Generation of Chimeric P-glycoprotein for Functional and Structural Investigations



Thesis submitted for the degree of Doctor of Philosophy

Kristen M. Pluchino

Keble College

Nuffield Department of Medicine

University of Oxford

Trinity Term, 2015

Generation of Chimeric P-glycoprotein for Functional and Structural Investigations

Kristen M. Pluchino

Nuffield Department of Medicine, Keble College

Thesis submitted for the degree of Doctor of Philosophy, Trinity Term, 2015

Abstract

A major challenge in cancer treatment is acquired or intrinsic multidrug resistance (MDR) to chemotherapeutics. A notorious mediator of MDR is P-glycoprotein (P-gp, ABCB1), product of the human *MDR1* gene, which actively effluxes cytotoxic drugs from cancer cells, resulting in sub-therapeutic intracellular concentrations. Understanding how P-gp interacts with drugs has been severely limited by the lack of high-resolution structures of P-gp. Although numerous efforts to obtain an X-ray crystal structure of P-gp have been attempted, human P-gp has never been crystallized. However, mouse P-gp (87% homologous to human P-gp) has been crystallized, and several structures of mouse P-gp have been recently reported. Despite a high degree of homology, it is currently unknown why mouse P-gp can be crystallized while human P-gp cannot. The studies presented in this thesis describe the creation of novel chimeras of mouse and human P-gp as an approach to investigate whether specific protein domains are responsible for differences in the ability to form crystals between mouse and human P-gp. A range of chimeras, created by protein domain swapping, were expressed in mammalian cells and all were found to retain MDR transport function demonstrating that P-gp can tolerate major structural changes. High-level expression of all chimeras was achieved by baculovirus-mediated heterologous protein expression. Chimeric proteins were purified by a multi-step process including immobilized metal affinity chromatography and size exclusion chromatography. Crystallization screening obtained protein crystals for two of the chimeras, indicating the approach adopted is a successful strategy, and an advance along the path towards a high-resolution structure of human P-gp.

Acknowledgments

I would like to wholeheartedly thank my mentors Michael Gottesman, Deborah Gill, and Steven Hyde, for giving me the opportunity to work in their labs in the OxCam program. Michael, you've supported my every scientific endeavor, a fact that I cannot thank you enough for. Also, thank you for conceiving the original concept of this project, which has allowed me to collaborate with (at least!) four different laboratories on two different continents. Deborah and Steve, I'd like to thank you for adopting me into the Gene Medicine group and giving me a home at Oxford. I entered into your lab with no molecular biology knowledge, and returned to the NIH confident in my newly learned research skills.

To Matthew Hall, I will always appreciate the time, commitment, support, and knowledge that you've shared with me. Without your expertise, enthusiasm and brilliant ideas, it is impossible for me to imagine the attainment of my DPhil. In addition to your scientific mentorship, you've been a source of laughter and friendship over the years. Thank you for the countless chats, meetings, and coffees. I am truly grateful for all you have given me.

One of my favorite aspects of this thesis project has been the ability to collaborate with the laboratories of Suresh Ambudkar and Di Xia. I would like to thank Suresh Ambudkar, who I consider a walking dictionary of all things ABC transporter related. You've been an immense help throughout the project. To Dia Xia, this collaboration would never have been possible without your support and the various resources and expertise in your laboratory.

Fei Zhou, I would like to thank you for your patient mentoring. I never would have thought I would enjoy protein expression and purification so much, but your never-ending optimism and cheerful smile made these experiences fun. You've been both a joy to work with and an enormous source of encouragement and support.

To Suneet Shukla, who I've been bothering with questions from the minute I started in the lab. I always enjoy our conversations, whether they are about P-gp, future careers, or travel. In addition to being a source of scientific support, you've also been a great friend over the years, and it's been a joy working with you.

I would like to thank Eduardo Chufan for the many hours spent with me on crystal contact analysis, whose kind patience has helped me decrypt the mystery of protein crystallization. I'd also like to thank Lothar Esser, who has been an immense help with both FPLC and crystallization screening. Thank you for always answering my (many) questions about crystallography.

I would like to thank Dom Esposito, who has kept me sane over the years by corroborating all of the crazy data I've observed while cloning mouse P-gp. Whether I was in England or Bethesda, you were always a phone call or email away with helpful advice and guidance.

Also, thank you to the many labmates, past and present, that have made it fun to be in lab over the years: Lora Weidner, Asia Poprawski, Jeyan Kumar, Jim Madigan, and all of the members of GeneMedicine. A special thank you to Janna Moen, who has been a true asset to this project. Also, thank you to Ian Pringle, who taught me many molecular biology tricks, and also gave me the best going-away present I've ever received (liquid chalk!).

I am grateful for my friends on both sides of the Atlantic that have made my journey an amazing experience. To Donah, Dave, India Clique, and many other Kebelites for giving me the best Oxford experience I could have imagined. Also, I am grateful for my fellow OxCammers, who have become a second family to me over the years.

George, I am ever grateful for your help proof-reading (and hyphenating!) this beast of a thesis.

I am thankful for my many past scientific mentors who have always supported and encouraged my graduate studies. I'd like to thank Eric James, Stephen Hoffman, and Kim Lee Sim at Sanaria for letting me share their dream of a malaria vaccine. To Richard Callaghan, who first gave me a home in Oxford. Also, a special thanks to Thierry Olivry, who gave me my first laboratory experiences and sparked my passion for scientific research.

To my parents, thank you for always supporting my endeavors (even if those endeavors sent me across an ocean).

Declarations

I, Kristen Pluchino, declare that I am the sole author of this thesis, and it has not been used in any previous application for a degree. All sources of published or unpublished information have been specifically acknowledged by means of a reference or personal communication notation. All results presented here were obtained by myself, with the following exceptions:

Chapter 3: GFP fusion plasmids were designed in collaboration with Dominic Esposito (Frederick National Laboratory for Cancer Research, FNLCR) and assembled by the Protein Expression Laboratory at FNLCR.

Chapter 4: Determination of crystal contacts was completed in collaboration with Eduardo Chufan (Laboratory of Cell Biology, NIH).

Chapter 5: The generation of BacMam virus used for infection of mammalian cells and baculovirus used for infection of insect cells was completed by the Protein Expression Laboratory at FNLCR.

Chapter 6: Crystallization trials were undertaken by Lothar Esser (NIH, USA).

Table of Contents

Chapter 1: Introduction

1.1 ABC transporters..... 1

1.1.1 The ABC transporter superfamily..... 1

1.1.2 ABC transporters in human genetic disorders..... 2

1.1.3 Excretory and protective ABC transporters..... 4

1.1.4 ABC transporters and the blood-brain-barrier..... 5

1.1.5 ABC transporters and cancer multidrug resistance..... 8

1.2 P-glycoprotein..... 10

1.2.1 Multidrug resistance..... 10

1.2.2 Clinical relevance of P-glycoprotein in multidrug resistance..... 11

1.2.3 Intrinsic and acquired multidrug resistance..... 12

1.2.4 P-glycoprotein nomenclature..... 16

1.2.5 Species differences and P-glycoprotein substrate affinities..... 16

1.2.6 Terminology: substrates, inhibitors and competitive inhibitors..... 19

1.2.7 P-glycoprotein mechanism of transport..... 20

1.3 P-glycoprotein structure..... 22

1.3.1 P-glycoprotein structure overview..... 22

1.3.2 *Staphylococcus aureus* Sav1866..... 23

1.3.3 Mouse P-glycoprotein (*mdr1a*)..... 25

1.3.4 *Caenorhabditis elegans* P-glycoprotein..... 27

1.3.5 Structure of human ABCB10..... 29

1.4 Expression, purification, and crystallization of membrane proteins..... 30

1.4.1 Expression systems for large-scale purification of membrane proteins..... 30

1.4.2 Heterologous protein expression in *P. pastoris*..... 32

1.4 3 Codon optimization of P-glycoprotein cDNA..... 33

1.4.4 Heterologous protein expression in insect cells..... 37

1.4.5 Baculovirus-based expression in mammalian cells..... 38

1.4.6 Purification of membrane proteins..... 38

1.4.7 Crystallization of membrane proteins..... 41

1.5 Chimeras..... 42

1.5.1 Chimera nomenclature..... 42

1.5.2 Chimeras for structural investigations..... 42

1.5.3 Functional chimeras of ABC transporters..... 44

1.6 Study aims…… 45

Chapter 2: Materials and Methods

2.1 Materials…… 46
2.1.1 Chemicals…… 46
2.1.2 Cell culture reagents and solutions…… 46
2.1.3 Antibodies…… 46
2.1.4 Additional reagents…… 47
2.1.5 Equipment…… 48
2.1.6 Centrifugation…… 49
2.1.7 Software…… 49
2.1.8 Plasmids…… 50
2.1.9 Enzymes…… 51
2.1.10 Human *MDR1* and mouse *mdr1a* cDNA…… 51
2.1.11 Primers…… 51
2.2 Methods…… 55
2.2.1 Amplification of cDNA by polymerase chain reaction…… 55
2.2.2 Agarose gel electrophoresis…… 55
2.2.3 Agarose gel purification and extraction of DNA fragments…… 55
2.2.4 Topocloning of PCR products…… 56
2.2.5 Restriction enzyme digests…… 56
2.2.6 Ligation of DNA fragments…… 56
2.2.7 Preparation of antibiotic stock solution…… 57
2.2.8 Bacterial transformation of plasmid DNA…… 57
2.2.9 Preparation of selective LB media and LB agar plates…… 57
2.2.10 Inoculation of bacterial colonies into LB selective media…… 58
2.2.11 Large- and small-scale purification of plasmid DNA from transformed bacteria…… 58
2.2.12 Gateway cloning…… 58
2.2.13 DNA quantification…… 59
2.2.14 Codon optimization of cDNA…… 59
2.2.15 Sequencing and sequence analysis of DNA…… 59
2.2.16 Mammalian cell culture…… 59
2.2.17 Mammalian cell lysates…… 60
2.2.18 Protein estimation…… 60
2.2.19 Sodium dodecyl sulfate polyacrylamide gel electrophoresis…… 60
2.2.20 Western blotting…… 61

2.2.21 Colloidal blue stain…… 61
2.2.22 Blue native polyacrylamide gel electrophoresis…… 61
2.2.23 MRK-16 P-gp cell surface expression assay…… 62
2.2.24 Substrate efflux and inhibition assay…… 62
2.2.25 Flow cytometry…… 63
2.2.26 ATPase assay…… 63
2.2.27 Cytotoxicity assay…… 64
2.2.28 Fluorescence spectroscopy…… 64
2.2.29 Confocal microscopy…… 65
2.2.30 Generation of baculovirus and BacMam baculovirus…… 65
2.2.31 Baculovirus-mediated protein expression in insect cells…… 65
2.2.32 BacMam transfection…… 66
2.2.33 Lipofectamine transfection…… 66
2.2.34 *P. pastoris* cell culture…… 66
2.2.35 Measurement of optical density…… 67
2.2.36 Preparation of *P. pastoris* competent cells…… 67
2.2.37 *P. pastoris* transformation…… 68
2.2.38 Small- and large-scale protein expression in *P. pastoris*…… 68
2.2.39 Preparation of *P. pastoris* cell lysates…… 68
2.2.40 Preparation of crude membranes from insect cells…… 69
2.2.41 Protein purification by immobilized metal affinity chromatography…… 70
2.2.42 Size-exclusion chromatography…… 71
2.2.43 Crystallization screening…… 71
2.2.44 Statistical analysis…… 71

Chapter 3: Identification of a cryptic bacterial promoter in mouse P-glycoprotein cDNA

3.1 Introduction…… 73
3.2 Results…… 78
3.2.1 Cloning of *mdr1a* cDNA in *E. coli*…… 78
3.2.2 Literature review of *mdr1a* cDNA toxicity in bacterial hosts…… 81
3.2.3 Personal communications regarding *mdr1a* toxicity…… 83
3.2.4 FlexProm *in silico* analysis of bacterial promoter elements…… 83
3.2.5 Construction of *mdr1a*-GFP fusion constructs…… 86
3.2.6 Fluorescence screening of transcriptional *mdr1a*-GFP fusion constructs in *E.coli*…… 88
3.2.7 Transformation of *E. coli* with *mdr1a*_M107L cDNA…… 91

3.2.8 Functional characterization of M107L P-gp..... 92
3.3 Discussion..... 97
3.4 Conclusions..... 99

Chapter 4: Generation of P-glycoprotein Chimeras

4.1 Introduction..... 100
4.2 Results..... 104
4.2.1 Generation of HHHH_WT and HHMM P-gp..... 104
4.2.1.1 Definition of domain boundaries in *mdr1a* and *MDR1* P-gp cDNA..... 104
4.2.1.2 Strategy to introduce a silent restriction enzyme site into the NBD1-TMD2 domain boundary in *mdr1a* and *MDR1* cDNA..... 106
4.2.1.3 Design of PCR primers to introduce a ClaI restriction enzyme site at the NBD1-TMD2 domain boundary of *MDR1* and *mdr1a* cDNA..... 108
4.2.1.4 Generation and validation of pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm, pTOPO_mdr1a_Nterm, and pTOPO_mdr1a_Cterm plasmids..... 110
4.2.1.5 Generation of pTOPO_HHHH_WT and pTOPO_HHMM plasmids..... 114
4.2.2 Generation of HHHM, HMHH, and HMHM P-gp Chimeras..... 118
4.2.2.1 *In silico* design of codon optimized *MDR1* and *mdr1a* cDNA..... 118
4.2.2.2 Generation of pTOPO_HHHH_Opti and pTOPO_MMMM_Opti plasmids..... 120
4.3.2.3 PCR, ligation, and cloning of pTOPO_M1_Opti, pTOPO_M2_Opti, pTOPO_M3_Opti and pTOPO_M4_Opti..... 121
4.2.2.4 Generation of pTOPO_HHHM, pTOPO_HMHH, and pTOPO_HMHM plasmids..... 122
4.2.2.5 Chimera generation concluding remarks..... 124
4.2.3 Analysis of crystal contacts in the mouse P-glycoprotein crystal structure..... 125
4.2.3.1 Strategy for crystal contact analysis..... 125
4.2.3.2 Analysis of crystal contacts in the asymmetric unit of mouse P-gp..... 126
4.2.3.3 Analysis of crystal contacts between the asymmetric unit of P-gp and neighboring P-gp molecules..... 128
4.2.3.4 Retention of crystal contact residues in human P-gp..... 131
4.2.3.5 Retention of crystal contact residues in P-gp chimeras..... 133
4.3 Discussion..... 135
4.4 Conclusions..... 139

Chapter 5: Functional Characterization of P-glycoprotein Chimeras

5.1 Introduction…….. 140

5.1.1 Flow cytometry assays for P-gp cell membrane expression and measurement of efflux capacity…… 141

5.1.2 ATPase assays…… 143

5.2 Results…… 146

5.2.1 Transient transduction of HeLa cells with BacMam baculovirus…….. 146

5.2.2 Effect of sodium butyrate on transient transduction by BacMam…….. 146

5.2.3 Reproducibility of transduction of HeLa cells by BacMam virus…….. 149

5.2.4 Confirmation of chimeric P-gp by PEPG-13 western blotting…….. 150

5.2.5 Expression of human wild-type, codon-optimized and chimeric P-gp in HeLa cells…… 151

5.2.6 Cell surface expression of P-gp chimeras by MRK-16 binding…… 153

5.2.7 Steady state Rh123 123, CAM, and JC-1 efflux assays…… 156

5.2.8 Inhibition of transport function of chimeras with TQR, CsA, verapamil, and elacridar. 159

5.2.9 Efflux rate of Rh123 by chimeras…… 161

5.2.10 Ability of chimeras to confer a MDR phenotype to HeLa cells…… 163

5.2.11 ATPase Assays…… 166

5.3 Discussion…… 170

5.4 Conclusions…… 172

Chapter 6: Expression, purification, and crystallization of P-glycoprotein chimeras

6.1 Introduction…… 173

6.2 Results…… 178

6.2.1 Expression of HHMM P-gp in *P. pastoris*…… 178

6.2.1.1 Generation pPICZ_HHMM and transformation into *P. pastoris*…… 178

6.2.1.2 Small-scale optimization of expression conditions in *P. pastoris*…… 181

6.2.1.3 Western blotting of small-scale HHMM chimeric expression…… 182

6.2.1.4 Large-scale expression of HHMM P-gp in *P. pastoris*…… 185

6.2.2 Protein expression, purification, and crystallization screening from chimeric P-gp expressed in insect cells…… 187

6.2.2.1 Expression of P-gp chimeras in insect cells…… 187

6.2.2.2 Analysis of small- and large-scale expression of chimeras in insect cells…… 187

6.2.2.3 Purification of chimeras from insect cell membranes: overview…… 189

6.2.2.4 Purification analysis of HHMM, HMHM, HMHH, and HHHM chimeras…… 191

6.2.2.5 SEC of HHMM, HMHM, HMHH, and HHHM chimeric P-gp…… 194

6.2.2.6 Yield of HHMM, HMHM, HMHH, and HHHM purification…… 198

6.2.2.7 Assessment of homogeneity of purified P-gp chimeras..... 200
6.2.2.8 Crystallization screening of P-gp chimeras..... 202
6.3 Discussion..... 204
6.4 Conclusions..... 206

Chapter 7: Discussion and Future Directions

7.1 Study overview..... 207
7.2 Study limitations..... 209
7.3 Selected advances in membrane protein purification and crystallization..... 211
7.4 Future studies..... 214
7.5 Concluding remarks..... 216

References

References..... 217

Appendix

Appendix Figure 1..... 238
Appendix Figure 2..... 245

List of Figures

Figure 1.1 Diagram showing the characteristic architecture of an ABC transporter…… 1

Figure 1.2 Diagram showing the direction of P-gp transport of substrates from endogenous expression in various organs……. 5

Figure 1.3 Positron emission tomography and magnetic resonance imaging showing the distribution of a P-gp substrate in the human brain…… 7

Figure 1.4 ABC transporters expressed in the cell membrane are key mediators of multidrug resistance. …… 9

Figure 1.5 Hydrophobic vacuum cleaner model of P-gp transport…… 11

Figure 1.6 Scheme showing Darwinian and tumor-initiating models of acquired MDR…… 14

Figure 1.7 Crystal structure of Sav1866 in a closed, ADP-bound conformation…… 24

Figure 1.8 Crystal structures of mouse P-gp co-crystallized with bound inhibitors…….. 26

Figure 1.9 Codon bias displayed in the genomes of various species…….. 35

Figure 1.10 Schematic representation of the solubilization of membrane proteins by detergents…… 40

Figure 3.1 Simplified scheme of a cryptic bacterial promoter in a plasmid containing heterologous cDNA…… 76

Figure 3.2 Gel electrophoresis of *MDR1* and *mdr1a* N- and C-terminal PCR products…….. 79

Figure 3.3 Mutations and their position in *md1a* cDNA after transformation into *E. coli*…….. 80

Figure 3.4 *In silico* identification of transcriptional and translational start sites in *mdr1a* cDNA.. 84

Figure 3.5 Alignment of *mdr1a* and *MDR1* cDNA sequences at predicted transcriptional and translational start sites…… 85

Figure 3.6 GFP plasmid constructs generated to test for cryptic promoter activity of *mdr1a*……. 87

Figure 3.7 Mouse P-gp model with methionine residue at position 107 highlighted…… 88

Figure 3.8 Confocal images of *E. coli* transformed with GFP fusion plasmids……. 89

Figure 3.9 Fluorescence of *E. coli* transformed with GFP fusion plasmids……. 90

Figure 3.10 Western blot of HEK 293 cells transfected with pDEST_Human and pDEST_M107L.. 93

Figure 3.11 Rh123 efflux assay of HEK 293 cells transfected with pDEST_Human and pDEST_M107L……. 95

Figure 3.12 ATPase of human P-gp and M107L P-gp……. 96

Figure 4.1 Crystal structure of mouse P-gp with non-homologous human amino acid residues highlighted…… 101

Figure 4.2 Nomenclautre of P-gp chimeras…….. 102

Figure 4.3 P-gp chimeras generated in this study…… 103

Figure 4.4 Location of domain boundaries of mouse P-gp used in the generation of chimeras…. 105
Figure 4.5 Amino acid and nucleic acid sequence alignment of the *MDR1* and *mdr1a* cDNA NBD1-TMD2 domain boundary…….. 108
Figure 4.6 Scheme of PCR primers used to generate *mdr1a* and *MDR1* cDNA amenable for HHMM chimera generation……. 109
Figure 4.7 Plasmid maps of pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm, pTOPO_mdr1a_Nterm, and pTOPO_mdr1a_Cterm…… 111
Figure 4.8 EcoRI diagnostic digest of pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm, pTOPO_mdr1a_Nterm, and pTOPO_mdr1a_Cterm plasmids……. 113
Figure 4.9 Scheme displaying restriction enzyme digestion of pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm and pTOPO_mdr1a_Cterm plasmids…… 115
Figure 4.10 Analysis of pTOPO_HHHH_WT plasmid by agarose gel electrophoresis…… 116
Figure 4.11 Analysis of potential pTOPO_HHMM plasmid by agarose gel electrophoresis…… 117
Figure 4.12 Plasmid maps of pTOPO_HHHH_WT and pTOPO_HHMM……. 117
Figure 4.13 Silent mutations to introduce unique restriction enzyme sites at domain boundaries in *mdr1a* and *MDR1* cDNA……. 120
Figure 4.14 Plasmid maps of pTOPO_HHHH_Opti, pTOPO_M1_Opti, pTOPO_M2_Opti, pTOPO_M3_Opti and pTOPO_M4_Opti…… 123
Figure 4.15 Plasmids maps of pTOPO_HHHM, pTOPO_HMHH, and pTOPO_HMHM……. 124
Figure 4.16 Crystal contacts in the asymmetric unit of mouse P-gp…… 127
Figure 4.17 Analysis of crystal contacts between P-gp A and P-gp B with neighboring P-gp molecules…… 129
Figure 4.18 Alignment of the first 120 amino acid residues of human and mouse P-gp……. 132
Figure 5.1 Scheme of flow cytometry cell surface expression and efflux assays…… 142
Figure 5.2. Cell surface expression and efflux flow cytometry assays with and without the addition of 10 mM NaBT during transduction of HeLa cells with BacMam baculovirus……. 147
Figure 5.3 PEPG-13 western blotting of chimeric P-gp……. 151
Figure 5.4 C219 western blot of chimeric P-gp……. 152
Figure 5.5 Cell surface expression of chimeras as detected by MRK-16……. 155
Figure 5.6 Structures of the fluorescent P-gp substrates CAM, JC-1 and Rh123…… 156
Figure 5.7 Efflux of the fluorescent P-gp substrates Rh123, JC-1 and CAM by chimeric P-gp…….. 158
Figure 5.8 Inhibition of chimeras with TQR, Elacridar, CsA, and Verapamil……. 160
Figure 5.9 Time-course of Rh123 efflux by transduced HeLa cells……. 162
Figure 5.10 Sodium butyrate toxicity on non-transduced HeLa cells…… 164

Figure 5.11 Transduction of HeLa cells with chimeric P-gp is able to confer a resistant phenotype..... 165
Figure 5.12 Stimulation of P-gp ATPase activity of chimeras by verapamil and kinetic parameters..... 167
Figure 5.13 Effect of selected compounds on the ATPase activity of HHHH_WT versus HHMM P-gp..... 169
Figure 6.1 Flowchart of heterologous protein expression of P-gp chimeras in *P. pastoris* and insect cells..... 176
Figure 6.2 pPICZ_HHMM plasmid..... 178
Figure 6.3 Western blot of transformed G/500/1 and G/500/2. 182
Figure 6.4 Western blot of transformed K/1000/1 and K/1000/2. 183
Figure 6.5 Western blot of transformed X/1000/1 and X/1000/2. 183
Figure 6.6 Western blot of transformed K/2000/1 and K/2000/2..... 184
Figure 6.7 Western blot of transformed G/500/1 and K/2000/2 after large-scale expression..... 185
Figure 6.8 Colloidal blue stain and western blot of insect cell membranes..... 188
Figure 6.9 Flowchart of protein purification protocol..... 190
Figure 6.10 Western blotting and colloidal blue staining of protein samples collected during IMAC-mediated purification..... 193
Figure 6.11 Protein standard SEC chromatogram..... 194
Figure 6.12 SEC chromatograms of purified HMHH, HHMM, HMHM and HHHM chimeric P-gp. 196
Figure 6.13 Western blots of SEC fractions. 197
Figure 6.14 Colloidal blue staining of SEC fractions. 198
Figure 6.15 BN-PAGE of SEC fractions of HMHM and HHHM chimeric P-gp. 201
Figure 6.16 BN-PAGE of purified HHMM, HMHM, HMHH, and HHHM chimeric P-gp..... 202
Figure 6.17 Protein crystals of HHMM and HHHM chimeras. 203
Appendix Figure 1..... 238
Appendix Figure 2..... 245

List of Tables

Table 1.1 ABC transporters responsible for heritable disorders and their functions…… 3

Table 1.2 Sequence identity of P-gp homologs from selected mammalian species…… 17

Table 1.3 Correlation coefficients of the efflux ratio between P-gp stably expressing LLC-PK1 cell lines transfected with P-gp homologs…… 18

Table 2.1 Antibodies…….. 46

Table 2.2 Additional reagents……. 47

Table 2.3 Equipment…… 48

Table 2.4 Centrifugation…… 49

Table 2.5 Software…….. 49

Table 2.6 Plasmids…….. 50

Table 2.7 Sequencing primers for wild-type mouse and human cDNA……. 52

Table 2.8 Sequencing primers for codon optimized mouse and human cDNA…….. 53

Table 2.9 Primers used to introduce ClaI silent restriction enzyme site……. 54

Table 2.10 Media compositions used for *P. pastoris* cell culture…… 67

Table 4.1 Domain boundaries in mouse and human P-gp…….. 105

Table 4.2 *MDR1* and *mdr1a* non-cutter enzymes and their corresponding restriction enzyme sites and amino acid residues…….. 107

Table 4.3 Crystal contacts between P-gp A and P-gp B in the asymmetric unit of mouse P-gp…. 127

Table 4.4 Crystal contacts between the P-gp asymmetric unit and neighboring P-gp molecules. 130

Table 4.5 Crystal contacts retained in P-gp chimeras…….. 134

Table 5.1 Expression and efflux assays to determine reproducibility of BacMam transduction experiments…… 149

Table 5.2 CellTiter-Glo cytotoxicity assay of chimeric P-gp with the chemotherapeutic drug paclitaxel…….. 165

Table 6.1 Genotype of *P. pastoris* host strains…….. 179

Table 6.2 Transformants selected for small-scale induction of protein expression…… 180

Table 6.3 *P. pastoris* small-scale expression conditions…… 181

Table 6.4 Total purified proteins obtained for crystallization screening of chimeric P-gp…….. 199

Table 6.5 Concentration of purified P-gp before and after ultracentrifugation to precipitate protein aggregates…… 199

List of Abbreviations

[^{125}I]IAAP	[^{125}I]iodoarylazidoprazosin
2-DG	2-deoxy-d-glucose
ABC Transporter	ATP-Binding Cassette Transporter
AcMNPV	*Autographa californica multicapsid nucleopolyhedrovirus*
ADLP	adrenoleukodystrophy protein
ADP	adenosine diphosphate
AEBSF	4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride
AML	acute myelogenous leukemia
AMP-PCP	β-γ-methyleneadenoside 5'-triphosphate
AMP-PNP	adenosine-5^{1}-(β,γ-imido)tri-phosphate
ANOVA	analysis of variance
AOX	alcohol oxidase
ATP	adenosine triphosphate
AU	arbitrary units
BacMam	baculovirus-based expression in mammalian cells
BBB	blood-brain barrier
BN-PAGE	blue native polyacrylamide gel electrophoresis
bp	base pair
C-term	C-terminal
CAM	calcein AM
CAT	chloramphenicol acetyltransferase
CFTR	cystic fibrosis transmembrane conductance regulator
CMV	cytomegalovirus
Cryo-EM	Cryo-electron microscopy
CsA	cyclospoin A
CSF	cerebrospinal fluid
CTG	CellTiter-Glo
DDM	*n*-Dodecyl β-D-maltoside
DENV	dengue virus
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
E. coli	*Escherichia coli*
ECL1	extracellular loop 1
EDTA	ethylenediaminetetraacetic acid

EGTA	ethylene glycol tetraacetic acid
EM	electron microscopy
FBS	fetal bovine serum
FNLCR	Frederick National Laboratory for Cancer Research
FPLC	fast protein liquid chromatography
GFP	green fluorescent protein
HDAC	histone deacetylase inhibitor
His_6	poly-histidine tag
HRP	horseradish peroxidase
IMAC	immobilized metal affinity chromatography
IMDM	Iscove's Modified Dulbecco's Medium
KanR	kanamycin resistance
Kir channel	inward rectifier potassium channel
LB	Luria-Bertani
LDS	lithium dodecyl sulfate
MDR	multidrug resistance
MDR1	human gene encoding P-glycoprotein
mdr1a	mouse gene encoding P-glycoprotein
MDR2	Multidrug Resistance Protein 2
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
N-term	N-terminal
NaBT	sodium butyrate
NBD	nucleotide binding domain
NIH	National Institutes of Health
NTA	nitrilotriacetic acid
OD	optical density
ORF	open reading frame
P-gp	P-glycoprotein
P. pastoris	*Pichia pastoris*
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PET	positron emission tomography
P_i	inorganic phosphate
PIP_2	phosphatidylinositol 4,5-bisphosphate
PMSF	phenylmethylsulfonyl fluoride
pPolh	polyhedrin promoter

pUC Ori	bacterial origin of replication
RBS	ribosome binding site
RFU	relative fluorescence units
Rh123	rhodamine 123
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel
SEC	size-exclusion chromatography
Sf9	*Spodoptera frugiperda*
SNP	single nucleotide polymorphism
SOC Media	Super optimal growth media
TAE	Tris-acetate-EDTA
TBS	Tris buffered saline
TM	transmembrane
TMD	transmembrane domain
TQR	Tariquidar
UTR	untranslated region
V_i	sodium orthovanadate
VLCFA	very long chain fatty acids
WT	wild-type
X-ALD	X-linked adrenoleukodystrophy
ZeoR	zeocin resistance

Chapter 1: Introduction

1.1 ABC transporters

1.1.1 The ABC transporter superfamily

The ATP Binding Cassette (ABC) transporter superfamily is a highly conserved group of membrane proteins that are present in all kingdoms of life. They play a vital role in the transport of molecules across cell membranes, recognizing a diverse range of substrates using the energy from ATP hydrolysis to translocate molecules from one side of a membrane to another, irrespective of the prevailing concentration gradient. Prokaryotes express both importer and exporter ABC transporters. Eukaryotes, with the exception of plants that express both importer and exporter ABC transporters, express only exporters (1). ABC transporters are typified by a characteristic four-domain architecture consisting of two cytoplasmic nucleotide binding domains (NBDs) that bind and hydrolyze ATP and two transmembrane domains (TMDs) that recognize and transport substrates (Figure 1.1). They may be formed by the folding of a single polypeptide, or by the hetero- or homo-dimerization of two subunits. While the structure and function of NBDs are similar throughout the family consisting of numerous conserved motifs, TMDs are highly heterogeneous allowing for an astounding amount of substrate diversity (2).

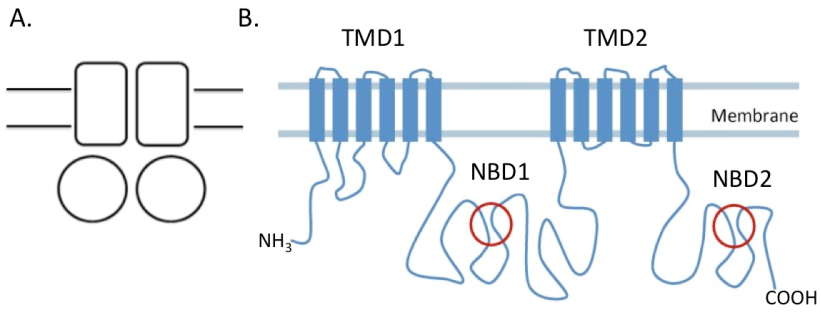


Figure 1.1 Diagram showing the characteristic architecture of an ABC transporter. A) The position of the transmembrane domains (TMD; rectangles) and the nucleotide binding domains (NBD; circles) represent the characteristic architecture of an ABC transporter. **B)** ABC transporters typically consist of two TMDs containing six α-helices and two cytosolic NBDs. Circles are shown to highlight areas of ATP binding and hydrolysis. Figure modified from Kavishe *et al.*, 2009 (3).

1.1.2 ABC transporters in human genetic disorders

There are 48 known human ABC transporter genes that have been classified into seven subfamilies (termed ABCA through ABCG) based on classification by phylogenic analysis and sequence homology of NBDs (4, 5). The diverse functions of ABC transporters in humans include antigen presentation, cholesterol and lipid transport, iron homeostasis in mitochondria, and chloride ion transport (2). These transporters play important physiological roles, and their dysfunction causes numerous diseases, such as cystic fibrosis (ABCC7, CFTR), pseudoxanthoma elasticum (ABCC6), Stargardt macular degeneration (ABCA4), Tangier disease (ABCA1), and harlequin ichthyosis (ABCA12) (Table 1.1) (6).

Table 1.1 ABC transporters responsible for heritable disorders and their functions. Treatments listed are employed to counteract symptoms; however, some strategies listed below partially resolve the disease either by off-setting the loss of function, or by correcting the defect via transplantation of organs with functional transporters (n.d. = not determined).

Gene	Function	Disorder	Primary Effect	Treatment	Ref
ABCA1 (CERP)	Lipid and cholesterol transport	Tangier disease	Blood high-density lipoprotein (HDL) lowered	None	(7, 8)
ABCA3	Surfactant storage	Respiratory distress syndrome	Lung function	Corticosteroids (anti-inflammatory), hydroxychloroquine (anti-fibrotic), lung transplant	(9)
ABCA4 (RmP, ABCR)	Retinoid flippase	Stargardt disease, cone rod dystrophy, retinitis pigmentosa	Vision	None	(10, 11)
ABCA12	Lipid transport	Harlequin type icthyosis	Thickening and scaling of strateum corneum	Anecdotal evidence for benefit from intensive skin sloughing and softening (liquid paraffin)	(12)
ABCB2/3 (Tap1/2)	Antigen presentation/delivery	Bare lymphocyte syndrome I, immune deficiencies	Compromised immune system	None, save vigilant watch for signs of infection	(13)
ABCB4 (MDR3)	Phosphatidylcholine transport	Progressive familial intrahepatic cholestasis type 3	Cholestasis	Liver transplant	(14)
ABCB7	Mitochondrial iron transport	X-linked sideroblastosis and anemia and ataxia	Anemia	Pyridoxine (stimulates heme synthesis), red blood cell transfusions, allogenic bone marrow transplant	(15, 16)
ABCB11 (BSEP)	Bile salt transport	Progressive familial intrahepatic cholestasis type 2	Cholestasis	Liver transplant	(14)
ABCC2 (MRP2)	Bile flow	Dubin-Johnson Syndrome	Conjugated hyperbilirubinemia	Usually not necessary	(17, 18)
ABCC6	n.d.	Pseudoxanthoma elasticum	Mineralization of connective tissue	None	(19, 20)
ABCC7	Chloride transport	Cystic fibrosis	Dehydration and mucus formation	Antibiotics, mucus clearance, pancreatic enzyme supplements, lung transplant, drugs that increase cell surface expression	(21, 22)
ABCC8 (SUR1)	Potassium transport	Familial persistent hyperinsulinemic hypoglycemia of infancy	Hypoglycemia	Controlled diet, near-total pancreatectomy	(23)
ABCD1 (ALDP)	Fatty acid transport	X-linked adrenoleukodystrophy	Demyelination	'Lorenzo's oil', very long chain saturated fatty acid (VLCSFA), adrenal steroid replacement therapy	(24)
ABCD3 (PMP70)	Fatty acid transport in peroxisomes	Zellweger syndrome	Absence of peroxisomes	None	(25)
ABCG5/8	Sterol/cholesterol absorption	Sitosterolemia	Hypercholesterolemia	Ezetimibe (lowers cholesterol absorption), diet	(26)

While some transporters have undergone extensive characterization, such as the ABC ion channel CFTR, other family members have unclear physiological roles and their endogenous functions and substrates remain unknown. For example, the human ABCD subfamily consists of four peroxisomal transporters, termed ABCD1-4, of which none has a clearly established function (27, 28). The best characterized transporter of this subfamily is ABCD1, also known as adrenoleukodystrophy protein (ADLP). Mutations in the *ABCD1* gene have been associated with X-linked adrenoleukodystrophy (X-ALD), a genetic disease that results in defective peroxisomal β-oxidation of fatty acids and thus the accumulation of very long chain fatty acids (VLCFAs) throughout the body (27, 29). Accumulation of VLCFAs leads to neurodegeneration, demyelination, and adrenocortical insufficiency that can progress to a vegetative state and early death (30). There is speculation that ABCD1 may have a role in the transport of VLCFAs to peroxisomes where they are metabolized. However, this hypothesis remains unproven and the role of ABCD1 in X-ALD remains unknown (27, 31). As illustrated by the ABCD subfamily, the current understanding of the role of numerous human ABC transporters is limited.

1.1.3 Excretory and protective ABC transporters

In addition to endogenous functions, many transporters function in either excretory and/or protective capacities by regulating concentration gradients of xenobiotics across biological membranes. ABC transporters are expressed at various physiological barriers, such as endothelial cells of capillaries of the blood-brain barrier (BBB), the blood-cerebrospinal fluid barrier (CSF), and the blood-testis and blood-placental barriers (Figure 1.2) (32, 33). ABC transporters localized on the apical membranes of these barriers prevent entry of exogenous molecules from the blood or lumen into organs, hence protecting the body from potentially harmful compounds. Additionally, ABC transporters expressed in the liver, kidneys, lungs, and gastrointestinal tract act to efflux exogenous compounds and their

metabolites into the bile, urine, mucus, and feces, respectively (Figure 1.2) (5, 33, 34). A consequence of these protective roles is that these transporters can impact pharmacokinetic parameters of drug absorption, distribution, metabolism, excretion and toxicity (ADMET) (4). While these functions have seemingly evolved to protect the body from potential toxins, the ability of ABC transporters to recognize and efflux a diverse variety of drugs creates a major impediment to drug efficacy (35, 36).

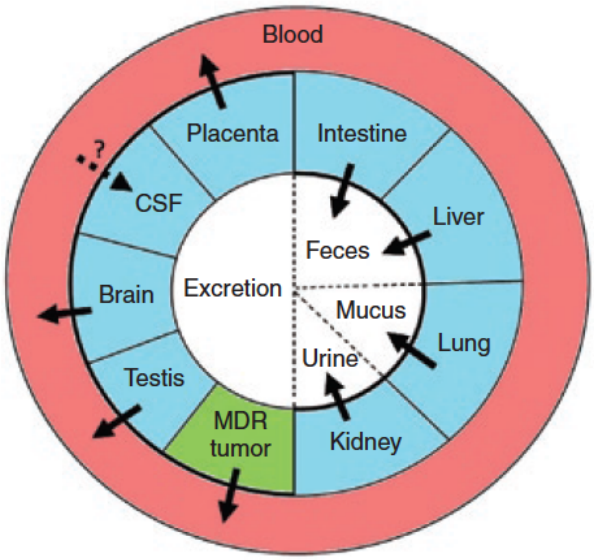


Figure 1.2 Diagram showing the direction of P-gp transport of substrates from endogenous expression in various organs. The left side of the circle represents the protective functions of P-gp, preventing entry of toxins from the blood crossing the BBB, the blood-CSF, the blood-testis and blood-placental barriers, and into MDR cancer cells. The right side of the figure exhibits how P-gp in the gastrointestinal tract, liver, lungs, and kidneys promotes the efflux of substrates into the feces, mucus, and urine. Figure reproduced with permission from Kannan *et al.*, 2009 (33).

1.1.4 ABC transporters and the blood-brain-barrier

The involvement of ABC transporters in drug efficacy is typified by expression of these transporters at the BBB. The BBB separates the central nervous system from peripheral blood circulation and is composed of specialized endothelial cells that act as a barrier of entry to molecules and ions through their connection by tight junctions (37, 38). These

capillaries also lack fenestrations normally present in vasculature, and the transport of nutrients, molecules, and ions into the brain is highly regulated according to the specific needs of neuronal tissues. The ability of a drug to reach its molecular target within cells is necessary to exert a therapeutic benefit. However, it is estimated that >98% of small molecule drugs are unable to cross the BBB (39). The BBB is able to prevent the passive diffusion of hydrophobic drugs across endothelial cells by the expression of ABC transporters on the apical membranes of endothelial cells.

The expression of ABC transporters at the BBB creates an impediment for drugs intended to be active in neuronal tissues. ABC transporters expressed at the BBB include P-glycoprotein (P-gp, MDR1, ABCB1), multidrug resistance protein 1 (MRP1, ABCC1) and breast cancer resistance protein (BCRP, ABCG2) (40). Due to the importance of ABC transporters in drug distribution, numerous *in vitro* and *in vivo* assays have been developed to assess whether compounds interact with ABC transporters. These assays include the use of inside-out membrane vesicles prepared from transporter-expressing cell lines to assess ATPase activity in the presence and absence of the drug of interest, and the use of *mdr1a*$^{-/-}$/*mdr1b*$^{-/-}$ knock-out mice (mice have two genes that encode P-gp) as compared to control mice (35, 41). These assays are routinely employed during the drug development process, as drugs that cannot reach their therapeutic target may be eliminated at early stages of development.

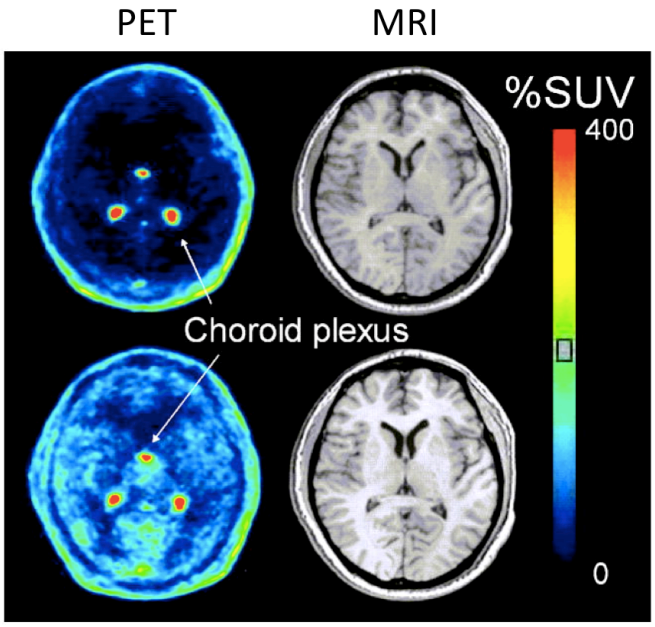


Figure 1.3 Positron emission tomography (PET) and magnetic resonance imaging (MRI) showing the distribution of a P-gp substrate in the human brain. The P-gp substrate [^{11}C]-*N-desmethyl loperamide* ([^{11}C]dLop) at baseline conditions (top row) and after coadministration with the P-gp inhibitor tariquidar (bottom row). When tariquidar is absent, the concentration of [^{11}C]dLop in the brain is low, with the exception of the choroid plexus that lies outside the BBB. Upon administration of tariquidar, P-gp is inhibited and [^{11}C]dLop passes the BBB. False color scale values represent mean concentration of radioactivity recorded by percentage standardized uptake volume (SUV). Figure reproduced with permission from Kriesl *et al.*, 2010 (42).

The involvement of ABC transporters in pharmacokinetic parameters at the BBB can be illustrated by a brief examination of the drug loperamide. Loperamide (Imodium®) is a potent opiate used for the treatment of diarrhea and has been listed as an essential medicine by the World Health Organization (43). Loperamide is an avid substrate of P-gp, and thus is unable to cross the BBB to exert the neurological effects typical of opiates (33, 44). However, when loperamide is coadministered with the P-gp inhibitor quinidine, an antiarrhythmic agent, loperamide does enter the brain, causing respiratory distress typical of opiates (33, 45). The ability of loperamide to be transported by P-gp has led to its

development as a radiotracer in positron emission tomography (PET) imaging, allowing for the quantification of P-gp function at the BBB (Figure 1.3) (42).

The case of loperamide shows how the expression of ABC transporters at the BBB can be viewed as a benefit, by keeping compounds with neurotoxic or adverse neurological effects out of the brain and thus making loperamide safe for over-the-counter use. However, this example also demonstrates the ability of ABC transporters to cause unexpected drug-drug interactions when in the presence of additional transporter modulators; a feat that is easily accomplished given the immense polyspecificity of these transporters. The drug-drug interactions mediated by P-gp also affect uptake of drugs at the level of the small intestine and excretion of drugs by the liver and kidney. To this end, the United States Food and Drug Administration (FDA) requires that all new drug be tested as substrates of P-gp, ABCG2, and ABCC1 (46).

1.1.5 ABC transporters and cancer multidrug resistance

Chemotherapy, surgery, and radiation are currently the three main cancer treatment strategies, with chemotherapy being the main strategy for the treatment of disseminated cancers such as metastatic and hematological malignancies. While the administration of chemotherapeutic drugs is often initially effective at reducing tumor bulk, many cancers will become resistant to chemotherapy throughout the course of treatment. It is estimated that resistance to chemotherapy may account for up to 90% of treatment failures for metastatic disease (47). Drug resistance can occur through various mechanisms, such as decreased drug uptake into cancer cells, increased drug efflux from cells, upregulation of detoxifying systems (cytochrome p450), or mutation of the chemotherapeutic target (48). Drug resistance due to increased efflux of drugs from cancer cells frequently results from the expression of ABC transporters (Figure 1.4) (Section 1.2.2).

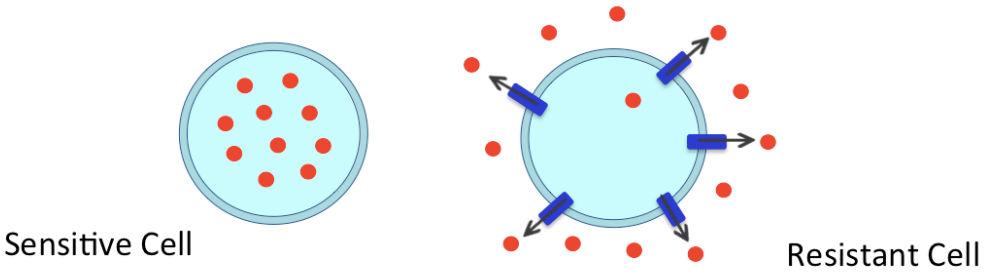


Figure 1.4 ABC transporters expressed in the cell membrane are key mediators of multidrug resistance. When ABC transporters are absent or expressed at low levels, amphipathic drugs (red circles) that diffuse through the cell membrane are able to reach intracellular targets to exert a therapeutic benefit (left). When ABC transporters are expressed at sufficient levels, drugs that are recognized by these transporters are actively effluxed from cells, typically against the prevailing concentration gradient (right) (35). The ABC transporters P-gp, ABCC1, and ABCG2 have been shown to efflux a panoply of anti-cancer drugs, thus expression of these transporters is able to induce multidrug resistance in cancer cells.

Three prominent ABC transporters that have been shown to modulate drug uptake are P-gp, ABCC1, and ABCG2 (48, 49). When present in cancer cells, these transporters can create a multidrug resistance (MDR) phenotype. MDR is characterized by resistance to drugs that differ both structurally and in their mechanism of action. A seminal paper describing MDR showed that cells selected for resistance to the natural product colchicine expressed P-gp, and were also cross-resistant to various drugs, such as vinblastine, puromycin, and daunomycin, all compounds that share no common chemical structure or mechanism (50). Given the characteristic polyspecificity of many ABC transporters, cross-resistance to a panoply of anti-cancer drugs is easily established once cells express P-gp, ABCC1, and/or ABCG2 at a level sufficient to decrease intracellular concentrations below a therapeutic threshold.

1.2 P-glycoprotein

1.2.1 Multidrug resistance

Since its identification nearly 35 years ago, P-gp has been extensively studied and characterized (50-52). P-gp recognizes a diverse range of substrates, including various therapeutically relevant natural product compounds, steroids, antiepileptic drugs, antibiotics, HIV protease inhibitors, immunosuppressive drugs, cardiac glycosides, and analgesics (53). P-gp substrates share limited characteristic traits; substrates are often hydrophobic compounds with a molecular mass between 300 to 2000 g/mol (5). The polyspecificity of P-gp can be illustrated by the numerous anti-cancer agents that have been identified as substrates, including anthracyclines (doxorubicin/daunorubicin), taxanes (paclitaxel/docetaxel), vinca alkaloids (vinblastine/vincristine), camptothecins (topotecan), and epipodophyllotoxins (etoposide, teniposide) (53). In the absence of P-gp, these hydrophobic drugs are believed to enter the cell via passive diffusion across the cell membranes as a means to reach the intracellular target. When expressed, P-gp prevents the entry of these compounds by using the energy from ATP hydrolysis to efflux them against a concentration gradient. The “hydrophobic vacuum cleaner” model has been proposed to describe the ability of P-gp to recruit hydrophobic substrates directly from the leaflet of the lipid bilayer into the drug-binding pocket (Figure 1.5) (Section 1.2.7) (54-56). This allows P-gp to intercept substrates before they access the intracellular space, ensuring that drugs cannot reach their intended targets.

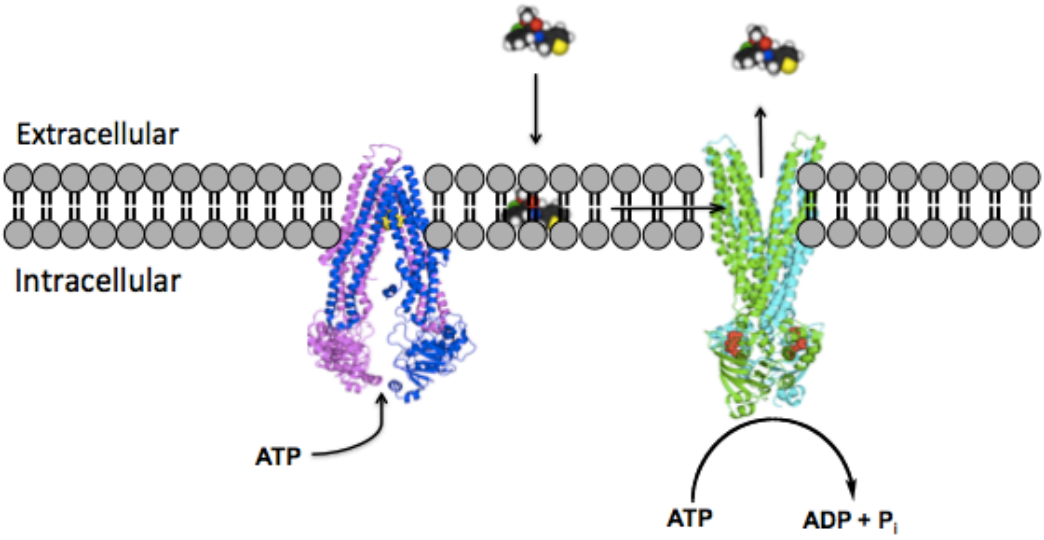


Figure 1.5 Hydrophobic vacuum cleaner model of P-gp transport. Hydrophobic compounds readily partition to the hydrophobic environment of the lipid bilayer. Instead of diffusing into the hydrophilic intracellular environment, compounds are instead directly recruited by P-gp through portals formed by TM helices into the drug-binding pocket. Once in the pocket, hydrolysis of ATP induces a conformational change that allows for the release of drug into the extracellular space. The open-inward conformation of mouse P-gp (left) (57) and the closed conformation of Sav1866 (right) (58) are shown.

1.2.2 Clinical relevance of P-glycoprotein in multidrug resistance

While MDR has been studied extensively with *in vitro* models, the clinical relevance of P-pg expression in MDR has been difficult to assess due to factors such as the lack of pre- and post-treatment tumor samples, poor specificity of anti-P-gp antibodies used in clinical studies, differential methods used to quantify expression, and infiltration of samples with non-tumor cells (5, 48). Animal models have been used to demonstrate that *de novo* expression of P-gp in cancer can confer MDR *in vivo* (59-61). For example, mouse models with inactivated of *Brca1* and *p53* genes (genetic modifications that induce "spontaneous" tumors akin to those acquired in hereditary breast cancer) showed that all tumors inevitably became resistant to the chemotherapeutic drug doxorubicin (a P-gp substrate). Microarray gene expression analysis showed that 11 of 13 doxorubicin-resistant tumors showed

increased *mdr1a* expression that ranged from 2.5- to 90-fold (59). These tumors also displayed cross-resistance to the chemotherapeutic drug docetaxel (also a P-gp substrate), a property that is a defining characteristic of MDR cells. Interestingly, modest increases in *mdr1a* upregulation were undetectable by the anti-P-gp antibody C219. This may support the finding that low levels of P-gp expression are able to induce a clinically relevant MDR phenotype, even when undetectable in assays routinely used for P-gp detection.

The most convincing clinical data correlating P-gp expression with MDR has been shown in studies of acute myelogenous leukemia (AML). Studies show that one third of patients had detectable P-gp expression at diagnosis, which increased to greater than 50% of patients at the time of relapse (48, 62). This increase in P-gp expression correlated with reduced remission rate and increased incidence of refractory disease (48, 62). However, efforts to improve treatment of AML by inhibiting P-gp have been generally unsuccessful, except for a study using cyclosporin A, which is a polyspecific ABC transporter inhibitor (63, 64). This suggests that other ABC transporters may also be involved in drug-resistant AML, and expression of other transporters has been demonstrated in post-relapse AML samples (65, 66). P-gp has also been shown to be expressed in various cancers derived from tissues that endogenously express P-gp, such as liver, colon, adrenocortical, hepatocellular, and kidney cancers (5, 48, 67, 68). However, it remains unclear to what extent P-gp influences prognosis in these cancers.

1.2.3 Intrinsic and acquired multidrug resistance

Drug resistance can be broadly classified as either intrinsic or acquired, with intrinsic drug resistance resulting from pre-existing resistance mechanisms and acquired resistance developing as a result of cancer treatment (69). As mentioned, P-gp is detectable in numerous cancers derived from tissues that endogenously express P-gp, potentially causing these cancers to be intrinsically resistant to chemotherapy. Intrinsic resistance may play a

role in certain cancers. For example, in a study investigating resistance in Ewing's sarcoma, seven of 15 tumors showed P-gp expression before chemotherapy, with an additional four tumor samples expressing P-gp in post-treatment samples (70). However, it is more commonly found that cancers that initially respond to therapy become resistant throughout the course of treatment.

How can the treatment of cancer cells with chemotherapy lead to MDR? It has been proposed that the application of cytotoxic drugs to cancer cells allows for the Darwinian selection of drug-resistant populations in which cells that are able to withstand treatment repopulate the tumor with a resistant phenotype (Figure 1.6.A). Tumor heterogeneity may partially account for this, as sub-populations of cancer cells with differing genotypes and phenotypes may give particular populations of cells a selective advantage to withstand chemotherapy. These subpopulations survive and pass the resistant phenotype to progenitor cells during tumor repopulation. Additionally, P-gp expression may be upregulated as a result of treatment with cytotoxic drugs. While the regulation of P-gp expression involves a complex interplay of cellular pathways that can be difficult to interpret, it seems that stress pathways, induced either directly by factors such as hypoxia or nutrient deprivation or indirectly by chemotherapy and/or radiation, play a role in the activation of P-gp expression (71). It has also been shown that the administration of cytotoxic drugs can induce epigenetic changes, including changes in DNA methylation and histone modification that can significantly increase P-gp expression (72, 73).

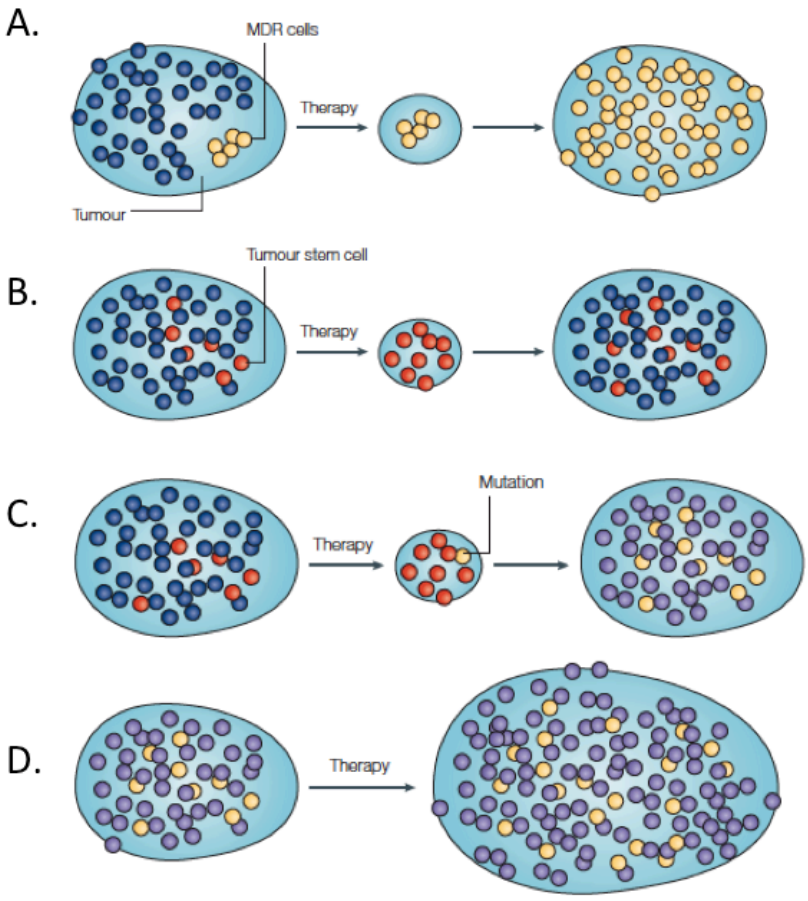


Figure 1.6 Scheme showing Darwinian and tumor-initiating models of acquired MDR. A) A Darwinian selection model of drug resistance shows that subpopulations of cells with genetic alterations that confer a selective advantage (yellow cells) survive chemotherapy and proliferate to generate a resistant tumor. **B)** A tumor-initiating model of drug resistance posits that tumor-initiating cells (red cells) that are resistant to chemotherapy, perhaps due to the intrinsic expression of P-gp and BCRP, will survive and proliferate to repopulate a heterogeneous tumor. **C)** By combining Darwinian selection and tumor-initiating models, tumor-initiating cells acquire mutations in the process of selection by chemotherapeutic drugs (which are often also mutagens; yellow cells), which can lead to mutated progenitor cells that also possess drug resistance mechanisms (purple cells). **D)** With both tumor-initiating cells and progenitor cancer cells now able to withstand chemotherapy, an intractable, resistant tumor has developed. In **A-C** the tumor bulk is greatly reduced after chemotherapy, as only subpopulations of cells are resistant. However, once tumor-initiating cells and progenitor cells have gained resistance mechanisms, as shown in **D**, chemotherapy no longer exerts any therapeutic benefit. Figure reproduced with permission from Dean *et al*., 2005 (74).

More recently, tumor-initiating cells (also termed cancer stem cells) have been implicated in drug-resistant cancers (Figure 1.6.B-D). Tumor-initiating cells can be broadly defined as cells that are able to populate cancers by their ability to both self-renew and self-differentiate (75). Early evidence supporting the presence of tumor-initiating cells comes from studies showing that only small subsets of cancer cells from leukemia and multiple myeloma are capable of tumor repopulation upon transplantation into NOD/SCID mice (76-78). The existence of tumor-initiating cells in solid tumors can has been acknowledged in medulloblastoma. Here, cells with the cell surface maker CD133 have been identified as brain tumor-initiating cells due to their ability to repopulate of the original tumor upon transplantation into immune compromised mice (79, 80). While it was the pluripotent nature of tumor-initiating cells that originally led them to be called cancer stem cells, another similarity between tumor-initiating cells and stem cells is the presence of enhanced defense mechanisms, including the expression of ABC transporters P-gp and ABCG2. In fact, a defining feature of tumor-initiating cells is their ability to efflux the fluorescent dyes Hoechst 33342 (a substrate of ABCG2) and Rhodamine 123 (a substrate of P-gp) (74, 81). These dyes normally accumulate in cells, but the presence of P-gp and ABCG2 in tumor-initiating cells leads to lower intracellular levels. Thus, a characteristic that allows for the identification of tumor-initiating cells (and other stem cells such as hematopoietic stem cells) in flow cytometry is their appearance as a side-population of cells on cell density plots due to their lower fluorescence (82). Overall, the expression of P-gp and ABCG2 in tumor-initiating cells has led to models where chemotherapy kills the majority of cancer cells while tumor-initiating cells are protected by the expression of ABC transporters. Tumor-initiating cells are thus able to survive chemotherapy and repopulate the tumor with cells that have acquired drug resistance mechanisms. The selective targeting of cytotoxic drugs to tumor-initiating cells may be necessary for the treatment of cancer.

1.2.4 P-glycoprotein nomenclature

Human, mouse, and hamster P-gp were initially isolated and cloned by three independent laboratories. This led to non-standardized and often confusing terminology referring to P-gp homologs in the literature. To address this, the Human Genome Organization Gene Nomenclature Committee held a meeting of ABC transporter investigators to adopt a standardized nomenclature for human ABC transporters, subdividing the family from ABCA to ABCG based on classification by phylogenic analysis and sequence homology (83, 84). Even after this attempt to standardize ABC nomenclature, there continues to be a lack of consensus when referring to P-gp homologs in the literature. For example, mouse P-gp is referred to as mdr3, mdr1a, and/or Abcb1a, and human P-gp as either MDR1 or ABCB1. For the purpose of this study, MDR1 refers to human P-gp from the *MDR1* gene, and mdr1a refers to mouse P-gp from the *mdr1a* gene.

1.2.5 Species differences and P-glycoprotein substrate affinities

Animal models are frequently used to assess various pharmacokinetic parameters of drugs in pre-clinical settings. *In vivo* models can be helpful to determine the biodistribution and metabolism of drugs, allowing drugs with undesirable pharmacokinetic parameters to be identified during drug development. A major limitation of the use of animal models arises from species differences in isoform composition, differential protein expression levels, and distinctive activities or substrate specificities of homologous proteins. For example, it has been shown that cytochrome p450 isoforms vary in expression levels, isoform distribution, and catalytic activity between humans, dogs, rats, mice, and monkeys, limiting the predictivity of animal models as it relates to drug metabolism by cytochrome p450 enzymes (85). The accurate translation of *in vivo* animal studies to the clinic depends on elucidation of species differences in the protein of interest. A complete understanding of species differences between P-gp, ABCG2, and ABCC1 is a prerequisite for using *in vivo* systems for

clinical modeling of pharmacokinetic parameters, as small changes in amino acid residues that line the drug-binding pocket may impact the affinity these transporters have for their numerous substrates (Table 1.2).

Table 1.2 Sequence identity of P-gp homologs from selected mammalian species. Sequence identity of P-gp from various species frequently utilized in pre-clinical *in vivo* assays compared to human P-gp. Table adapted from Kim *et al.,* 2008 (86, 87).

Species	Scientific Name	Gene	Number of Amino Acids	Sequence Identity (%)
Human	Homo spaiens	*MDR1*	1280	100
Cynomolgus Monkey	Macaca fascicularis	*MDR1*	1283	96.3
Rhesus Monkey	Macaca mulatta	*MDR1*	1283	96.2
Dog	Canis familiaris	*mdr1*	1281	90.9
Rat	Rattus norvegicus	*mdr1a* *mdr1b*	1272 1275	86.6 80.4
Mouse	Mus musculus	*mdr1a* *mdr1b*	1276 1276	87.0 80.6

Studies comparing cell lines transfected with P-gp homologs have been used to investigate variations in interspecies P-gp-mediated transport. In one study, a panel of cell lines transfected with human, monkey, canine, rat (MDR1a and MDR1b), and mouse (mdr1a and mdr1b) P-gp was established by stable transfection into the porcine kidney cell line LLC-PK1 (88). LLC-PK1 cells have been optimized for use in transport assays, as these cells form a polarized epithelial monolayer on Transwell inserts allowing the ratio of radiolabeled substrates from apical to basolateral (a to b) or basolateral to apical (b to a) to be measured (89). P-gp is expressed exclusively on the apical surface. This assay system can be used to assess parameters of substrate transport by P-gp across biological barriers by comparing levels of substrate in the basolateral or apical chambers. Twelve P-gp substrates were tested in Transwell transport assays to characterize interspecies and interisoform differences in P-

gp efflux. While human and monkey P-gp showed similar efflux ratios for all compounds tested, mouse and canine P-gp differed significantly from human P-gp, highlighting interspecies differences in P-gp transport (Table 1.3). *In vitro* cytotoxicity assays also support the presence of interspecies differences in substrate specificities, as one study found cells expressing mouse P-gp were approximately three- and 22-fold more resistant to actinomycin D and colchicine than human P-gp, respectively (90). Another study used ATPase assays using membranes isolated from insect cells infected with recombinant baculovirus expressing human, monkey, and canine P-gp screened 21 compounds for differential ATPase stimulation (86, 87). Screening revealed differential stimulation of numerous compounds, including several compounds that activated ATPase containing human and monkey P-gp but not canine P-gp. While there has been no comprehensive characterization of interspecies functional differences, numerous *in vitro* studies have shown both subtle and profound interspecies differences.

Table 1.3 Correlation coefficients of the efflux ratio between P-gp stably expressing LLC-PK1 cell lines transfected with P-gp homologs. Table adapted from Takeuchi *et al.*, 2006 (88)

	Human	Rat (MDR1a)	Rat (MDR1b)	Mouse (mdr1a)	Mouse (mdr1b)	Canine
Rat (MDR1a)	0.762	1				
Rat (MDR1b)	0.753	0.886	1			
Mouse (mdr1a)	0.609	0.786	0.838	1		
Mouse (mdr1b)	0.748	0.789	0.931	0.822	1	
Canine	0.665	0.794	0.842	0.948	0.888	1
Monkey	0.954	0.680	0.701	0.669	0.729	0.679

Interspecies differences in P-gp transport at the BBB has also been examined *in vivo* with the use of PET imaging. PET radioligands that are also substrates of P-gp are administered to different species and the distribution of the radiotracer in the brain is assessed. In one study, the PET radiotracer [^{11}C](S)-(2-methoxy-5-(5-trifluoromethyltetrazol-

1-yl)-phenylmethylamino)-2(S)-phenylpiperidine (GR205171) was administered to rats, guinea pigs, monkeys and humans and the radiotracer [^{18}F]altanserin was administered to rats, mini-pigs and humans in the presence and absence of the P-gp inhibitor cyclosporin A (91). Brain concentration and brain-to-plasma concentration ratios were compared and revealed interspecies differences to varying degrees. For example, the brain-to-plasma ratio of [^{11}C]GR205171 was approximately nine-fold higher in humans than in rats (91).

Lastly, it should be noted that the testing of candidate drugs in animal models is time consuming and costly, thus occurring late in the pre-clinical drug development pipeline. The availability of high-resolution structures to screen candidate drugs *in silico* for predicted ABC transporter interactions could save time and money, allowing for drugs with undesirable pharmacokinetic parameters to be identified early during drug screening. Currently, the crystal structure of mouse P-gp has not been solved to a resolution that allows *in silico* screening of drugs. Even if the resolution of mouse P-gp can be improved, elucidation of functional differences between mouse and human homologs would be required in order to allow for *in silico* modeling approaches of mouse P-gp to be translatable to the human homolog.

1.2.6 Terminology: substrates, inhibitors and competitive inhibitors

There are various terms used to categorize the molecular interactions between P-gp and compounds. Broadly, a modulator is any compound that alters the function of P-gp. A transport substrate can be defined as any molecule that is transported by P-gp. In contrast, a catalytic substrate can be defined as a molecule that can be bound and hydrolyzed by the NBDs of P-gp (i.e. ATP). P-gp-mediated efflux of the transport substrate is associated with ATP hydrolysis. Additionally, various molecules have shown to be inhibitors or P-gp. In this thesis, an inhibitor is defined as a molecule that interacts with P-gp to restrict either its catalytic or transport function. Common examples of P-gp inhibitors are the compounds

tariquidar (TQR, XR9576), cyclosporin A (CsA), elacridar, and zosuquidar, which have been shown to inhibit P-gp in various *in vitro* and *in vivo* assays (92-95). A compound may also be referred to as a competitive inhibitor, a term that is applicable when there is more than one P-gp substrate competing for transport. This scenario can produce various outcomes depending on the relative concentrations of each substrate in the system and the affinity of each substrate for transport by P-gp.

Terminology classifying molecules as either substrates or inhibitors is complicated by the fact that numerous compounds blur the lines between P-gp substrates and inhibitors. For example, the compound verapamil exhibits a dose-effect relationship with P-gp, acting as a P-gp substrate at low concentrations and as an inhibitor at high concentrations (96). Various *in vitro* assays have been developed to assess the molecular interactions between compounds and P-gp, including ATPase assays, flow cytometry assays, Transwell assays that assess the transport of compounds across a polarized cell monolayer, and photoaffinity labeling that determines interactions of test compounds with P-gp at the drug-binding site(s) (96-101).

1.2.7 P-glycoprotein mechanism of transport

Two models have been proposed for P-gp-mediated transport of substrates: the hydrophobic vacuum cleaner model and the lipid flippase model (102). Both models are based upon the finding that most P-gp substrates are hydrophobic and readily partition to the lipid bilayer. The hydrophobic vacuum cleaner model posits that drugs are recruited directly from the bilayer into the drug-binding pocket of P-gp (Figure 1.5). This mechanism of transport seems to be supported by recent structural data. For example, the structure of mouse P-gp displays portals formed by TM helices that would allow for drug recruitment from the bilayer into the internal cavity of P-gp (102, 103). In contrast, the lipid flippase model is based on observations that P-gp has flippase activity, and it has been proposed that

this activity could also be used to move drugs from the inner to outer leaflet of the lipid bilayer (104-106). Here, the flipping of drugs from the inner leaflet to the outer leaflet maintains a high concentration of drugs in the outer leaflet. Once this concentration gradient is achieved, drugs partition from the bilayer to the aqueous extracellular environment (105, 107). Current experimental data does not clearly support either the hydrophobic vacuum cleaner or flippase model, and thus the mechanism of P-gp-mediated drug transport remains the subject of discussion in the field (102).

While biochemical studies remain unable to determine the mechanism of P-gp-mediated transport, increasing structural evidence (Section 1.3.2 and Section 1.3.3) supports an "alternating access model," where P-gp adopts at least two distinct conformations: an inward facing conformation with the internal cavity of P-gp open to the cytoplasm (inverted V-shape), and an outward facing conformation with TMDs open to the extracellular space (V-shape) (108). P-gp adopts the inward conformation at the beginning of the transport cycle. In this conformation, the cytosolic NBDs are able to bind ATP. Substrates may access the pocket either by entry from the cytoplasm or directly through portals in the TMDs that allow for the entry of substrates directly from the membrane. P-gp's internal cavity is lined with hydrophobic residues presenting a high-affinity environment conducive to the binding of hydrophobic molecules. The "loading" of P-gp with ATP and substrate, then hydrolysis of ATP initiates the switch from an inward- to an outward-facing structure. The conformational change from inward to outward switches the drug-binding pocket from a high- to low-affinity state that allows for the release of substrates into the extracellular space (109, 110). The translocation process is divided into at least four general stages: I) the binding of ATP and substrates to NBDs and the drug-binding pocket, respectively, II) the "switch" of P-gp from an inward to outward conformation that results in substrate release, III) ATP hydrolysis, and IV) resetting of P-gp back to the inward facing conformation (111). The stoichiometry of ATP hydrolysis and molecules of substrates transported is a subject of discussion in the field.

The resetting of P-gp back to an open, inward conformation is thought to be achieved by the hydrolysis of a second ATP molecule (112).

1.3 P-glycoprotein structure

1.3.1 P-glycoprotein structure overview

P-gp consists of four domains; two NBDs involved in ATP binding and hydrolysis and two TMDs responsible for substrate recognition and efflux (Figure 1.1). TMDs are embedded in the lipid bilayer with NBDs located in the cytosol, thus having access to cellular stores of ATP that can fuel the export of substrates. Across the superfamily, NBDs are highly homologous, sharing several conserved motifs, including the Walker A (P-loop, phosphate-binding loop) motif, Walker B motif, ABC signature motif, and the H-loop (2, 102, 113, 114). In contrast, TMDs are highly heterogeneous, accounting for the diverse substrate specificities across the family. Human P-gp is a single 1280 amino acid polypeptide that is transcribed and translated in the order: (N-term) TMD1-NBD1-TMD2-NBD2 (C-term). P-gp folds to form two almost symmetrical halves each consisting of a TMD containing six α-helices and a NBD. The halves are connected via a flexible linker polypeptide approximately 75 amino acids in length, connecting NBD1 to TMD2 (115).

Studies using electron microscopy (EM) and cryo-EM have yielded low-resolution images of mouse and human P-gp (116-119). Early studies using EM showed that P-gp adopts at least two major conformations, with various intermediate conformations. Even at a low-resolution, it can be seen that P-gp undergoes a large-scale conformational change upon nucleotide binding, supporting the switch between inward- and outward-facing conformations in proposed translocation models (119, 120). While technical limitations have previously inhibited the elucidation of high-resolution structures using cryo-EM, recent advances have resulted in the attainment of sub-nanometer structures of several proteins, including the ABC exporter TmrAB (121, 122). Additionally, cryo-EM has recently been

utilized to probe for conformational changes in P-gp and MsbA in the presence and absence of nucleotide and substrate (123). While advances have greatly increased the utility of cryo-EM for structural investigations, currently, protein crystallography remains the gold standard for attainment of atomic resolution structures (124).

Despite attempts from various laboratories, a high-resolution crystal structure of human P-gp has remained elusive. Crystal structures of several homologs have been published including *Caenorhabditis elegans* P-gp and mouse P-gp with 46% and 87% identity to human P-gp, respectively (57, 103, 125). However, these structures have been reported in the absence of nucleotides, which, to some investigators, has cast doubt on their physiological relevance. Two additional crystal structures, the bacterial exporter Sav1866 and human ABCB10, are also relevant to understanding the structure and function of human P-gp, as Sav1866 is the only ABC exporter to be crystallized in a nucleotide-bound state and ABCB10 is the only human ABC transporter crystal structure reported to date (58, 126). The structures and scientific impact of Sav1866, mouse P-gp, *C. elegans* P-gp and ABCB10 are discussed below.

1.3.2 *Staphylococcus aureus* Sav1866

The structure of the *Staphylococcus aureus* Sav1866 was the first high-resolution crystal structure of an ABC exporter, solved to a 3.0 Å resolution by Dawson and Locher (Figure 1.7) (58). Sav1866 is expressed as a half transporter that dimerizes to translocate substrates. Functionally, Sav1886 shows similarity to members of the human ABCB subfamily, and is able to transport doxorubicin, vinblastine, and Hoechst 33342 (all substrates of human P-gp) (58). The N-terminal half of the Sav1866 dimer is 28% identical and 45% similar to human P-gp, while the C-terminal half is 28% identical and 48% similar (127). Sav1866 was crystallized in the presence of bound ADP, and thus is thought to reflect the nucleotide-bound closed conformation, with the central internal cavity facing outward.

