.21; tolerance values > 1-R^2 (1 – 0.936 = 0.064) were considered acceptable for this analysis. Multicollinearity was not an issue for this model since all variables demonstrated tolerance values beyond this threshold.

TABLE 5.16: Forward multiple linear regression analysis summary predicting *S. pneumoniae* ceftriaxone MICs

Model No.	Method[a]	Variables[b]		Model Summary[c]		ANOVA	
		Entered	Excluded	R^2	ΔR^2	F	*P*-value
1	Forward	PBP2x_pdis		0.854	0.853	1376.76	<0.001
2	Forward	+ PBP1a_comp1		0.908	0.907	1159.41	<0.001
3	Forward	+ PBP2x_comp1		0.921	0.920	913.55	<0.001
4	Forward	+ PBP1a_comp2		0.928	0.927	748.89	<0.001
5	Forward	+ PBP2b_pdis		0.935	0.933	664.45	<0.001
6	Forward	+ PBP2x_Q552E		0.937	0.935	532.436	<0.001
7	Forward	+ PBP2b_comp1	- PBP2x_comp2 - PBP2b_comp2 - PBP1 _pdis - PBP2b_T446A - PBP2x_S394L	0.938	0.936	464.71	<0.001

[a], Independent variables with the smallest *P*-values (≤0.05) were added one at a time using the forward multiple linear regression method. [b], Dependent variable: $\log_2$ (ceftriaxone MIC). [c], Coefficient of determination [R^2, adjusted (Δ) R^2] values range from 0 to 1.

TABLE 5.17: Unstandardized and standardized coefficients of the seven independent variables included in the final linear regression model predicting *S. pneumoniae* ceftriaxone MICs

Model No.	Model Variables	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		***B*[a]**	***SEB*[b]**	**β**	**t**	***P*-value[c]**	**Tolerance[d]**	**VIF[e]**
7	(constant)	-5.921	0.072		-82.808	<0.001		
	PBP2x_comp1	1.011	0.199	0.191	5.091	<0.001	0.191	5.234
	PBP2x_Q552E	-0.452	0.152	-0.056	-2.981	0.003	0.751	1.332
	PBP2x_pdis	28.533	2.491	0.443	11.457	<0.001	0.181	5.562
	PBP2b_comp1	-0.542	0.227	-0.108	-2.393	0.018	0.154	6.491
	PBP2b_pdis	15.882	2.775	0.156	5.723	<0.001	0.361	2.771
	PBP1a_comp1	1.209	0.216	0.222	5.586	<0.001	0.170	5.866
	PBP1a_comp2	0.779	0.154	0.149	5.062	<0.001	0.312	3.207

[a], Unstandardized beta coefficients, or slope of the associated independent variable (x) in the formula $Y = \beta_0 + \beta_i x_i + e$. [b], Standard error of beta (*B*). [c], *P*-values were considered significant if ≤0.05. [d], Tolerance = $1 - R^2$; multicollinearity is a problem if tolerance values are $< 1 - R^2$. [e], Variance inflation factor (VIF) = 1/Tolerance.

5.6.4 Ceftriaxone MIC Model Validation

Leave-one-out cross-validation was performed to evaluate the performance of the best-determined model (model 7) predicting *S. pneumoniae* ceftriaxone MICs. By this method, 238 rounds of cross-validation were conducted as previously described (Section 4.6.3), generating predicted ceftriaxone MIC values for all isolates in the dataset. The WGS-based predicted MIC values, displayed in Table 5.18, were subsequently compared to measured ceftriaxone MICs determined by broth microdilution. Overall, 70.6% (168/238) of predicted ceftriaxone MIC values were exact matches with measured MIC values. Comparatively, the percent exact match demonstrated by this model was 8% lower than what was demonstrated by the best-determined model for penicillin (78.6%, Section 5.6.2), indicating decreased precision in predicting ceftriaxone MICs. However, the majority of ceftriaxone MICs predicted by model 7 were within ± 1 doubling dilution of the measured MIC, demonstrating an EA of 94.5% (225/238). Importantly, this model demonstrated a slight tendency to under-predict the level of ceftriaxone resistance; 54.3% (38/70) of model predicted MICs that were non-exact matches ($\geq \pm 1$ doubling dilution of the measured MIC) were predicted at a MIC lower than the measured value. This trend was especially apparent at higher ceftriaxone MIC values, where 68.4% (13/19) of isolates with measured ceftriaxone MICs of 2 μg/mL were predicted at a MIC one doubling dilution lower.

The overall error rates demonstrated by model 7 are displayed in Figure 5.16 and were calculated according to the CLSI ceftriaxone IV meningitis breakpoint criteria (susceptible = ≤0.5 μg/mL, intermediate = 1 μg/mL and resistant = ≥2 μg/mL). Of note, the CLSI ceftriaxone non-meningitis breakpoints could not be applied to this model due to insufficient representation of ceftriaxone-resistant isolates (MICs ≥ 4 μg/mL, n = 3) by this criteria. A total of 25

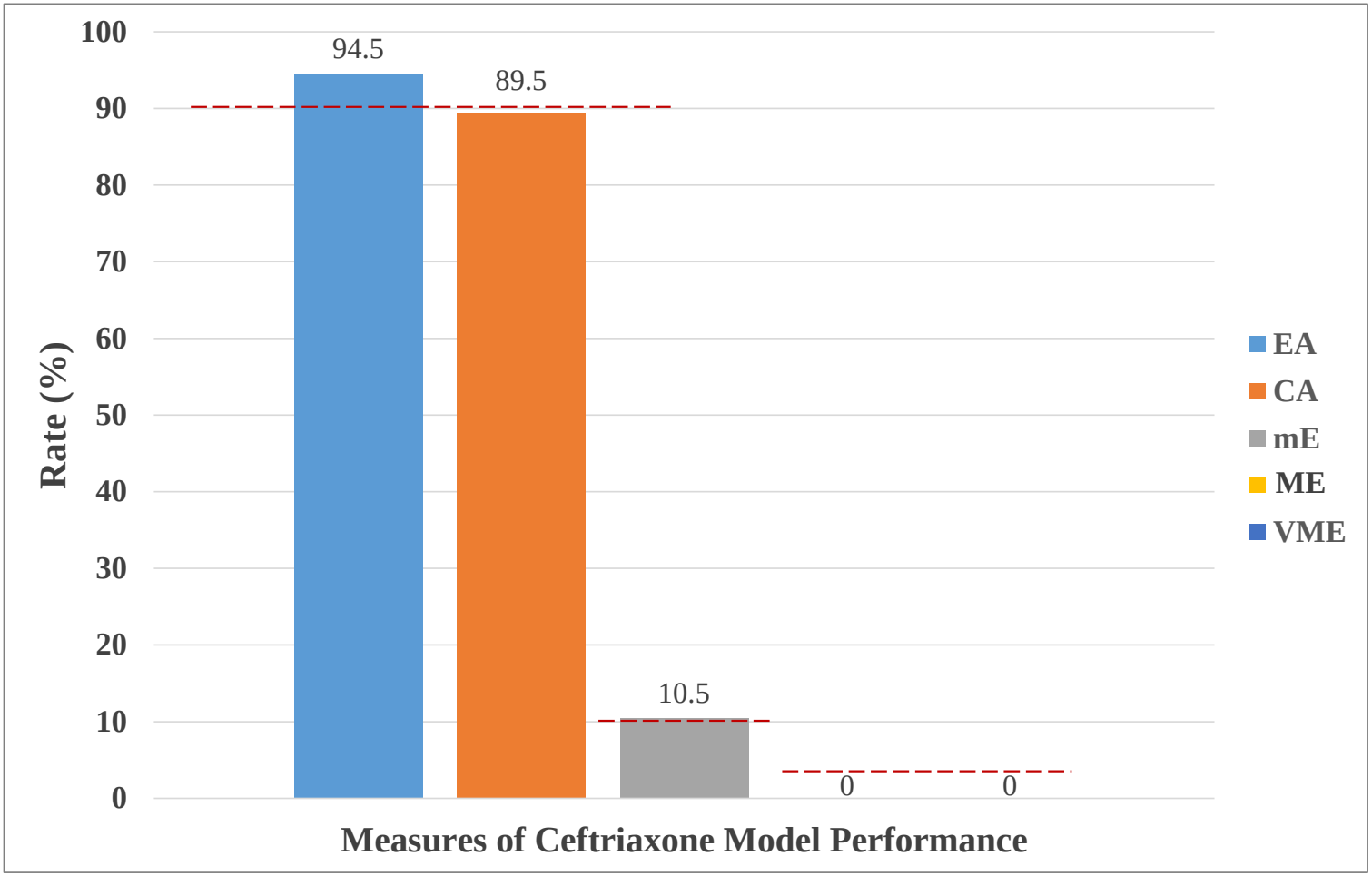
categorical discrepancies between the predicted and measured ceftriaxone MICs were demonstrated by model 7, according to the ceftriaxone IV meningitis breakpoint criteria, resulting in a CA rate of 89.5% (2013/238). Importantly, all categorical discrepancies were recognized as mEs (phenotypically intermediate, WGS-prediction susceptible/resistant and vice versa) since the predicted ceftriaxone MICs differed from measured MICs by only ± 1 doubling dilution. Model 7 demonstrated mE, ME and VME rates of 10.5% (25/238), 0% (0/171) and 0% (0/19), respectively, when the IV meningitis breakpoints were applied. The EA, CA, ME and VME rates were all within the acceptable performance specifications outlined by CLSI, whereas the mE (>10%) rate was slightly above acceptable threshold levels.

TABLE 5.18: Summary of model predicted ceftriaxone MIC values generated by leave-one-out cross-validation

		Predicted Ceftriaxone MIC[a] (µg/mL)								
Measured Ceftriaxone MIC[b] (µg/mL)		**≤0.015**	**0.03**	**0.06**	**0.12**	**0.25**	**0.5**	**1**	**≥2**	**Total Measured**
	≤0.015	**73**	2	2						77
	0.03	3	**15**		2					20
	0.06		11	**17**	7					35
	0.12		1		**4**		3			8
	0.25			1	6	**5**	5	2		19
	0.5				1	1	**2**	8		12
	1			1				**46**	1	48
	≥2							13	**6**	19
	Total Predicted	76	29	20	20	6	11	69	7	238

[a], WGS-based predicted ceftriaxone MIC values generated by leave-one-out cross-validation of model 8. [b], Phenotypic ceftriaxone MIC values determined by broth microdilution.

FIGURE 5.16: Accuracy of model predicted ceftriaxone MICs compared to measured ceftriaxone MICs for *S. pneumoniae*. The essential agreement (EA) and categorical agreement (CA), as well as minor errors (mE), major errors (ME) and very major errors (VME) between predicted and measured ceftriaxone MICs were calculated according to CLSI ceftriaxone IV meningitis breakpoint criteria. The three error types were defined as follows: mE = phenotypically intermediate, WGS-prediction susceptible or resistant (and vice versa); ME = phenotypically susceptible, WGS-prediction resistant; VME = phenotypically, WGS-prediction susceptible. Red dashed lines indicate threshold acceptable error rates: ≥90% for EA and CA, ≤10% for mE and < 3% for ME and VME.



5.6.5 Meropenem MIC Model Development

The most statistically significant combination of PBP-associated independent variables predicting *S. pneumoniae* meropenem MICs was determined using forward multiple linear regression. The starting input variables for this analysis were the same twelve independent variables that were used in the development of the two previously generated β-lactam MIC models (Section 5.6.1 and 5.6.3). By the forward approach, independent variables with *P*-values ≤0.05 were added individually in successive rounds of regression, generating a total of six statistically significant models with *P*-values <0.001. The model summaries of the six linear regression models predicting *S. pneumoniae* meropenem MICs are presented in Table 5.19.

Of the six models produced, model 6 [F (6,233) = 1283.35, P-value <0.001, adjusted R^2 = 0.971] was selected as the best predictive model as it explained the most variance in the dependent variable; this model explained 97.1% of the variance in $\log_2$ (meropenem MIC). The unstandardized beta coefficients of the six independent variables included in the final model, as well as the model constant term, are displayed in Table 5.20. In comparison to the best-determined models for penicillin (Section 5.6.1, adjusted R^2 = 0.964) and ceftriaxone (Section 5.6.3, adjusted R^2 = 0.936), meropenem model 6 required two less independent variables to predict 0.7% and 3.5% more variance in the dependent variable, respectively. In total, six of twelve input variables demonstrated *P*-values >0.05 and were excluded from this model. Interestingly, three of the six excluded variables were derived from PCA of PBP2x (*PBP2x_comp1*, *PBP2x_comp2* and *PBP2x*_S394L), suggesting decreased importance of PBP2x-associated genetic determinants in predicting *S. pneumoniae* meropenem MICs. The final linear regression equation derived from model 6 predicting *S. pneumoniae* meropenem MICs was:

$\log_2$ (meropenem MIC) = -6.89 - 0.81(PBP2x_Q552E) + 21.01(PBP2x_pdis) + 1.81(PBP2b_comp1) + 0.35(PBP2b_comp2) + 17.18(PBP2b_pdis) + 1.36(PBP1a_comp1)

Also displayed in Table 5.20 are the standardized beta coefficients (β-weights) of the six independent variables included in model 6. These values were included to allow for the comparison of weights between independent variables of different measurement scales. Overall, the variable *PBP2b_comp1* demonstrated the highest β-weight with a value of 0.313, indicating that this variable was the most significant contributor in predicting meropenem MICs. Conversely, the variables *PBP2b_comp2* and *PBP2x_Q552E* were the least significant contributors, displaying β-weights of 0.06 and -0.094, respectively. Of note, the *PBP2x_Q552E* variable was the only negative predictor of meropenem MICs out of the six variables included in model 6. Additionally, tolerance and VIF statistics were calculated for each of the independent variables included in model 6 to check for high intercorrelations amongst variables causing multicollinearity. These values are also presented in Table 5.20; tolerance values > 1-R^2 (1-0.971 = 0.029) were considered acceptable for this analysis. Multicollinearity was not an issue for this model since all variables demonstrated tolerance values >0.029.

