

The Basis of Substrate Promiscuity in Membrane Protein Drug Transporters Using the Small Multidrug Resistance (SMR) Family as Simple Models

by

Olive Escalicas Burata

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2023

Doctoral Committee:

Associate Professor Randy B. Stockbridge, Chair
Assistant Professor Ryan D. Baldridge
Professor Matthew R. Chapman
Professor Melanie D. Ohi

Olive Escalicas Burata

olib@umich.edu

ORCID iD: 0000-0002-8450-8930

© Olive Escalicas Burata 2023

Dedication

To all the female figures in my life whom I have been inspired and blessed to learn from.

To my late grandmother, Parfiria Abarquez Escalicas, who pushed me to see my dreams through.

Acknowledgements

This section to me was the most difficult to write not because of the sheer volume of individuals who have helped push me through this journey but also the quality of involvement they provided out of the limited times in their days, weeks, months, years just for my own progression. These individuals have shaped me, how I act and how I think. They have consistently humbled me and provided guidance in ways that I could not have ever imagined. Because of this, I have become a better version of my previous self in every grateful interaction I have had.

First, I would like to thank Dr. Alberto A. Rascon, Jr. who showed a different door to scientific research for someone who looks like me. To Mama Flor, who with Dr. R, provided another family for me at my lowest point. They accepted me for who I am and helped me focused on my strengths as I grew as a scientist. To the Burata family, for the constant, unwavering support despite not knowing full well what sort of sacrifices is required to become a scientist. To the Lucero family, for taking me into their family and loving me unconditionally as another daughter. I have gone through a lot of ups and downs with these individuals, but they stayed by my side no matter what, thank you so much! I know it's not easy to handle someone like me, but I am very privileged to have three families looking over me and just know that my thoughts constantly include the betterment of my family in every decision I make every day.

To Randy, who has provided a space for me to be intellectually challenged and be innovative. Thank you for always listening to my crazy ideas even though over half of them

never worked out. Thank you for constantly pushing me to my best every day, for being a great listener to my random drop-ins, helping me become a great scientific communicator, and challenging me to become a better leader. Many of my fears and weaknesses as an early scientist were overcome throughout my PhD study and I wouldn't have gotten far without your mentorship, thank you!

To the Stockbridge lab, it's been a great pleasure with you all. I have learned so much from the many conversations and laughs I've had with the various individuals who have joined our lab. The post-docs that I have learned from: Dr. Ben McIlwain, Dr. Ali Kermani, Dr. Jason Devlin, Dr. Achala Chittor, Dr. Aditya Banerjee. The graduate students (both permanent and rotating) I have had the pleasure of meeting and working with: Kamirah Demouchet, Rachael Lucero, Joseph Loomis, Fox Baudelaire, Chia-Yu Kang, Trevor Yeh, Saffron Little, Carla Peralta, Alex Missman, Katherine Wentworth, and Kaitrin Funckes, To the many undergraduates who I worked with: Vivek Parikh, Alex Davis, Alain Sullivan, Michal Ruprecht, Troy Cao, Brett Dash, Kemal Demirer, Minjun An, Sabrina Kolb, Allen Zhu. To Ben Koff and Isabella Reacher, our techs who have assisted me during our many purification campaigns. Special shoutout to the undergrads who have worked directly with me and have endured this crazy journey I call a scientific mentorship: Victoria 'Ever' O'Donnell, who people have mentioned is a carbon copy of me, but I hope will be better than me! Started from a very limited experience in scientific research, with very little support and guidance from her family. She overcame a lot of challenges in just a short amount of time, got accepted into many high-ranking PhD schools, even got a personal email from Dr. Francis Arnold herself, and is now a PhD graduate student in UCSF's Biophysics program. I am very proud of you and the journey that you have had to go through to get this far, and I hope you continue to see how far you've come. Jeonghoon Hyun, my second

undergrad who started off with shaky hands when pipetting and ended up becoming a leader on his part of the project after two years with me. It's been amazing seeing you transform into a confident scientist. I know it was very obvious that I was trying to get you to join the dark side (aka PhD school) but I know you'll also do great in Dentistry school! Last, but not least to Ethan Gibbs, thank you for bringing your lively presence into the last bit of my PhD study. It has been super refreshing and super fun to be consistently on my toes when it comes to discussing science with you. Although you're still early in your undergraduate career, know that I too have learned so much from our conversations. I look forward to hearing about your future success, I know it's coming! Altogether I have had high expectations of these three, to always come to lab prepared and to be intellectually challenged, they have always made me proud by coming to lab matching my own enthusiasm. Thank you guys!

To my committee members, Dr. Matt Chapman and Dr. Ryan Baldrige, both of whom have welcomed me during recruitment week at UMich and to Dr. Melanie Ohi, who I didn't get to meet until she became a part of my dissertation committee but became one of the most instrumental career mentors in my PhD studies. Thank you for pushing me to the great opportunities that awaited me after PhD and for believing in me to go far!

To my Wolverine Venture Fund family: Prof. Eric Gordon, Everardo, Jill, Maria, and TJ. You guys are amazing and thank you for being so patient with me as I learned the fundamentals of business and investing. To my Nucleate Michigan family, Desnor, Shree, and Akash, thank for your helping hand and guidance in ensuring we get the resources we need. Thank you for paving the way in building up the entrepreneurial ecosystem here in Michigan, especially for us students. To Tyler, my partner in crime throughout our journey in the activator program in Nucleate Michigan and co-founder of FluorGent. It has been amazing to have met you, sir!

To my colleagues, especially my cohort: Yanira, Junius, Mandi, Nolan, Maria, Sarah, Alaina, Saffron, Will, Josh, Annie, Taslima, and Rachael. Shoutout to my PCB big brother/sisters who made sure I didn't go too far off the rails and welcomed me to Ann Arbor: Dr. Anthony Balistreri, Dr. Hannah Chia, and Dr. Monika Franco. To Dr. Arti Dumbrepatil (Marsh Lab), Dr. Alex Vizurraga (Tall Lab), Dr. Victor Sosa (Lehnert Lab), Dr. April Lukowski (Narayan Lab) for the scientific collaborations and mentorship you have provided me early on in my PhD studies.

Thank you to my friends who have made the PhD journey bearable. Thank you for the laughs, the check-ins, the tears, support, and motivation (aka do better, Olive! crew). The Masters crew from SJSU: Diane, Angelina, Rebecca, Thanh, Lauren, Anthony, Saira, Carla & James. To the irreplaceable individuals who have made me better every day. Every conversation I have had with these folks is the therapy I sorely needed on some days: Saffron Little, Junius Thomas, Chia-Yu Kang, & Jaimin Rana.

To my wonderful wife, my partner, Dr. Rachael Lucero. 'Thank you for being my constant support' is such a big understatement. Thank you for being my better half and pushing me every day to be better. You're there by my side even when you're dealing with your own challenges from PhD life, and yet you never falter when I ask for help. You've no idea how proud I am of you since we first started our scientific journey together, and I know I joke a lot and never want to get in my feelings, but I am very excited of our future. I am very lucky to be married to you and I know we'll move mountains together in the future. I love you so much!

Table of Contents

Dedication.....	ii
Acknowledgements	iii
List of Tables	xii
List of Figures.....	xiii
List of Appendices.....	xviii
Abstract.....	xix
Chapter 1 Introduction.....	1
1.1 Addressing the ambiguities of promiscuity	1
1.2 Significance of substrate promiscuity in transporters	2
1.3 Why the SMRs?.....	2
1.4 A Closer at the SMR Family: ‘Still Rocking in the Structural Era: A Molecular Overview of the Small Multidrug Resistance (SMR) Transporter Family’	3
1.4.1 Introduction	4
1.4.2 SMR Family Topology	7
1.4.3 Distribution of SMR genes among bacterial genomes	8
1.4.4 Guanidinium exporters	12
1.4.5 Drug and antiseptic exporters (Qacs)	18
1.4.6 Polyamine transporters	27
1.4.7 Lipid transport proteins	28
1.4.8 Structural relationship between the SMR and the SLC35/DMT folds.....	29
1.4.9 Conclusions and perspective	33