In this conformation, the internal cavity was found to contain relatively hydrophilic amino acids, resulting in a low-affinity state for hydrophobic substrates leading to extrusion (58, 114). In this conformation, it seems plausible that substrates are shielded from the cytoplasm, allowing for unidirectional transport into the extracellular space.

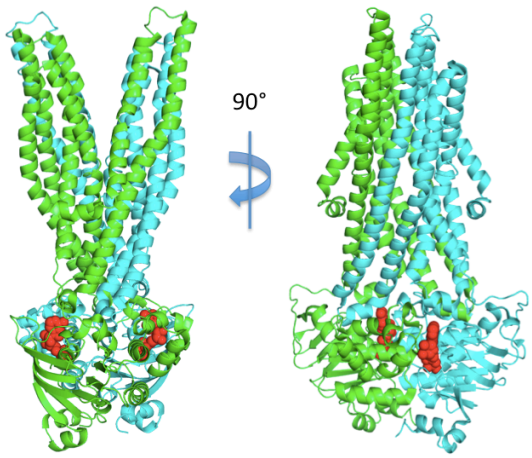


Figure 1.7 Crystal structure of Sav1866 in a closed, ADP-bound conformation. Helices from the N-terminal (green) and the C-terminal (cyan) halves form two distinct bundles that open to the extracellular space, promoting the extrusion of substrates. Bound ADP is shown in red. Figures generated using PyMOL from Protein Data Bank identification 2HYD.

Due to limited understanding of the events in the translocation cycle, it is unclear if ADP-bound P-gp should adopt an inward- or outward-facing conformation. Since the hydrolysis of ATP to ADP may be associated with the re-setting of P-gp back to an inward-facing state, concerns were raised about the physiological relevance of the Sav1866 structure. To address this, a follow-up study crystallized Sav1866 with the non-hydrolysable analogue of ATP, adenosine-5^{1}-(β,γ-imido)tri-phosphate (AMP-PNP) and found no significant differences in structure between ADP- and AMP-PNP-bound Sav1866 (128). The reported structures of Sav1866 agree with biochemical studies of ABC exporters and are thought to represent a physiologically relevant state. An additional, important contribution of Sav1866 to the field was that its publication cleared the confusion originating from prior published

work on MsbA, another bacterial exporter, with respect to the organization of transmembrane helices (129, 130). The reported structures of MsbA were retracted after the publication of Sav1866.

1.3.3 Mouse P-glycoprotein (*mdr1a*)

The first mammalian ABC transporter structure, reported by Aller *et al.* in 2009 solved the structure of mouse P-gp to a resolution of 3.8 Å (103). However, the later publication of *C. elegans* P-gp by Jin *et al.* (Section 1.3.4) showed significant deviations in structure between P-gp homologs, raising concerns about errors in model building of mouse P-gp (125). Two papers were published to address the inaccuracies of the original mouse P-gp structure; a paper re-analyzing the original mouse P-gp dataset leading to refined structures (57), and another paper that generated new crystals of mouse P-gp for analysis, including P-gp bound to a nanobody inhibitor (136).

To address issues raised regarding inaccurate model building, Li *et al.* used single-wavelength anomalous dispersion signals from bound mercury atoms to generate a new density modified electron density map from the original crystal dataset for model rebuilding (57). The new map required *de novo* model building for several TM helices and registry correction for six TM helices. The profound structural differences between mouse P-gp structures reported can be seen in the change in Ramachandran plots from 57% to 95% of residues in favored regions. Additionally, R_{work}, a measure of how well the structure matches the dataset, decreased from 0.306 to 0.259 and R_{free}, a parameter used to describe the quality of the model, decreased from 0.347 to 0.303 between the original and refined P-gp structures, respectively, indicating a more accurate model. The refined structures report up to an 8 Å shift of TM domains when compared with the original structure. For example, TM4, a helix that is thought to form a component of the drug entry portal, shifted >6 Å in the refined structure. Overall, the refined mouse P-gp structure shows significant improvement

in modeling and registry shifts, more accurately matching biochemical data and the higher resolution structure of *C. elegans* P-gp (Figure 1.8).

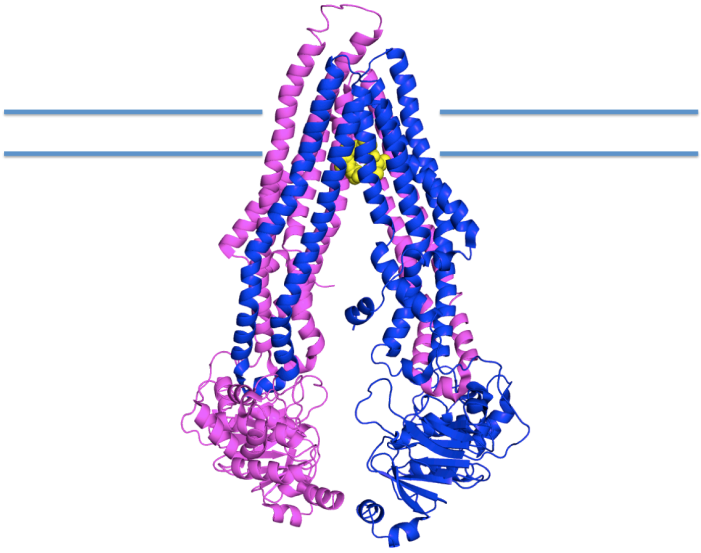


Figure 1.8 Crystal structure of mouse P-gp Mouse P-gp is composed of two symmetrical domains, N-terminal (magenta) and C-terminal (blue), each consisting of a TMD and NBD. Overall, all reported structures show that the TMDs of mouse P-gp form two separate bundles of helices, one bundle consisting of TM helices 1, 2, 3, 6, 10, and 11, and a second bundle consisting of TM helices 4, 5, 7, 8, 9, and 12. P-gp was crystallized in the open-inward conformation, with the NBDs in a conformation able to bind ATP and TMDs open to drug entry from both the cytoplasm and the lipid bilayer. Blue lines represent the putative position of the lipid bilayer. Figures were generated using PyMOL from Protein Data Bank identification 4M1M.

Three additional structures of mouse P-gp were published by Ward *et al.* (136). All structures reported are in the open, inward facing conformation, two of which are in the absence of bound inhibitor and a third structure co-crystallized with a bound nanobody, a single domain antibody fragment that retains the ability to selectively bind a specific antigen, that is able to inhibit ATP hydrolysis. Crystals obtained were validated by the introduction of 17 single-site cysteine substitution mutants which were labeled with ethyl mercury for enhanced model determination. These structures show a wide ~29-36 Å separation

separation between NBDs. Additionally, this study uses a nanobody to selectively inhibit P-gp ATP hydrolysis. Due to the ability for nanobodies to potentially be used as therapeutic agents, their ability to specifically inhibit the ATP hydrolysis of P-gp may provide a starting point for the development of therapeutically relevant nanobody P-gp inhibitors. However, logistical aspects of nanobody delivery would need to be investigated before therapeutic use, as they do not diffuse through the cell membrane and thus cannot gain access to NBDs.

While undoubtedly valuable to the ABC transporter field, the publication of these structures arguably raised more questions than they answered (54). For example, mouse P-gp was crystallized in a nucleotide-free environment; a situation that would be physiologically unlikely given that intracellular concentrations of ATP are 3-5 mM, a range that exceeds the affinity of ATP for P-gp (K_{mATP} = 0.3-1.0 mM) (54, 132, 133). Additionally, NBDs are separated by ~30 Å, a surprisingly far distance considering NBDs are thought to communicate directly (108, 133, 134). However, a study using molecular dynamics simulations and double electron-electron resonance spectroscopy has more recently shown that P-gp NBDs can, in fact, accommodate a high degree of NBD separation (>20 Å) in the presence of lipid bilayers both with and without nucleotide present (135). The physiological relevance of the mouse P-gp structure is highly debated, as the structure may be an artifact of crystallization due to consequences of crystal packing forces. Lastly, the low resolutions of the reported structures are insufficient to identify amino acid locations and therefore cannot answer questions regarding P-gp polyspecificity or mechanism of transport.

1.3.4 *Caenorhabditis elegans* P-glycoprotein

C. elegans P-gp is 46% identical to human P-gp. However, it is estimated that the drug-binding pocket shares only ~20% identity with human P-gp (137). To test the functional similarity between human and *C. elegans* P-gp, cytotoxicity assays on *Spodoptera frugiperda* (Sf9) insect cells infected with baculovirus containing *C. elegans* P-gp were conducted. These

assays showed the ability of *C. elegans* P-gp-expressing cells to efflux the drugs paclitaxel and actinomycin D, both known human P-gp substrates. Additionally, over 30 substrates of human P-gp were tested for *C. elegans* P-gp ATPase stimulation, and five substrates emerged as shared substrates, indicating a partial overlap in substrate specificity that would be expected between P-gp homologs. While it appears that there is a degree of substrate overlap between *C. elegans* and human P-gp, the endogenous function of *C. elegans* P-gp is unknown and the ability of the *C. elegans* structures to elucidate the mechanism of human P-gp is debated (137).

The *C. elegans* structure revealed important similarities when compared to the mouse P-gp structure reported by Ward e*t al.* Mouse and C. *elegans* P-gp were both resolved to a low-resolution (3.8 Å for mouse and 3.4 Å for *C. elegans*) in a nucleotide-free state in the open inward conformation with a large separation between NBDs (125). This supports the idea that in the absence of bound nucleotide, P-gp adopts an open, inward-facing conformation, which would allow for substrate binding recruitment either from the cytoplasm or directly from the membrane into P-gp through portals comprised of TM helices. The structural similarities between the reported *C. elegans* and mouse P-gp structures can either be interpreted as a confirmation of the pre-transport state or they may simply reinforce the hypothesis that crystallization in a nucleotide-free environment creates crystallization artifacts.

An important contribution of *C. elegans* P-gp to the field was that it highlighted errors in model building of the first published structure of mouse P-gp by Aller *et al*. For example, upon superimposition of mouse and *C. elegans* structures, TM helices 1, 2, 6, 7, 8, and 11 have comparable structures. However, TM helices 3, 4, and 5 show differences in structure due to register shifts in model building (TM helices 9, 10, and 12 were non-comparable in structure). These register differences lead to differences in amino acids that line the drug-binding pocket and are more compatible with biochemical data than the

reported mouse structure. This prompted the publication of refined mouse P-gp structures based on improved model building (57).

1.3.5 Structure of human ABCB10

The structure of human ABCB10 is unique in two distinct ways; it is both the first human ABC transporter to be crystallized and it was crystallized in nucleotide-free and nucleotide-bound states (126). Unlike the aforementioned P-gp homologs, ABCB10 is not a full-length cell surface drug transporter but rather is found in the inner membrane of mitochondria as a homo-dimer. While the function and endogenous substrate(s) of ABCB10 are currently unknown, ABCB10 has been shown to be crucial in erythropoiesis and protection against oxidative stress (126, 138, 139). Embryonic lethality is seen in *ABCB10*$^{-/-}$ knockout mice, with mice anemic at day 10.5 of gestation and dead at day 12.5 of gestation (138).

Shintre *et al*. reported four structures of ABCB10: one nucleotide-free or "apo" structure and three nucleotide-bound structures (126). ABCB10 crystallized in the absence of bound nucleotide adopts an open, inward conformation, similar to the previously described mouse and *C. elegans* P-gp structures. However, in contrast to mouse and *C. elegans* P-gp, ABCB10 shows a shorter distance between NBDs. When the distance between the γ-phosphate of the nucleotide and the C-α of the first glycine in the ABC signature motif was measured (a distance that would be required for NBDs to move in order for ATP hydrolysis to occur), a 17 Å versus 41 Å separation of NBDs was shown for ABCB10 and *C. elegans* P-gp, respectively.

In contrast to the subtle structural differences between ABCB10 and previously reported ABC transporters in the nucleotide free state, significant structural differences were observed between nucleotide-bound ABCB10 structures and the previously reported ADP- and AMP-PNP bound Sav1866 structures. ABCB10 was co-crystallized with either the

non-hydrolysable ATP analogue AMP-PNP or β-γ-methyleneadenoside 5'-triphosphate (AMP-PCP). While crystallization of Sav1866 with AMP-PNP is able to switch the protein from an inward to outward conformation, this is not observed with ABCB10, which remains in the open, inward conformation. It is unclear why the AMP-PNP is able to induce a closed conformation in Sav1886 but not ABCB10. However, the authors suggest that the presence of substrate may be required to induce the vast conformational change from inward- to outward-facing for ABCB10. The authors proposed a model for the ABCB10 translocation cycle, where either nucleotide or substrate can bind ABCB10 initially, but the switch from inward to outward conformation will not occur unless both components are present. However, this hypothesis has yet to be validated with experimental data.

1.4 Expression, purification, and crystallization of membrane proteins

1.4.1 Expression systems for large-scale purification of membrane proteins

Before crystallization can be attempted, expression and large-scale purification of the desired protein must occur. Many proteins, including membrane proteins, are not often expressed at high levels in their native environments and thus cannot be obtained in sufficient quantities for structural investigations. This conundrum has led to the development of numerous heterologous protein expression systems, all with a common goal of high-level expression of foreign proteins. Various systems are used for the expression of membrane proteins, including expression in the bacteria *Escherichia coli* and *Lactococcus lactis*, the yeasts *Pichia pastoris* and *Saccharomyces cerevisiae*, insect cells, and cultured mammalian cells (140). The choice of the best expression system is dependent on the individual protein to be expressed. For example, while bacterial systems can generate large amounts of protein both quickly and cheaply, if the protein of interest requires eukaryotic systems of post-translational modification, bacterial expression can be eliminated as an option as they do not accomplish these. Also, various foreign proteins can be toxic to the

host organism, as is sometimes observed with the expression of various membrane proteins in bacteria. Expressed proteins need to be trafficked to and folded appropriately in the membrane, factors that can be affected by the expression system. Additionally, membrane lipid composition can affect the stability of the expressed protein, with less stable proteins having a lower likelihood of producing diffraction quality crystals (141). The choice of expression system can often require the testing of multiple hosts, a process that can be time consuming and costly.

Recently, both individual laboratories and structural genomics consortia have adopted a high-throughput approach to strategically produce large quantities of a membrane protein. In contrast to labor-intensive expression optimization processes for a single protein, various homologs of a protein are screened for high expression levels in a small number of expression conditions. This strategy is based upon the fact that homologous proteins often have the same function, but differ, potentially significantly, in their expression levels and stability in detergents (142, 143). These differences are attributed to small changes in amino acid residues between homologs, that could either be favorable or detrimental to their expression in a host system. This strategy has been successfully used to crystallize various membrane proteins, including the bacterial ABC transporter Sav1866 and mouse P-gp (58, 144, 145).

As the production of large quantities of high quality protein is a prerequisite for structural studies, various systems for the expression of P-gp homologs have been tested and optimized. To take advantage of the high protein production capacity and cost-effective nature of bacterial expression, early studies by Bibi *et al.* attempted to express mouse P-gp in *E. coli* (146, 147). They note that many attempts to express P-gp in *E.coli* were unsuccessful, and a variety of different strains were screened before successful expression was attained. Many attempts to express P-gp in *E.coli* may have failed due to lethal or "toxic" effects of P-gp expression in *E.coli*, a property that is not uncommon for eukaryotic

membrane proteins expressed in bacteria (148). These observations were perhaps foretelling, as others were not able to reproduce expression of P-gp in bacterial systems, and the focus shifted from bacterial to yeast and insect cell expression systems for ABC transporter heterologous expression (148). Of the mammalian ABC exporter structures discussed, both mouse and *C. elegans* P-gp were expressed in *P. pastoris* and ABCB10 was expressed in insect cells (58, 103, 125, 126). Salient features of both yeast and insect cell expression will be discussed below.

1.4.2 Heterologous protein expression in *P. pastoris*

Since its discovery as a methylotrophic yeast strain over 40 years ago, *P. pastoris* has become a successful system for the expression of heterologous proteins (149). Heterologous protein expression in *P. pastoris* contains numerous advantages. It is able to produce foreign proteins at high levels and is able to perform eukaryotic post-translational modification, a necessity for the correct expression of some eukaryotic proteins. Additionally, *P. pastoris* is highly similar to the well characterized yeast species *S. cerevisiae* and numerous molecular biology techniques have been readily adapted from one yeast species to another, such as electroporation transformation methods. Lastly, commercially available kits have made use of *P. pastoris* as a heterologous system standard practice (150, 151). Proteins hailing from most kingdoms have been expressed in *P. pastoris*, including various bacterial, fungal, protist, plant, virus, vertebrate and non-vertebrate proteins (150, 151).

P. pastoris is able to utilize methanol as its sole source of carbon and energy (150). Methanol utilization requires the expression of the enzyme alcohol oxidase (AOX), encoded by the genes *AOX1* and *AOX2*. AOX is expressed at high levels (up to approximately 30% of total soluble protein) when *P. pastoris* is grown exclusively on methanol as the carbon source; in the absence of methanol these enzymes are undetectable (150-152). The *AOX1* gene is responsible for almost all of the AOX that is expressed in wild-type yeast. However,

some yeast strains have been genetically modified by deleting the *AOX1* gene so that the yeast relies on the weaker *AOX2* promoter. This creates a slow growth phenotype (termed Mut^S) that is favorable for the expression of certain foreign proteins.

Expression of foreign proteins in *P. pastoris* begins with the construction of a suitable expression vector. *P. pastoris* expression vectors share a handful of common features. Vectors must contain a bacterial origin of replication for the construction and amplification of vectors in *E.coli*. Additionally, the vector must contain fragments of *P. pastoris* DNA to allow for homologous recombination with the host genome. This is typically achieved by the incorporation of the 5' *AOX1* promoter in the expression vector flanking the gene to be expressed. Therefore, when the gene of interest is incorporated into the *P. pastoris* genome by homologous recombination, it occurs so that the gene of interest is under the control of the *AOX1* promoter. Expression of the foreign protein is achieved only when *P. pastoris* is grown in the presence of methanol, allowing for tight regulation of gene expression. Lastly, all expression vectors must contain genes encoding selection markers for growth in both *E.coli* and *P. pastoris*. The expression of mouse P-gp in *P. pastoris* for structural investigation has been well established (153, 154) and protein purified from *P. pastoris* was used in the crystallization of both mouse and *C. elegans* P-gp (103, 125, 136).

1.4 3 Codon optimization of P-glycoprotein cDNA

Codon degeneracy is a defining feature of the genetic code, characterized by the ability of multiple codons to encode the same amino acid. Redundancy is a product of the fact that 61 codons (plus three stop codons) correspond to only 20 amino acids, creating the potential for one (methionine and tryptophan) to six (leucine, arginine, serine) codons to encode for a single amino acid. Multiple codons that code for the same amino acid are referred to as synonymous codons while codons that code for different amino acids are referred to as non-synonymous codons. It is traditionally accepted that synonymous codon

changes do not affect the structure or function of proteins, while non-synonymous codon changes lead to amino acid changes that have variable effects on final protein structure and function[1] (157, 158). The potentially profound effect of non-synonymous codon changes is exemplified in Sickle cell disease. A single point mutation creates a non-synonymous amino acid change in the gene encoding the β-globin chain of hemoglobin. This is the sole factor leading to rigid, abnormally shaped red blood cells that cause various physiological phenotypes characteristic of Sickle cell anemia (159).

Codon bias is the phenomenon whereby different species display preferred usage of certain codons for a particular amino acid, thus leading to codons that are more and less frequent in certain genomes (Figure 1.9) (160, 161). Correspondingly, this leads to differences in the abundance of cognate tRNA molecules, with more tRNA molecules available for more common codons and less tRNA molecules for rare codons (162). The usage of particular codons in a gene can dictate the translational speed, as the use of more common codons can increase the translational rate as the translational machinery has an abundant pool of cognate tRNA molecules available (163). In contrast, when less abundant codons are used, the rate of translation slows and rare cognate tRNAs must be recruited to the translational complex. The use of common and rare codons to control translational speed allows folding of the nascent protein to occur at an optimal speed (163).

[1] While it has been traditionally accepted that amino acid composition is the sole determinant of protein structure and function, it is emerging that synonymous single nucleotide polymorphisms (SNPs) are not inconsequential, and may play a role in controlling both protein structure and function. For example, a synonymous 3435 C>T SNP in the *MDR1* gene from a common to rare codon has been shown to correlate with higher digoxin plasma levels suggesting either altered P-gp protein levels or altered function (155, 156). The substitution from common to rare codons may lead to varying degrees of ribosomal pausing or stalling, as rare cognate tRNAs need to become available in proximity to the ribosomal complex for continued translation. Changes in translational speed are thought to account for subtle differences in protein folding, which begins immediately after amino acids leave the ribosomal complex. Kimchi-Sarfaty *et al*. have shown that a common haplotype associated with the C3435T synonymous polymorphism results in altered folding of P-gp, possibly due to a change in translational speed resulting from altered codon usage (156). Prevalence studies have revealed that the prevalence of the 3435 C>T mutation varies significantly according to ethnicity, suggesting evolutionary discrimination based on function (155).

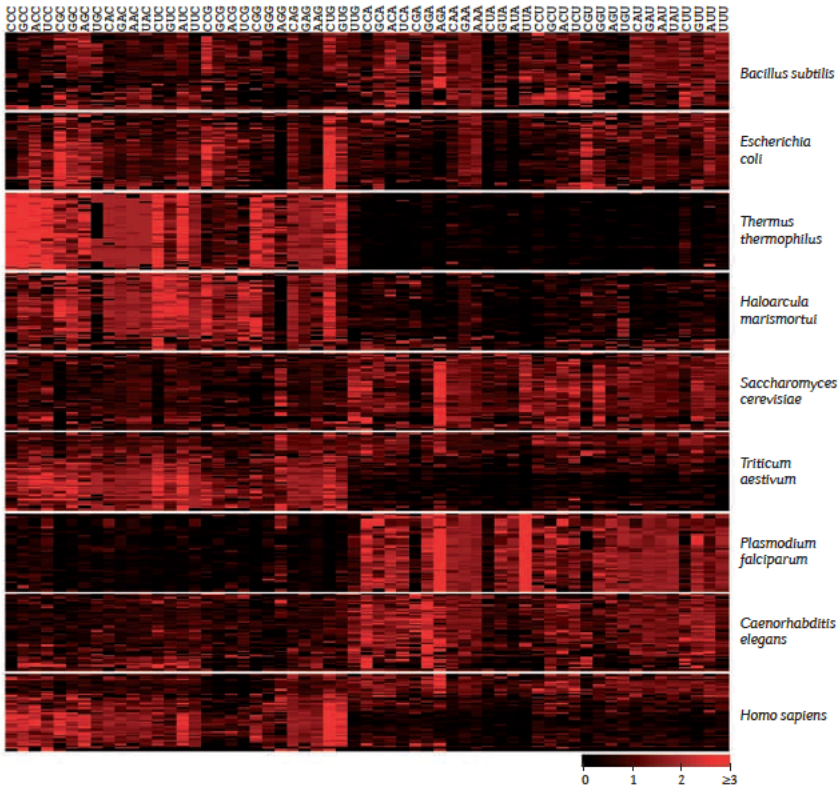


Figure 1.9 Codon bias displayed in the genomes of various species. Codon usage between the genomes of nine different organisms is shown for 50 randomly selected genes. Rows represent genes and columns represent codons (methionine, tryptophan and stop codons were omitted from analysis). Relative synonymous codon usage ranges from one, when there is no bias between codons, to six, when a single codon is used in a six-codon family. A value of zero indicates that the codon was not present in any of the genes selected for analysis. Some organisms, such as *Thermus thermophilus,* show an extreme degree of codon bias, where a large number of codons are absent from the selected genes. In contrast, some organisms exhibit a high degree of codon heterogeneity, such as seen with the genome of *Homo sapiens*. Figure reproduced with permission from Plotkin *et al.*, 2011 (160).

Codon bias can have profound effects on heterologous protein expression, as translational speed and efficiency can be hampered by cognate tRNAs that are present at low levels within the host (164). Strategies to increase protein production of foreign genes in heterologous systems include “optimizing” codon usage according to the codon bias preferred by the intended host organism. The optimization process replaces rare codons with common codons according to the codon bias of the host organism. Optimizing a foreign gene for codon usage in the expression system of choice has been shown to increase protein

expression. Codon optimization of foreign genes for heterologous expression in *P. pastoris* has been reported successfully with genes from various species, including mouse P-gp (165-170). Bai *et al.* reported a codon-optimized cDNA for mouse P-gp that increased protein expression in *P. pastoris* greater than three-fold (166). Although it is possible that increasing protein translation by codon optimization might lead to inaccuracies in nascent protein folding or processing, it can be an effective strategy for increase yield of expressed proteins. For example, codon-optimized mouse P-gp was used for the expression, purification, and crystallization of second-generation mouse P-gp structures (136). Codon optimization for P-gp is therefore an effective method for increasing protein expression in *P. pastoris.*

1.4.4 Heterologous protein expression in insect cells

Insect cells have been used as a system for heterologous protein expression, as they can be infected with recombinant baculoviruses. Baculovirus-mediated methods of heterologous protein expression were first established over 30 years ago and continue to be commonly used in many academic and industrial settings (171-173). Baculoviridae is a large family of rod-shaped, double-stranded DNA viruses. Wild-type baculoviruses are only capable of infecting arthropod hosts, and thus can be safely used in laboratory manipulations due to their inability to replicate in mammalian cells. The most commonly exploited baculovirus for the production of foreign proteins is *Autographa californica multicapsid nucleopolyhedrovirus* (AcMNPV), as its genome has been well characterized (174). Heterologous protein expression by baculoviruses in insect cells was based on early visual observations that infected cells accumulate occlusion bodies in their nuclei at a density that is visible during the very-late phase of viral infection. Infected cells accumulate from two- to three-dozen occlusion bodies, which are comprised of progeny virons in a paracrystalline array of polyhedrin, a virus encoded protein. Polyhedrin is produced in vast amounts, accounting for up to half of the total cellular protein during late phases of

infection (171). Furthered by the observation that polyhedrin was not necessary for viral replication, the idea that the strong polyhedrin promoter (pPolh) could potentially be harnessed for expression of a foreign protein enabled the development of baculovirus as a heterologous protein expression method.

The main goal of recombinant baculovirus generation is to introduce the foreign gene into the baculovirus genome under the polyhedrin promoter. Early methods used homologous recombination between the AcMNPV genome and a plasmid containing the gene of interest when cotransfected into insect cells. However, this method resulted in approximately 0.1% homologous recombination efficiency. Since its initial use, numerous developments have increased the efficiency of recombinant baculovirus generation, such as the use of specialized *E.coli* host strains in which transposition generates recombinant baculovirus at nearly 100% efficiency. Numerous commercial kits are now available, leading to the routine and standardized use of baculovirus for protein expression. The most commonly used insect cell host lines cultured for baculovirus infection include Sf9 cells, derived from *Spodoptera frugiperda* pupal ovarian tissue, and "High-Five" cells derived from *Trichoplusia ni* egg cell homogenates (175-177). The use of baculoviruses for expression of ABC transporters has been well characterized, as membranes used for ATPase assays are often generated by heterologous expression in baculovirus-infected insect cells (178). To date, structural investigations of mouse P-gp purified from insect cells have not been published. However, recent studies from the Laboratory of Cell Biology (Shukla *et al.*, unpublished results) have been able to crystallize mouse P-gp expressed in and purified from insect cells.

1.4.5 Baculovirus-based expression in mammalian cells

Following the first use of recombinant baculovirus for heterologous protein expression in insect cells, it was shown that baculoviruses that have been modified to include mammalian promoters are able to transduce mammalian cells and express foreign proteins (179, 180). The term "BacMam" is used to denote baculovirus-based expression in mammalian cells, and methods to generate recombinant BacMam virus are now commercially available. Prior to the development of BacMam, the use of recombinant vaccinia virus expression was a popular mammalian system for transient expression of heterologous proteins, although its use in a laboratory setting raises numerous safety concerns, as vaccinia virus is infectious via parenteral inoculation, injection and aerosol exposure (181, 182). The BacMam virus is unable to replicate in mammalian cells and is safe for routine laboratory use (175).

The use of BacMam for the expression of P-gp and P-gp mutants has been previously reported (183, 184). The mammalian cell type has been shown to greatly impact the expression level of foreign proteins (early studies showed that hepatocytes were easily transduced and produced high levels of foreign protein, while cell lines from other lineages displayed much lower expression levels). Therefore, a panel of 14 cell lines was screened for high level expression of P-gp when transduced with BacMam (179, 184). Results showed that the HeLa cervical cancer cell line expressed high levels of P-gp 24 hours after BacMam transduction. Overall, BacMam is a safe and effective method for transient transduction of mammalian cells for various functional investigations that require high levels of foreign protein expression.

1.4.6 Purification of membrane proteins

Protein purification refers to the processes required to isolate pure protein of interest from its native environment. To be purified, membrane proteins must be extracted

from the lipid bilayer and stabilized in an environment that mimics the plasma membrane and can support the presence of hydrophobic transmembrane domains (Figure 1.10). This is often accomplished by the removal of membrane proteins from the crude membranes with detergents and their subsequent formation into micelles (140). Detergents are amphipathic molecules comprised of a hydrophobic chain or "tail" and a polar head group (185). Detergents at concentrations above their critical micelle concentration in aqueous solutions spontaneously form spherical micelles, allowing for the solubilization of membrane proteins in micelles, creating an environment mimicking their endogenous membranes. Purification may require a considerable amount of optimization to determine the most effective detergent and the best detergent to protein ratio for protein extraction. While membrane proteins are stable in the environment of their native membrane, they often lose function and/or aggregate in a detergent-solubilized state, therefore making the detergent optimization a critical step as structural techniques require proteins that have maintained their structural integrity (186). When first attempting purification, a series of detergents are often screened to probe for the detergent that recovers the majority of functionally active, homogeneous stable protein (141). The detergent *n*-Dodecyl β-D-maltoside (DDM) is frequently used to extract membrane proteins, and several crystal structures have been obtained from proteins purified using DDM (103, 136, 187-189). Other detergents used for membrane protein purification include *N,N*-dimethyldodecylamine-*N*-oxide (LDAO), *n*-octyl-β-D-glucopyranoside (OG), octyltetraoxyethylene (C_8E_4), and n-decyl- β -d-maltopyranoside (DM) (186). The commonly used detergents DM and DDM are comprised of a hydrophilic maltose head group and a hydrophobic alkyl chain of varying length.

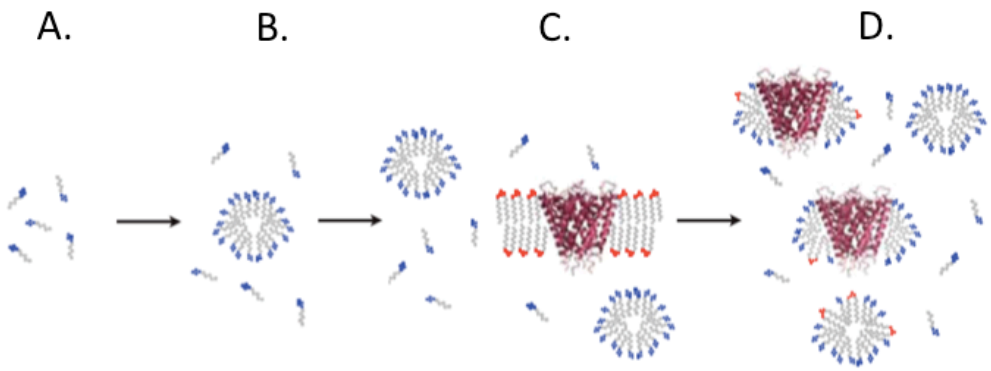


Figure 1.10 Schematic representation of the solubilization of membrane proteins by detergents. Detergent monomers (**A**) form micelles at a concentration above the critical micelle concentration (**B**). When detergents and membranes are mixed (**C**), detergents are able to extract membrane proteins from their native lipid environment. **D)** Solubilization products include the protein-detergent-lipid-complex of interest, and also a mixture of lipid-detergent micelles and leftover detergent monomers. Figure reproduced with permission from Newby *et al.*, 2009 (190).

After the transfer of membrane proteins from cell membranes to micelles, the membrane protein of interest must be isolated from the numerous other solubilized membrane proteins that have also been transferred to micelles. Before expression, many proteins are modified by molecular biology techniques to include a polyhistidine-tag either at their N- or C- terminus. The polyhistidine-tag is comprised of six to twelve histidine residues (191). As histidine binds strongly to divalent metal ions, such as nickel and cobalt, polyhistidine tags can be a useful method to separate the expressed protein from the mix of membrane proteins present after solubilization with detergents. Various commercial methods of nickel affinity chromatography are available which provide nickel ions immobilized with nitrilotriacetic acid (NTA) resin that can be packed into a column for immobilized metal affinity chromatography (IMAC). Detergent-solubilized membrane proteins are passed through the nickel column, allowing the protein of interest tagged with a polyhistidine tail to be captured on the column. Washing of the column with buffers allows for unbound proteins to be eluted, leaving only the polyhistidine-tagged protein bound to

the nickel resin. The protein of interest can be eluted and collected by the addition of a buffer containing imidazole, a compound that can compete with the polyhistidine-nickel interaction and thus displace the protein of interest. Once eluate containing the membrane protein of interest has been collected, the purification process often continues with a series of steps to concentrate the protein in a volume suitable for various techniques such as size exclusion chromatography (SEC) (192). Concentrated, purified protein is often also subjected to high-speed ultracentrifugation, typically before SEC, to remove protein aggregates, as only monodispersed proteins are suitable for structural studies.

1.4.7 Crystallization of membrane proteins

The term crystallization refers to the process of screening a large number of possible conditions that would favor the formation of protein crystals. Protocols usually consist of an initial screen to find a condition hit, followed by numerous optimization screens to improve crystal quality. Despite the fact that approximately 25-30% of human proteins are membrane proteins, there are currently fewer than 500 entries in the Protein Data Bank of high-resolution crystal structures of eukaryotic membrane proteins (141, 193). Membrane proteins have been notoriously difficult to crystallize for various reasons. Successful protein crystallization requires the preparation of a pure, monodispersed, stable protein in milligram quantities (190). As described earlier, heterologous expression of membrane proteins and optimization of purification conditions are bottlenecks for the crystallization. Even if the protein of interest can be expressed and purified in milligram quantities, the physiological nature of certain membrane proteins may make them intrinsically difficult to crystallize. For example, P-gp is a highly flexible protein, a property that is generally thought to be deleterious for its crystallization, as P-gp can adopt various conformations that may prohibit the formation of crystals. Ultimately, the quality of protein crystals is measured by their ability to diffract x-rays. Data collection and structural solution may also be more difficult for

membrane proteins, as crystals have a high solvent concentration, a factor that stems from the presence of the detergent micelle that shields the hydrophobic domains of the protein (141). Crystals with high solvent concentrations are generally fragile and suffer damage from radiation to a greater extent than crystals with lower solvent concentrations.

Another hurdle in the field of crystallization is that crystallization trials that fail to produce crystals are rarely reported. Even if a crystal structure is solved and published, authors rarely detail the evolution of the screening process that was used to generate the crystal (190). As crystallization is often a process of trial and error, knowing the trials and errors of others that have attempted membrane protein crystallization is valuable information that is unfortunately not freely available.

1.5 Chimeras

1.5.1 Chimera nomenclature

Depending on the scientific discipline, the term "chimera" may relate to a variety of either natural or genetically modified proteins or organisms. Unless otherwise noted, in this thesis a chimera is defined as a protein that is encoded by a combination of DNA from two distinct species. A brief introduction to the utilization of chimeras for structural and functional investigations is discussed below.

1.5.2 Chimeras for structural investigations

Mutagenic studies have shown that small changes in the amino acid composition of proteins can greatly affect properties of heterologous protein expression, purification, and crystallization. For example, the β1-adrenergic receptor was unable to be crystallized. However, six point mutations were introduced to increase the stability of the protein in detergent; a factor that allowed for crystal formation and subsequent structural elucidation (194, 195). Due to the profound impact that even subtle differences in amino acid

composition can have on the attainment of high-quality, homogenous protein for crystallization, a recent strategy utilized by both some individual laboratories and structural genomics consortia has been to screen homologs for favorable pre-crystallization properties (196-199). Screening of homologs in this way is a rational approach to obtain structural information for a given protein, as homologs typically retain similar functions but may have subtle residue differences that are more favorable for structural studies than the protein from the species of interest (typically the human homolog). However, for this approach to be relevant, the crystallized homolog must retain the salient structural and functional features of the human homolog. For example, while homolog screening was successfully utilized to determine the structure of mouse P-gp, its physiological relevance to human P-gp may be debated due to their differential substrate profiles.

In cases where the homologous protein shares low sequence identity compared with the human homolog, the rational design of chimeras may be used to exploit the favorable pre-crystallographic or crystallographic properties of the homolog while incorporating segments, domains, or individual residues of interest (200-203). A notable example of protein chimerization as a strategy for crystallization was utilized for the inward rectifier potassium (Kir) channel Kir3.1 (203). Kir channels serve various physiological functions, with Kir3.1 participating in heart rate control through regulation by G-protein and lipid phosphatidylinositol 4,5-bisphosphate (PIP_2) signaling pathways (204). Kir3.1 is composed of four subunits that form a central ion conductance pathway. A characteristic of all Kir channels is the presence of a cytoplasmic pore that is crucial for proper functioning and regulation of the protein. Prior to this study, the structure of a prokaryotic Kir (KirBac1.3) from the gram-negative bacterium *Burkholderia pseudomallei* was reported (205). However, the cytoplasmic pore of KirBac1.3 was only 19% identical to Kir3.1, leaving much to be desired about the structure of the regulatory domains in the eukaryotic intracellular pore. Utilizing the fact that KirBac1.3 could be successfully expressed, purified,

and crystallized, authors generated a KirBac1.3-Kir3.1 prokaryotic-eukaryotic chimera, where the cytoplasmic pore region of KirBac1.3 was substituted with the murine Kir3.1 cytoplasmic pore (99% identical to the human sequence). The chimeric Kir3.1 structure, described at 2.2 Å resolution, indicates a possible mechanism of action of how PIP_2 acts to regulate the channel. Chimeric Kir3.1 exemplifies how strategies of homolog screening and chimera generation can be used in combination to obtain relevant crystal structures.

1.5.3 Functional chimeras of ABC transporters

The generation of chimeras as a strategy for insight into numerous mechanistic and biochemical properties has been used to study ABC transporters (206-212). Studies investigating residues crucial for drug recognition and transport used human P-gp chimerized with domains or helices of human Multidrug Resistance Protein 2 (MDR2, ABCB4) (here, a chimera is defined as a protein that is comprised of cDNA from two distinct proteins rather than two distinct species). MDR2, while 78% homologous to P-gp, is a phosphatidylcholine flippase that does not transport the panoply of compounds characteristic of P-gp. However, due to the overall similarity in domain structure and function, chimeras of P-gp and MDR2 are functional albeit with limited drug transport capacity. A series of chimeric proteins was generated by substituting helices of MDR2 into P-gp (213, 214). Using site-directed mutagenesis to re-introduce residues of P-gp into the MDR2 regions of chimeras, the ability of chimeras to transport P-gp-specific substrates was restored, highlighting the importance of particular residues for drug recognition by P-gp. For example, replacing TM6 in P-gp with the equivalent helix of MDR2 abrogated transport by the chimera of various P-gp substrates. However, the mutagenesis of three residues (Q330, V331, and L332) in TMD6 from MDR2 residues to the equivalent P-gp residues was sufficient to restore the ability of the chimera to transport the P-gp substrates colchicine, bisantrene, rhodamine 123, and vinblastine (214).

1.6 Study aims

It is currently not known why mouse P-gp can be crystallized but human P-gp cannot, despite their high homology. Dissimilarities may be attributable to differential stability in detergents, propensity for purified human P-gp to aggregate, differences in the flexibility of structure, and/or variations in crystal contacts and crystal packing forces. However, it is unclear which non-homologous amino acid residues between mouse and human P-gp contribute these differences. It can be hypothesized that non-homologous amino acids contribute unequally to crystallizability of P-gp homologs, with some non-homologous residues having a significant impact on pre-crystallographic and crystallographic properties while other non-homologous residues are innocuous interspecies differences.

This thesis project aimed to exploit the ability to crystallize mouse P-gp through the generation of human-mouse P-gp chimeras as an approach to elucidate the structure of human P-gp. To this end, a panel of chimeras was generated by the exchange of homologous domains between mouse and human P-gp. As will be seen, mouse P-gp cDNA exerts toxic effects on bacterial growth, thus interfering with the use of molecular cloning techniques. To study the basis of this toxicity a cryptic bacterial promoter in *mdr1a* cDNA was investigated. After resolving toxicity, a panel of chimeras was generated by a domain swapping approach, and chimeras were functionally characterized to thoroughly assess their physiological relevance. Functional chimeras were selected for protein expression in both *P. pastoris* and insect cell heterologous expression systems. Chimeras that were functional, and able to be expressed and purified were screened for crystallization.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Chemicals

The chemicals used in this thesis were sourced from Sigma Aldrich (St. Louis, MO, USA) unless otherwise noted.

2.1.2 Cell culture reagents and solutions

The bacterial, mammalian, and *P. pastoris* cell culture reagents used in this thesis were sourced from Sigma Aldrich (St. Louis, MO, USA) unless otherwise noted.

2.1.3 Antibodies

The names and suppliers of antibodies used in this study are listed in Table 2.1.

Table 2.1 Antibodies

Name (Target)	Species Epitope/Antigen	Source
Anti-mouse IgG HRP linked antibody	Horse anti-mouse IgG antibody conjugated to HRP	Cell Signaling Technology, Danvers, MA, USA
Anti-mouse IgG FITC linked antibody	Rabbit anti-mouse IgG antibody conjugated to FITC	BD PharMingen, San Jose, CA, USA
C219 (P-gp)	Mouse monoclonal antibody; Epitope: VVQEALDKAREGRT	Fujirebio Diagnostics, Malvern, PA, USA
GAPDH (GAPDH)	Mouse monoclonal antibody; Epitope: GAPDH (whole molecule)	Life Technologies, Carlsbad, CA, USA
MRK-16 (P-gp)	Mouse monoclonal antibody; Epitope: extracellular/undefined	Kamiya Biomedical Company, Seattle, WA, USA
PEPG-13 (P-gp)	Rabbit polyclonal antibody recognizing a peptide sequence between residues 592 to 636 of human P-gp	Laboratory of Cell Biology, NIH, MD, USA
Purified Mouse IgG2a κ Isotype Control	Mouse monoclonal antibody with non-specific epitope	BD PharMingen, San Jose, CA, USA

2.1.4 Additional reagents

The names and suppliers of additional reagents used in this study are listed in Table 2.2.

Table 2.2 Additional reagents and materials

Name:	Source:
Gel filtration standard	Bio-Rad, Hercules, CA, USA
Lithium dodecyl sulfate (LDS) sample buffer (4x)	Life Technologies, Carlsbad, CA, USA
MES buffer (10X)	Sigma Aldrich, St. Louis, MO, USA
MOPS (10X)	Life Technologies, Carlsbad, CA, USA
Native PAGE 4-16% Bis Tris Gel	NuPAGE, Novex, Life Technologies, Carlsbad, CA, USA
Novex Sharp Pre-Stained Protein Standard	Life Technologies, Carlsbad, CA, USA
Precast Bis-Tris Protein Gel	NuPAGE, Novex, Life Technologies, Carlsbad, CA, USA
SDS polyacrylamide gels	NuPAGE, Novex, Life Technologies, Carlsbad, CA, USA
TAE Buffer (10X)	Life Technologies, Carlsbad, CA, USA
Tris Buffer (1 M)	KD Medical, Columbia, MD, USA
Tris Buffered Saline (10X)	Corning CellGro, Corning, NY, USA
Tween-20	Bio-Rad, Hercules, CA, USA

2.1.5 Equipment

Equipment used in this study is listed in Table 2.3.

Table 2.3 Equipment

Name:	Manufacturer:
AKTA Fast protein liquid chromatography	GE Healthcare Life Sciences, Piscataway, NJ, USA
Applied Biosystems GeneAmp PCR System 9700	Life Technologies, Paisley, UK
Autoclave	Boxer Laboratory Equipment LTD, UK
Baffled Flasks	Bellco, Vineland, NJ, USA Glass
Bioflow 3000 Fermentor	New Brunswick Scientific, NJ, USA
Cellometer Cell Counter	Nexcelom Bioscience, Lawrence, MA, USA
Confocal Zeiss LSM 710 NLO Microscope	Zeiss, Jena, Germany
FACSCanto II Flow Cytometer	Becton Dickinson Biosciences, San Jose, CA, USA
FLUOstar Omega (Fluorometer)	BMG Labtech, Ortenberg, Germany
GeneFlash Gel Documentation System	Syngene, Cambridge, UK
Hand-held UV light	UVP Inc., CA, USA
Homogenizing Pestle	Thomas Scientific, Swedesboro, NJ, USA
MicroPulser Electroporator	Bio-Rad, Hercules, CA, USA
Mosquito Crystal	TTP Labtech, Cambridge, UK
Nanodrop ND1000 Spectrophotometer	Thermo Scientific, Waltham, MA, USA
Shaking Incubator	Innova 4300 New Brunswick Scientific, NJ, USA
Sonicator Branson 1510	Sonicsonline, Richmond, VA, USA
Superdex 200	GE Healthcare Life Sciences, Piscataway, NJ, USA
Tecan Infinite M200 pro Luminometer	Tecan Group Ltd., Switzerland
Ultrospec 3100 Pro Spectrophotometer	GE Healthcare Life Sciences, Piscataway, NJ, USA

2.1.6 Centrifugation

Centrifugation was carried out using the centrifuges and rotors listed in Table 2.4.

Table 2.4 Centrifugation

Name:	Manufacturer:
Eppendorf Centrifuge 5430R Microcentrifuge	Eppendorf, Hauppauge, NY, USA
Sorvall WX Floor Ultra Centrifuge Beckman Ti 45 Rotor	Thermo Scientific, Waltham, MA, USA Beckman Coulter, Brea, CA, USA
Sorvall MX Micro-Ultra Centrifuge Sorvall S120-AT2 Rotor	Thermo Scientific, Waltham, MA, USA
Sorvall RC3BP+ Low Speed Centrifuge H-6000A 6 x 1000mL Swinging-Bucket Rotor	Thermo Scientific, Waltham, MA, USA

2.1.7 Software

All software used is listed in Table 2.5.

Table 2.5 Software

Name:	Manufacturer:
FloJo Single Cell Analysis Software	FlowJo, Ashland, OR, USA
GeneArt Optimizer Software	GeneArt Gene Synthesis, Life Technologies, Carlsbad, CA, USA
GraphPad Prism version 6.0	GraphPad Software, La Jolla, CA, USA
MacVector version 12.5.1	MacVector, Cary, NC, USA
PyMOL Molecular Graphics System version 1.7.2	PyMOL, Portland, OR, USA
Rock Imager 1000	Formulatrix, Bedford, MA, USA
Unicorn 5.20 (FPLC software)	GE Healthcare Life Sciences, Piscataway, NJ, USA
Zen 2012 Confocal Imaging Software	Zeiss, Jena, Germany

2.1.8 Plasmids

All plasmids used or generated in this study are listed in Table 2.6.

Table 2.6 Plasmids

Name:	Source:
M06	FLNCR, MD, USA
M07	FLNCR, MD, USA
p300-GFP	The current study
p321-GFP	The current study
p321-M107L-GFP	The current study
p321-stop-GFP	The current study
pCR-Blunt Topo	Life Technologies, Carlsbad, CA, USA
pDEST_Human	The current study
pDEST_M107L	The current study
pDEST_mdr1a_M107L	The current study
pDest-008	FLNCR, MD, USA
pDest-625	FLNCR, MD, USA
pDest-720	FLNCR, MD, USA
pDonr-225	FLNCR, MD, USA
pGEM7	Promega, Madison, WI, USA
pGEM7mdr1a	Molecular Oncology Group, Netherlands Cancer Institute, Netherlands
pPICZ	Life Technologies, Carlsbad, CA, USA
pPICZ_HHMM	The current study
pTOPO_HHHH_Opti	The current study
pTOPO_HHHH_WT	The current study
pTOPO_HHHM	The current study
pTOPO_HHMM	The current study
pTOPO_HMHH	The current study
pTOPO_HMHM	The current study
pTOPO_M1_Opti	The current study
pTOPO_M2_Opti	The current study
pTOPO_M3_Opti	The current study
pTOPO_M4_Opti	The current study
pTOPO_MDR1_Cterm	The current study
pTOPO_MDR1_Nterm	The current study

pTOPO_mdr1a_Cterm	The current study
pTOPO_mdr1a_Nterm	The current study
pTOPO_MMMM_Opti	The current study

2.1.9 Enzymes

All enzymes were supplied by New England Biolabs (Hitchin, UK) unless otherwise stated.

2.1.10 Human *MDR1* and mouse *mdr1a* cDNA

Human *MDR1* cDNA (Genebank identification NM_000927.4) was obtained from the Laboratory of Cell Biology, Multidrug Resistance Section, National Institutes of Health, USA (215). Mouse *mdr1a* cDNA (Genbank identification NM_011076.2) was a kind gift of Drs. Alfred Schinkel (Netherlands Cancer Institute, Netherlands) and Philippe Gros (McGill University, Canada) (41, 216). Amino acid and nucleotide sequences of *MDR1* and *mdr1a* are provided in Appendix Figure 1 (*MDR1*) and Figure 2 (*md1a*).

2.1.11 Primers

All primers were ordered from Sigma Aldrich (St. Louis, MO, USA). Primers used are listed in Tables 2.7, 2.8, and 2.9.

Table 2.7 Sequencing primers for wild-type mouse and human cDNA

Primer Name:	Species:	Forward/Reverse:	Location in cDNA (bp):	Primer (5′→3′):
GM616	Human/Mouse	Forward	397-415	GTTTCATTTTGGTGCCTGG
GM617	Human	Reverse	1041-1058	TCAATGCTTGGAGATGCC
GM618	Human/Mouse	Reverse	1512-1529	TAGGCATTGGCTTCCTTG
GM619	Human	Forward	2390-2409	TCAGACAGGATGTGAGTTGG
GM620	Human	Reverse	3414-3433	TCTCTTCCTGTGACACCACC
GM621	Human	Reverse	2872-2891	ACCAAGTAGGCTCCAAACCG
GM622	Mouse	Reverse	1130-1148	TTGTCTGGTTTGTGCCCAC
GM623	Mouse	Forward	2405-2422	ACCCTAAAAACACCACCG
GM624	Mouse	Reverse	3408-3429	CCTCACAATCTCCTCATAAGAC
GM625	Mouse	Reverse	2975-2992	TGGCTTTCGCATAGTCAG
GM627	Mouse	Reverse	156-174	AATAGCAGCCAGAGTTCCC
GM628	Mouse	Forward	3716-3735	TCGTGGTGATTCAGAACGGC
GM630	Mouse	Forward	1372-1392	AATGTGAGGTATCTGAGGGAG
GM631	Mouse	Reverse	2509-2526	TCCTGTCCCAAGATTTGC
GM642	Mouse	Reverse	2974-2992	TGGCTTTCGCATAGTCAGG
GM643	Mouse	Forward	2760-2781	GCAGATACCATACAGAAATGCG
GM644	Mouse	Reverse	2862-2881	TCACCAAGTAGGCACCGAAC
GM645	Human	Reverse	2418-2436	TCCAGTGGTGTTTTTAGGG
GM646	Human	Forward	1682-1700	TGGACACAGAAAGCGAAGC
GM647	Human	Reverse	1810-1829	CCTTTCTCCACAATGACTCC

Table 2.8 Sequencing primers for codon optimized mouse and human cDNA

Primer Name:	Species:	Forward/Reverse:	Location in cDNA (bp):	Primer (5′→3′):
GM655	Human	Forward	38-57	ATGGACTTGGAGGGTGACAG
GM656	Human	Forward	443-462	TGGTGTTTGGCTGCTGGTAG
GM657	Human	Forward	900-921	GTATCAAGAAGGCTATCACTGC
GM658	Human	Forward	1462-1479	AGAGCCAGTTTTGTTCGC
GM659	Human	Forward	1950-1968	AAAACGCTGCTGACGAGTC
GM660	Human	Forward	2487-2504	GATTGGCTAACGATGCTG
GM661	Human	Forward	2932-2951	TCACAAGTTGATGTCCTTCG
GM662	Human	Forward	3491-3512	GCTAACATCCACGCTTTCATTG
GM663	Human	Forward	3842-3861	TCTATGGTTTCTGTTCAGGC
GM671	Human	Reverse	228-271	TTACCAGCGTTAGCGAAG
GM670	Human	Reverse	814-797	AGCCAAAACCTCTTCAGC
GM669	Human	Reverse	1207-1189	TCCCTTGATGTTGTCTGGC
GM668	Human	Reverse	1826-1807	CGTCAGCGTTTCTAACAGTG
GM667	Human	Reverse	2292-2269	GAGTTTTGTCTCTTAGTCTCTGGG
GM666	Human	Reverse	2829-2809	GAGTTTCTGTAAGGAACCTGC
GM665	Human	Reverse	3286-3267	CAACAACTGAACAACTGTGG
GM664	Human	Reverse	3861-3841	GCCTGAACAGAAACCATAGAG
GM672	Mouse	Forward	82-102	CTCCAAGATGGGTAAGAAGTC
GM673	Mouse	Forward	468-486	GACAAAAGTTCTTCCACGC
GM674	Mouse	Forward	888-909	GTATCAAGAAGGCTATCACTGC
GM675	Mouse	Forward	1438-1455	AGTTGTTTCCCAAGAGCC
GM676	Mouse	Forward	1896-1917	AGTTGGTTATGACTCAGACTGC
GM677	Mouse	Forward	2477-2495	TTGGCTAACGATGCTGCTC
GM678	Mouse	Forward	2994-3012	AGGTTTCCTCTTTTGCTCC
GM679	Mouse	Forward	3491-3510	CAGTTCATTGATTCCTTGCC

Table 2.9 Primers used to introduce ClaI silent restriction enzyme site

Primer Name:	Species:	Forward/Reverse:	Primer (5′→3′):
GM608	Human	Forward	GGGGACAAGTTTGTACAAAAAAGTTGGCACCATGGATCTTGAAGGGGACCGCAATG
GM609	Human	Reverse	GGGGCTAGCATCGATTTCACTTTTGGATTCATCAGCTGC
GM610	Human	Forward	GGCATCGATGCCTTGGAAATGTCTTCAAATG
GM641	Human	Reverse	GGCCTAGGGGGGACCACTTTGTACAAGAAA
GM612	Mouse	Forward	GGGGACAAGTTTGTACAAAAAAGTTGGCACCATGGAACTTGAAGAGGACCTTAAG
GM613	Mouse	Reverse	GGGGCTAGCATCGATTTCATCCTTAGATTTACAAGCTTC
GM614	Mouse	Forward	GGCATCGATAATTTAGACATGTCTTCAAAAG
GM615	Mouse	Reverse	GGGCTAGCGGGGACCACTTTGTACAAGAAAGTTGATTAGTGATGGTGATGGTGATGTGCTA GTGAGCGCTTTGCTCCAGCCTGCAC

2.2 Methods

2.2.1 Amplification of cDNA by polymerase chain reaction

DNA was amplified using the PCR primers listed in Table 2.12. Vent Polymerase was used according to the standard protocol. Briefly, 50-250 ng of template DNA to be amplified was mixed with 1x ThermoPol reaction buffer, 10 µM forward and reverse primers, a mixture of 2 mM deoxynucleotides, 0-4 mM $MgSO_4$, 1 Unit of DNA Polymerase and deionized water to a total volume of 50 µL. Annealing temperatures for PCR reactions were determined using analysis on MacVector. A typical PCR reaction consisted of 35 cycles, including a 1-minute initial denaturation at 95°C followed by a 45-second hold at the pre-determined annealing temperature and a 3-minute elongation period at 72°C. PCR products were confirmed by agarose gel electrophoresis and gel purified as described in Sections 2.2.2 and 2.2.3, respectively.

2.2.2 Agarose gel electrophoresis

DNA (0.5-2.0 µg/well) was size-fractionated by agarose gel electrophoresis. Gels consisted of 0.7% agarose, 1 µg/mL ethidium bromide in 1x Tris-acetate-EDTA (TAE) buffer. DNA samples were mixed with 1x DNA gel loading dye containing bromophenol blue then loaded onto the gel and electrophoresed at 70 to 100 volts for 1 to 2 hours. DNA fragment size estimation was carried out by comparison to 500 ng of either a supercoiled or linear DNA ladder (range 0.5-10.0 kb) (New England Biolabs, Hitchin, UK). DNA products were visualized by UV excitation (365 nm). Images were either printed or saved electronically.

2.2.3 Agarose gel purification and extraction of DNA fragments

DNA to be gel purified was mixed with 1x DNA gel loading dye containing bromophenol blue and electrophoresed on a 0.7% agarose gel at 70 volts for 2 hours to allow for optimal band separation (Section 2.2.2). DNA was visualized using a hand-held UV light (365 nm). DNA

bands of interest were excised using a sterile scalpel and purified using the QIAEXII Gel Extraction Kit (Qiagen, Limburg, The Netherlands) according to the manufacturer's protocol. Purified DNA fragments were eluted with nuclease-free water and stored at -20°C. A sample of each gel purification was subjected to agarose gel electrophoresis to confirm the purification processes produced the expected result (Section 2.2.2)

2.2.4 Topo-cloning of PCR products

Cloning of agarose gel purified blunt-end PCR products was carried out using the Zero Blunt TOPO PCR Cloning Kit (Life Technologies, Carlsbad, CA, USA) into the supplied pCR-Blunt TOPO cloning plasmid according to the manufacturer's protocols. Blunt-end PCR products ligated into pCR-Blunt TOPO plasmids were transformed into One Shot Top10 Chemically Competent *E.coli* cells (Section 2.2.8) and cultured on Luria-Bertani (LB) agar plates containing 50 µg/mL kanamycin for 24 hours at 34-37°C to yield single colonies.

2.2.5 Restriction enzyme digests

All digests were performed using New England Biolabs Ltd. restriction enzymes according to the manufacturer's instructions. Typically, a 15-50 µL reaction volume containing 0.5-2.0 µg of plasmid DNA was incubated from 2 to 12 hours at 37 °C unless otherwise noted. Restriction digests were assessed by agarose gel electrophoresis and visualization of cut DNA fragments compared with uncut, supercoiled parent plasmid and a supercoiled ladder (Section 2.2.2).

2.2.6 Ligation of DNA fragments

DNA fragments to be ligated were incubated with T4 DNA ligase and 1x ligation buffer in approximately a 1:7 vector to insert ratio for a minimum of 2 hours at 18°C. Ligations were

transformed into One Shot Top10 Chemically Competent *E.coli* cells as described in section 2.2.8.

2.2.7 Preparation of antibiotic stock solution

Antibiotic stocks of 100 mg/mL ampicillin, 100 mg/mL zeocin (InvivoGen, San Diego, CA, USA), and 50mg/mL kanamycin were prepared in water. Once dissolved, stock solutions were filter sterilized using a 0.22 µm filter and stored at 4°C for up to 2 months.

2.2.8 Bacterial transformation of plasmid DNA

Chemically competent One Shot Top10 competent *E.coli* cells (genotype: F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG) (Life Technologies, Carlsbad, CA, USA) were transformed according to the manufacturer's protocol. Briefly, cells were thawed on ice and mixed with DNA to be transformed for 5 to 15 minutes. Cells were heat shocked at 42°C for 30 seconds and immediately placed back on ice for an additional 5 minutes. Cells were then transferred to sterile tubes containing 200 µL of room temperature SOC media and incubated on an orbital shaking incubator at 37°C for 1 hour. After incubation, cells were plated on LB agar plates containing selective media for 16-24 hours at 34-37°C.

2.2.9 Preparation of selective LB media and LB agar plates

A volume of 400 mL of deionized water was prepared with 1% bacto-tryptone, 0.5% bacto-yeast extract, and 0.5-1% sodium chloride to generate LB media. 1.5% bacto-agar was added to the above solution for the generation of LB agar. Once formulated, LB broth/agar was sterilized in an autoclave and allowed to cool to 50°C before addition of antibiotic selective agents. Antibiotic stock solutions (Section 2.2.7) were diluted to 1x working concentrations of 100 µg/mL ampicillin, 100 µg/mL zeocin, and 50 µg/mL kanamycin.

2.2.10 Inoculation of bacterial colonies into LB selective media

Single transformed *E. coli* colonies, grown overnight on selective LB agar plates, were inoculated using a sterile loop into 5 or 400 mL of LB media containing the appropriate antibiotic. Cultures were incubated in an orbital shaking incubator for 16-24 hours at 34-37°C.

2.2.11 Large- and small-scale purification of plasmid DNA from transformed bacteria

For small-scale plasmid purification, 2-5 mL of culture was purified using the Wizard Plus SV Miniprep DNA Purification (Promega, Madison, WI, USA) system according to the manufacturer's protocol. For large-scale plasmid purification, 200-500 mL of culture was purified using the HiSpeed Plasmid Maxi Kit (Qiagen, Limburg, The Netherlands) according to the manufacturer's protocol. Purified plasmid DNA was eluted in sterile water and stored at -20°C.

2.2.12 Gateway cloning

Gateway cloning (Life Technologies, Carlsbad, CA, USA) was performed according to manufacturer's protocol. All plasmids to be used for gateway cloning in this thesis project were flanked by *attB1* and *attB2* sites. A BP clonase reaction was performed with the pDonr-225 plasmid to generate entry clones (flanked by *attL* sites). Next, and LR clonase reaction was carried out between the entry clone and the desired expression vector (termed pDest vectors; Table 2.9) to generate the final expression vector. LR clonase reactions of entry clones into pDest-008 and pDest-625 vectors were performed to generate shuttle vectors for baculovirus and BacMam expression, respectively (Section 2.2.30). LR clonase reactions of entry clones into the pDest-720 vector was used to generate plasmids used for lipofectamine-mediated transient transfections (Section 2.2.33).

2.2.13 DNA quantification

DNA was quantified using the ThermoScientific Nanodrop ND1000 Spectrophotometer. Absorbance was measured at 260 nm. Sample purity was assessed by the ratio of absorbance of the sample at 260 and 280 nm ($A_{260/280}$).

2.2.14 Codon optimization of cDNA

Codon optimization for *P. pastoris* expression was performed using GeneOptimizer software and completed by GeneArt Gene Synthesis services (Life Technologies, USA).

2.2.15 Sequencing and sequence analysis of DNA

Sequencing was carried out either by Source Bioscience (Oxford, UK) or the NIH Center for Cancer Research Genomics Core (NIH, USA) using Big Dye Terminator sequencing kits (Life Technologies, Carlsbad, CA, USA). Analysis of sequencing chromatograms was carried out using MacVector alignment software. Sequencing chromatograms of wild-type *mdr1a* and *MDR1* were aligned to references sequences from Genbank identifications NM_011076.2 (*mdr1a*, Appendix Figure 2) and NM_000927.4 (*MDR1*, Appendix Figure 1). Sequencing chromatograms of codon optimized *MDR1* and *mdr1a* were aligned to reference sequences displayed in the corresponding Appendix figures.

2.2.16 Mammalian cell culture

All cell lines were grown at 37°C in 5% CO_2 and cultured in high-glucose DMEM, supplemented with 10% fetal bovine serum, 5 mM L-glutamine, 50 units/mL penicillin, and 50 µg/mL streptomycin. Additionally, the P-gp-expressing cell lines KB-8-5-11 and C3M were cultured in 100 ng/mL and 1 µg/mL colchicine, respectively, to maintain P-gp expression (217, 218).

2.2.17 Mammalian cell lysates

Cells were trypsinized, counted, and 1×10^6 cells were isolated. Cells were centrifuged in a bench top Eppendorf 5430R microcentrifuge to obtain a cell pellet and supernatant was removed. Cell pellets were resuspended in 100 µL of pre-cooled buffer composed of 10 mM Tris, 0.1% Triton-X 100, 10 mM $MgSO_4$, 2 mM $CaCl_2$, 1% aprotinin, 1 mM AEBSF, 2 mM DTT and 1 Unit of DNase. Resuspended cells were flash frozen on dry ice and subsequently thawed on wet-ice. Lysis was achieved by three cycles of a 1-minute sonication followed by a 30-second cool-down on ice. Supernatant from centrifuged lysates were stored at -80°C.

2.2.18 Protein estimation

Two methods were used in this thesis to quantify protein. Protein estimation on whole-cell lysates, crude membranes, and solubilized membrane proteins were completed using the Bio-Rad protein estimation Bradford assay (Bio-Rad, Hercules, CA, USA) according to the manufacturer's protocol. Protein estimation of purified protein was performed using a ThermoScientific Nanodrop ND1000 Spectrophotometer. Absorbance was measured at 280 nm.

2.2.19 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Samples were prepared by combining the desired amount of protein and 1x lithium dodecyl sulfate (LDS) sample buffer, to a total volume of 10-30 µL. Samples were loaded into each well of a Bis-Tris or Tris- Acetate SDS-PAGE gel. Gels were electrophoresed at 150 V either in 1x MOPS (for Bis-Tris gels) or 1x Tris-Acetate (for Tris-Acetate gels) running buffer. Electrophoresis was stopped when the dye-front reached the bottom of the gel. Gels were subsequently used either for western blotting (Section 2.2.20) or colloidal blue staining (Section 2.2.21)

2.2.20 Western blotting

After SDS-PAGE (Section 2.2.19), protein was transferred to a PVDF membrane using the iBlot Dry Blotting System (7 minute transfer) (Life Technologies, Carlsbad, CA, USA). Following transfer, membranes were blocked in 10% non-fat milk (American Bioanalytical, Natick, MA USA) dissolved in 1x Tris buffered saline (TBS) supplemented with 0.1% Tween-20 (termed TTBS buffer) for at least 1 hour. Blocking solution was discarded and membranes were washed with 15 mL of TTBS buffer 4 times for 5 minutes each. Membranes were incubated with primary antibody in 20 mL of TTBS buffer for an additional 1 hour. Membranes were washed as previously stated, and incubated with the secondary antibody for 1 hour. Membranes were washed six times with TTBS buffer before addition of chemiluminescent horseradish peroxidase (HRP) antibody detection reagent (HyGLO Chemiluminescent HRP Antibody Detection Reagent, Denville Scientific, South Plainfield, NJ, USA) and exposure to chemiluminescence film (Carestream Health, Rochester, New York, USA). Antibodies used are listed in Table 2.3. The primary antibodies C219 and GAPDH were used at a dilution of 1:10,000 and the PEPG-13 antibody was used at a 1:5,000 dilution. The anti-mouse IgG HRP-linked secondary antibody was used at a dilution of 1:10,000.

2.2.21 Colloidal blue stain

The Colloidal Blue Staining Kit (Life Technologies, Carlsbad, CA, USA) was used according to the manufacturer's protocol. SDS-PAGE gels (Section 2.2.19) were allowed to incubate with blue stain overnight and were de-stained in deionized water for an additional 48-96 hours.

2.2.22 Blue native polyacrylamide gel electrophoresis

Blue native PAGE (BN-PAGE) was carried out with the NativePAGE Novex Bis-Tris Gel System (NuPAGE, Novex, Life Technologies, Carlsbad, CA, USA) according to the manufacturer's protocol. A native 4-16% Bis-Tris gel was loaded with 6 µg of purified protein, 1x NativePAGE

sample buffer, G-250, and deionized water to a total volume of 10 µL. The amount of G-250 added to the sample is dependent on the amount of DDM in the purified protein sample, with G-250 concentrations added being at least 25% of the detergent concentration. Samples were electrophoresed at 150 V for 1 hour and then for an additional 45 minutes at 250 V. After electrophoresis, gels were destained in deionized water for 48 to 72 hours.

2.2.23 MRK-16 P-gp cell surface expression assay

Cells were trypsinized, counted, and 2×10^5 cells per assay condition were transferred into polystyrene flow cytometry tubes. Cells were centrifuged, supernatant removed, and resuspended in 1 mL of IMDM containing either 1 µg MRK-16 primary antibody or the control primary antibody isotype control (mouse IgG2a Kappa) and incubated at 37˚C for 30 minutes. After incubation, cells were centrifuged and supernatant containing MRK-16 and control antibody was removed. One mL of IMDM media containing 1 µg secondary antibody (FITC-labeled anti-mouse IgG2a) was added to tubes and incubated in the dark at 37˚C for an additional 20 minutes. After incubation, cells were centrifuged, supernatant was removed, and cells were resuspended in 200 µL of cold PBS and placed on ice. Flow cytometry was carried out as described in Section 2.2.25.

2.2.24 Substrate efflux and inhibition assay

Cells were trypsinized, counted, and 2×10^5 cells per assay conditions were transferred into polystyrene flow cytometry tubes. Cells were centrifuged and supernatant was removed. Cells were resuspended in IMDM containing either 10 µM CAM (Life Technologies, Carlsbad, CA, USA), 10 µM JC-1 (Cayman Chemical, Ann Arbor, Michigan, USA), or 1.5 µM Rh123 (Life Technologies, Carlsbad, CA, USA) with or without a P-gp inhibitor and briefly vortexed before incubating at 37˚C for 45 minutes. Inhibitors used in this study were used at the following concentrations: 100 nM TQR (MedKoo Biosciences, Chapel Hill, NC, USA), 10 µM CsA, and 10

µM elacridar (MedKoo Biosciences, Chapel Hill, NC, USA). After incubation, cells were centrifuged, supernatant was removed, and resuspended in 200 µL of cold PBS and placed on ice to stop transporter activity. Flow cytometry was carried out as described in Section 2.2.25.

2.2.25 Flow cytometry

A total of 10,000 individual cell-counting events were recorded using a FACSCanto II flow cytometer. A 488 nM blue laser was used to excite all fluorophores and emission was measured either at 530/30 nm (Rh123 and CAM) or 575/26 nm (JC-1). FACS data were analyzed using FlowJo Single Cell Analysis software.

2.2.26 ATPase assay

A 2x stock of assay buffer, composed of 100 mM MES-Tris buffer (pH 6.8), 100 mM KCl, 10 mM sodium azide, 2 mM EGTA, 2 mM ouabain, 20 mM MgCl2, and 4 mM DTT was generated. Insect cell crude membranes (Section 2.2.40) infected with the required recombinant baculovirus to be used were thawed on ice. All compounds to be tested for ATPase activity were performed in the presence and absence of 0.3 mM sodium orthovanadate to determine basal P-gp ATPase activity. Sodium orthovanadate (10 mM) was prepared by dissolving powdered sodium orthovanadate in deionized water and heating at 99˚C for 5 minutes to depolymerize oligomers. Assays were carried out by the addition of 10 µg of insect cell membranes to 1x assay buffer, the desired amount of the P-gp modulator to be tested, 0.2 mM of sodium orthovanadate (if appropriate) and deionized water to a total assay volume of 95 µL. The reaction was incubated at room temperature for 5 minutes, after which 5 µL of 100 mM ATP was added to each tube to yield a 5 mM final ATP concentration. Each assay condition was allowed to incubate at 37˚C for 20 minutes before the addition of 100 µL of 5% SDS to each assay condition to stop P-gp ATPase activity. To develop the assay,

500 µL of deionized water, 400 µL of P_i reagent (composed of 1% ammonium molybdate in 2.5 M sulfuric acid and 0.014% potassium antimonyl tartrate trihydrate) and 200 µL of 1% ascorbic acid was added into each assay condition. Color was allowed to develop for 10 minutes before samples were measured spectrophotometrically at 880 nm. Absorbance values were used to calculate the rate of ATPase activity (P_i/mg/min).

2.2.27 Cytotoxicity assay

Cytotoxicity assays were carried out using the CellTiter-Glo Luminescent Cell Viability Assay (Promega, Madison, WI, USA). Cells were trypsinized and counted, and plated into 96-well, white-walled plates at a density of 4,000 cells per well in 100 µL of media. After a 24-hour incubation, cells were dosed with the compound to be tested for toxicity. Serial dilutions of compounds to be tested for cytotoxicity were carried out. Drugs solubilized in DMSO were diluted so that the final concentration of DMSO did not exceed 0.0025 %, a concentration that is non-toxic to the cells used in this thesis. Cells were incubated with drug for an additional 24 to 72 hours. Viability was assessed by the removal of cell culture media, and the addition of 50 µL of CellTiter-Glo Reagent. Viability was assessed by the quantification of luminescent signal using a Tecan Infinite M200 pro Luminometer.

2.2.28 Fluorescence spectroscopy

Fluorescence spectroscopy was used to determine levels of GFP fluorescence in bacteria transformed with GFP plasmids (Table 2.9). One Shot Top10 Chemically Competent *E.coli* cells were transformed with GFP plasmids, and transformants were plated LB agar containing zeocin (Section 2.2.8). Colonies were selected for growth in 5 mL of LB media for 12 hours. Cultures were centrifuged and LB media was removed. Cell pellets were resuspended in PBS and transferred to a 96-welled, black walled plate for fluorescence analysis. Fluorescence intensity was measured by the FLUOstar Omega fluorometer.

2.2.29 Confocal microscopy

Transformed *E.coli* to be analyzed for GFP fluorescence by confocal microscopy were transferred to chambered coverslips. Cells were imaged with a Zeiss LSM 710 NLO confocal microscope and images were processed by Zen 2012 software.

2.2.30 Generation of baculovirus and BacMam baculovirus

Baculovirus and BacMam baculovirus were generated in the Protein Expression Laboratory at the Frederick National Lab for Cancer Research (FNLCR). The Bac-to-Bac Baculovirus Expression System (Life Technologies, Carlsbad, CA, USA) was used according to the manufacturer's protocol to generate recombinant baculovirus and BacMam baculovirus. Briefly, Gateway cloning (Section 2.2.12) was used to transfer cDNA from bacterial cloning plasmids to the pDest-625 and pDest-008 (Table 2.9) expression vectors for BacMam baculovirus and baculovirus expression, respectively. Expression vectors were transformed into *E.coli* DH10Bac cells where site-specific transposition of the gene of interest into a baculovirus shuttle vector (bacmid) occurred. Blue-white screening was carried out, white colonies were selected, and bacmid DNA was purified by alkaline lysis. Bacmid DNA was transfected into insect cells to generate recombinant baculovirus and BacMam baculovirus.

2.2.31 Baculovirus-mediated protein expression in insect cells

Protein expression in insect cells was carried out by the Protein Expression Laboratory at the FNLCR using a previously optimized protocol (219). Briefly, a 1 L culture of insect cells was infected (MOI of 10) with recombinant baculovirus containing the P-gp construct to be expressed. Approximately 60-72 hours post-infection, cells were examined for signs of baculovirus infection by trypan blue viability staining. After approximately 30 to 40% cell death was observed, cells were harvested, centrifuged, and stored at -80°C.

2.2.32 BacMam transfection

BacMam transfections were carried out according to a previously optimized protocol (184). In short, HeLa cells were counted, centrifuged, and resuspended to a density of 2.5×10^6 cells in 3 mL of DMEM media. BacMam virus was added to HeLa cells at a titer of 50 virus particles per 1 cell, and incubated at 37°C for 1 hour. After 1 hour, 23 mL of DMEM media was added and incubated for an additional 3 hours. Sodium butyrate was added to a final concentration of 10 mM and transfections were incubated for 20 to 24 hours before harvesting for assays.