TABLE 5.19: Forward multiple linear regression analysis summary predicting *S. pneumoniae* meropenem MICs

Model No.	Method[a]	Variables[b]		Model Summary[c]		ANOVA	
		Entered	Excluded	R^2	ΔR^2	F	*P*-value
1	Forward	PBP2b_comp1		0.876	0.875	1630.59	<0.001
2	Forward	+ PBP2x_pdis		0.938	0.938	1752.86	<0.001
3	Forward	+ PBP1a_comp1		0.952	0.952	1527.53	<0.001
4	Forward	+ PBP2b_pdis		0.964	0.964	1511.42	<0.001
5	Forward	+ PBP2x_Q552E		0.970	0.970	1479.27	<0.001
6	Forward	+ PBP2b_comp2	- PBP2x_comp1 - PBP2x_comp2 - PBP2x_S394L - PBP2b_T446A - PBP1a_pdis - PBP1a_comp2	0.971	0.971	1283.35	<0.001

[a], Independent variables with the smallest *P*-values (≤0.05) were added one at a time using the forward multiple linear regression method. [b], Dependent variable: $\log_2$ (meropenem MIC). [c], Coefficient of determination [R^2, adjusted (Δ) R^2] values range from 0 to 1.

TABLE 5.20: Unstandardized and standardized coefficients of the six independent variables included in the final linear regression model predicting *S. pneumoniae* meropenem MICs

Model No.	Model Variables	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		***B*[a]**	***SEB*[b]**	**β**	**t**	***P*-value[c]**	**Tolerance[d]**	**VIF[e]**
6	(constant)	-6.894	0.051		-136.612	<0.001		
	PBP2x_Q552E	-0.812	0.106	-0.094	-7.643	<0.001	0.832	1.202
	PBP2x_pdis	21.017	1.634	0.307	12.863	<0.001	0.222	4.504
	PBP2b_comp1	1.813	0.174	0.313	10.429	<0.001	0.140	7.158
	PBP2b_comp2	0.345	0.109	0.060	3.165	0.002	0.345	2.896
	PBP2b_pdis	17.177	2.398	0.159	7.165	<0.001	0.256	3.911
	PBP1a_comp1	1.362	0.155	0.234	8.805	<0.001	0.179	5.585

[a], Unstandardized beta coefficients, or slope of the associated independent variable (x) in the formula $Y = \beta_0 + \beta_i x_i + e$. [b], Standard error of beta (*B)*. [c], *P*-values were considered significant if ≤0.05. [d], Tolerance = $1 - R^2$; multicollinearity is a problem if tolerance values are $< 1 - R^2$. [e], Variance inflation factor (VIF) = 1/Tolerance.

5.6.6 Meropenem MIC Model Validation

The performance of the best-determined multiple linear regression model (model 6, Section 5.6.5) predicting *S. pneumoniae* meropenem MICs was evaluated using the leave-one-out cross-validation method. By this method, 233 rounds of cross-validation were conducted as previously described (Section 4.6.3), generating model predicted meropenem MIC values for all tested isolates. The WGS-based model predicted meropenem MIC values are presented in Table 5.21 along with the measured MIC values determined by broth microdilution for comparison.

Overall, the percentage of exact matches between predicted and measured meropenem MIC values was 78.5% (183/233). An increased proportion of meropenem MICs were predicted within ±1 doubling dilution of the measured MIC by this model. Of the three β-lactam MIC models that were generated, meropenem model 6 demonstrated the highest EA (98.3%, 229/233) between predicted and measured MICs. Overall, there were no trends demonstrated by this model in regards to direction of predicted MIC values; 50% (25/50) of the model predicted MICs that were non-exact matches (≥ ±1 doubling dilution of the measured MIC) were predicted at a MIC lower than the measured value. Importantly, model 6 demonstrated increased precision at higher meropenem MIC values; 88.9% (64/72) predicted meropenem MICs ranging from 0.5-1 µg/mL were exact matches with measured MIC values.

The overall error rates demonstrated by model 6 are shown in Figure 5.17. The CA, mE, ME and VME rates were calculated as previously described (Section 4.6.3) according to the CLSI meropenem breakpoint criteria (susceptible = ≤0.25 µg/mL, intermediate = 0.5 µg/mL and resistant = ≥1 µg/mL). A total of 11 categorical discrepancies were identified, resulting in a CA rate of 95.3% (222/233) for this model. Similarly to the best-determined model for ceftriaxone (Section 5.6.4), all categorical discrepancies were considered mEs (phenotypically intermediate,

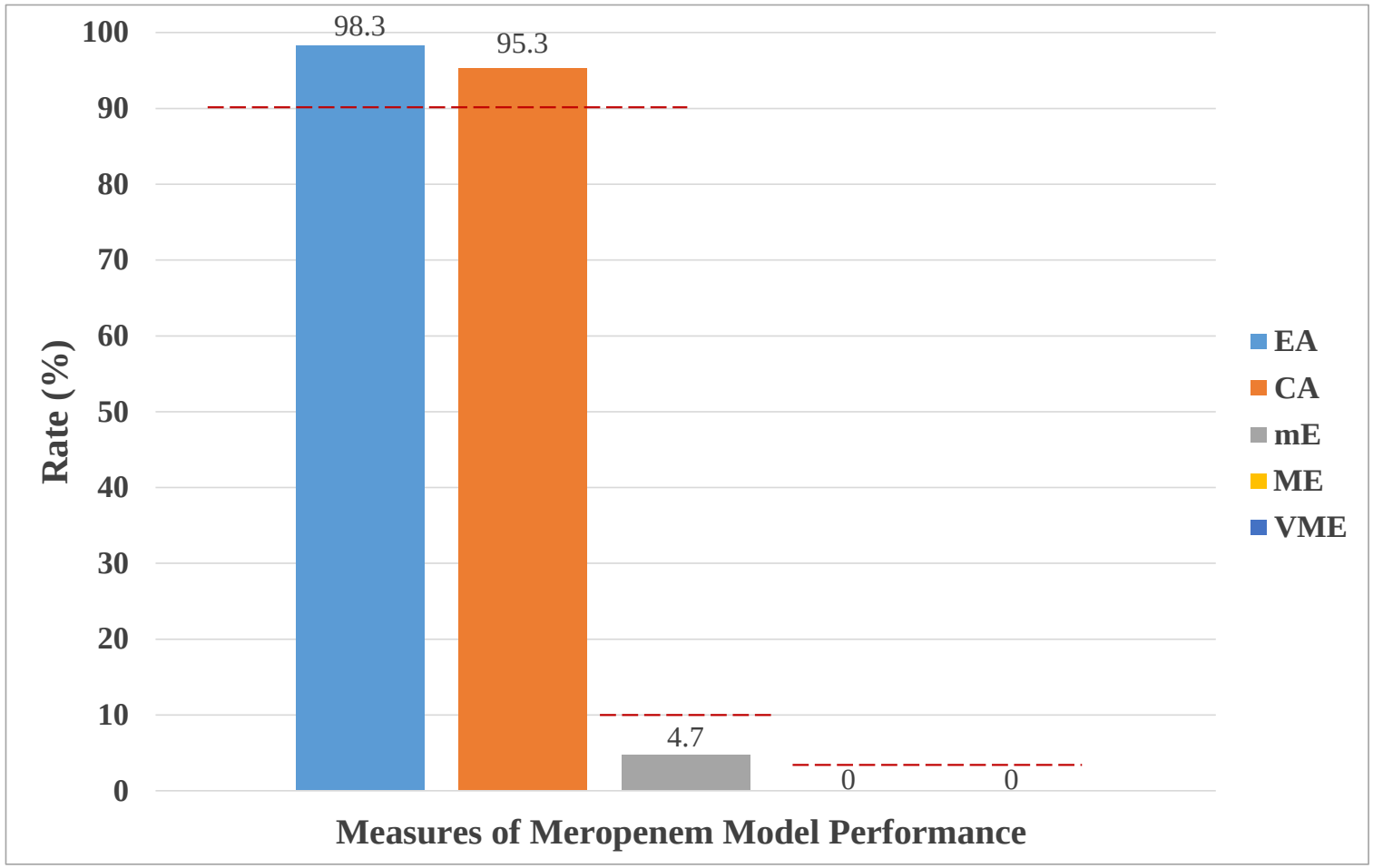
WGS-prediction susceptible/resistant and vice versa). Therefore, the EA rate of this model was increased compared to the CA rate since the difference between the predicted and measured MICs of the nine mEs was only ± 1 doubling dilution. The mE, ME and VME rates demonstrated by model 6 were 4.7% (11/233), 0% (0/161) and 0% (0/39), respectively. Overall, the EA and CA rates (≥ 90%), as well as mE (≤10%), ME and VME (<3%) rates demonstrated by this model were within the acceptable performance specifications outlined by CLSI.

TABLE 5.21: Summary of model predicted meropenem MIC values generated by leave-one-out cross-validation

		Predicted Meropenem MIC[a] (µg/mL)								
		≤0.008	**0.015**	**0.03**	**0.06**	**0.12**	**0.25**	**0.5**	**≥1**	**Total Measured**
Measured Meropenem MIC[b] (µg/mL)	**≤0.008**	**71**	7							78
	0.015	7	**33**		1					41
	0.03		7	**10**	3					20
	0.06		1	1	**0**	3	1			6
	0.12				1	**3**	4			8
	0.25					3	**2**	3		8
	0.5					1		**29**	3	33
	≥1							4	**35**	39
	Total Predicted	78	48	11	5	10	7	37	38	233

[a], WGS-based predicted meropenem MIC values generated by leave-one-out cross-validation of model 8. [b], Phenotypic meropenem MIC values determined by broth microdilution.

FIGURE 5.17: Accuracy of model predicted meropenem MICs compared to measured meropenem MICs for *S. pneumoniae*. The essential agreement (EA) and categorical agreement (CA), as well as minor errors (mE), major errors (ME) and very major errors (VME) between predicted and measured meropenem MICs were calculated according to CLSI meropenem breakpoint criteria. The three error types were defined as follows: mE = phenotypically intermediate, WGS-prediction susceptible or resistant (and vice versa); ME = phenotypically susceptible, WGS-prediction resistant; VME = phenotypically, WGS-prediction susceptible. Red dashed lines indicate threshold acceptable error rates: ≥90% for EA and CA, ≤10% for mE and <3% for ME and VME.



6. DISCUSSION

6.1 Putting it all Together: Characterization of Pneumococcal PBPs

6.1.1 Associating Sequence Variability with Serotype and β-lactam Resistance

Of the six pneumococcal PBPs, PBP2x, PBP2b and PBP1a have long been considered the primary β-lactam resistance determinants in *S. pneumoniae*; sequencing has revealed that they are highly modified in most β-lactam-resistant clinical isolates[24,32]. Their role in the development of β-lactam resistance was initially confirmed by Barcus *et al*., whereby the transformation of β-lactam-susceptible *S. pneumoniae* R6 with mosaic *pbp2x*, *pbp2b*, and *pbp1a* genes cloned from penicillin-resistant strains conferred the same high-level resistance[101]. Numerous combinations of variant PBP sequences have been documented, including in this study, and are associated with a wide range of β-lactam MICs[24]. PBP2x, PBP2b and PBP1a amino acid sequences derived from β-lactam-susceptible, -intermediate and -resistant invasive *S. pneumoniae* isolates were thoroughly characterized in this study prior to the development of our β-lactam MIC models. Sequence characterization was particularly challenging due to the complexity of PBP mutational patterns; therefore, we conducted separate phylogenetic analyses to assist in sequence organization according to PBP type.

An important finding from this work was that most β-lactam-susceptible isolates contained unaltered forms of the three PBPs; PBP2x, PBP2b and PBP1a amino acid sequences were highly conserved in these isolates, comprising the largest clusters in each of the three phylogenies. This observation is consistent with what was reported by Li *et al*., where most penicillin-susceptible isolates included in their study were of a single "PBP type" according to their classification system[51]. In the current study, β-lactam MICs of susceptible isolates

containing three unaltered PBPs were relatively consistent regardless of serotype, thereby enabling the establishment of base-line MIC values representing wild-type susceptibility; these were the lowest predicted values of each β-lactam MIC model developed in Part C of this study (penicillin=≤0.015 µg/mL, ceftriaxone=≤0.015 µg/mL and meropenem=≤0.008 µg/mL).

As described in Part A of the results, there was a clear correlation between the level of PBP sequence variability and β-lactam MICs. PBP2x and PBP2b sequences demonstrating <1% amino acid difference compared to reference strain *S. pneumoniae* R6 predominantly originated from β-lactam-susceptible isolates (penicillin MICs ≤0.06 µg/mL). In contrast, unaltered PBP1a sequences were commonly derived from both β-lactam-susceptible and -nonsusceptible isolates (penicillin MICs ≤0.25 µg/mL). This supports the notion that low-affinity PBP1a variants are involved in higher levels of β-lactam resistance[32,36,93,101]. PBP2x and PBP1a sequences originating from highly β-lactam-resistant isolates were similarly variable, demonstrating >10% amino acid difference compared to the corresponding wild-type sequences from *S. pneumoniae* R6. The maximum sequence variability displayed by PBP2b was lower than that of PBP2x and PBP1a, differing from PBP2b of *S. pneumoniae* R6 by up to 7% in β-lactam-resistant isolates. Previous genetic analysis of *pbp2x, pbp2b* and *pbp1a* sequences from penicillin-susceptible and -resistant *S. pneumoniae* isolates from Canada reported similar ranges of sequence variability[75].