1.5 References.....	34
Chapter 2 Structural Basis of Promiscuity in Small Multidrug Transporters	45
2.1 Introduction	45
2.1.1 Personal contributions to the published researched article.....	45
2.1.2 General overview.....	47
2.2 Results	51
2.2.1 Overlapping, promiscuous substrate transport by Qac and Gdx subtypes	51
2.2.2 Crystal structure of Gdx-Clo	55
2.2.3 The structural basis for conformational exchange.....	58
2.2.4 The substrate binding site.....	59
2.2.5 A membrane portal accommodates hydrophobic substrate substituents.....	63
2.3 Results	64
2.4 Methods	65
2.4.1 Sequence-similarity network.....	65
2.4.2 Transport expression, purification, and proteoliposome reconstitution.....	66
2.4.3 Radioactive flux assays.....	67
2.4.4 SSM electrophysiology.....	67
2.4.5 Monobody development.....	68
2.4.6 Monobody expression and purification.....	69
2.4.7 Crystal preparation.....	70
2.4.8 Structure determination.....	71
2.5 Acknowledgments	72
2.6 Competing Interests.....	73
2.7 References	73
Chapter 3 Crystal Structures of Bacterial Small Multidrug Resistance Transporter EmrE in Complex With Structurally Diverse Substrates.....	77

3.1 Introduction	79
3.1.1 Personal contributions to the published research article	79
3.1.2 General overview	81
3.2 Results	83
3.2.1 Engineering of EmrE to introduce a monobody binding site	83
3.2.2 Structure of EmrE3 without ligand at pH 5.2	86
3.2.3 Structures of substrate-bound EmrE ₃	91
3.2.4 Structure of Gdx-Clo at pH 5 and comparison to the substrate binding site of EmrE	94
3.2.5 EmrE is tolerant of mutations that eliminate hydrogen bonding in the binding pocket	98
3.3 Discussion	100
3.3.1 The application of multipurpose chaperones for crystallization	100
3.3.2 Comparison to the NMR model of EmrE S64V	102
3.3.3 Comparison to the NMR model of EmrE S64V	103
3.3.4 movements accommodate diverse substrates	105
3.3.5 Binding of benzalkonium ⁺ and other substrates with alkyl chains	107
3.4 Conclusion	110
3.5 Materials and Methods	110
3.5.1 Bioinformatics and sequence analysis	110
3.5.2 Protein purification and crystallization	111
3.5.3 Structure determination and analysis	112
3.5.4 Microscale thermophoresis	113
3.5.5 SSM electrophysiology	114
3.5.6 NMR chemical shift prediction	115
3.6 Acknowledgements	115

3.7 Competing Interests	115
3.8 References	116
Chapter 4 Necessary and Sufficient Molecular Determinants of Quaternary Ammonium Antiseptic Export by Small Multidrug Resistance Transporters	121
4.1 Introduction	121
4.2 Results	125
4.2.1 Engineering Gdx-Clo for Resistance Against Quaternary Ammonium Antiseptics	125
4.2.2 Gdx-Clo-7X gains the ability to bind and transport quaternary ammoniums	127
4.2.3 Screening a combinatorial library to identify necessary and sufficient mutations for quaternary ammonium resistance	131
4.2.4 Cluster 1 and 2 mutations are essential for TPA ⁺ binding and quaternary ammonium transport, whereas Cluster 3 mutations contribute to faster export kinetics	132
4.3 Discussion.....	135
4.4 Materials and Methods	138
4.4.1 Quaternary Ammonium Resistance Assays	138
4.4.2 Directed evolution	139
4.4.3 Combinatorial Library Construction, Selection Assays, and Illumina Sequencing	139
4.4.4 Protein Purification and Liposome Reconstitution.....	140
4.4.5 Synthesis of NBD-CTA ⁺ (5).....	141
4.4.6 NBD-CTA ⁺ Transport Assay.....	142
4.4.7 Tryptophan fluorescence	143
4.4.8 Solid supported membrane (SSM) electrophysiology.....	143
4.5 References	143
Chapter 5 Conclusions & Future Works	147
5.1 The Ancestral Reconstruction of the SMR Family	148
5.2 References	151

Appendices	152
------------------	-----

List of Tables

Table 1: Structural data and models available for SMR transporters Gdx-Clo and EmrE.....	6
Table 2: Selected scanning mutagenesis studies of EmrE.....	19
Table 3: Data collection, phasing and refinement statistics for Gdx-Clo complexes.....	163
Table 4: Data collection, phasing, and refinement statistics for EmrE and Gdx-Clo complexes.	176
Table 5: SSM electrophysiology peak currents (nA) for EmrE3 and Gdx-Clo mutants summarized by experimental replicate.	177
Table 6: Key resources table	177
Table 7:Generated datasets.....	178
Table 8: Designed oligos for engineering of 128 combinatorial variants of Gdx-Clo 7X.....	186
Table 9: Sequencing statistics for replicate NGS experiments.....	193
Table 10: Correlation between enrichment coefficients of replicate selection experiments.....	193
Table 11: Overrepresented variants (>10-fold) in CTA ⁺ selection	194
Table 12: Overrepresented variants (>2-fold) in TPA ⁺ selection.....	195

List of Figures

Figure 1.1: Proper categorization of promiscuity for transporters. Modified from Atkins 2015	2
Figure 1.2: Transport scheme and transporter topologies	5
Figure 1.3: Sequence conservation of the four major SMR transporter subtypes.....	10
Figure 1.4: Identification and annotation of SMR-coding genes from GEBA genomes.	12
Figure 1.5: Structure and mechanism of Gdx.....	15
Figure 1.6: Substrate binding to EmrE.	21
Figure 1.7: Structural relationship between the SMR and SLC35 folds.	31
Figure 2.1: Colocalization of SMR genes with guanidine riboswitches and horizontal gene transfer elements.....	50
Figure 2.2: Substrate transport of Gdx-Clo and EmrE.	54
Figure 2.3: Substrate transport of Gdx-Clo and EmrE.	57
Figure 2.4: Substrate binding by Gdx-Clo.	62
Figure 3.1: Introduction of monobody binding epitope to EmrE.....	85
Figure 3.2: Crystal structure of EmrE ₃	88
Figure 3.3: Substrate binding to EmrE ₃	93
Figure 3.4: Structure and sequence conservation of substrate binding site residues in Qac and Gdx subtypes.	96
Figure 3.5: Representative SSM electrophysiology recordings for EmrE ₃ and Gdx-Clo mutants.	99
Figure 3.6: Comparisons of NMR and crystallography models of EmrE.	104
Figure 3.7: Hypothetical model of benzalkonium binding to EmrE.	109
Figure 4.1: An engineered variant of Gdx-Clo confers bacterial resistance to quaternary ammonium antiseptics.	127

Figure 4.2:In vitro analysis shows that the substrate binding profile has changed for Gdx-Clo 7X.	128
Figure 4.3:Gdx-Clo 7X transports quaternary ammonium compounds.	129
Figure 4.4:Necessary and sufficient mutants identified from combinatorial library.....	132
Figure 4.5:Functional characterization of the three variant clusters.	134
Figure 4.6: Fitness schematic of Gdx-Clo 7X combinatorial variants in the presence of QACs.	136
Figure 5.1:Preliminary ancestral reconstruction of the SMRs.	150

Supplementary Figure A.1: SMR sequence similarity networks with additional annotation.	153
Supplementary Figure A.2: Sequence alignment of functionally characterized homodimeric SMRs.	154
Supplementary Figure A.3: Representative SSM electrophysiology recordings.	156
Supplementary Figure A.4: Uptake of ^{14}C Gdm $^{+}$ into Gdx-Clo proteoliposomes in exchange for the indicated substrate.	157
Supplementary Figure A.5: Venn diagram showing overlapping transport specificities of Gdx-Clo and EmrE.	158
Supplementary Figure A.6: Experimental electron density maps for Gdx-Clo.	159
Supplementary Figure A.7: Binding interface between monobody Clo-L10 and Gdx-Clo.	160
Supplementary Figure A.8: Currents mediated by Gdx-Clo in the presence and absence of monobody L10.	160
Supplementary Figure A.9: Electron density between E13 and E13' in 3.2 Å structure solved with 10 mM Gdm $^{+}$	161
Supplementary Figure A.10: Surface rendering of exposed TM3 GxxxG motifs.	161
Supplementary Figure A.11: Structural alignment of Gdx-Clo.	162
Supplementary Figure B.1: Crystal lattice for Gdx-Clo/L10 monobody complex (PDB: 6WK8).	164
Supplementary Figure B.2: K_m values for TPA $^{+}$ and PheGdm $^{+}$ transport by EmrE $_3$ (blue) and WT EmrE (red).	165
Supplementary Figure B.3: Representative microscale thermophoresis traces for monobody L10 in the presence of 30 nM – 10 μM EmrE $_3$	165
Supplementary Figure B.4: EmrE $_3$ maps.	166
Supplementary Figure B.5: Structural comparison of EmrE $_3$ crystal structure with electron microscopy maps, theoretical model, and Gdx-Clo.	167
Supplementary Figure B.6: Sidechain density in the EmrE $_3$ binding site.	168
Supplementary Figure B.7: Electron density maps of methyl viologen in different EmrE $_3$ protomers in the asymmetric unit.	168
Supplementary Figure B.8: Electron density maps for W63A modeled in different positions.	169
Supplementary Figure B.9: Electron density maps for E14B modeled in different positions.	170