2.2.33 Lipofectamine transfection

Lipofectamine transfections were performed using Lipofectamine 2000 Reagent (Life Technologies, Carlsbad, CA, USA) according to manufacturer's protocol. Briefly, 5×10^5 human embryonic kidney (HEK) 293 cells, cultured in DMEM media, were transfected with 2 µg of plasmid DNA (Table 2.9). HEK 293 cells were incubated for 24 hours, after which media was replaced with media containing 100 nM paclitaxel and incubated for an additional 48 hours before harvesting.

2.2.34 *P. pastoris* cell culture

The *P. pastoris* yeast strains GS115, KM71H, SMD1168 and X-33 were obtained from Life Technologies and cultured according to the supplied protocol (see Table 6.1 for a description of yeast strain genotypes). Media compositions for the culturing of yeast are shown in Table 2.13. X-33, KM71H, and SMD1168 are routinely cultured in YP or YPG media. Due to the deletion of the histidinol dehydrogenase gene, GS115 must have histidine supplemented in the media, and can be maintained on either MMH or YPG media supplemented with histidine. *P. pastoris* suspension cell culture was maintained in baffled flasks in a shaking incubator maintained at 30˚C unless otherwise stated.

Table 2.10 Media compositions used for *P. pastoris* cell culture

MMH Media	1.34% yeast nitrogenous base, 4×10^{-5}% biotin, 0.004% histidine
YP Media	1% yeast extract, 2% peptone
YPD Media	1% yeast extract, 2% peptone, 2% dextrose
YPDS Agar	1% yeast extract, 2% peptone, 2% dextrose, 1 M sorbitol, 2% agar
YPG Media	1% yeast extract, 2% peptone, 1% glycerol
YPGM Media	1% yeast extract, 2% peptone, 1% glycerol, supplemented with methanol
YPM Media	1% yeast extract, 2% peptone, supplemented with methanol

2.2.35 Measurement of optical density

Optical density (OD) of bacterial and yeast cell culture was measured spectrophotometrically at 660 nm. A sample of media was taken (typically 1 mL) and transferred to disposable plastic cuvette for OD determination. Dense cultures were diluted with the appropriate media if necessary. OD was measured twice for each sample, and the average of duplicate OD measurements was used as the final OD value.

2.2.36 Preparation of *P. pastoris* competent cells

Competent *P. pastoris* cells were prepared immediately preceding transformation, as the viability of competent *P. pastoris* is greatly decreased by freeze/thaw processes required for long-term storage. The *P. pastoris* host strains GS115, KM71H, SMD1168 and X-33 were grown to an OD of 1.5 in 50 mL of the appropriate media. Cells were centrifuged to obtain a cell pellet washed twice with pre-chilled deionized sterile water to remove residual growth media. Competency was achieved by a 5-minute incubation of cells with 2 mL of 1 M sorbitol.

2.2.37 *P. pastoris* transformation

Competent *P. pastoris* cells were transformed by the electroporation of 3-5 µg of linearized plasmid DNA. Competent cells and linearized DNA were co-incubated on ice for 5 minutes before electroporation. Cells were electroporated at 1.5 kV for 1 pulse. After electroporation, 300 µL of 1 M sorbitol was added to cells and incubated at 30°C in an orbital shaking incubator for 1 hour. After 1 hour, 500 µL of YPD media was added and transformed cells were incubated for an additional 2 hours. Transformed cells were streaked onto YPDS agar plates containing 0.1 to 2.0 mg/mL of zeocin. Recipes for YPDS agar and YPD media can be sourced from Table 2.13.

2.2.38 Small- and large-scale protein expression in *P. pastoris*

Single colonies of *P. pastoris* transfected with chimeric constructs were picked for small-scale (15 mL) or large-scale (8 L) culturing. Cells were grown to the optimal OD before protein expression was induced. Protein expression was induced by the addition of 0.1 to 5.0% methanol to cultures. Cultures were grown for 5 days and methanol was supplemented every 24 hours to account for methanol consumption and evaporation. After 5 days, cells were collected and centrifuged at 5,000 x g (Sorvall RC3BP+ Low Speed Centrifuge, H-6000A 6 x 1000mL Swinging-Bucket Rotor). Supernatant was discarded and cells were washed of residual media with washing buffer composed of 50 mM Tris and 2 mM PMSF. Washed pellets were frozen and stored at -80°C.

2.2.39 Preparation of *P. pastoris* cell lysates

P. pastoris cell samples were centrifuged to obtain cell pellets. Supernatant was removed and cells were resuspended in 50 µL of cell-breaking buffer composed of 50 mM sodium phosphate, 1 mM PMSF, 1 mM EDTA, and 5% glycerol. Glass beads were added so that approximately 50% of the volume contained beads, and samples were vortexed for 30

seconds then cooled on ice for 30 seconds for 10 cycles. Samples were centrifuged and lysates were transferred to sterile microcentrifuge tubes and stored at -80°C.

2.2.40 Preparation of crude membranes from insect cells

All components (buffers, homogenizers, centrifuges, etc.) were pre-cooled to 4°C. Frozen insect cell pellets were thawed on ice and resuspended in 30-40 mL of homogenization buffer composed of 50 mM Tris (pH 7.5), 50 mM mannitol, 2 mM EGTA, 0.5% aprotinin, 2 mM DTT, and 1 mM AEBSF and incubated for 30 minutes. Cells were homogenized in a tight-fitting homogenizer 20-30 times, first with pestle A and then with pestle B. Homogenized cells were centrifuged at 5,000 x g (Sorvall RC3BP+ Low Speed Centrifuge, H-6000A 6 x 1000mL Swinging-Bucket Rotor) for 5 minutes to pellet cell debris. Supernatant was transferred to ultracentrifuge tubes and ultracentrifugation at 165,000 x g (Sorvall WX Floor Ultra Centrifuge, Beckman Ti 45 Rotor) was conducted for 40 minutes to isolate cell membranes. Supernatant was discarded and 10 mL of wash buffer composed of 50 mM Tris (pH 7.5), 300 mM mannitol, 1 mM EGTA, 0.5% aprotinin, 1 mM DTT, and 1 mM AEBSF was added to the cell pellet. The cell pellet was washed by passage through an 18-gauge needle five times. Ultracentrifugation was completed to re-pellet washed cells. Wash buffer was discarded and membranes were resuspended in storage buffer composed of 20 mM Tris (pH 7.5), 300 mM mannitol, 1 mM EGTA, 0.5% aprotinin, 1 mM DTT, 1 mM AEBSF, and 10% glycerol by passage through an 18-gauge needle and subsequently through a 22-gauge needle five times. Crude membranes were aliquoted, flash frozen on dry ice, and stored at -80°C.

2.2.41 Protein purification by immobilized metal affinity chromatography

Crude membranes from insect cells (Section 2.2.40) were thawed on ice and all equipment and buffers were pre-cooled to 4°C. Crude membranes at a concentration of 5-10 mg/mL were solubilized with a solubilization buffer composed of 50 mM Tris (pH 7.5), 15% glycerol, 200 mM NaCl, 5 mM β-mercaptoethanol, 20 mM imidazole and 2% DDM (Anatrace Inc., Maumee, Ohio, USA) and incubated for 30 minutes at 4°C while gently stirring. Solubilized membranes were centrifuged at 165,000 x g (Sorvall WX Floor Ultra Centrifuge, Beckman Ti 45 Rotor) for 30 minutes and the supernatant collected. Nickel-NTA resin to be used for IMAC was washed in wash buffer composed of 50 mM Tris (pH 7.5), 15% glycerol, 200 mM salt, 5 mM imidazole, 0.0675% DDM, and 0.04% β-mercaptoethanol. Washed resin and solubilized membranes were co-incubated for 12 hours at 4°C gently stirring to allow for binding of His-tagged proteins to nickel-NTA resin. After incubation, resin was transferred to a 15 mL column and washed with approximately 10 column volumes of wash buffer or until no protein could be detected from the column eluate. The presence or absence of protein in the column eluate was determined by collecting samples of the eluate for immediate analysis by Bradford protein assay. Protein was eluted with an elution buffer composed of wash buffer supplemented with 300 mM imidazole to displace His-tagged protein from nickel NTA resin. Protein was collected and concentrated in a 100 kDa protein concentrator (Amicon Millipore, MA, USA) to a volume of 300 to 500 µL. Concentrated, purified protein was subjected to ultracentrifugation at 400,000 x g (Sorvall MX Micro-Ultra Centrifuge, Sorvall S120-AT2 Rotor) for 1.5 hours to remove protein aggregates. Size exclusion chromatography (SEC, Section 2.2.42) was completed and fractions 15 to 28 were collected for subsequent analysis by western blotting (Section 2.2.20), colloidal blue staining (Section 2.2.21), and BN-PAGE (Section 2.2.22). Fractions containing monomeric protein were collected and concentrated to 5-10 mg/mL for crystallography screening (Section 2.2.43). Purified protein was flash frozen in liquid nitrogen and stored at -80°C.

2.2.42 Size-exclusion chromatography

SEC buffer composed of 50 mM Tris (pH 7.5), 2% glycerol, 200 mM NaCl, 0.0675% DDM, and 0.04% β-mercaptoethanol was filter sterilized and degassed overnight. A Superdex 200 column was prepared by equilibration with two column volumes (approximately 50 mL) of SEC buffer. Purified protein concentrated to a volume less than 500 µL was loaded into a glass syringe and injected into a 500 µL injection loop. Protein was eluted from the column in SEC buffer at a rate of 0.5 mL per minute for 1.2 column volumes (~27 mL). Eluted protein was collected in 0.5 mL fractions. Protein elution from the column was measured by an UV detector and recorded in absorbance arbitrary units (AU).

2.2.43 Crystallization screening

Crystallization screening was completed by Lothar Esser (NIH, USA) using the MemGold (Molecular Dimensions, Sufolk, UK) and MembFrac (Hampton Research, Aliso Viejo, CA, USA) crystallization screening conditions. Hanging drop vapor diffusion in 96-well plate format was used to screen for protein crystals of chimeric P-gp. Crystallization conditions to be tested were dispensed in 200 nL droplets by a Mosquito Crystal liquid handling machine. Approximately 200 nL of purified protein concentrated to 5-10 mg/mL was added to each droplet. Hanging drop plates were kept at 4°C to allow for crystal formation. Plates were imaged twice a week by a high-resolution plate imager and assessed for crystal formation.

2.2.44 Statistical analysis

Curve-fitting and all statistical analysis was performed using GraphPad Prism 6.0. Multiple groups were analyzed using one-way analysis of variance (ANOVA) where significance was defined as $p < 0.05$. Data from dose-response curves of ATPase assay on crude membranes were fitted with the Michaelis-Menten curve from which the K_m and V_{max} values were calculated. Determination of 50% inhibitory concentrations and half-life times were

generated by curve fitting of data, and generating a linear fit of data on the curve slope for each data set to determine the concentration at 50% (Microsoft Excel). The data for each experiment were then averaged.

Chapter 3: Identification of a cryptic putative bacterial promoter in mouse P-glycoprotein cDNA

3.1 Introduction

Molecular cloning encompasses a set of experimental techniques in which genetic material is placed in a vector, such as a bacterial plasmid or viral vector, and transformed into a host organism for replication. The most common host organism utilized for cloning is the gram-negative bacterium *Escherichia coli*. *E. coli* is a popular host organism for numerous reasons. It can be grown safely in most laboratory environments with minimal equipment and materials. Importantly, wild-type *E. coli* replicates approximately every 20 minutes under exponential growth conditions, allowing for the amplification of large quantities of plasmid DNA. Studies characterizing ABC transporters have utilized cloning to address various experimental aims. For these reasons, a strategy for chimera generation was devised that required cloning of mouse (*mdr1a*) and human (*MDR1*) P-gp cDNA in *E. coli*.

An early impediment to the generation of chimeras in this thesis project was the inability to propagate mouse P-gp cDNA in *E. coli* (*vide infra*). As will be discussed in this chapter, *E. coli* transformed with *mdr1a* cDNA invariably yielded various insertion and deletion mutations. Plasmid DNA mutations can occur due to extrinsic factors, such as UV radiation or chemical mutagens, or intrinsic factors, such as inaccuracies during DNA replication and the transposition of insertion sequences from chromosomal to plasmid DNA. Thus, the occurrence of mutations in plasmid DNA upon replication in bacteria can be expected at low frequencies. However, bacteria transformed with *mdr1a* cDNA yielded 100% mutant plasmids, a frequency much higher than would be expected due to the aforementioned factors. The repeated appearance of mutated *mdr1a* plasmid DNA upon transformation into bacteria suggested that the presence *mdr1a* cDNA is exerting a selective pressure against *E. coli*, allowing only sub-clones with randomly occurring mutations in *mdr1a* cDNA that mitigate this selective pressure to survive transformation.

One mechanism by which the introduction of plasmid DNA during cloning can exert lethal effects is by the expression of a protein that is toxic to the bacterial host. Expression of foreign genes in *E. coli* requires a bacterial promoter upstream of the gene to be expressed, a principle that is the basis of heterologous protein expression of eukaryotic proteins in prokaryotic hosts. However, expression of a foreign protein can have deleterious effects on the host organism. For example, expression of certain membrane proteins in *E. coli* can result in membrane perturbation and subsequent cell death (220-222). Previous research in the Laboratory of Cell Biology (NIH) has attempted to express human P-gp in *E. coli* as a method for high-level protein expression for crystallographic studies (148). However, the expression of human P-gp was lethal to *E. coli*, reducing the colony forming ability 500- to 1000-fold (G. Evans, M. Gottesman, and I. Pastan, unpublished data). While these data indicate that the toxicity seen when cloning *mdr1a* may be due to mouse P-gp protein expression, the plasmids used for cloning in this thesis did not contain a bacterial promoter. Expression of mouse P-gp in *E. coli* should not, theoretically, occur in the absence of a bacterial promoter.

An alternative mechanism by which protein expression can occur is through transcription initiation by promoter-like sequences present in the transformed plasmid DNA (Figure 3.1). Eukaryotic and prokaryotic promoters consist of different sequence elements that must be present to initiate transcription. Eukaryotic promoters are diverse, consisting of at least ten different classes in addition to various regulatory sequences, making them sometimes difficult to identify and/or characterize (223). In contrast, bacterial promoters have been more extensively characterized. In bacteria, a protein known as a sigma factor associates with RNA polymerase and guides RNA polymerase to the promoter, where the initiation of transcription can occur. Often, regulatory protein(s) also play a role in the recognition of a promoter by RNA polymerase. *E. coli* encodes several different sigma factors (224), the most well known of which is sigma-70. Known as the 'housekeeping' sigma factor,

it is the primary sigma factor involved in the transcription of most genes under normal growth conditions. It associates with RNA polymerase and binds to promoters consisting of two main sequences: a -35 element and a -10 element. Both elements consist of a six nucleic acid consensus sequence, TTGACA for the -35 element and TATAAT for the -10 element, that denote the most frequently occurring base at each position (224, 225). The -35 and -10 elements are optimally separated by 17 bps (226, 227). By chance, a plasmid containing eukaryotic cDNA may contain regions that are similar to -35 and -10 elements, and thus may be recognized by bacterial proteins able to initiate transcription. This is especially likely from eukaryotic genetic material, as eukaryotes have different promoters and transcriptional regulation and thus sequences similar to bacterial -35 and -10 consensus sequences may coincidentally naturally occur. Therefore, foreign genetic material such as cDNAs from eukaryotic organisms may harbor sequences that are able to act as bacterial promoters and self-initiate their expression upon introduction into bacteria. Sequences from eukaryotic genomes that are able to induce transcription in a bacterial host are termed cryptic bacterial promoters (228, 229).

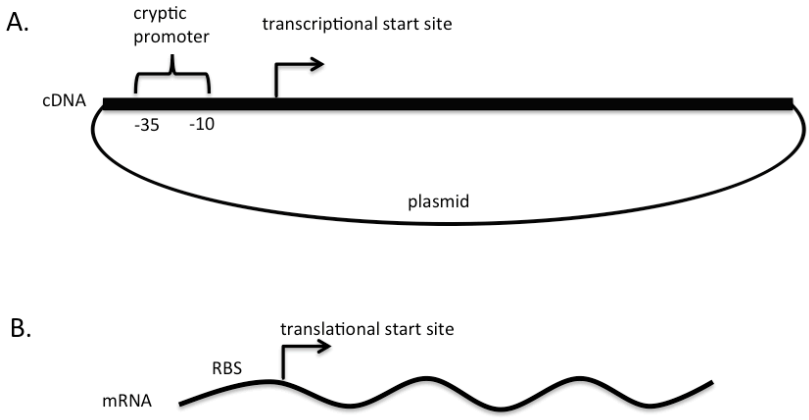


Figure 3.1 Simplified scheme of a cryptic bacterial promoter in a plasmid containing heterologous cDNA. A) Sequences intrinsic to the cDNA that are similar to -35 and -10 elements of bacterial promoters compose a cryptic bacterial promoter. The plasmid backbone lacks promoter sequences, so transcription can only occur through a cryptic promoter. **B)** Protein translation occurs when a start codon (AUG, GUG, UUG) precedes a ribosome binding site (RBS) in a transcript.

After a cryptic promoter induces transcription of an mRNA transcript, translation of the mRNA transcript into a primary amino acid sequence is possible. Bacterial translation occurs when, most commonly, an AUG (start methionine codon) optimally spaced from a ribosome binding site (RBS) on a transcript is encountered by the translational machinery. There are three reading frames for any transcript. The translation of the transcript resulting from a cryptic bacterial promoter can result in a random polypeptide or truncated protein. For example, if translation occurs in the same reading frame as the eukaryotic cDNA, a truncated protein product would be produced. In contrast, if translation occurs from a start codon that is in a different reading frame than the cDNA, a random polypeptide could be produced. Expression of such sequences could be detrimental to the bacterial host, and provide the selective pressure for accumulation of mutated cDNA sequences.

The ABC transporter CFTR provides a notable example of the cryptic promoter phenomenon (230). A seminal paper by Gregory *et al*. characterized such a cryptic bacterial promoter in CFTR cDNA (229). The authors observed that *E. coli* transformed with plasmids containing human CFTR cDNA often yielded plasmids with extensive rearrangements,

indicating that transfected bacteria might be selecting against full length CFTR cDNA. To investigate if any portion of the CFTR cDNA was being unintentionally expressed due to the presence of a cryptic bacterial promoter, plasmids containing segments of CFTR fused to chloramphenicol acetyltransferase (CAT), a gene that confers resistance to chloramphenicol, were generated. Constructs were transformed into *E. coli* and screened for chloramphenicol resistance, as resistance would indicate the expression of CAT due to a promoter present in CFTR cDNA. The authors reported a region of CFTR cDNA between 908 and 936 bp that was able to generate chloramphenicol-resistant *E. coli* clones due to the expression of CAT. To abrogate protein expression of CFTR cDNA in *E. coli*, a T936C silent mutation was introduced into the identified -10 promoter element, therefore generating a cDNA mutant capable of propagating in bacteria (231-234). More recently, a cryptic bacterial promoter was characterized in the cDNA encoding the dengue virus (DENV) 5' untranslated region (UTR) (228). Li *et al.* described the high mutational rate of plasmids containing DENV 5'UTR cDNA and identified sequences of cDNA highly homologous to bacterial -35 and -10 promoter elements.

This chapter investigates the presence of a hypothetical cryptic bacterial promoter intrinsic to *mdr1a* cDNA. Toxicity is surmised by the acquisition of insertion and/or deletion mutations after transformation into bacteria. *In silico* promoter prediction software and GFP fusion proteins are used to identify possible cryptic promoter elements, with the aim of generating mutant *mdr1a* cDNA less toxic to *E. coli*.

3.2 Results

3.2.1 Cloning of *mdr1a* cDNA in *E. coli*

The inability to clone *mdr1a* cDNA in bacteria was repeatedly encountered when performing experiments in this thesis project in the context of chimera generation (Chapter 4). At the onset of this thesis, plasmids containing *mdr1a* (216) and *MDR1* (235) cDNA were obtained from the Laboratory of Cell Biology (NIH). An initial aim of this project was to undertake a pilot study to generate two chimeras and test their functionality as a proof-of-principle before further chimera generation. The chimeras designed for the pilot study consisted of the N-terminal half of human P-gp fused with the C-terminal half of mouse P-gp (termed HHMM), and the reverse chimera consisting of the N-terminal half of mouse P-gp fused with the C-terminal half of human P-gp (termed MMHH) (Figure 4.2). HHMM and MMHH chimeras were to be generated by the introduction of a silent substitution by PCR-mediated mutagenesis to create a ClaI restriction enzyme site in the P-gp linker region of both mouse and human P-gp cDNA (Section 4.2.1.2, 4.2.1.3). Four PCR products were generated (N-terminal *mdr1a*/*MDR1* and C-terminal *mdr1a*/*MDR1*), containing the introduced ClaI site. PCR reactions were analyzed by gel electrophoresis to screen for the presence of an approximately 2 kb PCR fragment. Visual inspection of PCR products indicated that under all PCR conditions assessed, the amplified N-terminal half of *mdr1a* cDNA was approximately 1.5 kb larger than anticipated from the reference DNA sequence (Figure 3.2).

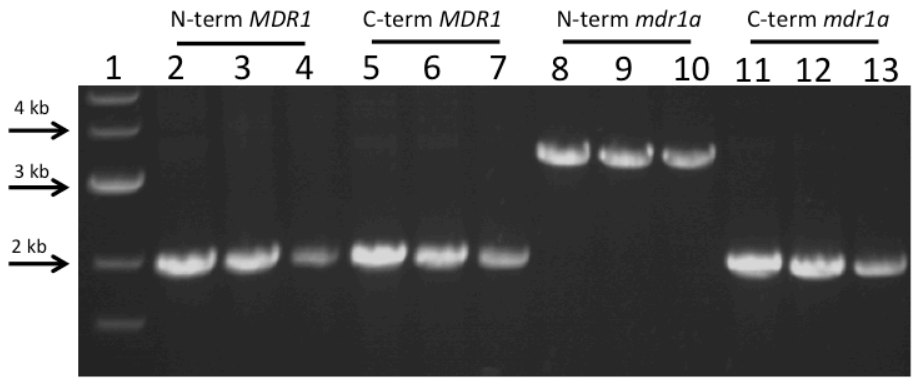


Figure 3.2 Gel electrophoresis of *MDR1* and *mdr1a* N- and C-terminal PCR products. Three PCR reactions (Section 2.2.1) were performed for each half of *MDR1* and *mdr1a* P-gp cDNA to optimize PCR conditions (indicated by the numbers above the lanes) and analyzed by gel electrophoresis (Section 2.2.2). Wild-type *MDR1* and *mdr1a* cDNA are approximately 4 kb. N-term and C-term halves of the cDNA are approximately 2 kb in size. All three PCR reactions of N-term *mdr1a* are 1.5 kb larger than anticipated. Lane 1 contains a DNA size ladder with selected bands labeled with their corresponding sizes.

Investigation into the origin of the 1.5 kb insertion by sequencing revealed the presence of the IS10 insertion sequence (Genbank reference AY583239.1) in N-terminal *mdr1a* cDNA. Thus, it is likely that transposition of IS10 occurred during previous cloning efforts in the Laboratory of Cell Biology (NIH) before the plasmid was obtained. An alternative plasmid containing *mdr1a* cDNA, originally generated by Dr. Philippe Gros (McGill University) (216), was obtained as a kind gift from Dr. Alfred Schinkel (Netherlands Cancer Institute) (41). The *mdr1a* cDNA was provided in a pGEM7 backbone, thus yielding the plasmid pGEM7*mdr1a*.

Upon receipt of the pGEM7*mdr1a* plasmid, *mdr1a* cDNA was sequenced confirmed to contain the expected sequence. The N-terminal and C-terminal *mdr1a* cDNA PCR amplifications described above were repeated and the PCR products analyzed by visual inspection after gel electrophoresis. The PCR products were all approximately 2 kb in size, and thus did not appear to contain any large insertion or deletion mutations. N-terminal and C-terminal *mdr1a* cDNA PCR products were ligated into a pCR-Blunt II-TOPO plasmid,

generating the plasmids pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm, pTOPO_*mdr1a*_Nterm, and pTOPO_*mdr1a*_Cterm (Figure 4.2.1.4), and transformed into *E.coli*. Multiple independent *E. coli* colonies transformed with each plasmid were analyzed first by a series of restriction enzyme digests and subsequently by DNA sequence analysis. Over 20 colonies of *E. coli* transformed with pTOPO_*mdr1a*_Nterm were analyzed by a series of restriction enzyme digests, but only four plasmids demonstrated the expected band pattern and sizes. However, upon sequence analysis, all four were shown to contain mutations that were undetectable by gel electrophoresis and subsequent visual inspection (Figure 3.3). Three contained frameshift mutations (insertion or deletion mutations that would generate an out-of-frame protein if expressed). One contained a single point mutation that introduced a stop codon into *mdr1a* cDNA. All were unsuitable for chimera generation.

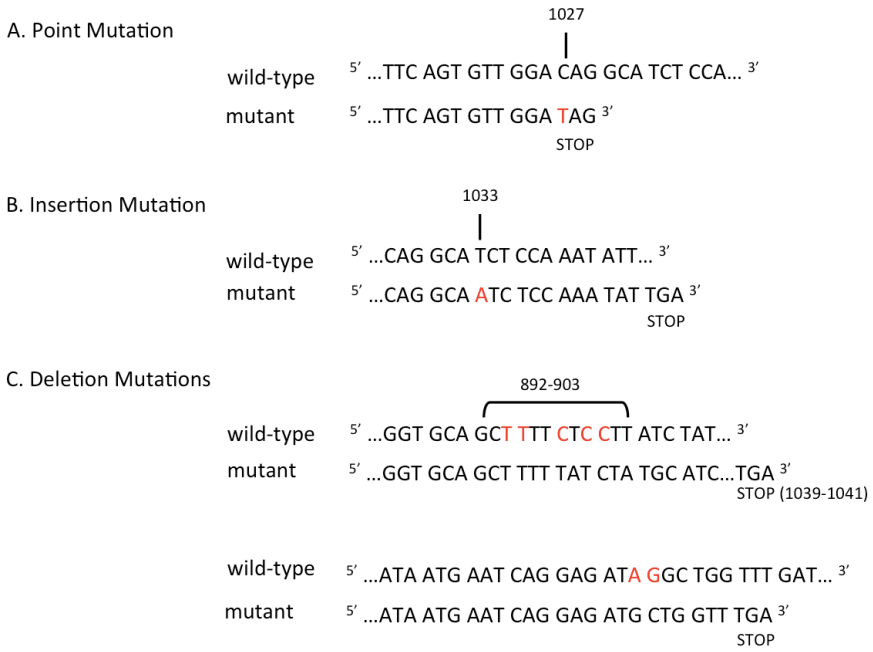


Figure 3.3 Mutations and their position in *md1a* cDNA after transformation into *E. coli*. Mutations observed can be classified as **A)** point, **B)** insertion, or **C)** deletion mutations (red). The location of the mutation is indicated above the wild-type *mdr1a* sequence. All mutations resulted either directly (A) or indirectly (B and C) in the introduction of a stop codon into *mdr1a* cDNA. Mutations were detected by analysis of sequencing chromatograms as compared to the wild-type *mdr1a* sequence (Section 2.2.15).

The inability to find a clone that harbored the expected *mdr1a* cDNA sequence suggested that wild-type *mdr1a* cDNA was somehow exerting a selective pressure on *E.coli*. This is supported by the fact that the original cDNA received for this study had acquired the IS10 element, presumably during routine culturing that occurred prior to its acquisition. Additionally, the nature of these mutations indicates a possible basis for *mdr1a* cDNA toxicity. Due to alteration of the open reading frame (ORF), frameshift mutations often result in a truncated protein product, as any naturally occurring sequences of TGA, TAA, or TAG (stop codons) are now in frame and able to terminate translation (Figure 3.3). Therefore, if *mdr1a* were expressed, all of the observed mutations would result in a severely truncated protein. The observation that only sub-clones with frameshift mutations or inserted stop codons in *mdr1a* cDNA were able to survive transformation may indicate that expression of mouse P-gp occurred and caused deleterious effects to the host bacteria. As the cloning plasmid used was not known to contain a bacterial promoter, it was hypothesized that expression was due to a cryptic bacterial promoter present in *mdr1a* cDNA.

3.2.2 Literature review of *mdr1a* cDNA toxicity in bacterial hosts

As efforts to clone *mdr1a* in bacteria were unsuccessful at the onset of this project, other attempts to clone *mdr1a* cDNA in *E. coli* were investigated. A literature search was conducted to determine if other laboratories have 1) been able to successfully clone *mdr1a* in *E. coli*, 2) determine their methods for successful cloning, and 3) investigate any reported toxicity of *mrd1a* in bacteria. Several early papers used cloning as a method for the propagation and amplification of plasmids containing *mdr1a* (41, 90, 206). Unfortunately, published literature often did not report the exact cloning conditions (bacterial host strain used, bacterial growth conditions, etc.) and it was not possible to repeat the cloning strategy employed for the generation of chimeras. Additionally, most literature characterizing *mdr1a* was published 20 to 30 years prior to this thesis, and it may be that various factors that

allowed for the cloning of *mdr1a* in those reports are no longer applicable. For example, plasmids used in earlier studies are now often considered to be 'low-copy number vectors' with plasmids present in bacterial cells in lower numbers than current commercially available plasmid vectors; and lower plasmid copies per cell would be expected to decrease the deleterious effects of a cryptic promoter if present in *mdr1a*. Also, at their time of publication, sequence analysis was not standard practice, and it is possible that the constructs described may have contained unidentified or uncharacterized mutations. Nonetheless, the limited published literature does not describe or characterize difficulty in cloning *mdr1a*.

A series of papers by Bibi *et al*. report methods for the expression of *mdr1b* (also termed mdr1), which is 83% homologous to *mdr1a*, using *E. coli* (146, 147, 236). This suggests that, even if *mdr1a* was expressed in the presence of a cryptic promoter, the protein product would not be harmful to bacteria (237). However, close inspection of these papers reveals multiple insights that would suggest the converse is true. For example, the authors screened a variety of bacterial host strains to identify a strain that was able to withstand *mdr1b* expression (147). Additionally, the authors noted that 'high-copy number' plasmids were not suitable for the expression of *mdr1b*, stating that high copy vectors increase toxic effects of *mdr1b* in bacteria (147). The authors also noted the presence of two distinct colony morphologies after bacterial transformation with plasmids containing *mdr1b*: large colonies that contain mutated *mdr1b* and small, slow-growing colonies containing wild-type *mdr1b*. Taken together, these data suggest that while certain bacterial strains and growth conditions may allow for the expression of *mdr1b*, its expression does exert a toxic effect when expressed in bacteria. Lastly, the authors attempted the expression of *mdr1a* in *E. coli* with various host strains, but several rearrangements of *mdr1a* cDNA were observed and expression could not be obtained (personal communication, E. Bibi).

3.2.3 Personal communications regarding *mdr1a* toxicity

While the literature is unclear regarding potential toxicity issues with *mdr1a*, various personal communications with investigators who have attempted cloning of *mdr1a* was conducted. Alfred Schinkel (Netherlands Cancer Institute) and Ina Urbatsch (Texas Tech University), two investigators who have published studies sub-cloning *mdr1a*, described that they also observed *mdr1a* toxicity in bacteria (A. Schinkel, personal communication; I. Urbatsch, personal communication) (41, 103). Both investigators described protocols that were optimized in their laboratories to circumvent *mdr1a* toxicity. For example, Urbatsch was able to successfully clone *mdr1a* in low-salt formulated LB agar and LB medium. Additionally, only small colonies of transformed bacteria were selected (as large colonies most likely contain mutated cDNA) and shaking cell culture was harvested at a low OD (<1.0) to decrease the likelihood of acquiring mutations. Schinkel also indicated that a prolonged growth period increased the frequency of mutations, and their optimized protocol consisted of selecting various sub-clones, growing them to a low density of bacterial culture, and then pooling the sub-clones to generate suitable quantities of plasmid DNA. While these personal communications indicate that variations of standard cloning protocols can be used to successfully clone *mdr1a* in *E. coli*, the basis of *mdr1a* experimental toxicity in *E. coli* has never been characterized and these protocols have never been published.

3.2.4 FlexProm *in silico* analysis of bacterial promoter elements

To investigate the presence of a cryptic promoter in *mdr1a* cDNA, the FlexProm algorithm was used to search for potential bacterial promoters (238). Additionally, the algorithm can identify sequences capable of initiating translation by identifying sequences similar to the bacterial RBS consensus sequence, also known as the Shine-Dalgarno sequence (AGGAGG with an optimal spacing 4-8 bps upstream of the translational start methionine) (239). By the identification of predicted transcriptional and translational

initiation sites, the FlexProm algorithm detects potential ORFs and specifies the resulting polypeptide.

The algorithm identified a sequence in *mdr1a* cDNA capable of acting as a bacterial promoter (Figure 3.4.A). This sequence consists of a weak -35 element in combination with a strong -10 element, separated by 17 bps. While this promoter is theoretically capable of initiating transcription, to support the hypothesis that *mdr1a* toxicity is due to the expression of P-gp, a RBS and translation start site must also be present in the cDNA. The algorithm identified a RBS ideally spaced upstream of a methionine start (ATG) codon (Figure 3.4.B). Additionally, the methionine is in-frame with regard to the ORF of *mdr1a* cDNA. In other words, if translation were initiated at this location, P-gp would be expressed in *E. coli* initiating at amino acid 107 and terminating at the native stop codon after amino acid 1276.

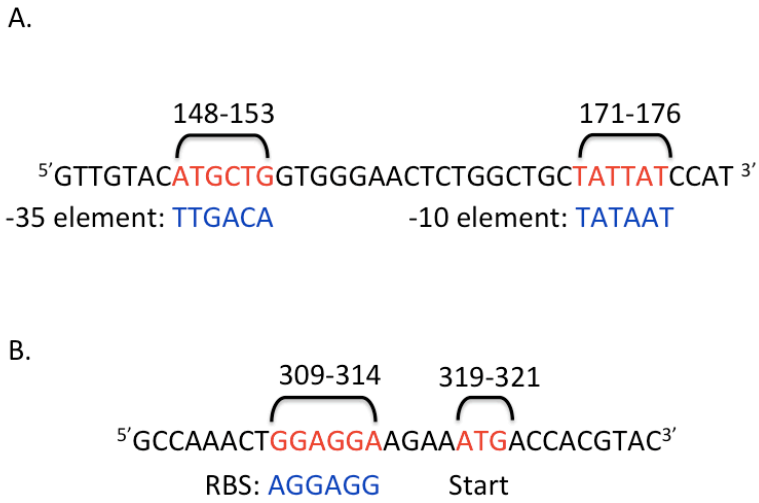


Figure 3.4 *In silico* identification of transcriptional and translational start sites in *mdr1a* cDNA. A) Sequence of *mdr1a* cDNA shown with predicted -35 and -10 elements (red) and corresponding consensus sequences (blue). The location in the cDNA for each element is noted above the sequence. **B)** Predicted RBS (red) and consensus sequence (blue) upstream, of a methionine that is in-frame with regards to the *mdr1a* ORF.

As human *MDR1* cDNA did not acquire mutations during cloning of PCR products in *E. coli*, it was of interest to compare *mdr1a* and *MDR1* cDNA sequences in the region of the predicted -35 and -10 elements (Figure 3.5.A). While sequence similarity is observed, the FlexProm algorithm did not predict that the corresponding *MDR1* sequence would act as a bacterial promoter. This is due to sequence differences between *MDR1* and *mdr1a* in the region of the predicted -10 element that result in the elimination of predicted promoter. The potential RBS and translational start site are retained in both *MDR1* and *mdr1a* (3.5.B). However, with the absence of an upstream promoter, there would be no transcription and therefore no protein translation.

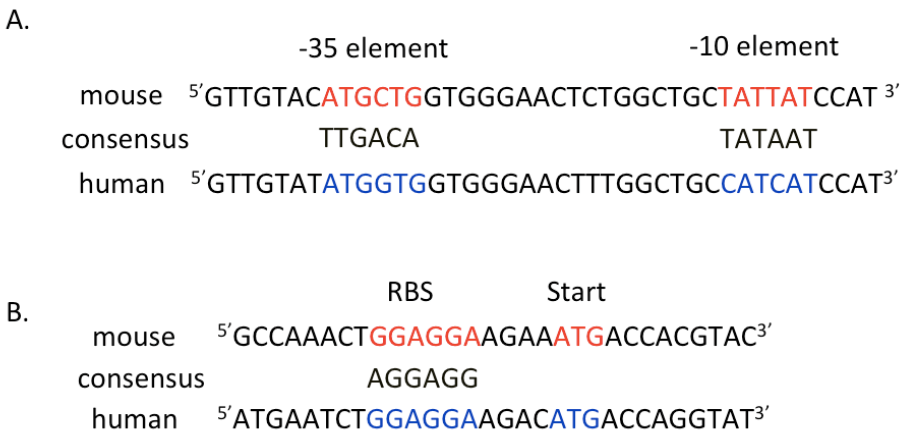


Figure 3.5 Alignment of *mdr1a* and *MDR1* cDNA sequences at predicted transcriptional and translational start sites. A) Sequence alignment of *mdr1a* (top, red) and *MDR1* (bottom, blue) cDNA at the predicted -35 and -10 bacterial promoter elements. **B)** Sequence alignment of *mdr1a* and *MDR1* cDNA at the predicted transcriptional start site. Consensus sequences for -35 and -10 promoter elements and RBS sequences are indicated.

3.2.5 Construction of *mdr1a*-GFP fusion constructs

To investigate the cryptic promoter predicted by the *in silico* analysis, a series of plasmids were generated containing truncations of *mdr1a* cDNA fused to GFP (Figure 3.6). Fusion constructs were designed in collaboration with Dominic Esposito (Frederick National Laboratory for Cancer Research, FNLCR) and assembled by the Protein Expression Laboratory at FNLCR. Plasmids were designed to investigate several aspects of the ability of *mdr1a* cDNA to initiate transcription and translation in bacteria. As the FlexProm algorithm predicted a bacterial promoter between 141 and 177 bp, plasmid p300-GFP was generated to determine the ability of sequences in the first 300 bps of *mdr1a* cDNA to promote GFP expression. Note that this construct contains an introduced RBS and start methionine to initiate GFP translation, and is therefore only reporting on the ability of the first 300 bps of *mdr1a* cDNA to function as a bacterial promoter. In addition to a cryptic promoter that would initiate transcription, translation must also occur for the expression of mouse P-gp. GFP fusions p321-GFP and p321-stop-GFP were generated to investigate the ability of *mdr1a* to express GFP (Figure 3.6). p321-GFP is composed of *mdr1a* cDNA 1-321 bp, which includes the predicted cryptic promoter (*mdr1a* bases 148-176), RBS (*mdr1a* bases 309-314), and the mouse P-gp methionine at amino acid 107 (*mdr1a* bases 319-321) fused in-frame with GFP. GFP expression observed from this construct after transformation into bacteria would suggest that both the transcriptional and translational sequences identified by the FlexProm analysis are functional. The p321-stop-GFP construct includes all of the elements of p-321-GFP, but with the addition of a stop codon directly preceding the mouse P-gp methionine at amino acid 107. This would be expected to abrogate GFP expression.

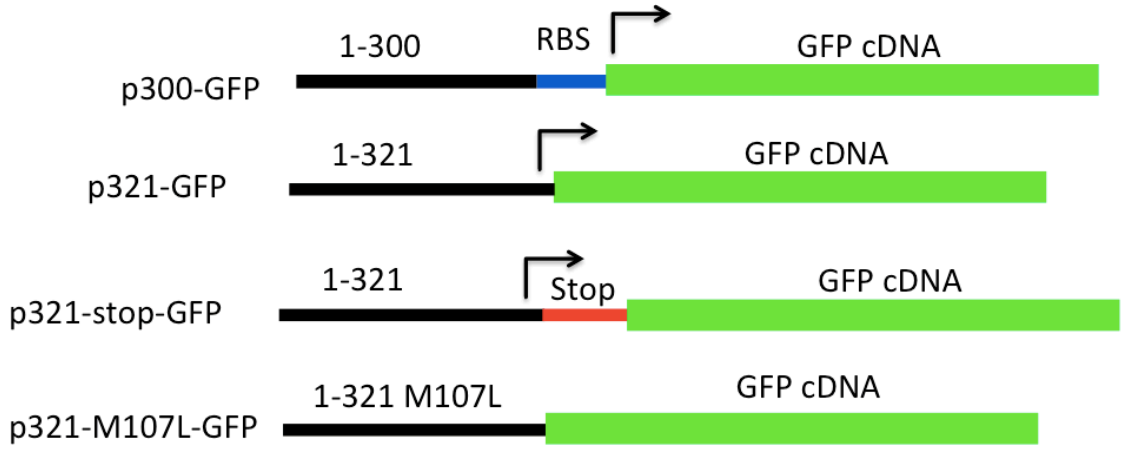


Figure 3.6 GFP plasmid constructs generated to test for cryptic promoter activity of *mdr1a*. Black lines represent *mdr1a* cDNA sequences. Black arrows represent the presence of a start methionine (ATG) codon. The presence of introduced RBS and Stop sites are shown where applicable. Numbering relates to the sequences of *mdr1a* fused upstream of GFP. Plasmids were constructed by the Protein Expression Laboratory at FNLCR.

Additionally, p321-M107L-GFP was constructed to encode a leucine rather than a methionine at position 107 in the amino acid sequence (Figure 3.6). This was generated in an attempt to abrogate GFP expression by removing the translational start site, and investigate if construction of the mutated form of *mdr1a* harboring a M107L mutation would be desirable. This mutational strategy makes the assumption that it is the expression of the truncated mouse P-gp protein, as opposed to another function of the cryptic *mdr1a* transcript, which is leading to undesirable effects in *E. coli* during sub-cloning. To determine if a mutation at residue 107 in *mdr1a* would affect protein function, the residue was located in the mouse P-gp crystal structure (Figure 3.7). Residue 107 is located at the boundary of extracellular loop 1 (ECL1) and transmembrane helix 1 (TM1). In this model, residue 107 does not face the drug-binding pocket, and would not be expected to influence function. Leucine was chosen as a replacement residue for methionine as both are residues with hydrophobic side chains that are relatively similar in size, and the codon for leucine (CTG) is not expected to efficiently initiate translation (Figure 3.7.C).

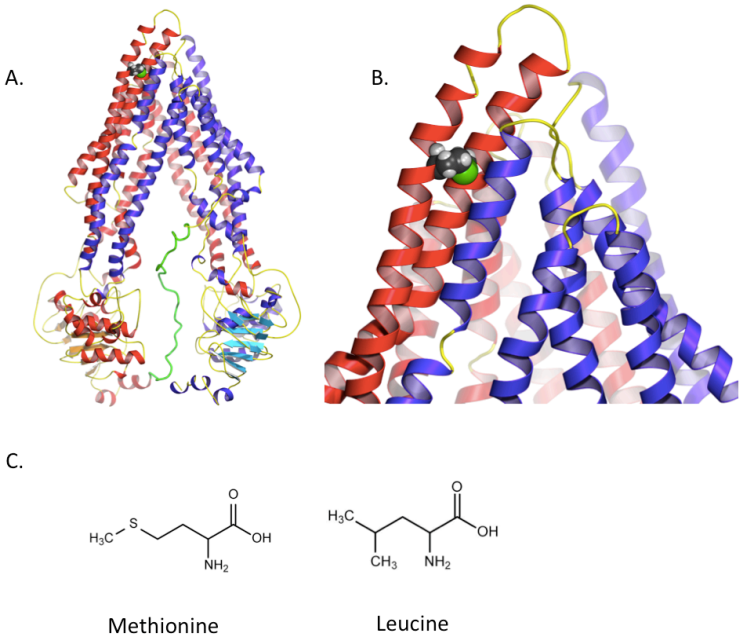


Figure 3.7 Mouse P-gp model with methionine residue at position 107 highlighted. A) Mouse P-gp is shown with the methionine to leucine mutation highlighted in TM1 by the presence of a space filled leucine residue. **B)** Enlarged view of M107L mutation. **C)** Structures of methionine and leucine. Figures were generated using PyMOL from Protein Data Bank identification 4M1M.

3.2.6 Fluorescence screening of transcriptional *mdr1a*-GFP fusion constructs in *E. coli*

The p300-GFP, p321-GFP, p321-stop-GFP and p321-M107L-GFP plasmids were used to transform *E. coli* along with plasmid M06 as a negative control plasmid containing GFP and a stuffer region in lieu of a promoter and a positive control plasmid M07 that contains GFP downstream from the strong lambda pR promoter (M07). Five colonies were selected from each transformation and were grown in 5 mL small-scale culture in LB media for 18 hours, after which cells were pelleted, LB media removed, and resuspended in PBS to eliminate any potential background fluorescence due to growth media. Cells were analyzed by confocal microscopy (Figure 3.8) and fluorescence spectroscopy (Figure 3.9).

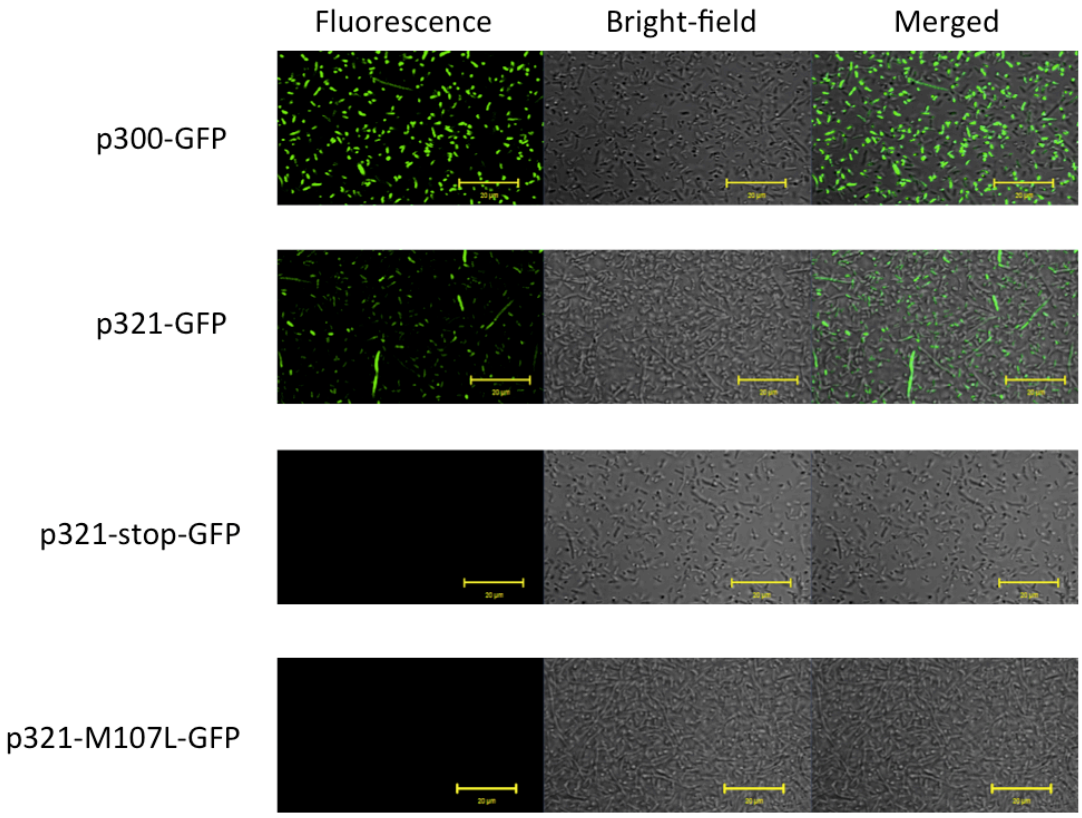


Figure 3.8 Confocal images of *E. coli* transformed with GFP fusion plasmids. Single colonies of bacteria transformed with the GFP fusion plasmids shown were selected for overnight growth in 5 mL culture. Cells were transferred to chambered coverslips and GFP expression was analyzed by confocal microscopy (Section 2.2.29). From left to right, columns represent fluorescence, bright field, and a merged image. Yellow bars indicate 20 µm.

Confocal microscopy was used to qualitatively determine levels of GFP expression in bacteria transformed with the p300-GFP, p321-GFP, p321-stop-GFP, and p321-M107L-GFP plasmids (Figure 3.8). Bacteria transformed with the p300-GFP and p321-GFP plasmids exhibited GFP expression. In contrast, fluorescence was not observed in bacteria transformed with the p321-stop-GFP and p321-M107L-GFP plasmids. Confocal images of transformed bacteria were in qualitative agreement with fluorescence levels observed by fluorescence spectroscopy.

Fluorescence spectroscopy was used to quantify the amount of GFP expression in transformed bacteria (Figure 3.9). Bacteria transformed with the p300-GFP plasmid

displayed significantly higher GFP expression as compared to M06 ($p \leq 0.0001$), indicating that the *mdr1a* cDNA sequence from 1-300 bp contains promoter elements able to initiate GFP transcription. The p321-GFP plasmid was used to determine if sequences 1-321 bp of *mdr1a* were able to express GFP by both the production of a transcript and subsequent translation into a protein product. Fluorescence intensity of the bacteria transformed with p321-GFP was also significantly higher than M06 ($p \leq 0.0001$). However, p321-GFP was also significantly lower than observed with p300-GFP ($p \leq 0.0001$). This is most likely due to the fact that the p300-GFP plasmid contains an introduced strong RBS with ideal positioning relative to the transcriptional start site. In contrast, the p321-GFP is predicted by the FlexProm analysis to contain a weaker RBS with less ideal spacing between the RBS and the transcriptional start site. The presence of an introduced stop codon after 1-321 bp of *mdr1a* cDNA in plasmid p321-stop-GFP abrogated GFP expression.

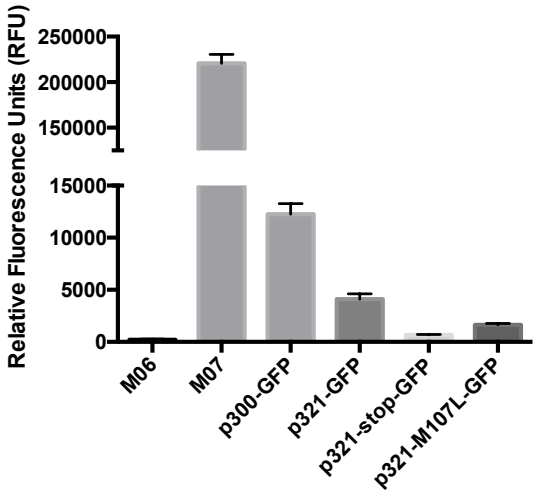


Figure 3.9 Fluorescence of *E. coli* transformed with GFP fusion plasmids. Fluorescence spectroscopy was used to determine levels of GFP fluorescence in bacteria transformed with GFP plasmids (Section 2.2.28). Fluorescence intensities of bacteria transformed with the named GFP fusion plasmids are shown. Data represent the mean and standard deviation of five individual colonies from one transformation event. Multiple groups were analyzed using one-way ANOVA where significance was defined as $p < 0.05$ in GraphPad Prism (Section 2.2.44).

GFP fusion plasmids containing the M107L mutation had significantly lower GFP expression than bacteria transformed with the p300-GFP and p321-GFP plasmids ($p \leq 0.0001$), indicating that introduction of the M107L mutation to full-length *mdr1a* may reduce toxicity and generate a more stable *mdr1a* cDNA. However, the M107L mutation did not completely abrogate GFP expression, as GFP levels in bacteria transformed with p321-M107L-GFP contained higher GFP expression than bacteria transformed with the negative control M06 plasmid DNA ($p \leq 0.01$). Various explanations could account for GFP expression by the p321-M107L-GFP plasmid, such as the presence of another codon that may be able to initiate translation. For example, while ATG is the most common start codon, used to initiate translation in approximately 83% of genes in *E. coli*, GTG and TTG initiate translation in 14% and 3% of genes, respectively (240). Additionally, translation of at least one gene in *E. coli* uses CTG (the codon used to generate the leucine mutation) as a start codon for translation (241).

3.2.7 Transformation of *E. coli* with *mdr1a*_M107L cDNA

While *mdr1a*-GFP fusion plasmids containing the M107L mutation had significantly less GFP expression, the critical question was whether the M107L mutation would reduce *mdr1a* cDNA-associated toxicity. Full-length *mdr1a* cDNA was mutated to introduce the M107L mutation to generate *mdr1a*_M107L cDNA. Next, *mdr1a*_M107L was ligated into a pDest-008 plasmid to generate the pDEST_*mdr1a*_M107L plasmid (constructed by the Protein Expression Laboratory at FNLCR). The pDEST_*mdr1a*_M107L plasmid was used to transform *E. coli* and transformed bacteria were streaked onto LB agar plates containing a selection agent. Three colonies were selected for small-scale culturing and subsequent plasmid purification and sequencing. All three pDEST_*mdr1a*_M107L purified plasmids contained non-mutated cDNA, with the exception of the introduced M107L mutation (data

not shown). Elimination of the translational start site associated with the cryptic bacterial promoter appeared to reduce the toxicity of *mdr1a* cDNA to bacteria.

3.2.8 Functional characterization of M107L P-gp

To show that M107L P-gp was functional similar to wild-type P-gp, a full functional characterization was completed. All plasmids generated for the functional characterization of M107L P-gp were designed by the author and generated by the Protein Expression Laboratory at FNLCR. Additionally, the generation of recombinant baculovirus (Section 2.2.30) and protein expression in insect cells (Section 2.2.31) was completed by the Protein Expression Laboratory at FNLCR. All subsequent steps were completed by the author in the Laboratory of Cell Biology (NIH).

Two main systems were used for the characterization of P-gp mutants and chimeras in this chapter: transient transfection of P-gp into HEK 293 mammalian cells for whole cell-based flow cytometry functional assays and ATPase assays performed with P-gp-containing crude membranes isolated from insect cells infected with recombinant baculovirus. For characterization of M107L P-gp in mammalian cells, *MDR1* and *mdr1a*_M107L were sub-cloned into a plasmid containing a cytomegalovirus (CMV) promoter to generate pDEST_Human and pDEST_M107L plasmids (Section 2.2.12). pDEST_Human and pDEST_M107L plasmids were transiently transfected into HEK 293 cells by lipofectamine transfection (Section 2.2.33). Western blotting (Figure 3.10) and flow cytometry assays (Figure 3.11) were carried out on transfected cells to determine the expression and functional activity of M107L P-gp, respectively.

Characterization of M107L P-gp was carried out in comparison to human P-gp. A more accurate assessment of M107L function would be to compare the functional activity of M017L P-gp to wild-type mouse P-gp, as species differences may occur naturally between mouse and human P-gp, and these variables may confound functional differences resulting

from the M107L mutation. However, due to mutations in wild-type *mdr1a* cDNA aquired during cloning, the plasmids required to carry out functional assays could not be generated. Therefore, all functional assays carried out used human P-gp as a control for M107L P-gp expression and function.

Western blotting was performed to determine if M107L P-gp was expressed at a level comparable to wild-type human P-gp. The C219 primary antibody was used for western blotting, as it recognizes an epitope conserved in all mammalian P-gps, including mouse and human P-gp (242). P-gp was not detected in untransfected HEK 293 cells. HEK 293 cells transfected with pDEST_Human and pDEST_M107L highly expressed human and M107L P-gp, respectively (Figure 3.10).

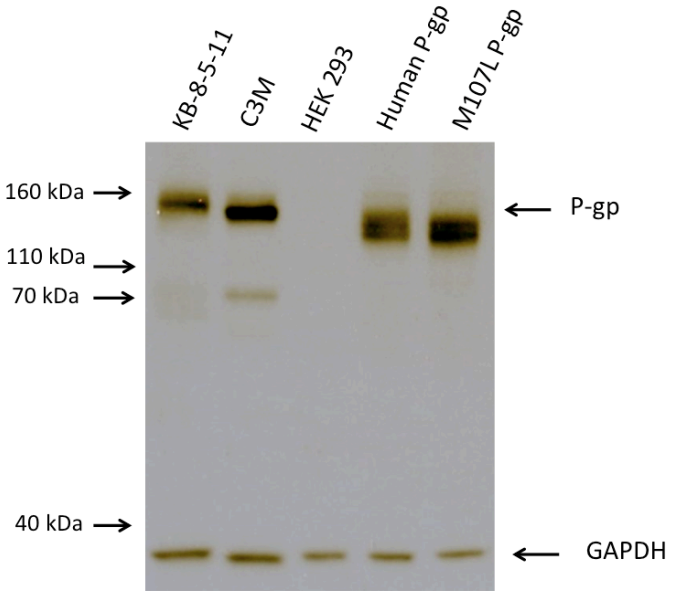


Figure 3.10 Western blot of HEK 293 cells transfected with pDEST_Human and pDEST_M107L P-gp. HEK 293 cells were transiently transfected with pDEST_Human and pDest_M107L plasmids (Section 2.2.33). Protein (5 µg) was subjected to SDS-PAGE (Section 2.2.19) and subsequent western blotting (Section 2.2.20) with the C219 primary antibody. Additionally, the drug-selected cell lines C3M and KB-8-5-11 over-expressing wild-type mouse and human P-gp, respectively, were included. GAPDH was used as a loading control (37 kDa).

Flow cytometry efflux assays were completed to determine functionality of M107L P-gp compared to human P-gp (described in Section 5.1.1). Rh123 is a fluorescent compound that is a substrate of mouse and human P-gp (Figure 5.8). HEK 293 cells transfected with human and M107L P-gp should efflux similar amounts of Rh123. Non-transfected HEK 293 cells should accumulate high intracellular levels of Rh123, as these cells lack P-gp-mediated efflux. The extent of P-gp-mediated efflux can be observed by the co-incubation of transfected cells with both Rh123 and the P-gp inhibitor TQR.

Non-transfected HEK 293 cells, which do not express P-gp (Figure 3.10), accumulated high levels of Rh123 (Figure 3.11). HEK 293 cells transfected with human and M017L mouse P-gp effluxed Rh123 to similar extents, as indicated by the low level of cellular fluorescence (a shift of the histogram towards the left). Additionally, when cells transfected with human or M107L P-gp were co-incubated with Rh123 and TQR, P-gp-mediated efflux was abrogated, allowing Rh123 to accumulate in cells, as observed by the increased intracellular fluorescence to non-transfected levels. Thus, M107L P-gp was shown to be expressed and functional at a level similar to that of wild-type human P-gp.

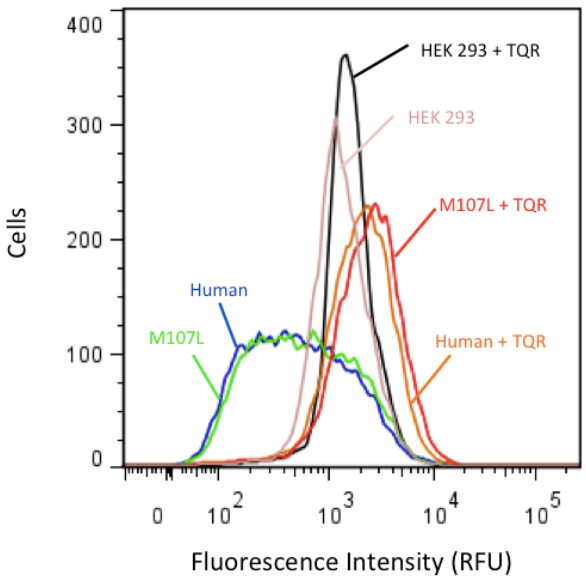


Figure 3.11 Rh123 efflux assay of HEK 293 cells transfected with pDEST_Human and pDEST_M107L P-gp. HEK 293 cells were transfected with pDEST_Human and pDest_M107L plasmids expression M107L P-gp and human P-gp (Section 2.2.33). Cells were harvested and Rh123 efflux assays were performed (Section 2.2.24). Transfected HEK 293 cells were incubated with Rh123 either alone (M107L P-gp, green; human P-gp, blue) or in combination with 100 nM of the P-gp inhibitor TQR (M107L P-gp, red; human P-gp, orange). Untransfected HEK cells were also treated with either Rh123 alone (HEK 293, pink) or in combination with TQR (HEK 293, black). Histograms shown are representative of three independent studies. The mean fluorescence intensity of the histogram was determined by FloJo Single Cell Analysis Software.

ATPase activity is required for the transport of substrates by P-gp. ATPase assays were carried out to determine biochemical differences in the ATP binding and hydrolysis between human and M107L P-gp (ATPase assays are described in Section 5.1.2). ATPase assays were conducted on cell crude membranes from insect cells infected with recombinant baculovirus expressing either human P-gp or M107L P-gp (Figure 3.12). Various concentrations of verapamil were tested to measure ATPase activity of M107L and human P-gp (Figure 3.12.A). Interestingly, M107L P-gp exhibited higher maximal fold-stimulation than human P-gp. As the M107L mutation is not located in the drug-binding pocket and therefore not predicted to influence drug transport (Figure 3.7), this result may reflect a species

difference between mouse and human P-gp. To determine if differences in ATPase activity were observed with additional compounds, a panel of seven known P-gp ATPase stimulators were screened. The compounds cis-flupenthixol, CAM, nicardipine, paclitaxel, vincristine, Rh123, and loperamide were used to probe for activity of M107L P-gp compared to human P-gp (Figure 3.12.B). Human and M017L P-gp showed similar levels of ATPase stimulation, indicating that M017L is functional and can be stimulated to levels similar to those of human P-gp. Overall, while M107L P-gp displayed higher levels of ATPase activity with certain substrates, M107L P-gp functionality was deemed similar to that of wild-type P-gp.

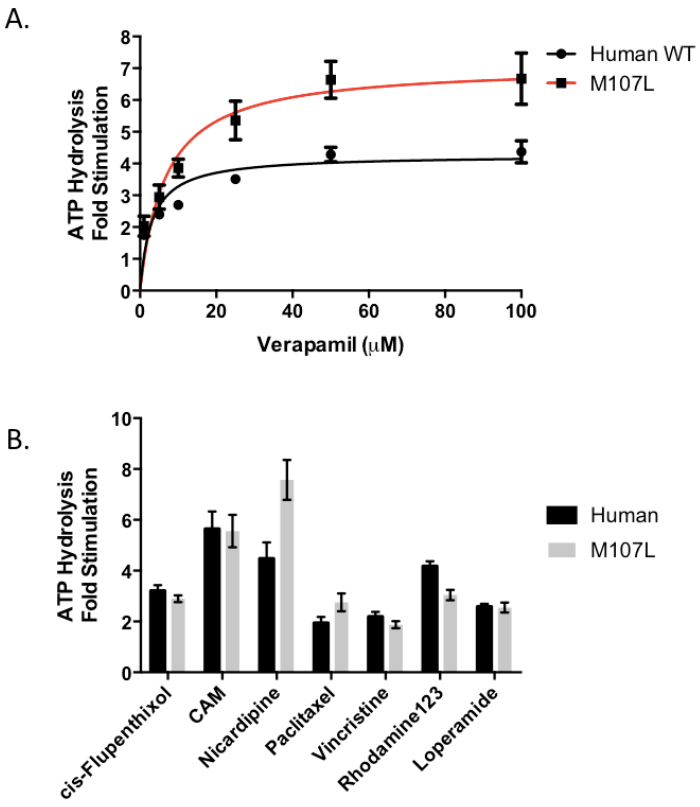


Figure 3.12 ATPase of human P-gp and M107L P-gp. A) Dose response ATPase assays of M107L P-gp compared to human wild-type P-gp in the presence of 0, 1, 5, 10, 25, 50 and 100 µM of verapamil (Section 2.2.26). ATPase activity is reported as the fold-change in the rate of ATP hydrolysis over basal levels (Section 5.1.2). **B)** A panel of seven compounds that are known substrates of P-gp were selected to determine activity of M107L P-gp compared to human P-gp. The following concentrations were used for each assay: 20 µM cis-flupenthixol, 10 µM CAM, 20 µM nicardipine, 20 µM paclitaxel, 20 µM vincristine, 500 µM Rhodamine 123, and 20 µM loperamide. Data represents average and standard deviation of technical duplicates from three individual assays.

3.3 Discussion

The results described in this chapter demonstrate the presence of a cryptic bacterial promoter present in *mdr1a* cDNA. The existence of a cryptic bacterial promoter in *mdr1a* cDNA was hypothesized due to the propensity of plasmids containing *mdr1a* to become mutated upon transformation into *E. coli*. The repeated appearance of mutated *mdr1a* plasmid DNA upon transformation suggested that there was a selection against *E. coli* transformants containing wild-type cDNA. Selection of transformed *E. coli* against plasmids carrying wild-type *mdr1a* indicated that the cDNA contained an element that is toxic or lethal to bacteria, and mutation of the harmful element mitigates this toxicity. One mechanism by which eukaryotic cDNA can exert toxic effects in bacteria is by the expression of a foreign protein that is detrimental to bacteria. Experiments described in this chapter investigated the presence of a cryptic bacterial promoter in *mdr1a* cDNA that causes unintentional expression of truncated mouse P-gp. *In silico* promoter prediction software was used to identify both transcriptional and translational start sites in *mdr1a* cDNA (Figure 3.4). These data provided the basis for the generation of various *mdr1a*-GFP fusion constructs (Figure 3.6), which showed that sequences within the first 321 bps of *mdr1a* are able to express GFP upon transformation into *E. coli*. Additionally, mutation of the translation start site from methionine to leucine (M107L) in *mdr1a* cDNA was able to reduce mutations generated during cloning.

While the generation of M107L-*mdr1a* cDNA may have reduced overall toxicity, it is likely that low-level expression from the cryptic promoter still occurs. This is indicated by the inability of the M107L mutation to completely abrogate GFP expression, as seen by bacteria transformed with p321-M107L-GFP having GFP expression levels significantly higher than negative control plasmids ($p \leq 0.01$) (Figure 3.9). In this study, plasmids purified from *E. coli* transformed with *mdr1a*-M107L did not contain mutations typically seen with plasmids containing wild-type *mdr1a*. However, only three colonies were screened and plasmids were

purified from a small-scale (5 mL) culture. As mutations may be more likely to occur under certain cloning conditions, such as during large-scale culturing or in different plasmids or bacterial host strains, it may be necessary to mutate additional sequences of *mdr1a* cDNA, such as the promoter itself, to fully abrogate expression.

One potential shortcoming of this study is that it assumes expression of mouse P-gp is harmful to bacteria. While this study did not investigate toxic effects of mouse P-gp expression, previous unpublished studies in the Laboratory of Cell Biology (NIH) attempted to express human P-gp in *E. coli* and observed a 500 to 1000-fold reduction in the ability of transformed bacteria to form colonies, indicating that expression of human P-gp is lethal to *E. coli*. Additionally, personal communications with E. Bibi, who reported expression of *mdr1b* in *E. coli*, indicated the expression of *mdr1a* in *E. coli* with various host strains was attempted but could not be obtained. Lastly, it would be difficult to account for how the presence of a plasmid in bacteria could exert toxic effects if it were not for expression of a foreign protein or polypeptide, as the presence of plasmid DNA itself is known not to be detrimental to the strain of bacteria used in this study.

Personal communications with investigators who have reported cloning *mdr1a* cDNA have indicated unpublished, specialized protocols were required to reduce the harmful effect of *mdr1a* cDNA toxicity in *E. coli*. Although helpful for the purposes of sub-cloning, these protocols do not address the root cause of the genetic instability of plasmids containing *mdr1a* cDNA. This study identifies the presence of a cryptic bacterial promoter in *mdr1a* cDNA able to initiate protein expression in the absence of a promoter upstream of the cDNA. Reduction of toxicity is observed by a single point mutation from a methionine to a leucine residue at position 107. These studies are the basis for the generation of a *mdr1a* cDNA mutant that, upon transformation, is benign to bacteria and can be propagated in bacteria with standard molecular cloning protocols.

3.4 Conclusions

This chapter presents the identification and characterization of a cryptic bacterial promoter in *mdr1a* cDNA. While the genetic instability of plasmids containing *mdr1a* cDNA has been previously observed by numerous laboratories, no study has formally 1) reported the detrimental effects of plasmids containing *mdr1a* cDNA upon transformation into *E. coli*, or 2) characterized the basis of such effects. The experiments described in this chapter have identified the presence of sequences intrinsic to *mdr1a* cDNA that are able to initiate transcription and translation and thus induce protein expression. Lastly, this study reports an *mdr1a*-M107L mutant that is more genetically stable during cloning and retains the functional characteristics of wild-type P-gp. This mutant may prove useful for laboratories wishing to study *mdr1a* in the future.

Chapter 4: Generation of P-glycoprotein chimeras

4.1 Introduction

The first stage of this thesis project was the generation of chimeras from mouse (*mdr1a*) and human (*MDR1*) P-gp cDNA. Despite a high degree of homology, it is currently unknown why mouse P-gp can be crystallized while human P-gp cannot (Figure 4.1). This thesis project uses the generation of chimeras as an approach to investigate whether specific protein domains are responsible for differences in crystallizability between mouse and human P-gp. Various approaches could be employed to generate chimeras. For example, they could be made by mutating specific residues thought to be important for crystallization in human P-gp to the homologous mouse residue. However, this approach assumes knowledge of which residues contribute to the favorable crystallographic properties of mouse P-gp; knowledge that is not currently available. In this chapter, chimeras were generated by exchanging homologous mouse and human P-gp domains. By the crystallization screening of chimeras consisting of variable amounts of mouse P-gp, the aim was to identify essential domains that contribute to differences in the crystallizability between mouse and human P-gp.

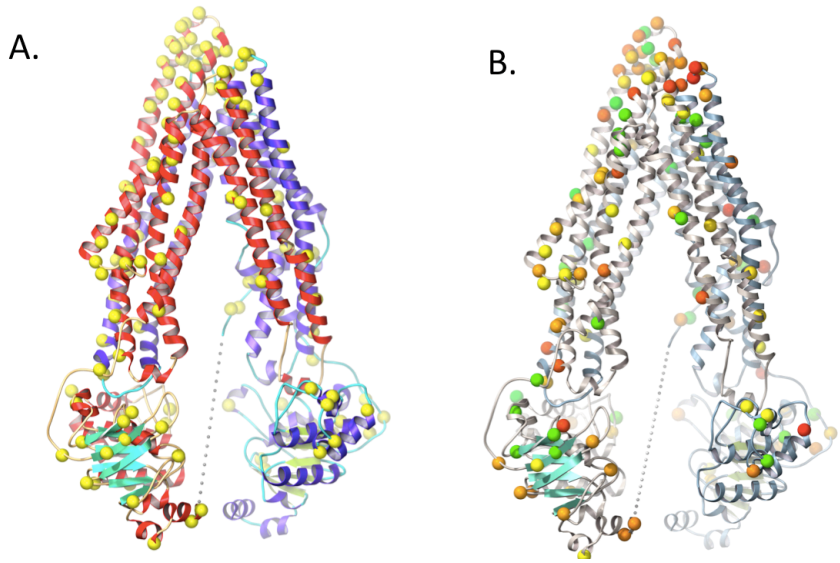


Figure 4.1 Models of mouse P-gp with non-homologous human amino acid residues highlighted. A) Non-homologous mouse and human P-gp amino acid residues are indicated by the presence of a yellow sphere. **B)** Non-homologous residues can be either similar or dissimilar in properties such as size, polarity, and charge. Green and yellow spheres indicate similar non-homologous residues, while orange and red spheres indicate dissimilar non-homologous residues. Color-coding was calculated based on the PAM30 substitution matrix and figures were generated in PyMOL using PDB deposit 4M1M.

Mouse and human P-gp consist of two TMDs and two NBDs (Sections 1.1.1 and 1.3.1). Human and mouse P-gp proteins are each formed by the folding of a single amino acid polypeptide, which is transcribed and translated in the following order: (N-term) TMD1-NBD1-TMD2-NBD2 (C-term). In this thesis project, a nomenclature to describe chimeras was devised by attributing the first letter from the cDNA of the species of origin, H or M for human or mouse P-gp, respectively, in a particular domain to the order that it is present in the polypeptide. For example, a chimera that is composed of human TMD1 and NBD1 and mouse TMD2 and NBD2 would be termed a HHMM chimera. Wild-type human and mouse P-gp are termed HHHH_WT and MMMM_WT, respectively. Figure 4.2 illustrates various chimeras and their assigned names. By the exchange of homologous domains, a series of 2^4 = 16 constructs can be generated, consisting of 14 chimeras in addition to wild-type human and mouse P-gp (Figure 4.2).

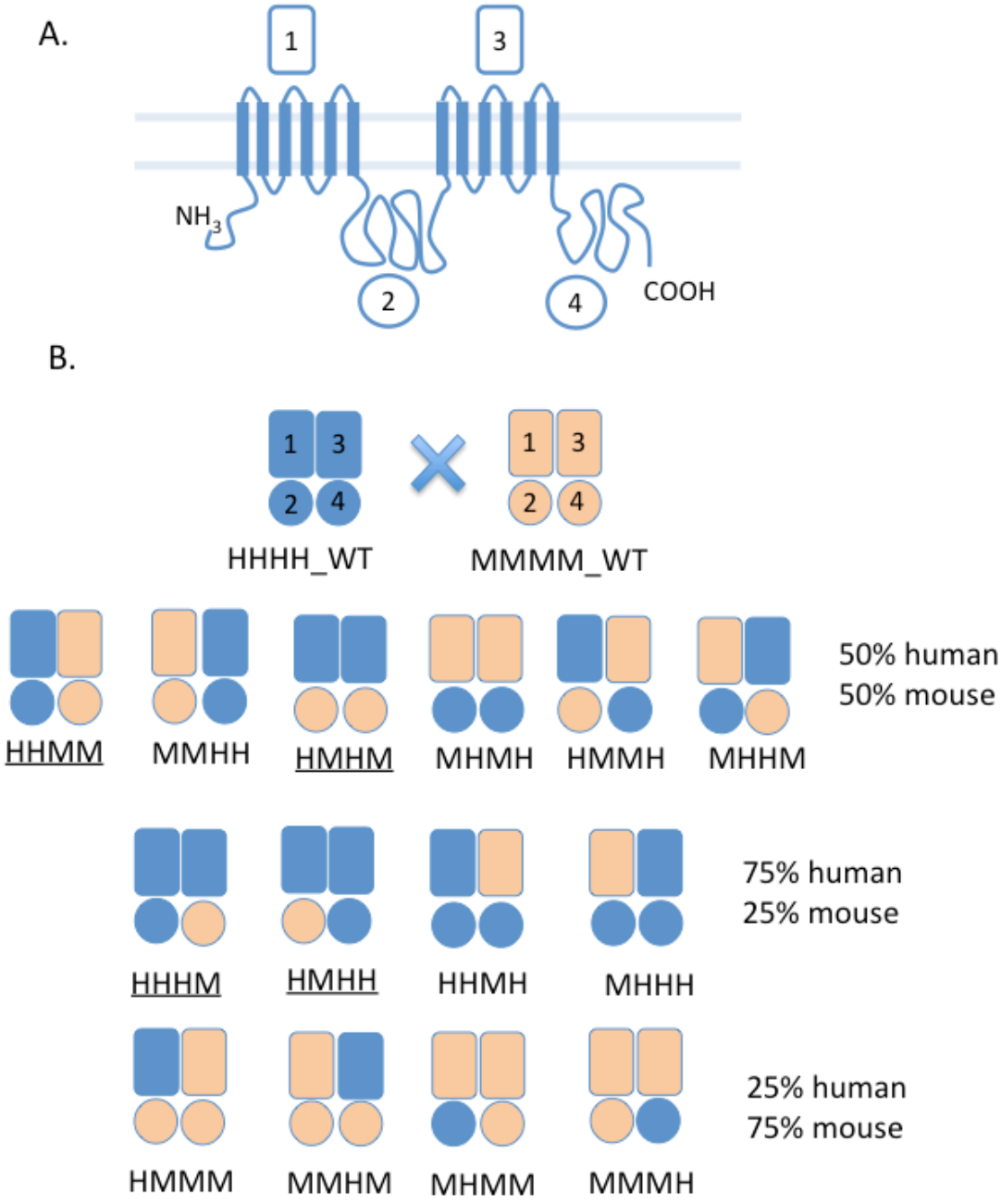


Figure 4.2 Nomenclature of P-gp chimeras. **A)** Human and mouse P-gp consist of four domains: two TMDs and two NBDs. Both human and mouse P-gp are formed by the folding of a single polypeptide chain that is transcribed and translated in the order: TMD1(1)-NBD1(2)-TMD2(3)-NBD2(4). **B)** A total of 14 chimeras can be made by domain swapping between mouse (orange) and human (blue) P-gp. Chimeras are divided into three groups based on the percentage of mouse and human domains that comprise each chimera. The underlined chimeras were generated in this study.