Also noted in the results, the majority of PBP sequence clusters were associated with specific serotypes and levels of β-lactam non-susceptibility in each of the three phylogenies. To our knowledge, few studies exist that assess phylogenetic relationships of a large number of PBP sequences while relating sequence clusters to serotype and β-lactam MIC as this study. Of the three PBPs, cluster analysis of PBP2x produced the largest number of sequence clusters, demonstrating predominant serotype/MIC combinations; similar well-defined clustering by

serotype was also noted in PBP2b sequences. Conversely, cluster analysis of PBP1a generated the fewest number of sequence clusters that were commonly comprised of sequences from multiple serotypes. The differences in clustering patterns suggest that PBP2x and PBP2b demonstrate more diverse mutational patterns than PBP1a, perhaps as a result of an increased number of genetic recombination events involving *pbp2x* and *pbp2b*. This is consistent with a previous recombination study, where Stanhope *et al.* detected two recombination breakpoints in *pbp2x* and *pbp2b* and only a single recombination breakpoint in *pbp1a*[102]. Since the majority of β-lactam-resistant pneumococcal infections are caused by a relatively small number of multidrug-resistant (MDR) lineages[24], serotype-specific PBP sequence clusters are likely related to specific MDR clonal complexes. MLST was not a focus of this study since it has been shown previously that MLST type is a poor predictor of β-lactam MIC[51]; however, it would be interesting to assess the relationship between PBP sequence clusters and MLST types in future studies.

6.1.2 Associating Regions of Variability with Mechanisms of β-lactam Resistance

Pneumococcal PBPs have been the subject of several previously conducted structural and functional studies that have provided valuable insight into the PBP-associated mechanism of β-lactam resistance in *S. pneumoniae*[81-91,93-96]. The crystal structures of PBP2x, PBP2b and PBP1a originating from both wild-type β-lactam-susceptible and -resistant *S. pneumoniae* strains have all been solved to date[82,85,89,95]; the availability of these structures facilitated evolutionary conservation analysis in Part A of this study. Based on this analysis, we determined that PBP structural core was highly conserved, containing distinct regions of variability throughout that were formed by the clustering of multiple highly variable amino acid residues. In all three PBPs, the most prominent region of variability was similarly located in the vicinity of the TP domain

active site; however, other mutational hotspots were noted in other structural domains. This is consistent with the hypothesis by Li *et al.* that amino acid positions are mutated nonrandomly and that certain structural constraints limit PBP alterations to specific regions to retain viability[51].

6.1.2.1 Variability within the TP Domain Active Site

Remodelling of the PBP TP domain active site is the primary mutational mechanism contributing to β-lactam resistance in *S. pneumoniae*. Previous structural studies report that the majority of mutations in low-affinity PBP2x, PBP2b and PBP1a variants are concentrated within this region[82,89,95]. According to these studies, in general, PBP active site remodelling confers resistance in two ways[82,89,95]: (a) mutations within the active site inhibit antibiotic binding and (b) mutations surrounding the entrance impede antibiotic active site access. In this study, distinct regions of variability were observed within the TP domain active site of the three PBPs, indicating that this region is commonly evolving in our cohort of *S. pneumoniae* isolates. By further examining these regions, we found many variable residues buried deep within the active site cleft, as well as surrounding the active site entrance of each PBP; this observation is consistent with previously described patterns of PBP mutations[82,89,95]. Importantly, all mutations that were selected as potential predictors of β-lactam MICs in Part C of this study were mapped within this region of the three PBPs, further supporting their role in pneumococcal β-lactam resistance (in addition to the listed rationale in Appendix A) and their inclusion as variables in our models.

6.1.2.2 Regions of Variability in Other Structural Domains

An interesting finding in this study was that the C-terminal domain of PBP2x was similarly as or increasingly variable compared to the TP domain in β-lactam-nonsusceptible

isolates (Section 5.2.3). The increased variability of the C-terminal domain was further reflected in the conservation analysis of PBP2x, where two distinct hotspots of variability were noted in this region in the 3D structural representation of the protein (Section 5.2.1). Earlier studies also recognized the C-terminal domain of PBP2x as a mutational hotspot in β-lactam-resistant isolates[82,102]. Currently, the function of the PBP2x C-terminal domain is not well understood; however, a recent study by Bernardo-Garcia *et al.* provided evidence that this domain serves as an allosteric binding site for unlinked peptidoglycan[86]. Its potential function as a β-lactam binding domain was earlier described by Gordon *et al.*, who first solved the crystal structure of PBP2x in complex with cefuroxime[85]. The authors observed one antibiotic molecule bound covalently to the TP domain active site; unexpectedly, however, another molecule was found noncovalently associated with the C-terminal domain. More recently, Stanhope *et al.* identified several sites in the C-terminal domain of PBP2x in a large international collection of *S. pneumoniae* isolates that were under significant positive selection pressure and were highly correlated with amoxicillin resistance[102]. This evidence, along with the association between sequence variability and increased β-lactam MICs in this study, strongly suggest that the C-terminal domain of PBP2x is additionally involved in resistance.

As described in the results, the C-terminal region of PBP1a was also shown to be highly variable in select β-lactam-nonsusceptible isolates (Section 5.3.3). Contreras-Martel *et al.* described this region previously as a mutational hotspot in β-lactam-resistant clinical *S. pneumoniae* isolates from France, South Africa and Japan[95]. The C-terminal region of PBP1a is generally considered a unit or tail, as opposed to an actual structural domain, and its function is currently unknown[95]. However, similarly to the C-terminal domain of PBP2x, Stanhope e*t al*. identified several positively selected amino acid sites within this region that were significantly

associated with amoxicillin resistance in their diverse cohort of isolates[102]. The authors hypothesized that the C-terminal unit of PBP1a is involved in pneumococcal β-lactam resistance in an unknown manner similar to that of the C-terminal domain of PBP2x. Most biochemical and functional analyses of pneumococcal PBPs have primarily focused on a small number of resistance-conferring point mutations within the TP domain region (Appendix A, Tables 10.1-10.3); however, this narrowed focus may have resulted in incomplete knowledge of the PBP-associated mechanism of β-lactam resistance in *S. pneumoniae*. The results from this study combined with evidence from previously described studies suggest that other PBP regions outside of the TP domain are rapidly evolving in *S. pneumoniae* and are potentially additional contributors to the development of β-lactam resistance.

6.2 PBP Mutational Patterns and Serotype-specific Mechanisms of β-lactam Resistance

The primary objective of this thesis was to develop WGS-based statistical models predicting *S. pneumoniae* β-lactam MICs. To address this, we identified and analyzed several previously reported PBP-associated β-lactam resistance-conferring point mutations (Section 4.5, Table 4.2), subsequently employing them as predictors of MIC in Part C of this study. However, by analyzing them collectively as 11-, 9- and 11-residue mutational profiles for PBP2x, PBP2b and PBP1a, respectively, we observed patterns of mutations that were uniquely associated with different serotypes and levels of β-lactam resistance. Overall, the stepwise process of the pneumococcal β-lactam resistance pathway has proven difficult to ascertain in the laboratory; however, low-level β-lactam resistance is generally associated with low-affinity PBP2x and PBP2b variants, whereas high-level resistance is said to require altered PBP1a[24,29,32,36-39,93]. It is important to note that most early PBP functional studies employed only a few isolates in their analyses.

6.2.1 PBP2b: Not Just for Low-level β-lactam Resistance?

Consistent with the literature, several of the nine PBP2b mutations selected for analysis in this study were frequently identified in isolates demonstrating lower levels of β-lactam resistance, namely, T446A, D625G, Q628E and T630N (Section 5.2.4). All four substitutions have been detected previously in clinical β-lactam-resistant strains[89]. T446A is well characterized biochemically, and its role in β-lactam resistance has been confirmed in several published studies (described in detail in Appendix A). Conversely, only structural evidence implicates D625G, Q628E and T630N in the resistance process[89]. Similarly to their mapped location in Section 5.2.1 of this study, Contreras-Martel *et al.* demonstrated that this group of substitutions imparts flexibility to an important loop structure at the entrance of the TP domain active site, potentially altering antibiotic access to this region[89]. Interestingly, 3 out of 4 serotypes in Part B of this study were shown to employ this lesser-known mechanism of resistance in combination with select PBP2x substitutions, potentially conferring low-level β-lactam resistance in these isolates (Section 5.4).

Conversely, an interesting finding of this work was that select PBP2b substitutions were commonly detected at higher β-lactam MIC values, indicating a role for PBP2b in high-level β-lactam resistance. Of the nine PBP2b substitutions analyzed, the S412P, N422Y, T426K and Q427L group of substitutions were identified in β-lactam-resistant serotype 15A and 35B isolates, subsequent to maximum alteration of both PBP2x and PBP1a (Section 5.4). Contreras-Martel *et al.* suggested that this group of substitutions modifies the polarity of the TP domain active site entrance[89]; these substitutions were mapped in a similar location in this study (Section 5.2.1). Conversely, β-lactam-resistant serotype 19A isolates demonstrated the opposite stepwise pattern of PBP2b mutations, where the D625G, Q628E and T630N group of substitutions were

detected subsequent to maximum alteration of PBP2x and PBP1a (Section 5.4). Despite the serotype-specific differences, these observations potentially suggest that second step PBP2b mutations are required for high-level β-lactam-resistance in select *S. pneumoniae* strains. This hypothesis is supported by a previous transformation study by Sauerbier *et al.*[103]; the authors demonstrated that the transfer of a portion of the *pbp2b* gene from a highly penicillin-resistant *S. mitis* donor strain was the last essential step to achieve high-level penicillin-resistance in laboratory transformants. Further supporting this notion, Stanhope e*t al.* identified by multiple statistical analyses, a large proportion of variant sites in PBP2b that were significantly associated with high-level amoxicillin resistance in their cohort of *S. pneumoniae* isolates[102].

6.2.2 PBP1a: Not Just for High-level β-lactam Resistance?

Of the three PBPs, PBP1a has long been considered the most clinically important PBP type due to its association with high-level β-lactam resistance in *S. pneumoniae*[29,32]. Interestingly, however, several of the eleven selected PBP1a mutations in this study were frequently identified in isolates demonstrating lower levels of β-lactam resistance, namely TSQF574-577NTGY substitutions; this block of substitutions has been observed in all mosaic PBP1a sequences analyzed to date[32,95,102]. The role of TSQF574-577NTGY substitutions in pneumococcal β-lactam resistance has been confirmed in previous studies; Smith *et al.* demonstrated that substitution reversion by site-directed mutagenesis decreased the additional resistance afforded by PBP1a in *S. pneumoniae* R6 transformants cloned with mosaic *pbp1a* from a β-lactam resistant donor strain[96]. Structurally, Contreras-Martel e*t al.* determined that these substitutions introduce flexibility on a specific loop region that encapsulates the TP domain active site cleft, thereby altering antibiotic active site accessibility[95]. Interestingly, the mutation

of a similar loop structure was noted previously in PBP2b (Section 6.2.1), indicating a shared mutational mechanism between the two PBP types.

In this study, the TSQF574-577NTGY substitutions were the most frequently identified PBP1a mutations and were commonly identified in combination with N546G and N609D substitutions (Section 5.3.4). This combination of substitutions was the mutational profile uniquely displayed by a subset of serotype 19A isolates with equivalent penicillin and ceftriaxone MICs (0.06µg/mL, Sections 5.3.4 & 5.4.1); importantly, these isolates contained unaltered PBP2b sequences. Contradictory to traditional PBP literature, our observations suggest a role for PBP1a rather than PBP2b, in combination with PBP2x, in lower levels of β-lactam resistance in select *S. pneumoniae* serotypes. Based on the previously described differences in β-lactam binding affinities (Section 1.3.3), perhaps this unique PBP1a-associated mechanism evolved in response to selective pressure by cephalosporins and could represent an important first mutational step required for the development of cephalosporin resistance. This notion is supported by the absence of this mutational profile in serotype 19A isolates with non-equivalent penicillin and ceftriaxone MICs (Section 5.4.1). To our knowledge, this study is the first to relate specific first step PBP1a mutations to low-level pneumococcal β-lactam resistance; however, this hypothesis is based mainly on observations of a few isolates and thus requires further investigation.

6.3 Evaluation of WGS-based Statistical Models Predicting *S. pneumoniae* β-lactam MICs

6.3.1 Assessment of β-lactam MIC Model Performance

In this study, we demonstrate the successful application of a "rules-based" model approach in quantitatively predicting β-lactam susceptibility to the level of MICs in *S.*

pneumoniae. Individual statistical models were generated by forward multiple linear regression for three β-lactam antibiotics (penicillin, ceftriaxone and meropenem) commonly used for the treatment of IPD. Each model classifier (final linear regression equation) was trained on a set of PBP-associated genetic variables obtained by WGS that were identified *a priori* as potential predictors of resistance. The current modelling approach ensured that our three best-determined models contained the genetic resistance determinants that were the most statistically informative in predicting β-lactam MIC values.