Supplementary Figure B.10: Gdx-Clo and EmrE substrate binding sites.	171
Supplementary Figure B.11: Sequence alignments of five representative Gdx proteins (from top to bottom: Clostridiales bacterium oral taxon 876, <i>E. coli</i> , <i>Micromonospora</i> , <i>Streptomyces tsukubensis</i> , and <i>Leifsonia aquatica</i>) and five representative Qac proteins (from top to bottom: <i>E. coli</i> , <i>Klebsiella pneumoniae</i> , <i>Pseudomonas aeruginosa</i> , <i>Mycobacterium bovis</i> , and <i>Bordetella avium</i>).	172
Supplementary Figure B.12: (A) Crystal structure of EmrE ₃ (orange and blue cartoon) overlaid with experimental electron microscopy density (cyan mesh contoured at 1.5 σ) (Ubarretxena-Belandia et al., 2003).	173
Supplementary Figure B.13: Comparison of experimental chemical shifts for EmrE (BMRB accession number 50411) with chemical shifts predicted from the crystallography model and NMR ensemble using LARMORC α (Frank et al., 2015).	174
Supplementary Figure B.14: Top down structures of EmrE in complex with benzyltrimethylammonium (PDB:7T00; model for benzalkonium headgroup binding) and Gdx-Clo in complex with octylguanidinium (PDB:6WK9; model for alkyl tail positioning). ...	175
Supplementary Figure C.1: No-drug control for bacterial dilution assays in Figure 4A.	179
Supplementary Figure C.2: Spotting assay of individual mutations found in Gdx-Clo 7X variant when grown in 120 μ M CTA ⁺	180
Supplementary Figure C.3: FPLC profiles for Gdx-Clo WT and Gdx-Clo-7X purification	181
Supplementary Figure C.4: Second independent protein preparation of Gdx-Clo-7X: binding to Gdm ⁺ , CTA ⁺ , or TPA ⁺ , as indicated.	182
Supplementary Figure C.5: Solid supported membrane electrophysiology: no protein controls.	183
Supplementary Figure C.6: LCMS analysis of NBD-CTA ⁺ synthesis.	184
Supplementary Figure C.7: NMR analysis of NBD-CTA ⁺ synthesis	185
Supplementary Figure C.8: Heat maps showing raw counts for combinatorial library with no drug (also shown in Figure 4), CTA ⁺ selection, and TPA ⁺ selection.	187
Supplementary Figure C.9: Correlations between replicate selection experiments.	188
Supplementary Figure C.10: Dilution assays for 5X, 6X, and 7X variants.	189
Supplementary Figure C.11: Dilution assays of Gdx-Clo cluster 1, cluster 2, and cluster 3 mutations: no-drug control (left) and 120 μ M CTA ⁺ (right).	190
Supplementary Figure C.12: FPLC profiles of mutant Gdx-Clo proteins used for functional analysis: cluster 1, cluster 2, cluster 3, cluster 1+2 (4X), 5X, 6X.	191

Supplementary Figure C.13: Tryptophan fluorescence spectra as a function of TPA⁺ titration. 192

List of Appendices

Appendix A: Supporting Information for Chapter 2	153
Appendix B: Supporting Information for Chapter 3	164
Appendix C: Supporting Information for Chapter 4	179

Abstract

This dissertation showcases the mechanisms underlying substrate promiscuity in the Small Multidrug Resistance (SMR) proteins, pivotal in microbial resistance to biocides. This body of work focuses on the structural and functional dissection of the two major SMR subtypes: the guanidinium-specific exporters (SMR_{Gdx}) and the quaternary ammonium compound exporters (SMR_{Qac}). Our primary models are SMR_{Qac} EmrE from *E. coli* and SMR_{Gdx} Gdx-Clo from *Clostridiales* sp., which share a conserved architecture yet exhibit distinct substrate specificities. Initial comparative analysis of EmrE and Gdx-Clo identified critical binding site residues, informing the design of Gdx-Clo variants with altered substrate preferences. Through a combination of rational design and directed evolution, I engineered Gdx-Clo-7X, a Gdx variant capable of exporting quaternary ammonium compounds, a function native to EmrE in the SMR family. The engineered Gdx-Clo-7X contains seven mutations, conferring bacterial resistance and functional transport of quaternary ammonium antiseptics, a significant retreat from its native guanidinium-specific transport. Functional assays revealed that majority of these mutations expand the substrate binding pocket and disrupt the strict hydrogen bond network essential for Gdx-Clo's selectivity. Subsequent screenings with a combinatorial library of Gdx-Clo-7X variants identified mutations indispensable for quaternary ammonium resistance, underscoring the multifaceted contributions of individual residues to the transporter's specificity shift. The findings highlighted in this thesis indicate that promiscuity in SMR transporters is not dictated by single mutations but a concert of changes that collectively redefine substrate specificity. This

study not only provides insights into the evolutionary mechanics of substrate polyspecificity but also establishes a methodological framework for engineering transport proteins with desired transport capabilities. The implications of this work extend beyond a molecular understanding of SMR transporters. By delineating how promiscuous transporters like EmrE handle diverse substrates, strategies can be devised to combat antimicrobial resistance. Furthermore, the principles gleaned from the SMR family can inform the evolution of substrate specificity in other transporter families, potentially aiding in biotechnological applications such as bioremediation.

Chapter 1 Introduction

1.1 Addressing the ambiguities of promiscuity

Categorically, we can define polyspecificity in two realms: promiscuity and moonlighting (Gupta & Uversky, 2023). While moonlighting refers to the ability of an individual protein to carry out multiple functions, promiscuity simply refers to the ability of a protein to recognize multiple substrates. The differences between the two is ideally the distinction between multi-tasking (moonlighting) and multi-recognition (moonlight and promiscuity), where moonlighting requires both. Extensive studies behind catalytic enzymes have established enzyme-centric categories of multi-specificity (Karl & Per, 2007; William, 2015) (**Figure 1.1**), but in the case of promiscuous transporter systems, the definition is simplified. Many transporters display less restrictions in their specificity, as is the case with multidrug efflux transporters, and since transporters never really alter their substrates upon a completed transport cycle, only substrate promiscuity is truly relevant to their overall function. This would be analogous to the enzyme-centric category of ‘enzyme substrate promiscuity’, Therefore, this body of work will only address this category when referring to promiscuous transporters.

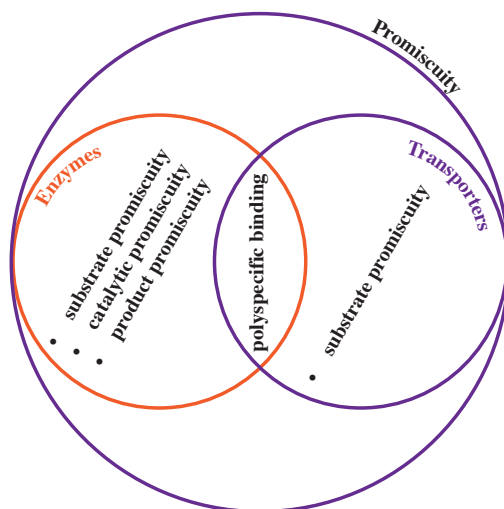


Figure 1.1: Proper categorization of promiscuity for transporters. Modified from Atkins 2015

1.2 Significance of substrate promiscuity in transporters

Substrate promiscuity has been the hallmark of antimicrobial resistance. Unfortunately, this has been a constant challenge in the medical field, where simple microbial infections have become difficult to treat. Only two solutions have been truly utilized as a stopgap for this issue: 1.) educating society on how to properly use antibiotics, 2) creating new drugs. The former is difficult to regulate as it relies heavily on human compliance. The latter eventually circles back to the original problem by introducing a new synthetic selection upon microbial fitness. Because of this vicious cycle, the pipeline to drug development has been narrowed as we encounter limits to drug diversity.