The work in this chapter describes the molecular biology and cloning strategies used to generate P-gp chimeras and the results are divided into three sections. Generation of HHMM and HHHH_WT P-gp (Section 4.2.1) and generation of HHHM, HMHH, and HMHM P-gp chimeras (Section 4.2.2) are discussed separately to account for the different molecular biology and cloning strategies employed (Figure 4.3). Additionally, this chapter includes an analysis of the crystal contacts present in the mouse P-gp crystal structure and the retention of those contacts in human P-gp and selected chimeras generated in this chapter (Section 4.2.3).

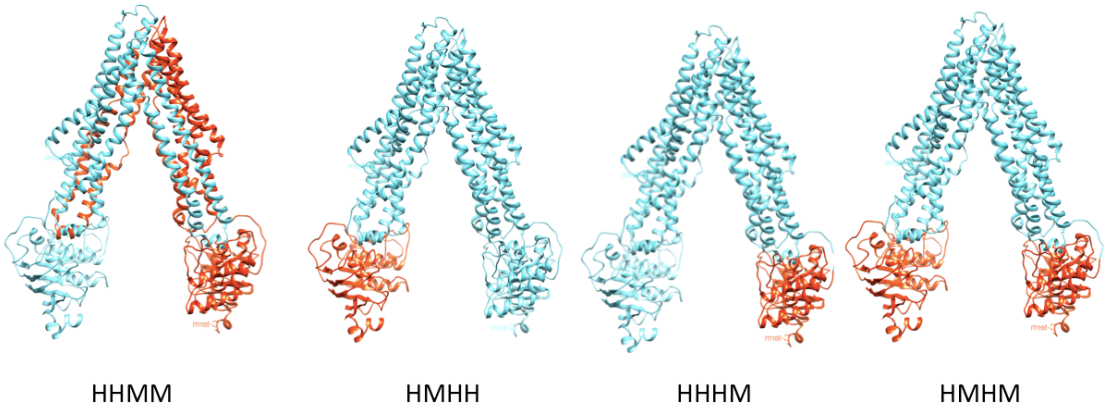


Figure 4.3 P-gp chimeras generated in this study. Four P-gp chimeras consisting of human (cyan) and mouse (orange) domains and their associated names. According to the nomenclature described in Figure 4.2, the HHMM (consisting of human TMD1/NBD1 and mouse TMD2/NBD2), HMHH (consisting of human TMD1/TMD2/NBD2 and mouse NBD1), HHHM (consisting of human TMD1/NBD1/TMD2 and mouse NBD2) and HMHM (consisting of human TMD1/TMD2 and mouse NBD1/NBD2) are depicted.

4.2 Results

4.2.1 Generation of HHHH_WT and HHMM P-gp

At the onset of this thesis project, it was unknown if chimeras comprised of mouse and human P-gp would be functional. Only chimeras that retained transport properties similar to wild-type human P-gp would be considered as candidates for crystallographic studies. Therefore, initial efforts aimed to generate two chimeras as a proof-of-principle for functional analysis before the generation of additional chimeras. The chimeras initially designed for the pilot study consisted of the N-terminal half of human P-gp fused with the C-terminal half of mouse P-gp (HHMM), and the reverse chimera consisting of the N-terminal half of mouse P-gp fused with the C-terminal half of human P-gp (MMHH). Additionally, wild-type human (HHHH_WT) and mouse (MMMM_WT) constructs would be re-generated as experimental controls to validate the PCR, cloning, and ligation strategies employed to create the chimeras. As *mdr1a* cDNA could not be cloned due to the presence of a cryptic bacterial promoter (see results presented in Chapter 3), MMHH and MMMM_WT P-gp could not be generated. Thus, Section 4.2.1 describes the generation of HHHH_WT and HHMM P-gp.

4.2.1.1 Definition of domain boundaries in *mdr1a* and *MDR1* P-gp cDNA

It was necessary to define the domain boundaries where restriction enzyme sites were to be inserted that would allow for subsequent chimera generation. As P-gp is formed by the folding of a single polypeptide, domain boundaries are not well defined and no assigned formal definition of domain boundaries exists. Domain boundaries were therefore determined by visual inspection of the crystal structure of mouse P-gp from PDB deposit 4M1M (Table 4.1) (57). Domain boundaries between TMD1 and NBD1 and TMD2 and NBD2 consist of the loops between TM6 and NBD1 and TM12 and NBD2, respectively (Figure 4.4). The NBD1 to TMD2 boundary, also termed the linker region, was defined as the flexible

string of amino acids that connect NBD1 to TMD2. Domain boundaries in human P-gp were determined by an alignment of mouse and human P-gp to identify the analogous boundaries. Boundaries defined here agree with the 2-D sequence display reported in PDB deposit 4M1M (57, 103).

Table 4.1 Domain boundaries in mouse and human P-gp. Domain boundaries were determined in this thesis project by the visual inspection of the mouse P-gp crystal structure and subsequent determination of the analogous residues in human P-gp. Domain boundaries between TMD1 and NBD1, NBD1 and TMD2, and TMD2 and NBD2 for mouse and human P-gp are shown.

Species	TMD1-NBD1	NBD1-TMD2	TMD2-NBD2
Human	374-392	644-674	1017-1035
Mouse	370-388	640-670	1013-1031

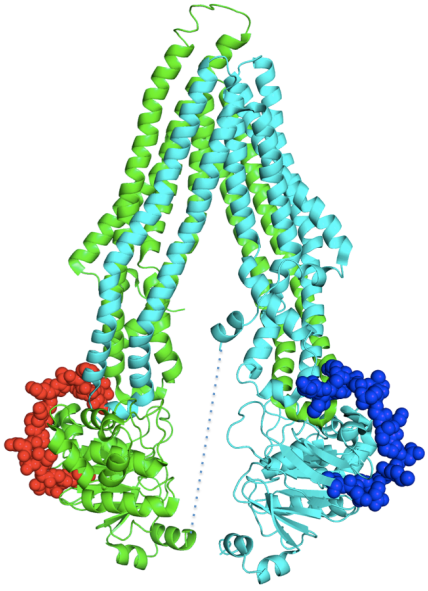


Figure 4.4 Location of domain boundaries of mouse P-gp used in the generation of chimeras. Structure of mouse P-gp with N-terminus (green) and C-terminus (cyan) shown. Residues of the TMD1-NBD1 and TMD2-NBD2 domain boundaries are shown with space-filled amino acid residues in red and blue, respectively. The NBD1-TMD2 domain boundary is composed of residues in the flexible linker region of P-gp that are disordered in the mouse P-gp crystal structure. The dotted line represents the NBD1-TMD2 domain boundary.

4.2.1.2 Strategy to introduce a silent restriction enzyme site into the NBD1-TMD2 domain boundary in *mdr1a* and *MDR1* cDNA

To generate the HHMM chimera, a unique restriction enzyme site was introduced into the NBD1-TMD2 domain boundary of *MDR1* and *mdr1a* cDNA. To ensure that the restriction enzyme site to be inserted at the domain boundary would be unique, a list of restriction enzymes that do not recognize sequences in *mdr1a* and *MDR1* cDNA (also termed "non-cutter" enzymes) was generated by an analysis with the nucleic acid sequence analysis software program MacVector. A feature of MacVector is the ability to compare input sequences (here *MDR1* and *mdr1a* cDNA) to a library of restriction enzymes and generate a list of non-cutter enzymes. It was important that the restriction enzyme site to be introduced was unique, as it ensures that the restriction enzyme will only cut at the desired location. For example, EcoRI restriction enzyme sites naturally occur in wild-type *mdr1a* and *MDR1* cDNA. The insertion of an EcoRI site at the NBD1-TMD2 domain boundary would not be ideal, as digestion would result in cutting the cDNA into fragments, thus inhibiting chimera generation. Typically, restriction enzymes recognize a hexameric sequence of nucleic acids. For example, the enzyme EcoRI recognizes the sequence GAATTC. When a hexameric restriction enzyme site is located in-frame with an ORF, the site will be translated into two amino acids, as a hexameric sequence contains two codons. Thus, the GAA and TTC codons that comprise the EcoRI restriction enzyme site would be translated into a glutamic acid and phenylalanine residues. A list of *MDR1* and *mdr1a* non-cutter enzymes and their corresponding restriction enzyme sites and amino acid residues is shown in Table 4.2.

Table 4.2 *MDR1* and *mdr1a* non-cutter enzymes and their corresponding restriction enzyme sites and amino acid residues. A list of restriction enzymes that do not recognize sequences in both *mdr1a* and *MDR1* cDNA and their corresponding hexameric restriction enzyme sites is shown. The amino acids that correspond to the two codons intrinsic to each hexameric sequence are indicated (amino acids are represented by their one letter codes).

Restriction Enzyme	Restriction Site	Translated amino acid
AclI	AACGTT	NV
AgeI	ACCGGT	TG
AvrII	CCTAGG	PR
AflII	CTTAAG	LK
BsiWI	CGTACG	RT
BspEI	TCCGGA	SG
BsrGI	TGTACA	CT
BssHII	GCGCGC	AR
ClaI	ATCGAT	ID
EagI	CGGCCG	RP
Eco53kI	GAGCTC	EL
FspI	TGCGCA	CA
HpaI	GTTAAC	VN
KasI	GGCGCC	GA
MluI	ACGCGT	TR
NaeI	GCCGGC	AG
NarI	GGCGCC	GA
NgoMIV	GCCGGC	AG
NruI	TCGCGA	SR
SacI	GAGCTC	EL
SacII	CCHCGG	PR
SalI	GTCGAC	VD
SnaBI	TACGTA	YV
SpeI	ACTAGT	TS
SphI	GCATGC	AC
SmaI	CCCGGG	PG
TspMI	CCCGGG	PG
XmaI	CCCGGG	PG

NBD1-TMD2 domain boundaries in mouse and human P-gp were screened for a pair of conserved residues that were also present in Table 4.2. Conserved isoleucine (I) and aspartic acid (D), residues that comprise a ClaI restriction enzyme site, were located at positions 644-645 and 648-649 of mouse and human P-gp, respectively (Figure 4.5). A ClaI restriction enzyme site is composed of the hexameric sequence ATCGAT, while the wild-type *MDR1* and *mdr1a* sequence coding for the isoleucine and aspartic acids residues is ATTGAT. As the codons ATT and ATC both encode an isoleucine residue, a silent T to C mutation was selected for the introduction of a ClaI restriction site. The strategy employed here (Section

4.2.1.2) allows for the introduction of restriction enzyme sites into *mdr1a* and *MDR1* cDNA that are silent with respect to the final amino acid sequence.

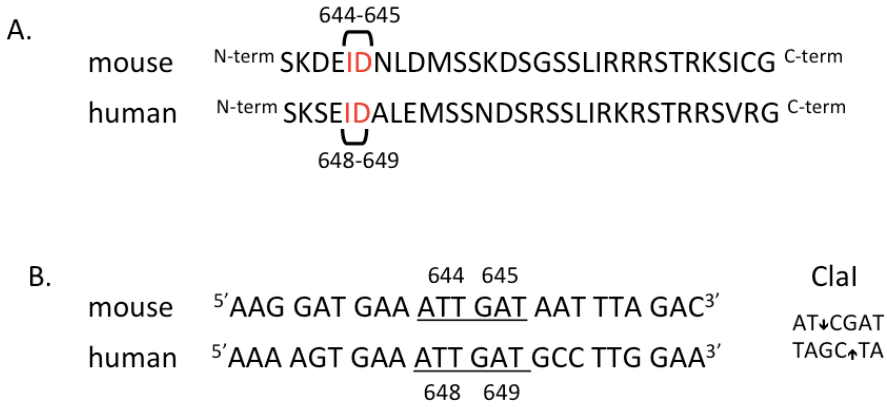


Figure 4.5 Amino acid and nucleic acid sequence alignment of the *MDR1* and *mdr1a* cDNA NBD1-TMD2 domain boundary. **A)** Amino acid residues 640 to 670 and 644 to 674 in mouse and human P-gp, respectively, are shown. Conserved isoleucine and aspartic acid residues are highlighted in red and their position in mouse and human P-gp is indicated above and below the alignment, respectively. **B)** Nucleic acid sequence alignment of codons in the NBD1-TMD2 domain boundary with wild-type isoleucine and aspartic acid codons underlined. A ClaI recognition site can be introduced by a silent T to C mutation in the isoleucine codon.

4.2.1.3 Design of PCR primers to introduce a ClaI restriction enzyme site at the NBD1-TMD2 domain boundary of *MDR1* and *mdr1a* cDNA

PCR amplification was used to introduce a silent T to C mutation at position 1944 bp (*MDR*1) and 1932 bp (*mdr1a*) at the isoleucine residue in the NBD1-TMD2 domain boundary to create a ClaI restriction enzyme site. Additionally, the PCR primers were designed to introduce various elements/restriction enzyme sites to enable downstream sub-cloning and expression in multiple heterologous expression systems (Figure 4.6). Firstly, as chimeras generated would be expressed and functionally characterized in mammalian cell-based assays, primers were designed to introduce a Kozak sequence (243) directly upstream of the *MDR1* and *mdr1a* cDNA to aid translational initiation during mammalian expression. Secondly, the chimera generation strategy to be used required additional restriction enzyme

sites to facilitate the ligation of chimeras (to be described in Section 4.2.1.5); thus, NheI and AvrII restriction enzyme sites (NheI and AvrII are isoschizomers) were introduced. Thirdly, chimeras found to be functional would be expressed and purified for crystallographic structural studies. Purification of chimeras from membranes would be completed by IMAC, which requires a poly-histidine (His_6) tag for binding to nickel resin; thus the primer designs introduced a His_6 tag at the C-terminus of both the *MDR1* and *mdr1a* cDNAs prior to the stop codon.

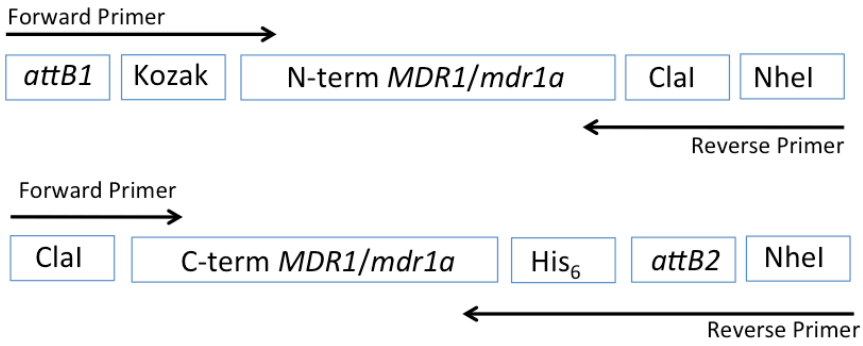


Figure 4.6 Scheme of PCR primers used to generate *mdr1a* and *MDR1* cDNA amenable for HHMM chimera generation. Primers were designed (indicated by arrows) to introduce a ClaI restriction enzyme site at the NBD1-TMD2 domain boundary. Additionally, a Kozak sequence and a His_6 tag were introduced at the N-terminus and C-terminus, respectively. A NheI restriction enzyme site was included to allow for HHHH_WT and HHMM P-gp generation. Lastly, *attB1* and *attB2* sequences were included to allow for downstream Gateway cloning of chimeras.

Lastly, PCR primers were designed to include *attB1* and *attB2* Gateway cloning sites to allow for the downstream sub-cloning of chimeras upon their generation into various expression vectors for expression in mammalian, yeast, and insect cells for both characterization of function and heterologous protein expression for crystallographic studies (Section 2.2.12). Gateway cloning is based on the site-specific recombination of bacteriophage lambda into a bacterial host genome during its lysogenic life cycle (244-246). Integration is achieved when specific attachment sequences in the phage genome (termed *attP1* and *attP2* sites) recognize and bind to attachment sites in the bacterial host genome

(termed *attB1* and *attB2*) (244). Site-specific recombination occurs in the presence of various enzymes and integration factors produced by the both the phage and host bacterium (a BP reaction). During the lytic cycle, the opposite reaction occurs, and phage DNA is excised from the host genome for subsequent replication of viral DNA and production of progeny phage particles (a LR reaction). Various expression vectors are commercially available that contain *attP* sites (termed pDonr vectors; Table 2.9). Additionally, the various enzymes and components required for recombination at *attP* and *attB* sites are commercially available. Therefore, the inclusion of *attB1* and *attB2* sites flanking *MDR1* and *mdr1a* cDNA allowed for the facile introduction of the various P-gp chimeras into the mammalian, yeast, and insect cell expression vectors required for this study. PCR amplification was carried out as described in Section 2.2.1.

4.2.1.4 Generation and validation of pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm, pTOPO_mdr1a_Nterm, and pTOPO_mdr1a_Cterm plasmids

PCR fragments were ligated into pCR-Blunt Topo plasmid vectors using the Zero Blunt TOPO PCR cloning kit (Section 2.2.4). Blunt-ended PCR fragments were ligated into a linearized pCR-Blunt Topo plasmid by topoisomerase I. Four plasmids, designated pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm, pTOPO_mdr1a_Nterm, and pTOPO_mdr1a_Cterm containing N-term *MDR1*, C-term *MDR1*, N-term *mdr1a*, and C-term *mdr1a*, respectively, were generated by the ligation of PCR products into the pCR-Blunt Topo plasmid (Table 2.9; Figure 4.7). Plasmids were transformed into chemically competent One Shot Top10 *E.coli* and spread onto LB agar containing kanamycin as a selection reagent to allow for clonal growth of transformed bacteria (Section 2.2.8).

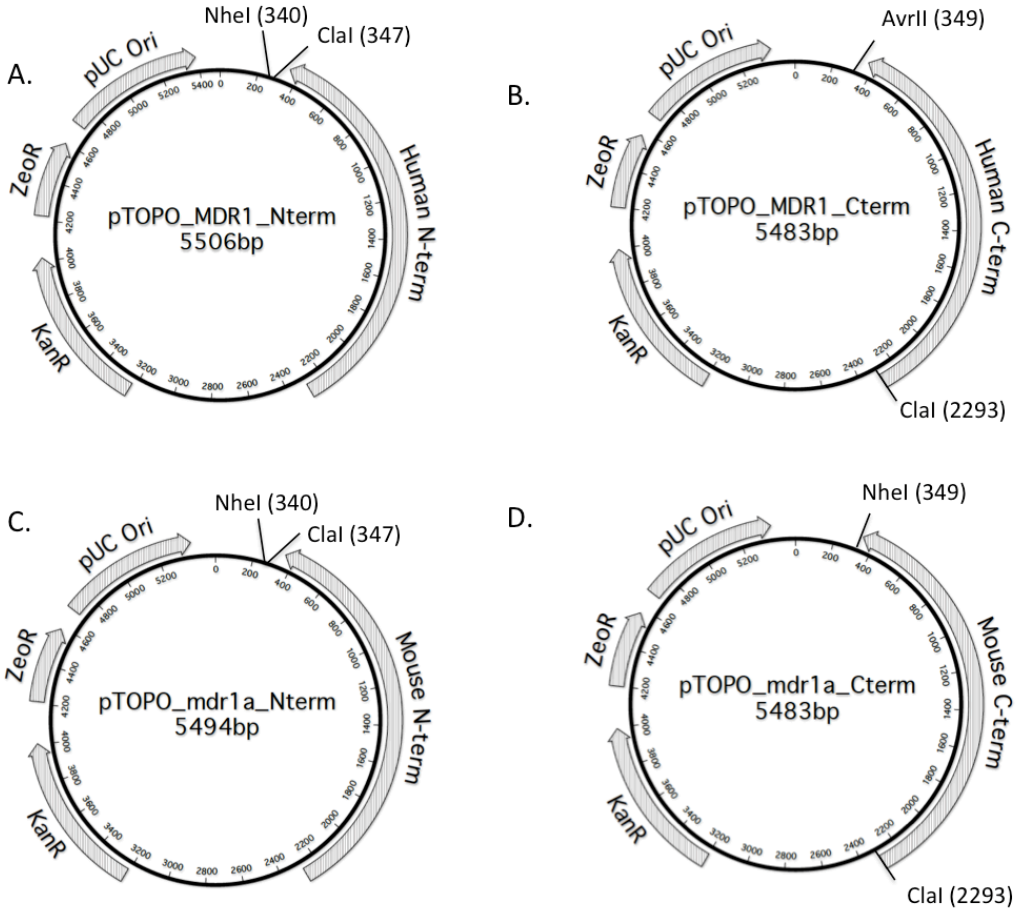
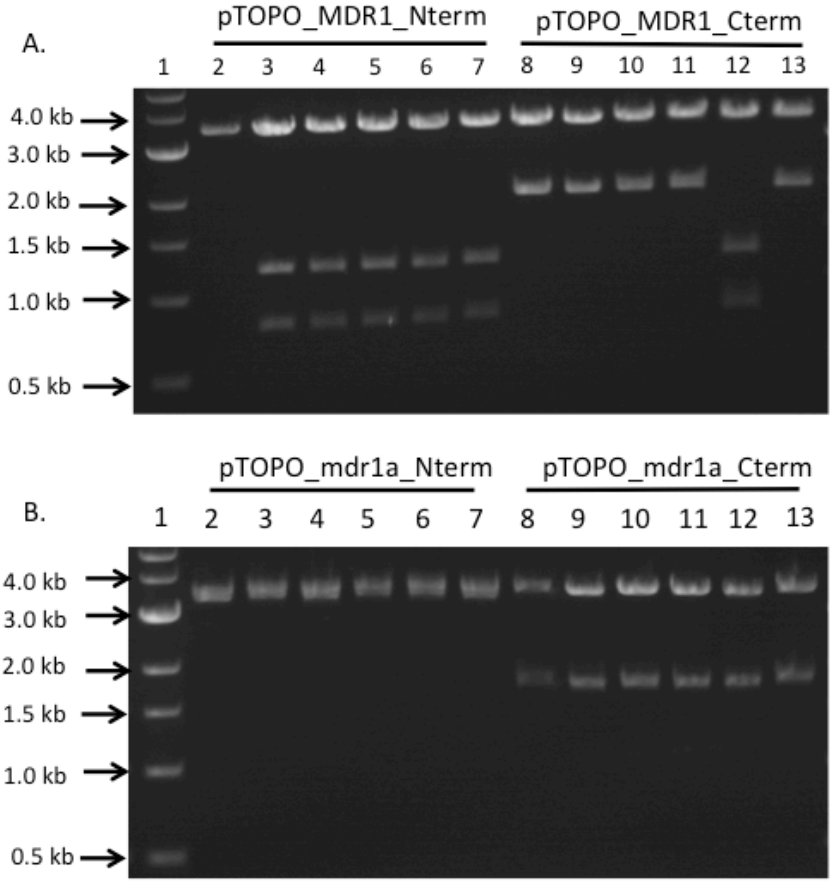


Figure 4.7 Plasmid maps of pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm, pTOPO_mdr1a_Nterm, and pTOPO_mdr1a_Cterm. N-terminal and C-terminal halves of *MDR1* and *md1ra* cDNA PCR products were ligated into pCR-Blunt Topo plasmids (Section 2.2.4). The name and location of the restriction enzyme sites introduced by PCR are shown. All plasmids are approximately 5.5 kb in size, composed of a 3.5 kb pCR-Blunt Topo plasmid backbone and a PCR product of approximately 2.0 kb. The pCR-Blunt Topo plasmid backbone confers kanamycin (KanR) and zeocin (ZeoR) resistance and contains a bacterial origin of replication (pUC Ori) (Table 2.9).

Typically, at least 5% of ligation reactions result in the self-ligation of the linearized pCR-Blunt Topo plasmid without the PCR product of interest (Section 2.2.4). To address this, the manufacturer's protocol recommends the screening of four to six transformed colonies to determine which plasmids have the PCR product successfully ligated into the pCR-Blunt Topo plasmid. Six colonies from each transformation were selected for small-scale culture and purification of plasmid DNA. To screen for plasmids that contained a correct ligation of the PCR fragment into the pCR-Blunt Topo plasmid, purified plasmid DNA was subjected to a

"diagnostic" digest with EcoRI and agarose gel electrophoresis (Figure 4.8). Successful ligations were identified when the size of the purified plasmid increased from 3.5 kb (the size of a self-ligated pCR-Blunt Topo plasmid) to 5.5 kb, indicating the insertion of the approximately 2.0 kb PCR fragment. Additionally, purified plasmid subjected to EcoRI digestion should result in a predictable DNA fragmentation pattern. For example, the pTOPO_MDR1_Nterm plasmid contains three EcoRI sites, and digested plasmid should yield three DNA fragments of 3.5, 1.2, and 0.7 kb. By comparing the DNA fragments produced from an EcoRI digest of pTOPO_MDR1_Nterm plasmid DNA to the expected DNA fragmentation pattern, successful ligations of PCR products into pCR-Blunt Topo plasmids can be confirmed. A table showing the DNA fragmentation patterns to be expected from EcoRI digest of the pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm, pTOPO_mdr1a_Nterm, and pTOPO_mdr1a_Cterm plasmids is shown in Figure 4.8.C. The comparison of digested plasmid DNA to predicted DNA fragmentation patterns is used throughout this study as a screening strategy to determine the outcome of various plasmid manipulations.

Five out of six pTOPO_MDR1_Nterm and pTOPO_MDR1_Cterm plasmids comprised successful ligations of the required PCR product into the pCR-Blunt Topo plasmid (Figure 4.8.A). Additionally, all purified pTOPO_mdr1a_Cterm plasmids screened comprised successful ligation of the C-term *mdr1a* PCR product into the pCR-Blunt Topo plasmid (Figure 4.8.B). There were no examples of successful ligation of N-terminal *mdr1a* into the pCR-Blunt Topo plasmid, as a 1.9 kb DNA band was not observed (Figure 4.8.B). The failure to obtain a pTOPO_mdr1a_Nterm plasmid after multiple attempts was attributed to the presence of a cryptic bacterial promoter in *mdr1a* cDNA inhibiting the growth of transformed *E. coli* (see experiments described in Chapter 3).



C.

Plasmid	EcoRI digest
pTOPO_MDR1_Nterm	3.5, 1.2, 0.7 kb
pTOPO_MDR1_Cterm	3.5, 1.9 kb
pTOPO_mdr1a_Nterm	3.5, 1.9 kb
pTOPO_mdr1a_Cterm	3.5, 1.7, 0.17 kb

Figure 4.8 EcoRI diagnostic digest of pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm, pTOPO_mdr1a_Nterm, and pTOPO_mdr1a_Cterm plasmids. PCR products were ligated into pCR-Blunt Topo plasmids and transformed into *E. coli*. Six colonies from each transformation were selected for small-scale growth and plasmid purification. Purified plasmids were digested with EcoRI and subjected to agarose gel electrophoresis. Gel electrophoresis of **A)** digested pTOPO_MDR1_Nterm and pTOPO_MDR1_Cterm plasmids, and **B)** pTOPO_mdr1a_Nterm and pTOPO_mdr1a_Cterm plasmids. **C)** Table showing DNA fragmentation expected from EcoRI digestion of pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm, pTOPO_mdr1a_Nterm, and pTOPO_mdr1a_Cterm plasmids.

One of each pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm and pTOPO_mdr1a_Cterm plasmid was selected for large-scale culturing and plasmid purification. While EcoRI diagnostic digests confirmed successful ligation of the PCR product into the pCR-Blunt Topo plasmid, small mutations may have occurred either during PCR or cloning of plasmid DNA in bacteria. Therefore, sequencing of each plasmid containing *mdr1a* and *MDR1* cDNA was completed (Section 2.2.15). Analysis of sequence chromatograms confirmed that each plasmid contained non-mutated *mdr1a* or *MDR1* cDNA.

4.2.1.5 Generation of pTOPO_HHHH_WT and pTOPO_HHMM plasmids

After pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm and pTOPO_mdr1a_Cterm plasmids were confirmed by sequence analysis, generation of HHMM and HHHH_WT P-gp commenced. pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm and pTOPO_mdr1a_Cterm plasmids were subjected to either a ClaI/NheI or ClaI/AvrII double digestion (Figure 4.9). As pTOPO_MDR1_Nterm contains consecutive NheI and ClaI sites, restriction enzyme digestion excised a 7bp fragment of DNA, resulting in the formation of a linearized plasmid with ClaI and NheI cohesive ends (Figure 4.9). In contrast, pTOPO_MDR1_Cterm and pTOPO_mdr1a_Cterm plasmids contain restriction enzyme sites flanking C-term *MDR1* and *mdr1a* cDNA. Thus, ClaI/AvrII and ClaI/NheI digestion of pTOPO_MDR1_Cterm and pTOPO_mdr1_Cterm, respectively, resulted in the production of two DNA fragments: a 1.9 kb fragment containing *MDR1* and *mdr1a* C-term and the 3.5 kb plasmid backbone (Figure 4.9). Restriction enzyme digestion reactions were subjected to agarose gel electrophoresis for the separation of DNA fragments and subsequent gel purification of the desired DNA fragment containing C-term *mdr1a* and *MDR1* cDNA.

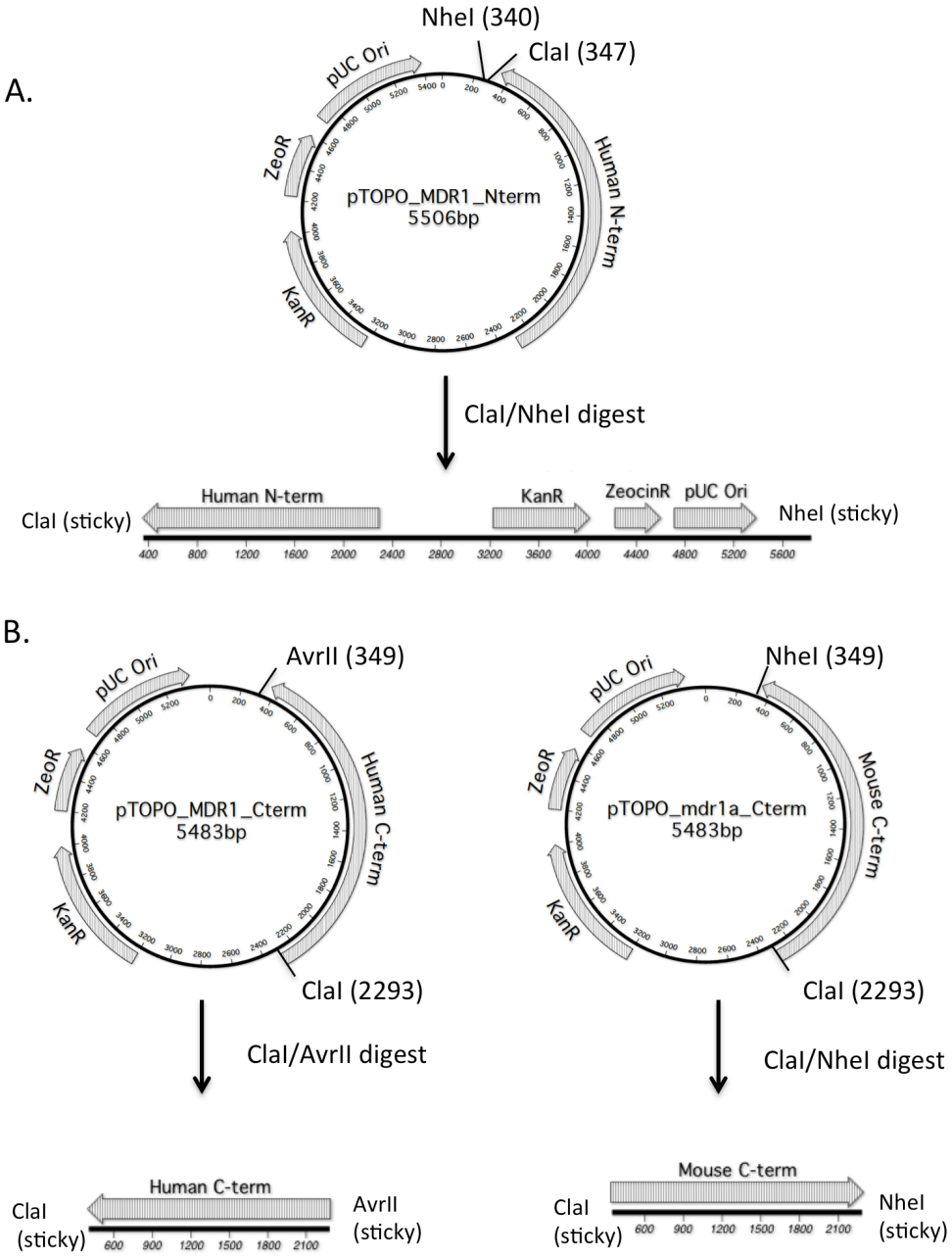
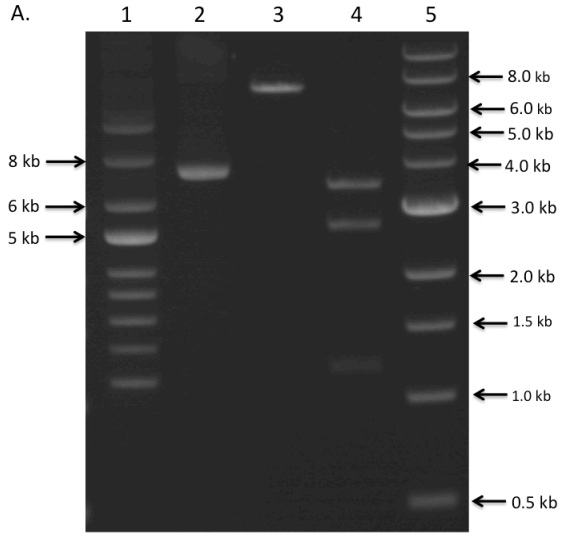


Figure 4.9 Scheme displaying restriction enzyme digestion of pTOPO_MDR1_Nterm, pTOPO_MDR1_Cterm and pTOPO_mdr1a_Cterm plasmids. A) A ClaI/NheI double digestion of the pTOPO_MDR1_Nterm plasmid resulted in a linearized plasmid containing N-term *MDR1* cDNA. **B)** A ClaI/AvrII and ClaI/NheI double digestion of pTOPO_MDR1_Cterm and pTOPO_mdr1a_Cterm plasmids, respectively, resulted in a 1.9 kb fragment containing *MDR1* and *mdr1a* C-term and the 3.5 kb plasmid backbone (not shown). All restriction enzymes used result in complimentary-ended DNA.

The ligation of linearized pTOPO_MDR1_Nterm to linearized *MDR1* C-term and *mdr1a* C-term resulted in the generation of pTOPO_HHHH_WT and pTOPO_HHMM plasmids containing human wild-type and HHMM P-gp, respectively. Plasmids were transformed into *E. coli* and spread onto LB agar containing kanamycin as a selection agent to allow for clonal growth of transformed bacteria. Colonies of each transformation were selected for small-scale growth and purification of plasmid DNA. Purified plasmids were subjected to agarose gel electrophoresis and various diagnostic restriction enzyme digests (Figure 4.10 and Figure 4.11). Both supercoiled (uncut) plasmids and linearized plasmids showed an increase in size from 5.5 kb to 7.5 kb, indicating the ligation of C-term *MDR1* and *mdr1a* cDNA into the pTOPO_MDR1_Nterm plasmid. Purified pTOPO_HHHH_WT and pTOPO_HHMM plasmids were confirmed by subsequent sequence analysis. Plasmid maps for pTOPO_HHHH_WT and pTOPO_HHMM are shown in Figure 4.12.

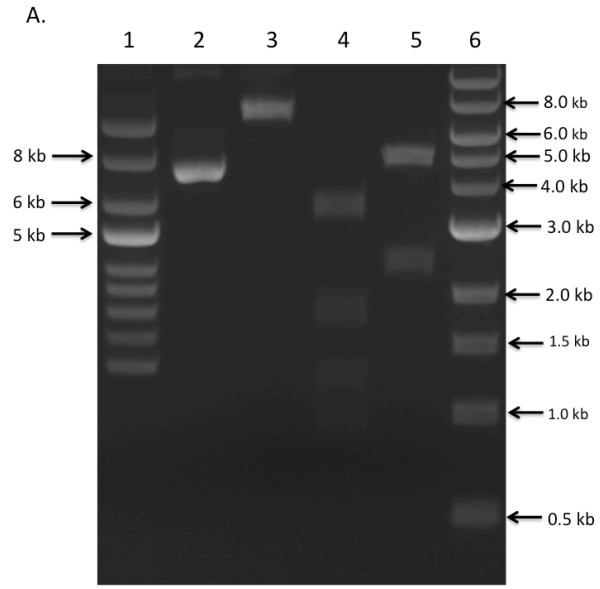
A.



B.

Lane	
1	Supercoiled ladder
2	Uncut HHHH_WT plasmid
3	ClaI digest (7.5 kb)
4	EcoRI digest (3.5, 2.8, 1.2 kb)
5	Linear ladder

Figure 4.10 Analysis of pTOPO_HHHH_WT plasmid by agarose gel electrophoresis. A) A potential ligation to yield a pTOPO_HHHH_WT plasmid was analyzed by agarose gel electrophoresis of supercoiled plasmid DNA (lane 2) and by ClaI (lane 3) and EcoRI (lane 4) diagnostic restriction enzyme digestion. A successful ligation event was indicated by the increase in size from 5.5 to 7.5 kb. DNA ladders for both supercoiled (lane 1) and linear (lane 5) DNA are included. **B)** Table showing the expected DNA fragmentation pattern expected from either supercoiled or digested pTOPO_HHHH_WT plasmid DNA.



B.

Lane	
1	Supercoiled ladder
2	Uncut HHMM plasmid
3	FseI digest (7.5 kb)
4	EcoRI digest (3.5, 1.8, 1.2, 0.9 kb)
5	PciI digest (5.0, 2.4 kb)
6	Linear ladder

Figure 4.11 Analysis of potential pTOPO_HHMM plasmid by agarose gel electrophoresis. A) A potential ligation to yield the pTOPO_HHMM was analyzed by agarose gel electrophoresis of supercoiled plasmid DNA (lane 2) and by FseI (lane 3), EcoRI (lane 4), and PciI (lane 5) diagnostic restriction enzyme digestion. A successful ligation event was indicated by the increase in size from 5.5 to 7.5 kb. DNA ladders for both supercoiled (lane 1) and linear (lane 6) DNA are included. **B)** Table showing the expected DNA fragmentation pattern expected from either supercoiled or digested pTOPO_HHMM plasmid DNA.

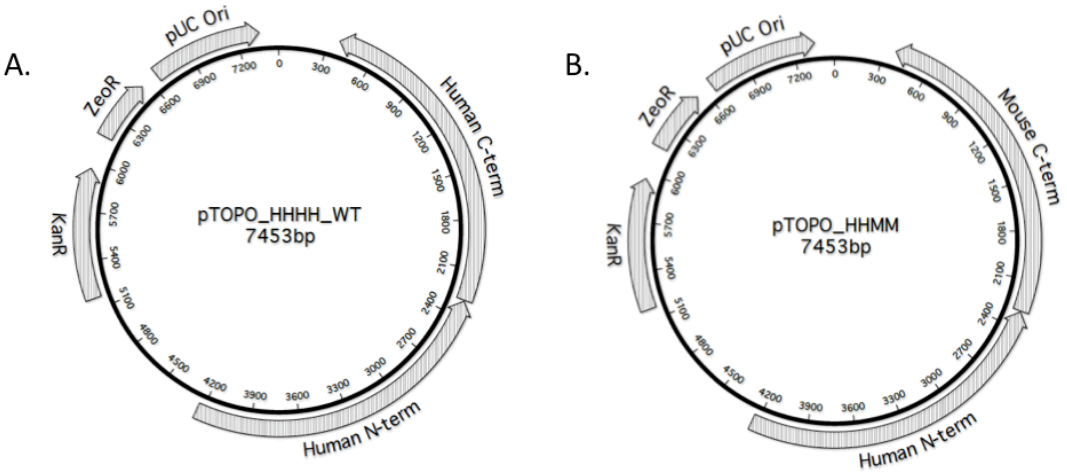


Figure 4.12 Plasmid maps of pTOPO_HHHH_WT and pTOPO_HHMM. Plasmids containing HHHH_WT and HHMM chimeric cDNA were termed **A)** pTOPO_HHHH_WT and **B)** pTOPO_HHMM, respectively (Table 2.9). Plasmids generated were approximately 7.5 kb in size, consisting of a 3.5 kb TOPO backbone, and either HHHH_WT or HHMM cDNA of approximately 4 kb.

4.2.2 Generation of HHHM, HMHH, and HMHM P-gp Chimeras

Both the proof-of-principle HHMM chimera and the HHHH_WT control construct were shown to be functional, having transport properties similar to wild-type human P-gp (Chapter 5). Therefore, this study continued with the generation of additional chimeras. The generation of the HHHM, HMHH, and HMHM chimeras is discussed in the following sections. Similar to the method used to produce the HHMM chimera, the HHHM, HMHH, and HMHM chimeras were generated by the addition of unique restriction enzyme sites into the domain boundaries of *MDR1* and *mdr1a* cDNA. However, due to difficulties cloning *mdr1a* cDNA (Chapter 3), these chimeras were created using a modified cloning strategy. To avoid cloning of full-length *mdr1a* cDNA, PCR was used to isolate individual NBD and TMD domains of *mdr1a*. The isolated domains of *mdr1a* were inserted into individual plasmids, and domain swapping between a plasmid containing full-length *MDR1* and the desired *mdr1a* domain was carried out.

4.2.2.1 *In silico* design of codon-optimized *MDR1* and *mdr1a* cDNA

Codon optimization is a commonly used strategy to increase heterologous protein expression through the substitution of rare codons with common codons preferred by the host organism (Section 1.4.3). At the onset of this project, heterologous protein expression in *P. pastoris* was the anticipated method for high-level expression of chimeras for structural studies. Given this, *mdr1a* and *MDR1* cDNA were codon optimized for expression in *P. pastoris* for two main reasons. Firstly, codon optimization strategies have been previously shown to increase mouse P-gp expression in *P. pastoris* greater than three-fold (166). Secondly, due to the difficulties encountered in cloning *mdr1a* cDNA, codon optimization was used as a strategy to possibly reduce *mdr1a* toxicity by changing sequences that might contain a cryptic bacterial promoter. (At the time of chimera generation, the cryptic promoter had not yet been identified and therefore a targeted mutational approach to

abrogate toxicity could not be carried out). Codon-optimized *MDR1*, termed HHHH_Opti, and *mdr1a*, termed MMMM_Opti, cDNA was designed *in silico* and assembled by GeneArt Gene Synthesis services offered through Life Technologies.

Synthesized HHHH_Opti and MMMM_Opti cDNA were designed with unique restriction enzyme sites at each domain boundary (Figure 4.13). The three restriction enzyme sites chosen were based on previous analysis completed and described in Section 4.2.1.2 and shown in Table 4.2. A BspEI site was introduced at conserved serine (S) and glycine (G) residues in the TMD1-NBD1 domain boundary at positions 377 and 378 in mouse P-gp and at positions 381 and 382 in human P-gp (Figure 4.13.A). Additionally, an AclI site was introduced at conserved asparagine (N) and valine (V) residues in the TMD2-NBD2 domain boundary at positions 1030 and 1031 in mouse P-gp and 1034 and 1035 in human P-gp (Figure 4.13.C). Lastly, the ClaI site introduced for the generation of the HHMM chimera was also included in codon-optimized cDNA (Figure 4.13.B). Synthesized cDNA also contained various elements that were previously introduced by PCR, such as a Kozak sequence, *attB1* and *attB2* sites, and a His_6 tag (Section 4.2.1.3). Two additional unique restriction enzyme sites, SalI and AvrII, were also included at the 5' and 3' termini to facilitate chimera construction and subsequent use. Thus, both synthesized mouse and human cDNA consisted of the following general sequence:

(N-term) SalI-*attB1*-Kozak-TMD1-BspEI-NBD1-ClaI-TMD2-AclI-NBD2-His_6-*attB2*-AvrII (C-term)

Overall, by the design of mouse and human cDNA that included the elements described, all possible configurations of domain-swapped P-gp could, theoretically, be achieved (Figure 4.2). Additionally, inclusion of the *attB1*/*attB2*, Kozak, and His_6 elements allows for downstream expression of chimeras in various hosts and subsequent purification of expressed protein.

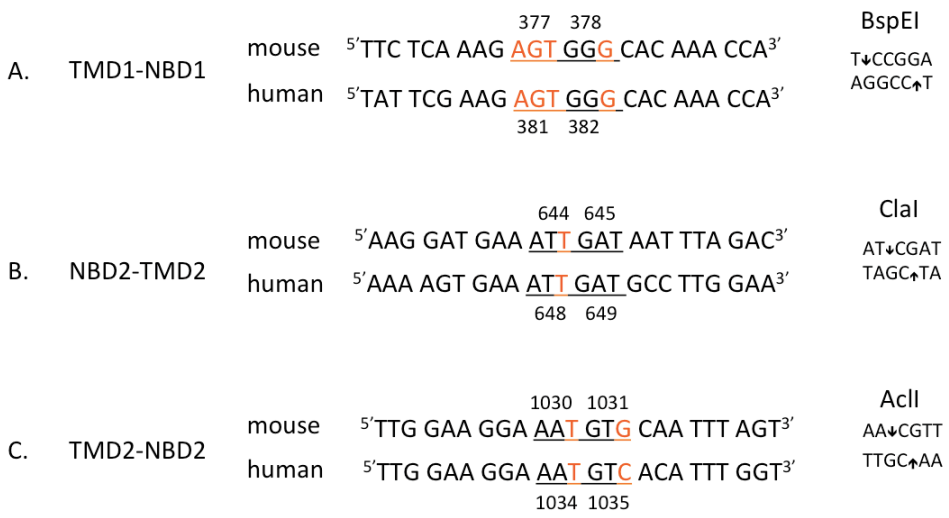


Figure 4.13 Silent mutations to introduce unique restriction enzyme sites at domain boundaries in *mdr1a* and *MDR1* cDNA. Sequence alignment of wild type *mdr1a* and *MDR1* at the **A)** TMD1-NBD1, **B)** NBD1-TMD2, and **C)** TMD2-NBD2 domain boundaries. The conserved wild-type residues that were mutated to restriction enzyme sites during gene synthesis are underlined and the amino acid residue positions are indicated above (mouse) and below (human) the sequences. The specific nucleic acids that were mutated to introduce the restriction enzyme site are noted in red. The restriction enzyme site introduced into each boundary appears to the right of the sequence alignments.

4.2.2.2 Generation of pTOPO_HHHH_Opti and pTOPO_MMMM_Opti plasmids

Upon receipt of plasmids containing codon-optimized HHHH_Opti and MMMM_Opti cDNA, cDNA was sequenced and confirmed to be correct. HHHH_Opti and MMMM_Opti cDNA was received in a commercial vector that could not be used directly for chimera generation due to the presence of additional AclI and AvrII restriction sites in the plasmid backbone (thus making those sites no longer unique in the *mdr1a* and *MDR1* cDNA). Therefore, PCR of HHHH_Opti and MMMM_Opti and ligation of PCR products into the pCR-Blunt Topo plasmid were carried out to generate pTOPO_HHHH_Opti and pTOPO_MMMM_Opti plasmids. Plasmids were transformed into *E. coli* and streaked on kanamycin selective media for clonal growth.

Bacteria transformed with pTOPO_MMMM_Opti plasmids showed a mutation rate similar to that observed with plasmids containing wild-type *mdr1a* cDNA (data not shown). This was also confirmed by personal communications with GeneArt representatives involved in the gene synthesis process. For example, while gene synthesis is typically completed in 2-4 weeks (depending on the size of the DNA sequence), MMMM_Opti took over three months to synthesize; a process that required the ligation and screening of MMMM_Opti into various plasmids to select one that would support cloning. These findings supported previous evidence that standard plasmids and growth conditions are not suitable for the replication of plasmids containing *mdr1a* cDNA in *E. coli* (Chapter 3).

4.3.2.3 PCR, ligation, and cloning of pTOPO_M1_Opti, pTOPO_M2_Opti, pTOPO_M3_Opti and pTOPO_M4_Opti

Given the propensity for both wild-type and codon-optimized *mdr1a* cDNA to acquire mutations upon transformation into bacteria, a cloning strategy to generate chimeras was devised that avoided cloning of full-length mouse cDNA. Four PCR products containing each domain of the mouse P-gp domains were prepared and ligated into pCR-Blunt Topo plasmids to generate four plasmids: pTOPO_M1_Opti containing TMD1, pTOPO_M2_Opti containing NBD1, pTOPO_M3_Opti containing TMD2, and pTOPO_M4_Opti containing NBD2 (Figure 4.14, Table 2.9). Plasmids were transformed into *E. coli* and streaked on LB agar plates containing kanamycin for selection of clones. Four to six clones were selected for small-scale culture and plasmid purification. Purified plasmids were subjected to gel electrophoresis to determine colonies that contained the ligated PCR products. The high mutation rate of full-length *mdr1a* cDNA was not observed with the pTOPO_M1_Opti, pTOPO_M2_Opti, pTOPO_M3_Opti and pTOPO_M4_Opti plasmids. Large-scale culture of each plasmid was performed, plasmid DNA was purified, and the *mdr1a* cDNA sequence in each plasmid was confirmed by sequence analysis.

4.2.2.4 Generation of pTOPO_HHHM, pTOPO_HMHH, and pTOPO_HMHM plasmids

HHHM, HMHH, and HMHM chimeras were generated by the restriction enzyme digestion of pTOPO_HHHH_Opti and the analogous restriction enzyme digestion of either pTOPO_M2_Opti or pTOPO_M4_Opti (Figures 4.14 and 4.15). A BspEI/ClaI double digestion of pTOPO_HHHH_Opti and pTOPO_M2_Opti was carried out to generate DNA fragments with complementary cohesive ends. Ligation of the linearized pTOPO_HHHH_Opti plasmid with mouse NBD1 resulted in the generation of the pTOPO_HMHH plasmid. Similarly, an AclI/AvrII double digestion of pTOPO_HHHH_Opti and pTOPO_M4_Opti was carried out and subsequent ligation of products generated pTOPO_HHHM. To construct the HMHM chimera, pTOPO_HMHH and pTOPO_M4_Opti were subjected to AclI/AvrII double digestion and ligation to produce the pTOPO_HMHM plasmid (Figure 4.15).

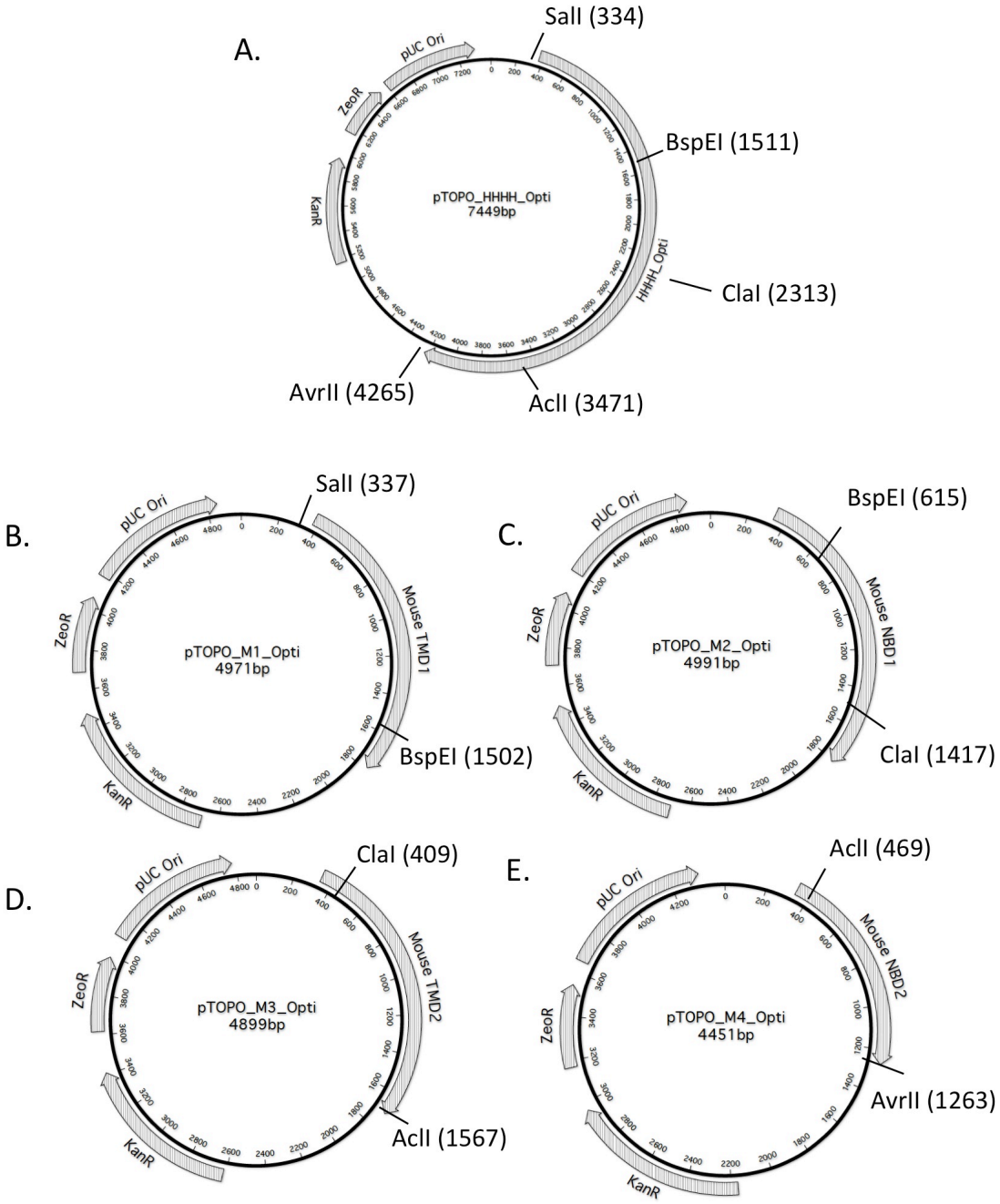


Figure 4.14 Plasmid maps of pTOPO_HHHH_Opti, pTOPO_M1_Opti, pTOPO_M2_Opti, pTOPO_M3_Opti and pTOPO_M4_Opti. A) PCR to sub-clone HHHH_Opti from a commercial vector into the pCR-Blunt Topo plasmid was performed, generating the pTOPO_HHHH_Opti plasmid. **B-E)** PCR was performed to isolate each domain of mouse P-gp and PCR products were ligated into pCR-Blunt Topo plasmids to generate **B)** pTOPO_M1_Opti, **C)** pTOPO_M2_Opti, **D)** pTOPO_M3_Opti, and **E)** pTOPO_M4_Opti (Table 2.9). Silent restriction enzyme sites introduced for chimera generation are shown.

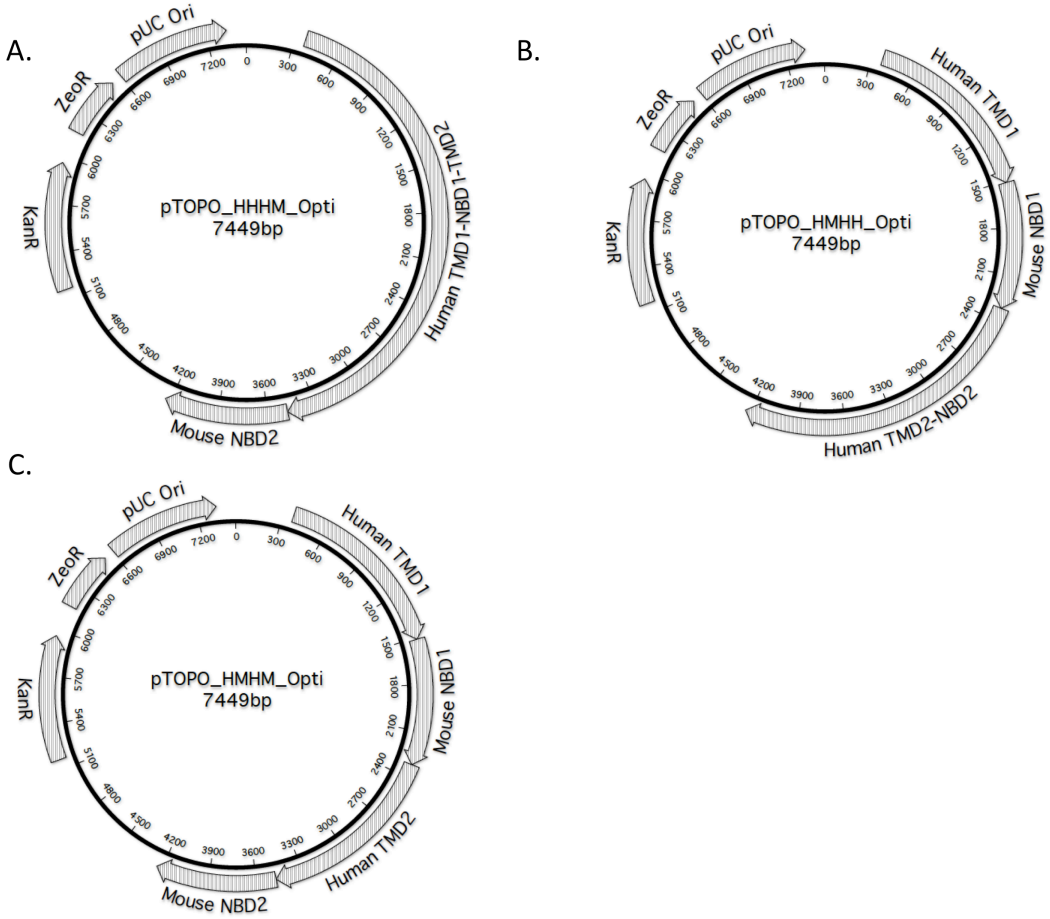


Figure 4.15 Plasmid maps of pTOPO_HHHM, pTOPO_HMHH, and pTOPO_HMHM. Plasmid maps are shown for **A)** pTOPO_HHHM, **B)** pTOPO_HMHH, and **C)** pTOPO_HMHM containing the chimeras HHHM, HMHH, and HMHM, respectively (Table 2.9).

4.2.2.5 Chimera generation concluding remarks

Overall, the generation of four chimeras is described in this chapter: HHMM, HMHM, HHHM, and HMHH (Figure 4.3). While the molecular cloning strategy used could be pursued to generate other chimeras, ultimately, each chimera needed to be individually functionally characterized, optimized for expression conditions, and screened for crystallization. Thus there were limitations to the number of chimeras that could be investigated concurrently. A pragmatic decision was made to restrict the number of chimeras to be functionally characterized, expressed, and screened for crystallization, to the four chimeras described above.

4.2.3 Analysis of crystal contacts in the mouse P-glycoprotein crystal structure

Despite the high homology between mouse and human P-gp, it is currently unknown why mouse P-gp crystallizes and human P-gp does not. One hypothesis is that the amino acid residues of mouse P-gp that partake in crystal packing forces are not conserved in human P-gp. During crystallization, protein molecules form intermolecular contacts, also termed crystal contacts, which facilitate crystal formation (247). To our knowledge, an analysis of the crystal contacts of mouse P-gp and their comparison with human P-gp has not been reported. As favorable contacts between P-gp molecules are crucial for crystal formation, the crystal contacts between mouse P-gp molecules in the reported crystal structure were analyzed (Section 4.2.3.1 and 4.2.3.2). After the residues involved in the crystal contacts of mouse P-gp were determined, these residues were compared to the equivalent residues in human P-gp to determine if human P-gp possesses similar crystal packing forces (Section 4.2.3.4). Lastly, chimeras generated here were analyzed to determine which chimeras retain the crystal contacts observed in mouse P-gp. All analyses reported herein are based on the refined and corrected X-ray crystal structure of mouse P-gp from PDB deposit 4M1M (57).

4.2.3.1 Strategy for crystal contact analysis

Mouse P-gp crystal contacts were determined in collaboration with Eduardo Chufan (Laboratory of Cell Biology, NIH). A crystal contact was defined as a pair of amino acid residues present on different P-gp molecules that interact non-covalently at a molecular level during crystal packing. Such contacts may comprise electrostatic interactions, hydrogen bonds, and van der Waals interactions. Contacts were determined by visual inspection of amino acid residue proximity to adjacent P-gp molecules with the three-dimensional structure. In general, contacts between residues separated by 3.5-4.0 Å distances are considered weak interactions (247). Because of the low resolution of the available X-ray

structure (3.8 Å), contacts up to 4.3 Å were considered. Distances between residues were assessed by measurement in the molecular visualization software PyMOL. Interactions reported are between the side chains of amino acid residues unless otherwise noted.

4.2.3.2 Analysis of crystal contacts in the asymmetric unit of mouse P-gp

The asymmetric unit of mouse P-gp contains two P-gp molecules, termed P-gp A and P-gp B (Figure 4.16). The asymmetric unit is defined as the smallest portion of a crystal structure to which operations of symmetry can be applied to generate the crystal repeating unit (248). There are two regions of contacts between P-gp A and P-gp B. Each region of contact is composed of numerous interacting amino acid residues. Crystal contacts at each region of contact between P-gp A and P-gp B are shown in Table 4.3.

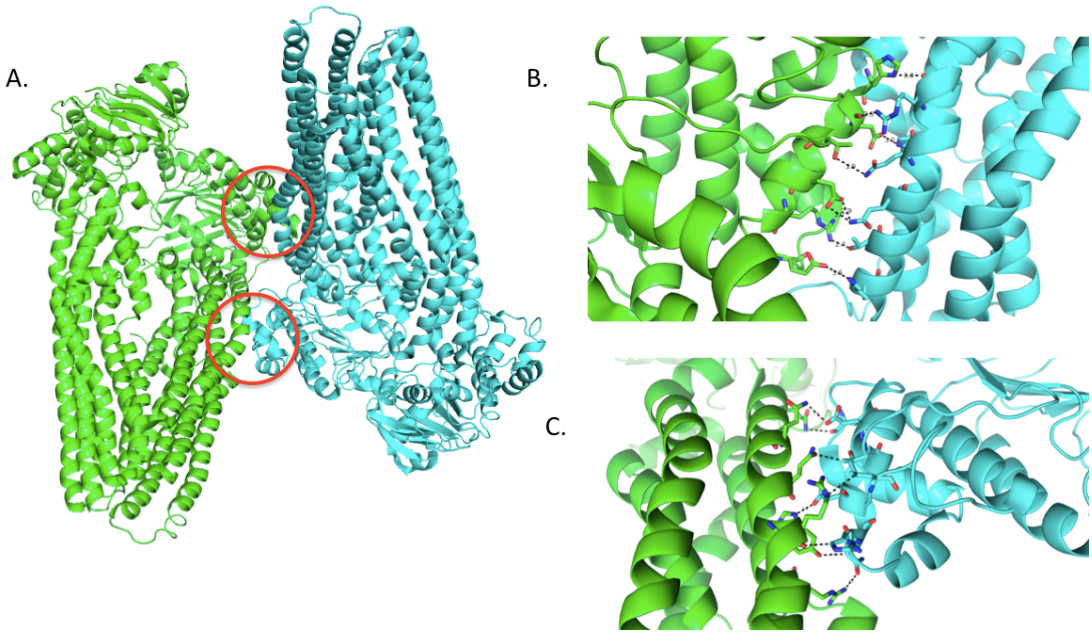


Figure 4.16 Crystal contacts in the asymmetric unit of mouse P-gp. A) The asymmetric unit of mouse P-gp consists of two P-gp molecules termed P-gp A (green) and P-gp B (cyan). Red circles indicate two regions of contact between P-gp A and P-gp B. **B-C)** Crystal contacts in the two regions of contact between P-gp A and P-gp B molecules. Contacts are indicated by dotted lines between interacting amino acid residue side-chains.

Table 4.3 Crystal contacts between P-gp A and P-gp B in the asymmetric unit of mouse P-gp.

Amino acid residue pairs that interact to form the two points of contact between P-gp A and P-gp B. Each table represents residues that comprise each point of contact, with Table A representing the top contact and Table B representing the bottom contact displayed in Figure 4.16, respectively. Residues are indicated by their one letter codes and their positions in either P-gp A or P-gp B are noted both by the exact location and also by the domain or TM helix in which the residue presides. Single asterisks are included when the interaction occurs between a side chain on one molecule of P-gp and the carbonyl group of the polypeptide backbone on the interacting molecule of P-gp.

A.

P-gp A (green)	P-gp B (cyan)	Distance
R138 (TM2)	Q515 (NBD1)	3.1 Å
H141 (TM2)	D517 (NBD1)	2.8 Å
R925 (TM11)	L519 (NBD1)	3.5 Å
Y924(TM11*)	H514 (NBD1)	4.2 Å
Y924 (TM11)	Q515 (NBD1)	3.1 Å
Q921 (TM11)	T479 (NBD1)	3.1 Å
Q918 (TM11)	E482 (NBD1)	3.3 Å
Q910 (TM11)	V491 (NBD1*)	3.2 Å

B.

P-gp A (green)	P-gp B (cyan)	Distance
R485 (NBD1)	T914 (TM11)	2.4 Å
E489 (NBD1)	K911 (TM11)	3.6 Å
E482 (NBD1)	Q918 (TM11)	2.9 Å
T479 (NBD1)	Q921 (TM11)	3.9 Å
D517 (NDB1*)	R925 (TM11)	2.5 Å
D517 (NBD1)	H141 (TM2)	3.0 Å
H514 (NBD1)	R925 (TM11*)	3.0 Å

There are two regions of contact (Figure 4.16) composed of 15 crystal contacts between P-gp A and P-gp B (Table 4.3). The intracellular end of TM11 between residues 910 and 925 and areas of NBD1 between residues 483-519 interact to form the crystal contacts in the asymmetric unit. Note that while it may appear that P-gp A is a 180-degree vertical rotation of P-gp B with symmetry existing between the two regions of contact, this is not the case; each point of contact is composed of different individual intermolecular interactions. While contacts occur between similar regions of P-gp (TM11 and NBD1), the specific contacts that comprise each region of contact are unique.

4.2.3.3 Analysis of crystal contacts between the asymmetric unit of P-gp and neighboring P-gp molecules

Each molecule of P-gp in the asymmetric unit (P-gp A and P-gp B) interacts with four neighboring P-gp molecules (Figure 4.17). Thus, each asymmetric unit of P-gp interacts with eight neighboring P-gp molecules. Each interaction between either P-gp A and P-gp B with a neighboring molecule was characterized. There were 67 crystal contacts characterized between the asymmetric unit and neighboring P-gp molecules with 33 and 34 contacts occurring between P-gp A and P-gp B with neighboring P-gp molecules, respectively. Due to symmetry properties of the crystal, there were various individual crystal contacts repeated between the asymmetric unit and neighboring P-gp molecules. Therefore, there are only 40 unique crystal contacts between the asymmetric unit and neighboring P-gp molecules.

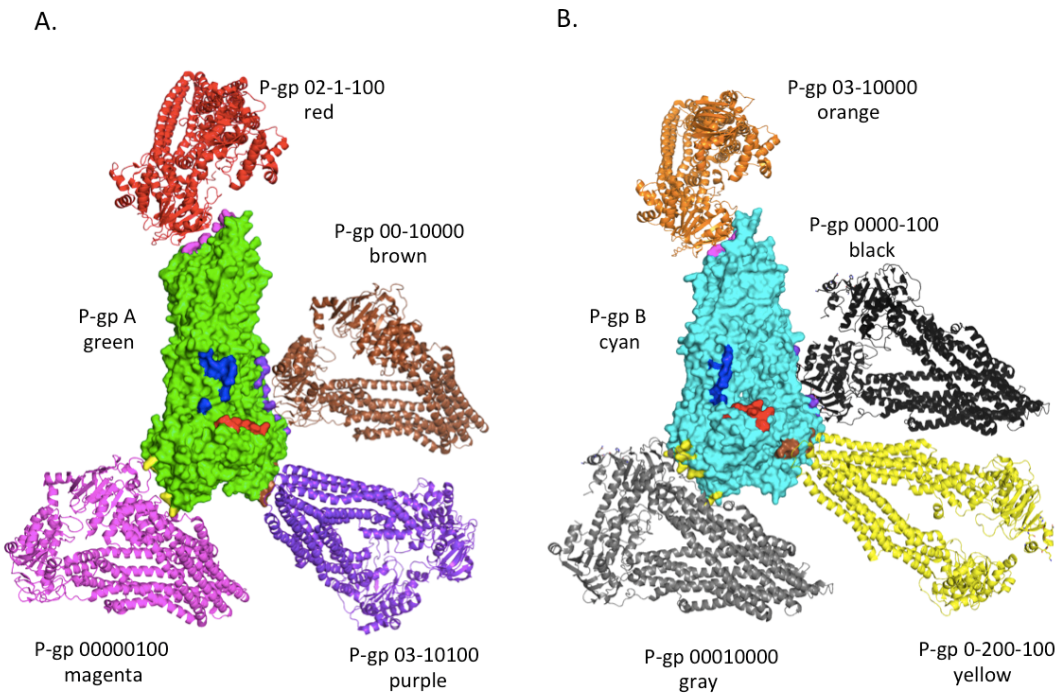


Figure 4.17 Analysis of crystal contacts between P-gp A and P-gp B with neighboring P-gp molecules. **A)** P-gp A (green) and **B)** P-gp B (cyan) both have crystal contacts with four neighboring P-gp molecules. Both P-gp A and P-gp B are shown in a space-filled model with residues involved in crystal contacts highlighted. Blue and red highlighted regions correspond to the crystal contacts between the molecules of the asymmetric unit. Yellow residues on P-gp A and P-gp B interact with P-gp 00000100 and P-gp 00010000, respectively. Brown residues on P-gp A and P-gp interact with P-gp 03-10100 and P-gp 0-200-100, respectively. Purple residues on P-gp A and P-gp B interact with P-gp 00-10000 and P-gp 0000-100, respectively. Pink residues on P-gp A and P-gp B interact with P-gp 02-2-100 and 03-10000, respectively. Names of neighboring P-gp molecules are adopted from the corresponding identification in PBD deposit 4M1M.

When both contacts within the asymmetric unit and between the asymmetric units and neighboring P-gp molecules are considered, there are a total of 82 crystal contacts of which 55 are unique (Table 4.4). An overall theme of the contacts observed is that they occur between areas of P-gp that comprise the “soluble” regions of P-gp (in other words, regions of P-gp with hydrophilic residues exposed to in the intracellular or extracellular environment, such as NBD1, NBD2, and the extracellular loops). This is likely due to the fact that transmembrane regions of P-gp are shielded by detergents used to stabilize and facilitate the retention of conformation in a membrane-free environment (Section 1.4.6).

Table 4.4 Crystal contacts between the P-gp asymmetric unit and neighboring P-gp molecules. Crystal contacts between molecules of P-gp in the asymmetric unit (P-gp A and P-gp B) with neighboring molecules of P-gp. **A-H)** P-gp A and P-gp B each interact with four other P-gp molecules, yielding eight total points of contact. The individual residues that comprise each crystal contact are shown in each table. Asterisks are included when the interaction occurs between a side chain on one molecule of P-gp and the carbonyl (*) or amino group (**) of the polypeptide backbone on the interacting molecule of P-gp. Names of neighboring P-gp molecules are adopted from the corresponding identification in PBD deposit 4M1M. Distances were determined in PyMOL.

A.

P-gp B (cyan)	P-gp 0000-100 (black)	Distance
L1022 (NBD2*)	K391 (NBD1)	3.3 Å
Q1020 (NBD2)	N410 (NBD1)	3.1 Å
T1019 (NBD2)	I602 (NBD1)	3.8 Å
K790 (TM8)	E614 (NBD1)	4.1 Å
Y786 (TM8)	R613 (NBD1)	3.3 Å
R783 (TM8)	R613 (NBD1)	4.1 Å
K698 (TM7)	K615 (NBD1)	3.0 Å

B.

P-gp A (green)	P-gp 00000100 (magenta)	Distance
K391 (NBD1)	L1022 (NBD2*)	3.3 Å
N410 (NBD1)	Q1020 (NBD2)	3.1 Å
I602 (NBD1)	T1019 (NBD2)	3.8 Å
E614 (NBD1)	K790 (TM8)	4.1 Å
R613 (NBD1)	Y786 (TM8)	3.3 Å
R613 (NBD1)	R783 (TM8)	4.1 Å
K615 (NBD1)	K698 (TM7)	3.0 Å

C.

P-gp A (green)	P-gp 02-1-100 (red)	Distance
M91 (ECL1)	N1244 (NBD2)	3.3Å
Q90 (ECL1)	K1246 (NBD2)	2.9Å
Q90 (ECL1)	Q1050 (NBD2)	2.6Å
Q87 (ECL1)	Q1050(NBD2*)	4.3Å
E742 (ECL4)	S1089(NBD2)	3.4Å
E742 (ECL4)	S1089 (NBD2**)	4.0Å
P741 (ECL4)	S1034(NBD2)	3.7Å
K320 (ECL3)	S1034 (NBD2)	3.4Å
I318 (ECL3*)	Q1032 (NBD2)	3.1Å

D.

P-gp B (cyan)	P-gp 0-200-100 (yellow)	Distance
N1244 (NBD2)	M91 (ECL1)	3.3Å
K1246 (NBD2)	Q90 (ECL1)	2.9Å
Q1050 (NBD2)	Q90 (ECL1)	2.6Å
Q1050(NBD2*)	Q87 (ECL1)	4.3Å
S1089(NBD2)	E742 (ECL4)	3.4Å
S1089 (NBD2**)	E742 (ECL4)	4.0Å
S1034(NBD2)	P741 (ECL4)	3.7Å
S1034 (NBD2)	K320 (ECL3)	3.4Å
Q1032 (NBD2)	I318 (ECL3*)	3.1Å

E.

P-gp B (cyan)	P-gp 03-10000 (orange)	Distance
Q742 (ECL4)	U1247 (NBD2*)	4.1 Å
P741 (ECL4)	K1248 (NBD2*)	3.7 Å
K320 (ECL3)	Q1255 (NBD2)	3.2 Å
K320 (ECL3)	H1250 (NBD2)	3.8 Å
Q83 (ECL1)	Q1259 (NBD2)	3.6 Å
V84 (ECL1)	Q1259 (NBD2*)	3.2 Å
V84 (ECL1)	K1260 (NBD2)	3.7 Å

F.

P-gp A (green)	P-gp 03-10100 (purple)	Distance
K1260 (NBD2)	Q90 (ECL1)	4.1Å
Q1259 (NBD2*)	V84 (ECL1)	3.2Å
Q1259 (NBD2)	Q83 (ECL1)	3.9Å
Q1255 (NBD2)	K320 (ECL3)	3.2Å
H1250 (NBD2)	K320 (ECL3)	3.4Å
V1247 (NBD2)	E742 (ECL4)	4.1Å

G.