The goodness-of-fit of each model to the input data was assessed by computing the model adjusted R^2 values; this is an important regression metric that represents the proportion of variance in the dependent variable (β-lactam MIC) that is predicted by the independent variables while adjusting for the number of terms in the model[98]. The adjusted R^2 values of 0.94-0.97 indicate that all three models developed in this study explained most (>90%) of the variation in β-lactam MIC; this is a substantial effect, according to Leech *et al*[98]. Furthermore, the high adjusted R^2 values suggest that the condition of model overfitting was likely not an issue for the present study; model adjusted R^2 values were slightly decreased in comparison to the corresponding model unadjusted R^2 value, indicating that all included variables improved the overall model fit. Factors that may be contributing to the small amount of unexplained variation (<10% for each model) potentially include the exclusion of unidentified genetic resistance determinants, as well as the intrinsic variation associated with phenotypic AST methods. The latter factor was partially corrected for by the performance of replicate MIC testing prior to model development; however, inherent variability can still exist even with the reference AST method (broth microdilution) across multiple replicate MIC measurements[45].

Although the adjusted R^2 value is a common metric used to assess overall model performance, it is limited in that it only measures the model fit to the entire input (training) data set, defined as the training error rate; it is, therefore, not a reliable estimate of the true test error rate of a particular model[104]. In this study, we performed the leave-one-out cross-validation method of resampling in order to directly estimate the test error rates of the three β-lactam MIC models. By this method, predicted β-lactam MIC values were produced for all study isolates by computing the test error (difference in fit) that resulted from the exclusion of each isolate individually (n) from the training data set in successive rounds of cross-validation. Leave-one-out was selected as our method of choice for model validation in large part because our sample size prohibited the use of a designated dataset for validation; therefore, all available samples (n-1) were combined to make up the training dataset to maximize the available power. The major advantage of this method compared to other cross-validation methods that employ the validation set approach is that there is significantly less selection bias since the model is repeatedly fit *n* number of times[99]. Furthermore, without designated training and validation datasets, less randomness is introduced as there is only *n* number of ways the dataset can be split[99].

In addition to evaluating the different model R^2 values, we assessed model performance by comparing model-predicted β-lactam MIC values to measured MIC values determined by broth microdilution. Overall, β-lactam MICs were predicted by all three models to the exact value of the reference AST method for >70% of study isolates and were predicted within one doubling dilution for an even larger proportion (>90%). Furthermore, very few marked differences (>±2 doubling dilutions) between predicted and measured β-lactam MICs were observed (≤1 instance in each model). The closeness of the predicted MICs to the phenotypically measured values indicates that all three models performed reliably and demonstrate a high

degree of predictive accuracy. We evaluated the true predictive accuracies by computing model EA and CA rates (defined in Section 4.6.3); these are two measures of accuracy commonly used for the evaluation of new antimicrobial susceptibility tests[45]. Importantly, all three models developed in this study demonstrated EA rates of ≥95%, which is well above the CLSI acceptable performance threshold of ≥90%[45].

To determine model CA rates, we applied CLSI penicillin (oral penicillin V and IV meningitis breakpoints), ceftriaxone and meropenem breakpoint criteria to the corresponding β-lactam MIC model. Comparing model-predicted β-lactam MIC values to clinical breakpoints yielded CA rates of ≥90% for all three models. Despite being within regulatory targets (≥90%), the CA rates demonstrated by our best-determined penicillin (by oral penicillin V breakpoints) and ceftriaxone models approached unacceptable performance levels. Interestingly, however, when the CLSI IV meningitis breakpoints were applied to our penicillin model, its CA rate was notably increased (>97%), indicating that the measured accuracy of this model was dependent on where the clinical breakpoints fell in the MIC distribution. Conversely, the CA and EA rates demonstrated by our best-determined meropenem MIC model were similarly increased (≥95%); this suggests better overall predictive capabilities since its accuracy was less affected by the application of breakpoints compared to the other two β-lactam models. Comparison of model adjusted R^2 values further supports this notion; of the three models, our meropenem MIC model demonstrated the highest adjusted R^2 value and was thereby the model that explained the most variation in the dependent variable ($\log_2$ β-lactam MIC).

Application of clinical breakpoints also allowed us to determine error rates for three types of error (mE, ME and VME errors, described in Section 4.6.3), where discrepancies between model-predicted and measured MIC values resulted in a change in categorical interpretation.

The acceptable rate of error threshold established by CLSI for the validation of new antimicrobial susceptibility tests differs according to error type due to the potential clinical impact (≤10% for mE and <3% for ME and VME)[45]. Importantly, we observed that the majority of errors in model-predicted β-lactam MICs resulted from discrepancies near clinical breakpoints that were within one doubling dilution of measured MIC values. Since these discrepancies were by definition in essential agreement, model EA rates were notably increased compared to the corresponding CA rate of each model. However, the total number and types of errors resulting from these "minor" discrepancies differed for each model based on the clinical breakpoints that were applied and the resultant categorical changes. Overall, the error rates demonstrated by all three models were within acceptable performance specifications outlined by CLSI, with the exception of VME rates (>3%) for penicillin by the IV meningitis breakpoints and mE (>10%) rates for ceftriaxone.

The unacceptably high VME rate demonstrated by our penicillin MIC model resulted from a small number of MIC discrepancies below the single IV meningitis breakpoint. Importantly, all VMEs arose from discrepancies in MIC predictions that were within one doubling dilution of the measured MIC value. Despite the essential agreement, discrepancies near the breakpoint resulted in changes to predicted categorical interpretation from resistant to susceptible in the absence of an intermediate zone by this criteria. Another factor potentially contributing to the increased VME rate of this model was the relatively small sample size of this study. The proportion of penicillin MIC discrepancies that were classified as VMEs was relative to the total number of measured penicillin-resistant isolates; therefore, the small denominator value perhaps falsely increased the overall proportion of these errors. Furthermore, a larger sample size would increase the number of times the model classifier was trained on specific

genetic determinants associated with lower levels of penicillin resistance, thereby increasing the model predictive power near the breakpoint. However, the erroneous categorization of any true-resistant isolate as susceptible (false susceptibility) by this model is of great concern clinically due to the increased likelihood of treatment failure in IPD patients by guiding the administration of ineffective antimicrobial therapies.

Conversely, the increased mE rate demonstrated by our ceftriaxone model is of lesser concern as this type of error has limited therapeutic consequences[45]. Similarly to our penicillin model, all errors arose from discrepancies in ceftriaxone MIC prediction that were in essential agreement with measured values; however, by comparison, discrepancies resulted in only minor changes in categorical assignment due to the presence of an intermediate interpretive category by this criteria. Most mE errors resulted from discrepancies in MIC prediction directly below the upper ceftriaxone IV meningitis MIC breakpoint, thereby changing the categorical interpretation of these isolates from resistant to intermediate. The remaining proportion of mEs arose from single doubling dilution MIC discrepancies directly above the lower ceftriaxone IV meningitis MIC breakpoint (predicted intermediate but were true-susceptible isolates). The increased number of discrepancies around these breakpoints suggests that the individual variable or combination of variables with the greatest impact on ceftriaxone resistance were either not identified or weighted appropriately by this model.

Although mEs are of lesser clinical importance in comparison to the other error types, it is valuable to monitor the direction of mEs generated by the AST method under evaluation[45]. Any tendencies of a new AST method to produce results that read consistently higher or lower than the reference method indicates the potential for error rates to increase above acceptable threshold levels in future assessments of method performance. Interestingly, each of our three

models demonstrated differing predictive tendencies. In general, our best-determined model for penicillin tended to over-predict the level of resistance; the majority of MICs that were non-exact matches were consistently predicted as a higher value than phenotypically measured MICs. This trend suggests that some of the variables included in this model were weighted too highly as predictors of penicillin resistance, or that the additive effect of the multiple resistance determinants was statistically greater than their combined mechanistic effect in some isolates. In comparison, our best-determined model for ceftriaxone displayed the tendency to under-predict the level of resistance, particularly in the 0.06-0.25 μg/mL range and at MICs ≥2 μg/mL. This scenario is less desirable than the previously described alternative due to an increased potential for VMEs to occur and may have resulted from the exclusion of unidentified genetic determinants that are important predictors of ceftriaxone resistance.

Of the three β-lactam models that were produced, our best-determined model for meropenem did not display predictive tendencies in any one direction. The lack of any observable trends suggests that the total number variables and their assigned predictive weights were determined appropriately by this model, resulting in the reliable prediction of *S. pneumoniae* meropenem MICs. Overall, this model was considered our best performing model since it demonstrated the highest predictive accuracy (highest adjusted R^2 value with the highest EA and CA rates), coupled with the lowest rates of error. The better overall predictive performance and the requirement of a fewer number of variables indicate that meropenem resistance in *S. pneumoniae* is statistically less complex than penicillin and ceftriaxone resistance.

An advantage of multiple linear regression over other WGS-based AST approaches, specifically, rules-based approaches are that multiple mechanisms of resistance can be accounted

for concurrently in the prediction of β-lactam MICs. This is an important consideration for the present study as multiple PBP-associated genetic determinants have been identified that are associated with varying levels of β-lactam nonsusceptibility in clinical *S. pneumoniae* isolates. Interestingly, the total number of included variables and the overall proportion of variables associated with the three PBP types differed for each model. Half of the variables included in our best-determined model for penicillin were PBP2x-associated genetic determinants, indicating that PBP2x was the most statistically informative PBP type in predicting *S. pneumoniae* penicillin MICs. Conversely, half of the variables included in our best-determined model for meropenem were genetic determinants derived from this PBP2b. Furthermore, all three PBP types were equally as statistically important in the prediction of *S. pneumoniae* ceftriaxone MICs; this model included an approximately equal number of variables derived from each PBP. Of the twelve variables used as starting input in our regression analysis, four variables representing each of the three PBPs (*PBP2x_pdis*, *PBP2x_Q552E*, *PBP2b_pdis* and *PBP1a_comp1*) were statistically significant predictors of MIC in all three models.

The different model variable components highlight an important limitation of the modelling approach employed in this study; by the forward regression method, only PBP-associated genetic determinants that were statistically significant predictors of β-lactam MICs were included in each model, without accounting for differences in their mechanistic importance. Our final model for ceftriaxone best demonstrates this limitation; if model variables were selected while taking into account previously published observations relating differences β-lactam binding affinities, we would expect an increased proportion of PBP2x-associated variables in this model and potentially the absence of any PBP2b-associated variables since 3rd generation cephalosporins preferentially bind PBP2x over PBP2b[32,36-39]. However, two

statistically significant PBP2b-associated variables were included in this model, despite their potentially limited mechanistic importance in the development of ceftriaxone resistance in *S. pneumoniae*.

6.3.2 Assessment of Model Performance Compared to Other WGS-based AST Models Predicting β-lactam MICs

6.3.2.1 A Different WGS-based AST Model Applied to *S. pneumoniae*

Over the course of this study, other WGS-based AST models have been developed for the purposes of predicting β-lactam resistance in *S. pneumoniae*[50,51,105,106]. The most notable example of successful implementation is the WGS-based PBP typing system developed by the Centers for Disease Control and Prevention - Active Bacterial Core surveillance (ABCs) team. Briefly, by their methodology, *S. pneumoniae* isolates are assigned a three-number identifier ("PBP type") representing their composite PBP2x, PBP2b and PBP1a TP domain amino acid sequence signatures[107]. In one of the original published studies employing this system, Li *et al*. developed three different WGS-based AST models predicting β-lactam MICs from PBP type for over 2,000 *S. pneumoniae* isolates divided into separate training and testing datasets[51]. Their work demonstrated that MICs of six β-lactams could be accurately predicted (EA and CA rates >90%) for isolates of trained PBP types using both rules-based and ML models. However, all three models were notably less accurate in their prediction of MICs for isolates of nontrained PBP types; critically, ME and VME rates were well outside of regulatory targets for some β-lactams. The small sample size of isolates of nontrained PBP types was a noted limitation of their study. The authors then performed a follow-up study for the purposes of validating their model-based β-lactam MIC predictions using a larger cohort of isolates of nontrained PBP types[105]. Importantly, their ML model-predicted MICs satisfied criteria of acceptable

performance for all six β-lactams, including the three β-lactams of interest in the current study. In comparison, their rules-based model demonstrated a generally lower percentage of EA and CA with unacceptably high VME rates for four of six β-lactams, including ceftriaxone.

Overall, the β-lactam MIC models developed in the current study performed similarly as well to the models developed by Li e*t al.* for the prediction of β-lactam MICs for isolates of trained PBP types. We demonstrated that our best-determined models for penicillin, ceftriaxone and meropenem performed with high levels of predictive accuracy with combined EA and CA rates of ≥90%. These results are in accordance with the work by Li e*t al.*, whereby both their rules-based and machine learning models demonstrated EA and CA rates well above acceptable performance specifications (>90% for isolates of trained PBP types)[51]. Additionally, both the current modelling framework and the previously described PBP typing system produced similarly low rates of error in MIC predictions of ceftriaxone and meropenem, yielding ME and VME rates of <3%. Interestingly, similarly to our penicillin MIC model (VME rate of >3%), all three predictive models developed by Li *et al*. demonstrated increased VME rates for penicillin by the CLSI IV meningitis breakpoint criteria. Increased VME rates for the same antibiotic suggests the involvement of other non-PBP-associated genetic determinants in reducing penicillin susceptibility in select strains of *S. pneumoniae* that were unaccounted for in both studies.