1.3 Why the SMRs?

We explore the basis of promiscuity within the context of antimicrobial resistance. Although there are multiple mechanisms that cause resistance, the most formidable method of

resistance is the efflux pump. Efflux pumps are membrane-embedded proteins that act as selective gate keepers of both essential and toxic molecules. Efflux pumps are known for their broad-spectrum activity, pumping out a wide range of drugs making them versatile as front-line defenses in cells. There are many families of efflux pumps, classified by a multitude of criteria: size, energy source, substrates, where and how they are localized in the membrane, and in which organism they're found. In bacteria, multiple types of transporters can be found, showcasing the power of functional redundancy in their resistance arsenal (Qinghu & Ian, 2007). The smallest efflux pumps known to date are from the Small Multidrug Resistance (SMR) family found in bacteria, archaea, and fungi. They are rudimentary versions of the complex efflux pumps we commonly see today, therefore are presumed to be simpler models. They also contain a promiscuous and a selective subtype within the same family while maintaining high conservation of structure and sequence. This helps us utilize the subtypes as models of both specificities and can focus on the narrowed sequence space that are different, allowing us to investigate the molecular basis of how different specificities have evolved within the same family.

1.4 A Closer at the SMR Family: ‘Still Rocking in the Structural Era: A Molecular Overview of the Small Multidrug Resistance (SMR) Transporter Family’

This section is adapted from the following published review article:

Burata, O.E.[#], Yeh, T.J.[#], Macdonald, C.B., Stockbridge, R.B. Still rocking in the structural era: A molecular overview of the small multidrug resistance (SMR) transporter family. J Biol Chem. 2022 Oct;298(10):102482. doi: 10.1016/j.jbc.2022.102482.

[#] equal contributions

C. B. M. data curation; O. E. B., T. J. Y., and R. B. S. writing—original draft; O. E. B., T. J. Y., C. B. M., and R. B. S. writing—review & editing; O. E. B., T. J. Y., and R. B. S. visualization; R. B. S. project administration; R. B. S. funding acquisition.

1.4.1 Introduction

From atomistic descriptions of membrane transport mechanism to global spread of multidrug resistance over the last century, small multidrug resistance (SMR) proteins have provided broad insights along multiple research fronts since the family's discovery in the mid 1990s (Gillings, 2017; Zhu, 2017). With just four transmembrane helices and ~100 residues, SMR proteins are among nature's smallest membrane transport proteins, making them ideal systems for biochemical and biophysical investigation. These same properties impeded high resolution structural characterization for many years, however, since the proteins are almost entirely embedded in the membrane, too small for cryo-EM, and with little polar surface area to form crystal contacts. Recently, new high resolution crystal structures have been determined for two functionally distinct SMR subtypes (Pal et al., 2015; Wales & Davies, 2015). These complement the body of mechanistic data that have been assembled over the years and provide an opportunity to consider the molecular underpinnings of functional diversity among SMR transporters.

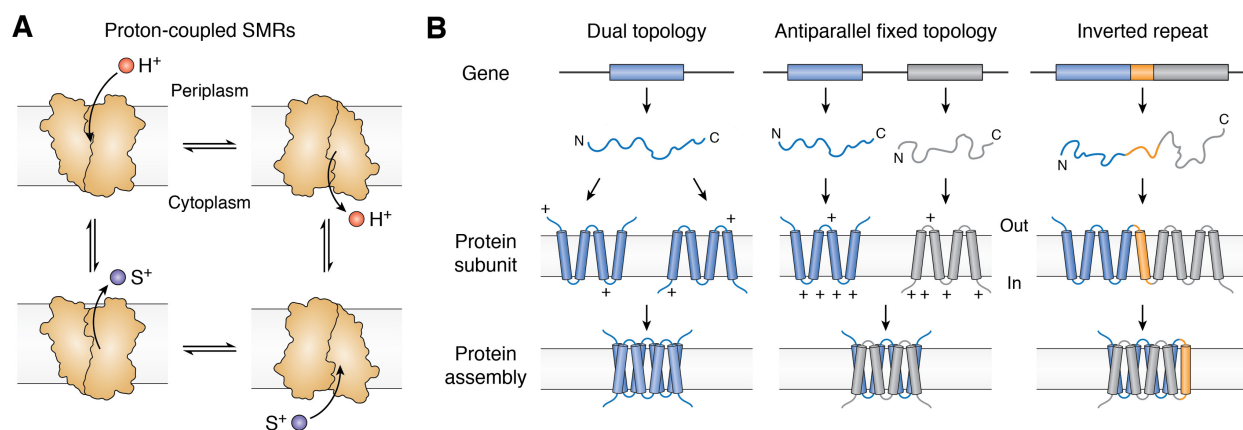


Figure 1.2: Transport scheme and transporter topologies

A, simplified scheme of the transport cycle for proton-coupled SMRs. *B*, cartoons showing some potential transporter topologies. SMR proteins are found as dual topology homodimers and fixed topology heterodimers (*left* and *center*). Dual topology proteins are characterized by a balanced distribution of positively charged residues (arginines and lysines, indicated by + symbols) on the extramembrane loops and termini, whereas subunits that assemble as antiparallel fixed dimers have oppositely biased positive charge distributions. For a protein with an even number of helices, evolution of inverted repeat topology (*right*) requires the insertion of a transmembrane helix (*orange*) to enforce antiparallel architecture of fused 4-TM subunits.

In general, the SMRs transport positively charged solutes across the membrane coupled to the antiport of protons (**Fig. 1.2a**). The resting membrane potential and pH gradient of most bacteria implies that they typically function in the active efflux of substrates. Four major functional subtypes have been described within the SMR family, and according to our bioinformatic analysis described later in this review, at least 97% of bacterial SMR genes correspond to one of these four subtypes. The first, and likely primal, SMR subtype transports guanidinium ion, a small cationic byproduct of nitrogen metabolism, and is referred to as Gdx (**guanidinium export**) (Ubarretxena-Belandia et al., 2003). These are also known by the name of the gene encoding them, *sugE*. The second subtype, which we refer to as Qac (**quaternary ammonium cation**), are promiscuous exporters of hydrophobic cationic compounds, including quaternary ammonium antiseptics like benzalkonium and cetyltrimethylammonium (also known as cetrimonium), and polyaromatic cationic biocides like methyl viologen (also known as paraquat), acriflavine, and ethidium (Breaker et al., 2017; Ali A. Kermani et al., 2018;

Zhu, 2017). The promiscuous transport phenotype of the Qac subtype gave the SMR family its name (Abhinav & William, 2008; Gillings, 2017), and this subtype includes the well-studied multidrug exporter from *Escherichia coli*, EmrE. Associated gene names for the Qac transporters include *emrE*, *ebrA/ebrB*, *qacE*, *qacG*, *qacH*, and others. The third subtype (gene name *mdtI/mdtJ*) has been implicated in the transport of small polyamine metabolites like spermidine and putrescine (Battaglia & Ke, 2018), and the fourth subtype (gene name *arnE/arnF*) acts as a glycolipid flippase (Russell, 2002). In this review, we will first describe unique topological considerations shared by all four SMR subtypes, then analyze the occurrence and distribution of the different SMR subtypes among bacterial genomes. We will review recent advances in our understanding of each SMR subtype, with particular emphasis on recent high resolution structures (**Table 1**) and finally, analyze structural homology between the SMRs and a distantly related family of transporters, SLC35 (also known as Drug/Metabolite Transport (DMT)).

Protein	Substrate	Method(max. resolution)	PDB	Reference
EmrE	Tetraphenylphosphonium(TPP)	Electron microscopy with 2D crystals (7.5 Å)	Data: EMD-1087 Model: 2I68	Data (12) Model (58)
Gdx-Clo	Gdm ⁺	Crystallography (3.50 Å)	6WK5	(4)
Gdx-Clo	PhenylGdm ⁺	Crystallography (2.53 Å)	6WK8	(4)
Gdx-Clo	OctylGdm ⁺	Crystallography (2.32 Å)	6WK9	(4)
Gdx-Clo	None (pH 5.0)	Crystallography (2.32 Å)	7SZT	(3)
^a EmrE ₃	None (pH 5.2)	Crystallography (2.85 Å)	7MH6	(3)
^a EmrE ₃	Methylviologen	Crystallography (3.13 Å)	7MGX	(3)
^a EmrE ₃	TPP	Crystallography (3.36 Å)	7SV9	(3)
^a EmrE ₃	Methyltriphenylphosphonium	Crystallography (3.22 Å)	7SSU	(3)
^a EmrE ₃	Benzyltrimethylammonium	Crystallography (3.91 Å)	7T00	(3)
^a EmrE ₃	Harmane	Crystallography (3.91 Å)	7SVX	(3)
^b EmrE S64V	Tetra(4-fluorophenyl)phosphonium/pH 5.8	NMR	7JK8	(80)
^b EmrE S64V	Tetra(4-fluorophenyl)phosphonium/pH 8.0	NMR	7SFQ	(81)

Abbreviation PDB, Protein Data Bank.