P-gp A (green)	P-gp 00-10000 (brown)	Distance
V688 (TM7)	R613 (NBD1)	2.9Å, 3.0Å
A691 (TM7)	D609 (NBD1)	3.1Å
P690 (TM7)	D609 (NBD1)	4.2Å
T1011 (TM12)	Q417 (NBD1)	3.2Å
T1011(TM12*)	K414 (NBD1)	3.2Å
E1013 (TM12)	K412 (NBD1)	4.3Å
A1087 (NBD2)	K391 (NBD1)	3.5Å
K1098 (NBD2)	D450 (NBD1)	3.4Å
K1098(NBD2*)	Q385 (NBD1)	2.4Å
Q1020 (NBD2)	Q385 (NBD1)	3.2Å
G1021 (NBD2)	N383 (NBD1)	2.5Å

H.

P-gp B (cyan)	P-gp 00010000 (grey)	Distance
R613 (NBD1)	V688 (TM7)	2.9Å, 3.0Å
D609 (NBD1)	A691 (TM7)	3.1Å
D609 (NBD1)	P690 (TM7)	4.2Å
Q417 (NBD1)	T1011 (TM12)	3.2Å
K414 (NBD1)	T1011 (TM12*)	3.2Å
K412 (NBD1)	E1013 (TM12)	4.3Å
K391 (NBD1)	A1087 (NBD2)	3.5Å
D450 (NBD1)	K1098 (NBD2)	3.4Å
Q385 (NBD1)	K1098 (NBD2*)	2.4Å
Q385 (NBD1)	Q1020 (NBD2)	3.2Å
N383 (NBD1)	G1021 (NBD2)	2.5Å

4.2.3.4 Retention of crystal contact residues in human P-gp

An important overall aim of this thesis was to investigate crystallographic differences between mouse and human P-gp. Therefore, it was of interest to compare the residues involved in mouse P-gp crystal contacts to the equivalent human P-gp residues, as it might yield insight into crystallization differences between species. To complete this analysis, an alignment of mouse and human P-gp was performed. It is important to note that there is a four-amino-acid length difference in overall length between mouse and human P-gp, with human P-gp consisting of 1280 residues and mouse P-gp consisting of 1276 residues (Figure 4.18). Three of the four residues absent from mouse P-gp are clustered in the first extracellular loop (ECL1) of human P-gp, making ECL1 of mouse and human P-gp different in size and structure. In addition to the species differences that account for differences in ECL1, the reported mouse P-gp crystal structure was generated using a glycosylation-deficient mutant of mouse P-gp. There are three N-linked glycosylation sites present in ECL1 located at positions N83, N87, and N90 (249) in mouse P-gp (all three residues are involved in crystal contacts in the mouse P-gp structure). While these glycosylation sites are conserved between wild-type mouse and human P-gp, the cDNA that was used to express protein for

crystallization in the reported structures of mouse P-gp contain asparagine to glutamine mutations at each glycosylation site to prevent the protein from being glycosylated in the expression system. Due to these factors, residues in ECL1 that were involved in crystal contacts of mouse P-gp were not considered to have a homologous human P-gp counterpart, as the structure and composition of ECL1 varies significantly between species.

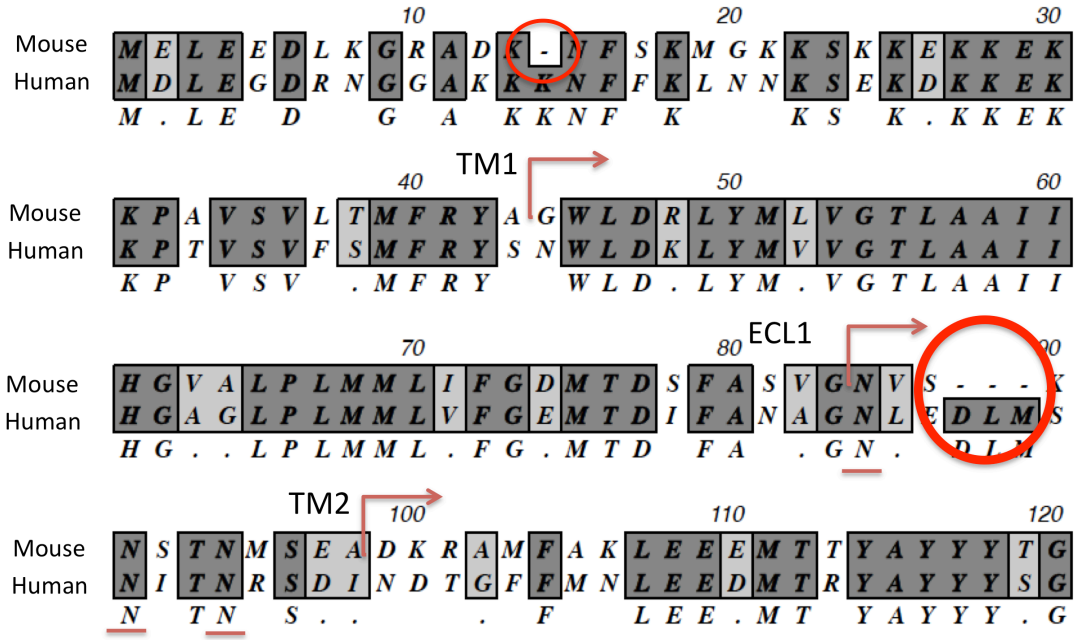


Figure 4.18 Alignment of the first 120 amino acid residues of human and mouse P-gp. Human and mouse P-gp consist of 1280 and 1276 amino acid residues, respectively. The four different residues between human and mouse P-gp are indicated by dash marks and are circled in red. The approximate boundaries for TM1, ECL1, and TM2 are indicated above the alignment. Additionally, N-linked glycosylation sites are underlined in red.

Pairs of amino acid residues that constitute a mouse P-gp crystal contact were compared to the analogous residues in human P-gp to determine if the pair of residues was conserved between species. For example, in the asymmetric unit of mouse P-gp, residues R138 (P-gp A) and Q515 (P-gp B) interact. While an alignment showed that the homologous residue of R138 is conserved in human P-gp (R142), the homologous residue to Q515 was not conserved in human P-gp (K519). Thus, this crystal contact was considered non-

conserved between mouse and human P-gp. Using this approach, it was possible to compare pairs of residues involved in mouse P-gp crystal contacts to the analogous pairs of residues in human P-gp.

Twelve of 15 pairs of residues that constitute the asymmetric unit crystal contacts of P-gp are retained in human P-gp. However, only nine of 40 contacts between the asymmetric unit of P-gp and neighboring P-gp molecules are conserved in human P-gp. Overall, out of the 55 unique pairs of crystal contacts (15 contacts in the asymmetric unit and 40 between the asymmetric unit and neighboring P-gp molecules) present in mouse P-gp, 21 contacts are hypothetically conserved in human P-gp. While it is unlikely that all crystal contacts present in the mouse P-gp crystal structure are crucial for crystallization, this analysis reveals that only 38% of contacts would be retained in human P-gp, a fact that could significantly contribute the differences in crystallizability between mouse and human P-gp. As noted previously, one striking difference between mouse and human P-gp is the length and composition in ECL1. ECL1 of mouse P-gp appears crucial for the interactions between P-gp A and P-gp 02-1-100, P-gp B and P-gp 0-200-100, and P-gp A and P-gp 03-10100 (Table 4.4). This alone may contribute to the differences in crystallization between species.

4.2.3.5 Retention of crystal contact residues in P-gp chimeras

Chimeras were generated as a strategy to investigate the crystallographic differences between mouse and human P-gp. Therefore, it was of interest to assess to what extent chimeras retained the crystal contacts present in mouse P-gp. Similar to the analysis performed between wild-type human and mouse P-gp, crystal contacts between mouse P-gp and the HHMM, HMHM, HMHH, and HHHM chimeras were also analyzed for each of the 55 unique crystal contacts present both in the mouse P-gp asymmetric unit and between neighboring P-gp molecules. Table 4.5 displays the percentage of contacts conserved in the P-gp chimeras generated in this thesis project.

Table 4.5 Crystal contacts retained in P-gp chimeras. Chimeras were analyzed to determine if amino acid residues that participate in crystal contacts in the wild-type mouse P-gp crystal structure would be hypothetically conserved in chimeras. The percentage of conserved contacts is indicated for each chimera.

Chimera	Conserved crystal contacts (%)
HHMM	53%
HMHM	62%
HMHH	55%
HHHM	43%

Crystal contact residues in chimeric P-gp are retained to varying degrees (Table 4.5). The HMHM chimera displays the highest retention (62%) of the crystal contact residues observed in wild-type mouse P-gp. The HHHM chimera displays the lowest retention (43%) of contacts, a percentage that is only slightly higher than the amount of contacts retained between wild-type human and mouse P-gp (38%). As none of the chimeras generated contained mouse TMD1 and therefore mouse ECL1, a significant amount of contacts (18% of all contacts observed in mouse P-gp contain residues of ECL1) were not retained in any of the chimeras studied. Overall, while the chimeras introduced various amino acid residues of mouse P-gp into human P-gp that may be important for the crystallization of P-gp, the absence of mouse ECL1 in the chimeras may inhibit the formation of crystals.

4.3 Discussion

This chapter describes the generation of human-mouse chimeras by the exchange of homologous domains. *MDR1* and *mdr1a* cDNA were first modified to include silent restriction enzyme sites located between domain boundaries of mouse and human P-gp. Restriction enzyme digestion and ligation of various domains of mouse and human P-gp resulted in the generation of HHMM, HHHM, HMHH, and HMHM chimeras. HHHH_WT P-gp was re-generated as a positive control construct to validate the PCR, cloning, and ligation strategies used to generate chimeras. Additionally, this chapter analyzed the crystal contacts observed in the reported mouse P-gp structure. The amino acid residue pairs that participate in crystal contacts of mouse P-gp were compared to the analogous residues in both wild-type human P-gp and the chimeras under study.

Results presented in this chapter demonstrate the inability to obtain plasmids containing *mdr1a* cDNA upon transformation into *E. coli*. For example, pTOPO_mdr1a_Nterm and pTOPO_MMMM_Opti could not be cloned in *E. coli.* Generation of plasmids containing individual domains of mouse P-gp was used as a strategy to circumvent the high mutational rate observed during cloning of full-length *mdr1a* in *E. coli* (Section 4.2.2). At the time of chimera generation, a parallel investigation into the presence of a cryptic bacterial promoter as a potential cause of toxicity of *mdr1a* cDNA to bacteria was under examination; Chapter 3 describes the presence of a cryptic bacterial promoter in the first 200 bp of *mdr1a*, capable of protein expression in bacteria (Section 3.2.6). Additionally, an *mdr1a* cDNA containing a methionine to leucine mutation at residue position 107 (M107L) that was able to reduce *mdr1a* cDNA toxicity, possibly through the elimination of a translational start site, was characterized (Section 3.2.7). As chimera generation and investigation into *mdr1a* cDNA toxicity were carried out in parallel, chimeras generated in this chapter did not use M107L *mdr1a* cDNA. However, future chimera generation may incorporate the M107L mutation as this reduces *mdr1a* toxicity.

The refined mouse P-gp crystal structure was analyzed to determine the amino acid residues that form the crystal contacts both within the asymmetric unit and between the asymmetric unit and neighboring P-gp molecules (Section 4.2.3.2 and Section 4.2.3.3). Crystal contacts are of crucial importance to crystallization, as the intermolecular forces that comprise the contacts are the basis of crystal formation. One hypothesis to account for differences in crystallization between mouse and human P-gp is that amino acid residues that are important for the intermolecular forces required to form crystals are reduced or absent in human P-gp. Therefore, the characterization of the crystal contacts occurring in the reported mouse P-gp crystal structure and a comparison to see if those contacts would be, hypothetically, conserved in human and chimeric P-gp was of interest to this thesis. The analysis revealed at least 55 unique crystal contacts within the P-gp asymmetric unit or between the asymmetric unit and neighboring P-gp molecules. Furthermore, it revealed that only 38% of these interacting residue pairs are retained in human P-gp. In particular, this analysis highlighted significant differences in the size and conformation of ECL1 between mouse and human P-gp. While intriguing, it is however, unknown to what extent contacts in ECL1 influence the crystallization of mouse P-gp. Future experimental investigation of this hypothesis is warranted. Importantly, the chimeras generated in this chapter displayed increased retention of residue pairs that are important for crystal contacts, perhaps favorably impacting the ability of chimeras to crystallize.

There are several potential shortcomings of this analysis that should be noted. Firstly, this analysis assumes a conserved pair of homologous residues between mouse P-gp and human or chimeric P-gp would form a similar crystal contact during crystallization screening. However this may not be the case, as there may be other differences between mouse and human/chimeric P-gp besides the individual residues that influence crystal formation, such as differences in protein aggregation between homologs (see Section 6.2.2.4). Secondly, this type of analysis assumes that both human and chimeric P-gp would

crystallize in the same conformation as mouse P-gp. As P-gp is known to undergo a large conformational change between the open and closed states, it can be assumed that different conformations of P-gp would have dramatically different crystal contacts, and thus the analysis completed herein would not be relevant. Lastly, the analysis presented in this chapter compares amino acid residue pairs that comprise crystal contacts in mouse P-gp to the analogous residues in human and chimeric P-gp. Residue pairs that were not identical between mouse and human/chimeric P-gp were considered non-conserved. However, this type of analysis does not consider the similarity between non-conserved residues. Amino acids may be considered either similar or dissimilar due to properties of size, polarity, and charge. It is likely that numerous non-conserved crystal contacts between mouse and human/chimeric P-gp may actually be composed of amino acid residues with similar properties, therefore retaining the crystal contact of interest. Thus, both identity and similarity of residues on mouse P-gp involved in intermolecular crystal packing forces between human and chimeric P-gp should be analyzed to more fully understand factors that may contribute to differences in the ability to crystallize mouse versus human P-gp.

While certain limitations to this analysis should be kept in mind, the analysis of residues that may be important for the crystallization of mouse P-gp also provides a basis for future chimera generation efforts. For example, chimeras incorporating mouse ECL1 into human P-gp may increase the likelihood of chimeric crystal formation. Or, a chimera composed of a human P-gp background containing all of the residues involved in mouse crystal contacts could be generated. Additionally, a study analyzing the similarity of non-conserved contacts between human and chimeric P-gp could be completed. It is likely that not all crystal contacts contribute equally to differences in crystallization between mouse and human P-gp, and there may be a few key dissimilar residues between mouse and human P-gp that may more significantly contribute to differences in crystallizability between species. Therefore, a future study may include the generation a chimera in a human P-gp background

where the dissimilar crystal contact residues in human P-gp are mutated to the analogous residues in mouse P-gp.

4.4 Conclusions

This study generated four chimeras by the exchange of homologous domains between mouse and human P-gp. These chimeras were next functionally characterized, and used to investigate crystallographic differences between mouse and human P-gp. Additionally, an analysis of the crystal contacts present in the reported mouse P-gp crystal structure was completed. This analysis yielded insight into the regions of mouse P-gp that are involved in crystallization, and information reported herein may act a basis for the generation of future chimeras.

Chapter 5: Functional characterization of P-glycoprotein chimeras

5.1 Introduction

Functional characterization of chimeras is a necessary prerequisite to structural investigations. Only chimeras that are functional and display physiological properties will be expressed, purified, and screened for crystallization in this thesis project. Functional characterization is typically assessed in advance of crystallization. Proteins that are found to be non-functional are often triaged or excluded from crystallization studies, as the relevance of their structures may be of questionable significance.[2] At the start of this project, it was not known whether chimeric P-gp would be functional. Therefore, a full functional characterization was carried out to determine the physiological relevance of the chimeric proteins constructed.

In addition to functional characterization as it pertains to crystallographic studies, the construction and characterization of P-gp mutants is often employed as a strategy to investigate mechanistic aspects of P-gp-mediated transport. For example, mutational studies have investigated the importance of conserved glutamic acid residues at positions E556 (in NBD1) and E1201 (in NBD2) in the Walker B domains of P-gp. When both glutamic acid residues were mutated to glutamine (E556/1201Q; also referred to as the "EQ mutant"), ATP hydrolysis and drug transport were abrogated, highlighting the necessity of the conserved glutamic acid residues for functional P-gp (260, 261). Mutational studies can also be used to determine residues that contribute to drug recognition and binding in the central cavity. For example, one study used site-directed mutagenesis to mutate various residues in TM11 (262). When mutant P-gp was exposed to the P-gp ATPase stimulators verapamil, vinblastine, and colchicine, ATPase activity varied significantly among P-gp mutants,

[2] Structures of non-functional proteins have been reported (250-252). However, these are typically proteins that have been purposefully altered to be non-functional as a means to lock the protein in one conformation, as conformationally homogenous protein may increase the likelihood of crystal formation. Additionally, structures of protein domains or subunits have been crystallized as a strategy to gain insight into structures of proteins that have proven difficult to crystallize (253-259).

indicating residues that may be important in drug recognition and translocation of these compounds (262). While chimeras differ from the abovementioned examples in that they are not mutants generated by single amino acid residue substitutions, they may be viewed as P-gp mutants that could potentially elucidate mechanistic characteristics of P-gp. For example, constructs with chimerized NBDs and chimerized TMDs have been generated, and thus functional differences that occur may provide insight into how TMDs and NBDs comprised of different species are able to communicate to transport substrates.

Various methods have been established to assess the transport function of P-gp. In this thesis project, several assays were used to characterize human wild-type and codon-optimized P-gp and chimeric P-gp constructs. These studies aimed to 1) assess whether P-gp chimeras were expressed and trafficked to the plasma membrane upon transduction, 2) determine if chimeras were functional and able to confer a MDR phenotype, and 3) elucidate functional differences, if any, between chimeras and wild-type P-gp homologs. These aims were addressed by the use of flow cytometry assays, ATPase assays, and western blotting. Flow cytometry and ATPase assays are introduced below.

5.1.1 Flow cytometry assays for P-gp cell membrane expression and measurement of efflux capacity

Flow cytometry assay protocols have been developed to determine various aspects of ABC transporter function (263-265). The flow cytometry assays used in this study can be broadly divided into cell membrane expression assays and transporter efflux assays (Figure 5.1). Cell membrane expression assays use anti-P-gp antibodies that recognize extracellular epitopes, such as the mouse monoclonal antibody MRK-16, to quantify cell surface P-gp expression (Figure 5.1.A). Thus, the amount of P-gp expressed on the cell surface is proportional to the fluorescence intensity of the cells. Flow cytometry-based efflux assays are used to assess P-gp transport capacity. Here, cell permeable fluorophores that are also

substrates of P-gp are used to measure the ability of P-gp to efflux substrates (Figure 5.1.B). Cells that do not express P-gp, or do not express functional P-gp, at the cell surface will accumulate higher intracellular concentrations of the fluorophore. In contrast, cells that express functional P-gp will efflux the fluorophore leading to lower intracellular fluorescence levels (a shift to the left in the histogram). Common fluorescent probes used for this purpose are Rhodamine 123 (Rh123), Calcein AM (CAM), and JC-1.[3] The extent of P-gp-mediated efflux is observed by the co-incubation of cells with both fluorophores and P-gp inhibitors. In this case, inhibited P-gp is no longer able to efflux fluorophores and cells accumulate levels of intracellular fluorescence to levels equal to non-P-gp-expressing cells (Figure 5.1.B).

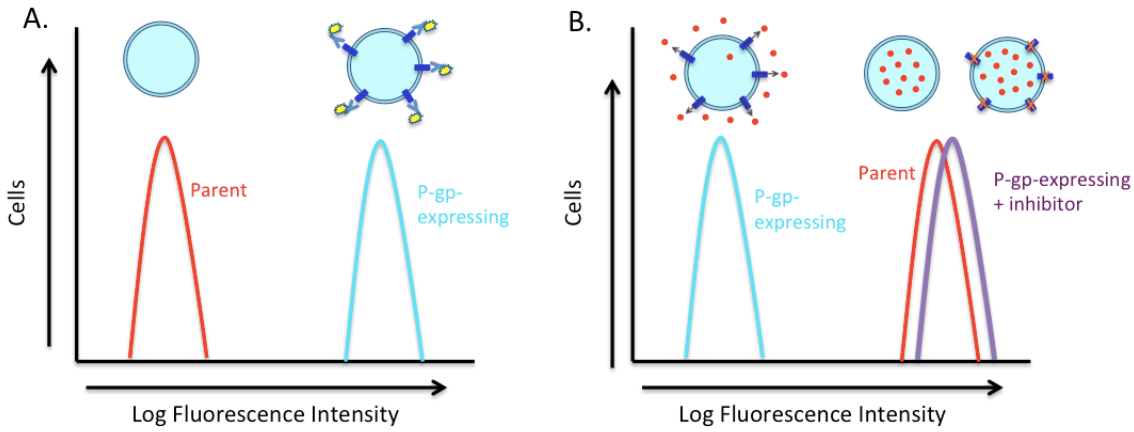


Figure 5.1 Scheme of flow cytometry cell surface expression and efflux assays. **A)** Schematic showing a typical cell membrane expression assay. Fluorescence intensity is proportional to P-gp-specific antibody binding and therefore P-gp expression. **B)** Schematic showing a typical transporter efflux assay. Membrane-permeable fluorophores that are also P-gp substrates are used to assess P-gp function. P-gp-expressing cells have low intracellular fluorescence and cells that do not express P-gp accumulate high levels of cell-permeable fluorophores. The presence of a P-gp inhibitor is able to block fluorophore efflux, allowing for the intracellular accumulation of fluorescence and a rightward histogram shift.

[3] Rh123 and JC-1 are both cationic fluorescent compounds that readily diffuse into cell in the absence of P-gp. In the presence of P-gp, both compounds are actively effluxed, resulting in low intracellular fluorescence. While Rh123 is efficient at detecting high P-gp expression, it may be insufficient at detecting low levels of P-gp expression often found in clinical samples (266, 267). JC-1 was developed as a high-affinity fluorescent probe to study P-gp expression in AMLs that express low, but clinically relevant, levels of P-gp (266).

5.1.2 ATPase assays

ATPase assays are often employed to assess the biochemical properties of ATP binding and hydrolysis in the presence or absence of P-gp modulators (268-270). ATPase activity is required for the transport of substrates by P-gp. Thus, these assays can be used to determine the inhibitory or stimulatory properties of compounds on P-gp ATPase activity. ATPase activity can be assayed using several methods, including a colorimetric method to quantify inorganic phosphate (P_i) released as a result of the hydrolysis of ATP to ADP and P_i. Compounds that are known ATPase stimulators were used to probe for functional activity of chimeric and codon-optimized P-gp compared to human wild-type P-gp. It is important to note that ATPase assays do not measure the transport of substrates, only the ability of a compound to stimulate or inhibit P-gp ATPase, in turn demonstrating that the compound binds to the drug-binding pocket and the communication between the drug-binding pocket and NBDs is operating similarly to wild-type protein. Thus, ATPase assays are used in combination with other functional assays, such as flow cytometry efflux assays, to more fully understand the interaction of a transporter with modulators.

To perform an ATPase assay, P-gp is first overexpressed in insect cells from which crude membranes are generated (Section 2.2.40). While membranes can be prepared from various expression systems (mammalian, insect, yeast, etc.), insect cells infected with baculovirus encoding P-gp are often used for the preparation of crude membranes for several reasons. Firstly, insect cells can be grown to high densities in suspension culture, thus generating a large quantity of membranes. Secondly, baculovirus infection as a method of heterologous P-gp expression generates high levels of P-gp expression (Section 1.4.4). Thirdly, the presence of F- and P-type ATPases in the cell membrane of insect cells is lower than in mammalian cells and thus there is less P_i generated from “background” ATPases, therefore reducing noise in the assay (268). The limitations of mammalian membranes further support the use of insect cell crude membranes for ATPase assays.

To perform an ATPase assay, crude membranes are incubated in the presence of assay buffer containing the F- and P-type ATPase inhibitors sodium azide, ouabain, and ethylene glycol tetraacetic acid (EGTA) (268). These compounds have been shown to inhibit background ATPase activity but not P-gp, and are used to decrease background levels of non-specific ATP hydrolysis. P-gp exhibits a basal level of ATP hydrolysis in the absence of P-gp stimulators. Basal ATPase activity is measured by the incubation of crude membranes with and without sodium orthovanadate (V_i), a phosphate analog that has been shown to prevent the release of ADP from various ATPases including P-gp, thus inhibiting all ATPase activity (271, 272). Basal ATPase activity is calculated by the following equation:

$$P_{i(Basal)} = P_i - P_{i\ (Vi)}$$

Where P_i is the amount of inorganic phosphate generated by crude membranes in the absence of a P-gp stimulator and $P_{i\ (Vi)}$ is the amount of inorganic phosphate generated in the presence of sodium orthovanadate.

To assess the ability of compounds to stimulate P-gp ATPase, compounds are incubated with crude membranes with and without V_i. Incubation with V_i is required to determine P-gp-specific hydrolysis of ATP. After a 20-minute incubation of crude membranes with a P-gp modulator and ATP, the reaction is stopped by the addition of sodium dodecyl sulfate (SDS), a process that inactivates P-gp and stops further ATP hydrolysis. The amount of P_i is measured by absorbance based on a colorimetric signal generated by the addition of ammonium molybdate, which, in the presence of P_i and a reducing agent, produces a blue color.

The ATPase-specific activity is expressed as nanomoles P_i/minute/mg of protein and the stimulation of this activity is reported in the form of fold stimulation as calculated by the following equation:

$$\text{Fold Stimulation} = \frac{P_{i\ (\text{Stimulator})} - P_{i\ (\text{Stimulator+ Vi})}}{P_{i\ (\text{Basal})}}$$

Where $P_{i\ (\text{Stimulator})}$ is the amount of inorganic phosphate generated in the presence of a P-gp stimulator, $P_{i\ (\text{Stimulator + Vi})}$ is the amount of inorganic phosphate generated in the presence of both a P-gp stimulator and sodium orthovanadate, and $P_{i\ (\text{Basal})}$ is the amount of basal ATP hydrolysis in the absence of a P-gp modulator.

5.2 Results

5.2.1 Transient transduction of HeLa cells with BacMam baculovirus

Transient expression of P-gp in HeLa cells was achieved by the use of BacMam virus genetically modified to express wild-type, codon-optimized, and chimeric P-gps (Section 2.2.32). P-gp constructs in the pTOPO_HHMM, pTOPO_HMHM, pTOPO_HMHH, and pTOPO_HHHM plasmids generated in this thesis project (Section 4.2.1.5 and Section 4.2.2.4) were cloned into pDest-625 vectors by Gateway cloning for generation of BacMam virus by the Protein Expression Laboratory at the FNLCR (Section 2.2.12). Subsequently, BacMam virus was generated by FNLCR (Section 2.2.30). All subsequent steps were completed by the author in the Laboratory of Cell Biology (NIH).

Previous studies have shown that the host mammalian cell line can greatly influence BacMam infection and expression levels of the transduced protein. For example, cervical adenocarcinoma-derived HeLa cells exhibit high levels of P-gp expression, while mouse embryonic fibroblast-derived NIH 3T3 cells showed little or no detectable expression of newly introduced proteins after transduction (184). Additionally, it was necessary to use a cell line with no detectable expression of ABC transporters that could confound the results obtained during functional characterization. Thus, the HeLa cell line was chosen as an ideal candidate for BacMam transduction, as it has been extensively characterized and is known not to express P-gp or ABCG2 (184). Transductions were carried out by incubating BacMam virus in a 1:50 cell to virus particle ratio in the presence of 10 mM sodium butyrate (NaBT). HeLa cells were incubated for 24 hours after the addition of virus and subsequently harvested for expression and functional assays.

5.2.2 Effect of sodium butyrate on transient transduction by BacMam

Previous reports of BacMam transduction have used NaBT to increase the expression of foreign proteins in mammalian cells (184, 273, 274). NaBT is known to have

various effects on mammalian cells, such as growth inhibition and altered gene expression, due to its activity as a non-specific histone deacetylase (HDAC) inhibitor (275, 276). Concentrations exceeding 5-10 mM NaBT are typically cytotoxic to mammalian cells (277). However, the addition of NaBT during BacMam transduction has been shown to increase protein expression due to its ability to de-condense the baculovirus genome and therefore enhance transcription of the foreign protein (274). Given that NaBT can be cytotoxic to mammalian cells, and has variable effects on expression levels of transduced protein, it was necessary to determine the effect of NaBT on chimeric P-gp expression in HeLa cells.

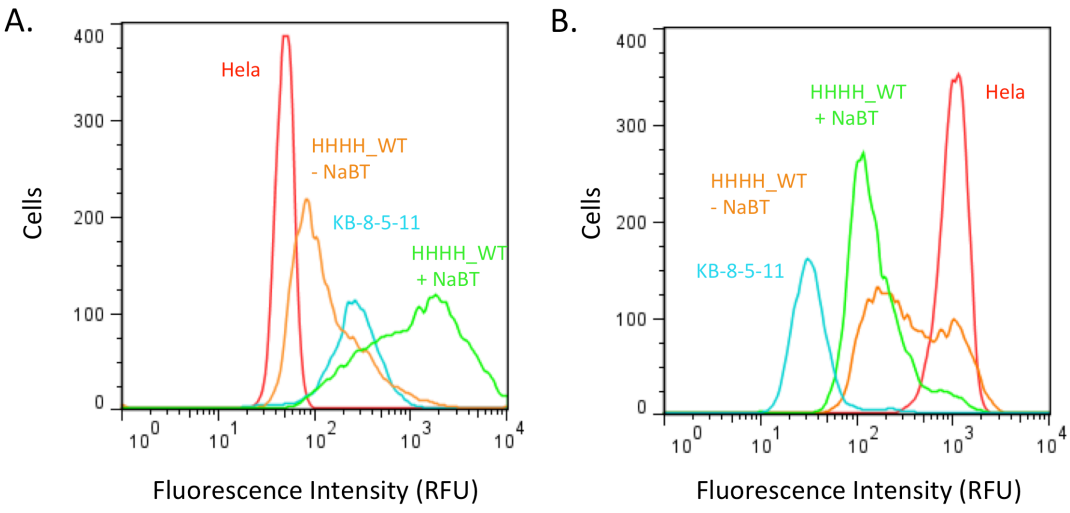


Figure 5.2. Cell surface expression and efflux flow cytometry assays with and without the addition of 10 mM NaBT during transduction of HeLa cells with BacMam baculovirus. A) MRK-16 expression assays (Section 2.2.23) show that the addition of 10 mM NaBT increases the level of HHHH_WT cell surface expression (HHHH + NaBT, green) versus cells that were transduced in the absence of NaBT (HHHH – NaBT, orange). Broad histogram peaks of both transductions indicate a range in MRK-16 antibody binding and therefore HHHH_WT expression. **B)** Efflux assay (Section 2.2.24) completed in parallel with MRK-16 assays to determine the efflux capacity of HeLa cells transduced with and without 10 mM NaBT. Histograms shown are representative of three independent studies. The mean fluorescence intensity of the histogram was determined by FloJo Single Cell Analysis Software.

The ability of NaBT to increase chimeric protein expression was tested by parallel transductions with and without the addition of 10 mM NaBT (Figure 5.2.A). This thesis project uses the mean fluorescence intensity of the histogram as expressed in relative fluorescence units (RFU) to quantify P-gp expression or transport activity. HeLa cells, which do not express P-gp, had a mean fluorescence intensity of 48 RFU. HeLa cells that were transduced with the HHHH_WT BacMam virus in the presence of 10 mM NaBT expressed higher levels of P-gp, as shown by increased fluorescence intensity due to MRK-16 binding (mean fluorescence intensity of 1023 RFU). In contrast, HeLa cells transduced in the absence of NaBT had markedly less MRK-16 binding (mean fluorescence intensity of 132 RFU). These findings indicate that the addition of NaBT during transduction increases levels of protein expression over 20-fold compared to non-transduced HeLa cells assuming linearity response. Transductions both with and without NaBT show a wide range of MRK-16 binding, as reflected by the broad histogram. This indicates variability in the level of HHHH_WT expression in HeLa cells after transduction with BacMam virus.

Efflux assays were performed to determine the functionality of cells transduced with the HHHH_WT construct with and without 10 mM NaBT (Figure 5.2.B). Non-transduced HeLa cells in the presence of Rh123 had a mean fluorescence intensity of 961 RFU. Cells transduced with NaBT expressed high levels of P-gp at the cell surface, and were thus able to efflux Rh123 from cells resulting in a mean fluorescence intensity of 157 RFU. Interestingly, cells transduced without NaBT demonstrated varying degrees of Rh123 efflux, ranging from virtually no Rh123 (~1000 RFU) efflux to efflux comparable to cells that were transduced in the presence of NaBT (~100 RFU). This can be seen visually by the wide histogram that spans fluorescence intensities from the fluorescence values of non-transduced HeLa cells to HeLa cells transduced with the HHHH_WT construct in combination with NaBT. Therefore, even though the expression level of the HHHH_WT construct is significantly lower when transduced in the absence of NaBT (Figure 5.2.A), even low levels of HHHH_WT expression

were able to efficiently efflux Rh123 from cells. This observation can be explained by the spare transporter effect, which posits that when transporters are expressed in excess, decreased amounts of transporter expression are still able to achieve maximal transport of substrates (33). Thus, decreases in the cellular accumulation of substrates can only be observed after a substantial percentage of the transporter is absent or inhibited.

5.2.3 Reproducibility of transduction of HeLa cells by BacMam virus

Multiple transductions were required to complete a full characterization of chimeras and it was necessary that the transduction efficiency was consistent between experiments. The reproducibility of transduction efficiency was assessed by comparing P-gp expression and efflux capacity of HHHH_WT P-gp from triplicate transduction experiments (Table 5.1). Triplicate assays were carried out by parallel transductions of the HHHH_WT BacMam virus into HeLa cells (Section 2.2.32). After 24 hours, each transduction was harvested and prepared separately for analysis by flow cytometry. Additionally, three technical replicates of each transduction were prepared to probe for variations that might occur during the assay (pipetting error, technical error from the flow cytometer, etc).

Table 5.1 Expression and efflux assays to determine reproducibility of BacMam transduction experiments. Triplicate transductions and flow cytometry assays were carried out to determine P-gp expression (Section 2.2.23) and efflux capacity (Section 2.2.24). **A)** Table showing the average and standard deviations of MRK-16 expression assays from technical replicates. **B)** Table showing functional data of Rh123 efflux capacity for technical replicates. Data is represented as the mean fluorescence intensity of the histogram as expressed in relative fluorescence units (RFU).

A.

	Mean Fluorescence Intensity (RFU)
Transduction 1	1710 ± 32
Transduction 2	1947 ± 80
Transduction 3	2028 ± 68

B.

	Mean Fluorescence Intensity (RFU)
Transduction 1	680 ± 32
Transduction 2	709 ± 27
Transduction 3	706 ± 98

Transduction of HeLa cells with BacMam virus expressing HHHH_WT P-gp resulted in P-gp expression to equivalent levels, as shown by MRK-16 binding (Table 5.1.A). Additionally, transduced cells were able to efflux Rh123 from cells to a similar extent (Table 5.1.B). Technical errors for both expression and efflux assays were similarly low. It was concluded that there is low variability among transduction experiments both in amount of P-gp expressed and subsequent functionality of expressed P-gp. In the remainder of this study, flow cytometry efflux assays are the result of a single transduction event prepared in triplicate to account for technical error.

5.2.4 Confirmation of chimeric P-gp by PEPG-13 western blotting

Chimeras were generated by a series of restriction enzyme digests, ligations, and subsequent analysis by DNA sequencing to validate that the correct sequence was achieved (Chapter 4). Subsequently, chimeric cDNAs were sub-cloned into shuttle vectors for the generation of BacMam virus. Western blotting using the anti-P-gp antibody PEPG-13 was carried out to confirm that all chimeras were transferred correctly from plasmids used for cloning and molecular biology in *E. coli* to vectors used to generate recombinant BacMam virus. The PEPG-13 antibody recognizes an epitope in NBD1 between amino acid residues 592 to 636 that is specific to human P-gp (268, 271). Given this, the PEPG-13 antibody would be expected to recognize human wild-type and codon-optimized P-gp and the HHHM and HHMM chimeras. As the HMHH and HMHM chimeras contain mouse NBD1, reactivity with PEPG-13 would not be expected.

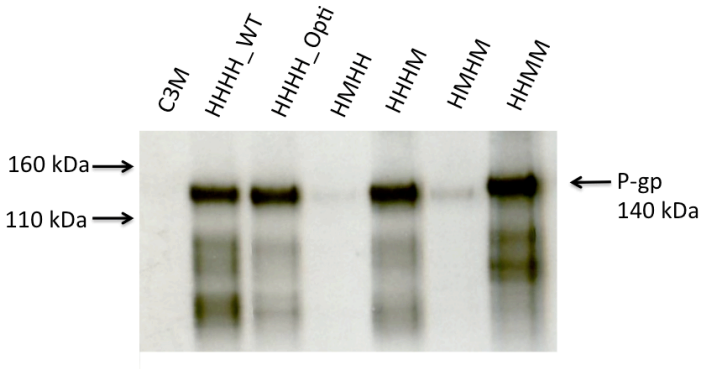


Figure 5.3 PEPG-13 western blotting of chimeric P-gp. Crude membranes were isolated from insect cells infected with recombinant baculovirus expressing P-gp chimeras (Section 2.2.40). Crude membranes (1.0 µg/lane) were subjected to SDS-PAGE (Section 2.2.19) and subsequence PEPG-13 western blotting (Section 2.2.20). Membranes from the drug-selected cell line C3M expressing mouse P-gp (218) were included as a negative control. Additionally, PEPG-13 western blotting was repeated with lysates harvested from transduced HeLa cells (not shown), which yielded equivalent results.

Western blotting of wild-type, codon-optimized and chimeric P-gp generated the expected pattern of PEPG-13 antibody reactivity (Figure 5.3). As expected, wild-type mouse P-gp and the HMHH and HMHM chimeras were not recognized by PEPG-13, while wild-type, codon-optimized, and the HHHM and HHMM showed strong reactivity. Faint bands are observed in the HMHH and HMHM lanes of the gel, but this is most likely due to "spill-over" of the protein from neighboring lane. These results confirmed that BacMam vectors allow for chimeric P-gp expression.

5.2.5 Expression of human wild-type, codon-optimized and chimeric P-gp in HeLa cells

Transient expression of chimeric, codon-optimized and wild-type human P-gp was assessed by western blotting. The C219 antibody binds to a 14-residue epitope in the ATP-binding regions of NBDs, consisting of the sequence VVQEALDKAREGRT, that is highly conserved among P-gp homologs (242). C219 recognizes all mammalian P-gp, including mouse and human P-gp. Thus, it was expected that C219 should also bind to all P-gp chimeras generated, and would be an ideal antibody to confirm expression of the wild-type

mouse P-gp and the HMHH and HMHM chimeras that were not recognized by the PEPG-13 antibody (Figure 5.3). Two cell lines known to express mouse and human P-gp were also included as positive controls: the human KB-8-5-11 cell line and the mouse C3M cell line. These cell lines have been previously characterized and are known to possess a MDR phenotype due to P-gp expression (217, 278).

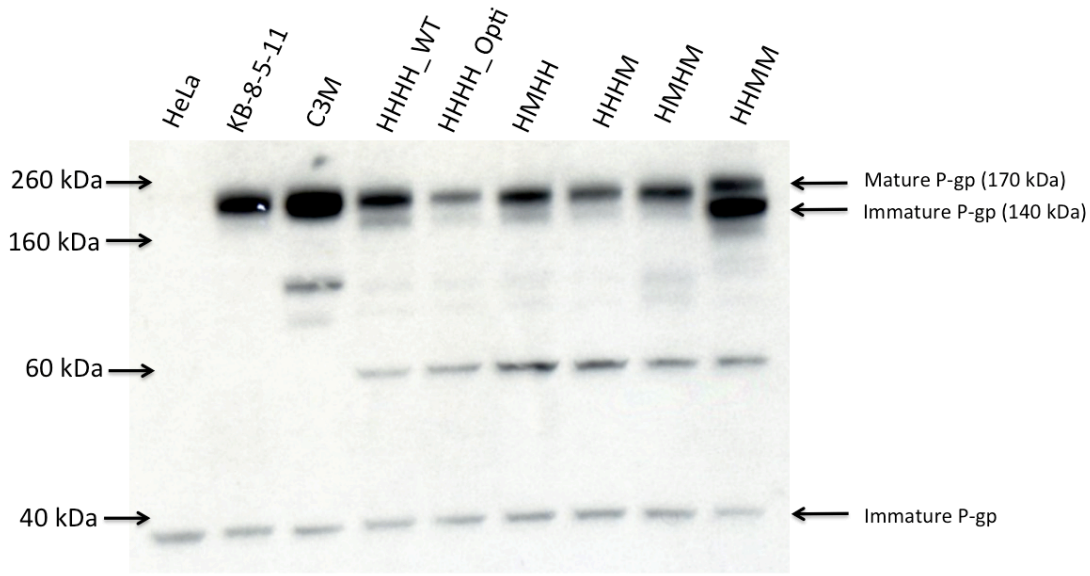


Figure 5.4 C219 western blot of chimeric P-gp. SDS-PAGE was performed with untransduced HeLa cells and HeLa cells transduced with wild-type human P-gp (HHHH_WT), codon-optimized P-gp (HHHH_Opti), and the HHHM, HMHH, HMHM, and HHMM chimeras (5.0 µg/lane) (Section 2.2.19) Western blotting with the C219 monoclonal antibody was carried out (Section 2.2.20). Additionally, the drug-selected cell lines C3M and KB-8-5-11 expressing wild-type mouse and human P-gp, respectively, were included. Size markers from a protein standard were used to assess molecular weight. Fully glycosylated P-gp (indicated as mature P-gp) is a 170 kDa protein and non-glycosylated P-gp is 140 kDa (indicated as immature P-gp). GAPDH was used as a loading control (37 kDa). The blot shown is representative of lysates generated from three independent transduction events.

Several differences were observed in the expression levels and molecular weights of the P-gp chimeras (Figure 5.4). At least two bands between 140 and 170 kDa were present for all HeLa cells that were transduced with chimeric, codon-optimized, and wild-type human P-gp. Wild-type mouse and human P-gp contains three N-linked glycosylation sites in

the first extracellular loop of TMD1. Fully glycosylated P-gp has a molecular weight of 170 kDa, while non-glycosylated P-gp is approximately 140 kDa (249). Previous studies have shown that the degree of glycosylation does not affect P-gp transport function (249). It is likely that the two distinct bands appearing at 170 and 140 kDa are both glycosylated (often referred to as "mature") and non-glycosylated ("immature") P-gp, respectively. While all transduced HeLa cells have a dominant glycosylated form migrating at 170 kDa and a secondary form migrating at 140 kDa, the HHMM chimera does not show this glycosylation pattern, having a large amount of non-glycosylated P-gp. Both wild-type mouse or human P-gp from the C3M or KB-5-8-11 drug-selected cell lines appear to contain fully glycosylated P-gp, lacking the secondary band that is present in all transfection conditions (P-gp was not detected in untransfected HeLa cells). It is not currently known why BacMam transductions yield P-gp with variable glycosylation. However, previous reports using BacMam for the expression of human P-gp have indicated the presence of both glycosylated and non-glycosylated P-gp, and this was therefore not pursued further here (184). Levels of total P-gp expression also differ between chimeras. Codon-optimized P-gp (HHHH_Opti) and chimeras generated using codon-optimized cDNA (HHHM, HMHH, and HMHM) showed lower expression than wild-type P-gp (HHHH_WT). This could indicate that the codon optimization process for *P. pastoris* reduces P-gp expression in mammalian cells.

5.2.6 Cell surface expression of P-gp chimeras by MRK-16 binding

Subtle mutations in the primary amino acid sequence of proteins can potentially influence protein folding, trafficking, and degradation. A notorious example is found in cystic fibrosis, where the most common mutation in the *CFTR* gene (ABCC7) that leads to a disease state is due to a three base pair deletion resulting in the loss of a phenylalanine (ΔF508) residue from CFTR (279). The loss of this single amino acid residue leads to misfolded CFTR that is retained in the endoplasmic reticulum and subsequently degraded, therefore unable

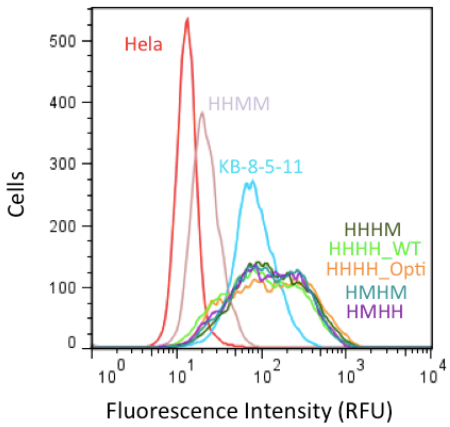
to reach the cell surface to carry out its physiologic function. The degradation of conformationally altered P-gp has also been observed. In one study, P-gp mutants containing amino acid substitutions or single amino acid deletions were retained in the endoplasmic reticulum and subsequently degraded (183, 280). Since the codon-optimization or chimerization of P-gp may have resulted in conformational changes that would target P-gp for degradation, it was necessary to determine if chimeric proteins were affected in P-gp processing and trafficking to the cell membrane. Cell surface expression of chimeras by flow cytometry analysis was used to determine differences in trafficking of codon-optimized and chimeric proteins.

Cell surface protein detection by flow cytometry generally requires the use of an antibody with reactivity to an external epitope. The P-gp antibody MRK-16 was used to determine the cell surface expression of transduced HeLa cells. MRK-16 is a conformationally-sensitive antibody that recognizes an external epitope composed of the extracellular loops of P-gp (281). MRK-16 binding is also specific to human P-gp, and thus C3M cells expressing mouse P-gp were excluded for this assay. It was unknown if MRK-16 would bind to P-gp with chimerized TMDs, such as the HHMM chimera. A literature search was unable to find a reported antibody with reactivity towards an external epitope of mouse P-gp. The P-gp antibody F4 has reported reactivity towards human and hamster P-gp, and thus it was hypothesized that the F4 antibody might also display reactivity towards mouse P-gp. However, under the experimental conditions tested, the F4 antibody did not show reactivity towards mouse P-gp therefore could not be utilized (data not shown).

Wild-type P-gp (HHHH_WT) and all chimeras containing both human TMDs (HHHH_Opti, HHHM, HMHM, and HHHM chimeras) were recognized by MRK-16 (Figure 5.5). This can be appreciated by the approximately 9-fold increase in mean fluorescence intensity indicating MRK-16 binding compared with non-transduced HeLa cells (Figure 5.5.B). The HHMM chimera, which contains chimerized TMDs and therefore a mixture of extracellular

loops from both mouse and human P-gp, showed a 1.6-fold increase in fluorescence as compared with non-transduced HeLa cells; this indicated a possible restricted recognition of the HHMM chimera by MRK-16. There were no noteworthy differences in MRK-16 binding between the HHHH_WT control construct and the HHHH_Opti, HHHM, HMHH, and HMHM chimeras, indicating that both wild-type human P-gp and chimeras recognized by MRK-16 were expressed to a similar extent on the cell surface. It can be concluded that the chimerization process did not greatly affect cell surface expression for the chimeras that showed reactivity with MRK-16.

A.



B.

Cell line	Mean Fluorescence Intensity (RFU)	Fold-Increase
KB-8-5-11	80.1	6.2
Hela	13.0	1.0
-HHHH_WT	95.0	7.3
-HHHH_Opti	121	9.3
-HHHM	124	9.5
-HMHH	111	8.5
-HMHM	117	9.0
-HHMM	21.6	1.7

Figure 5.5 Cell surface expression of chimeras as detected by MRK-16. A) Non-transduced HeLa cells (red) do not express detectable P-gp. KB-8-5-11 cells (light blue), known to express high levels of P-gp, were used as a positive control. Histograms shown are representative of three independent studies. **B)** The mean fluorescence intensity values of the histograms and the fold-change in mean fluorescence values as compared with non-transduced HeLa cells. The mean fluorescence intensity of the histogram was determined by FloJo Single Cell Analysis Software.

All transduced cells displayed heterogeneous P-gp-expression, as indicated by the wide histogram peaks (Figure 5.5.A), indicating a distribution of MRK-16 binding. As each flow cytometry event records 10,000 individual cells, it can be seen that a population of transduced cells has an array of P-gp expression levels that range from low expression to

expression that greatly exceeds levels of P-gp expression in KB-8-5-11 cells. Variation in P-gp-expression may indicate differences in transduction efficiency of individual cells. Broad P-gp expression in transduced HeLa cells has been observed in the literature, and results presented here are consistent with these published studies (183, 184).

5.2.7 Steady state Rh123 123, CAM, and JC-1 efflux assays

Rh123, CAM, and JC-1 efflux assays were used to establish chimera functionality. All three fluorophores are membrane-permeable and substrates of wild-type mouse and human P-gp. Rh123, CAM, and JC-1 are structurally diverse compounds, and are ideal candidates to use as probes for transport function of chimeras (Figure 5.6). It was not known whether the chimeras would have the same substrate profile as wild-type P-gp homologs. This was especially true for the HHMM chimera, which contains a chimeric drug-binding pocket. While there is no detailed understanding of where each substrate interacts inside the drug-binding pocket, it is probable that, due to the structural diversity of these compounds, they interact with distinct sites within the pocket (282). Therefore, the use of Rh123, CAM, and JC-1 may yield insight into distinct substrate profiles of the chimeras versus wild-type P-gp homologs.

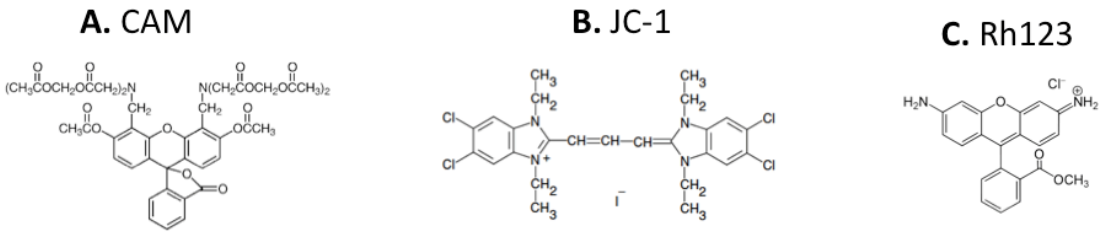


Figure 5.6 Structures of the fluorescent P-gp substrates CAM, JC-1, and Rh123. A) CAM, **B)** JC-1, and **C)** Rh123 are structurally diverse, exhibiting the ability of P-gp to transport a variety of molecules (283-285). Structures are reproduced from the open chemistry database PubChem.

Transduced HeLa cells were incubated for 45 minutes with each fluorophore. Previous studies have shown that cells expressing wild-type human P-gp are able to reach a steady state of Rh123 efflux within 5 minutes (183). As these experiments used an incubation period 15 times longer, these "steady state" assays were used to assess overall transporter function and differences in substrate specificities of chimeras, but are unable to reveal differences in the efflux rate of chimeras.

Various control conditions were also included for analysis. It is important to note that there have been few comprehensive studies characterizing the differences between P-gp homologs (87, 88, 90, 91). Thus, it was unknown if there would be differences in Rh123, CAM, and JC-1 transport between wild-type mouse and human P-gp homologs. The drug-selected cell lines KB-8-5-11 and C3M expressing human and mouse P-gp, respectively, were included. Additionally, non-functional EQ mutant P-gp (described in Section 5.1) was included to ensure that differences in fluorophore accumulation were independent of the transduction process.

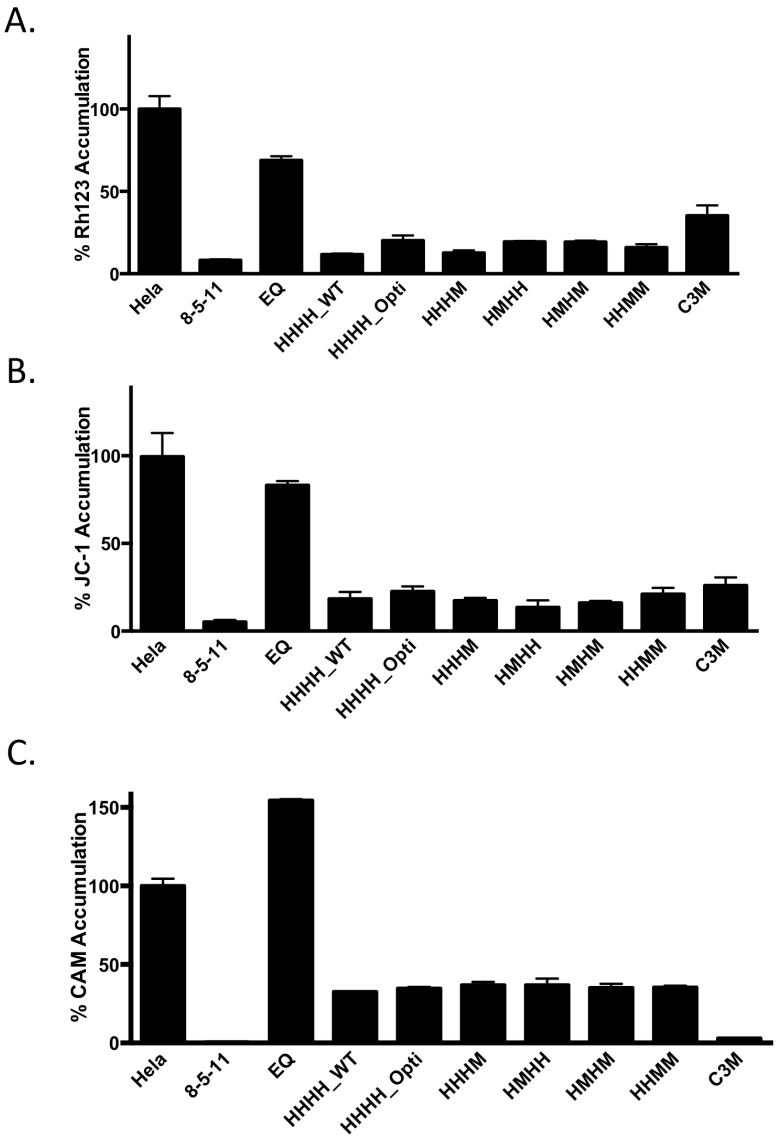


Figure 5.7 Efflux of the fluorescent P-gp substrates Rh123, JC-1 and CAM by chimeric P-gp. All chimeras were functional and able to efflux the fluorescent substrates **A)** Rh123, **B)** JC-1, and **C)** CAM, significantly as compared with non-transduced HeLa cells. All data was normalized to 100% accumulation in non-transduced HeLa cells. Data represent the mean fluorescence intensity and standard deviation of technical triplicates resulting from a single transduction event.

All chimeras were functional and able to efflux Rh123, CAM, and JC-1 compared with parental HeLa cells (Figure 5.7). There were no appreciable differences in the efflux of these compounds between chimeras and the HHHH_WT P-gp at steady state. Additionally, the codon optimization process did not influence fluorophore transport since there were no differences between HHHH_WT and HHHH_Opti efflux. Wild-type mouse and human P-gp in

the drug-selected cell lines KB-8-5-11 and C3M were also able to efflux all fluorophores efficiently. While it is tempting to draw conclusions regarding species differences between mouse and human P-gp from these data (for example, C3M cells appear less efficient at Rh123 efflux than KB-8-5-11 cells), species differences in fluorophore affinity cannot be deduced from these assays. This is due to variable P-gp expression levels between cell lines and, perhaps more importantly, differences in cellular backgrounds (the C3M cell line is derived from the mouse embryonic fibroblast cell line NIH/3T3 while the KB-8-5-11 cell line is derived from a sub-line of the human adenocarcinoma HeLa cell line) that may affect baseline fluorophore accumulation in cells.

5.2.8 Inhibition of transport function of chimeras with TQR, CsA, verapamil, and elacridar

P-gp inhibitors were used in flow cytometry efflux assays for two reasons. Firstly, P-gp contains multiple drug-binding sites within its central cavity, a fact that can be seen by the two distinct drug-binding sites of QZ59-RRR and QZ59-SSS inhibitors in the crystal structure of mouse P-gp (Figure 1.8) (103). Therefore, it was unknown if or how the chimerization process would affect the ability of inhibitors to interact with P-gp chimeras. Secondly, inhibitors can be used to confirm that decreased fluorophore accumulation is attributable to P-gp expression, as the presence of P-gp inhibitors during the course of the assay should fully inhibit fluorophore efflux and thus restore fluorescence intensities to levels of non-transduced HeLa cells (Figure 5.1.B). The P-gp inhibitors TQR, CsA, verapamil, and elacridar were utilized for their capacity to inhibit P-gp and allow for Rh123 accumulation within transduced cells (Figure 5.8).

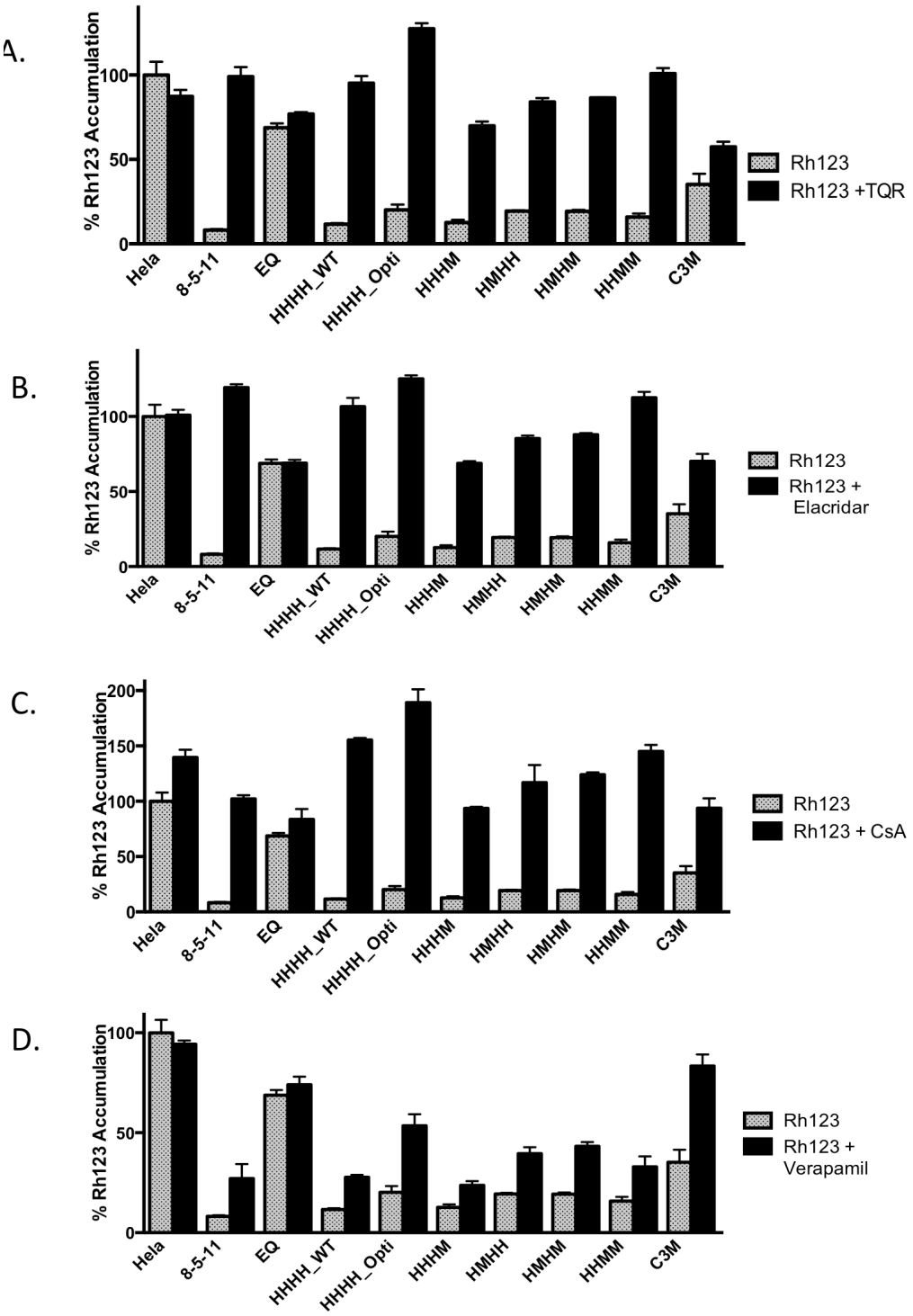


Figure 5.8 Inhibition of chimeras with TQR, elacridar, CsA, and verapamil. Rh123 efflux and inhibition with compounds **A)** TQR (100 nM), **B)** elacridar (100 nM), **C)** CsA, (10 µM) and **D)** verapamil (10 µM). All data were normalized to 100% accumulation in non-transduced HeLa cells. Data represent the mean fluorescence intensity and standard deviation of technical triplicates resulting from a single transduction event.

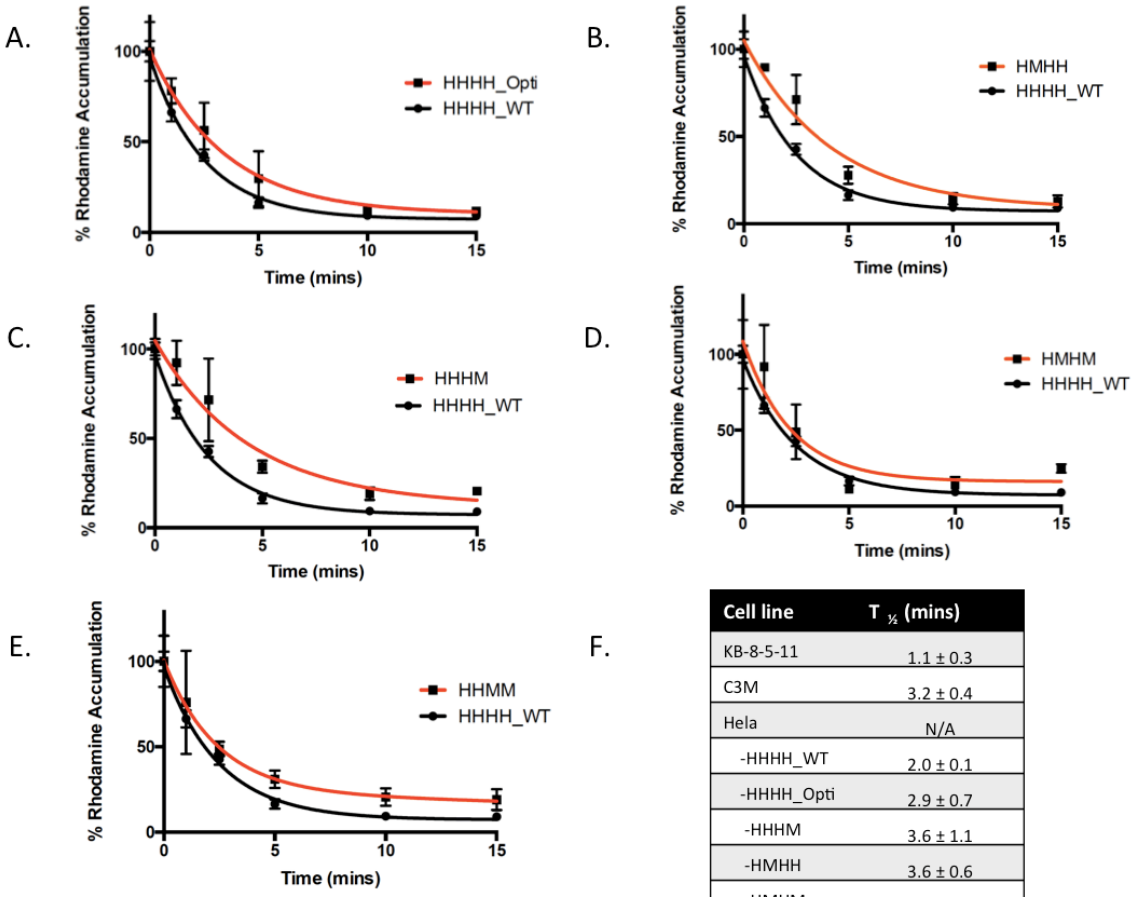
TQR, elacridar and CsA were all able to significantly increase Rh123 accumulation compared with uninhibited treatments for all chimeras. The presence of inhibitors did not significantly change Rh123 accumulation in cells transduced with the EQ mutant, non-transporting P-gp. At the concentration tested, verapamil was not able to fully inhibit chimeric P-gp or wild-type P-gp homologs to achieve levels of intracellular Rh123 equal to those achieved by other inhibitors. It can be appreciated that in some instances the presence of an inhibitor results in the intracellular accumulation of Rh123 above 100%. For example, various chimeras and P-gp homologs inhibited by CsA accumulate Rh123 to a greater level than non-transduced HeLa cells (Figure 5.8.C). This effect seems to be independent of P-gp expression, as even non-transduced Hela cells that do not express P-gp accumulate higher levels of intracellular Rh123 in the presence of CsA. This effect has been previously observed in this laboratory, but it is unclear why some inhibitors can increase levels of intracellular fluorescence. However, this effect may be attributed to off-target effects of the inhibitor at concentrations used in this assay that have yet to be elucidated.

5.2.9 Efflux rate of Rh123 by chimeras

Time course experiments were undertaken to determine if chimeras showed differences in the rate of Rh123 efflux (Figure 5.9). Briefly, cells were depleted of ATP by incubation in PBS containing 20 mM 2-deoxy-d-glucose (2-DG), a glycolysis inhibitor, and 5 mM sodium azide, a mitochondrial uncoupler, for 20 minutes (286). [4] After cellular energy depletion, cells were centrifuged, the PBS containing 2-DG and sodium azide was removed, and cells were incubated on ice in the presence of Rh123 for an additional 20 minutes. This process allowed Rh123 to "load" into cells while P-gp-mediated transport was inhibited by both low temperature and energy depletion. Transport of Rh123 by P-gp was initiated when

[4] 2-DG acts as a glycolytic inhibitor is due to its ability to compete with glucose for hexokinase phosphorylation (once phosphorylated 2-DG cannot be further metabolized) (287). Sodium azide acts as a mitochondrial inhibitor by binding to and inhibiting cytochrome c oxidase (complex IV) (288).

PBS containing 50 mM glucose was added to cells and the incubation temperature was adjusted to 37 °C. Cells were allowed to efflux Rh123 for 1, 2.5, 5, 10, 15, and 45 minute(s) before being washed with ice cold PBS and transferred to ice to stop P-gp transport activity prior to quantification via flow cytometry analysis.



F.

Cell line	$T_{½}$ (mins)
KB-8-5-11	1.1 ± 0.3
C3M	3.2 ± 0.4
Hela	N/A
-HHHH_WT	2.0 ± 0.1
-HHHH_Opti	2.9 ± 0.7
-HHHM	3.6 ± 1.1
-HMHH	3.6 ± 0.6
-HMHM	2.1 ± 1.1
-HHMM	2.3 ± 0.6

Figure 5.9 Time-course of Rh123 efflux by transduced HeLa cells. Rh123 efflux of HeLa cells transduced with BacMam constructs expressing **A)** HHHH_Opti, **B)** HMHH, **C)** HHHM, **D)** HMHM and **E)** HHMM as compared with wild-type P-gp (HHHH_WT). Each time point represents the average and standard deviation of technical triplicates from a single transduction event. All data was normalized to 100% accumulation of energy-depleted cells (zero minutes). **F)** $T_{1/2}$ values for all chimeras, and wild-type human and mouse P-gp from KB-8-5-11 and C3M drug selected cell lines.

There were no noteworthy differences in the rate of Rh123 efflux between codon-optimized (HHHH_Opti), chimeric (HHHM, HMHH, HMHM, HHMM), and wild-type (HHHH_WT) P-gp (Figure 5.9). All chimeras reached a steady state of efflux within 10 minutes. Rh123 efflux for up to 45 minutes was also measured. However, no changes in Rh123 efflux occurred after 15 minutes (data not shown). Non-transduced HeLa cells and HeLa cells expressing the EQ mutant P-gp were also assessed for Rh123 efflux and did not exhibit efflux activity (data not shown). Additionally, the time required for P-gp chimeras and homologs to efflux 50% of intracellular Rh123 ($T_{1/2}$) was calculated (Figure 5.9.F). Slight variations in $T_{1/2}$ were observed with the HHHM and HMHH chimeras having a 1.8-fold decrease in efflux rate compared to HHHH_WT P-gp. While this difference may indicate that the HMHH and HHHM are slightly less efficient at Rh123 efflux, differences in efflux are subtle at best.

5.2.10 Ability of chimeras to confer a paclitaxel-resistant phenotype to HeLa cells

In order to assess the ability of chimeras to induce a MDR phenotype, a suitable cytotoxicity assay was optimized. Previous attempts at using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) cytotoxicity assay protocol on transduced HeLa cells were not able to generate a full dose response curve (184). Failure was attributed to the inability of transduced cells to survive the 72 hour incubation required for a detectable colorimetric signal, at least in part due to BacMam-induced transduction toxicity that increases with time due to high levels of foreign protein expression and progression of viral infection (184). Therefore, an ideal cytotoxicity assay would produce a strong signal as a measure of toxicity and would permit a reduced incubation of transduced cells with cytotoxic drug. Additionally, in previous attempts to utilize MTT assays, cells were incubated in the presence of 10 mM NaBT, which has been reported toxic in the 5 to 10 mM range (277). Cytotoxicity assays showed approximately 35% of HeLa cells were viable after a 72

hour incubation of non-transduced HeLa cells with 10 mM NaBT (Figure 5.10). Therefore, cytotoxicity assays must be conducted in the absence of NaBT due to its confounding cytotoxic effect. As such, cytotoxicity assays conducted in this study rely on the lower, but detectable, expression of P-gp in cells transduced without the addition of NaBT (Figure 5.2).

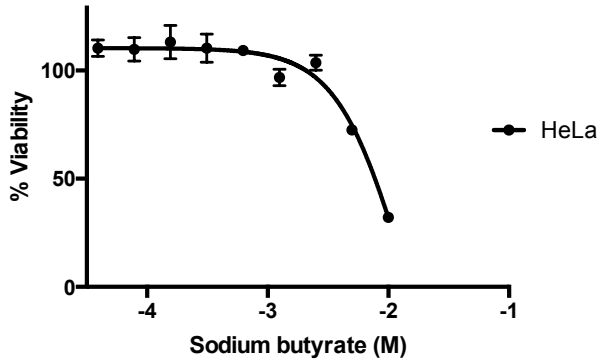


Figure 5.10 Sodium butyrate toxicity on non-transduced HeLa cells. CellTiter-Glo cytotoxicity assay of non-transduced HeLa cells after a 72 hour incubation with NaBT (Section 2.2.27). The highest NaBT concentration used was 10 mM, as this is the same concentration that is used during viral transduction. Cells were less than 40% viable after 72 hours when exposed to 10 mM NaBT. Data points represent the mean and standard deviation of three triplicate assays. A vehicle only control was not carried out as the final concentration of DMSO was less than 0.0025% for all conditions tested and thus below a cytotoxic threshold.

The CellTiter-Glo (CTG) assay system was tested to determine if cytotoxicity could be assessed after a 48 hour incubation of transduced HeLa cells with drug. The CTG assay system is based upon the catalysis of D-luciferin to oxyluciferin in the presence of ATP by the enzyme firefly luciferase, an enzymatic reaction that produces light. An advantage of the CTG assay system is that it produces a luminescent signal that is able to measure viable cells at numbers below the detection limits of the colorimetric MTT assay. HeLa cells were transduced with chimeric, codon-optimized, and wild-type P-gp without NaBT and incubated for 24 hours to allow for expression. After 24 hours, media containing BacMam virus was removed, and replaced with media containing the chemotherapeutic drug paclitaxel. Cells were incubated for an additional 48 hours before determination of cell viability. Paclitaxel

was chosen for this study because it is known to be an avid substrate of P-gp such that even cells expressing low levels of P-gp are resistant to its cytotoxic effects. Cytotoxicity was measured by the addition of the CTG reagent, which contains a lysis agent in addition to the reagents required to produce a luminescent signal (i.e. luciferase and luciferin). The amount of ATP liberated from lysed cells generates a luminescent signal that is proportional to cell viability.

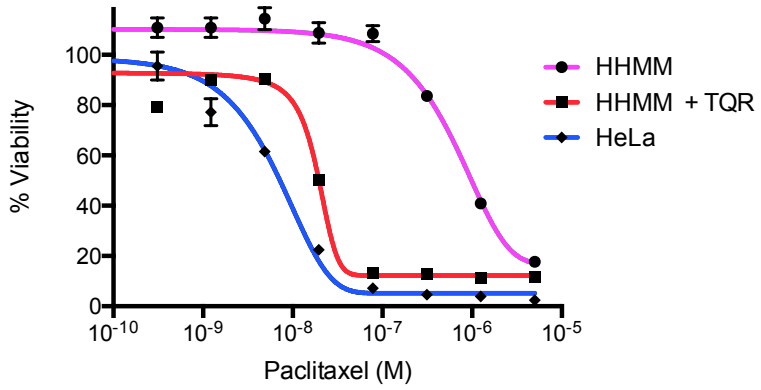


Figure 5.11 Transduction of HeLa cells with chimeric P-gp is able to confer a resistant phenotype. Cells transduced with the construct expressing the HHMM chimera (IC_{50} = 1049 ± 68 nM) showed an increased resistance as compared with non-transduced HeLa cells (IC_{50} = 9.2 ± 1.4 nM). This resistance was able to be reversed when cells were incubated with the P-gp inhibitor TQR (IC_{50} = 20.0 ± 0.6 nM), confirming that resistance was due to expression of the HHMM chimera. Data represents the mean and the standard deviation of three independent observations. Note that some deviation bars are obscured due to the consistency of data from triplicate measurements.

HeLa cells transduced with HHMM P-gp displayed a paclitaxel-resistant phenotype (Figure 5.11). HHMM-transduced cells were approximately 114-fold more resistant to paclitaxel than the parental HeLa cells. Resistance of HHMM-transduced cells can be attributed to expression of chimeric P-gp by abrogation of resistance when co-incubated with the P-gp inhibitor TQR during the course of the cytotoxicity assay. Cytotoxicity data for all chimeric and wild-type P-gp homologs are shown in Table 5.2. All chimeras were able to confer resistance to paclitaxel. The degree of P-gp-mediated resistance can be analyzed by

the ratio of IC_{50} values with and without the inhibitor TQR. This ratio is often referred to as a resistance ratio, calculated by the following equation:

$$\text{Resistance Ratio} = \frac{IC_{50\ (drug)}}{IC_{50\ (drug + inhibitor)}}$$

A resistance ratio of 1 indicates that there is no resistance attributable to P-gp expression, while a resistance ratio > 1 indicates resistance that is reversible upon P-gp inhibition (49).