The differences in predictive performance for isolates of trained and nontrained PBP types demonstrated in the study by Li *et al.* highlights an important strength of the modelling approach used in this study. As previously described, we performed model validation using the leave-one-out method; this method of cross-validation is advantageous compared to other validation set approaches as it reduces selection bias by repeatedly fitting the model *n* number of

times[99]. By employing this method, we increased the number of times that specific genetic predictors of β-lactam MIC were observed, thereby decreasing the likelihood that test isolates were misclassified. In comparison, the validation set approach used by Li *et al.* applied to their rules-based model caused the misclassification of several test isolates that were of newly encountered (nontrained) PBP types, resulting in unacceptably high VME rates for four of six β-lactams[51]. The poor predictive capability of their rules-based model reflects its inability to detect signatures of resistance in PBP types that demonstrate significant sequence divergence from PBP types within the training set.

6.3.2.2 A Similar WGS-based AST Model Applied to *N. gonorrhoeae*

To further evaluate model performance, we compared the results of this study to WGS-based AST models developed by Eyre *et al.* that employ a similar modelling approach for the prediction of *N. gonorrhoeae* β-lactam MICs[108]. For their study, the authors identified several *a priori* genetic determinants conferring reduced β-lactam susceptibility in *N. gonorrhoeae* and applied them as potential predictor variables in multivariate linear regression analysis, in much the same way as this study. Two β-lactam MIC models (penicillin and cefixime) were developed for >600 isolates using the backward regression method and were validated by leave-one-out cross-validation, generating model-predicted MIC values.

Overall, the models of the current study and the models developed by Eyre e*t al*. performed similarly in predicting β-lactam MICs. The model adjusted R^2 values presented in the study by Eyre *et al.* were slightly lower than the adjusted R^2 values of the current study, accounting for approximately 10% less variation in β-lactam MIC. However, when clinical breakpoints were applied to MIC predictions, the EA rates of both studies were similarly increased and were above the acceptable performance threshold. The only notable differences in performance between the

two studies were the model rates of error; Eyre *et al*. cefixime MIC model demonstrated an unacceptably high VME rate, whereas no errors of this type were produced by our ceftriaxone MIC model. Conversely, our best-determined penicillin MIC model demonstrated an unacceptably high VME rate (IV meningitis breakpoints only), whereas no VMEs were noted for penicillin by Eyre *et al*. Despite these differences, both the present study and the work by Eyre *et al.* demonstrate how similar rules-based models can be applied to different bacterial species for the reliable prediction of β-lactam MICs from WGS data.

6.3.3 Clinical Applications and Future Model Improvements

Overall, the results of this study support the rejection of our hypothesis that WGS cannot be used exclusively as a tool to infer antimicrobial susceptibilies in the determination of *S. pneumoniae* β-lactam MIC values. Despite the high predictive accuracies, several previously described technical limitations associated with WGS-based AST methods (Section 1.4.2) currently limit the utility of our models in the clinical setting. One of the major factors limiting the routine use of WGS-based AST in clinical laboratories is the increased time to results[48,49]. According to the current standard of care paradigm for phenotypic AST, for nonfastidious organisms such as *S. pneumoniae*, the average time to susceptibility results from specimen plating is approximately 48 hours[46]; therefore, delaying the administration of targeted antimicrobial therapy for IPD patients. In comparison, the current timeframe for obtaining WGS-based AST results for most bacterial pathogens is similarly or increasingly delayed[46,49]; this is in large part due to the requirement of a cultured organism for sequencing, which is an assumed limitation in most WGS-based AST studies, including the present study.

Assuming an initial culture step and the use of 2nd generation sequencing technologies, the models generated in this study could potentially be applied to clinical IPD isolates for the

prediction of β-lactam MICs, providing results within 48 hours[109]. This proposed timeline is comparable to the time to results of phenotypic AST; therefore, the current modelling approach would not provide additional benefit to IPD patient care over standard AST methods. This also assumes unbatched sequencing runs, which is impractical given the current laboratory workflow and the total number of positive cultures per day. In order to justify the added costs of sequencing in the clinical setting, our WGS-based AST models must offer the potential for improved IPD patient treatment outcomes[49]. Two ways in which to improve the clinical utility of our models in future applications include (a) sequencing of cultured IPD isolates using 3rd generation sequencing technologies and (b) sequencing of blood culture specimens directly using a metagenomics approach[46,48,109]. These methods would potentially enable earlier model application in downstream analysis and increase the speed at which predicted AST results are generated compared to phenotypic methods. However, the use of both metagenomics and long-read sequencing for WGS-based AST is relatively early in development for the majority of bacterial species[46,48]. Therefore, additional studies are required to evaluate the ability of these methods used in combination with the models developed in this study to predict *S. pneumoniae* β-lactam MICs and to assess the overall applicability to IPD patient care.

In addition to the suggested technical improvements described above employing alternative sequencing technologies and methods to increase the speed of resistance detection, other modifications could be made that would potentially improve predictive model performance. Due to the relatively small sample size of this study, the lack of available power limited us to include only relatively common PBP-associated β-lactam resistance determinants in the formulation of our models. Overall, we were able to limit model overfitting by restricting the initial number of potential predictor variables, performing variable reduction using PCA and

selecting variables in our final models that were statistically significant. However, it is possible that variables with fewer numbers of "positive" variant observations were excluded from our final models due to a lack of statistical significance rather than a lack of mechanistic importance. It is important to note that, in some cases, variables that are not statistically significant can make a better overall model when included[104]. An alternative approach would have been to use the simultaneous method to compute multiple regression, whereby all variables derived from PCA would have been retained in the final models. However, this method would increase the likelihood of model over-fitting and increase the risk of multicollinearity; therefore, a larger sample size would be required, along with manual corrections for multicollinearity in downstream analysis.

In relation to the previously described limitation, multiple unknown resistance determinants were also likely excluded from our final models. The contributions of unknown, non-PBP-associated β-lactam resistance determinants was suggested in early transformation studies by Chesnel *et al.*, where transformation of *S. pneumoniae* R6 with variant PBP genes from high-level β-lactam-resistant strains conferred resistance[110]; however, the resultant level was reduced compared to the PBP donor. Furthermore, the authors showed that isolates containing identical PBP sequences exhibited wide-ranging β-lactam MICs. In this study, the decreased accuracy of MIC predictions demonstrated by our final penicillin model at lower MIC values, as well as our final ceftriaxone model at higher MIC values, perhaps suggests that non-PBP-associated determinants are contributing to low- and high-level penicillin and ceftriaxone-resistance, respectively.

To date, several non-PBP proteins have been implicated in the development of β-lactam resistance in select clinical *S. pneumoniae* strains including other enzymes involved in

peptidoglycan biosynthesis (MraY, MraW, MurM and MurN), as well as enzymes with different cellular functions (StkP, CiaH, CpoA and ClpL)[24,29]. Recently, a genome-wide association study of *S. pneumoniae* identified specific polymorphisms in other non-PBP genes showing significant associations with β-lactam resistance[111]. The impact of these novel resistance determinants on *S. pneumoniae* β-lactam MICs has not been well-established and requires further investigation. Nonetheless, the inclusion of non-PBP-associated variables in future model applications could potentially improve predictive performance, particularly for the prediction of penicillin and ceftriaxone MICs.

6.4 Study Limitations

The relatively small sample size was the primary limitation of this study. Ideally, a larger number of β-lactam-nonsusceptible isolates would have been included in WGS analysis; however, this was not possible due to time and cost constraints. This factor limited us to include only four different serotypes in the analysis of serotype-specific PBP mutation profiles in Part B of this study. In general, the inclusion of more isolates would enable better characterization of the three PBP proteins responsible for β-lactam resistance in *S. pneumoniae*. Furthermore, in Part C of this study, the limited sample size meant that all available isolates had to be included as a single large training dataset to maximize the available statistical power; therefore, we were unable to designate an independent dataset for the purposes of model validation. Despite the combining of samples and the performance of leave-one-out cross-validation, the limited sample size meant that, overall, our study lacked sufficient statistical power for the number of variables that were included in each of our final β-lactam MIC models. Inclusion of more isolates for model development, particularly more β-lactam-nonsusceptible isolates, would increase

statistical power, increase the number of times less common PBP variants were observed, as well as would enable us to include more predictor variables without risking model over-fitting.

Since we drew our isolate collection from the larger SAVE study, the inherent limitations associated with SAVE are also relevant to this study. Most importantly, this study was likely impacted by regional bias due to the country-specific sampling frame. This potentially resulted in the overrepresentation of PBP-associated mechanisms of β-lactam resistance common to Canadian IPD isolates, thereby increasing their statistical significance in the development of our predictive models. Therefore, a larger and more diverse dataset containing IPD isolates from multiple countries is required to determine the association between country of isolation and the different β-lactam resistance determinants analyzed in this study.

7. CONCLUSIONS

This thesis presents WGS data on three pneumococcal PBPs (PBP2x, PBP2b and PBP1a) commonly implicated in β-lactam resistance. From this data, we developed separate multiple linear regression models to predict the MICs of three β-lactam antibiotics commonly used for the empiric and directed treatment of IPD. The modelling approach taken in this study focused on a select number of conserved resistance-conferring substitutions, thereby preventing the inclusion of mechanistically unimportant hitchhiking mutations that are present in high frequencies in mosaic PBP sequences. Our best-determined models were validated using the leave-one-out cross-validation method, generating WGS-based predicted β-lactam MICs. By comparing model-predicted and phenotypically measured MIC values, we tested our hypothesis that WGS cannot be used exclusively for the determination of *S. pneumoniae* β-lactam MIC values. The MICs are critical for patient care as they are routinely used to guide clinical decision making for the treatment of IPD.

The work presented in this thesis demonstrates the capability of our modelling approach to accurately predict AST results to the MIC level for Canadian IPD isolates to three different β-lactam antibiotics using WGS data. All three models developed in this study predicted MICs within ± one doubling dilution of measured MIC values for >90% of study isolates, surpassing the minimum threshold of acceptable performance set by CLSI for the validation of new antimicrobial susceptibility tests. Furthermore, our models demonstrated a maximum of one instance per model where MIC discrepancies were greater than two doubling dilutions between the two AST methods. We observed notable differences in predictive accuracies when clinical breakpoints were applied in the assessment of model performance; however, all three models demonstrated acceptable rates of CA (≥90%). Importantly, the majority of discrepancies

resulting in unacceptable rates of error (VME rate for penicillin by IV meningitis breakpoints and mE rate for ceftriaxone) were due to MIC discrepancies that were in essential agreement (within ± one doubling dilution). Therefore, model error rates were not reflective of true predictive inaccuracies, but rather reflected where the clinical breakpoints fell in the model MIC distributions. Overall, the results of this thesis support both the rejection of our hypothesis and the future use of WGS as an alternative AST tool to infer *S. pneumoniae* β-lactam antimicrobial susceptibilities.

This study illustrated the complexity of the PBP-associated mechanism of β-lactam resistance in *S. pneumoniae*. Overall, the models that were developed in this study successfully minimized this, thereby enabling the accurate and reliable prediction of β-lactam MICs. Although the results indicate that our predictive models could conceptually replace traditional phenotypic AST methods following a subsequent validation step with a large designated testing dataset, several factors related to WGS currently limit model utility in the clinical setting. Most notably is the requirement of an initial culture step, yielding turnaround times similar to phenotypic AST; therefore, model implementation would provide limited benefit to IPD patient care over standard methods. Furthermore, most primary diagnostic microbiology laboratories lack the equipment and bioinformatics knowledge for WGS, as well as, optimal WGS-based AST timelines require samples to be tested as they are positive or as received, which is impractical in the clinical setting. Despite these limitations, the capability of WGS to determine other important pneumococcal strain features warrants further investigation into its use as a diagnostic tool to infer antimicrobial susceptibilities. As technical advances continue, earlier application of our models following WGS directly from clinical samples using 3rd generation sequencing technologies offers the potential to reduce turnaround times. In that setting, the rapid

detection of resistance using our WGS-based AST methodology for cases of IPD caused by β-lactam-resistant strains would reduce the time to effective therapy, resulting in improved patient treatment outcomes.

8. FUTURE DIRECTIONS

(A) Evaluation of model performance using an independent validation dataset

As discussed previously, the primary limitation of this study was the relatively small sample size, therefore requiring us to combine all available isolates into a single training dataset to maximize the available power for model development. Although we employed leave-one-out cross-validation, which is advantageous over other validation methods for the previously described reasons (Section 6.3.1), the statistical conclusions made from this study were underpowered for the number of variables included in our models. Therefore, the performance of a second validation step using a large independent testing dataset is imperative for clinical implementation. Importantly, the supplementary dataset should be comprised of a larger number of β-lactam-nonsusceptible isolates from diverse geographic regions. The inclusion of more isolates, as well as employing two different validation methods for comparison, would increase the statistical power of this study beyond proof-of-concept principles.

(B) Development of a universal β-lactam MIC model

Individual predictive models were developed for the three β-lactams of interest in this study; however, the development of a single model that could infer *S. pneumoniae* antimicrobial susceptibilities for multiple β-lactam antibiotics would be useful in future clinical applications. As described previously, a current limitation of WGS-based AST methods is the overall complexity of downstream analyses, requiring individuals with specialized training to interpret the results[48,49]. The development of a universal β-lactam MIC model that generates easily interpretable results would reduce the overall computing time, as well as would potentially reduce the time that the clinical microbiology laboratory reports results.