^a The construct EmrE₃ bears three functionally neutral mutations, E25N, W31I, and V34M, to facilitate crystal formation (3).

^b The S64V mutation preserves substrate binding but reduces the rate of conformational change by 8-fold (88).

Table 1: Structural data and models available for SMR transporters Gdx-Clo and EmrE

1.4.2 SMR Family Topology

In general, bacterial membrane proteins are inserted into the membrane according to the ‘positive inside rule’, in which the cytoplasmic face of the protein has an excess of positively charged residues arginine and lysine relative to the periplasmic face (Jeong & Nasir, 2017). The SMR proteins were among the first membrane proteins to be identified as possessing unusual “dual topology” architecture (Partridge et al., 2009; Slipski, 2020; Carmine J. Slipski et al., 2020). Dual topology proteins lack the typical biased charge distribution and are thus inserted into the membrane in both inward- and outward-facing orientations (**Fig. 1.2b**) where they can oligomerize with antiparallel subunits (Moura, 2009; Nelson et al., 2017). Experimental evidence suggests that individual EmrE subunits achieve their topology cotranslationally or *via* limited posttranslational annealing (Moura, 2009; Nelson et al., 2017; Robinson et al., 2017). The subunits can interact with each other during the immediate posttranslational protein folding stage (Chen et al., 2007; Rotem & Schuldiner, 2004a) but do not undergo major reorientations within the membrane after insertion to form the antiparallel homodimers (Moura, 2009).

In addition to dual topology homodimers, there are also numerous examples of SMR gene duplications that have given rise to co-expressed genes within a single operon that assemble as obligate heterodimers (Battaglia & Ke, 2018; Koide et al., 2012; Ovchinnikov et al., 2018a; Vermaas et al., 2018) (**Fig. 1.2b, middle**). Sometimes called “paired SMRs” or PSMRs (Bay et al., 2008), SMRs with this arrangement are found among all four functional subtypes and likely evolved *via* multiple independent duplication events (Kermani et al., 2020). In the great majority of these cases, the paired protomers exhibit opposite charge biases, which determine the orientation of each subunit in the membrane and enforces the antiparallel assembly (Rapp et al., 2006)

Although dual topology and fixed antiparallel topology are only rarely observed among membrane proteins (Assur Sanghai et al., 2018; Rapp et al., 2006; Stockbridge et al., 2015; Stockbridge et al., 2013) the assembly may be an evolutionary antecedent to an architecture that is extremely common among membrane transport proteins, the inverted repeat (**Fig 1.2B, right**), in which a single protein possesses structurally homologous domains arranged antiparallel with respect to each other (Forrest, 2015; Keller et al., 2014; Sarti et al., 2019). Unlike other membrane protein families that include dual topology members (Lolkema et al., 2008; Macdonald & Stockbridge, 2017), no simple inverted repeat representatives have been detected among the SMRs, suggesting that internally fused SMR proteins might not be evolutionarily advantageous (Lloris-Garcerá et al., 2014). Alternatively, the fusion of 4-transmembrane (TM) dual topology proteins might simply be an evolutionarily rare event, since this process requires addition of a transmembrane linker helix to connect the N and C termini of the two monomers (Macdonald & Stockbridge, 2017) (**Fig 1.2B, right**).

1.4.3 Distribution of SMR genes among bacterial genomes

To gauge the distribution of SMR genes across diverse microbes, we evaluated bacterial genomes from the Joint Genome Institute's curated set of ~1000 Genomic Encyclopedia of Bacteria and Archaea (GEBA) genomes (Mukherjee et al., 2017). This set of genomes was selected to reduce sampling bias and maximize phylogenetic diversity in microbial sequences used for evolutionary studies. Available sequencing data tend to be biased toward pathogenic isolates (Mukherjee et al., 2017), and because many SMRs play a role in multidrug resistance, they are particularly prone to horizontal gene transfer *via* plasmids and other transposable sequences (An et al., 2018; Domingues et al., 2015; Gaze et al., 2005; Kermani et al., 2020; Zhu,

2017). It is therefore particularly important to use a phylogenetically representative dataset to gain a balanced view of SMR distribution among bacteria. Although archaea do possess SMR transporters (Bay & Turner, 2009; Ninio & Schuldiner, 2003), we excluded archaea from this analysis.

SMR genes were identified from GEBA genomes with HMMER3.3.2 (Eddy, 2011) using a profile Hidden Markov Model (profile HMM) constructed for the SMR family (pfam 00893). Profile HMMs for each subtype (Gdx, Qac, polyamine transport, and lipid transport) were constructed from functionally annotated clusters in a sequence similarity network of reference SMR proteins (Kermani et al., 2020), and SMR sequences were assigned to the subtype that corresponded to the lowest e-value calculated by HMMR. SMR sequences were annotated “other” if the e-value was $>10^{-20}$. Consensus sequences for each SMR subtype are shown in **Figure 1.3A and B**, and sequence information and annotations for individual SMR proteins from this set of genomes is available for download from the Deep Blue Data repository hosted by the University of Michigan with unique identifier doi.org/10.7302/0ynd-b343.

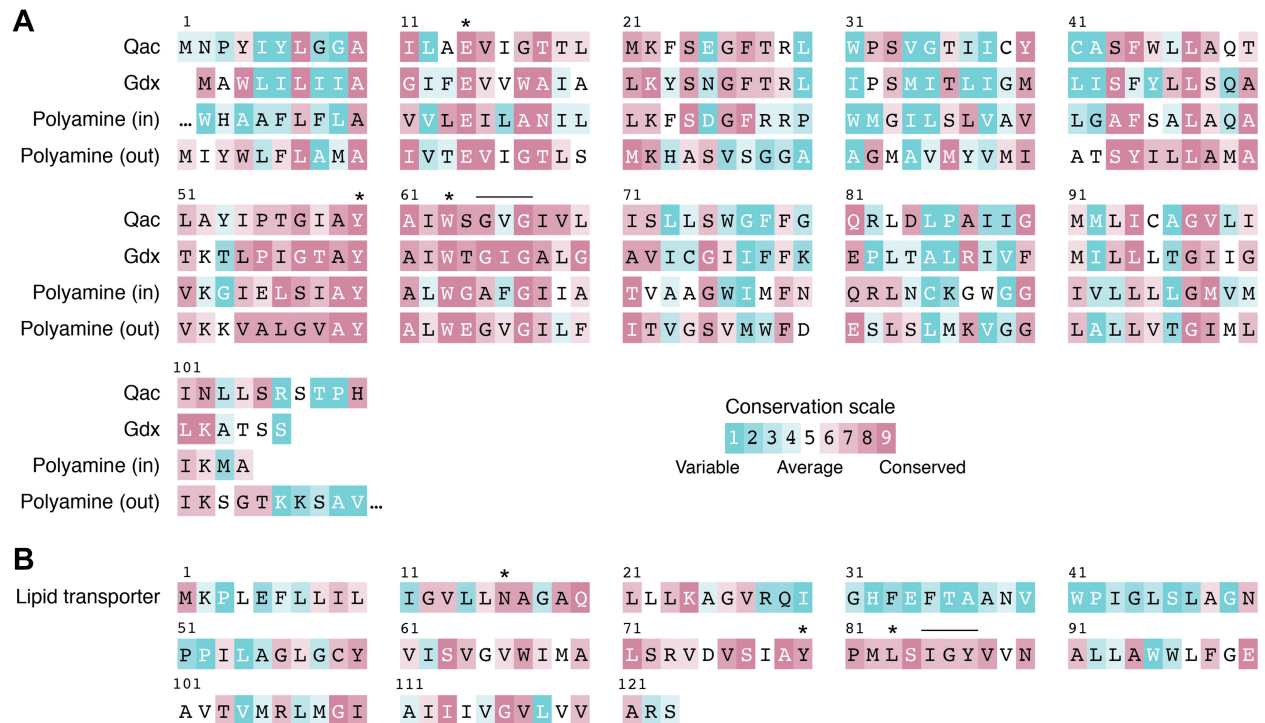


Figure 1.3: Sequence conservation of the four major SMR transporter subtypes.