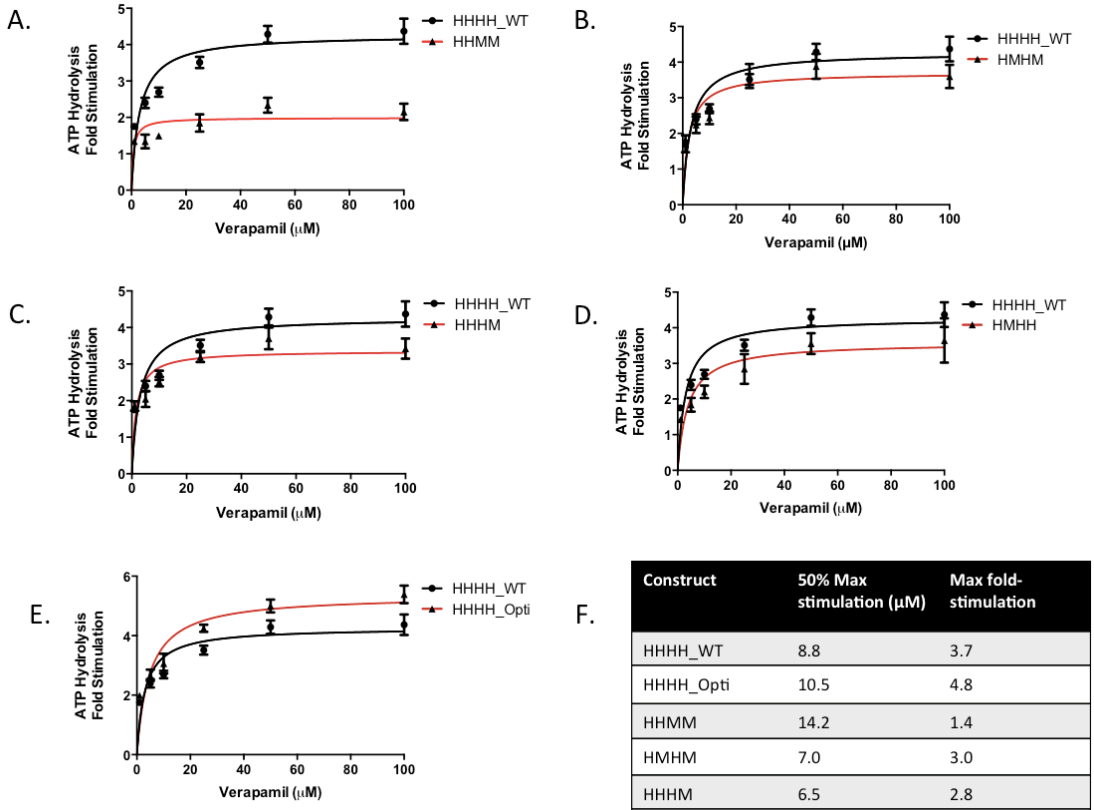
Table 5.2 CellTiter-Glo cytotoxicity assay of chimeric P-gp with the chemotherapeutic drug paclitaxel. The ability of chimeras to confer a paclitaxel-resistant phenotype can be assessed by the increased IC_{50} values of HeLa cells expressing chimeras as compared to parental HeLa cells. Increased IC_{50} values indicate that it takes a higher concentration of paclitaxel to achieve 50% cell death. Resistance to paclitaxel can be attributed to expression of chimeric P-gp, as resistance is reversed when cells are co-incubated with the P-gp inhibitor TQR (100 nM) during the course of the assay. Resistance ratios are shown to indicate the degree of P-gp-mediated resistance.

Cell Line	IC_{50} Paclitaxel (nM)	IC_{50} Paclitaxel +TQR (nM)	Resistance Ratio
KB-8-5-11	3718 ± 193	4.0 ± 0.1	929
Hela	9.2 ± 1.4	5.5 ± 1.7	1.67
-HHHH_WT	936 ± 81	19.4 ± 2.8	48.2
-HHHH_Opti	284 ± 19	11.9 ± 0.7	23.8
-HHMM	1049 ± 68	20.0 ± 0.6	52.5
-HMHH	421 ± 126	14.0 ± 1.9	30.1
-HHHM	773 ± 22	12.0 ± 1.0	64.4
-HMHM	221 ± 7	15.3 ± 1.2	14.4

5.2.11 ATPase Assays

ATPase assays were carried out to determine biochemical differences in the ATP binding and hydrolysis between chimeric, codon-optimized and wild-type human P-gp. Crude membranes were isolated from insect cells infected with baculovirus genetically

modified to express P-gp chimeras (Section 2.2.40). ATPase assays were completed with 1, 5, 10, 25, 50, and 100 µM verapamil (Section 2.2.26). Higher concentrations of verapamil inhibited P-gp and decreased ATPase stimulation and were therefore excluded. Dose-response curves were used to determine maximal fold stimulation and concentration required for 50% max stimulation.



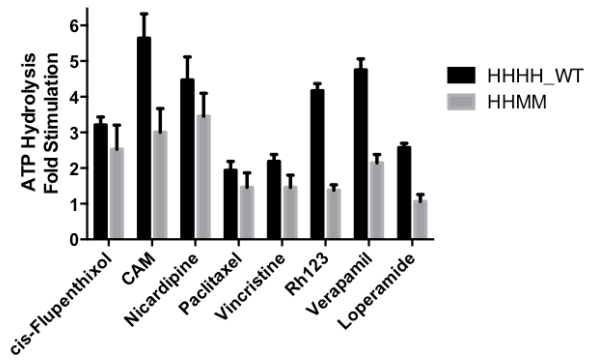
Construct	50% Max stimulation (µM)	Max fold-stimulation
HHHH_WT	8.8	3.7
HHHH_Opti	10.5	4.8
HHMM	14.2	1.4
HMHM	7.0	3.0
HHHM	6.5	2.8
HMHH	14.1	3.1

Figure 5.12 Stimulation of P-gp ATPase activity of chimeras by verapamil and kinetic parameters. Dose response ATPase assays of **A)** HHHH_Opti, **B)** HMHH, **C)** HHHM, **D)** HMHM and **E)** HHMM compared with wild-type human P-gp (HHHH_WT) in the presence of 0, 1, 5, 10, 25, 50 and 100 µM of verapamil (Section 2.2.26). ATPase activity is reported as the fold-change in the rate of ATP hydrolysis over basal levels. Data represent average and standard deviation of technical duplicates from three individual assays. **F)** Table showing values of maximal fold stimulation and 50% maximal fold stimulations derived from the dose-response verapamil curves. Values were determined by Michaelis–Menten analysis in GraphPad Prism.

A range of maximal fold stimulation values was observed, suggesting altered function among chimeras (Figure 5.12). HHHH_Opti showed an approximately 30% increase in maximal fold stimulation as compared to wild-type P-gp (Figure 5.12.F). In contrast, the HHMM and HHHM chimeras displayed a 60% and 25% decrease in maximal fold stimulation, respectively, compared with wild-type P-gp. The HMHM and HMHH chimeras also showed smaller differences in maximal fold stimulation compared to wild-type P-gp, having a 16% and 18% decrease for HMHH and HMHM chimeras, respectively. While differences in maximal fold stimulation were observed, values of 50% maximal fold stimulation were similar for all chimeras, ranging from 6.5 to 14 μM.

It is difficult to determine the reason variations in maximal fold stimulation would occur with chimeras, given other studies presented in this thesis do not show differences in transport activity among chimeric, codon-optimized, and wild-type human P-gp. Several explanations are possible to account for differences in maximal fold stimulation observed. Firstly, chimeras may be differentially stimulated by verapamil due to changes in the drug-binding pocket. Changes in the pocket may arise directly, from chimerization of the TMDs as seen in the HHMM chimera, or indirectly, by subtle conformational changes that may have arisen from either the chimerization or codon optimization process even when both TMDs are comprised of human P-gp. Secondly, all chimeras contain a mixture between human and mouse TMDs and NBDs, a factor that may affect the way NBDs and TMDs communicate to hydrolyze ATP and transport substrates. Thirdly, chimerization of the NBDs may affect the rate of ATP hydrolysis, a characteristic of chimeras that could be detected in this assay system but not in whole-cell based flow cytometry assays. Lastly, this effect could be specific to the substrate verapamil.

A.



B.

Compound	p-value
cis-Flupenthixol	n.s
CAM	p ≤ 0.0001
Nicardipine	p ≤ 0.001
Paclitaxel	n.s
Vincristine	p ≤ 0.05
Rh 123	p ≤ 0.0001
Verapamil	p ≤ 0.0001
Loperamide	p ≤ 0.0001

Figure 5.13 Effect of selected compounds on the ATPase activity of HHHH_WT versus HHMM P-gp. A) A panel of eight compounds that are known substrates of P-gp were selected to determine if there were differences in ATPase activity of HHHH_WT and HHMM P-gp. The following concentrations were used for each assay: 20 µM cis-flupenthixol, 10 µM CAM, 20 µM nicardipine, 20 µM paclitaxel, 20 µM vincristine, 500 µM Rh123, 25 µM verapamil, and 20 µM loperamide. **B)** Multiple groups were analyzed using one-way ANOVA where significance was defined as $p < 0.05$ in GraphPad Prism (n/s = not significant).

To determine if differences in ATPase activity were specific to verapamil, a panel of eight known P-gp ATPase stimulators were tested with HHHH_WT and HHMM P-gp. The HHMM chimera was chosen for this analysis, as it displayed the greatest difference in maximal fold stimulation compared to wild-type human P-gp. The compounds cis-flupenthixol, CAM, nicardipine, paclitaxel, vincristine, Rh123, and loperamide were used to probe for differential ATPase activity between HHHH_WT and HHMM P-gp (Figure 5.13). No significant differences in ATPase stimulation between HHHH_WT and HHMM P-gp were seen with cis-flupenthixol or paclitaxel. However, CAM, nicardipine, vincristine, Rh123, and loperamide were all able to stimulate ATPase activity to a greater extent in HHHH_WT P-gp than with the HHMM chimera (Figure 5.13.B). These data indicate that the differences seen previously were not specific to verapamil, rather the ATPase activity of the HHMM chimera is less stimulated by various P-gp stimulators compared with wild-type human P-gp.

5.3 Discussion

The studies described in this chapter investigated the function of chimeric P-gp compared with wild-type human P-gp. Numerous assays were employed to obtain a full functional characterization of chimeras, including flow cytometry, cytotoxicity, western blotting, and ATPase assays. Chimeras were expressed and present on the cell surface at equivalent levels (Figure 5.5), and were able to confer a paclitaxel-resistant phenotype characteristic of P-gp-mediated drug resistance (Table 5.2). No differences were detected in the ability of chimeras to efflux the fluorophores CAM, Rh123, and JC-1 as compared to wild-type P-gp (Figure 5.7). Additionally, the P-gp inhibitors CsA, TQR, verapamil and elacridar were able to reverse P-gp-mediated Rh123 efflux from cells equally in cells transduced with chimeric and wild-type P-gp (Figure 5.8).

One notable observation was the decreased ATPase activity of the HHMM chimera compared with wild-type P-gp ATPase activity. HHMM P-gp showed decreased ATPase activity for six out of eight compounds tested and a 60% decrease in maximal fold stimulation upon verapamil stimulation. While the HHMM chimera showed decreased P-gp ATPase activity to Rh123 and CAM, these substrates were also used as fluorophores in flow cytometry efflux assays where Rh123 and CAM were not differentially effluxed from cells. As HHMM P-gp contains a chimerized drug-binding pocket, it is feasible that the pocket in this chimera is altered and therefore stimulated less efficiently by various compounds. While differences in drug transport are observed in sensitive biochemical ATPase assays, whole cell-based assays such as flow cytometry and cytotoxicity assays did not indicate functional differences between chimeric and wild-type P-gp.

Findings that the HHMM chimera may contain an altered drug-binding pocket are supported by a parallel investigation completed by a collaborator (E. Chufan *et al.* unpublished results) using [^{125}I]iodoarylazidoprazosin (IAAP) labeling. [^{125}I]IAAP is a compound known to bind P-gp in the drug-binding pocket (100, 101, 289). [^{125}I]IAAP can also

be crosslinked to P-gp upon exposure to UV light. Studies have shown that [^{125}I]IAAP binding is displaced by various known P-gp modulators in a concentration-dependent manner (101). Collaborators have investigated the ability of the P-gp inhibitors CsA, TQR, and zosuquidar to displace [^{125}I]IAAP binding in human and mouse P-gp, and the HHHM, HMHH, HMHM, and HHMM chimeras generated in this thesis. Interestingly, [^{125}I]IAAP binding was displaced with each modulator tested and with all P-gp homologs and chimeras except the HHMM chimera. These data suggest that CsA, TQR, and zosuquidar may not interact with the HHMM chimera at the primary site. This is consistent with recent work showing that there are multiple binding sites for a given substrate in human P-gp (290). However, it is known from the functional assays presented in this chapter that both CsA and TQR do, in fact, inhibit the activity of P-gp to levels comparable to those of wild-type P-gp in whole-cell based assay (Figure 5.8). Taken together, these data suggest that the conformation of the drug-binding pocket of the HHMM chimera may be altered in a way that affects the interaction between P-gp and modulators. Nonetheless, these changes do not affect the ability of P-gp in whole-cell based assays to efflux various substrates at maximal rate (Figure 5.9) and to efficiently induce a drug-resistant phenotype (Figure 5.11 And Table 5.2).

Together, the studies presented in this chapter indicate that human P-gp can tolerate a variety of residue changes, such as those that directly impact the drug-binding pocket, and remain functional and able to interact with a variety of compounds. These findings are in line with various pharmacogenomic investigations into P-gp (5). Numerous single nucleotide polymorphisms (SNPs) have been reported in P-gp, many occurring as part of a haplotype that varies between ethnic backgrounds (155). While SNPs may contribute to subtly altered protein function, there are currently no known individuals harboring non-functional P-gp, indicating a variety of SNPs can be tolerated and result in functional P-gp.

5.4 Conclusions

Functional characterization of chimeras was carried out to determine their physiological relevance as a prerequisite to structural investigations. Chimeras and wild-type P-gp were indistinguishable in whole-cell based assays. Biochemical assays, such as ATPase and photoaffinity labeling assays, indicate differences in substrate binding and transport kinetics. However, these differences do not affect the ability of chimeric proteins to efflux substrates or confer a MDR phenotype to HeLa cells. Overall, all chimeras constructed in this chapter were deemed functionally similar to wild-type P-gp and suitable for structural studies.

Chapter 6: Expression, purification, and crystallization of P-glycoprotein chimeras

6.1 Introduction

Crystallization requires the generation of large quantities of pure protein. Protein expression and subsequent purification of membrane proteins are notoriously time-consuming, costly, and laborious pre-crystallographic processes (Section 1.4). Heterologous protein expression techniques often rely on trial-and-error strategies to identify protocols that yield high-level expression of the foreign protein. Additionally, optimization is typically required on a protein-by-protein basis to determine favorable protein expression conditions. Subtle changes in the amino acid sequence of the protein to be expressed can greatly affect the subsequent level of expression. Therefore, re-optimization is often required for the expression of mutants or homologs of the same protein. After high-level protein expression has been accomplished, purification of a membrane protein from its native environment must be achieved. Optimization of purification processes is crucial to prevent inactivation, unfolding, and/or aggregation of the protein in a detergent-solubilized state (Section 1.4.6). Immobilized metal affinity chromatography (IMAC) is typically employed for the purification of the desired protein (191). Size exclusion chromatography (SEC) can also be employed as a second step to isolate monomeric protein for structural studies (291).

The reported crystal structures of mouse P-gp were obtained from protein that was expressed in and purified from *P. pastoris* (57, 103, 136). *P. pastoris* is methylotrophic yeast able to integrate vector DNA into the host genome by homologous recombination, typically under the control of the strong *AOX1* promoter (Section 1.4.2). Expression of foreign proteins in *P. pastoris* may be advantageous for several reasons, including the low cost of growth media, the limited equipment required for *P. pastoris* culture, and the ability to attain large-scale production in fermenters (150, 151). However, experience in the Laboratory of Cell Biology expressing protein in *P. pastoris* has proven labor-intensive, with a

high degree of optimization required to determine the best conditions for protein expression.

Heterologous protein expression in insect cells infected with recombinant baculovirus is another method commonly used to express P-gp (Section 1.4.4) (148, 171, 271, 292). Typically, insect cell expression is used for the generation of P-gp-containing crude membranes for use in ATPase assays (Section 2.2.26 and Section 5.2.1) (219, 271). During the course of this thesis project, the Laboratory of Cell Biology has been able to reproduce the crystallization of mouse P-gp from protein expressed in and purified from insect cells (Shukla *et al.*, unpublished results). Additionally, high-level expression of a human P-gp mutant was also achieved using the same insect cell expression protocol employed to express mouse P-gp without further optimization efforts (Shukla *et al.*, unpublished results). Therefore, baculovirus-mediated expression in insect cells is an attractive prospect, as initial observations indicate high-level protein expression of P-gp homologs and mutants may be achieved with a common protocol. However, large-scale protein expression in insect cells may be more costly than expression in *P. pastoris*, as this laboratory is not equipped for insect cell culturing, thus protein expression must be commercially sourced.

The experimental stages of protein expression, purification, and crystallization in *P. pastoris* and insect cells are outlined in Figure 6.1. To express protein in *P. pastoris* (Figure 6.1.A), chimeric cDNA is first sub-cloned into a vector amenable for homologous recombination in yeast and used to transform various *P. pastoris* host strains. Single colonies of *P. pastoris* are then screened for chimera expression in small-scale (15 mL) cultures following the addition of methanol to induce protein expression. Expression of the protein chimera is assessed by western blotting and additional optimization carried out. Large-scale (8 L) protein expression and optimization of expression conditions is then performed and recombinants screened for chimeric expression by western blotting.

The expression of protein chimeras in insect cells is shown in Figure 6.1.B. The cDNA expressing chimeric proteins is first sub-cloned into an appropriate vector for the generation of recombinant baculovirus. A small-scale (250 mL) culture of insect cells is infected with recombinant baculovirus and screened for expression of chimeric protein by western blotting and once confirmed, large-scale (4 L) culturing and infection of insect cells is carried out, and cells screened for expression of protein chimeras.

Subsequent procedures are similar for both insect cell and *P. pastoris* expression systems; crude membranes are isolated and membrane proteins solubilized with a solution containing the detergent n-Dodecyl β-D-maltoside (DDM) (Section 1.4.6). Chimeric P-gp is then purified by IMAC through the binding of chimeric His_6 tagged protein to Ni-NTA resin (all chimeras contain a His_6 tag, Section 4.2.1.3). SEC is completed with the use of a Superdex 200 fast protein liquid chromatography (FPLC) column to separate proteins by size and purified monomeric protein was subsequently concentrated to 5-10 mg/mL using centrifugal protein concentrators with a 100 kDa cutoff and screened for crystallization.

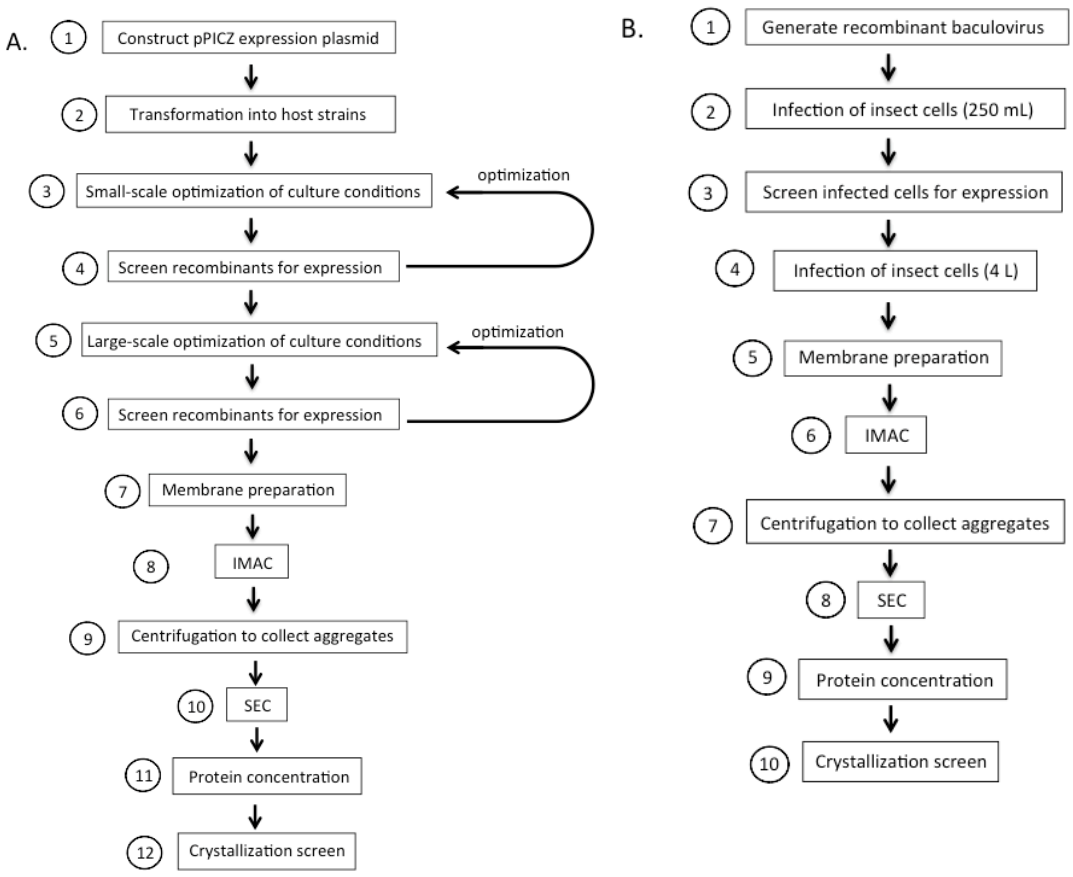


Figure 6.1 Flowchart of heterologous protein expression of P-gp chimeras in *P. pastoris* and insect cells. The salient steps of expression, purification, and crystallization in **A)** *P. pastoris* and **B)** insect cells are outlined.

This studies described in this chapter used both *P. pastoris* and insect cells for the expression of chimeras. This chapter is divided into two sections: protein expression of HHMM P-gp in *P. pastoris* (Section 6.2.1) and protein expression, purification, and crystallization screening of chimeric P-gp expressed in insect cells (Section 6.2.2). An extensive small-scale optimization process was carried out to express the HHMM chimera in *P. pastoris*. While various strategies were used to attain large-scale, high-level expression of the HHMM chimera, this study failed to identify conditions that expressed chimeric P-gp in large-scale culturing. In contrast, HHMM, HMHM, HHHM, and HMHH chimeric P-gp were expressed in insect cells and high-level expression of all chimeras was achieved. An analysis

of the purification of chimeric P-gp from insect cells is presented. An initial crystallization trial of chimeric P-gp using the MemGold and MembFrac crystallization screening conditions was completed, and protein crystals from the HHMM and HHHM crystallization screen were obtained.

6.2 Results

6.2.1 Expression of HHMM P-gp in *P. pastoris*

6.2.1.1 Generation pPICZ_HHMM and transformation into *P. pastoris*

A yeast expression vector containing the HHMM chimera was generated by Gateway cloning into an Invitrogen pPICZ expression plasmid, generating the pPICZ_HHMM plasmid (Figure 6.2) (Section 2.2.12) (293). Transformation of the pPICZ_HHMM plasmid into four different *P. pastoris* host strains was carried out to probe for strain-specific differences in expression of chimeric P-gp. The *P. pastoris* host strain used can influence the level of expression of a foreign protein due to differences in the host genotype. Therefore, various strains of *P. pastoris* are commercially available to screen for expression of the protein of interest. The GS115, X-33, KM71H, and SMD168H host strains were made competent (Section 2.2.36) and transformed by electroporation (Section 2.2.37). Host strain *P. pastoris* genotypes are listed and described in Table 6.1.

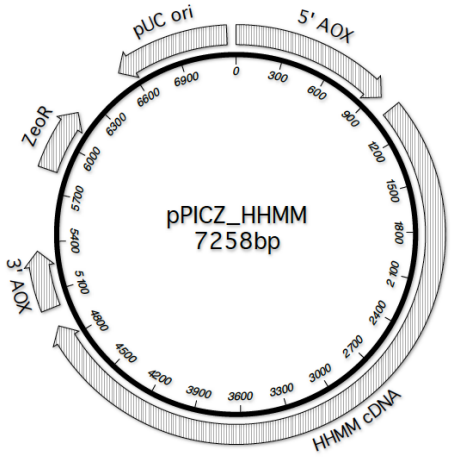


Figure 6.2 pPICZ_HHMM plasmid. HHMM chimeric cDNA is flanked by the 5' *AOX1* promoter and 3' *AOX1* terminal sequences. The 942 bp 5' *AOX1* promoter is upstream of the HHMM cDNA; a feature that allows for homologous recombination and subsequent expression of the chimera upon induction with methanol (Section 1.4.2). Additionally, the 260 bp 3' *AOX1* sequence downstream of the HHMM chimera allows for transcriptional termination and polyadenylation that permits efficient 3' mRNA processing and increased mRNA stability. The plasmid contains a zeocin resistance gene (ZeoR) and a bacterial origin of replication (pUC Ori).

Table 6.1 Genotype of *P. pastoris* host strains. The GS115 strain of yeast has been genetically modified to be deficient in histidinol dehydrogenase (*his4*), a gene required for histidine synthesis (294). Therefore, only vectors that introduce *his4* into the strain will be able to synthesize histidine (in this study, *his4* was not used as a selection marker, and thus histidine was introduced into the media for survival of transformants). KM71H is a yeast strain where the *AOX1* gene is deleted and replaced with the *S. cerevisiae arg4* gene. This creates a slow methanol utilization (Mut^S) phenotype (295). SMD1168 is deficient in proteinase A, a product of the *pep4* gene, and thus may produce favorable conditions for foreign proteins that are rapidly degraded (296). The strain X-33 is considered wild-type, as it contains wild-type *AOX1* and *AOX2* genes, and is not a histidine- or protease-deficient mutant (295).

Strain	Genotype
X-33	wild-type
GS115	*his4*
KM71H	*aox1::ARG4, arg4*
SMD1168	*pep4*

Transformed *P. pastoris* was plated onto YPDS agar plates supplemented with four different concentrations of zeocin: 100 µg/mL (1x), 500 µg/mL (5x), 1000 µg/mL (10x), and 2000 µg/mL (20x) (Section 2.2.34). *P. pastoris* is capable of incorporating multiple copies of transformed DNA into the genome through homologous recombination; an occurrence that can be exploited for higher levels of protein expression than could be achieved by a single insertion event. Therefore, transformed *P. pastoris* was streaked on YPDS agar plates containing increasing concentrations of zeocin to screen for colonies that may harbor multiple integration events and thus increased zeocin resistance. Overall, four host strains were transformed and plated onto YPDS agar containing four zeocin concentrations, generating 16 unique conditions to screen for HHMM expression.

All host strains transformed with the pPICZ_HHMM plasmid yielded zeocin-resistant colonies that grew on YPDS agar plates containing up to 20x concentrations of zeocin, potentially indicating multiple homologous recombination events. While the acquisition of zeocin resistance implies the integration of chimeric cDNA into the host genome and the

ability of *P. pastoris* to express chimeric protein, *P. pastoris* must be induced with methanol to confirm expression. Four transformed host strains were chosen for small-scale growth and methanol induction: X-33 with 1000 µg/mL zeocin resistance, GS115 with 500 µg/mL zeocin resistance, and KM71H with 1000 and 2000 µg/mL zeocin resistance. These clones were chosen as they represent a variety of host strains and zeocin resistance levels. Transformants from SMD1168 were not selected for further analysis based on past experience that this host strain does not typically express wild-type mouse P-gp at levels adequate for crystallization. Additionally, previous efforts in the Laboratory of Cell Biology have screened multiple transformants from a single transformation event, and have characterized significant clone-to-clone variation of foreign protein expression levels upon methanol induction. To address potential clone-to-clone variability, two clones from each transformation were selected for induction and screening for HHMM expression. The name, host strain, concentration of zeocin, and clone number of each transformant selected for small-scale expression and induction is listed in Table 6.2.

Table 6.2 Transformants selected for small-scale induction of protein expression. Transformants were named based on their *P. pastoris* host strain of origin, the concentration of zeocin supplemented into media (ZeoR), and clone number.

Name	Strain	ZeoR	Clone #
G/500/1	GS115	500 µg/mL	1
G/500/2	GS115	500 µg/mL	2
X/1000/1	X-33	1000 µg/mL	1
X/1000/2	X-33	1000 µg/mL	2
K/1000/1	KM71H	1000 µg/mL	1
K/1000/2	KM71H	1000 µg/mL	2
K/2000/1	KM71H	2000 µg/mL	1
K/2000/2	KM71H	2000 µg/mL	2

6.2.1.2 Small-scale optimization of expression conditions in *P. pastoris*

To avoid time-consuming and costly optimization of expression conditions at a scale that is required to generate protein quantities necessary for crystallization, an extensive small-scale (15 mL) optimization of expression conditions was carried out. Three growth conditions were tested for each of the eight clones selected for analysis generating a total of 24 conditions to be tested (Table 6.3). These conditions were selected based on expert advice regarding conditions that have been used to express various P-gp homologs and mutants. For example, Condition 1 is used in the Laboratory of Cell Biology for the expression of wild-type mouse P-gp in *P. pastoris*. The ability to translate the favorable expression conditions of mouse P-gp to the expression of chimeric P-gp would be advantageous, as it would bypass the lengthy optimization process typically employed when first expressing a new protein. Transformants were grown to the OD (Section 2.2.35) required for each condition and then induced with the specified amount of methanol; 1 mL samples of *P. pastoris* culture were collected every 24 hours for 5 days and stored at -80°C for analysis by C219 western blotting. In total, 120 samples were collected and analyzed (eight clones tested in three media conditions for 5 days = 120 samples).

Table 6.3 *P. pastoris* small-scale expression conditions. Three conditions were tested to optimize HHMM chimeric protein expression in *P. pastoris*. Information regarding the recipes for YPM, MMH, and YPGM medium can be found in Section 2.2.34 and Table 2.13. OD was determined as described in Section 2.2.35.

	Condition 1	Condition 2	Condition 3
Media	YPM	MMH	YPGM
Methanol	0.5%	0.5%	0.2%
OD	25	1	25
Temp	30°C	30°C	20°C

6.2.1.3 Western blotting of small-scale HHMM chimeric expression

Western blotting analysis was used to determine the growth conditions that expressed the HHMM chimera (Figures 6.3, 6.4, 6.5, and 6.6). *P. pastoris* cell lysates were generated by breaking yeast with glass beads (Section 2.2.39). The C219 antibody was previously shown to recognize the HHMM chimera (Section 5.2.5) and thus was suitable for this study.

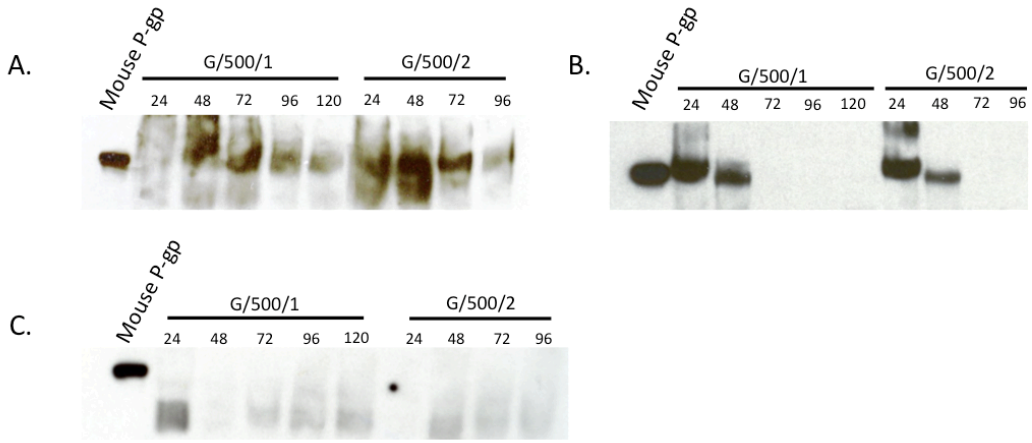


Figure 6.3 Western blot of transformed G/500/1 and G/500/2. A) Condition 1, **B)** Condition 2, and **C)** Condition 3 were screened for expression of the HHMM chimera in G/500/1 and G/500/2 transformants upon induction with methanol. Samples of *P. pastoris* were collected every 24 hours after induction, as indicated above the image (24 = 24hrs after induction, etc). Due to lane limitations of the polyacrylamide gel, only four time points were examined for HHMM expression of the G/500/2 clone. Purified mouse P-gp (6.5 ng) was used as both a size marker and positive control.

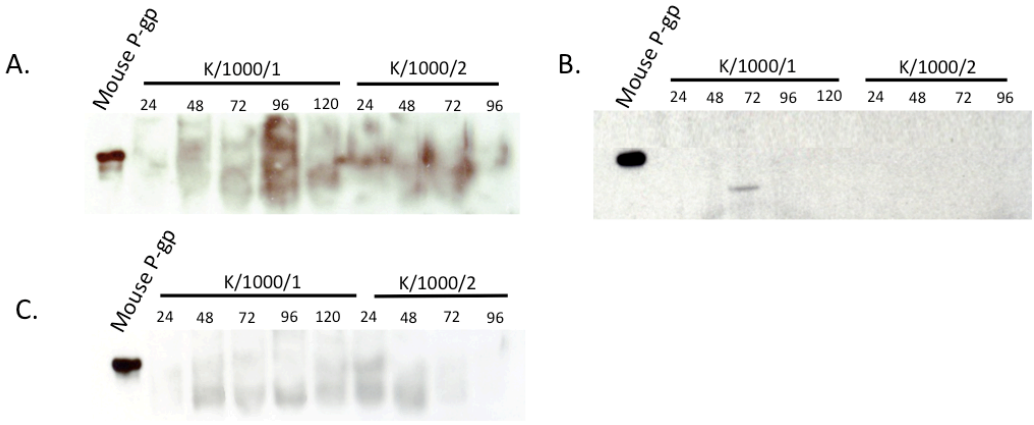


Figure 6.4 Western blot of transformed K/1000/1 and K/1000/2. A) Condition 1, **B)** Condition 2, and **C)** Condition 3 were screened for expression of the HHMM chimera in K/1000/1 and K/1000/2 transformants upon induction with methanol. Samples of *P. pastoris* were collected every 24 hours after induction, as indicated above the image (24 = 24hrs after induction, etc). Due to lane limitations of the polyacrylamide gel, only four time points were examined for HHMM expression of the K/1000/2 clone. Purified mouse P-gp (6.5 ng) was used as both a size marker and positive control.

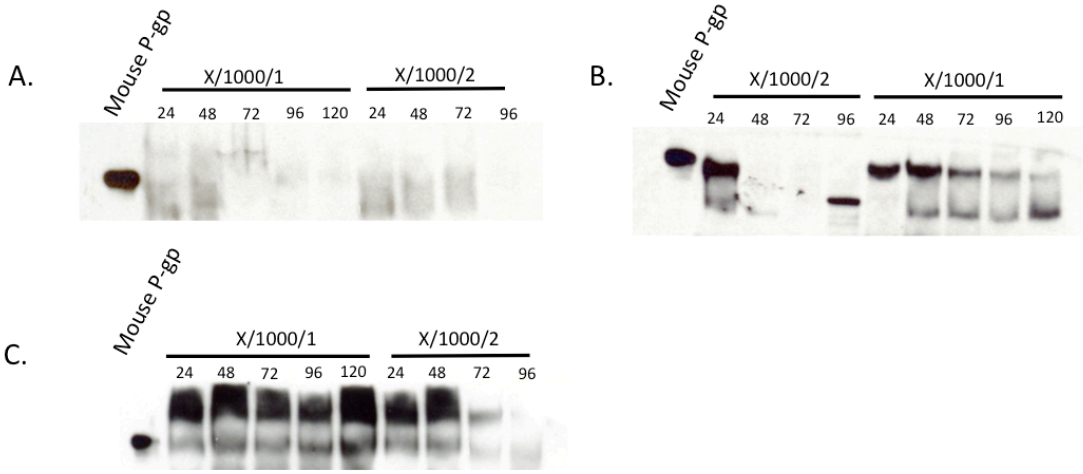


Figure 6.5 Western blot of transformed X/1000/1 and X/1000/2. A) Condition 1, **B)** Condition 2, and **C)** Condition 3 were screened for expression of the HHMM chimera in X/1000/1 and X/1000/2 transformants upon induction with methanol. Samples of *P. pastoris* were collected every 24 hours after induction, as indicated above the image (24 = 24hrs after induction, etc). Due to lane limitations of the polyacrylamide gel, only four time points were examined for HHMM expression of the X/1000/2 clone. Purified mouse P-gp (6.5 ng) was used as both a size marker and positive control.

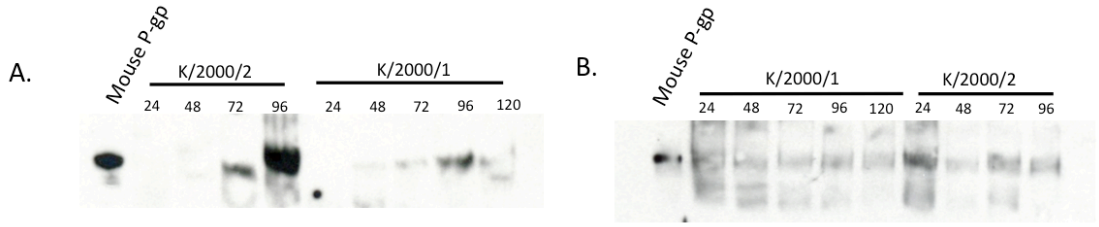


Figure 6.6 Western blot of transformed K/2000/1 and K/2000/2. A) Condition 1 and **B)** Condition 3 were screened for expression of the HHMM chimera in K/2000/1 and K/2000/2 transformants upon induction, with methanol. Samples of *P. pastoris* were collected every 24 hours after induction as indicated above the image (24 = 24hrs after induction, etc). Due to lane limitations of the polyacrylamide gel, only four time points were examined for HHMM expression of the K/2000/2 clone. Purified mouse P-gp (6.5 ng) was used as both a size marker and positive control.

Western blotting of the G/500, K/2000, X/1000 and K/1000 transformants showed differential expression of HHMM P-gp that was highly variable under different growth conditions. Neither K/1000/1 or K/1000/2 showed HHMM P-gp expression under any growth conditions tested (Figure 6.4). Both X/1000 clones showed HHMM expression under Conditions 2 and 3, but not Condition 1 (Figure 6.5). Clone-to-clone variation was also observed between X/1000/1 and X/1000/2 under Condition 2 (Figure 6.5). For example, in Condition 2, X/1000/2 showed high HHMM expression after 24 hours that became undetectable at 48 hours while X/1000/1 expressed protein for 72 hours after induction. Dramatic differences in protein expression due to growth conditions were also observed in G/500/1 and G/500/2 transformants (Figure 6.3). Here, Condition 3 did not lead to the expression of the chimera at detectable levels, while under both Conditions 1 and 2 the protein was expressed to variable degrees. Lastly, K/2000/2 yielded high levels of protein after 96 hours when grown under Condition 1, which was not observed with K/2000/1 (Figure 6.6). Both K/2000/1 and K/2000/2 expressed low levels of HHMM P-gp for all periods of time under Condition 3 (Figure 6.6). Overall, this small-scale optimization screen confirms how differences in various media compositions, growth conditions, and clone-to-clone

variation can dramatically affect protein expression in *P. pastoris*. These factors indicate that an extensive small-scale screening including the screening of multiple clones in various growth conditions was appropriate to determine optimal chimeric P-gp expression conditions.

6.2.1.4 Large-scale expression of HHMM P-gp in *P. pastoris*

Based on the results of the small-scale optimization of protein expression conditions for the HHMM chimera (Section 6.2.1.3), two transformants were selected for large-scale optimization efforts: G/500/1 and K/2000/2. Large-scale expression (8 L) was conducted with the G/500/1 and K/2000/2 transformants grown under Condition 2 and Condition 1, respectively. Large-scale cultures were induced with methanol, and samples were collected approximately every 6 hours for 2 days after induction.

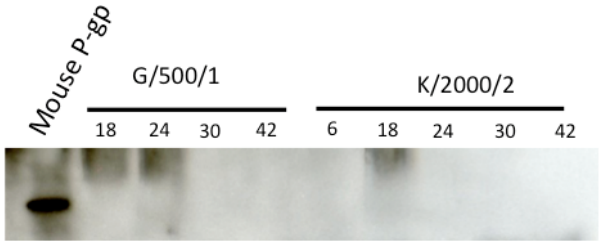


Figure 6.7 Western blot of transformed G/500/1 and K/2000/2 after large-scale expression. The G/500/1 and K/2000/2 transformants were induced for protein expression at a large scale. Samples of transformed *P. pastoris* were collected approximately every 6 hours after induction with methanol. Cell lysates were generated (Section 2.2.39) and C219 western blotting was completed (Section 2.2.20). Purified mouse P-gp was used as a size marker and positive control.

As previously described, western blotting with the C219 antibody was carried out to determine the expression of chimeric P-gp (Figure 6.7). While small-scale expression with G/500/1 and K/2000/2 transformants expressed HHMM P-gp, this was not observed at a large scale. It is not clear why small-scale expression conditions could not be translated to a

large scale. However, this effect has been observed with other P-gp homologs and mutants previously expressed in this laboratory. One possible explanation is that certain variables affecting protein expression were not controlled for. These variables may include differential levels of aeration, age of *P. pastoris* culture, etc. When large-scale expression does not occur, a decision is typically made to either 1) optimize large-scale expression conditions, or 2) select another clone that exhibited protein expression at a small-scale and attempt large-scale expression. At the time this work was completed, heterologous expression of chimeras in insect cells was also being conducted in parallel. As an insect cell expression system was able to generate quantities of chimeric protein suitable for crystallization, efforts to optimize protein expression in *P. pastoris* were discontinued.

6.2.2 Protein expression, purification, and crystallization screening from chimeric P-gp expressed in insect cells

6.2.2.1 Expression of P-gp chimeras in insect cells

All recombinant baculovirus production and insect cell expression of chimeric P-gp was completed by the Protein Expression Laboratory at the Frederick National Laboratory for Cancer Research (FNLCR). The pTOPO_HHMM, pTOPO_HMHM, pTOPO_HMHH, and pTOPO_HHHM plasmids generated in this study (Section 4.2.1.5 and Section 4.2.2.4) were sub-cloned into the pDest-008 shuttle vector by Gateway cloning (Section 2.2.12). Recombinant vectors were transformed into DH10Bac cells where site-specific transposition of the gene of interest into a baculovirus shuttle vector (bacmid) occurred (Section 2.2.30). Recombinant baculovirus was generated for the wild-type (HHHH_WT) and codon-optimized (HHHH_Opti) P-gp, and for HHMM, HHHM, HMHH, and HMHM chimeric P-gp. Baculovirus-infected Hi-Five insect cells were produced in either 250 mL (small-scale) or 4 L (large-scale) cell culture volumes (Section 2.2.31). Unlike heterologous expression in *P. pastoris*, protein expression conditions were not optimized on a chimera-to-chimera basis, but rather a general protocol that was previously optimized for the expression of human P-gp in insect cells was used (219). Insect cells infected with wild-type, codon-optimized, or chimeric P-gp was harvested and stored at -80 °C. All subsequent steps were completed by the author in the Laboratory of Cell Biology at the NIH.

6.2.2.2 Analysis of small- and large-scale expression of chimeras in insect cells

It was necessary to assess if wild-type human, codon-optimized, and chimeric P-gp were expressed in insect cells at a level that would be suitable for protein purification and subsequent crystallization screens. This was additionally important given a previously optimized protocol for the expression of human P-gp in insect cells (219) was used and no further optimization was carried out for the expression of P-gp chimeras. Crude membranes

were generated from a small-scale culture of insect cells infected with recombinant baculovirus containing the HHMM, HMHM, HMHH, and HHHM chimeras (Section 2.2.40). Colloidal blue staining of membranes showed that small-scale cultures expressed P-gp at a level suitable for scale-up (data not shown). Large-scale culturing of insect cells and infection with baculovirus was completed, and insect cell crude membranes were isolated. Colloidal blue staining and C219 western blotting of insect cell crude membranes were carried out to probe for overexpression of chimeric P-gp (Figure 6.8).

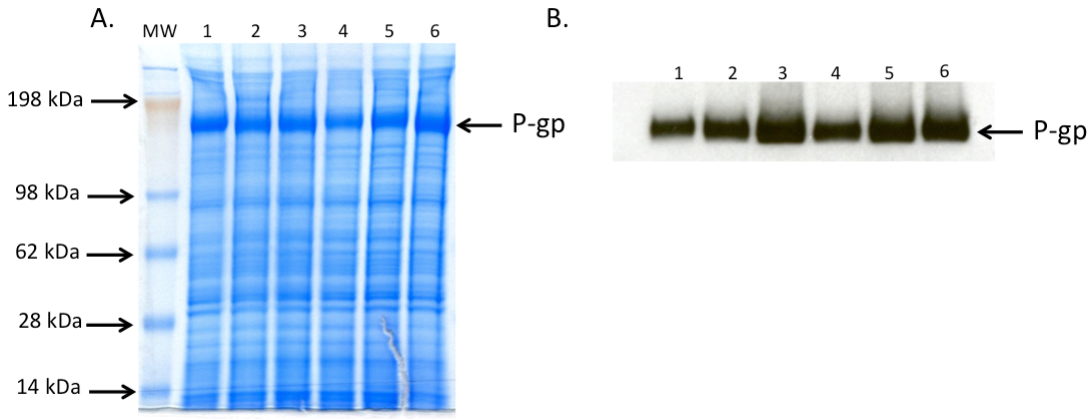


Figure 6.8 Colloidal blue staining and western blot of insect cell membranes. A) Colloidal blue stain of crude membranes isolated from infected insect cells (30 µg protein/lane) (Section 2.2.21). A band of over-expressed P-gp, indicated by the arrow, was seen in all P-gp containing baculovirus infections. **B)** C219 western blotting (Section 2.2.20) was completed to confirm that the overexpressed band seen in the colloidal blue stain was chimeric P-gp (1 µg protein/lane). Lanes for both the blue stain and western blot are indicated: lane 1) HHHH_WT, 2) HHHH_Opti, 3) HHHM, 4) HMHH, 5) HMHM, and 6) HHMM. A pre-stained molecular weight standard was also included for colloidal blue staining (MW, left lane).

Colloidal blue staining showed a large band of P-gp, indicating all chimeric and codon-optimized P-gp had been highly expressed in insect cell membranes (Figure 6.8.A). These results were confirmed by western blotting (Figure 6.8.B). Note that a loading control for the C219 western blotting was not included, as most commonly used loading control

proteins (GADPH, actin, etc.) are cytosolic and thus not present in crude membranes. Blotting of mammalian Na,K-ATPase, a membrane protein that is often used as a loading control for mammalian membranes, was attempted. However, Na,K-ATPase could not be detected in insect cell crude membranes, possibly due to the lack of crossreactivity because of species differences. The level of expression of chimeras can be compared with the wild-type P-gp. Crude membranes expressed an approximately equivalent amount of chimeric protein compared to human wild-type P-gp. Thus, it was concluded that no optimization of the baculovirus-mediated protein expression protocol used was necessary for the expression of chimeras in insect cells.

6.2.2.3 Purification of chimeras from insect cell membranes: overview

Similar to methods of heterologous protein expression, protein purification may also require a high degree of optimization to obtain high-quality, homogenous protein that has retained structural integrity throughout the purification process. While membrane proteins are stable in the plasma membrane environment, they often lose function, structure, and/or aggregate in a detergent-solubilized state. Thus, optimization of both the type and concentration of detergent used during purification is a critical step. Due to both the publication of the purification methods used to obtain protein for the reported mouse P-gp structures and expertise in the Laboratory of Cell Biology at purifying various P-gp homologs and mutants, a previously optimized protocol was used for the purification of chimeric P-gp (Section 2.2.41). A stepwise overview of the purification protocol used is shown in Figure 6.9.

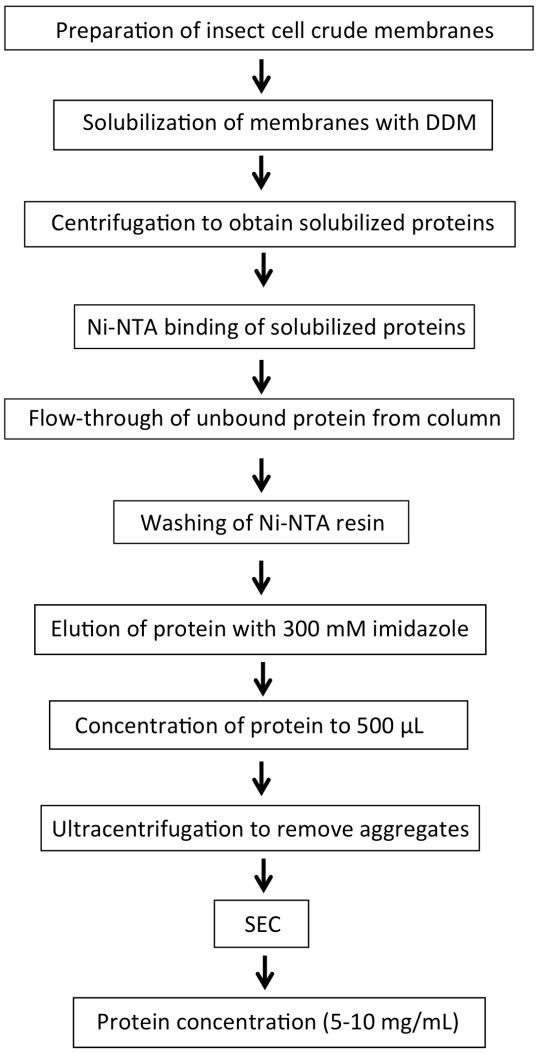


Figure 6.9 Flowchart of protein purification protocol. Steps of the purification process used to obtain purified, monomeric chimeric P-gp are outlined. Crude membranes were generated from insect cells infected with recombinant baculovirus. Membranes were solubilized with DDM, and subjected to ultracentrifugation to separate solubilized membrane proteins from membrane debris. Supernatant containing solubilized protein was mixed with Ni-NTA resin for binding of chimeric His_6 tagged proteins to resin. The protein-resin mixture was transferred into a column to allow for the elution of unbound proteins. After unbound proteins were eluted from the column, wash buffer was added to the column until no protein was detected in the flow-through. An elution buffer containing 300 mM imidazole was then used to elute chimeric proteins from the Ni-NTA resin. SEC required a sample of less than 500 µL in volume. Therefore, eluted protein was concentrated with the use of a 100 kDa protein concentrator to a volume suitable for SEC. Ultracentrifugation was carried out to remove protein aggregates and SEC was carried out to isolate monomeric P-gp chimeras. Fractions containing monomeric protein were collected and concentrated to 5-10 mg/mL protein for crystallization screening.

6.2.2.4 Purification analysis of HHMM, HMHM, HMHH, and HHHM chimeras

All chimeras were purified as outlined in Figure 6.9 and described in Section 2.2.41. One challenge concerning membrane protein purification is the retention of the protein in its native conformation in the absence of the plasma membrane (186). An important factor for the preservation of a membrane protein in its physiologic conformation is the choice and concentration of detergent used during purification. An "ideal" detergent for membrane protein purification would be one that most closely mimics the plasma membrane. However, no single universal detergent exists for the extraction and stabilization of membrane proteins, and often a series of detergents and detergent concentrations must be screened to increase the protein stability during purification. As a starting point for purification trials, this chapter used a protocol previously optimized in the Laboratory of Cell Biology for the purification of P-gp homologs and mutants from *P. pastoris.* However, this protocol might not be optimal for the purification of either 1) chimeric P-gp, and/or 2) P-gp expressed in insect cells.

Analysis of various samples collected during the purification process was conducted to gain insight into the efficiency of the purification and to assess the need for further optimization of the purification protocol. Samples collected should reflect the corresponding stage in purification from which the sample was obtained. For example, samples taken during the solubilization of membranes in DDM (Sample 1), and after centrifugation to obtain detergent-solubilized membrane proteins (Sample 2) should reflect a heterogeneous population of membrane proteins upon colloidal blue staining. After detergent-solubilized proteins are incubated with Ni-NTA resin to bind His_6 tagged chimeric protein, samples collected from the column flow-through (Sample 3) should contain unbound proteins that are not His_6 tagged P-gp chimeras, as chimeras should be completely immobilized on Ni-NTA resin. Similarly, samples collected during the washing of the column (Sample 4) should not contain P-gp. After elution of P-gp with 300 mM imidazole (Sample 5), analysis should reflect

the isolation of purified P-gp. To concentrate protein to a volume suitable for SEC, a protein concentrator with a 100 kDa cutoff was used. Samples of both the concentrated, purified protein (Sample 6) and the concentrator flow-through (Sample 7) were collected for analysis, as the presence of P-gp in the concentrator flow-through may indicate degraded P-gp (less than 100 kDa) in the sample. Ultracentrifugation was performed to remove aggregated protein. After ultracentrifugation, samples were collected from both the supernatant (Sample 8) and the pellet (Sample 9), which was collected and solubilized in buffer. All fractions produced by SEC were collected for analysis. SEC fractions that contained monomeric P-gp were collected, combined, and concentrated to 5-10 mg/mL for crystallization screening (Sample 10).

Both western blotting and colloidal blue staining were used to generate a complete picture of IMAC-mediated purification (Figure 6.10). Western blotting can be used to confirm the presence of P-gp and can identify P-gp degradation products. However, western blotting with C219 cannot detect the presence of proteins that are not P-gp. Thus, colloidal blue staining was used to assess the purity of the samples, as this will elucidate the presence of protein contaminants that were not eliminated during purification.

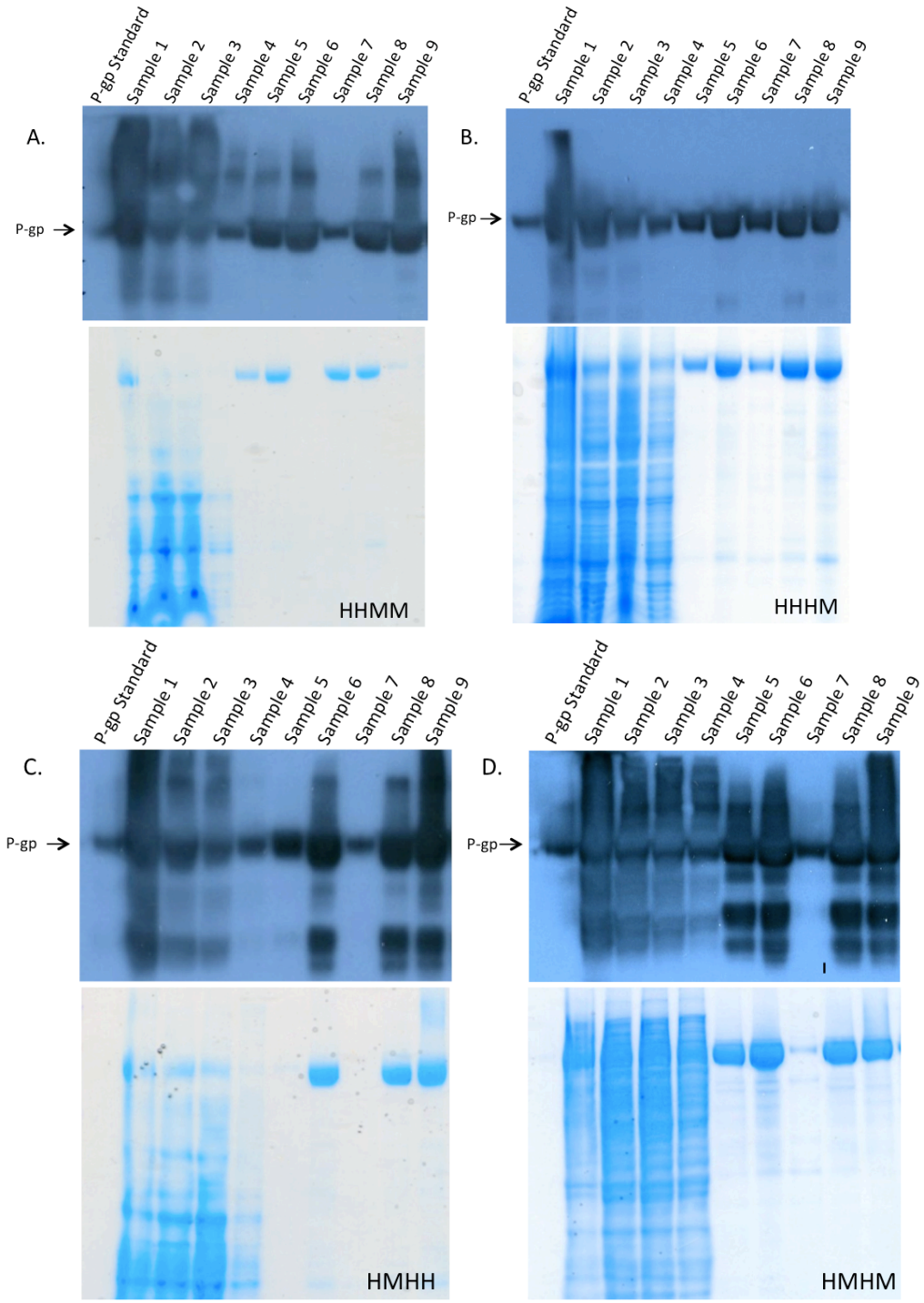
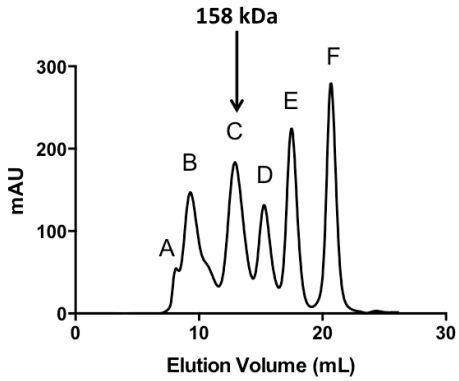


Figure 6.10 Western blotting and colloidal blue staining of protein samples collected during IMAC-mediated purification. Western blotting and colloidal blue staining of **A)** HHMM, **B)** HHHM, **C)** HMHH, and **D)** HMHM P-gp chimeras. SDS-PAGE was carried out (5 µg protein/lane) and protein was transferred onto nitrocellulose membranes for C219 western blotting (Section 2.2.20). Subsequently, SDS-PAGE gels were subjected to colloidal blue staining (Section 2.2.21). Samples were collected as described in the text (Section 6.2.2.4). Purified mouse P-gp was used as a positive control and size marker (left lane).

Colloidal blue staining confirmed the purification of P-gp chimeras as visualized by the comparison of samples collected before IMAC (Samples 1-4) with samples collected after elution of His_6 tagged protein from Ni-NTA resin (Samples 5-9) (Figure 6.12). The presence of P-gp was confirmed by western blotting with the C219 monoclonal antibody. Additionally, the presence of chimeric P-gp in the pellet after ultracentrifugation (Sample 9) was confirmed, indicating the aggregation of P-gp.

6.2.2.5 SEC of HHMM, HMHM, HMHH, and HHHM chimeric P-gp

SEC is employed to assess and isolate monomeric protein, separating proteins by size, with higher molecular weight proteins eluting before low molecular weight proteins. To identify both monomeric and aggregated P-gp, a protein standard was used (Figure 6.11). P-gp expressed in insect cells is not glycosylated (249). Therefore, monomeric P-gp should be approximately 140 kDa. Monomeric P-gp can be identified by comparison to the peak resulting from γ-globulin (158 kDa) in the protein standard. As P-gp aggregates are a higher molecular weight than monomeric P-gp, aggregates are observed at the peak representing thyroglobulin (670 kDa) or at a lower elution volume.



	Protein	Molecular Weight (kDa)
A	Aggregates	-
B	Thyroglobluin	670
C	γ-globulin	158
D	Ovalbumin	44
E	Myoglobin	17
F	Vitamin B_{12}	1.35

Figure 6.11 Protein standard SEC chromatogram. The Bio-Rad gel filtration standard was used as a size marker during SEC of chimeric proteins. Proteins included in the standard are labeled A through F and their corresponding molecular weights are displayed.

SEC was conducted after IMAC to isolate monomeric protein for crystallization screening (Figure 6.12). SEC chromatograms displayed a peak of purified monomeric P-gp, which was collected in SEC fractions 21 to 23 for all chimeras. Additionally, all purified chimeras contained aggregated protein, as indicated by the broad secondary peak ranging from approximately 300 to 700 kDa. Chimeras showed subtle differences in the amount of aggregated protein present. For example, more aggregated protein was observed with the HHHM chimera than the HHMM chimera (Figure 6.11). SEC fractions 21-23 were collected and concentrated to 5-10 mg/mL to be used in crystallization screening. SEC fractions from 15 to 19 and 24 to 29 were also collected for all chimeras for subsequent analysis by SDS PAGE (western blotting and colloidal blue staining) and blue native PAGE (BN-PAGE).

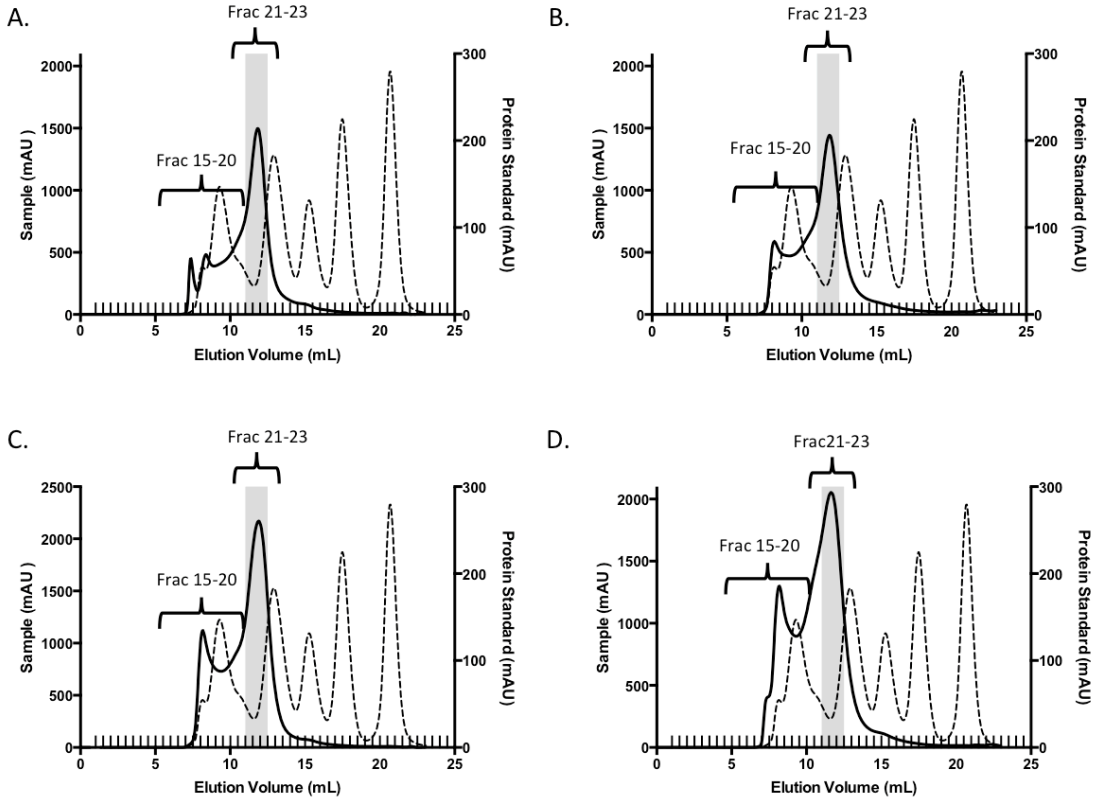


Figure 6.12 SEC chromatograms of purified HMHH, HHMM, HMHM and HHHM P-gp chimeras. SEC chromatograms for **A)** HMHH, **B)** HHMM, **C)** HMHM, and **D)** HHHM, P-gp chimeras are shown (Section 2.2.42). All graphs display the chromatograms of purified chimeric P-gp (left y-axis, mAU Sample) and of a protein standard (right y-axis, dotted line, mAU Protein Standard). Eluted protein was collected in 0.5 mL fractions, indicated by upward tick marks on the x-axis. Fractions containing P-gp aggregates (Frac 15-20) and monomeric P-gp (Frac 21-23) are indicated. Shaded gray areas indicate the fractions that were collected and concentrated to generate protein used in crystallization screening.

All SEC fractions were collected for subsequent analysis by western blotting (Figure 6.13) and colloidal blue staining (Figure 6.14) also confirming the presence of purified P-gp absent contaminants. Sample 10 represents SEC fractions 21-23 that were collected, pooled, and concentrated for use in crystallographic screening.

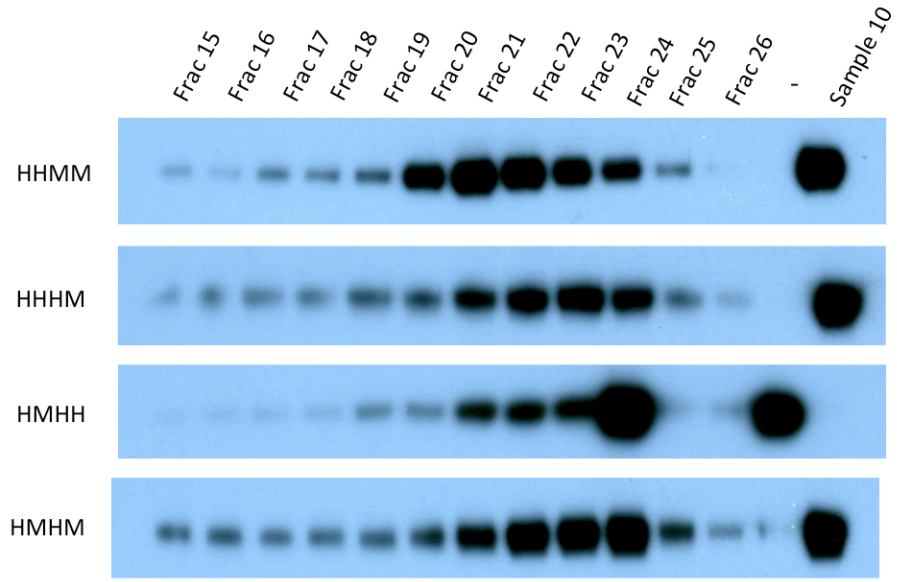


Figure 6.13 Western blots of SEC fractions. SEC fractions were collected and protein concentration was estimated. All SEC fractions were subjected to a 200-fold dilution in SEC buffer to obtain final concentrations of protein ranging from 0.1 to 0.5 µg protein. Equal volumes of protein (not concentration) was subjected to SDS-PAGE and subsequent C219 western blotting (Section 2.2.20). Sample 10 represents SEC fractions 21-23 that were collected, pooled, and concentrated for use in crystallographic screening. Note that a blank lane (-) was left between Frac 26 and Sample 10 to account for signal due to possible spill-over.

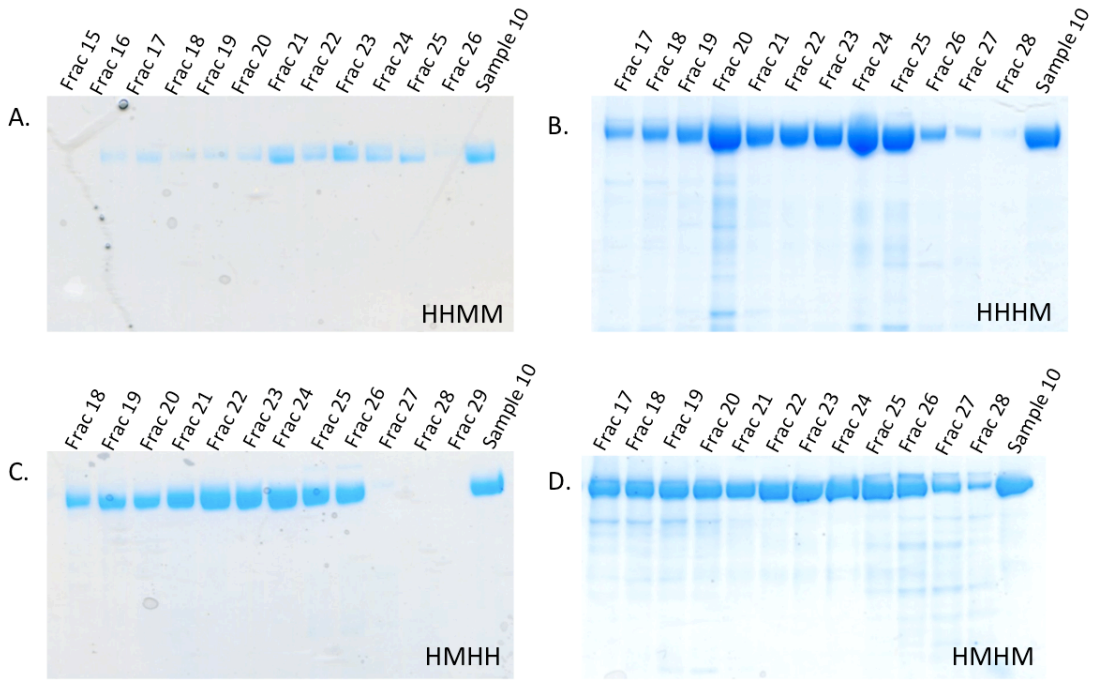


Figure 6.14 Colloidal blue staining of SEC fractions. SEC fractions were collected and protein estimation was performed. Colloidal blue staining of the **A)** HHMM, **B)** HHHM, **C)** HMHH, and **D)** HMHM chimeras was completed (5 µg protein/ well). Sample 10 represents SEC fractions 21-23 that were collected, pooled, and concentrated for use in crystallographic screening.

6.2.2.6 Yield of HHMM, HMHM, HMHH, and HHHM purification

Protein concentration was estimated to assess the yield of purified protein at various stages (Table 6.4). The concentration of crude membranes resulting from a large-scale (4 L) culture of infected insect cells was determined at the outset of chimeric purification. Protein concentration was also estimated after crude membranes were incubated with DDM to solubilize membrane proteins and ultracentrifugation was carried out to remove excess lipids and insoluble components. IMAC purification of chimeric proteins was performed, and protein concentration of the eluted protein was estimated. Additionally, it was of interest to determine protein concentration both before and after ultracentrifugation to remove aggregated P-gp protein (Table 6.5). One of the main differences between mouse and human P-gp is the proclivity for human P-gp to aggregate. For example, approximately 80-90% of purified human P-gp typically aggregates during

ultracentrifugation to form a precipitate (Fei Zhou, personal communication). In contrast, approximately 10% of mouse P-gp typically aggregates during ultracentrifugation (Fei Zhou, personal communication). It is not known why human and mouse P-gp display different aggregation properties, although these differences are thought to be a major factor contributing to the ability to crystallize mouse but not human P-gp. After ultracentrifugation, SEC was carried out to isolate monomeric P-gp, an estimation of protein concentration was determined for the SEC fractions that were collected and concentrated for crystallization screening (fractions 21 to 23).

Table 6.4 Total purified protein obtained for crystallization screening of chimeric P-gp. Protein concentration (Section 2.2.28) was determined at various steps during the purification of chimeras.

Chimera	Crude membranes (mg)	Detergent-solubilized protein (mg)	Eluate from Ni-NTA resin (mg)	After ultra-centrifugation (mg)	SEC fractions 21-23 (mg)	Purification Yield
HHMM	1,426	1,179.2	11.0	3.7	1.38	0.096%
HMHM	1,091	958.4	15.5	12.0	2.96	0.27%
HHHM	1,145	1,099.2	13.8	8.7	2.05	0.18%
HMHH	1,305	1,172.8	8.0	4.89	1.16	0.088%

Table 6.5 Concentration of purified P-gp before and after ultracentrifugation to remove protein aggregates. Protein concentration (Section 2.2.28) was assessed before and after ultracentrifugation. The decrease in protein concentration is attributed to P-gp aggregation.

Chimera	Before ultra-centrifugation (mg)	After ultra-centrifugation (mg)	% Loss
HHMM	11.0	3.7	66.4%
HMHM	15.5	12.0	22.6%
HHHM	13.8	8.7	37.0%
HMHH	8.0	4.89	38.9%

The overall yield of the purification process was determined by comparing the total quantity of crude membranes with the final quantity of purified, monomeric protein obtained from purification. Purification yields ranged from 0.088% (HMHH chimera) to 0.27% (HMHM chimera). These values are comparable with previously obtained yields from 4 L insect cell cultures expressing wild-type mouse P-gp (S. Shukla, personal communication). Additionally, compared with human and mouse P-gp, chimeras displayed an intermediate amount of protein aggregation, ranging from 22.6% to 66.4% loss due to aggregation. It is not known whether an intermediate phenotype may favorably impact crystallization of chimeric P-gp.

6.2.2.7 Assessment of homogeneity of purified P-gp chimeras

For crystallization, it is key to have monomeric, homogenous, purified protein. Both ultracentrifugation and SEC were employed to remove protein aggregates, but these processes are not 100% effective. One limitation of SDS-PAGE is that protein aggregates cannot be characterized due to denaturing conditions, including the use of SDS. Therefore BN-PAGE, which uses non-denaturing conditions allowing for protein separation by size, was carried out to assess if P-gp monomers, dimers, and aggregates were present in various SEC fractions.

Three SEC fractions of the HMHM and HHHM chimeras were subjected to BN-PAGE to demonstrate the ability to visualize P-gp monomers, dimers, and aggregates (Figure 6.15). P-gp monomers and dimers migrate into the gel during electrophoresis and can be distinguished by the presence of two separate bands, whereas protein aggregates above a certain size (greater than 1,000 kDa) cannot enter the gel, and can be seen trapped in the loading well of the polyacrylamide gel. BN-PAGE of SEC fraction 15 from HMHM and HHHM contains P-gp aggregates that cannot enter the gel, corresponding to the high molecular weight secondary peak observed in the SEC chromatograms. Fractions 21 and 18 collected

from SEC of HMHM and HHHM P-gp, respectively, contain a mixture of P-gp monomers and dimers. While the presence of a P-gp dimer is still present in SEC fractions 24, mainly monomeric P-gp is present. Mouse P-gp has been crystallized in the Laboratory of Cell Biology with SEC fractions that contain both monomeric and dimeric P-gp, suggesting that the presence of dimers may not unduly affect the crystallization process.

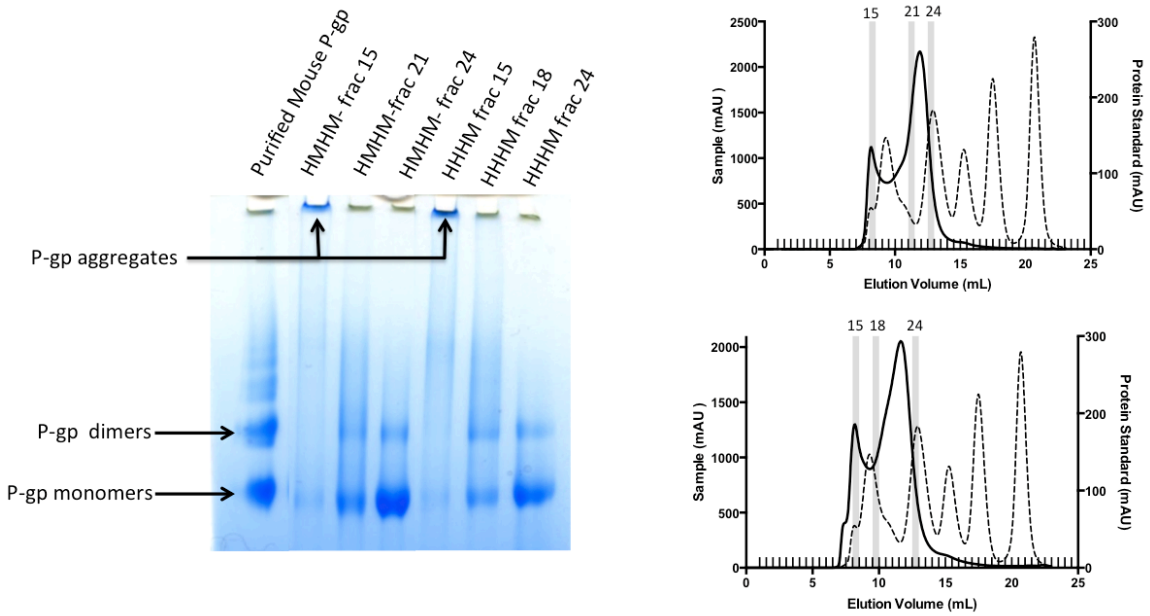


Figure 6.15 BN-PAGE of SEC fractions of HMHM and HHHM chimeric P-gp. Three different SEC fractions of the HMHM and HHHM chimeras (6 µg protein/well) were analyzed by BN-PAGE (Section 2.2.22). SEC chromatograms from the purification of the HMHM (top chromatogram) and HHHM (bottom chromatogram) chimeras are shown with the corresponding fractions used in BN- PAGE indicated by gray bars. Purified, wild-type mouse P-gp (6 µg) was used as a control.