(C) Biochemical characterization of conserved substitutions in the C-terminal regions of PBP2x and PBP1a

The results of this study suggest the additional involvement of regions outside of the PBP TP domain in the development of pneumococcal β-lactam resistance. As discussed previously, the TP domain of the three PBPs has primarily been the focus of most previous studies, potentially resulting in gaps in knowledge of the PBP-associated mechanism of resistance. In this study, the C-terminal domains of PBP2x and PBP1a were shown to be similarly as or increasingly variable compared to the TP domain in select β-lactam-nonsusceptible *S. pneumoniae* strains. These observations, combined with the finding that the C-terminal domain of PBP2x serves as an allosteric binding site for both peptidoglycan and β-lactams[85,86], makes this region of PBP2x and PBP1a an interesting alternative site to focus on in future biochemical studies. A potential approach for future studies would be to perform site-directed mutagenesis on a select number of conserved mutations within this region and subsequently determine the β-lactam MICs of the resulting transformants.

9. REFERENCES

1. Brandt, C. & Spellerberg, B. Streptococcus in. *Manual of Clinical Microbiology, 11th Edition* 383–402 (American Society of Microbiology, 2015).

2. Facklam, R. What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clin. Microbiol. Rev.* **15,** 613–30 (2002).

3. Geno, K. A. *et al.* Pneumococcal Capsules and Their Types: Past, Present, and Future. *Clin. Microbiol. Rev.* **28,** 871–99 (2015).

4. Bogaert, D., de Groot, R. & Hermans, P. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect. Dis.* **4,** 144–154 (2004).

5. Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat. Rev. Microbiol.* **6,** 288–301 (2008).

6. Weiser, J. N., Ferreira, D. M. & Paton, J. C. *Streptococcus pneumoniae*: transmission, colonization and invasion. *Nat. Rev. Microbiol.* **16,** 355–367 (2018).

7. Donkor, E. S. Understanding the pneumococcus: Transmission and evolution. *Front. Cell. Infect.* **4,** 7 (2013).

8. Simell, B. *et al.* The fundamental link between pneumococcal carriage and disease. *Expert Rev. Vaccines* **11,** 841–855 (2012).

9. Bridy-Pappas, A. E., Margolis, M. B., Center, K. J. & Isaacman, D. J. *Streptococcus pneumoniae* : Description of the Pathogen, Disease Epidemiology, Treatment, and Prevention. *Pharmacotherapy* **25,** 1193–1212 (2005).

10. Henriques-Normark, B., Blomberg, C., Dagerhamn, J., Bättig, P. & Normark, S. The rise and fall of bacterial clones: *Streptococcus pneumoniae*. *Nat. Rev. Microbiol.* **6,** 827–837 (2008).

11. Tettelin, H. *et al.* Genomics, Genetic Variation and Regions of Differences in. *Streptococcus Pneumoniae* 81–107 (Academic Press, 2015).

12. Donati, C. *et al.* Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol.* **11,** (2010).

13. Claverys, J. P., Martin, B. & Håvarstein, L. S. Competence-induced fratricide in streptococci. *Molecular Microbiology* **64,** 1423–1433 (2007).

14. Hoskins, J. *et al.* Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J.*

Bacteriol. **183,** 5709–17 (2001).

15. van der Poll, T. & Opal, S. M. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet* **374,** 1543–1556 (2009).

16. Lanie, J. A. *et al.* Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. *J. Bacteriol.* **189,** 38–51 (2007).

17. Lynch, J. P. & Zhanel, G. G. *Streptococcus pneumoniae*: Epidemiology, risk factors, and strategies for prevention. *Seminars in Respiratory and Critical Care Medicine* **30,** 189–209 (2009).

18. Bergenfelz, C. & Hakansson, A. P. *Streptococcus pneumoniae* Otitis Media Pathogenesis and How It Informs Our Understanding of Vaccine Strategies. *Curr. Otorhinolaryngol. Rep.* **5,** 115–124 (2017).

19. CDC: Pneumococcal Disease-Clinical Features. Available at: https://www.cdc.gov/pneumococcal/clinicians/clinical-features.html.

20. WHO. *Pneumococcus: Vaccine-Preventable Diseases Surveillance Standards*. (WHO, 2018).

21. ECDC: Disease Factsheet about Pneumococcal Disease. Available at: https://www.ecdc.europa.eu/en/pneumococcal-disease/facts.

22. Invasive Pneumococcal Disease - Government of Canada. Available at: https://www.canada.ca/en/public-health/services/immunization/vaccine-preventable-diseases/invasive-pneumococcal-disease/health-professionals.html.

23. Fitzgerald, D. & Waterer, G. W. Invasive Pneumococcal and Meningococcal Disease. *Infect. Dis. Clin. North Am.* **33,** 1125–1141 (2019).

24. Kim, L., McGee, L., Tomczyk, S. & Beall, B. Biological and Epidemiological Features of Antibiotic-Resistant *Streptococcus pneumoniae* in Pre- and Post-Conjugate Vaccine Eras: a United States Perspective. *Clin. Microbiol. Rev.* **29,** 525–52 (2016).

25. Tunkel, A. R. *et al.* Infectious Diseases Society of America's Clinical Practice Guidelines for Healthcare-Associated Ventriculitis and Meningitis. *Clin. Infect. Dis.* **64,** e34–e65 (2017).

26. Page, M. G. P. Beta-lactam Antibiotics in. *Antibiotic Discovery and Development* 79–117 (Springer US, 2012).

27. Cojocel, C. Beta-lactam Antibiotics in. *Clinical Nephrotoxins* 293–321 (Springer US, 2008).

28. Kuti, J. L. Optimizing Antimicrobial Pharmacodynamics: A Guide For Your Stewardship Program. *Rev. Médica Clínica Las Condes* **27,** 615–624 (2016).

29. Hakenbeck, R., Brückner, R., Denapaite, D. & Maurer, P. Molecular mechanisms of β-lactam resistance in *Streptococcus pneumoniae*. *Future Microbiology* **7,** 395–410 (2012).

30. Lovering, A. L., Safadi, S. S. & Strynadka, N. C. J. Structural Perspective of Peptidoglycan Biosynthesis and Assembly. *Annu. Rev. Biochem.* **81,** 451–478 (2012).

31. Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A. & Charlier, P. The penicillin-binding proteins: Structure and role in peptidoglycan biosynthesis. *FEMS Microbiology Reviews* **32,** 234–258 (2008).

32. Zapun, A., Contreras-Martel, C. & Vernet, T. Penicillin-binding proteins and β-lactam resistance. *FEMS Microbiology Reviews* **32,** 361–385 (2008).

33. Macheboeuf, P., Contreras-Martel, C., Job, V., Dideberg, O. & Dessen, A. Penicillin binding proteins: Key players in bacterial cell cycle and drug resistance processes. *FEMS Microbiology Reviews* **30,** 673–691 (2006).

34. Kell, C. M. *et al.* Deletion analysis of the essentiality of penicillin-binding proteins 1A, 2B and 2X of *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.* **106,** 171–175 (1993).

35. Hoskins, J. *et al.* Gene disruption studies of penicillin-binding proteins 1a, 1b, and 2a in *Streptococcus pneumoniae*. *J. Bacteriol.* **181,** 6552–5 (1999).

36. Coffey, T. J. *et al.* Genetic analysis of clinical isolates of *Streptococcus pneumoniae* with high-level resistance to expanded-spectrum cephalosporins. *Antimicrob. Agents Chemother.* **39,** 1306–13 (1995).

37. Krauss, J., van der Linden, M., Grebe, T. & Hakenbeck, R. Penicillin-binding proteins 2x and 2b as primary PBP targets in *Streptococcus pneumoniae*. *Microb. Drug Resist.* **2,** 183–6 (1996).

38. Grebe, T. & Hakenbeck, R. Penicillin-binding proteins 2b and 2x of *Streptococcus pneumoniae* are primary resistance determinants for different classes of antibiotics. *Antimicrob. Agents Chemother.* **40,** 829–834 (1996).

39. Reichmann, P., Konig, A., Marton, A. & Hakenbeck, R. Penicillin-binding Proteins as Resistance Determinants in Clinical Isolates of *Streptococcus pneumoniae*. *Microb. Drug Resist.* **2,** 177–181 (1996).

40. Sibold, C. *et al.* Mosaic *pbpX* genes of major clones of penicillin-resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of a penicillin-sensitive *Streptococcus oralis*. *Mol. Microbiol.* **12,** 1013–1023 (1994).

41. Jensen, A., Valdórsson, O., Frimodt-Møller, N., Hollingshead, S. & Kilian, M. Commensal streptococci serve as a reservoir for β-lactam resistance genes in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **59,** 3529–40 (2015).

42. Boolchandani, M., D'Souza, A. W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat. Rev. Genet.* **20,** 356–370 (2019).

43. Balouiri, M., Sadiki, M. & Ibnsouda, S. K. Methods for in vitro evaluating antimicrobial activity: A review. *J. Pharm. Anal.* **6,** 71–79 (2016).

44. Jorgensen, J. H. & Ferraro, M. J. Antimicrobial Susceptibility Testing: A Review of General Principles and Contemporary Practices. *Clin. Infect. Dis.* **49,** 1749–1755 (2009).

45. Humphries, R. M. *et al.* CLSI methods development and standardization working group best practices for evaluation of antimicrobial susceptibility tests. *J. Clin. Microbiol.* **56,** (2018).

46. Yee, R. & Simner, P. J. Next-Generation Sequencing Approaches to Predicting Antimicrobial Susceptibility Testing Results. *Adv. Mol. Pathol.* **2,** 99–110 (2019).

47. Maurer, F. P., Christner, M., Hentschke, M. & Rohde, H. Advances in Rapid Identification and Susceptibility Testing of Bacteria in the Clinical Microbiology Laboratory: Implications for Patient Care and Antimicrobial Stewardship Programs. *Infect. Dis. Rep.* **9,** 6839 (2017).

48. Su, M., Satola, S. W. & Read, T. D. Genome-Based Prediction of Bacterial Antibiotic Resistance. *J. Clin. Microbiol.* **57,** (2019).

49. Ellington, M. J. *et al.* The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin. Microbiol. Infect.* **23,** 2–22 (2017).

50. Metcalf, B. J. *et al.* Using whole genome sequencing to identify resistance determinants and predict antimicrobial resistance phenotypes for year 2015 invasive pneumococcal disease isolates recovered in the United States. *Clin. Microbiol. Infect.* **22,** 1002.e1-1002.e8 (2016).

51. Li, Y. *et al.* Penicillin-Binding Protein Transpeptidase Signatures for Tracking and Predicting β-Lactam Resistance Levels in *Streptococcus pneumoniae*. *MBio* **7,** (2016).

52. Bettinger, J. A. *et al.* The effect of routine vaccination on invasive pneumococcal infections in Canadian children, Immunization Monitoring Program, Active 2000–2007. *Vaccine* **28,** 2130–2136 (2010).

53. Hicks, L. A. *et al.* Incidence of Pneumococcal Disease Due to Non–Pneumococcal Conjugate Vaccine (PCV7) Serotypes in the United States during the Era of Widespread

PCV7 Vaccination, 1998–2004. *J. Infect. Dis.* **196,** 1346–1354 (2007).

54. Demczuk, W. H. B. *et al.* Serotype distribution of invasive *Streptococcus pneumoniae* in adults 65 years of age and over after the introduction of childhood 13-valent pneumococcal conjugate vaccination programs in Canada, 2010–2016. *Vaccine* **36,** 4701–4707 (2018).

55. Demczuk, W. H. B. *et al.* Serotype distribution of invasive *Streptococcus pneumoniae* in Canada after the introduction of the 13-valent pneumococcal conjugate vaccine, 2010–2012. *Can. J. Microbiol.* **59,** 778–788 (2013).

56. Desai, S., McGeer, A., Quach-Thanh, C., Elliott, D. & approved by NACI. National Advisory Committee on Immunization (NACI) - Update on the Use of Conjugate Pneumococcal Vaccines in Childhood. *Canada Commun. Dis. Rep.* **36,** 1–21 (2010).

57. Streptococcus and STI Unit, Bacteriology and Enteric Diseases Program, National Microbiology Laboratory. National Surveillance of Invasive Streptococcal Disease in Canada - Annual Summary 2016. Public Health Agency of Canada (PHAC). Available at: https://www.researchgate.net/publication/324088684_National_Laboratory_Surveillance_of_Invasive_Streptococcal_Disease_in_Canada_Annual_Summary_2016.

58. Clinical and Laboratory Standards Institute (CLSI). Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically - Eleventh Edition: Approved Standard M07-A11. CLSI, Wayne, PA. (2018).

59. Clinical and Laboratory Standards Institute (CLSI). Performance Standards for Antimicrobial Susceptibility Testing - Twenty-Eighth Edition: M100-S28. CLSI, Wayne, PA. (2018).

60. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27,** 2957–63 (2011).

61. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–77 (2012).

62. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069 (2014).

63. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).

64. Larsson, A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30,** 3276–3278 (2014).

65. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35,** 1547–

1549 (2018).

66. MEGA-X Help. Available at: https://www.megasoftware.net/web_help_10/index.htm#t=First_Time_User.htm&rhsearch=p-distance&ux=search.

67. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44,** W344-50 (2016).

68. Pupko, T., Bell, R. E., Mayrose, I., Glaser, F. & Ben-Tal, N. Rate4Site: An algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. in *Bioinformatics* **18,** S71–S77 (Narnia, 2002).

69. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8,** 275–282 (1992).

70. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33,** 299–302 (2005).

71. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

72. Ragonnet-Cronin, M. *et al.* Automated analysis of phylogenetic clusters. *BMC Bioinformatics* **14,** 317 (2013).

73. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** W320–W324 (2014).