A, alignment of representative proteins of the Gdx, Qac, and polyamine transport subtypes. Sequences from *E. coli* EmrE for Qac, Clostridiales Gdx-Clo for Gdx, and the inward- and outward-facing polyamine transporter subunits from proteobacterium *Photobacterium* *australis*. Sequences are numbered to correspond to the EmrE sequence and colored according to sequence conservation within that subtype using ConSurf (142). Highly conserved and mechanistically important residues including the central Glu, the Tyr switch, and the binding site Trp are indicated with asterisks. The TM3 GXG fulcrum motif is indicated by the horizontal line. B, representative sequence of lipid transporter from proteobacterium *Microvirgula aerodenitrificans* with sequence conservation analyzed and colored as in panel (A). Residues that align with the central Glu, the Tyr switch, the binding site Trp, and the GXG fulcrum are indicated.

Approximately 2/3 of the bacterial genomes from the GEBA set have at least one gene encoding an SMR protein, and ~1/3 of the GEBA genomes encode two or more SMR genes (Fig. 1.4A). This count of genomes with multiple SMR genes reflects both paired SMR genes that encode heterodimers, as well as genomes with more than one SMR functional subtype. The majority of SMR genes have no other SMR gene within 100 base pairs, suggesting they are expressed independently. As expected for dual topology proteins (Rapp et al., 2006), the Arg/Lys bias distribution for these genetic singletons is centered at 0, and only 3% encode protomers with an Arg/Lys bias greater than ± 2 (Fig. 1.4B). SMR genes are also found as adjacent gene pairs. In

our dataset, >95% of adjacent gene pairs encode subunits with opposite Arg/Lys biases. The Arg/Lys distributions are centered around +4 and -4 for inward- and outward-facing protomers, respectively, which assemble to form heterodimeric transporters (**Fig. 1.4B**). Approximately 10% of Qac transporters and ~20% of Gdx transporters are encoded by such paired genes, as are all polyamine transporters (**Fig. 1.4C**). In the GEBA genome set, all the SMR lipid transporters are encoded by singleton genes, although functional pairs have been identified in some bacteria (9).

Most bacterial phyla possess genes encoding SMRs. The most prevalent SMRs are Gdx, which are found in about ~50% of all bacterial genomes, including 80% of Actinobacteria, half of Proteobacteria and *Bacteroides*, and ~30% of Firmicutes (**Fig. 1.4D and E**). Qac genes are also frequent, found in ~25% of bacterial genomes overall, including ~50% of Proteobacteria, ~30% of Actinobacteria, and ~25% of Firmicutes. Many species possess both Gdx and Qac transporters. The lipid and polyamine transporters are less common and found mainly in Proteobacteria, where they are found in <10% of species (**Fig. 1.4F**). Thus, the SMR transporters are widespread among bacteria and dual topology Gdx and Qac transporters are the predominant SMR variants.

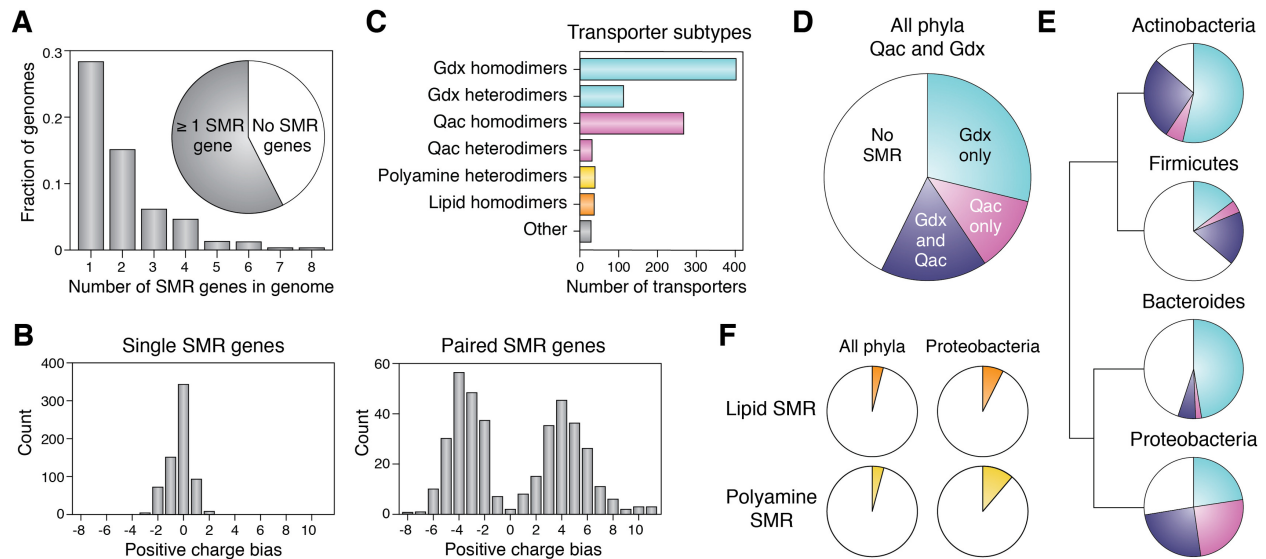


Figure 1.4: Identification and annotation of SMR-coding genes from GEBA genomes.

A, proportion of GEBA genomes that possess one or more SMR genes. B, left, positive charge (Arg/Lys) bias (cytoplasmic face) for SMR subunits encoded by genetic singletons (defined as no other SMR genes within 100 base pairs). Right, positive charge (Arg/Lys) bias (cytoplasmic face) for SMR subunits encoded by adjacent gene pairs (within 100 base pairs). The positive charge bias is given by (Arg+Lys+N-terminal amino group) termini, loop 2 – (Arg+Lys)loop1, loop3. C, transporter subtypes identified from GEBA genomes. Functional annotation is based on sequence comparison to functionally annotated gene clusters using HMMER (Eddy, 2011) as described in the text. For this annotation, proteins encoded by singleton genes with unbiased positive charge distribution are annotated as homodimers, and proteins encoded by adjacent SMR genes with opposite charge biases are annotated as heterodimers. D, proportion of all GEBA genomes that encode Gdx, Qac, or both subtypes. E, frequency of genes encoding Gdx (light blue), Qac (magenta), or both subtypes (dark blue) for major bacterial phyla. Phylogenetic relationships and distances between clades from (Coleman et al., 2021). F, proportion of all genomes and Proteobacterial genomes in the GEBA genome set that encode polyamine and lipid transport SMRs (yellow and orange slices, respectively). GEBA, Genomic Encyclopedia of Bacteria and Archaea.

1.4.4 Guanidinium exporters

Although they are the most common SMRs encoded in bacterial genomes, the Gdx transporters were also the last to be functionally annotated. The proteins were originally reported to play a role in activity of the chaperone GroEL and called SUG (**S**uppressor of **G**roEL mutations) (Greener et al., 1993). However, this phenotype was later shown to be artefactual (Bay et al., 2008; Bay & Turner, 2009; Chung & Saier, 2002). With high sequence similarity to the multidrug export Qac subtype and frequent association with horizontally transferred multidrug resistance gene arrays (Jeong & Nasir, 2017; Kermani et al., 2020; Slipski, 2020),

early characterization efforts focused on resistance to antiseptics and biocides (Bay & Turner, 2011; Chung & Saier, 2002; Nishino & Yamaguchi, 2001; Sikora & Turner, 2005). The subtype appeared to contribute to low levels of resistance to a narrow subset of drugs, but the activity was not robust, and the proteins from this subtype remained poorly characterized until their physiological role in export of guanidinium (Gdm^+) was established (Kermani et al., 2020; Nelson et al., 2017). These proteins were renamed Gdx (Guanidinium export) rather than SUG to reflect their proper functional annotation (Kermani et al., 2020).

Gdm^+ has been recognized as a byproduct of bacterial metabolism since the late 1800s, when high concentrations of Gdm^+ were found in spoiled meats (Vaughan & Novy, 1891). But the molecular players have only begun to emerge since 2017, beginning with the discovery of riboswitch-controlled operons dedicated to Gdm^+ metabolism and transport (Nelson et al., 2017). Four unrelated classes of Gdm^+ riboswitches have been identified (Lenkeit et al., 2020; Nelson et al., 2017; Salvail et al., 2020; Sherlock & Breaker, 2017; Sherlock et al., 2017), along with three distinct enzymatic pathways for utilizing Gdm^+ as a nitrogen source (Funck et al., 2022; Schneider et al., 2020; Wang et al., 2019; Wang et al., 2021), including as a sole nitrogen source by some bacteria (Funck et al., 2022; Sinn et al., 2021). The bacteria that do not consume Gdm^+ —about half of those with Gdm^+ riboswitches—instead produce and export endogenous Gdm^+ , likely as a metabolic waste product (A. A. Kermani et al., 2018; Nelson et al., 2017). The riboswitches bind Gdm^+ with K_D values between ~ 60 to $200 \mu\text{M}$ (Nelson et al., 2017) to upregulate expression of the associated transporters and enzymes. These proteins have somewhat higher K_m values for Gdm^+ , between $\sim 200 \mu\text{M}$ – 1 mM (Funck et al., 2022; Kermani et al., 2020; Nelson et al., 2017; Schneider et al., 2020), suggesting that Gdm^+ accumulation becomes toxic to cells and must be mitigated within this range.