As previously mentioned, SEC fractions 21 to 23 were collected for all chimeras and concentrated to 5-10 mg/mL for crystallization screening. BN-PAGE of the combined fractions was carried out to determine the presence of aggregated P-gp (Figure 6.16). All samples showed a dominant band of monomeric P-gp. While P-gp dimers were also present, they were present at amounts equal to or less than the quantity of dimers observed in

purified mouse P-gp, which is known to yield diffraction quality crystals. Large protein aggregates were not observed, as assessed by the absence of protein in the wells of the polyacrylamide gel. Overall, these data indicate that the chimeric proteins obtained were of sufficient purity and quality to attempt crystallization screening.

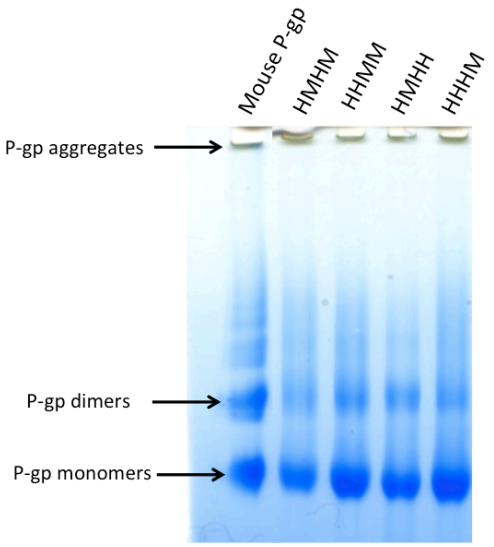


Figure 6.16 BN-PAGE of purified HHMM, HMHM, HMHH, and HHHM chimeric P-gp. BN-PAGE was carried out (6 µg protein/well) to determine the presence of monomeric, dimeric, and aggregated chimeric P-gp. Samples represent pooled SEC fractions 21 to 23 that were combined and concentrated for use in crystallization screening. Purified, wild-type mouse P-gp (6 µg) was used as a control.

6.2.2.8 Crystallization screening of P-gp chimeras

Crystallization trials were completed by Lothar Esser in the Laboratory of Cell Biology. Hanging drop vapor diffusion crystallization was set up in a 96-well format. Initial crystallization screening of purified HHMM, HMHM, HMHH, and HHHM was carried out using MembFrac and MemGold crystallization screening conditions. These conditions include a range of pH, polyethylene glycol (PEG) concentrations, and salt additives that have been successfully used for the crystallization of various membrane proteins. While these studies are currently ongoing, two initial hits were obtained with the HHMM and HHHM

chimeras (Figure 6.17). Crystals of both HHMM and HHHM were obtained with the following conditions: 0.05 M NaCl, 0.1 M tris-Sodium citrate (pH 5.5), and 26% v/v PEG400. These conditions are similar to conditions typically used in this laboratory to crystallize mouse P-gp, which are 0.05 M NaCl, 0.15 M Tris (pH 7.5), and 20% v/v PEG400. HHMM and HHHM protein crystals were identified by imaging with UV light and visual confirmation of fluorescence (297).

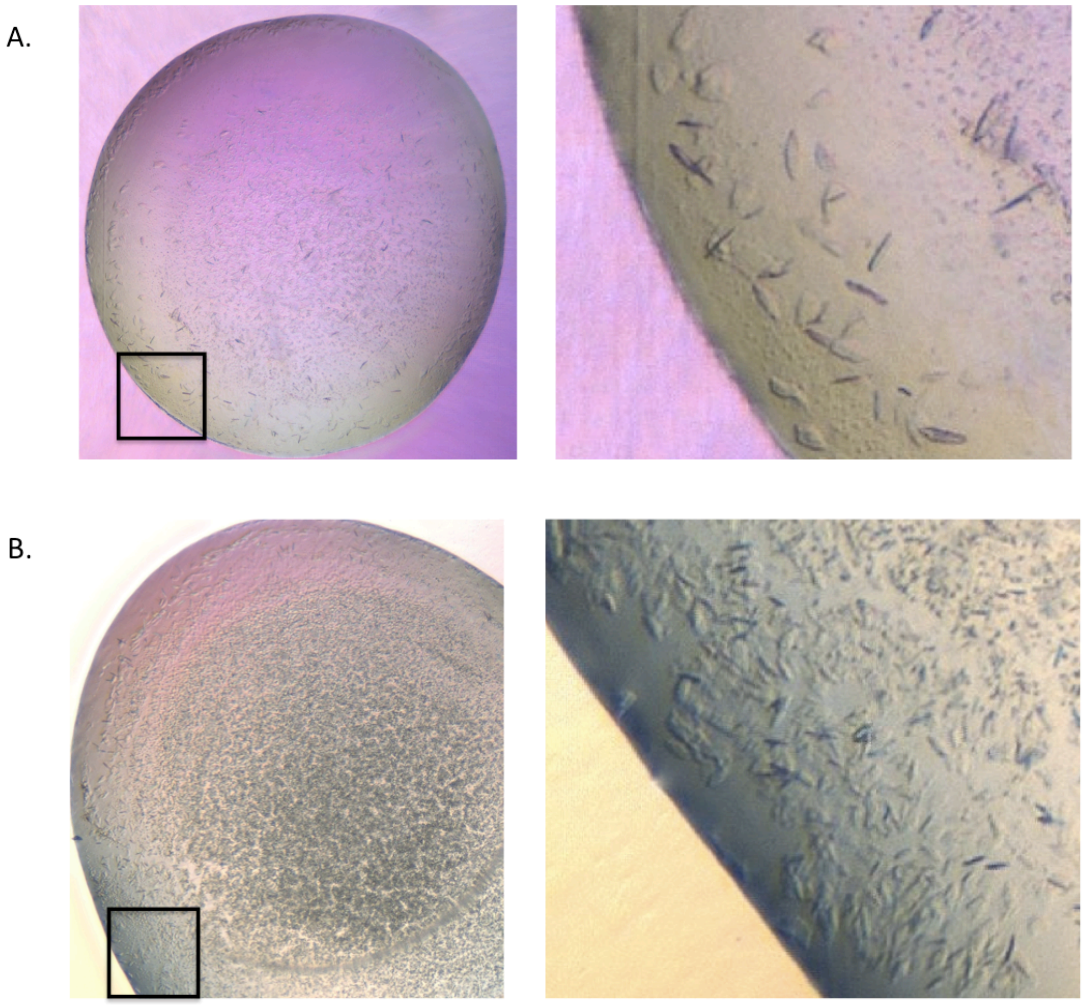


Figure 6.17 Protein crystals of HHMM and HHHM chimeras. Top-down view of droplets containing protein crystals of the **A)** HHMM or **B)** HHHM chimera. Columns represent images of the droplet (left column) and enlarged images of the protein crystals (right column). Black boxes indicate the area displayed in the enlarged images. Crystals are approximately 50 µm in size.

6.3 Discussion

This chapter describes heterologous expression of P-gp chimeras in *P. pastoris* and insect cells. Various growth conditions were tested at a small-scale to express the HHMM chimera in *P. pastoris* (Section 6.2.1.3) but unfortunately the expression conditions identified did not translate to a larger-scale (Section 6.2.1.4), and thus expression in *P. pastoris* was not continued in this study. In contrast, infection of insect cells with recombinant baculovirus using a previously developed protocol (219) was able to achieve high-level expression of chimeras in insect cells (Figure 6.8). Purification of the HHMM, HMHM, HMHH, and HHHM chimeras was completed (Section 6.2.2.4). Samples collected during the purification process of all chimeras were subjected to analysis by C219 western blotting, colloidal blue staining, and BN-PAGE to determine if purified, monomeric protein was obtained (Section 6.2.2.5 and 6.2.2.7). Lastly, crystallization screening was initiated, and protein crystals of the HHMM and HHHM chimeras were obtained (Section 6.2.2.8).

While previous reported structures of mouse P-gp have been obtained from protein expressed in *P. pastoris*, this study was not able to optimize large-scale expression conditions for the HHMM chimera (Figure 6.7). High-level expression of wild-type mouse P-gp has been achieved through transformation *mdr1a* cDNA (in a pPICZ vector) into the GS115 host strain and growth and methanol induction with Condition 1. However, expression conditions for mouse P-gp were not translatable to HHMM P-gp, which highlights the need to optimize expression conditions in *P. pastoris* on a case-by-case basis, as even proteins that are highly homologous can have very different optimal expression conditions.

In contrast, high-level expression of chimeras in insect cells was achieved using a previously optimized protocol without further optimization. As previously mentioned, large-scale protein expression in insect cells was commercially sourced due to the lack of equipment, expertise, and facilities in the Laboratory of Cell Biology to support insect cell expression. Thus, a main drawback to the use baculovirus-mediated insect cell expression in

this laboratory was cost. However, the advantages to insect cell expression in this study far outweighed the drawbacks, as protein expression could be achieved without variation of expression conditions. An additional benefit to baculovirus-mediated expression of chimeric P-gp is that infected insect cell crude membranes can be also used for biochemical characterization of chimeras (Section 5.1.2).

One of the main differences between mouse and human P-gp observed in the Laboratory of Cell Biology is the propensity for purified human P-gp to aggregate. It is not known why human and mouse P-gp display different aggregation properties, although these differences are thought to be a major factor contributing to the ability of mouse but not human P-gp to crystallize. To assess aggregation of chimeric proteins, protein estimation was completed before and after ultracentrifugation (Table 6.5). Chimeras displayed variable amounts of aggregation, ranging from approximately 66%, for the HHMM chimera, to 23%, for the HHHM chimera. Finally, this study included an initial crystallization screening of all purified chimeras using MembFrac and MemGold crystallization screening conditions. Protein crystals of the HHMM and HHHM chimeras were obtained from this initial screening (Figure 6.17), and expression, purification and crystallization screening cycles continue.

6.4 Conclusions

This study examined *P. pastoris* and insect cell heterologous expression systems for the expression of chimeric P-gp. Whereas each expression system offers a unique set of advantages and drawbacks, the studies in this chapter showed that insect cell expression, using a generic protocol previously established for the expression of mouse P-gp in insect cells, was able to achieve high-level expression of all P-gp chimeras without further optimization. As exemplified by the expression of the HHMM chimera in *P. pastoris*, protein expression can often be a time- and resource-consuming process. Therefore, the ability to express chimeric protein in insect cells using a common protocol is extremely valuable. Purification of chimeric P-gp expressed in insect cells yielded milligram quantities of purified, monomeric protein that was used in crystallization. Protein crystals of the HHMM and HHHM chimeras were generated, and crystallization screening of all chimeras is currently ongoing.

Chapter 7: Discussion and Future Directions

7.1 Study overview

While chemotherapy is currently the standard treatment for metastatic and hematological malignancies, drug resistance to chemotherapy, be it intrinsic or acquired, is often inevitable (48). Multidrug resistance (MDR) is characterized by the simultaneous resistance to drugs that differ both structurally and mechanistically. A notorious mediator of MDR is the ABC transporter P-glycoprotein (P-gp, ABCB1), product of the human *MDR*1 gene (5, 102, 137). P-gp actively effluxes a wide range of amphiphilic drugs irrespective of the prevailing concentration gradient, thereby lowering intracellular concentrations below therapeutic levels. P-gp has been characterized extensively *in vitro,* with many drugs used in the clinic having been identified as substrates of P-gp. Overexpression of P-gp has been shown to correlate with overall poor chemotherapy response and prognosis and is a barrier to the successful treatment of cancer (48, 62).

Understanding how P-gp interacts with drugs has been severely limited by the lack of high-resolution structures of P-gp. Although numerous efforts to obtain an X-ray crystal structure of P-gp have been attempted, human P-gp has never been crystallized. However, mouse P-gp (87% homologous to human P-gp) has been crystallized, and several structures of mouse P-gp have been recently reported (57, 103, 136). While undoubtedly informative, the reported structures of mouse P-gp leave several issues unresolved. Firstly, all reported mouse P-gp structures are in the same conformation, therefore limiting the ability to resolve questions regarding P-gp's mechanism of drug transport. Secondly, species differences in substrate recognition between mouse and human P-gp have been reported, thus a high-resolution structure of human P-gp remains highly desirable to understand its physiologic role in cancer treatment (86-88). Lastly, the 3.8 Å resolution reported for the mouse P-gp structure is insufficient to identify amino acid locations and therefore cannot answer questions regarding P-gp polyspecificity or mechanism of transport.

The studies presented in this thesis describe the creation of novel chimeras of mouse and human P-gp as an approach to investigate whether specific protein domains are responsible for differences in the ability to form crystals between mouse and human P-gp. A range of chimeras were generated using a molecular biology strategy designed to introduce silent restriction enzyme sites at the domain boundaries of mouse and human P-gp, and subsequent domain swapping between species (Sections 4.2.1 and 4.2.2). Four chimeras, termed HHMM, HMHM, HHHM, and HMHH (Figure 4.2) were successfully created. All chimeras were functionally characterized, and showed functional activity similar to wild-type human P-gp thus warranting structural investigations (Chapter 5). A parallel comparison of heterologous protein expression in insect cells and *P. pastoris* was conducted, identifying baculovirus-mediated expression in insect cells as an expression system able to achieve high-level protein expression without optimization on a chimera-to-chimera basis (Sections 6.2.1 and 6.2.2.1). Lastly, purification of all chimeras and an initial crystallization screen was completed, yielding crystals of the HHMM and HHHM chimeras (Sections 6.2.2.4 and 6.2.2.8). While structural data from chimeras has yet to be realized, the attainment of initial crystals of two chimeras validates the generation of P-gp chimeras as a rational approach to structural investigations.

The chimera strategy employed here is not without precedent. A chimerization strategy has been successfully utilized to determine the eukaryotic cytoplasmic pore region of the inward rectifier potassium channel Kir3.1, a region that is crucial to protein function and regulation. Prior to the structural elucidation of the eukaryotic pore of Kir3.1, the structure of a prokaryotic Kir (KirBac1.3) from the gram-negative bacterium *Burkholderia pseudomallei* was reported (205). However, the bacterial cytoplasmic pore of KirBac1.3 is only 19% identical to the pore of human Kir3.1, leaving much to be desired about the structure of the eukaryotic intracellular pore. Utilizing the fact that KirBac1.3 could be successfully expressed, purified, and crystallized, the authors generated a KirBac1.3-Kir3.1

prokaryotic-eukaryotic chimera, where the cytoplasmic pore region of KirBac1.3 was substituted with the murine Kir3.1 cytoplasmic pore (99% identical to the human sequence). The chimeric Kir3.1 structure, described at 2.2 Å resolution, indicates a possible mechanism of action of how PIP_2 acts to regulate the channel.

Chimeric Kir3.1 exemplifies how strategies of homolog screening and chimera generation can be used in combination to obtain relevant crystal structures. Homolog screening was previously utilized to identify mouse P-gp as a P-gp homolog able to be highly expressed in and purified from *P. pastoris*, and was therefore selected for crystallization screening. This thesis project builds upon the successful expression purification, and crystallization observed with mouse P-gp. Similarly to Kir3.1, this thesis aimed to use chimerization of mouse and human P-gp as a strategy to crystallize domains of human P-gp.

7.2 Study limitations

Limitations to various strategies employed this thesis project can be acknowledged. The ability for mouse, but not human, P-gp to crystallize displays how subtle changes in the primary amino acid sequence is a key factor in determining whether a protein will crystallize or not. Small changes in the amino acid composition of proteins may greatly affect properties of heterologous protein expression, purification, and crystallization. For example, the introduction of six point mutations into the β1-adrenergic GPCR was shown to increase the stability of the protein in detergent; a factor that allowed for crystal formation and subsequent structural elucidation (194, 195). It is likely that non-homologous residues between mouse and human P-gp contribute unequally to differences in crystallization, with differences in crystallization most likely due to a handful of key residues that disproportionately impact crystal formation. Therefore, a domain-swapping approach may be viewed as unsystematic, as this strategy relies on the chance that residues which favorably impact crystal formation in mouse P-gp can be introduced to human P-gp by the

introduction of particular domains. A more targeted approach to the crystallization of human P-gp could use the identification of key residues likely to contribute to differences in crystallization between species, and subsequent mutation of those specific residues in human P-gp to the homologous mouse residue. While a targeted approach may leave less to chance compared to a domain swapping approach, currently, it is not known which residues impact crystal formation. Thus, a targeted mutational approach was not carried out in this thesis.

To gain insight into which residues may impact P-gp crystal formation, an analysis of the crystal contacts present in the reported mouse P-gp structure was conducted (Section 4.2.3) (57). Intermolecular forces that comprise crystal contacts are of significance, as they are an important factor to crystal formation. To determine if a crystal contact present in the mouse P-gp structure may theoretically be present in human or chimeric P-gp, the analogous residues in the crystal contact needed to be identical to the corresponding mouse residues to be considered a conserved contact. Amino acid residues vary in properties such as size, polarity, and charge, and non-conserved residues can be either similar or dissimilar in such properties. A potential shortcoming of the crystal contact analysis is that it does not take into account the fact that non-conserved crystal contacts could be composed of amino acid residues with similar properties to the homologous mouse P-gp residues, and thus the contact would, in fact, be conserved. An analysis investigating the similarity of non-conserved crystal contacts between mouse and human P-gp was conducted (data not shown). All crystal contacts in mouse P-gp involving ECL1 were excluded from this analysis, as ECL1 was not considered homologous between mouse and human P-gp (Figure 4.18). When similarity was considered, there were various pairs of amino acid residues involved in crystal packing of mouse P-gp that were similar in human P-gp. However, this was not always the case. For example, mouse P-gp has various crystal contacts composed of uncharged, polar residues, such as glutamine or serine, where the analogous human residue

is the positively charged residue lysine. It may be that a handful of dissimilar residues may be responsible for the inability to crystallize human P-gp. A future study investigating the ability of these dissimilar residues to contribute to differences in crystallizability between homologs is warranted.

This study also revealed that an important difference between mouse and human P-gp is the size and conformation of ECL1, which is involved in approximately 18% of all mouse P-gp crystal contacts. An additional shortcoming of this project is that chimeras generated did not incorporate information uncovered in the analysis of mouse P-gp crystal contacts in chimera generation. This is due to the prioritization of chimera generation in the timeline of this project (molecular biology was performed in Oxford prior to time spent at the NIH investigating crystallographic aspects of mouse and human P-gp). Therefore, the analysis of crystal contacts was completed after chimeras were generated, and chimeras containing mouse ECL1 were not investigated. While the lack of chimeras containing mouse ECL1 was predicted to possibly impede crystal formation of the chimeras, crystals of both the HHMM and HHHM chimeras, which contain human ECL1, have been grown. Nonetheless, future studies should incorporate data generated from the crystal contact analysis presented in Section 4.2.3 in the next round of chimera generation.

7.3 Selected advances in membrane protein purification and crystallization

Membrane proteins are notoriously difficult to crystallize for various reasons. Successful protein crystallization requires the preparation of pure, monodispersed, stable protein in milligram quantities. Protein purification of membrane proteins may be especially difficult, as membrane proteins are often unstable in a detergent-solubilized state. Additionally, the physiological nature of certain membrane proteins may make them intrinsically difficult to crystallize. For example, membrane proteins, such as P-gp, are often highly flexible and the ability to adopt numerous conformations may impede crystal

formation. Also, hydrophobic transmembrane regions of purified membrane proteins are shielded by detergents, a fact that may inhibit crystal packing and limit the amount of available hydrophilic residues that can partake in crystal packing forces.

Various noteworthy advances have allowed for the determination of several high-resolution G-protein-coupled receptor (GPCR) crystal structures (298). GPCRs constitute a large family of eukaryotic membrane protein receptors that are involved in a diverse range of signal transduction pathways. GPCRs possess many of the aforementioned challenges relating to membrane protein crystallography. For example, GPCRs are dynamic proteins that undergo large conformational changes, particularly in the absence of a bound substrate or ligand. Additionally, purified GPCRs are unstable in detergents, tending to denature in a detergent-solubilized state. To address protein instability in detergents, various optimization efforts have been carried out. However, while the optimization of detergents increased protein stability, GPCRs were nonetheless unable to be crystallized, presumably due to the size of resulting micelle and the inability for the comparatively small regions of soluble protein to interact to form the crystal contacts necessary for crystallization.

At least three main strategies have been employed to generate high-resolution crystal structures of several GPCRs. Two of these methods used the addition of either a antigen-binding fragment (Fab) or the insertion of T4 lysozyme to the cytoplasmic loop 3 of the β_2 adrenergic GPCR to both increase area of hydrophilic residues able to partake in intermolecular packing forces and also to conformationally stabilize the protein, allowing for structural elucidation (299, 300). While these methods relied on the addition of a soluble element to increase the surface area available for intermolecular packing, a third strategy that has also been successfully employed for structural elucidation is the generation of GCPR mutants with increased stability in a detergent-solubilized state. As previously mentioned, a series of six point mutations was shown to increase the stability of the β1 adrenergic GPCR, subsequently leading to its structural elucidation (194).

While ABC transporters and GPCRs are not structurally or functionally similar, strategies that have enabled the structural elucidation of various GPCRs may be translatable to the crystallization of human P-gp and/or P-gp chimeras. For example, the extracellular, hydrophilic regions of P-gp are composed of four ECLs of approximately 5 to 10 residues. While the ECLs comprise only a small area of the total surface area of P-gp, ECLs are crucial for crystal formation of mouse P-gp, participating in numerous crystal contacts between neighboring P-gp molecules in the reported crystal structure. Similar to strategies using Fab and T4 lysozyme for the structural elucidation of the β_2 adrenergic GPCR, it may be advantageous for the crystallization of P-gp chimeras to increase the amount of hydrophilic surface area in the ECLs. Therefore, future studies involving P-gp chimeras generated in this study should include co-crystallization with antibodies. The HMHM, HMHH, and HHHM chimeras also present a unique opportunity to use the mouse anti human P-gp monoclonal antibodies MRK-16 and UIC2, as these antibodies only recognize a conformational epitope composed of the ECLs of human P-gp and thus have been previously unable to be utilized in crystallization trials of P-gp homologs. Additionally, these antibodies may stabilize the conformation of P-gp, increasing the ability for chimeras to be crystallized.

Structural elucidation of particular membrane proteins by x-ray crystallography may be unattainable. As an alternative to x-ray crystallography, recent technological advances in cryo-electron microscopy (cryo-EM) have made it an attractive method for structural elucidation of membrane proteins (124). Notably, a subnanometer structure of the ABC exporter TmrAB, generated using cryo-EM, was recently reported (122). Cryo-EM has several distinct advantages over x-ray crystallography. Firstly, cryo-EM images are obtained from membrane proteins in the native environment of the lipid bilayer. Therefore, crystallization screening can be avoided, as protein crystal are not required for structural determination. Additionally, as membrane proteins are retained in their native environment, images obtains from cryo-EM are typically considered physiologically relevant. In contrast, crystal formation

frequently occurs in non-physiologic conditions, and structures obtained by x-ray crystallography may be considered artifacts of the crystallization process. Cryo-EM can also image the protein of interest in various conformations, which could potentially yield insight into mechanistic aspects of various membrane proteins. For these reasons, a parallel study in the Laboratory of Cell Biology is using cryo-EM as a method to gain structural insight into P-gp is currently being conducted. However, despite the numerous advantages of cryo-EM, it cannot currently achieve the resolution capable with x-ray crystallography. Therefore, while cryo-EM merits investigation, x-ray crystallography remains the gold standard for unambiguous structural determination at an atomic level.

7.4 Future Studies

Of four chimeras generated in this study, two (HHMM and HHHM) produced protein crystals during initial crystallization screening. While protein crystals of the HHHM and HHMM chimeras were obtained, they are not yet of a sufficient size or quality to obtain structural data. Future studies should be conducted to optimize crystallization conditions so that diffraction quality crystals can be obtained. Additionally, future chimeras should be generated based on the outcome of these experiments. For example, if a high resolution structure of a chimera is obtained, future rounds of chimera generation may generate a new series of chimeras aiming to "zone-in" on the residues of human P-gp that are responsible for species differences observed during crystallization. Additionally, while crystals of the HMHH and HMHM chimera were not obtained during initial screening, future studies should continue to attempt to crystallize these constructs. Strategies that should be used to crystallize these chimeras include the screening of additional commercially available membrane protein crystallization screening kits. Furthermore, this laboratory has a previously established set of crystallization conditions similar to the conditions used to crystallize mouse P-gp that should be used to screen all chimeras generated in this thesis.

Lastly, as previously mentioned, future studies will use antibody and Fab binding as a strategy to increase the hydrophilic surface area available to form crystal contacts and potentially lock the protein into a single conformation. This is particularly attractive for the HMHM, HMHH, and HHHM chimeras as the monoclonal antibodies MRK-16 and UIC2 can be utilized.

While the molecular cloning strategy described in this thesis could have generated additional chimeras, ultimately, each chimera needed to be individually created, functionally characterized, optimized for expression conditions, and screened for crystallization, creating a bottleneck to the number of chimeras that could be investigated concurrently. A pragmatic decision was made to restrict the number of chimeras to be functionally characterized, expressed, and screened for crystallization, to the four described in Chapter 4. However, future studies should generate new chimeric constructs that were not investigated here. For example, a chimera composed of a human P-gp background containing all of the residues involved in mouse crystal contacts should be generated and tested for functionality and crystallization. Also, a chimera composed of human P-gp with the surface residues mutated to the homologous mouse P-gp residues should be generated and screened for crystallization. Lastly, future chimeras generated should incorporate the M107L mutation in mouse P-gp to aid in cloning of *mdr1a* during chimera generation. Additionally, this thesis carried out protein expression in two commonly used heterologous expression systems: *P. pastoris* and insect cells. While chimeric protein was unable to be expressed at a large-scale in *P. pastoris*, a previously optimized protocol was used to attain high-level expression of all chimeric P-gps in insect cells. Therefore, future chimeras should only be expressed using baculovirus-mediated expression in insect cells, which would be expected to expedite the attainment of purified protein suitable for crystallization screening.

7.5 Concluding remarks

Membrane proteins are notoriously recalcitrant to methods of heterologous protein expression, purification, and crystallization. A popular strategy that is used to obtain membrane protein crystal structures is the screening of homologs from various species to, by chance, uncover a homolog with favorable pre-crystallographic and crystallographic properties. However, this strategy may be viewed as a starting point to structural elucidation of the human homolog when the crystallized homolog either 1) shares low homology with the human homolog, as is the case with *C. elegans* P-gp, or 2) has characterized functional differences compared to the human homolog, as observed with mouse P-gp. This thesis built upon the previously laid foundation of the ability to purify and crystallize mouse P-gp. An initial crystallization screen obtained protein crystals for two of the chimeras, indicating the approach adopted is a successful strategy, and an advance along the path towards a high-resolution structure of human P-gp.

References

1. Kang J*, et al.* (2011) Plant ABC Transporters. *The Arabidopsis book / American Society of Plant Biologists* 9:e0153.

2. Rees DC, Johnson E, & Lewinson O (2009) ABC transporters: the power to change. *Nature reviews. Mol. Cell Biol.* 10(3):218-227.

3. Kavishe RA*, et al.* (2009) Localization of the ATP-binding cassette (ABC) transport proteins PfMRP1, PfMRP2, and PfMDR5 at the Plasmodium falciparum plasma membrane. *Malaria J.* 8:205.

4. Szakacs G, Paterson JK, Ludwig JA, Booth-Genthe C, & Gottesman MM (2006) Targeting multidrug resistance in cancer. *Nat. Rev. Drug Disc.* 5(3):219-234.

5. Ambudkar SV, Kimchi-Sarfaty C, Sauna ZE, & Gottesman MM (2003) P-glycoprotein: from genomics to mechanism. *Oncogene* 22(47):7468-7485.

6. Dean M, Rzhetsky A, & Allikmets R (2001) The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res.* 11(7):1156-1166.

7. Zarubica A, Trompier D, & Chimini G (2007) ABCA1, from pathology to membrane function. *Pflugers Arch.* 453(5):569-579.

8. Oram JF (2000) Tangier disease and ABCA1. *Biochim. Biophys. Acta* 1529(1-3):321-330.

9. Bullard JE, Wert SE, & Nogee LM (2006) ABCA3 deficiency: neonatal respiratory failure and interstitial lung disease. *Sem. Perinatol.* 30(6):327-334.

10. Westerfeld C & Mukai S (2008) Stargardt's disease and the ABCR gene. *Sem. Ophthalmol.* 23(1):59-65.

11. Molday RS, Zhong M, & Quazi F (2009) The role of the photoreceptor ABC transporter ABCA4 in lipid transport and Stargardt macular degeneration. *Biochim. Biophys. Acta* 1791(7):573-583.

12. Akiyama M & Shimizu H (2008) An update on molecular aspects of the non-syndromic ichthyoses. *Exp. Dermatol.* 17(5):373-382.

13. Abele R & Tampe R (2004) The ABCs of immunology: structure and function of TAP, the transporter associated with antigen processing. *Physiology* 19:216-224.

14. Alissa FT, Jaffe R, & Shneider BL (2008) Update on progressive familial intrahepatic cholestasis. *J. Ped. Gastroenterol. Nutr.* 46(3):241-252.

15. Pondarre C*, et al.* (2007) Abcb7, the gene responsible for X-linked sideroblastic anemia with ataxia, is essential for hematopoiesis. *Blood* 109(8):3567-3569.

16. Camaschella C (2009) Hereditary sideroblastic anemias: pathophysiology, diagnosis, and treatment. *Sem. Hematol.* 46(4):371-377.

17. Nies AT & Keppler D (2007) The apical conjugate efflux pump ABCC2 (MRP2). *Pflugers Arch.* 453(5):643-659.

18. Jedlitschky G, Hoffmann U, & Kroemer HK (2006) Structure and function of the MRP2 (ABCC2) protein and its role in drug disposition. *Expert Opin. Drug Metab. Toxicol.* 2(3):351-366.

19. Finger RP*, et al.* (2009) Pseudoxanthoma elasticum: genetics, clinical manifestations and therapeutic approaches. *Survey Ohthalmol.* 54(2):272-285.

20. Bergen AA, Plomp AS, Hu X, de Jong PT, & Gorgels TG (2007) ABCC6 and pseudoxanthoma elasticum. *Pflugers Arch.* 453(5):685-691.

21. Gadsby DC, Vergani P, & Csanady L (2006) The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* 440(7083):477-483.

22. Hwang TC & Sheppard DN (2009) Gating of the CFTR Cl- channel by ATP-driven nucleotide-binding domain dimerisation. *J. Physiol.* 587(Pt 10):2151-2161.

23. James C, Kapoor RR, Ismail D, & Hussain K (2009) The genetic basis of congenital hyperinsulinism. *J. Med. Genet.* 46(5):289-299.

24. Kemp S & Wanders RJ (2007) X-linked adrenoleukodystrophy: very long-chain fatty acid metabolism, ABC half-transporters and the complicated route to treatment. *Mol. Genet. Metab.* 90(3):268-276.

25. Imanaka T, Aihara K, Suzuki Y, Yokota S, & Osumi T (2000) The 70-kDa peroxisomal membrane protein (PMP70), an ATP-binding cassette transporter. *Cell Biochem. Biophys.* 32 Spring:131-138.

26. Oram JF & Vaughan AM (2006) ATP-Binding cassette cholesterol transporters and cardiovascular disease. *Circ. Res.* 99(10):1031-1043.

27. Moser HW, Mahmood A, & Raymond GV (2007) X-linked adrenoleukodystrophy. *Nat. Clin. Prac. Neurol.* 3(3):140-151.

28. Wanders RJ, Visser WF, van Roermund CW, Kemp S, & Waterham HR (2007) The peroxisomal ABC transporter family. *Pflugers Arch.* 453(5):719-734.

29. Kemp S*, et al.* (2001) ABCD1 mutations and the X-linked adrenoleukodystrophy mutation database: role in diagnosis and clinical correlations. *Human Mut.* 18(6):499-515.

30. Berger J & Gartner J (2006) X-linked adrenoleukodystrophy: clinical, biochemical and pathogenetic aspects. *Biochim. Biophys. Acta* 1763(12):1721-1732.

31. van Roermund CW*, et al.* (2008) The human peroxisomal ABC half transporter ALDP functions as a homodimer and accepts acyl-CoA esters. *FASEB J.* 22(12):4201-4208.

32. de Lange EC (2004) Potential role of ABC transporters as a detoxification system at the blood-CSF barrier. *Adv. Drug Deliv. Rev.* 56(12):1793-1809.

33. Kannan P*, et al.* (2009) Imaging the function of P-glycoprotein with radiotracers: pharmacokinetics and in vivo applications. *Clin. Pharmacol. Ther.* 86(4):368-377.

34. van der Deen M*, et al.* (2005) ATP-binding cassette (ABC) transporters in normal and pathological lung. *Respir. Res.* 6:59.

35. International Transporter C*, et al.* (2010) Membrane transporters in drug development. *Nat. Rev. Drug Disc.* 9(3):215-236.

36. Yan Q (2010) Membrane transporters and drug development: relevance to pharmacogenomics, nutrigenomics, epigenetics, and systems biology. *Meth. Mol. Biol.* 637:1-21.

37. Obermeier B, Daneman R, & Ransohoff RM (2013) Development, maintenance and disruption of the blood-brain barrier. *Nat. Med.* 19(12):1584-1596.

38. ElAli A & Hermann DM (2011) ATP-binding cassette transporters and their roles in protecting the brain. *Neuroscientist* 17(4):423-436.

39. Pardridge WM (2007) Blood-brain barrier delivery. *Drug Disc. Today* 12(1-2):54-61.

40. Loscher W & Potschka H (2005) Blood-brain barrier active efflux transporters: ATP-binding cassette gene family. *NeuroRx* 2(1):86-98.

41. Schinkel AH, Wagenaar E, van Deemter L, Mol CA, & Borst P (1995) Absence of the mdr1a P-Glycoprotein in mice affects tissue distribution and pharmacokinetics of dexamethasone, digoxin, and cyclosporin A. *J. Clin. Inv.* 96(4):1698-1705.

42. Kreisl WC*, et al.* (2010) P-glycoprotein function at the blood-brain barrier in humans can be quantified with the substrate radiotracer 11C-N-desmethyl-loperamide. *J. Nuc. Med.* 51(4):559-566.

43. http://www.who.int/medicines/publications/essentialmedicines/en/ (2013) WHO Model List of Essential Medicines. (World Health Organization).

44. Kannan P*, et al.* (2010) N-desmethyl-loperamide is selective for P-glycoprotein among three ATP-binding cassette transporters at the blood-brain barrier. *Drug Metab. Dispos.* 38(6):917-922.

45. Sadeque AJ, Wandel C, He H, Shah S, & Wood AJ (2000) Increased drug delivery to the brain by P-glycoprotein inhibition. *Clin. Pharmacol. Ther.* 68(3):231-237.

46. Prueksaritanont T*, et al.* (2013) Drug-drug interaction studies: regulatory guidance and an industry perspective. *AAPS J.* 15(3):629-645.

47. Longley DB & Johnston PG (2005) Molecular mechanisms of drug resistance. *J. Pathol.* 205(2):275-292.

48. Gottesman MM, Fojo T, & Bates SE (2002) Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat. Rev. Cancer* 2(1):48-58.

49. Pluchino KM, Hall MD, Goldsborough AS, Callaghan R, & Gottesman MM (2012) Collateral sensitivity as a strategy against cancer multidrug resistance. *Drug Resist. Upd.* 15(1-2):98-105.

50. Juliano RL & Ling V (1976) A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochim. Biophys. Acta* 455(1):152-162.

51. Gottesman MM & Ling V (2006) The molecular basis of multidrug resistance in cancer: the early years of P-glycoprotein research. *FEBS Lett.* 580(4):998-1009.

52. Chen CJ*, et al.* (1986) Internal duplication and homology with bacterial transport proteins in the mdr1 (P-glycoprotein) gene from multidrug-resistant human cells. *Cell* 47(3):381-389.

53. Sharom FJ (2011) The P-glycoprotein multidrug transporter. *Essays Biochem.* 50(1):161-178.

54. Gottesman MM, Ambudkar SV, & Xia D (2009) Structure of a multidrug transporter. *Nat. Biotechnol.* 27(6):546-547.

55. Hall MD, Handley MD, & Gottesman MM (2009) Is resistance useless? Multidrug resistance and collateral sensitivity. *Trends Pharmacol. Sci.* 30(10):546-556.

56. Raviv Y, Pollard HB, Bruggemann EP, Pastan I, & Gottesman MM (1990) Photosensitized labeling of a functional multidrug transporter in living drug-resistant tumor cells. *J. Biol. Chem.* 265(7):3975-3980.

57. Li J, Jaimes KF, & Aller SG (2014) Refined structures of mouse P-glycoprotein. *Protein science : a publication of the Protein Soc.* 23(1):34-46.

58. Dawson RJ & Locher KP (2006) Structure of a bacterial multidrug ABC transporter. *Nature* 443(7108):180-185.

59. Borst P, Jonkers J, & Rottenberg S (2007) What makes tumors multidrug resistant? *Cell Cyc.* 6(22):2782-2787.

60. Rottenberg S*, et al.* (2007) Selective induction of chemotherapy resistance of mammary tumors in a conditional mouse model for hereditary breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* 104(29):12117-12122.

61. Pajic M*, et al.* (2009) Moderate increase in Mdr1a/1b expression causes in vivo resistance to doxorubicin in a mouse model for hereditary breast cancer. *Cancer Res.* 69(16):6396-6404.

62. Leith CP*, et al.* (1999) Frequency and clinical significance of the expression of the multidrug resistance proteins MDR1/P-glycoprotein, MRP1, and LRP in acute myeloid leukemia: a Southwest Oncology Group Study. *Blood* 94(3):1086-1099.

63. List AF*, et al.* (2001) Benefit of cyclosporine modulation of drug resistance in patients with poor-risk acute myeloid leukemia: a Southwest Oncology Group study. *Blood* 98(12):3212-3220.

64. Beattie JV & Petruska PJ (2002) Cyclosporine modulation in poor-risk acute myeloid leukemia. *Blood* 100(1):366; author reply 366-367.

65. Shaffer BC*, et al.* (2012) Drug resistance: still a daunting challenge to the successful treatment of AML. *Drug Resist. Upd.* 15(1-2):62-69.

66. Patel C*, et al.* (2013) Multidrug resistance in relapsed acute myeloid leukemia: evidence of biological heterogeneity. *Cancer* 119(16):3076-3083.

67. Fojo AT*, et al.* (1987) Expression of a multidrug-resistance gene in human tumors and tissues. *Proc. Natl. Acad. Sci. U.S.A.* 84(1):265-269.

68. Goldstein LJ*, et al.* (1989) Expression of a multidrug resistance gene in human cancers. *J Natl. Cancer Inst.* 81(2):116-124.

69. Kapoor K, Sim H, & Ambudkar SV (2013) *Molecular Mechanisms of Tumor Cell Resistance to Chemotherapy* (Spinger).

70. Hijazi YM*, et al.* (1994) Immunohistochemical detection of P-glycoprotein in Ewing's sarcoma and peripheral primitive neuroectodermal tumors before and after chemotherapy. *Am. J. Clin. Pharmacol.* 102(1):61-67.

71. Callaghan R, Crowley E, Potter S, & Kerr ID (2008) P-glycoprotein: so many ways to turn it on. *J. Clin. Pharmacol.* 48(3):365-378.

72. Pisco AO*, et al.* (2013) Non-Darwinian dynamics in therapy-induced cancer drug resistance. *Nat. Comm.* 4:2467.

73. Baker EK, Johnstone RW, Zalcberg JR, & El-Osta A (2005) Epigenetic changes to the MDR1 locus in response to chemotherapeutic drugs. *Oncogene* 24(54):8061-8075.

74. Dean M, Fojo T, & Bates S (2005) Tumour stem cells and drug resistance. *Nat. Rev. Cancer* 5(4):275-284.

75. Zhou BB*, et al.* (2009) Tumour-initiating cells: challenges and opportunities for anticancer drug discovery. *Nature Rev. Drug Disc.* 8(10):806-823.

76. Reya T, Morrison SJ, Clarke MF, & Weissman IL (2001) Stem cells, cancer, and cancer stem cells. *Nature* 414(6859):105-111.

77. Bonnet D & Dick JE (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* 3(7):730-737.

78. Lapidot T*, et al.* (1994) A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* 367(6464):645-648.

79. Singh SK*, et al.* (2004) Identification of human brain tumour initiating cells. *Nature* 432(7015):396-401.

80. Singh SK*, et al.* (2003) Identification of a cancer stem cell in human brain tumors. *Cancer Res.* 63(18):5821-5828.

81. Hirschmann-Jax C*, et al.* (2004) A distinct "side population" of cells with high drug efflux capacity in human tumor cells. *Proc. Natl. Acad. Sci. U.S.A.* 101(39):14228-14233.

82. Robinson SN, Seina SM, Gohr JC, Kuszynski CA, & Sharp JG (2005) Evidence for a qualitative hierarchy within the Hoechst-33342 'side population' (SP) of murine bone marrow cells. *Bone Marrow Transplant.* 35(8):807-818.

83. Anonymous (2002) ATP-Binding Cassette Transporters. ed Committee HGN.

84. Dean M (2002) The Human ATP-Binding Cassette (ABC) Transporter Superfamily. ed Information NCfB (Bethesda, MD).

85. Martignoni M, Groothuis GM, & de Kanter R (2006) Species differences between mouse, rat, dog, monkey and human CYP-mediated drug metabolism, inhibition and induction. *Expert Opin. Drug Met. Toxicol.* 2(6):875-894.

86. Kim IW, Booth-Genthe C, & Ambudkar SV (2008) Relationship between drugs and functional activity of various mammalian P-glycoproteins (ABCB1). *Mini Rev. Med. Chem.* 8(3):193-200.

87. Xia CQ*, et al.* (2006) Comparison of species differences of P-glycoproteins in beagle dog, rhesus monkey, and human using Atpase activity assays. *Mol. Pharm.* 3(1):78-86.

88. Takeuchi T*, et al.* (2006) Establishment and characterization of the transformants stably-expressing MDR1 derived from various animal species in LLC-PK1. *Pharm. Res.* 23(7):1460-1472.

89. Misfeldt DS & Sanders MJ (1981) Transepithelial transport in cell culture: D-glucose transport by a pig kidney cell line (LLC-PK1). *J. Memb. Biol.* 59(1):13-18.

90. Tang-Wai DF*, et al.* (1995) Human (MDR1) and mouse (mdr1, mdr3) P-glycoproteins can be distinguished by their respective drug resistance profiles and sensitivity to modulators. *Biochemistry* 34(1):32-39.

91. Syvanen S*, et al.* (2009) Species differences in blood-brain barrier transport of three positron emission tomography radioligands with emphasis on P-glycoprotein transport. *Drug Metab. Dispos.* 37(3):635-643.

92. Fox E & Bates SE (2007) Tariquidar (XR9576): a P-glycoprotein drug efflux pump inhibitor. *Exp. Rev. Anticancer Ther.* 7(4):447-459.

93. Liow JS*, et al.* (2007) Effect of a P-glycoprotein inhibitor, Cyclosporin A, on the disposition in rodent brain and blood of the 5-HT1A receptor radioligand, [11C](R)-(-)-RWAY. *Synapse* 61(2):96-105.

94. Kuntner C*, et al.* (2010) Dose-response assessment of tariquidar and elacridar and regional quantification of P-glycoprotein inhibition at the rat blood-brain barrier using (R)-[(11)C]verapamil PET. *Eur. J. Nuc. Med. Mol.* 37(5):942-953.

95. Kemper EM, Cleypool C, Boogerd W, Beijnen JH, & van Tellingen O (2004) The influence of the P-glycoprotein inhibitor zosuquidar trihydrochloride (LY335979) on the brain penetration of paclitaxel in mice. *Cancer Chemother. Pharmacol.* 53(2):173-178.

96. Ambudkar SV*, et al.* (1999) Biochemical, cellular, and pharmacological aspects of the multidrug transporter. *Annu. Rev. Pharmacol. Toxicol.* 39:361-398.

97. Hunter J, Jepson MA, Tsuruo T, Simmons NL, & Hirst BH (1993) Functional expression of P-glycoprotein in apical membranes of human intestinal Caco-2 cells. Kinetics of vinblastine secretion and interaction with modulators. *J. Biol. Chem.* 268(20):14991-14997.

98. Polli JW*, et al.* (2001) Rational use of in vitro P-glycoprotein assays in drug discovery. *J. Pharmacol. Exp. Ther.* 299(2):620-628.

99. Feng B*, et al.* (2008) In vitro P-glycoprotein assays to predict the in vivo interactions of P-glycoprotein with drugs in the central nervous system. *Drug Metab. Dispos.* 36(2):268-275.

100. Greenberger LM (1998) Identification of drug interaction sites in P-glycoprotein. *Meth. Enzymol.* 292:307-317.

101. Dey S, Ramachandra M, Pastan I, Gottesman MM, & Ambudkar SV (1998) Photoaffinity labeling of human P-glycoprotein: effect of modulator interaction and ATP hydrolysis on substrate binding. *Meth. Enzymol.* 292:318-328.

102. Kapoor K, Sim H, & Ambudkar S (2013) Multidrug Resistance in Cancer: A Tale of ABC Drug Transporters. *Molecular Mechanisms of Tumor Cell Resistance to Chemotherapy,* Resistance to Targeted Anti-Cancer Therapeutics, ed Bonavida B (Springer New York), Vol 1, pp 1-34.

103. Aller SG*, et al.* (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323(5922):1718-1722.

104. Eckford PD & Sharom FJ (2006) P-glycoprotein (ABCB1) interacts directly with lipid-based anti-cancer drugs and platelet-activating factors. *Biochem. Cell Biol.* 84(6):1022-1033.

105. Higgins CF & Gottesman MM (1992) Is the multidrug transporter a flippase? *Trends Biochem. Sci.* 17(1):18-21.

106. Romsicki Y & Sharom FJ (2001) Phospholipid flippase activity of the reconstituted P-glycoprotein multidrug transporter. *Biochemistry* 40(23):6937-6947.

107. Sharom FJ (2014) Complex Interplay between the P-Glycoprotein Multidrug Efflux Pump and the Membrane: Its Role in Modulating Protein Function. *Front. Oncol.* 4:41.

108. van Wonderen JH*, et al.* (2014) The central cavity of ABCB1 undergoes alternating access during ATP hydrolysis. *FEBS J.* 281(9):2190-2201.

109. Martin C, Higgins CF, & Callaghan R (2001) The vinblastine binding site adopts high- and low-affinity conformations during a transport cycle of P-glycoprotein. *Biochemistry* 40(51):15733-15742.

110. Martin C*, et al.* (2000) Drug binding sites on P-glycoprotein are altered by ATP binding prior to nucleotide hydrolysis. *Biochemistry* 39(39):11901-11906.

111. Callaghan R, Ford RC, & Kerr ID (2006) The translocation mechanism of P-glycoprotein. *FEBS Lett.* 580(4):1056-1063.

112. Sauna ZE & Ambudkar SV (2007) About a switch: how P-glycoprotein (ABCB1) harnesses the energy of ATP binding and hydrolysis to do mechanical work. *Mol. Cancer Ther.* 6(1):13-23.

113. Linton KJ & Higgins CF (2007) Structure and function of ABC transporters: the ATP switch provides flexible control. *Pflugers Arch.* 453(5):555-567.

114. Locher KP (2009) Review. Structure and mechanism of ATP-binding cassette transporters. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364(1514):239-245.

115. Sato T*, et al.* (2009) Functional role of the linker region in purified human P-glycoprotein. *FEBS J.* 276(13):3504-3516.

116. Lee JY, Urbatsch IL, Senior AE, & Wilkens S (2002) Projection structure of P-glycoprotein by electron microscopy. Evidence for a closed conformation of the nucleotide binding domains. *J. Biol. Chem.* 277(42):40125-40131.

117. Lee JY, Urbatsch IL, Senior AE, & Wilkens S (2008) Nucleotide-induced structural changes in P-glycoprotein observed by electron microscopy. *J. Biol. Chem.* 283(9):5769-5779.

118. Rosenberg MF, Callaghan R, Modok S, Higgins CF, & Ford RC (2005) Three-dimensional structure of P-glycoprotein: the transmembrane regions adopt an asymmetric configuration in the nucleotide-bound state. *J. Biol. Chem.* 280(4):2857-2862.

119. Rosenberg MF, Kamis AB, Callaghan R, Higgins CF, & Ford RC (2003) Three-dimensional structures of the mammalian multidrug resistance P-glycoprotein demonstrate major conformational changes in the transmembrane domains upon nucleotide binding. *J. Biol. Chem.* 278(10):8294-8299.

120. Rosenberg MF*, et al.* (2001) Repacking of the transmembrane domains of P-glycoprotein during the transport ATPase cycle. *EMBO J.* 20(20):5615-5625.

121. Bartesaghi A*, et al.* (2015) 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. *Science*.

122. Kim J*, et al.* (2015) Subnanometre-resolution electron cryomicroscopy structure of a heterodimeric ABC exporter. *Nature* 517(7534):396-400.

123. Moeller A*, et al.* (2015) Distinct conformational spectrum of homologous multidrug ABC transporters. *Structure* 23(3):450-460.

124. Kuhlbrandt W (2014) Cryo-EM enters a new era. *eLife* 3:e03678.

125. Jin MS, Oldham ML, Zhang Q, & Chen J (2012) Crystal structure of the multidrug transporter P-glycoprotein from Caenorhabditis elegans. *Nature* 490(7421):566-569.

126. Shintre CA*, et al.* (2013) Structures of ABCB10, a human ATP-binding cassette transporter in apo- and nucleotide-bound states. *Proc. Natl. Acad. Sci. U.S.A.* 110(24):9710-9715.

127. Zolnerciks JK, Wooding C, & Linton KJ (2007) Evidence for a Sav1866-like architecture for the human multidrug transporter P-glycoprotein. *FASEB J.* 21(14):3937-3948.

128. Dawson RJ & Locher KP (2007) Structure of the multidrug ABC transporter Sav1866 from Staphylococcus aureus in complex with AMP-PNP. *FEBS Lett.* 581(5):935-938.

129. Reyes CL & Chang G (2005) Structure of the ABC transporter MsbA in complex with ADP.vanadate and lipopolysaccharide. *Science* 308(5724):1028-1031.

130. Chang G & Roth CB (2001) Structure of MsbA from E. coli: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* 293(5536):1793-1800.

131. Tao H*, et al.* (2011) Design and synthesis of Selenazole-containing peptides for cocrystallization with P-glycoprotein. *Chembiochem* 12(6):868-873.

132. Shapiro AB & Ling V (1994) ATPase activity of purified and reconstituted P-glycoprotein from Chinese hamster ovary cells. *J. Biol. Chem.* 269(5):3745-3754.

133. Loo TW, Bartlett MC, & Clarke DM (2010) Human P-glycoprotein is active when the two halves are clamped together in the closed conformation. *Biochem. Biophys. Res. Comm.* 395(3):436-440.

134. O'Mara ML & Mark AE (2012) Effect of environment on membrane protein structure: P-glycoprotein under physiological conditions. *J. Chem. Theory Comp.* 8:3964-3976.

135. Wen PC, Verhalen B, Wilkens S, McHaourab HS, & Tajkhorshid E (2013) On the origin of large flexibility of P-glycoprotein in the inward-facing state. *J. Biol. Chem.* 288(26):19211-19220.

136. Ward AB*, et al.* (2013) Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain. *Proc. Natl. Acad. Sci. U.S.A.* 110(33):13386-13391.

137. Chufan EE, Sim HM, & Ambudkar SV (2015) Molecular Basis of the Polyspecificity of P-Glycoprotein (ABCB1): Recent Biochemical and Structural Studies. *Adv. Cancer Res.* 125:71-96.

138. Hyde BB*, et al.* (2012) The mitochondrial transporter ABC-me (ABCB10), a downstream target of GATA-1, is essential for erythropoiesis in vivo. *Cell Death Differ.* 19(7):1117-1126.

139. Liesa M, Qiu W, & Shirihai OS (2012) Mitochondrial ABC transporters function: the role of ABCB10 (ABC-me) as a novel player in cellular handling of reactive oxygen species. *Biochim. Biophys. Acta* 1823(10):1945-1957.

140. Junge F*, et al.* (2008) Large-scale production of functional membrane proteins. *Cell Mol. Life Sci.* 65(11):1729-1755.

141. Carpenter EP, Beis K, Cameron AD, & Iwata S (2008) Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.* 18(5):581-586.

142. Lundstrom K (2006) Structural genomics for membrane proteins. *Cell Mol. Life Sci.* 63(22):2597-2607.

143. Gutmann DA*, et al.* (2007) A high-throughput method for membrane protein solubility screening: the ultracentrifugation dispersity sedimentation assay. *Protein Sci.* 16(7):1422-1428.

144. Locher KP, Lee AT, & Rees DC (2002) The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* 296(5570):1091-1098.

145. Chang G, Spencer RH, Lee AT, Barclay MT, & Rees DC (1998) Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel. *Science* 282(5397):2220-2226.

146. Bibi E, Gros P, & Kaback HR (1993) Functional expression of mouse mdr1 in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* 90(19):9209-9213.

147. Bibi E, Edgar R, & Beja O (1998) Functional expression of mdr and mdr-like cDNAs in Escherichia coli. *Meth. Enzymol.* 292:370-382.

148. Evans GL*, et al.* (1995) Heterologous expression systems for P-glycoprotein: E. coli, yeast, and baculovirus. *J. Bioen. Biomemb.* 27(1):43-52.

149. Ogata K, Nishikawa H, & Ohsugi M (1969) A yeast capable of utilizing methanol. *Ag. Biol. Chem.* 33:1519-1590.

150. Cereghino JL & Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol. Rev.* 24(1):45-66.

151. Cregg JM, Cereghino JL, Shi J, & Higgins DR (2000) Recombinant protein expression in Pichia pastoris. *Mol. Biotechnol.* 16(1):23-52.

152. Roggenkamp R, Janowicz Z, Stanikowski B, & Hollenberg CP (1984) Biosynthesis and regulation of the peroxisomal methanol oxidase from the methylotrophic yeast Hansenula polymorpha. *Mol. Gen. Genet.* 194(3):489-493.

153. Urbatsch IL, Wilke-Mounts S, Gimi K, & Senior AE (2001) Purification and characterization of N-glycosylation mutant mouse and human P-glycoproteins expressed in Pichia pastoris cells. *Arch. Biochem. Biophys.* 388(1):171-177.

154. Beaudet L, Urbatsch IL, & Gros P (1998) High-level expression of mouse Mdr3 P-glycoprotein in yeast Pichia pastoris and characterization of ATPase activity. *Meth. Enzymol.* 292:397-413.

155. Fung KL & Gottesman MM (2009) A synonymous polymorphism in a common MDR1 (ABCB1) haplotype shapes protein function. *Biochim. Biophys. Acta* 1794(5):860-871.

156. Kimchi-Sarfaty C*, et al.* (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 315(5811):525-528.

157. Hershberg R & Petrov DA (2008) Selection on codon bias. *Ann. Rev. Genet.* 42:287-299.

158. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181(4096):223-230.

159. Malowany JI & Butany J (2012) Pathology of sickle cell disease. *Sem. Diag. Pathol.* 29(1):49-55.

160. Plotkin JB & Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12(1):32-42.

161. Gustafsson C, Govindarajan S, & Minshull J (2004) Codon bias and heterologous protein expression. *Trends Biotechnol.* 22(7):346-353.

162. Ikemura T (1981) Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* 146(1):1-21.

163. Novoa EM & Ribas de Pouplana L (2012) Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* 28(11):574-581.

164. Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Curr. Opin. Biotechnol.* 6(5):494-500.

165. Hu H*, et al.* (2013) Codon optimization significantly improves the expression level of a keratinase gene in Pichia pastoris. *PLoS One* 8(3):e58393.

166. Bai J*, et al.* (2011) A gene optimization strategy that enhances production of fully functional P-glycoprotein in Pichia pastoris. *PLoS One* 6(8):e22577.

167. Yadava A & Ockenhouse CF (2003) Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. *Infection Immun.* 71(9):4961-4969.

168. Sinclair G & Choy FY (2002) Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, Pichia pastoris. *Protein Exp. Pur.* 26(1):96-105.

169. Huang Y*, et al.* (2014) Expression of codon optimized human bone morphogenetic protein 4 in Pichia pastoris. *Biotechnol. App. Biochem.*61(2):175-183.

170. Yu P, Yan Y, Gu Q, & Wang X (2013) Codon optimisation improves the expression of Trichoderma viride sp. endochitinase in Pichia pastoris. *Sci. Rep.* 3:3043.

171. Jarvis DL (2009) Baculovirus-insect cell expression systems. *Meth. Enzymol.* 463:191-222.

172. Pennock GD, Shoemaker C, & Miller LK (1984) Strong and regulated expression of Escherichia coli beta-galactosidase in insect cells with a baculovirus vector. *Mol. Cell Biol.* 4(3):399-406.

173. Smith GE, Summers MD, & Fraser MJ (1983) Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell Biol.* 3(12):2156-2165.

174. Ayres MD, Howard SC, Kuzio J, Lopez-Ferber M, & Possee RD (1994) The complete DNA sequence of Autographa californica nuclear polyhedrosis virus. *Virology* 202(2):586-605.

175. Kost TA & Condreay JP (2002) Recombinant baculoviruses as mammalian cell gene-delivery vectors. *Trends Biotechnol.* 20(4):173-180.

176. Vaughn JL, Goodwin RH, Tompkins GJ, & McCawley P (1977) The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In vitro* 13(4):213-217.

177. Wickham TJ, Davis T, Granados RR, Shuler ML, & Wood HA (1992) Screening of insect cell lines for the production of recombinant proteins and infectious virus in the baculovirus expression system. *Biotechnol. Prog.* 8(5):391-396.

178. Germann UA (1998) Baculovirus-mediated expression of human multidrug resistance cDNA in insect cells and functional analysis of recombinant P-glycoprotein. *Meth. Enzymol.* 292:427-441.

179. Boyce FM & Bucher NL (1996) Baculovirus-mediated gene transfer into mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 93(6):2348-2352.

180. Hofmann C*, et al.* (1995) Efficient gene transfer into human hepatocytes by baculovirus vectors. *Proc. Natl. Acad. Sci. U.S.A.* 92(22):10099-10103.

181. Moss B & Earl PL (2002) Overview of the vaccinia virus expression system. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* Chapter 16:Unit16 15.

182. Ramachandra M, Gottesman MM, & Pastan I (1998) Recombinant vaccinia virus vectors for functional expression of P-glycoprotein in mammalian cells. *Meth. Enzymol.* 292:441-455.

183. Kapoor K, Bhatnagar J, Chufan EE, & Ambudkar SV (2013) Mutations in intracellular loops 1 and 3 lead to misfolding of human P-glycoprotein (ABCB1) that can be rescued by cyclosporine A, which reduces its association with chaperone Hsp70. *J. Biol. Chem.* 288(45):32622-32636.

184. Shukla S, Schwartz C, Kapoor K, Kouanda A, & Ambudkar SV (2012) Use of baculovirus BacMam vectors for expression of ABC drug transporters in mammalian cells. *Drug. Metab. Dispos.* 40(2):304-312.

185. Seddon AM, Curnow P, & Booth PJ (2004) Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta* 1666(1-2):105-117.

186. Prive GG (2007) Detergents for the stabilization and crystallization of membrane proteins. *Methods* 41(4):388-397.

187. Yernool D, Boudker O, Jin Y, & Gouaux E (2004) Structure of a glutamate transporter homologue from Pyrococcus horikoshii. *Nature* 431(7010):811-818.

188. Bhatt FH & Jeffery CJ (2010) Expression, detergent solubilization, and purification of a membrane transporter, the MexB multidrug resistance protein. *J. Vis. Exp.* (46).

189. Pollock N, Cant N, Rimington T, & Ford RC (2014) Purification of the cystic fibrosis transmembrane conductance regulator protein expressed in Saccharomyces cerevisiae. *J. Vis. Exp.* (87).

190. Newby ZE*, et al.* (2009) A general protocol for the crystallization of membrane proteins for X-ray structural investigation. *Nat. Protoc.* 4(5):619-637.

191. Loughran S & Walls D (2011) Purification of Poly-Histidine-Tagged Proteins. *Protein Chromatography,* Methods in Molecular Biology, eds Walls D & Loughran ST (Humana Press), Vol 681, pp 311-335.

192. Madadlou A, O'Sullivan S, & Sheehan D (2011) Fast Protein Liquid Chromatography. *Protein Chromatography,* Methods in Molecular Biology, eds Walls D & Loughran ST (Humana Press), Vol 681, pp 439-447.

193. Almen MS, Nordstrom KJ, Fredriksson R, & Schioth HB (2009) Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* 7:50.

194. Serrano-Vega MJ, Magnani F, Shibata Y, & Tate CG (2008) Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. U.S.A.* 105(3):877-882.

195. Warne T*, et al.* (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* 454(7203):486-491.

196. Mancia F & Love J (2010) High-throughput expression and purification of membrane proteins. *J. Struc. Biol.* 172(1):85-93.

197. Lewinson O, Lee AT, & Rees DC (2008) The funnel approach to the precrystallization production of membrane proteins. *J. Mol. Biol* 377(1):62-73.

198. Wiener MC (2004) A pedestrian guide to membrane protein crystallization. *Methods* 34(3):364-372.

199. Pollock N, Moran O, Baroni D, Zegarra-Moran O, & Ford RC (2014) Characterisation of the salmon cystic fibrosis transmembrane conductance regulator protein for structural studies. *AIMS Mol. Sci.* 1(4):141-161.

200. Seetharaman SV, Taylor AB, Holloway S, & Hart PJ (2010) Structures of mouse SOD1 and human/mouse SOD1 chimeras. *Arch. Biochem. Biophys.* 503(2):183-190.

201. Heaslet H*, et al.* (2007) Crystal structure of an FIV/HIV chimeric protease complexed with the broad-based inhibitor, TL-3. *Retrovirology* 4:1.

202. Chen JC*, et al.* (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl. Acad. Sci. U.S.A.* 97(15):8233-8238.

203. Nishida M, Cadene M, Chait BT, & MacKinnon R (2007) Crystal structure of a Kir3.1-prokaryotic Kir channel chimera. *EMBO J.* 26(17):4005-4015.

204. Stanfield PR, Nakajima S, & Nakajima Y (2002) Constitutively active and G-protein coupled inward rectifier K+ channels: Kir2.0 and Kir3.0. *Rev. Physiol. Biochem. Pharmacol.* 145:47-179.

205. Kuo A*, et al.* (2003) Crystal structure of the potassium channel KirBac1.1 in the closed state. *Science* 300(5627):1922-1926.

206. Dhir R & Gros P (1992) Functional analysis of chimeric proteins constructed by exchanging homologous domains of two P-glycoproteins conferring distinct drug resistance profiles. *Biochemistry* 31(26):6103-6110.

207. Bhatia A, Schafer HJ, & Hrycyna CA (2005) Oligomerization of the human ABC transporter ABCG2: evaluation of the native protein and chimeric dimers. *Biochemistry* 44(32):10893-10904.

208. Chen M, Abele R, & Tampe R (2004) Functional non-equivalence of ATP-binding cassette signature motifs in the transporter associated with antigen processing (TAP). *J. Biol. Chem.* 279(44):46073-46081.

209. Arora S, Lapinski PE, & Raghavan M (2001) Use of chimeric proteins to investigate the role of transporter associated with antigen processing (TAP) structural domains in peptide binding and translocation. *Proc. Natl. Acad. Sci. U.S.A.* 98(13):7241-7246.

210. Geillon F*, et al.* (2014) Structure-function analysis of peroxisomal ATP-binding cassette transporters using chimeric dimers. *J. Biol. Chem.* 289(35):24511-24520.

211. Saini P, Gaur NA, & Prasad R (2006) Chimeras of the ABC drug transporter Cdr1p reveal functional indispensability of transmembrane domains and nucleotide-binding domains, but transmembrane segment 12 is replaceable with the corresponding homologous region of the non-drug transporter Cdr3p. *Microbiology* 152(Pt 5):1559-1573.

212. Buschman E & Gros P (1991) Functional analysis of chimeric genes obtained by exchanging homologous domains of the mouse mdr1 and mdr2 genes. *Mol. Cell Biol.* 11(2):595-603.

213. Zhou Y, Gottesman MM, & Pastan I (1999) Domain exchangeability between the multidrug transporter (MDR1) and phosphatidylcholine flippase (MDR2). *Mol. Pharmacol.* 56(5):997-1004.

214. Zhou Y, Gottesman MM, & Pastan I (1999) Studies of human MDR1-MDR2 chimeras demonstrate the functional exchangeability of a major transmembrane segment of the multidrug transporter and phosphatidylcholine flippase. *Mol. Cell Biol.* 19(2):1450-1459.

215. Ueda K, Cardarelli C, Gottesman MM, & Pastan I (1987) Expression of a full-length cDNA for the human "MDR1" gene confers resistance to colchicine, doxorubicin, and vinblastine. *Proc. Natl. Acad. Sci. U.S.A.* 84(9):3004-3008.

216. Gros P, Croop J, & Housman D (1986) Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. *Cell* 47(3):371-380.

217. Akiyama S, Fojo A, Hanover JA, Pastan I, & Gottesman MM (1985) Isolation and genetic characterization of human KB cell lines resistant to multiple drugs. *Som. Cell Mol. Genet.* 11(2):117-126.

218. Currier SJ, Ueda K, Willingham MC, Pastan I, & Gottesman MM (1989) Deletion and insertion mutants of the multidrug transporter. *J. Biol. Chem.* 264(24):14376-14381.

219. Ramachandra M*, et al.* (1998) Human P-glycoprotein exhibits reduced affinity for substrates during a catalytic transition state. *Biochemistry* 37(14):5010-5019.

220. Zoonens M & Miroux B (2010) Expression of membrane proteins at the Escherichia coli membrane for structural studies. *Meth. Mol. Biol.* 601:49-66.

221. Schlegel S*, et al.* (2010) Revolutionizing membrane protein overexpression in bacteria. *Microb. Biotechnol.* 3(4):403-411.

222. Montigny C, Penin F, Lethias C, & Falson P (2004) Overcoming the toxicity of membrane peptide expression in bacteria by upstream insertion of Asp-Pro sequence. *Biochim. Biophys. Acta* 1660(1-2):53-65.

223. Gagniuc P & Ionescu-Tirgoviste C (2012) Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. *BMC Genomics* 13:512.

224. Gruber TM & Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Ann. Rev. Microbiol.* 57:441-466.

225. Harley CB & Reynolds RP (1987) Analysis of E. coli promoter sequences. *Nucleic Acids Res.* 15(5):2343-2361.

226. Lisser S & Margalit H (1993) Compilation of E. coli mRNA promoter sequences. *Nucleic Acids Res.* 21(7):1507-1516.

227. Hawley DK & McClure WR (1983) Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Res.* 11(8):2237-2255.

228. Li D, Aaskov J, & Lott WB (2011) Identification of a cryptic prokaryotic promoter within the cDNA encoding the 5' end of dengue virus RNA genome. *PLoS One* 6(3):e18197.

229. Gregory RJ*, et al.* (1990) Expression and characterization of the cystic fibrosis transmembrane conductance regulator. *Nature* 347(6291):382-386.

230. Quinton PM (1999) Physiological basis of cystic fibrosis: a historical perspective. *Physiol. Rev.* 79(1 Suppl):S3-S22.

231. Riordan JR*, et al.* (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245(4922):1066-1073.

232. Boyd AC*, et al.* (1999) Insertion of natural intron 6a-6b into a human cDNA-derived gene therapy vector for cystic fibrosis improves plasmid stability and permits facile RNA/DNA discrimination. *J. Gene Med.* 1(5):312-321.

233. Wei L*, et al.* (2000) Suppressive interactions between mutations located in the two nucleotide binding domains of CFTR. *FEBS Lett.* 473(2):149-153.

234. Cheng SH*, et al.* (1990) Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* 63(4):827-834.

235. Ueda K*, et al.* (1987) The human multidrug resistance (mdr1) gene. cDNA cloning and transcription initiation. *J. Biol. Chem.* 262(2):505-508.

236. Beja O & Bibi E (1996) Functional expression of mouse Mdr1 in an outer membrane permeability mutant of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* 93(12):5969-5974.

237. Devault A & Gros P (1990) Two members of the mouse mdr gene family confer multidrug resistance with overlapping but distinct drug specificities. *Mol. Cell Biol.* 10(4):1652-1663.

238. Shultzaberger RK, Chen Z, Lewis KA, & Schneider TD (2007) Anatomy of Escherichia coli sigma70 promoters. *Nucleic Acids Res.* 35(3):771-788.

239. Shine J & Dalgarno L (1974) The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U.S.A.* 71(4):1342-1346.

240. Blattner FR*, et al.* (1997) The complete genome sequence of Escherichia coli K-12. *Science* 277(5331):1453-1462.

241. Missiakas D, Georgopoulos C, & Raina S (1993) The Escherichia coli heat shock gene htpY: mutational analysis, cloning, sequencing, and transcriptional regulation. *J. Bacteriol.* 175(9):2613-2624.

242. van Den Elsen JM, Kuntz DA, Hoedemaeker FJ, & Rose DR (1999) Antibody C219 recognizes an alpha-helical epitope on P-glycoprotein. *Proc. Natl. Acad. Sci. U.S.A.* 96(24):13679-13684.

243. Kozak M (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* 266(30):19867-19870.

244. Hartley JL, Temple GF, & Brasch MA (2000) DNA cloning using in vitro site-specific recombination. *Genome Res.* 10(11):1788-1795.

245. Landy A (1989) Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Ann. Rev. Biochem.* 58:913-949.

246. Groth AC & Calos MP (2004) Phage integrases: biology and applications. *J. Mol. Biol.* 335(3):667-678.

247. Durbin SD & Feher G (1996) Protein crystallization. *Annu. Rev. Phys. Chem.* 47:171-204.

248. Rhodes G (2006) *Crystallography made crystal clear : a guide for users of macromolecular models* (Elsevier/Academic Press, Amsterdam ; Boston) 3rd Ed pp xxv, 306 p.

249. Gribar JJ, Ramachandra M, Hrycyna CA, Dey S, & Ambudkar SV (2000) Functional characterization of glycosylation-deficient human P-glycoprotein using a vaccinia virus expression system. *J. Memb. Biol.* 173(3):203-214.

250. Hunte C, Koepke J, Lange C, Rossmanith T, & Michel H (2000) Structure at 2.3 A resolution of the cytochrome bc(1) complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment. *Structure* 8(6):669-684.

251. Kwong PD*, et al.* (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393(6686):648-659.

252. Chen GQ, Sun Y, Jin R, & Gouaux E (1998) Probing the ligand binding domain of the GluR2 receptor by proteolysis and deletion mutagenesis defines domain boundaries and yields a crystallizable construct. *Protein Sci.* 7(12):2623-2630.

253. Hung LW*, et al.* (1998) Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature* 396(6712):703-707.

254. Yuan YR*, et al.* (2001) The crystal structure of the MJ0796 ATP-binding cassette. Implications for the structural consequences of ATP hydrolysis in the active site of an ABC transporter. *J. Biol. Chem.* 276(34):32313-32321.

255. Schmitt L, Benabdelhak H, Blight MA, Holland IB, & Stubbs MT (2003) Crystal structure of the nucleotide-binding domain of the ABC-transporter haemolysin B: identification of a variable region within ABC helical domains. *J. Mol. Biol.* 330(2):333-342.

256. Gaudet R & Wiley DC (2001) Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. *EMBO J.* 20(17):4964-4972.

257. Karpowich N*, et al.* (2001) Crystal structures of the MJ1267 ATP binding cassette reveal an induced-fit effect at the ATPase active site of an ABC transporter. *Structure* 9(7):571-586.

258. Hopfner KP*, et al.* (2000) Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* 101(7):789-800.

259. Diederichs K*, et al.* (2000) Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon Thermococcus litoralis. *EMBO J.* 19(22):5951-5961.

260. Sauna ZE, Nandigama K, & Ambudkar SV (2006) Exploiting reaction intermediates of the ATPase reaction to elucidate the mechanism of transport by P-glycoprotein (ABCB1). *J. Biol. Chem.* 281(36):26501-26511.

261. Sauna ZE, Muller M, Peng XH, & Ambudkar SV (2002) Importance of the conserved Walker B glutamate residues, 556 and 1201, for the completion of the catalytic cycle of ATP hydrolysis by human P-glycoprotein (ABCB1). *Biochemistry* 41(47):13989-14000.

262. Loo TW & Clarke DM (1999) Identification of residues in the drug-binding domain of human P-glycoprotein. Analysis of transmembrane segment 11 by cysteine-scanning mutagenesis and inhibition by dibromobimane. *J. Biol. Chem.* 274(50):35388-35392.

263. Wang EJ, Casciano CN, Clement RP, & Johnson WW (2000) In vitro flow cytometry method to quantitatively assess inhibitors of P-glycoprotein. *Drug Metab. Dispos.* 28(5):522-528.

264. Strouse JJ*, et al.* (2013) Fluorescent substrates for flow cytometric evaluation of efflux inhibition in ABCB1, ABCC1, and ABCG2 transporters. *Anal. Biochem.* 437(1):77-87.

265. Garcia-Escarp M*, et al.* (2004) Flow cytometry-based approach to ABCG2 function suggests that the transporter differentially handles the influx and efflux of drugs. *Cytometry* 62(2):129-138.

266. Legrand O, Perrot JY, Simonin G, Baudard M, & Marie JP (2001) JC-1: a very sensitive fluorescent probe to test Pgp activity in adult acute myeloid leukemia. *Blood* 97(2):502-508.

267. Kuhnel JM, Perrot JY, Faussat AM, Marie JP, & Schwaller MA (1997) Functional assay of multidrug resistant cells using JC-1, a carbocyanine fluorescent probe. *Leukemia* 11(7):1147-1155.

268. Ambudkar SV (1998) Drug-stimulatable ATPase activity in crude membranes of human MDR1-transfected mammalian cells. *Meth. Enzymol.* 292:504-514.

269. Sarkadi B, Price EM, Boucher RC, Germann UA, & Scarborough GA (1992) Expression of the human multidrug resistance cDNA in insect cells generates a high activity drug-stimulated membrane ATPase. *J. Biol. Chem.* 267(7):4854-4858.