74. ESPript 3: Command-line & Web Guide. Available at: http://espript.ibcp.fr/ESPript/ESPript/esp_userguide.php.

75. Nichol, K. A., Zhanel, G. G. & Hoban, D. J. Penicillin-binding protein 1A, 2B, and 2X alterations in Canadian isolates of penicillin-resistant *Streptococcus pneumoniae. Antimicrob. Agents Chemother.* **46,** 3261–4 (2002).

76. Asahi, Y., Takeuchi, Y. & Ubukata, K. Diversity of substitutions within or adjacent to conserved amino acid motifs of penicillin-binding protein 2X in cephalosporin-resistant *Streptococcus pneumoniae* isolates. *Antimicrob. Agents Chemother.* **43,** 1252–5 (1999).

77. Biçmen, M., Gülay, Z., Ramaswamy, S. V., Musher, D. M. & Gür, D. Analysis of mutations in the pbp genes of penicillin-non-susceptible pneumococci from Turkey. *Clin. Microbiol. Infect.* **12,** 150–155 (2006).

78. Cafini, F. *et al.* Alterations of the penicillin-binding proteins and murM alleles of clinical *Streptococcus pneumoniae* isolates with high-level resistance to amoxicillin in Spain. *J.*

Antimicrob. Chemother. **57,** 224–229 (2006).

79. Kosowska, K., Jacobs, M. R., Bajaksouzian, S., Koeth, L. & Appelbaum, P. C. Alterations of penicillin-binding proteins 1A, 2X, and 2B in *Streptococcus pneumoniae* isolates for which amoxicillin MICs are higher than penicillin MICs. *Antimicrob. Agents Chemother.* **48,** 4020–2 (2004).

80. Davies, T. A., Shang, W., Bush, K. & Flamm, R. K. Activity of doripenem and comparator β-lactams against US clinical isolates of *Streptococcus pneumoniae* with defined mutations in the penicillin-binding domains of *pbp1a*, *pbp2b* and *pbp2x*. *Journal of Antimicrobial Chemotherapy* **61,** 751–753 (2008).

81. Carapito, R., Chesnel, L., Vernet, T. & Zapun, A. Pneumococcal beta-lactam resistance due to a conformational change in penicillin-binding protein 2x. *J. Biol. Chem.* **281,** 1771–7 (2006).

82. Dessen, A., Mouz, N., Gordon, E., Hopkins, J. & Dideberg, O. Crystal structure of PBP2x from a highly penicillin-resistant *Streptococcus pneumoniae* clinical isolate: a mosaic framework containing 83 mutations. *J. Biol. Chem.* **276,** 45106–12 (2001).

83. Chesnel, L. *et al.* The structural modifications induced by the M339F substitution in PBP2x from *Streptococcus pneumoniae* further decreases the susceptibility to beta-lactams of resistant strains. *J. Biol. Chem.* **278,** 44448–56 (2003).

84. Smith, A. M. & Klugman, K. P. Amino acid mutations essential to production of an altered PBP 2X conferring high-level beta-lactam resistance in a clinical isolate of *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **49,** 4622–7 (2005).

85. Gordon, E., Mouz, N., Duée, E. & Dideberg, O. The crystal structure of the penicillin-binding protein 2x from *Streptococcus pneumoniae* and its acyl-enzyme form: implication in drug resistance. *J. Mol. Biol.* **299,** 477–485 (2000).

86. Bernardo-García, N. *et al.* Allostery, Recognition of Nascent Peptidoglycan, and Cross-linking of the Cell Wall by the Essential Penicillin-Binding Protein 2x of *Streptococcus pneumoniae*. *ACS Chem. Biol.* **13,** 694–702 (2018).

87. Mouz, N. *et al.* Mutations in the active site of penicillin-binding protein PBP2x from *Streptococcus pneumoniae*. Role in the specificity for β-lactam antibiotics. *J. Biol. Chem.* **274,** 19175–19180 (1999).

88. Pernot, L. *et al.* A PBP2x from a clinical isolate of *Streptococcus pneumoniae* exhibits an alternative mechanism for reduction of susceptibility to beta-lactam antibiotics. *J. Biol. Chem.* **279,** 16463–70 (2004).

89. Contreras-Martel, C., Dahout-Gonzalez, C., Martins, A. D. S., Kotnik, M. & Dessen, A. PBP active site flexibility as the key mechanism for beta-lactam resistance in

pneumococci. *J. Mol. Biol.* **387,** 899–909 (2009).

90. Calvez, P. *et al.* Substitutions in PBP2b from β-Lactam-resistant *Streptococcus pneumoniae* Have Different Effects on Enzymatic Activity and Drug Reactivity. *J. Biol. Chem.* **292,** 2854–2865 (2017).

91. Pagliero, E. *et al.* Biochemical characterization of *Streptococcus pneumoniae* penicillin-binding protein 2b and its implication in beta-lactam resistance. *Antimicrob. Agents Chemother.* **48,** 1848–55 (2004).

92. Asahi, Y. & Ubukata, K. Association of a thr-371 substitution in a conserved amino acid motif of penicillin-binding protein 1A with penicillin resistance of *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **42,** 2267–73 (1998).

93. Smith, A. M. & Klugman, K. P. Alterations in PBP 1A essential-for high-level penicillin resistance in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **42,** 1329–33 (1998).

94. Job, V., Carapito, R., Vernet, T., Dessen, A. & Zapun, A. Common alterations in PBP1a from resistant *Streptococcus pneumoniae* decrease its reactivity toward beta-lactams: structural insights. *J. Biol. Chem.* **283,** 4886–94 (2008).

95. Contreras-Martel, C. *et al.* Crystal Structure of Penicillin-binding Protein 1a (PBP1a) Reveals a Mutational Hotspot Implicated in β-Lactam Resistance in *Streptococcus pneumoniae. J. Mol. Biol.* **355,** 684–696 (2006).

96. Smith, A. M. & Klugman, K. P. Site-specific mutagenesis analysis of PBP 1A from a penicillin-cephalosporin-resistant pneumococcal isolate. *Antimicrob. Agents Chemother.* **47,** 387–9 (2003).

97. Golden, A. R. Molecular Characterization of the Twelve Predominant *Streptococcus pneumoniae* Serotypes Causing Invasive Infections in Canada from 2011 to 2015. (University of Manitoba, 2017).

98. Leech, N. L., Barrett, K. C. & Morgan, G. A. *IBM SPSS for Intermediate Statistics: Use and Interpretation*. (Taylor & Francis, 2015).

99. James, G., Witten, D., Hastie, T. & Tibshirani, R. Resampling Methods in. *An Introduction to Statistical Learning, vol 103* 175–201 (Springer, New York, NY, 2013).

100. Linear Regression: Saving New Variables. (2017). Available at: https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/base/idh_regr_sav.html.

101. Barkus, V., Ghanekar, K., Yeo, M., Coffey, T. & Dowson, C. Genetics of high level penicillin resistance in clinical isolates of *Streptococcus pneumoniae*. *FEMS Microbiol.*

Lett. **126,** 299–303 (1995).

102. Stanhope, M. J. *et al.* Positive selection in penicillin-binding proteins 1a, 2b, and 2x from *Streptococcus pneumoniae* and its correlation with amoxicillin resistance development. *Infect. Genet. Evol.* **8,** 331–339 (2008).

103. Sauerbier, J., Maurer, P., Rieger, M. & Hakenbeck, R. *Streptococcus pneumoniae* R6 interspecies transformation: genetic analysis of penicillin resistance determinants and genome-wide recombination events. *Mol. Microbiol.* **86,** 692–706 (2012).

104. James, G., Witten, D., Hastie, T. & Tibshirani, R. Linear Model Selection and Regularization in. *An Introduction to Statistical Learning*. 203–264 (Springer, New York, NY, 2013).

105. Li, Y. *et al.* Validation of β-lactam minimum inhibitory concentration predictions for pneumococcal isolates with newly encountered penicillin binding protein (PBP) sequences. *BMC Genomics* **18,** 621 (2017).

106. Davis, J. *et al.* Antimicrobial Resistance Prediction in PATRIC and RAST. *Sci. Rep.* **7,** (2016).

107. Metcalf, B. J. *et al.* Strain features and distributions in pneumococci from children with invasive disease before and after 13-valent conjugate vaccine implementation in the USA. *Clin. Microbiol. Infect.* **22,** 9–29 (2016).

108. Eyre, D. *et al.* WGS to predict antibiotic MICs for *Neisseria gonorrhoeae*. *J. Antimicrob. Chemother.* **72,** 1937–1947 (2017).

109. Tamma, P. *et al.* Applying Rapid Whole-Genome Sequencing To Predict Phenotypic Antimicrobial Susceptibility Testing Results among Carbapenem-Resistant *Klebsiella pneumoniae* Clinical Isolates. *J. Antimicrob. Chemother.* **63,** (2019).

110. Chesnel, L. *et al.* Identical Penicillin-Binding Domains in Penicillin-Binding Proteins of *Streptococcus pneumoniae* Clinical Isolates with Different Levels of β-Lactam Resistance. *Antimicrob. Agents Chemother.* **49,** 2895–902 (2005).

111. Mobegi, F. M. *et al.* Deciphering the distance to antibiotic resistance for the pneumococcus using genome sequencing data. *Sci. Rep.* **7,** 42808 (2017).

10. APPENDIX A

TABLE 10.1: Description of the eleven previously reported substitutions in PBP2x associated with reduced β-lactam-susceptibility in *S. pneumoniae*

Gene	Amino acid substitution			Summary/rationale	Reference (s)
	Allele 1	Position	Allele 2		
pbp2x	T	338	A	- The T338 residue is located within the conserved STMK active site motif. This substitution is commonly identified in penicillin-intermediate and -resistant isolates of *S. pneumoniae* globally.	[75,76,77,78,79,80]
				- Introduction of the T338A substitution in *pbp2x* of *S. pneumoniae* R6 resulted in reduced acylation efficiency of R6-PBP2x*T338A for cefotaxime.	[81]
				- Reversal of the T338A substitution in *S. pneumoniae* R6 transformed with *pbp2x* from β-lactam-resistant *S. pneumoniae* isolate 5204 decreased cefotaxime MICs from 1 µg/mL to 0.75 µg/mL.	[81]
				- Analysis of the high-resolution crystal structure of PBP2x from the penicillin-resistant *S. pneumoniae* isolate sp328[82] and PBP2x from *S. pneumoniae* R6 bearing the T338A substitution[83] revealed the loss of a stabilizing hydrogen bond between T338 and a conserved water molecule within the PBP2x active site.	[82, 83]
				- Reversal of the T338A substitution in *S. pneumoniae* R6 transformed with *pbp2x* from β-lactam-resistant *S. pneumoniae* isolate 3191 resulted in 8- and 4-fold decreases in penicillin and cefotaxime MICs, respectively.	[84]
	L	364	F	- Reversal of the L364F substitution in *S. pneumoniae* R6 transformed with *pbp2x* from β-lactam-resistant *S. pneumoniae* isolate 3191 resulted in 2-fold decreases in penicillin and cefotaxime MICs.	[84]
	A	369	V	- Reversal of the A369V substitution in *S. pneumoniae* R6 transformed with *pbp2x* from β-lactam-resistant isolate	[84]

				3191 resulted in 2-fold decreases in penicillin and cefotaxime MICs.	
	I	371	T	- Reversal of the I371T substitution in *pbp2x* of β-lactam resistant *S. pneumoniae* isolate 5204 resulted in an increased acylation efficiency of 5204-PBP2x*T371I for cefotaxime.	[81]
				- Reversal of the I371T substitution in *S. pneumoniae* R6 transformed with *pbp2x* from β-lactam-resistant *S. pneumoniae* isolate 5204 decreased cefotaxime MICs from 1 μg/mL to 0.5 μg/mL.	[81]
				- Reversal of the I371T substitution in *S. pneumoniae* R6 transformed with *pbp2x* from β-lactam-resistant *S. pneumoniae* isolate 3191 resulted in 8- and 4-fold decreases in penicillin and cefotaxime MICs, respectively.	[84]
	R	384	G	- Introduction of the R384G substitution in *pbp2x* of *S. pneumoniae* R6 resulted in a decrease in the acylation efficiency of R6-PBP2x*R384G for cefotaxime.	[81]
				- Reversal of the R384G substitution in *S. pneumoniae* R6 transformed with *pbp2x* from β-lactam-resistant *S. pneumoniae* isolate 5204 decreased cefotaxime MICs from 1 μg/mL to 0.5 μg/mL.	[81]
				- Reversal of the R384G substitution in *S. pneumoniae* R6 transformed with *pbp2x* from β-lactam-resistant *S. pneumoniae* isolate 3191 resulted in 2-fold decreases in penicillin and cefotaxime MICs.	[84]
	S	389	L	- Analysis of the high-resolution crystal structure of PBP2x from the penicillin-resistant *S. pneumoniae* isolate Sp328 revealed that the S389L substitution contributes to the displacement of the loop containing the conserved AHSSNV active site motif as a result of steric clash.	[82]
	H	394	L	- The H394 residue is located adjacent to the conserved SSN active site motif. This substitution is commonly identified in penicillin-intermediate and -resistant isolates of *S. pneumoniae* globally.	[75,76,79]

	N	514	H	-	Analysis of the high resolution crystal structure of PBP2x from the penicillin-resistant *S. pneumoniae* isolate Sp328 revealed that the N514H substitution contributes to the displacement of the α4 loop containing the SXN motif as a result of steric clash.	[82]
	L	546	V	-	The L546 residue is located adjacent to the conserved KSG active site motif. This substitution is commonly identified in penicillin-intermediate and -resistant isolates of *S. pneumoniae* globally.	[75,76,78,79,80]
	Q	552	E	-	Analysis of the high-resolution crystal structure of PBP2x from *S. pneumoniae* R6 in complex cefuroxime revealed hydrogen bonding between Q552 and the furan group of cefuroxime.	[85]
				-	Analysis of the high-resolution crystal structure of PBP2x from *S. pneumoniae* RX1 in complex with cefepime, along with modeling of peptidoglycan in complex with PBP2x revealed that Q552 anchors nascent peptidoglycan strands near catalytic residues within the PBP2x active site.	[86]
				-	Introduction of the Q552E substitution in *pbp2x* of *S. pneumoniae* R6 significantly reduced the acylation efficiency of R6-PBP2x*Q552E for penicillin G and cefotaxime.	[87, 88]
				-	Reversal of the Q552E substitution in *S. pneumoniae* R6 transformed with *pbp2x* from *S. pneumoniae* isolate 5259 resulted in 6- and 2-fold decreases in penicillin and cefotaxime MICs, respectively.	[88]
				-	Analysis of the high-resolution crystal structure of PBP2x from *S. pneumoniae* isolate 5259 revealed that Q552E is situated near the entrance of the PBP2x active site cleft and introduces a net-negative charge into this region.	[88]
	N	605	T	-	Introduction of the N605T substitution in *pbp2x* of *S. pneumoniae* R6 reduced the acylation efficiency of R6-PBP2x*N605T for cefotaxime.	[81]
				-	Reversal of the N605T substitution in *S. pneumoniae* R6 transformed with *pbp2x* from β-lactam resistant *S.*	[81]

pneumoniae isolate 5204 decreased cefotaxime MICs from 1 µg/mL to 0.75 µg/mL.