Like other SMR transporters, the Gdx harness the bacteria's proton motive force to drive transport, exporting Gdm^+ with strict 2:1 $\text{H}^+:\text{Gdm}^+$ antiport stoichiometry (A. A. Kermani et al., 2018; Thomas et al., 2021). To prevent export of useful guanidinylated metabolites, the Gdx are highly selective for Gdm^+ over other physiological compounds with guanidinyll moieties, such as arginine, agmatine, and creatine A. A. Kermani et al. (2018). However, electrophysiological transport experiments show that the Gdx are not exquisitely selective for Gdm^+ either—although Gdx proteins strictly exclude guanidinyll metabolites with polar substituents, like arginine, they transport guanidinyll compounds with single hydrophobic substitutions at WT-like levels (Kermani et al., 2020).

The Gdx subtype yielded the first high resolution crystal structures from the SMR family, of a protein from *Clostridiales* oral taxon 876, referred to as Gdx-Clo (Kermani et al., 2020). Many of the structural features observed for this homolog had been proposed for Qac protein EmrE based on prior biophysical and biochemical experiments, establishing common structural attributes of the SMR family. Gdx-Clo possesses the expected antiparallel topology, and the two subunits assemble as an asymmetric homodimer with an aqueous cavity opened to one side of the membrane (**Fig. 1.5A**). Dimerization is mediated primarily by TM helix 4. The extramembrane loops also form extensive hydrogen bonded cross-subunit interactions to seal the closed side of the transporter. Each subunit of the dimer is composed of two discrete lobes delineated by a conserved Gly-Ile-Gly motif that acts as a fulcrum in TM helix 3 (TM3). The subunits differ according to a $\sim 35^\circ$ rotation between N- and C-terminal lobes that stems from a difference in the angle of the Gly-Ile-Gly kink (**Fig. 1.5B**). The outward facing to inward facing conformational transition involves a structural swap between the two subunits, each changing the degree of rotation between the N- and C-terminal lobes. As a result, the inward- and outward-

open conformations are 2-fold symmetric to each other, and the structural swap opens an identical, symmetry-related aqueous cavity on the opposite side of the membrane (**Fig. 1.5A**).

This elegant mechanism for the alternating access of the substrate-binding site—a prototype for the “rocker-switch” mechanism used by many other transporters—was first predicted and demonstrated for EmrE (Fleishman et al., 2006; Morrison et al., 2011), and the same TM3 fulcrum motif (Gly-hydrophobic-Gly or GXG) is conserved in the Gdx, Qac, and polyamine subtypes (**Fig. 1.3**).

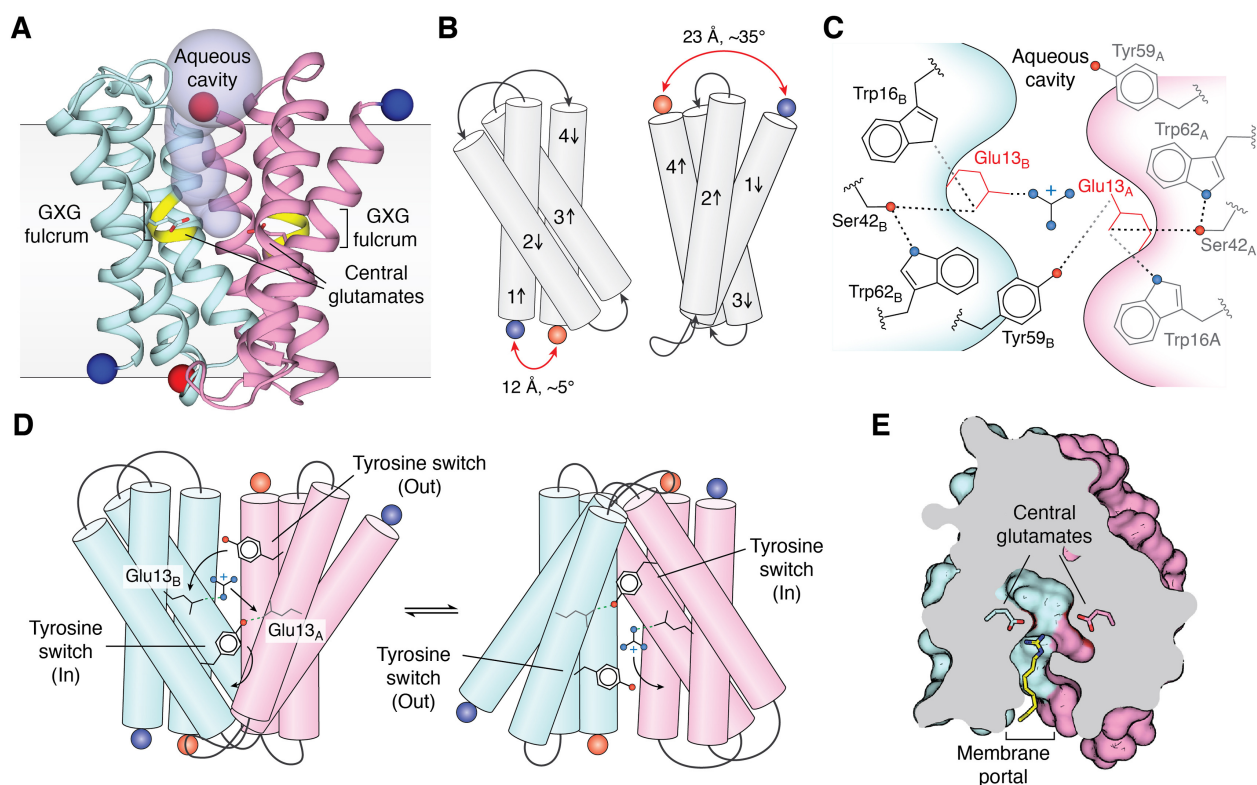


Figure 1.5: Structure and mechanism of Gdx.

A, crystal structure of Gdx-Clo (PDB:6WK9). Subunits are colored in cyan and pink, with N and C termini rendered as blue and red spheres, respectively. Central glutamates (E13_A and E13_B in Gdx-Clo) are shown as sticks, the conserved TM3 GXG fulcrum is colored yellow, and the aqueous vestibule is shown as a light blue surface. The approximate membrane boundaries are shown. B, structural comparison of individual subunits. To highlight the structural difference, the different angles between helix 1 and 4 and the different distances between the N and C termini (blue and red spheres) are indicated. C, diagram showing the hydrogen bond network in the substrate-binding site. D, proposed tyrosine switch mechanism. Rotameric movements of the tyrosine switch are indicated by arrows and hydrogen bonds indicated by dashed green lines. In this view, one of the central glutamates is occluded from view in each panel (E13_A in the left panel and E13_B in the right panel). The occluded glutamate is shown in a lighter gray color. E, top-down view of Gdx-Clo in complex with octylGdm⁺ (yellow stick rendering) sliced at the midpoint of the membrane (PDB:6WK9). The central glutamates are shown as sticks. The alkyl tail of the substrate extends out of the binding site through the membrane portal. PDB, Protein Data Bank.

The substrate-binding site is located at the bottom of the aqueous cavity, with the positively charged Gdm^+ situated between a pair of negatively charged glutamates, E13_A and E13_B, one contributed by each subunit. These “central glutamates” are conserved and essential in the Gdxs, the Qacs, and the polyamine transporters (Higashi et al., 2008; Kermani et al., 2020; Muth & Schuldiner, 2000). In addition to binding positively charged substrate, the central glutamates are protonatable at physiological pH and carry protons across the membrane during the opposing leg of the antiport cycle (Muth & Schuldiner, 2000; Yerushalmi & Schuldiner, 2000). This common binding site for the small molecule substrate and the two antiported protons favors alternating binding site occupancy by the substrates and sets the 2:1 $\text{H}^+:\text{Gdm}^+$ stoichiometry measured for Gdx (Kermani et al., 2020).