270. Calcagno AM, Kim IW, Wu CP, Shukla S, & Ambudkar SV (2007) ABC drug transporters as molecular targets for the prevention of multidrug resistance and drug-drug interactions. *Curr. Drug Deliv.* 4(4):324-333.

271. Hrycyna CA*, et al.* (1998) Mechanism of action of human P-glycoprotein ATPase activity. Photochemical cleavage during a catalytic transition state using orthovanadate reveals cross-talk between the two ATP sites. *J. Biol. Chem.* 273(27):16631-16634.

272. Loo TW & Clarke DM (2002) Vanadate trapping of nucleotide at the ATP-binding sites of human multidrug resistance P-glycoprotein exposes different residues to the drug-binding site. *Proc. Natl. Acad. Sci. U.S.A.* 99(6):3511-3516.

273. Dukkipati A, Park HH, Waghray D, Fischer S, & Garcia KC (2008) BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. *Protein Expr. Pur* 62(2):160-170.

274. Condreay JP, Witherspoon SM, Clay WC, & Kost TA (1999) Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proc. Natl. Acad. Sci. U.S.A.* 96(1):127-132.

275. Van Wijk R, Tichonicky L, & Kruh J (1981) Effect of sodium butyrate on the hepatoma cell cycle: possible use for cell synchronization. *In vitro* 17(10):859-862.

276. Candido EP, Reeves R, & Davie JR (1978) Sodium butyrate inhibits histone deacetylation in cultured cells. *Cell* 14(1):105-113.

277. Fornwald JA, Lu Q, Wang D, & Ames RS (2007) Gene expression in mammalian cells using BacMam, a modified baculovirus system. *Meth. Mol. Biol.* 388:95-114.

278. Hall MD*, et al.* (2011) Synthesis and structure-activity evaluation of isatin-beta-thiosemicarbazones with improved selective activity toward multidrug-resistant cells expressing P-glycoprotein. *J. Med. Chem.* 54(16):5878-5889.

279. Rowe SM, Miller S, & Sorscher EJ (2005) Cystic fibrosis. *N. Engl. J. Med.* 352(19):1992-2001.

280. Loo TW & Clarke DM (1997) Correction of defective protein kinesis of human P-glycoprotein mutants by substrates and modulators. *J. Biol. Chem.* 272(2):709-712.

281. Vasudevan S, Tsuruo T, & Rose DR (1998) Mode of binding of anti-P-glycoprotein antibody MRK-16 to its antigen. A crystallographic and molecular modeling study. *J. Biol. Chem.* 273(39):25413-25419.

282. Martin C*, et al.* (2000) Communication between multiple drug binding sites on P-glycoprotein. *Mol. Pharmacol.* 58(3):624-632.

283. Chaoui D*, et al.* (2006) JC-1, a sensitive probe for a simultaneous detection of P-glycoprotein activity and apoptosis in leukemic cells. *Cytometry* 70(3):189-196.

284. Essodaigui M, Broxterman HJ, & Garnier-Suillerot A (1998) Kinetic analysis of calcein and calcein-acetoxymethylester efflux mediated by the multidrug resistance protein and P-glycoprotein. *Biochemistry* 37(8):2243-2250.

285. Petriz J & Garcia-Lopez J (1997) Flow cytometric analysis of P-glycoprotein function using rhodamine 123. *Leukemia* 11(7):1124-1130.

286. Harvey J, Hardy SC, & Ashford ML (1999) Dual actions of the metabolic inhibitor, sodium azide on K(ATP) channel currents in the rat CRI-G1 insulinoma cell line. *Br. J. Pharmacol.* 126(1):51-60.

287. Pelicano H, Martin DS, Xu RH, & Huang P (2006) Glycolysis inhibition for anticancer treatment. *Oncogene* 25(34):4633-4646.

288. Ishii H, Shirai T, Makino C, & Nishikata T (2014) Mitochondrial inhibitor sodium azide inhibits the reorganization of mitochondria-rich cytoplasm and the establishment of the anteroposterior axis in ascidian embryo. *Devel. Growth Differ.* 56(2):175-188.

289. Safa AR (1998) Photoaffinity labels for characterizing drug interaction sites of P-glycoprotein. *Meth. Enzymol.* 292:289-307.

290. Chufan EE*, et al.* (2013) Multiple transport-active binding sites are available for a single substrate on human P-glycoprotein (ABCB1). *PLoS One* 8(12):e82463.

291. Cutler P (1996) Size-exclusion chromatography. *Meth. Mol. Biol* 59:255-267.

292. Kost TA, Condreay JP, & Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol.* 23(5):567-575.

293. Esposito D*, et al.* (2005) Gateway cloning is compatible with protein secretion from Pichia pastoris. *Protein Expr. Pur.* 40(2):424-428.

294. Cregg JM, Barringer KJ, Hessler AY, & Madden KR (1985) Pichia pastoris as a host system for transformations. *Mol. Cell Biol.* 5(12):3376-3385.

295. Higgins DR & Cregg JM (1998) Introduction to Pichia pastoris. *Meth. Mol. Biol.* 103:1-15.

296. Gleeson MA, White CE, Meininger DP, & Komives EA (1998) Generation of protease-deficient strains and their use in heterologous protein expression. *Meth. Mol. Biol.* 103:81-94.

297. Desbois S, Seabrook SA, & Newman J (2013) Some practical guidelines for UV imaging in the protein crystallization laboratory. *Acta crystallographica* 69(Pt 2):201-208.

298. Ghosh E, Kumari P, Jaiman D, & Shukla AK (2015) Methodological advances: the unsung heroes of the GPCR structural revolution. *Nat. Rev. Mol. Cell Biol.* 16(2):69-81.

299. Rasmussen SG*, et al.* (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 450(7168):383-387.

300. Rosenbaum DM*, et al.* (2007) GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* 318(5854):1266-1273.

Appendix

Appendix Figure 1. Amino acid and nucleotide sequence alignment of wild-type and codon optimized human (*MDR1*) P-gp.

Amino acid residues (blue) and the corresponding codons for wild-type (WT) and codon optimized (CO) MDR1 cDNA are shown.

Nucleotides changed during codon optimization are indicated in red.

		M	D	L	E	G	D	R	N	G	G	A	K	K	K	N	F	F	K	L	N
WT	1	ATG	GAT	CTT	GAA	GGG	GAC	CGC	AAT	GGA	GGA	GCA	AAG	AAG	AAG	AAC	TTT	TTT	AAA	CTG	AAC
CO	1	ATG	GAC	TTG	GAG	GGT	GAC	AGA	AAT	GGT	GGT	GCT	AAG	AAG	AAG	AAC	TTC	TTC	AAG	TTG	AAC
		N	K	S	E	K	D	K	K	E	K	K	P	T	V	S	V	F	S	M	F
WT	61	AAT	AAA	AGT	GAA	AAA	GAT	AAG	AAG	GAA	AAG	AAA	CCA	ACT	GTC	AGT	GTA	TTT	TCA	ATG	TTT
CO	61	AAC	AAG	TCC	GAG	AAG	GAC	AAG	AAA	GAG	AAG	AAG	CCA	ACT	GTT	TCC	GTT	TTC	TCC	ATG	TTC
		R	Y	S	N	W	L	D	K	L	Y	M	V	V	G	T	L	A	A	I	I
WT	121	CGC	TAT	TCA	AAT	TGG	CTT	GAC	AAG	TTG	TAT	ATG	GTG	GTG	GGA	ACT	TTG	GCT	GCC	ATC	ATC
CO	121	AGA	TAC	TCC	AAC	TGG	TTG	GAT	AAG	TTG	TAC	ATG	GTT	GTT	GGT	ACT	TTG	GCT	GCT	ATC	ATC
		H	G	A	G	L	P	L	M	M	L	V	F	G	E	M	T	D	I	F	A
WT	181	CAT	GGG	GCT	GGA	CTT	CCT	CTC	ATG	ATG	CTG	GTG	TTT	GGA	GAA	ATG	ACA	GAT	ATC	TTT	GCA
CO	181	CAC	GGT	GCT	GGT	TTG	CCA	TTG	ATG	ATG	TTG	GTT	TTC	GGT	GAG	ATG	ACT	GAC	ATC	TTC	GCT
		N	A	G	N	L	E	D	L	M	S	N	I	T	N	R	S	D	I	N	D
WT	241	AAT	GCA	GGA	AAT	TTA	GAA	GAT	CTG	ATG	TCA	AAC	ATC	ACT	AAT	AGA	AGT	GAT	ATC	AAT	GAT
CO	241	AAC	GCT	GGT	AAC	TTG	GAG	GAC	TTG	ATG	TCC	AAC	ATC	ACT	AAC	AGA	TCC	GAC	ATC	AAC	GAC
		T	G	F	F	M	N	L	E	E	D	M	T	R	Y	A	Y	Y	Y	S	G
WT	301	ACA	GGG	TTC	TTC	ATG	AAT	CTG	GAG	GAA	GAC	ATG	ACC	AGG	TAT	GCC	TAT	TAT	TAC	AGT	GGA
CO	301	ACT	GGT	TTC	TTC	ATG	AAC	TTG	GAA	GAG	GAC	ATG	ACT	AGA	TAC	GCT	TAC	TAC	TAC	TCC	GGT
		I	G	A	G	V	L	V	A	A	Y	I	Q	V	S	F	W	C	L	A	A
WT	361	ATT	GGT	GCT	GGG	GTG	CTG	GTT	GCT	GCT	TAC	ATT	CAG	GTT	TCA	TTT	TGG	TGC	CTG	GCA	GCT
CO	361	ATT	GGT	GCT	GGT	GTT	TTG	GTT	GCT	GCT	TAC	ATC	CAG	GTT	TCT	TTC	TGG	TGT	TTG	GCT	GCT
		G	R	Q	I	H	K	I	R	K	Q	F	F	H	A	I	M	R	Q	E	I
WT	421	GGA	AGA	CAA	ATA	CAC	AAA	ATT	AGA	AAA	CAG	TTT	TTT	CAT	GCT	ATA	ATG	CGA	CAG	GAG	ATA
CO	421	GGT	AGA	CAG	ATC	CAC	AAG	ATC	AGA	AAG	CAG	TTC	TTC	CAC	GCT	ATC	ATG	AGA	CAA	GAG	ATC

		G	W	F	D	V	H	D	V	G	E	L	N	T	R	L	T	D	D	V	S
WT	481	GGC	TGG	TTT	GAT	GTG	CAC	GAT	GTT	GGG	GAG	CTT	AAC	ACC	CGA	CTT	ACA	GAT	GAT	GTC	TCC
CO	481	GGT	TGG	TTC	GAC	GTT	CAC	GAC	GTT	GGT	GAG	TTG	AAC	ACT	AGA	TTG	ACT	GAC	GAC	GTT	TCC
		K	I	N	E	G	I	G	D	K	I	G	M	F	F	Q	S	M	A	T	F
WT	541	AAG	ATT	AAT	GAA	GGA	ATT	GGT	GAC	AAA	ATT	GGA	ATG	TTC	TTT	CAG	TCA	ATG	GCA	ACA	TTT
CO	541	AAG	ATC	AAC	GAG	GGT	ATC	GGT	GAC	AAG	ATC	GGT	ATG	TTC	TTC	CAG	TCC	ATG	GCT	ACT	TTC
		F	T	G	F	I	V	G	F	T	R	G	W	K	L	T	L	V	I	L	A
WT	601	TTC	ACT	GGG	TTT	ATA	GTA	GGA	TTT	ACA	CGT	GGT	TGG	AAG	CTA	ACC	CTT	GTG	ATT	TTG	GCC
CO	601	TTC	ACT	GGA	TTC	ATC	GTT	GGT	TTC	ACT	AGA	GGT	TGG	AAG	TTG	ACT	TTG	GTT	ATC	TTG	GCT
		I	S	P	V	L	G	L	S	A	A	V	W	A	K	I	L	S	S	F	T
WT	661	ATC	AGT	CCT	GTT	CTT	GGA	CTG	TCA	GCT	GCT	GTC	TGG	GCA	AAG	ATA	CTA	TCT	TCA	TTT	ACT
CO	661	ATC	TCC	CCA	GTT	TTG	GGT	TTG	TCT	GCT	GCT	GTT	TGG	GCT	AAG	ATC	TTG	TCC	TCA	TTC	ACT
		D	K	E	L	L	A	Y	A	K	A	G	A	V	A	E	E	V	L	A	A
WT	721	GAT	AAA	GAA	CTC	TTA	GCG	TAT	GCA	AAA	GCT	GGA	GCA	GTA	GCT	GAA	GAG	GTC	TTG	GCA	GCA
CO	721	GAC	AAA	GAG	TTG	TTG	GCT	TAC	GCT	AAG	GCT	GGT	GCT	GTT	GCT	GAA	GAG	GTT	TTG	GCT	GCT
		I	R	T	V	I	A	F	G	G	Q	K	K	E	L	E	R	Y	N	K	N
WT	781	ATT	AGA	ACT	GTG	ATT	GCA	TTT	GGA	GGA	CAA	AAG	AAA	GAA	CTT	GAA	AGG	TAC	AAC	AAA	AAT
CO	781	ATT	AGA	ACT	GTT	ATC	GCT	TTC	GGT	GGT	CAG	AAG	AAA	GAG	TTG	GAG	AGA	TAC	AAC	AAG	AAT
		L	E	E	A	K	R	I	G	I	K	K	A	I	T	A	N	I	S	I	G
WT	841	TTA	GAA	GAA	GCT	AAA	AGA	ATT	GGG	ATA	AAG	AAA	GCT	ATT	ACA	GCC	AAT	ATT	TCT	ATA	GGT
CO	841	TTG	GAG	GAA	GCT	AAG	AGA	ATC	GGT	ATC	AAG	AAG	GCT	ATC	ACT	GCT	AAC	ATC	TCC	ATC	GGT
		A	A	F	L	L	I	Y	A	S	Y	A	L	A	F	W	Y	G	T	T	L
WT	901	GCT	GCT	TTC	CTG	CTG	ATC	TAT	GCA	TCT	TAT	GCT	CTG	GCC	TTC	TGG	TAT	GGG	ACC	ACC	TTG
CO	901	GCT	GCT	TTT	TTG	TTG	ATC	TAC	GCT	TCC	TAC	GCT	TTG	GCT	TTC	TGG	TAC	GGT	ACT	ACT	TTG
		V	L	S	G	E	Y	S	I	G	Q	V	L	T	V	F	F	S	V	L	I
WT	961	GTC	CTC	TCA	GGG	GAA	TAT	TCT	ATT	GGA	CAA	GTA	CTC	ACT	GTA	TTC	TTT	TCT	GTA	TTA	ATT
CO	961	GTT	TTG	TCC	GGT	GAG	TAC	TCC	ATT	GGT	CAG	GTT	TTG	ACT	GTT	TTC	TTC	TCC	GTT	TTG	ATC
		G	A	F	S	V	G	Q	A	S	P	S	I	E	A	F	A	N	A	R	G
WT	1021	GGG	GCT	TTT	AGT	GTT	GGA	CAG	GCA	TCT	CCA	AGC	ATT	GAA	GCA	TTT	GCA	AAT	GCA	AGA	GGA
CO	1021	GGT	GCT	TTC	TCC	GTT	GGT	CAA	GCT	TCT	CCA	TCT	ATT	GAG	GCT	TTT	GCT	AAC	GCT	AGA	GGT
		A	A	Y	E	I	F	K	I	I	D	N	K	P	S	I	D	S	Y	S	K
WT	1081	GCA	GCT	TAT	GAA	ATC	TTC	AAG	ATA	ATT	GAT	AAT	AAG	CCA	AGT	ATT	GAC	AGC	TAT	TCG	AAG
CO	1081	GCT	GCT	TAC	GAG	ATC	TTC	AAG	ATC	ATC	GAC	AAC	AAG	CCA	TCC	ATC	GAC	TCC	TAC	TCT	AAG

		S	G	H	K	P	D	N	I	K	G	N	L	E	F	R	N	V	H	F	S
WT	1141	AGT	GGG	CAC	AAA	CCA	GAT	AAT	ATT	AAG	GGA	AAT	TTG	GAA	TTC	AGA	AAT	GTT	CAC	TTC	AGT
CO	1141	TCC	GGA	CAT	AAG	CCA	GAC	AAC	ATC	AAG	GGA	AAC	TTG	GAG	TTC	AGA	AAT	GTT	CAC	TTC	TCC
		Y	P	S	R	K	E	V	K	I	L	K	G	L	N	L	K	V	Q	S	G
WT	1201	TAC	CCA	TCT	CGA	AAA	GAA	GTT	AAG	ATC	TTG	AAG	GGC	CTG	AAC	CTG	AAG	GTG	CAG	AGT	GGG
CO	1201	TAC	CCA	TCC	AGA	AAA	GAG	GTT	AAG	ATT	TTG	AAG	GGT	TTG	AAC	TTG	AAG	GTT	CAG	TCC	GGT
		Q	T	V	A	L	V	G	N	S	G	C	G	K	S	T	T	V	Q	L	M
WT	1261	CAG	ACG	GTG	GCC	CTG	GTT	GGA	AAC	AGT	GGC	TGT	GGG	AAG	AGC	ACA	ACA	GTC	CAG	CTG	ATG
CO	1261	CAG	ACT	GTT	GCT	TTG	GTT	GGT	AAC	TCT	GGT	TGT	GGT	AAG	TCC	ACT	ACT	GTT	CAG	TTG	ATG
		Q	R	L	Y	D	P	T	E	G	M	V	S	V	D	G	Q	D	I	R	T
WT	1321	CAG	AGG	CTC	TAT	GAC	CCC	ACA	GAG	GGG	ATG	GTC	AGT	GTT	GAT	GGA	CAG	GAT	ATT	AGG	ACC
CO	1321	CAG	AGA	TTG	TAC	GAC	CCA	ACT	GAG	GGT	ATG	GTT	TCC	GTT	GAC	GGT	CAG	GAC	ATC	AGA	ACT
		I	N	V	R	F	L	R	E	I	I	G	V	V	S	Q	E	P	V	L	F
WT	1381	ATA	AAT	GTA	AGG	TTT	CTA	CGG	GAA	ATC	ATT	GGT	GTG	GTG	AGT	CAG	GAA	CCT	GTA	TTG	TTT
CO	1381	ATC	AAT	GTT	AGA	TTC	TTG	AGA	GAG	ATC	ATC	GGA	GTT	GTT	TCC	CAA	GAG	CCA	GTT	TTG	TTC
		A	T	T	I	A	E	N	I	R	Y	G	R	E	N	V	T	M	D	E	I
WT	1441	GCC	ACC	ACG	ATA	GCT	GAA	AAC	ATT	CGC	TAT	GGC	CGT	GAA	AAT	GTC	ACC	ATG	GAT	GAG	ATT
CO	1441	GCT	ACT	ACT	ATC	GCT	GAG	AAC	ATC	AGA	TAC	GGT	AGA	GAA	AAT	GTT	ACT	ATG	GAC	GAG	ATC
		E	K	A	V	K	E	A	N	A	Y	D	F	I	M	K	L	P	H	K	F
WT	1501	GAG	AAA	GCT	GTC	AAG	GAA	GCC	AAT	GCC	TAT	GAC	TTT	ATC	ATG	AAA	CTG	CCT	CAT	AAA	TTT
CO	1501	GAG	AAG	GCT	GTT	AAG	GAA	GCT	AAC	GCT	TAC	GAC	TTT	ATC	ATG	AAG	TTG	CCA	CAC	AAG	TTC
		D	T	L	V	G	E	R	G	A	Q	L	S	G	G	Q	K	Q	R	I	A
WT	1561	GAC	ACC	CTG	GTT	GGA	GAG	AGA	GGG	GCC	CAG	TTG	AGT	GGT	GGG	CAG	AAG	CAG	AGG	ATC	GCC
CO	1561	GAC	ACA	TTG	GTT	GGT	GAG	AGA	GGT	GCT	CAA	TTG	TCT	GGT	GGT	CAA	AAG	CAG	AGA	ATC	GCT
		I	A	R	A	L	V	R	N	P	K	I	L	L	L	D	E	A	T	S	A
WT	1621	ATT	GCA	CGT	GCC	CTG	GTT	CGC	AAC	CCC	AAG	ATC	CTC	CTG	CTG	GAT	GAG	GCC	ACG	TCA	GCC
CO	1621	ATT	GCT	AGA	GCT	TTG	GTT	AGA	AAC	CCA	AAG	ATT	TTG	TTG	TTG	GAC	GAG	GCT	ACT	TCC	GCT
		L	D	T	E	S	E	A	V	V	Q	V	A	L	D	K	A	R	K	G	R
WT	1681	TTG	GAC	ACA	GAA	AGC	GAA	GCA	GTG	GTT	CAG	GTG	GCT	CTG	GAT	AAG	GCC	AGA	AAA	GGT	CGG
CO	1681	TTG	GAC	ACT	GAA	TCT	GAA	GCT	GTT	GTT	CAG	GTT	GCT	TTG	GAC	AAG	GCT	AGA	AAG	GGT	AGA
		T	T	I	V	I	A	H	R	L	S	T	V	R	N	A	D	V	I	A	G
WT	1741	ACC	ACC	ATT	GTG	ATA	GCT	CAT	CGT	TTG	TCT	ACA	GTT	CGT	AAT	GCT	GAC	GTC	ATC	GCT	GGT
CO	1741	ACT	ACA	ATC	GTT	ATC	GCT	CAC	AGA	TTG	TCC	ACT	GTT	AGA	AAC	GCT	GAC	GTT	ATT	GCT	GGT

		F	D	D	G	V	I	V	E	K	G	N	H	D	E	L	M	K	E	K	G
WT	1801	TTC	GAT	GAT	GGA	GTC	ATT	GTG	GAG	AAA	GGA	AAT	CAT	GAT	GAA	CTC	ATG	AAA	GAG	AAA	GGC
CO	1801	TTC	GAC	GAC	GGT	GTT	ATC	GTT	GAG	AAG	GGT	AAC	CAC	GAC	GAG	TTG	ATG	AAG	GAA	AAG	GGT

		I	Y	F	K	L	V	T	M	Q	T	A	G	N	E	V	E	L	E	N	A
WT	1861	ATT	TAC	TTC	AAA	CTT	GTC	ACA	ATG	CAG	ACA	GCA	GGA	AAT	GAA	GTT	GAA	TTA	GAA	AAT	GCA
CO	1861	ATC	TAC	TTT	AAG	TTG	GTT	ACT	ATG	CAG	ACT	GCT	GGT	AAC	GAG	GTT	GAA	TTG	GAA	AAC	GCT

		A	D	E	S	K	S	E	I	D	A	L	E	M	S	S	N	D	S	R	S
WT	1921	GCT	GAT	GAA	TCC	AAA	AGT	GAA	ATT	GAT	GCC	TTG	GAA	ATG	TCT	TCA	AAT	GAT	TCA	AGA	TCC
CO	1921	GCT	GAC	GAG	TCC	AAG	TCT	GAG	ATC	GAT	GCT	TTG	GAA	ATG	TCC	TCC	AAC	GAC	TCC	AGA	TCC

		S	L	I	R	K	R	S	T	R	R	S	V	R	G	S	Q	A	Q	D	R
WT	1981	AGT	CTA	ATA	AGA	AAA	AGA	TCA	ACT	CGT	AGG	AGT	GTC	CGT	GGA	TCA	CAA	GCC	CAA	GAC	AGA
CO	1981	TCC	TTG	ATT	AGA	AAG	AGA	TCC	ACT	AGA	AGA	TCC	GTT	AGA	GGT	TCC	CAA	GCT	CAG	GAC	AGA

		K	L	S	T	K	E	A	L	D	E	S	I	P	P	V	S	F	W	R	I
WT	2041	AAG	CTT	AGT	ACC	AAA	GAG	GCT	CTG	GAT	GAA	AGT	ATA	CCT	CCA	GTT	TCC	TTT	TGG	AGG	ATT
CO	2041	AAG	TTG	TCC	ACA	AAA	GAG	GCT	TTG	GAC	GAA	TCC	ATC	CCA	CCA	GTT	TCA	TTT	TGG	AGA	ATT

		M	K	L	N	L	T	E	W	P	Y	F	V	V	G	V	F	C	A	I	I
WT	2101	ATG	AAG	CTA	AAT	TTA	ACT	GAA	TGG	CCT	TAT	TTT	GTT	GTT	GGT	GTA	TTT	TGT	GCC	ATT	ATA
CO	2101	ATG	AAG	TTG	AAT	TTG	ACT	GAG	TGG	CCA	TAC	TTC	GTT	GTT	GGT	GTT	TTC	TGT	GCT	ATT	ATC

		N	G	G	L	Q	P	A	F	A	I	I	F	S	K	I	I	G	V	F	T
WT	2161	AAT	GGA	GGC	CTG	CAA	CCA	GCA	TTT	GCA	ATA	ATA	TTT	TCA	AAG	ATT	ATA	GGG	GTT	TTT	ACA
CO	2161	AAC	GGT	GGA	TTG	CAG	CCA	GCT	TTC	GCT	ATC	ATC	TTC	TCC	AAG	ATT	ATC	GGT	GTT	TTC	ACA

		R	I	D	D	P	E	T	K	R	Q	N	S	N	L	F	S	L	L	F	L
WT	2221	AGA	ATT	GAT	GAT	CCT	GAA	ACA	AAA	CGA	CAG	AAT	AGT	AAC	TTG	TTT	TCA	CTA	TTG	TTT	CTA
CO	2221	AGA	ATC	GAC	GAC	CCA	GAG	ACT	AAG	AGA	CAA	AAC	TCC	AAC	TTG	TTC	TCT	TTG	TTG	TTC	TTG

		A	L	G	I	I	S	F	I	T	F	F	L	Q	G	F	T	F	G	K	A
WT	2281	GCC	CTT	GGA	ATT	ATT	TCT	TTT	ATT	ACA	TTT	TTC	CTT	CAG	GGT	TTC	ACA	TTT	GGC	AAA	GCT
CO	2281	GCT	TTG	GGT	ATC	ATC	TCC	TTT	ATT	ACA	TTC	TTC	TTG	CAG	GGT	TTC	ACT	TTC	GGT	AAA	GCT

		G	E	I	L	T	K	R	L	R	Y	M	V	F	R	S	M	L	R	Q	D
WT	2341	GGA	GAG	ATC	CTC	ACC	AAG	CGG	CTC	CGA	TAC	ATG	GTT	TTC	CGA	TCC	ATG	CTC	AGA	CAG	GAT
CO	2341	GGT	GAG	ATC	TTG	ACA	AAA	AGA	TTG	AGA	TAT	ATG	GTT	TTC	AGA	TCC	ATG	TTG	AGA	CAG	GAC

		V	S	W	F	D	D	P	K	N	T	T	G	A	L	T	T	R	L	A	N
WT	2401	GTG	AGT	TGG	TTT	GAT	GAC	CCT	AAA	AAC	ACC	ACT	GGA	GCA	TTG	ACT	ACC	AGG	CTC	GCC	AAT
CO	2401	GTT	TCC	TGG	TTC	GAT	GAC	CCA	AAG	AAC	ACT	ACT	GGT	GCT	TTG	ACT	ACT	AGA	TTG	GCT	AAC

		D	**A**	**A**	**Q**	**V**	**K**	**G**	**A**	**I**	**G**	**S**	**R**	**L**	**A**	**V**	**I**	**T**	**Q**	**N**	**I**
WT	2461	GAT	GCT	GCT	CAA	GTT	AAA	GGG	GCT	ATA	GGT	TCC	AGG	CTT	GCT	GTA	ATT	ACC	CAG	AAT	ATA
CO	2461	GAT	GCT	GCT	CAG	GTT	AAG	GGT	GCT	ATT	GGT	TCC	AGA	TTG	GCT	GTT	ATC	ACT	CAG	AAC	ATT
		A	**N**	**L**	**G**	**T**	**G**	**I**	**I**	**I**	**S**	**F**	**I**	**Y**	**G**	**W**	**Q**	**L**	**T**	**L**	**L**
WT	2521	GCA	AAT	CTT	GGG	ACA	GGA	ATA	ATT	ATA	TCC	TTC	ATC	TAT	GGT	TGG	CAA	CTA	ACA	CTG	TTA
CO	2521	GCT	AAC	TTG	GGT	ACT	GGT	ATC	ATC	ATC	TCA	TTC	ATC	TAC	GGA	TGG	CAA	TTG	ACT	CTG	TTG
		L	**L**	**A**	**I**	**V**	**P**	**I**	**I**	**A**	**I**	**A**	**G**	**V**	**V**	**E**	**M**	**K**	**M**	**L**	**S**
WT	2581	CTC	TTA	GCA	ATT	GTA	CCC	ATC	ATT	GCA	ATA	GCA	GGA	GTT	GTT	GAA	ATG	AAA	ATG	TTG	TCT
CO	2581	TTG	TTG	GCT	ATC	GTT	CCA	ATT	ATC	GCT	ATC	GCT	GGT	GTT	GTT	GAG	ATG	AAG	ATG	TTG	TCA
		G	**Q**	**A**	**L**	**K**	**D**	**K**	**K**	**E**	**L**	**E**	**G**	**A**	**G**	**K**	**I**	**A**	**T**	**E**	**A**
WT	2641	GGA	CAA	GCA	CTG	AAA	GAT	AAG	AAA	GAA	CTA	GAA	GGT	GCT	GGG	AAG	ATC	GCT	ACT	GAA	GCA
CO	2641	GGA	CAG	GCT	TTG	AAA	GAT	AAG	AAA	GAA	TTG	GAA	GGT	GCT	GGT	AAG	ATC	GCT	ACT	GAG	GCT
		I	**E**	**N**	**F**	**R**	**T**	**V**	**V**	**S**	**L**	**T**	**Q**	**E**	**Q**	**K**	**F**	**E**	**H**	**M**	**Y**
WT	2701	ATA	GAA	AAC	TTC	CGA	ACC	GTT	GTT	TCT	TTG	ACT	CAG	GAG	CAG	AAG	TTT	GAA	CAT	ATG	TAT
CO	2701	ATC	GAG	AAC	TTC	AGA	ACT	GTT	GTT	TCC	TTG	ACT	CAA	GAG	CAA	AAG	TTT	GAG	CAC	ATG	TAC
		A	**Q**	**S**	**L**	**Q**	**V**	**P**	**Y**	**R**	**N**	**S**	**L**	**R**	**K**	**A**	**H**	**I**	**F**	**G**	**I**
WT	2761	GCT	CAG	AGT	TTG	CAG	GTA	CCA	TAC	AGA	AAC	TCT	TTG	AGG	AAA	GCA	CAC	ATC	TTT	GGA	ATT
CO	2761	GCT	CAG	TCC	TTG	CAG	GTT	CCT	TAC	AGA	AAC	TCC	TTG	AGA	AAG	GCT	CAC	ATC	TTC	GGT	ATC
		T	**F**	**S**	**F**	**T**	**Q**	**A**	**M**	**M**	**Y**	**F**	**S**	**Y**	**A**	**G**	**C**	**F**	**R**	**F**	**G**
WT	2821	ACA	TTT	TCC	TTC	ACC	CAG	GCA	ATG	ATG	TAT	TTT	TCC	TAT	GCT	GGA	TGT	TTC	CGG	TTT	GGA
CO	2821	ACT	TTC	TCA	TTC	ACA	CAG	GCT	ATG	ATG	TAC	TTC	TCT	TAC	GCT	GGT	TGT	TTC	AGA	TTC	GGT
		A	**Y**	**L**	**V**	**A**	**H**	**K**	**L**	**M**	**S**	**F**	**E**	**D**	**V**	**L**	**L**	**V**	**F**	**S**	**A**
WT	2881	GCC	TAC	TTG	GTG	GCA	CAT	AAA	CTC	ATG	AGC	TTT	GAG	GAT	GTT	CTG	TTA	GTA	TTT	TCA	GCT
CO	2881	GCT	TAC	TTG	GTT	GCT	CAC	AAG	TTG	ATG	TCC	TTC	GAG	GAC	GTT	TTG	TTG	GTT	TTT	TCC	GCT
		V	**V**	**F**	**G**	**A**	**M**	**A**	**V**	**G**	**Q**	**V**	**S**	**S**	**F**	**A**	**P**	**D**	**Y**	**A**	**K**
WT	2941	GTT	GTC	TTT	GGT	GCC	ATG	GCC	GTG	GGG	CAA	GTC	AGT	TCA	TTT	GCT	CCT	GAC	TAT	GCC	AAA
CO	2941	GTT	GTT	TTC	GGA	GCT	ATG	GCT	GTT	GGA	CAG	GTT	TCC	TCT	TTT	GCT	CCT	GAC	TAC	GCT	AAA
		A	**K**	**I**	**S**	**A**	**A**	**H**	**I**	**I**	**M**	**I**	**I**	**E**	**K**	**T**	**P**	**L**	**I**	**D**	**S**
WT	3001	GCC	AAA	ATA	TCA	GCA	GCC	CAC	ATC	ATC	ATG	ATC	ATT	GAA	AAA	ACC	CCT	TTG	ATT	GAC	AGC
CO	3001	GCT	AAG	ATT	TCC	GCT	GCT	CAC	ATC	ATC	ATG	ATC	ATC	GAA	AAG	ACT	CCA	TTG	ATT	GAC	TCT
		Y	**S**	**T**	**E**	**G**	**L**	**M**	**P**	**N**	**T**	**L**	**E**	**G**	**N**	**V**	**T**	**F**	**G**	**E**	**V**
WT	3061	TAC	AGC	ACG	GAA	GGC	CTA	ATG	CCG	AAC	ACA	TTG	GAA	GGA	AAT	GTC	ACA	TTT	GGT	GAA	GTT
CO	3061	TAC	TCC	ACT	GAG	GGT	TTG	ATG	CCA	AAC	ACT	TTG	GAG	GGA	AAC	GTT	ACT	TTC	GGA	GAG	GTT

		V	F	N	Y	P	T	R	P	D	I	P	V	L	Q	G	L	S	L	E	V
WT	3121	GTA	TTC	AAC	TAT	CCC	ACC	CGA	CCG	GAC	ATC	CCA	GTG	CTT	CAG	GGA	CTG	AGC	CTG	GAG	GTG
CO	3121	GTT	TTC	AAC	TAC	CCA	ACT	AGA	CCA	GAC	ATC	CCT	GTT	TTG	CAA	GGT	TTG	TCC	TTG	GAA	GTT
		K	K	G	Q	T	L	A	L	V	G	S	S	G	C	G	K	S	T	V	V
WT	3181	AAG	AAG	GGC	CAG	ACG	CTG	GCT	CTG	GTG	GGC	AGC	AGT	GGC	TGT	GGG	AAG	AGC	ACA	GTG	GTC
CO	3181	AAG	AAG	GGT	CAG	ACA	TTG	GCT	TTG	GTT	GGA	TCT	TCC	GGT	TGT	GGA	AAA	TCC	ACA	GTT	GTT
		Q	L	L	E	R	F	Y	D	P	L	A	G	K	V	L	L	D	G	K	E
WT	3241	CAG	CTC	CTG	GAG	CGG	TTC	TAC	GAC	CCC	TTG	GCA	GGG	AAA	GTG	CTG	CTT	GAT	GGC	AAA	GAA
CO	3241	CAG	TTG	TTG	GAA	AGA	TTC	TAT	GAC	CCT	TTG	GCT	GGT	AAG	GTT	TTG	TTG	GAC	GGT	AAA	GAG
		I	K	R	L	N	V	Q	W	L	R	A	H	L	G	I	V	S	Q	E	P
WT	3301	ATA	AAG	CGA	CTG	AAT	GTT	CAG	TGG	CTC	CGA	GCA	CAC	CTG	GGC	ATC	GTG	TCC	CAG	GAG	CCC
CO	3301	ATC	AAG	AGA	TTG	AAT	GTT	CAG	TGG	TTG	AGA	GCT	CAC	TTG	GGT	ATC	GTT	TCT	CAA	GAA	CCT
		I	L	F	D	C	S	I	A	E	N	I	A	Y	G	D	N	S	R	V	V
WT	3361	ATC	CTG	TTT	GAC	TGC	AGC	ATT	GCT	GAG	AAC	ATT	GCC	TAT	GGA	GAC	AAC	AGC	CGG	GTG	GTG
CO	3361	ATC	TTG	TTC	GAC	TGT	TCC	ATT	GCT	GAG	AAT	ATT	GCT	TAC	GGT	GAC	AAC	TCC	AGA	GTT	GTT
		S	Q	E	E	I	V	R	A	A	K	E	A	N	I	H	A	F	I	E	S
WT	3421	TCA	CAG	GAA	GAG	ATC	GTG	AGG	GCA	GCA	AAG	GAG	GCC	AAC	ATA	CAT	GCC	TTC	ATC	GAG	TCA
CO	3421	TCT	CAA	GAA	GAG	ATC	GTT	AGA	GCT	GCT	AAA	GAG	GCT	AAC	ATC	CAC	GCT	TTC	ATT	GAG	TCC
		L	P	N	K	Y	S	T	K	V	G	D	K	G	T	Q	L	S	G	G	Q
WT	3481	CTG	CCT	AAT	AAA	TAT	AGC	ACT	AAA	GTA	GGA	GAC	AAA	GGA	ACT	CAG	CTC	TCT	GGT	GGC	CAG
CO	3481	TTG	CCA	AAC	AAG	TAC	TCC	ACT	AAG	GTT	GGT	GAC	AAG	GGT	ACT	CAG	TTG	TCC	GGT	GGA	CAG
		K	Q	R	I	A	I	A	R	A	L	V	R	Q	P	H	I	L	L	L	D
WT	3541	AAA	CAA	CGC	ATT	GCC	ATA	GCT	CGT	GCC	CTT	GTT	AGA	CAG	CCT	CAT	ATT	TTG	CTT	TTG	GAT
CO	3541	AAA	CAA	AGA	ATT	GCT	ATC	GCT	AGA	GCT	TTG	GTT	AGA	CAG	CCA	CAC	ATC	TTG	TTG	TTG	GAT
		E	A	T	S	A	L	D	T	E	S	E	K	V	V	Q	E	A	L	D	K
WT	3601	GAA	GCC	ACG	TCA	GCT	CTG	GAT	ACA	GAA	AGT	GAA	AAG	GTT	GTC	CAA	GAA	GCC	CTG	GAC	AAA
CO	3601	GAA	GCT	ACA	TCT	GCT	TTG	GAT	ACA	GAG	TCT	GAG	AAG	GTT	GTT	CAA	GAA	GCT	TTG	GAT	AAG
		A	R	E	G	R	T	C	I	V	I	A	H	R	L	S	T	I	Q	N	A
WT	3661	GCC	AGA	GAA	GGC	CGC	ACC	TGC	ATT	GTG	ATT	GCT	CAC	CGC	CTG	TCC	ACC	ATC	CAG	AAT	GCA
CO	3661	GCT	AGA	GAG	GGA	AGA	ACT	TGT	ATC	GTT	ATT	GCT	CAT	AGA	TTG	TCT	ACT	ATC	CAG	AAC	GCT
		D	L	I	V	V	F	Q	N	G	R	V	K	E	H	G	T	H	Q	Q	L
WT	3721	GAC	TTA	ATA	GTG	GTG	TTT	CAG	AAT	GGC	AGA	GTC	AAG	GAG	CAT	GGC	ACG	CAT	CAG	CAG	CTG
CO	3721	GAC	TTG	ATC	GTT	GTT	TTC	CAA	AAC	GGT	AGA	GTT	AAG	GAA	CAC	GGT	ACT	CAC	CAG	CAG	TTG

		L	A	Q	K	G	I	Y	F	S	M	V	S	V	Q	A	G	T	K	R	Q
WT	3781	CTG	GCA	CAG	AAA	GGC	ATC	TAT	TTT	TCA	ATG	GTC	AGT	GTC	CAG	GCT	GGA	ACA	AAG	CGC	CAG
CO	3781	TTG	GCT	CAA	AAG	GGA	ATC	TAT	TTC	TCT	ATG	GTT	TCT	GTT	CAG	GCT	GGT	ACA	AAA	AGA	CAA

Appendix Figure 2. Amino acid and nucleotide sequence alignment of wild-type and codon optimized mouse (*mdr1a*) P-gp. Amino acid residues (blue) and the corresponding codons for wild-type (WT) and codon optimized (CO) *mdr1a* cDNA are shown. Nucleotides changed during codon optimization are indicated in red.

		M	E	L	E	E	D	L	K	G	R	A	D	K	N	F	S	K	M	G	K
WT	1	ATG	GAA	CTT	GAA	GAG	GAC	CTT	AAG	GGA	AGA	GCA	GAC	AAG	AAC	TTC	TCA	AAG	ATG	GGC	AAA
CO	1	ATG	GAA	TTG	GAA	GAG	GAC	TTG	AAG	GGT	AGA	GCT	GAC	AAG	AAC	TTC	TCC	AAG	ATG	GGT	AAG
		K	S	K	K	E	K	K	E	K	K	P	A	V	S	V	L	T	M	F	R
WT	61	AAG	AGT	AAA	AAG	GAG	AAG	AAA	GAA	AAG	AAA	CCA	GCA	GTC	AGT	GTG	CTT	ACA	ATG	TTT	CGT
CO	61	AAG	TCC	AAG	AAA	GAA	AAG	AAA	GAG	AAA	AAG	CCT	GCT	GTT	TCC	GTT	TTG	ACT	ATG	TTC	AGA
		Y	A	G	W	L	D	R	L	Y	M	L	V	G	T	L	A	A	I	I	H
WT	121	TAT	GCA	GGT	TGG	CTG	GAC	AGG	TTG	TAC	ATG	CTG	GTG	GGA	ACT	CTG	GCT	GCT	ATT	ATC	CAT
CO	121	TAC	GCT	GGT	TGG	TTG	GAC	AGA	TTG	TAC	ATG	TTG	GTT	GGT	ACT	TTG	GCT	GCT	ATC	ATC	CAC
		G	V	A	L	P	L	M	M	L	I	F	G	D	M	T	D	S	F	A	S
WT	181	GGA	GTG	GCG	CTC	CCA	CTT	ATG	ATG	CTG	ATC	TTT	GGT	GAC	ATG	ACA	GAT	AGC	TTT	GCA	AGT
CO	181	GGT	GTT	GCT	TTG	CCA	TTG	ATG	ATG	TTG	ATC	TTC	GGT	GAT	ATG	ACT	GAC	TCC	TTC	GCT	TCC
		V	G	N	V	S	K	N	S	T	N	M	S	E	A	D	K	R	A	M	F
WT	241	GTA	GGA	AAC	GTC	TCT	AAA	AAC	AGT	ACT	AAT	ATG	AGT	GAG	GCC	GAT	AAA	AGA	GCC	ATG	TTT
CO	241	GTT	GGT	AAT	GTT	TCC	AAG	AAC	TCC	ACT	AAC	ATG	TCC	GAG	GCT	GAT	AAG	AGA	GCT	ATG	TTC
		A	K	L	E	E	E	M	T	T	Y	A	Y	Y	Y	T	G	I	G	A	G
WT	301	GCC	AAA	CTG	GAG	GAA	GAA	ATG	ACC	ACG	TAC	GCC	TAC	TAT	TAC	ACC	GGG	ATT	GGT	GCT	GGT
CO	301	GCT	AAG	TTG	GAG	GAA	GAG	ATG	ACT	ACT	TAC	GCT	TAC	TAC	TAC	ACT	GGT	ATC	GGT	GCT	GGT
		V	L	I	V	A	Y	I	Q	V	S	F	W	C	L	A	A	G	R	Q	I
WT	361	GTG	CTC	ATA	GTT	GCC	TAC	ATC	CAG	GTT	TCA	TTT	TGG	TGC	CTG	GCA	GCT	GGA	AGA	CAG	ATA
CO	361	GTT	TTG	ATC	GTT	GCT	TAC	ATC	CAG	GTT	TCT	TTC	TGG	TGT	TTG	GCT	GCT	GGT	AGA	CAG	ATC

		H	**K**	**I**	**R**	**Q**	**K**	**F**	**F**	**H**	**A**	**I**	**M**	**N**	**Q**	**E**	**I**	**G**	**W**	**F**	**D**
WT	421	CAC	AAG	ATC	AGG	CAG	AAG	TTT	TTT	CAT	GCT	ATA	ATG	AAT	CAG	GAG	ATA	GGC	TGG	TTT	GAT
CO	421	CAC	AAG	ATC	AGA	CAA	AAG	TTC	TTC	CAC	GCT	ATC	ATG	AAT	CAA	GAG	ATC	GGT	TGG	TTC	GAC
		V	**H**	**D**	**V**	**G**	**E**	**L**	**N**	**T**	**R**	**L**	**T**	**D**	**D**	**V**	**S**	**K**	**I**	**N**	**E**
WT	481	GTG	CAT	GAC	GTT	GGG	GAG	CTC	AAC	ACC	CGG	CTC	ACA	GAT	GAT	GTT	TCC	AAA	ATT	AAT	GAA
CO	481	GTT	CAC	GAC	GTT	GGT	GAG	TTG	AAC	ACT	AGA	TTG	ACT	GAC	GAC	GTT	TCT	AAG	ATC	AAC	GAG
		G	**I**	**G**	**D**	**K**	**I**	**G**	**M**	**F**	**F**	**Q**	**A**	**M**	**A**	**T**	**F**	**F**	**G**	**G**	**F**
WT	541	GGA	ATT	GGT	GAC	AAA	ATC	GGA	ATG	TTC	TTC	CAG	GCA	ATG	GCA	ACA	TTT	TTT	GGT	GGT	TTT
CO	541	GGT	ATC	GGA	GAC	AAG	ATC	GGT	ATG	TTT	TTC	CAG	GCT	ATG	GCT	ACT	TTC	TTC	GGT	GGT	TTC
		I	**I**	**G**	**F**	**T**	**R**	**G**	**W**	**K**	**L**	**T**	**L**	**V**	**I**	**L**	**A**	**I**	**S**	**P**	**V**
WT	601	ATA	ATA	GGA	TTT	ACC	CGT	GGC	TGG	AAG	CTA	ACC	CTT	GTG	ATT	TTG	GCC	ATC	AGC	CCT	GTT
CO	601	ATC	ATC	GGT	TTC	ACT	AGA	GGT	TGG	AAG	TTG	ACT	TTG	GTT	ATC	TTG	GCT	ATC	TCC	CCA	GTT
		L	**G**	**L**	**S**	**A**	**G**	**I**	**W**	**A**	**K**	**I**	**L**	**S**	**S**	**F**	**T**	**D**	**K**	**E**	**L**
WT	661	CTT	GGA	CTG	TCA	GCT	GGT	ATT	TGG	GCA	AAG	ATA	TTG	TCT	TCA	TTT	ACT	GAT	AAG	GAA	CTC
CO	661	TTG	GGT	TTG	TCC	GCT	GGT	ATT	TGG	GCT	AAG	ATC	TTG	TCC	TCA	TTC	ACT	GAC	AAA	GAG	TTG
		H	**A**	**Y**	**A**	**K**	**A**	**G**	**A**	**V**	**A**	**E**	**E**	**V**	**L**	**A**	**A**	**I**	**R**	**T**	**V**
WT	721	CAT	GCT	TAT	GCA	AAA	GCT	GGA	GCA	GTT	GCT	GAA	GAA	GTC	TTA	GCA	GCC	ATC	AGA	ACT	GTG
CO	721	CAC	GCT	TAC	GCT	AAG	GCT	GGT	GCT	GTT	GCT	GAA	GAA	GTT	TTG	GCT	GCT	ATT	AGA	ACT	GTT
		I	**A**	**F**	**G**	**G**	**Q**	**K**	**K**	**E**	**L**	**E**	**R**	**Y**	**N**	**N**	**N**	**L**	**E**	**E**	**A**
WT	781	ATT	GCG	TTT	GGA	GGA	CAA	AAG	AAG	GAA	CTT	GAA	AGG	TAC	AAT	AAC	AAC	TTG	GAA	GAA	GCT
CO	781	ATC	GCT	TTC	GGT	GGT	CAG	AAG	AAA	GAG	TTG	GAG	AGA	TAC	AAC	AAC	AAC	TTG	GAA	GAA	GCT
		K	**R**	**L**	**G**	**I**	**K**	**K**	**A**	**I**	**T**	**A**	**N**	**I**	**S**	**M**	**G**	**A**	**A**	**F**	**L**
WT	841	AAA	AGG	CTG	GGG	ATA	AAG	AAA	GCT	ATC	ACG	GCC	AAC	ATC	TCC	ATG	GGT	GCA	GCT	TTT	CTC
CO	841	AAG	AGA	TTG	GGT	ATC	AAG	AAG	GCT	ATC	ACT	GCT	AAC	ATC	TCC	ATG	GGT	GCT	GCT	TTC	TTG
		L	**I**	**Y**	**A**	**S**	**Y**	**A**	**L**	**A**	**F**	**W**	**Y**	**G**	**T**	**S**	**L**	**V**	**I**	**S**	**K**
WT	901	CTT	ATC	TAT	GCA	TCA	TAT	GCT	CTG	GCA	TTC	TGG	TAT	GGG	ACT	TCC	TTG	GTC	ATC	TCC	AAA
CO	901	TTG	ATC	TAC	GCT	TCC	TAC	GCT	TTG	GCT	TTC	TGG	TAC	GGT	ACT	TCC	TTG	GTT	ATT	TCC	AAA
		E	**Y**	**S**	**I**	**G**	**Q**	**V**	**L**	**T**	**V**	**F**	**F**	**S**	**V**	**L**	**I**	**G**	**A**	**F**	**S**
WT	961	GAA	TAC	TCT	ATT	GGA	CAA	GTG	CTC	ACT	GTC	TTC	TTT	TCC	GTG	TTA	ATT	GGA	GCA	TTC	AGT
CO	961	GAG	TAC	TCC	ATC	GGT	CAG	GTT	TTG	ACT	GTT	TTC	TTC	TCT	GTT	TTG	ATT	GGT	GCT	TTC	TCC

		V	G	Q	A	S	P	N	I	E	A	F	A	N	A	R	G	A	A	Y	E
WT	1021	GTT	GGA	CAG	GCA	TCT	CCA	AAT	ATT	GAA	GCC	TTC	GCC	AAT	GCA	CGA	GGA	GCA	GCT	TAT	GAA
CO	1021	GTT	GGT	CAG	GCT	TCC	CCA	AAC	ATT	GAA	GCT	TTC	GCT	AAC	GCT	AGA	GGT	GCT	GCT	TAC	GAG
		V	F	K	I	I	D	N	K	P	S	I	D	S	F	S	K	S	G	H	K
WT	1081	GTC	TTC	AAA	ATA	ATT	GAT	AAT	AAG	CCC	AGT	ATA	GAC	AGC	TTC	TCA	AAG	AGT	GGG	CAC	AAA
CO	1081	GTT	TTC	AAG	ATC	ATC	GAC	AAC	AAG	CCA	TCC	ATC	GAC	TCA	TTC	TCT	AAG	TCC	GGA	CAC	AAG
		P	D	N	I	Q	G	N	L	E	F	K	N	I	H	F	S	Y	P	S	R
WT	1141	CCA	GAC	AAC	ATA	CAA	GGA	AAT	CTG	GAA	TTT	AAG	AAT	ATT	CAC	TTC	AGT	TAC	CCA	TCT	CGA
CO	1141	CCA	GAC	AAC	ATC	CAA	GGT	AAC	TTG	GAG	TTC	AAG	AAT	ATC	CAC	TTC	TCC	TAC	CCA	TCC	AGA
		K	E	V	Q	I	L	K	G	L	N	L	K	V	K	S	G	Q	T	V	A
WT	1201	AAA	GAA	GTT	CAG	ATC	TTG	AAG	GGC	CTC	AAT	CTG	AAG	GTG	AAG	AGC	GGA	CAG	ACG	GTG	GCC
CO	1201	AAA	GAG	GTT	CAG	ATC	TTG	AAG	GGA	TTG	AAC	TTG	AAG	GTT	AAG	TCC	GGT	CAG	ACT	GTT	GCT
		L	V	G	N	S	G	C	G	K	S	T	T	V	Q	L	M	Q	R	L	Y
WT	1261	CTG	GTT	GGC	AAC	AGT	GGC	TGT	GGA	AAA	AGC	ACA	ACT	GTC	CAG	CTG	ATG	CAA	AGG	CTC	TAC
CO	1261	TTG	GTT	GGT	AAC	TCT	GGT	TGT	GGT	AAG	TCC	ACT	ACT	GTT	CAG	TTG	ATG	CAG	AGA	TTG	TAC
		D	P	L	D	G	M	V	S	I	D	G	Q	D	I	R	T	I	N	V	R
WT	1321	GAC	CCC	CTA	GAT	GGC	ATG	GTC	AGT	ATC	GAC	GGA	CAG	GAC	ATC	AGA	ACC	ATC	AAT	GTG	AGG
CO	1321	GAC	CCA	TTG	GAC	GGT	ATG	GTT	TCC	ATT	GAC	GGT	CAG	GAC	ATC	AGA	ACT	ATC	AAT	GTT	AGA
		Y	L	R	E	I	I	G	V	V	S	Q	E	P	V	L	F	A	T	T	I
WT	1381	TAT	CTG	AGG	GAG	ATC	ATT	GGT	GTG	GTG	AGT	CAG	GAA	CCT	GTG	CTG	TTT	GCC	ACC	ACG	ATC
CO	1381	TAC	TTG	AGA	GAG	ATC	ATC	GGA	GTT	GTT	TCC	CAA	GAG	CCA	GTT	TTG	TTC	GCT	ACT	ACT	ATC
		A	E	N	I	R	Y	G	R	E	D	V	T	M	D	E	I	E	K	A	V
WT	1441	GCC	GAG	AAC	ATT	CGC	TAT	GGC	CGA	GAA	GAT	GTC	ACC	ATG	GAT	GAG	ATT	GAG	AAA	GCT	GTC
CO	1441	GCT	GAG	AAC	ATC	AGA	TAC	GGT	AGA	GAA	GAT	GTT	ACT	ATG	GAC	GAG	ATC	GAG	AAG	GCT	GTT
		K	E	A	N	A	Y	D	F	I	M	K	L	P	H	Q	F	D	T	L	V
WT	1501	AAG	GAA	GCC	AAT	GCC	TAT	GAC	TTC	ATC	ATG	AAA	CTG	CCC	CAC	CAA	TTT	GAC	ACC	CTG	GTT
CO	1501	AAG	GAA	GCT	AAC	GCT	TAC	GAC	TTT	ATC	ATG	AAG	TTG	CCA	CAC	CAG	TTC	GAC	ACA	TTG	GTT
		G	E	R	G	A	Q	L	S	G	G	Q	K	Q	R	I	A	I	A	R	A
WT	1561	GGT	GAG	AGA	GGG	GCG	CAG	CTG	AGT	GGG	GGA	CAG	AAA	CAG	AGA	ATC	GCC	ATT	GCC	CGG	GCC
CO	1561	GGT	GAA	AGA	GGT	GCT	CAA	TTG	TCC	GGT	GGT	CAA	AAA	CAA	AGA	ATT	GCT	ATC	GCT	AGA	GCT

		L	V	R	N	P	K	I	L	L	L	D	E	A	T	S	A	L	D	T	E
WT	1621	CTG	GTC	CGC	AAT	CCC	AAG	ATC	CTT	TTG	TTG	GAC	GAG	GCC	ACC	TCA	GCC	CTG	GAT	ACA	GAA
CO	1621	TTG	GTT	AGA	AAC	CCA	AAG	ATT	TTG	TTG	TTG	GAC	GAG	GCT	ACT	TCC	GCT	TTG	GAC	ACT	GAA
		S	E	A	V	V	Q	A	A	L	D	K	A	R	E	G	R	T	T	I	V
WT	1681	AGT	GAA	GCT	GTG	GTT	CAG	GCC	GCA	CTG	GAT	AAG	GCT	AGA	GAA	GGC	CGG	ACC	ACC	ATT	GTG
CO	1681	TCT	GAG	GCT	GTT	GTT	CAA	GCT	GCT	TTG	GAT	AAG	GCT	AGA	GAG	GGT	AGA	ACT	ACA	ATC	GTT
		I	A	H	R	L	S	T	V	R	N	A	D	V	I	A	G	F	D	G	G
WT	1741	ATA	GCT	CAT	CGC	TTG	TCT	ACC	GTT	CGT	AAT	GCT	GAC	GTC	ATT	GCT	GGT	TTT	GAT	GGT	GGT
CO	1741	ATC	GCT	CAC	AGA	TTG	TCC	ACT	GTT	AGA	AAC	GCT	GAC	GTT	ATT	GCT	GGT	TTC	GAC	GGT	GGT
		V	I	V	E	Q	G	N	H	D	E	L	M	R	E	K	G	I	Y	F	K
WT	1801	GTC	ATT	GTG	GAG	CAA	GGA	AAT	CAT	GAT	GAG	CTC	ATG	AGA	GAA	AAG	GGC	ATT	TAC	TTC	AAA
CO	1801	GTT	ATC	GTT	GAG	CAA	GGT	AAC	CAC	GAC	GAA	TTG	ATG	AGA	GAG	AAG	GGT	ATC	TAC	TTC	AAG
		L	V	M	T	Q	T	A	G	N	E	I	E	L	G	N	E	A	C	K	S
WT	1861	CTT	GTC	ATG	ACA	CAG	ACA	GCA	GGA	AAT	GAA	ATT	GAA	TTA	GGA	AAT	GAA	GCT	TGT	AAA	TCT
CO	1861	TTG	GTT	ATG	ACT	CAG	ACT	GCT	GGT	AAC	GAG	ATT	GAG	TTG	GGT	AAC	GAG	GCT	TGT	AAG	TCC
		K	D	E	I	D	N	L	D	M	S	S	K	D	S	G	S	S	L	I	R
WT	1921	AAG	GAT	GAA	ATT	GAT	AAT	TTA	GAC	ATG	TCT	TCA	AAA	GAT	TCA	GGA	TCC	AGT	CTA	ATA	AGA
CO	1921	AAG	GAC	GAA	ATC	GAT	AAC	TTG	GAT	ATG	TCC	TCC	AAG	GAC	TCT	GGA	TCC	TCC	TTG	ATC	AGA
		R	R	S	T	R	K	S	I	C	G	P	H	D	Q	D	R	K	L	S	T
WT	1981	AGA	AGA	TCA	ACT	CGC	AAA	AGC	ATC	TGT	GGA	CCA	CAT	GAC	CAA	GAC	AGG	AAG	CTT	AGT	ACC
CO	1981	AGA	AGA	TCC	ACT	AGA	AAG	TCC	ATC	TGT	GGT	CCA	CAC	GAC	CAG	GAC	AGA	AAG	TTG	TCT	ACA
		K	E	A	L	D	E	D	V	P	P	A	S	F	W	R	I	L	K	L	N
WT	2041	AAA	GAG	GCC	CTG	GAT	GAA	GAT	GTA	CCT	CCA	GCT	TCC	TTT	TGG	CGG	ATC	CTG	AAG	TTG	AAT
CO	2041	AAA	GAG	GCT	TTG	GAC	GAG	GAC	GTT	CCA	CCA	GCT	TCA	TTT	TGG	AGA	ATC	TTG	AAG	TTG	AAC
		S	T	E	W	P	Y	F	V	V	G	I	F	C	A	I	I	N	G	G	L
WT	2101	TCA	ACT	GAA	TGG	CCT	TAT	TTT	GTG	GTT	GGT	ATA	TTC	TGT	GCC	ATA	ATA	AAT	GGA	GGC	TTA
CO	2101	TCT	ACT	GAG	TGG	CCA	TAC	TTC	GTT	GTT	GGT	ATC	TTC	TGT	GCT	ATT	ATC	AAC	GGT	GGA	TTG
		Q	P	A	F	S	V	I	F	S	K	V	V	G	V	F	T	N	G	G	P
WT	2161	CAG	CCA	GCA	TTC	TCC	GTA	ATA	TTT	TCA	AAA	GTT	GTA	GGG	GTT	TTT	ACA	AAT	GGT	GGC	CCC
CO	2161	CAG	CCA	GCT	TTC	TCT	GTT	ATT	TTC	TCC	AAG	GTT	GTT	GGT	GTT	TTC	ACT	AAC	GGT	GGT	CCA

		P	E	T	Q	R	Q	N	S	N	L	F	S	L	L	F	L	I	L	G	I
WT	2221	CCT	GAA	ACC	CAG	CGG	CAG	AAC	AGC	AAC	TTG	TTT	TCC	TTG	TTG	TTT	CTG	ATC	CTT	GGG	ATC
CO	2221	CCA	GAG	ACT	CAA	AGA	CAG	AAC	TCT	AAC	TTG	TTC	TCC	TTG	TTG	TTC	TTG	ATC	TTG	GGA	ATC
		I	S	F	I	T	F	F	L	Q	G	F	T	F	G	K	A	G	E	I	L
WT	2281	ATT	TCT	TTC	ATT	ACA	TTT	TTT	CTT	CAG	GGC	TTC	ACA	TTT	GGC	AAA	GCT	GGA	GAG	ATC	CTC
CO	2281	ATC	TCC	TTC	ATC	ACA	TTT	TTC	TTG	CAG	GGT	TTC	ACT	TTC	GGT	AAA	GCT	GGT	GAG	ATC	TTG
		T	K	R	L	R	Y	M	V	F	K	S	M	L	R	Q	D	V	S	W	F
WT	2341	ACC	AAG	CGA	CTC	CGA	TAC	ATG	GTT	TTC	AAA	TCC	ATG	CTG	AGA	CAG	GAT	GTG	AGC	TGG	TTT
CO	2341	ACA	AAA	AGA	TTG	AGA	TAT	ATG	GTT	TTT	AAG	TCC	ATG	TTG	AGA	CAG	GAC	GTT	TCC	TGG	TTC
		D	D	P	K	N	T	T	G	A	L	T	T	R	L	A	N	D	A	A	Q
WT	2401	GAT	GAC	CCT	AAA	AAC	ACC	ACC	GGA	GCA	CTG	ACC	ACC	AGG	CTC	GCC	AAC	GAT	GCT	GCT	CAA
CO	2401	GAT	GAC	CCA	AAG	AAC	ACT	ACT	GGT	GCT	TTG	ACT	ACT	AGA	TTG	GCT	AAC	GAT	GCT	GCT	CAG
		V	K	G	A	T	G	S	R	L	A	V	I	F	Q	N	I	A	N	L	G
WT	2461	GTG	AAA	GGG	GCT	ACA	GGG	TCT	AGG	CTT	GCT	GTG	ATT	TTC	CAG	AAC	ATA	GCA	AAT	CTT	GGG
CO	2461	GTT	AAG	GGT	GCT	ACT	GGT	TCC	AGA	TTG	GCT	GTT	ATC	TTC	CAG	AAC	ATT	GCT	AAC	TTG	GGT
		T	G	I	I	I	S	L	I	Y	G	W	Q	L	T	L	L	L	L	A	I
WT	2521	ACA	GGA	ATC	ATC	ATA	TCC	CTA	ATC	TAT	GGC	TGG	CAA	CTA	ACA	CTT	TTA	CTC	TTA	GCA	ATT
CO	2521	ACT	GGT	ATC	ATC	ATT	TCC	TTG	ATC	TAT	GGA	TGG	CAA	TTG	ACT	CTT	TTG	TTG	TTG	GCT	ATC
		V	P	I	I	A	I	A	G	V	V	E	M	K	M	L	S	G	Q	A	L
WT	2581	GTA	CCC	ATC	ATT	GCG	ATA	GCA	GGA	GTG	GTT	GAA	ATG	AAA	ATG	TTG	TCT	GGA	CAA	GCA	CTG
CO	2581	GTT	CCA	ATT	ATC	GCT	ATC	GCT	GGT	GTT	GTT	GAG	ATG	AAG	ATG	TTG	TCC	GGT	CAA	GCT	TTG
		K	D	K	K	E	L	E	G	S	G	K	I	A	T	E	A	I	E	N	F
WT	2641	AAA	GAT	AAG	AAG	GAA	CTA	GAA	GGT	TCT	GGA	AAG	ATT	GCT	ACG	GAA	GCA	ATT	GAA	AAC	TTC
CO	2641	AAG	GAC	AAA	AAA	GAA	TTG	GAG	GGT	TCC	GGT	AAG	ATC	GCT	ACT	GAG	GCT	ATT	GAG	AAC	TTC
		R	T	V	V	S	L	T	R	E	Q	K	F	E	T	M	Y	A	Q	S	L
WT	2701	CGC	ACT	GTT	GTC	TCT	TTG	ACT	CGG	GAG	CAG	AAG	TTT	GAA	ACC	ATG	TAT	GCC	CAG	AGC	TTG
CO	2701	AGA	ACT	GTT	GTT	TCC	TTG	ACT	AGA	GAG	CAA	AAG	TTC	GAG	ACT	ATG	TAC	GCT	CAG	TCC	TTG
		Q	I	P	Y	R	N	A	M	K	K	A	H	V	F	G	I	T	F	S	F
WT	2761	CAG	ATA	CCA	TAC	AGA	AAT	GCG	ATG	AAG	AAA	GCA	CAC	GTG	TTT	GGG	ATC	ACG	TTC	TCC	TTC
CO	2761	CAG	ATC	CCA	TAC	AGA	AAC	GCT	ATG	AAG	AAG	GCT	CAC	GTT	TTC	GGT	ATC	ACT	TTC	TCA	TTC

		T	**Q**	**A**	**M**	**M**	**Y**	**F**	**S**	**Y**	**A**	**A**	**C**	**F**	**R**	**F**	**G**	**A**	**Y**	**L**	**V**
WT	2821	ACC	CAG	GCC	ATG	ATG	TAT	TTT	TCT	TAT	GCT	GCT	TGT	TTC	CGG	TTC	GGT	GCC	TAC	TTG	GTG
CO	2821	ACA	CAG	GCT	ATG	ATG	TAC	TTC	TCT	TAC	GCT	GCT	TGT	TTC	AGA	TTC	GGT	GCT	TAC	TTG	GTT
		T	**Q**	**Q**	**L**	**M**	**T**	**F**	**E**	**N**	**V**	**L**	**L**	**V**	**F**	**S**	**A**	**I**	**V**	**F**	**G**
WT	2881	ACA	CAA	CAA	CTC	ATG	ACT	TTT	GAA	AAT	GTT	CTG	TTA	GTA	TTC	TCA	GCT	ATT	GTC	TTT	GGT
CO	2881	ACT	CAG	CAG	TTG	ATG	ACT	TTC	GAG	AAT	GTT	TTG	TTG	GTT	TTC	TCA	GCT	ATC	GTT	TTC	GGA
		A	**M**	**A**	**V**	**G**	**Q**	**V**	**S**	**S**	**F**	**A**	**P**	**D**	**Y**	**A**	**K**	**A**	**T**	**V**	**S**
WT	2941	GCC	ATG	GCA	GTG	GGG	CAG	GTC	AGT	TCA	TTC	GCT	CCT	GAC	TAT	GCG	AAA	GCC	ACA	GTG	TCA
CO	2941	GCT	ATG	GCT	GTT	GGA	CAG	GTT	TCC	TCT	TTT	GCT	CCT	GAC	TAC	GCT	AAA	GCT	ACT	GTT	TCC
		A	**S**	**H**	**I**	**I**	**R**	**I**	**I**	**E**	**K**	**T**	**P**	**E**	**I**	**D**	**S**	**Y**	**S**	**T**	**Q**
WT	3001	GCA	TCC	CAC	ATC	ATC	AGG	ATC	ATT	GAG	AAA	ACC	CCC	GAG	ATT	GAC	AGC	TAC	AGC	ACG	CAA
CO	3001	GCT	TCC	CAC	ATC	ATC	AGA	ATC	ATC	GAA	AAG	ACT	CCA	GAG	ATC	GAC	TCC	TAC	TCC	ACT	CAA
		G	**L**	**K**	**P**	**N**	**M**	**L**	**E**	**G**	**N**	**V**	**Q**	**F**	**S**	**G**	**V**	**V**	**F**	**N**	**Y**
WT	3061	GGC	CTA	AAG	CCG	AAT	ATG	TTG	GAA	GGA	AAT	GTG	CAA	TTT	AGT	GGA	GTC	GTG	TTC	AAC	TAT
CO	3061	GGT	TTG	AAG	CCA	AAC	ATG	TTG	GAG	GGT	AAC	GTT	CAG	TTC	TCC	GGT	GTT	GTT	TTC	AAC	TAC
		P	**T**	**R**	**P**	**S**	**I**	**P**	**V**	**L**	**Q**	**G**	**L**	**S**	**L**	**E**	**V**	**K**	**K**	**G**	**Q**
WT	3121	CCC	ACC	CGA	CCC	AGC	ATC	CCA	GTG	CTT	CAG	GGG	CTG	AGC	CTT	GAG	GTG	AAG	AAG	GGC	CAG
CO	3121	CCA	ACT	AGA	CCA	TCC	ATC	CCT	GTT	TTG	CAA	GGT	TTG	TCT	TTG	GAG	GTT	AAG	AAG	GGT	CAG
		T	**L**	**A**	**L**	**V**	**G**	**S**	**S**	**G**	**C**	**G**	**K**	**S**	**T**	**V**	**V**	**Q**	**L**	**L**	**E**
WT	3181	ACG	CTG	GCC	CTG	GTG	GGC	AGC	AGT	GGC	TGC	GGG	AAG	AGC	ACA	GTG	GTC	CAG	CTG	CTC	GAG
CO	3181	ACA	TTG	GCT	TTG	GTT	GGA	TCT	TCC	GGT	TGT	GGA	AAA	TCC	ACA	GTT	GTT	CAG	TTG	TTG	GAA
		R	**F**	**Y**	**D**	**P**	**M**	**A**	**G**	**S**	**V**	**F**	**L**	**D**	**G**	**K**	**E**	**I**	**K**	**Q**	**L**
WT	3241	CGC	TTC	TAC	GAC	CCC	ATG	GCT	GGA	TCA	GTG	TTT	CTA	GAT	GGC	AAA	GAA	ATA	AAG	CAA	CTG
CO	3241	AGA	TTC	TAT	GAC	CCT	ATG	GCT	GGA	TCC	GTT	TTC	TTG	GAT	GGT	AAA	GAG	ATC	AAG	CAG	TTG
		N	**V**	**Q**	**W**	**L**	**R**	**A**	**Q**	**L**	**G**	**I**	**V**	**S**	**Q**	**E**	**P**	**I**	**L**	**F**	**D**
WT	3301	AAT	GTC	CAG	TGG	CTC	CGA	GCA	CAG	CTG	GGC	ATT	GTG	TCC	CAA	GAG	CCC	ATT	CTC	TTT	GAC
CO	3301	AAT	GTT	CAG	TGG	TTG	AGA	GCT	CAG	TTG	GGT	ATC	GTT	TCT	CAA	GAA	CCT	ATC	TTG	TTC	GAC
		C	**S**	**I**	**A**	**E**	**N**	**I**	**A**	**Y**	**G**	**D**	**N**	**S**	**R**	**V**	**V**	**S**	**Y**	**E**	**E**
WT	3361	TGC	AGC	ATC	GCA	GAG	AAC	ATT	GCC	TAC	GGA	GAC	AAC	AGC	CGG	GTC	GTG	TCT	TAT	GAG	GAG
CO	3361	TGT	TCC	ATT	GCT	GAG	AAT	ATT	GCT	TAC	GGT	GAC	AAC	TCC	AGA	GTT	GTT	TCT	TAC	GAA	GAG
		I	**V**	**R**	**A**	**A**	**K**	**E**	**A**	**N**	**I**	**H**	**Q**	**F**	**I**	**D**	**S**	**L**	**P**	**D**	**K**

WT	3421	ATT	GTG	AGG	GCA	GCC	AAG	GAG	GCC	AAC	ATC	CAC	CAG	TTC	ATC	GAC	TCG	CTA	CCT	GAT	AAA
CO	3421	ATC	GTT	AGA	GCT	GCT	AAA	GAG	GCT	AAC	ATT	CAC	CAG	TTC	ATT	GAT	TCC	TTG	CCA	GAC	AAG
		Y	**N**	**T**	**R**	**V**	**G**	**D**	**K**	**G**	**T**	**Q**	**L**	**S**	**G**	**G**	**Q**	**K**	**Q**	**R**	**I**
WT	3481	TAC	AAC	ACC	AGA	GTA	GGA	GAC	AAA	GGC	ACT	CAG	CTG	TCG	GGT	GGG	CAG	AAG	CAG	CGC	ATC
CO	3481	TAC	AAC	ACT	AGA	GTT	GGT	GAC	AAG	GGA	ACT	CAG	TTG	TCA	GGT	GGA	CAG	AAG	CAG	AGA	ATC
		A	**I**	**A**	**R**	**A**	**L**	**V**	**R**	**Q**	**P**	**H**	**I**	**L**	**L**	**L**	**D**	**E**	**A**	**T**	**S**
WT	3541	GCC	ATC	GCA	CGC	GCC	CTC	GTC	AGA	CAG	CCT	CAC	ATT	TTA	CTT	CTG	GAC	GAA	GCA	ACA	TCA
CO	3541	GCT	ATT	GCT	AGA	GCT	TTG	GTT	AGA	CAG	CCA	CAC	ATC	TTG	TTG	TTG	GAT	GAA	GCT	ACA	TCT
		A	**L**	**D**	**T**	**E**	**S**	**E**	**K**	**V**	**V**	**Q**	**E**	**A**	**L**	**D**	**K**	**A**	**R**	**E**	**G**
WT	3601	GCT	CTG	GAT	ACA	GAA	AGT	GAA	AAG	GTT	GTC	CAG	GAA	GCG	CTG	GAC	AAA	GCC	AGG	GAA	GGC
CO	3601	GCT	TTG	GAT	ACT	GAG	TCC	GAG	AAG	GTT	GTT	CAA	GAA	GCT	TTG	GAC	AAA	GCT	AGA	GAA	GGT
		R	**T**	**C**	**I**	**V**	**I**	**A**	**H**	**R**	**L**	**S**	**T**	**I**	**Q**	**N**	**A**	**D**	**L**	**I**	**V**
WT	3661	CGC	ACC	TGC	ATT	GTG	ATC	GCT	CAC	CGC	CTG	TCC	ACC	ATC	CAG	AAC	GCG	GAC	TTG	ATC	GTG
CO	3661	AGA	ACA	TGT	ATC	GTT	ATT	GCT	CAT	AGA	TTG	TCT	ACT	ATC	CAG	AAC	GCT	GAC	TTG	ATC	GTT
		V	**I**	**Q**	**N**	**G**	**K**	**V**	**K**	**E**	**H**	**G**	**T**	**H**	**Q**	**Q**	**L**	**L**	**A**	**Q**	**K**
WT	3721	GTG	ATT	CAG	AAC	GGC	AAG	GTC	AAG	GAG	CAC	GGC	ACC	CAC	CAG	CAG	CTG	CTG	GCG	CAG	AAG
CO	3721	GTT	ATT	CAG	AAC	GGT	AAG	GTT	AAG	GAA	CAC	GGT	ACT	CAC	CAG	CAG	TTG	TTG	GCT	CAA	AAG
		G	**I**	**Y**	**F**	**S**	**M**	**V**	**S**	**V**	**Q**	**A**	**G**	**A**	**K**	**R**	**S**				
WT	3781	GGC	ATC	TAC	TTC	TCA	ATG	GTC	AGT	GTG	CAG	GCT	GGA	GCA	AAG	CGC	TCA				
CO	3781	GGA	ATC	TAC	TTT	TCC	ATG	GTT	TCC	GTT	CAA	GCT	GGT	GCT	AAG	AGA	TCC				