TABLE 10.2: Description of the nine previously reported substitutions in PBP2b associated with reduced β-lactam-susceptibility in *S. pneumoniae*

Gene	Amino acid substitution			Summary/rationale	Reference
	Allele 1	Position	Allele 2		
pbp2b	S	412	P	- Analysis of the high-resolution crystal structure of PBP2b from β-lactam-resistant *S. pneumoniae* strain 5204 revealed that these substitutions are concentrated around the entry of the PBP2b TP domain active site, thereby altering the charge and polarity of this region.	89,90
	N	422	Y		
	T	426	K		
	Q	427	L		
	T	446	A	- T446 is located within the conserved SSNT active site motif. This substitution is commonly identified in penicillin-intermediate and -resistant isolates of *S. pneumoniae* globally	75,78,79,80
				- Analysis of the high-resolution crystal structure of PBP2b from *S. pneumoniae* R6 showed that T446 forms multiple interactions that stabilize the α4/α5 loop containing the SSNT active site motif.	89
				- The *in vitro* transpeptidase activity of purified PBP2b-R6*T446A was significantly reduced compared to wild-type PBP2b-R6, generating a decreased amount of cross-linked peptidoglycan. The acylation efficiency of purified PBP2b-R6*T446A for multiple β-lactams including penicillin G, piperacillin and cefotaxime was reduced compared to PBP2b-R6.	90
				- Introduction of the T446A substitution in *pbp2b* of *S. pneumoniae* R6 resulted in a twofold increase in piperacillin MIC from 0.016 µg/mL to 0.032 µg/mL. The acylation efficiency of purified PBP2b-R6*T446A for penicillin G was reduced compared to PBP2b-R6.	91
	A	619	G	- The A619 residue is located directly adjacent to the KTG active site motif. This substitution is commonly identified in penicillin-intermediate and -resistant isolates of *S. pneumoniae*.	75,78,80
				- Analysis of the high resolution crystal structure of PBP2b from β-lactam-resistant *S. pneumoniae* strain 5204	89

					revealed that the A619G substitution contributes to the observed increase in flexibility of the β3a/β3b loop positioned at the entry of the PBP2b TP domain active site cleft.	
	D Q T	625 628 630	G E N	-	Analysis of the high-resolution crystal structure of PBP2b from β-lactam-resistant *S. pneumoniae* strain 5204 revealed that these substitutions increase the flexibility of the β3a/β3b loop positioned at the entry of PBP2b TP domain active site cleft.	[89,90]

TABLE 10.3: Description of the eleven previously reported substitutions in PBP1a associated with reduced β-lactam-susceptibility in *S. pneumoniae*

Gene	Amino acid substitution			Summary/rationale	Reference
	Allele 1	Position	Allele 2		
pbp1a	T	371	A/S	- The T371 residue is located within the STMK active site motif. This substitution is commonly identified in penicillin-intermediate and -resistant isolates of *S. pneumoniae* globally.	75,77,78,79,80,92
				- Reversal of the T371A/S substitution in *S. pneumoniae* R6 transformed with *pbp1a* from penicillin-resistant *S. pneumoniae* isolate 8303 decreased penicillin MICs from 1.5 µg/mL to 0.5 µg/mL.	93
				- Introduction of the T371A substitution in *pbp1a* of *S. pneumoniae* R6 decreased the acylation efficiency of R6-PBP1a*T371A for penicillin and cefotaxime.	94
				- Reversal of the T371A/S substitution in *S. pneumoniae* R6 transformed with *pbp1a* from β-lactam-resistant isolate 5204 decreased cefotaxime MICs from 1 µg/mL to 0.5 µg/mL.	94
				- Analysis of the high-resolution crystal structure of PBP1a from β-lactam-resistant *S. pneumoniae* isolate 5204 revealed that the T371A/S substitution causes a shift in the STMK active site motif as a result of the loss of a stabilizing hydrogen bond.	94
	P	432	T	- The P432 residue is located adjacent to the conserved SRN active site motif. This substitution is commonly identified in penicillin-intermediate and -resistant isolates of *S. pneumoniae*.	75,77,78,80
				- Analysis of the high-resolution crystal structure of PBP1a from β-lactam-resistant *S. pneumoniae* isolate 5204 revealed that the P432T substitution forms an alternative hydrogen bonding pattern within the PBP1a active site compared to *S. pneumoniae* R6, potentially resulting in active site destabilization.	94

	N	546	G	-	Analysis of the high resolution crystal structure of PBP1a from *S. pneumoniae* R6 and mapping of substitutions identified in β-lactam resistant strain 73311 revealed that the N546G substitution is located near the entry of the PBP1a active site cleft, potentially affecting active site accessibility. [95]
	TSQF	574-577	NTGY	-	Reversal of the TSQF574-577NTGY block of substitutions in *S. pneumoniae* R6 transformed with *pbp1a* from penicillin-resistant isolate 8303 decreased penicillin MICs from 1.5 µg/mL to 0.375 µg/mL [93]
				-	Introduction of the TSQF574-577NTGY block of substitutions in *pbp1a* of *S. pneumoniae* R6 significantly reduced the acylation efficiency of PBP1a-R6*TSQF574-577NTGY for penicillin and cefotaxime. [94]
				-	Reversal of the TSQF574-577NTGY block of substitutions in *S. pneumoniae* R6 transformed with *pbp1a* from β-lactam-resistant isolate 5204 decreased cefotaxime MICs from 1 µg/mL to 0.75 µg/mL. [94]
				-	Analysis of the high-resolution crystal structure of PBP1a from β-lactam-resistant *S. pneumoniae* isolate 5204 revealed that this block of substitutions increases the flexibility of the β3-β4 loop positioned at the entry of the TP domain active site, potentially affecting antibiotic active site access. [94]
				-	Reversal of the TSQF574-577NTGY block of substitutions in *S. pneumoniae* R6 transformed with *pbp1a* from β-lactam-resistant isolate 3191 decreased penicillin MICs from 4 µg/mL to 0.5 µg/mL and cefotaxime MICs from 2 µg/mL to 0.5 µg/mL. [96]
	L N L T	606 609 611 612	I D F L	-	Analysis of the high-resolution crystal structure of PBP1a from *S. pneumoniae* R6 and mapping of substitutions identified in β-lactam-resistant strain 73311 revealed that this group of substitutions are concentrated around the entrance of the PBP1a active site cleft. [95]

11. APPENDIX B

TABLE 11.1: Univariate correlations (r) between the eleven previously reported substitutions in PBP2x associated with reduced β-lactam-susceptibility in *S. pneumoniae*

		T338A/S	L364F	A369V	I371T	R384G	S389L	H394L	N514H	L546V	Q552E	N605T
Correlation[a]	**T338A/S**	**1.000**	0.736	0.269	0.739	0.823	0.561	-0.217	0.169	0.696	-0.410	0.587
	L364F	0.736	**1.000**	-0.336	0.982	0.632	0.836	-0.163	0.355	0.947	-0.309	0.871
	A369V	0.269	-0.336	**1.000**	-0.330	0.165	-0.194	0.441	-0.265	-0.262	-0.211	-0.239
	I371T	0.739	0.982	-0.330	**1.000**	0.621	0.821	-0.161	0.351	0.948	-0.303	0.856
	R384G	0.823	0.632	0.165	0.621	**1.000**	0.594	-0.259	0.225	0.571	-0.121	0.529
	S389L	0.561	0.836	-0.194	0.821	0.594	**1.000**	0.169	0.248	0.832	-0.260	0.877
	H394L	-0.217	-0.163	0.441	-0.161	-0.259	0.169	**1.000**	-0.096	0.019	-0.103	0.106
	N514H	0.169	0.355	-0.265	0.351	0.225	0.248	-0.096	**1.000**	0.326	0.466	0.314
	L546V	0.696	0.947	-0.262	0.948	0.571	0.832	0.019	0.326	**1.000**	-0.320	0.834
	Q552E	-0.410	-0.309	-0.211	-0.303	-0.121	-0.260	-0.103	0.466	-0.320	**1.000**	-0.289
	N605T	0.587	0.871	-0.239	0.856	0.529	0.877	0.106	0.314	0.834	-0.289	**1.000**
KMO Measure[b]		0.709										
Bartlett's Test[c]		Chi-Square = 4185.211 Significance value = <0.001										

[a], correlation coefficient (r) values, ranging from -1 to 1, are presented in the correlation matrix to show how the eleven selected PBP2x variables are associated with each of the other ten. [b], Measures the suitability of the input variables for PCA; value should be ≥ .70. [c], Tests for the assumption of correlation amongst variables required for PCA; variables are correlated enough if significance value is <0.05.

TABLE 11.2: Univariate correlations (r) between the nine previously reported substitutions in PBP2b associated with reduced β-lactam-susceptibility in *S. pneumoniae*

		T446A	S412P	N422Y	T426K	Q427L	A619G	D625G	Q628E	T630N
Correlation (r)[a]	**T446A**	**1.000**	0.647	0.554	0.554	0.589	0.334	0.669	0.703	0.669
	S412P	0.647	**1.000**	0.856	0.856	0.791	0.516	0.631	0.589	0.631
	N422Y	0.554	0.856	**1.000**	1.000	0.940	0.603	0.498	0.462	0.498
	T426K	0.554	0.856	1.000	**1.000**	0.940	0.603	0.498	0.462	0.498
	Q427L	0.589	0.791	0.940	0.940	**1.000**	0.566	0.438	0.400	0.438
	A619G	0.334	0.516	0.603	0.603	0.566	**1.000**	0.487	0.464	0.487
	D625G	0.669	0.631	0.498	0.498	0.438	0.487	**1.000**	0.953	1.000
	Q628E	0.703	0.589	0.462	0.462	0.400	0.464	0.953	**1.000**	0.953
	T630N	0.669	0.631	0.498	0.498	0.438	0.487	1.000	0.953	**1.000**
KMO Measure[b]		0.729								
Bartlett's Test[c]		Chi-Square = 787.86 Significance value = <0.001								

[a], correlation coefficient (r) values, ranging from -1 to 1, are presented in the correlation matrix to show how the eleven selected PBP2x variables are associated with each of the other ten. [b], Measures the suitability of the input variables for PCA; value should be ≥ .70. [c], Tests for the assumption of correlation amongst variables required for PCA; variables are correlated enough if significance value is <0.05.

TABLE 11.3: Univariate correlations (r) between the eleven previously reported substitutions in PBP1a associated with reduced β-lactam-susceptibility in *S. pneumoniae*

		T371A/S	P432T	N546G	TSQF574-577NTGY	L606I	N609D	L611F	T612L
Correlation (r)[a]	**T371A/S**	**1.000**	0.828	0.765	0.729	0.923	0.718	0.956	0.956
	P432T	0.828	**1.000**	0.690	0.658	0.849	0.663	0.879	0.879
	N546G	0.765	0.690	**1.000**	0.953	0.747	0.945	0.786	0.786
	TSQF574-577NTGY	0.729	0.658	0.953	**1.000**	0.776	0.992	0.749	0.749
	L606I	0.923	0.849	0.747	0.776	**1.000**	0.782	0.996	0.996
	N609D	0.718	0.663	0.945	0.992	0.782	**1.000**	0.755	0.755
	L611F	0.956	0.879	0.786	0.749	0.996	0.755	**1.000**	1.000
	T612L	0.956	0.879	0.786	0.749	0.996	0.755	1.000	**1.000**
KMO Measure[b]		0.784							
Bartlett's Test[c]		Chi-Square = 2903.48 Significance value = <0.001							

[a], correlation coefficient (r) values, ranging from -1 to 1, are presented in the correlation matrix to show how the eleven selected PBP2x variables are associated with each of the other ten. [b], Measures the suitability of the input variables for PCA; value should be ≥ .70. [c], Tests for the assumption of correlation amongst variables required for PCA; variables are correlated enough if significance value is <0.05.