In Gdx-Clo, the central glutamates are fixed in position by a polarized stack of alternating hydrogen bond donors and acceptors, including W16, S42, and W62 (Kermani et al., 2020) (**Fig. 1.5C**). Mutation of any one of these residues to remove H-bonding capacity substantially impairs transport function (Kermani et al., 2020). In the structures, the substrate Gdm^+ is directly coordinated by E13_B, whereas E13_A is deflected away from the Gdm^+ by a cross-subunit interaction with Y59_B (Kermani et al., 2020). Y59_A, in contrast, points away from the substrate-binding pocket and into the aqueous vestibule. The divergent poses of Y59 presented a mechanistic proposal for conformational change by the SMR transporters (Kermani et al., 2020; Vermaas et al., 2018) whereby the rotameric switch of Y59_A from the aqueous vestibule toward E13_B displaces the substrate Gdm^+ from its interaction with E13_B. The Gdm^+ , in turn, engages with E13_A, displacing Y59_B, which undergoes the converse rotameric switch, away from the central glutamates. This “tyrosine switch” has been proposed to trigger the global conformational swap that opens an aqueous cavity to the other side of the membrane, where Y59_B ultimately

rests (**Fig. 1.5D**) (Kermani et al., 2020). This tyrosine is almost perfectly conserved in all SMR subtypes and is mechanistically essential in all three SMR subtypes in which the effect of its mutation has been tested (Higashi et al., 2008; Kermani et al., 2022; Kermani et al., 2020; Rotem et al., 2006), suggesting that the tyrosine switch is fundamental to transport by the SMRs.

In addition to the central glutamates and the tyrosine switch, the structures of Gdx-Clo revealed a third structural feature that is likely to be conserved among other SMR subtypes, the membrane portal (Kermani et al., 2020). This portal is defined by TM2_A and TM2_B, which form one side of the binding pocket and splay apart on the open side of the transporter. The gap between these helices is lined by hydrophobic sidechains and could, in principle, permit substrate access between the aqueous substrate-binding site and the membrane interior (**Fig. 1.5E**). In EmrE, spectroscopic experiments lead to the suggestion that the hydrophobic residues lining this portal act as a gate that permits access for the lipophilic substrates to the binding site (Dastvan et al., 2016; Vermaas et al., 2018). Similar lateral openings are well-described features of both lipid and drug transport proteins (Aller, 2009; He, 2010; Payandeh et al., 2011), permitting hydrophobic or amphipathic substrates to diffuse between, or have simultaneous access to, the membrane and the substrate-binding pocket.

The significance of the portal for the Gdx subtype is less readily apparent, however, since the physiological substrate Gdm⁺ is small and hydrophilic and would be expected to access the binding pocket directly from aqueous solution. Nonetheless, structures of Gdx-Clo with phenyl-bound and octylGdm⁺-bound showed that these non-natural substrates utilize the membrane portal to accommodate their hydrophobic substituents, while their guanidinyll headgroups bind between the central glutamates in the binding pocket, similar to Gdm⁺ (Kermani et al., 2020). These structures rationalize prior observations that Gdx transport hydrophobic, singly substituted

guanidiniums (Kermani et al., 2020; A. A. Kermani et al., 2018). Moreover, the positioning of these non-natural substrates also suggested a mechanism to select against natural guanidinyll metabolites (Kermani et al., 2020). Should a compound such as arginine or agmatine enter the binding site in the same orientation, its polar tail would likewise be positioned to extend from the binding pocket through the membrane portal. But the hydrophobic membrane interior would not favorably interact with the polar substituents, and thus, the membrane itself could contribute to selectivity against natural guanidinyllated metabolites (Kermani et al., 2020). The membrane portal might also explain the association of Gdx-encoding genes with multidrug resistance gene arrays in environmental reservoirs (Jeong & Nasir, 2017; Kermani et al., 2020; C. J. Slipski et al., 2020): hydrophobic guanidinyll compounds that enter the biosphere *via* municipal wastewater or farm runoff present microbes with a persistent low-grade toxic threat (Peter, 2018; Roberts & Hutson, 1999; Zahn, 2019) that could be mitigated by a Gdx exporter. Examples of such common biocides include the agricultural antifungal dodine (decylGdm⁺) and pharmaceuticals like metformin, which is excreted into wastewater, where it is slow to degrade and accumulates to levels of environmental concern (Blair et al., 2013; Scheurer et al., 2012). Likewise, this portal may explain prior observations that cationic detergents bind to Gdx homologs (Chung & Saier, 2002).

1.4.5 Drug and antiseptic exporters (Qacs)

Transporters of the Qac subtype garnered early attention for their role in bacterial multidrug resistance. Frequently found in clinical and agricultural isolates (Bay et al., 2008; Gillings, 2017), this SMR subtype confers resistance to the quaternary ammonium compounds used as common hospital and household antiseptics. These antimicrobial agents were introduced

in the 1930s, and evolutionary analysis suggests that it was around this time that the immediate ancestor of the clinically important vector for multidrug resistance, the class I integron, emerged (Gillings et al., 2008). This ancestral class I integron likely consisted of an integron/integrase sequence to capture drug resistance genes, a transposable element to facilitate its spread among microbial populations, and a single resistance gene: an SMR transporter of the Qac subtype (Gillings et al., 2008). Sequence analysis suggests that Qac SMRs have been dynamically associated with these and other drug resistance gene arrays over the last hundred years, gained and lost multiple times as these elements have spread among both pathogenic and environmental bacteria (Gillings et al., 2009). Today, Qac transporters remain adaptive to subinhibitory concentrations of quaternary ammonium antiseptics found in wastewater and surface runoff and remain among the most common genes isolated from human-adjacent environments (An et al., 2018; Domingues et al., 2015; Gaze et al., 2005; Zhu, 2017). By conferring this selective advantage against ubiquitous environmental biocides, the Qac SMRs coselect for other resistance genes in the cassettes that provide resistance against more potent clinical antibiotics (Pal et al., 2015), contributing to the continued spread of multidrug resistance.

Mutation tested	Substrate tested	Assay	Reference
Scan: all residues to Ala, Gly, and Val	Ethidium, acriflavine, methyl viologen	Bacterial resistance	(88)
Scan: all residues to Cys	Ethidium	Bacterial resistance	(85)
33 variants of mechanistically important residues	12 drugs including polyaromatics, quaternary ammonium cations, and antiseptics with alkyl tails	Bacterial resistance	(89)
18 mutations in helix 1, loop 1, and helix 2 to Cys	TPP	Binding (purified, reconstituted protein)	(144)
48 mutations throughout protein to Cys	Methyl viologen, acriflavine, ethidium	Bacterial resistance	(145)
All Glycines (12) mutated to Cys, Ala, Pro	Methyl viologen, acriflavine, ethidium	Bacterial resistance	(146)
All Tyrosines (5) mutated to various residues. 18 total mutants tested.	Methyl viologen, TPP	Transport and binding experiments	(62)
All Tryptophans (4) mutated to Cys.	TPP, methyl viologen, acriflavine, ethidium	Transport, binding, and bacterial resistance experiments	(86)
All acidic and basic residues (10) mutated to various residues	Ethidium, acriflavine, methyl viologen	Transport and bacterial resistance experiments	(60)

Table 2: Selected scanning mutagenesis studies of EmrE

Meanwhile, the *E. coli* variant, EmrE, has become one of the best studied bacterial multidrug exporters over the last 25 years. EmrE was originally shown to transport a variety of polyaromatic, cationic antimicrobial compounds (Yerushalmi et al., 1995) (**Fig. 1.6A**). Early, low resolution electron microscopy (EM) of 2D crystals demonstrated the unusual antiparallel architecture and established an elementary understanding of the helical connectivity and protein fold (Fleishman et al., 2006; Ubarretxena-Belandia et al., 2003). Although high resolution structural information lagged, biochemical and biophysical studies provided a detailed molecular picture of the protein and its transport cycle. As a reference, we provide a summary of the scanning mutagenesis studies that have been performed for EmrE in **Table 2**. Although too extensive to discuss individually here (see reference (Schuldiner, 2009) for an in-depth review of EmrE mutagenesis), these studies provided a functional grounding for the interpretation of electron paramagnetic resonance (EPR) distance measurements (Dastvan et al., 2016), models based on the low-resolution EM data with computationally predicted sidechains (Jurasz et al., 2021; Ovchinnikov et al., 2018b; Vermaas et al., 2018), models derived from NMR chemical shifts and substrate/protein distance restraints (Shcherbakov et al., 2021; Shcherbakov et al., 2022), and ultimately, the high resolution crystal structures, as discussed later (Kermani et al., 2020).

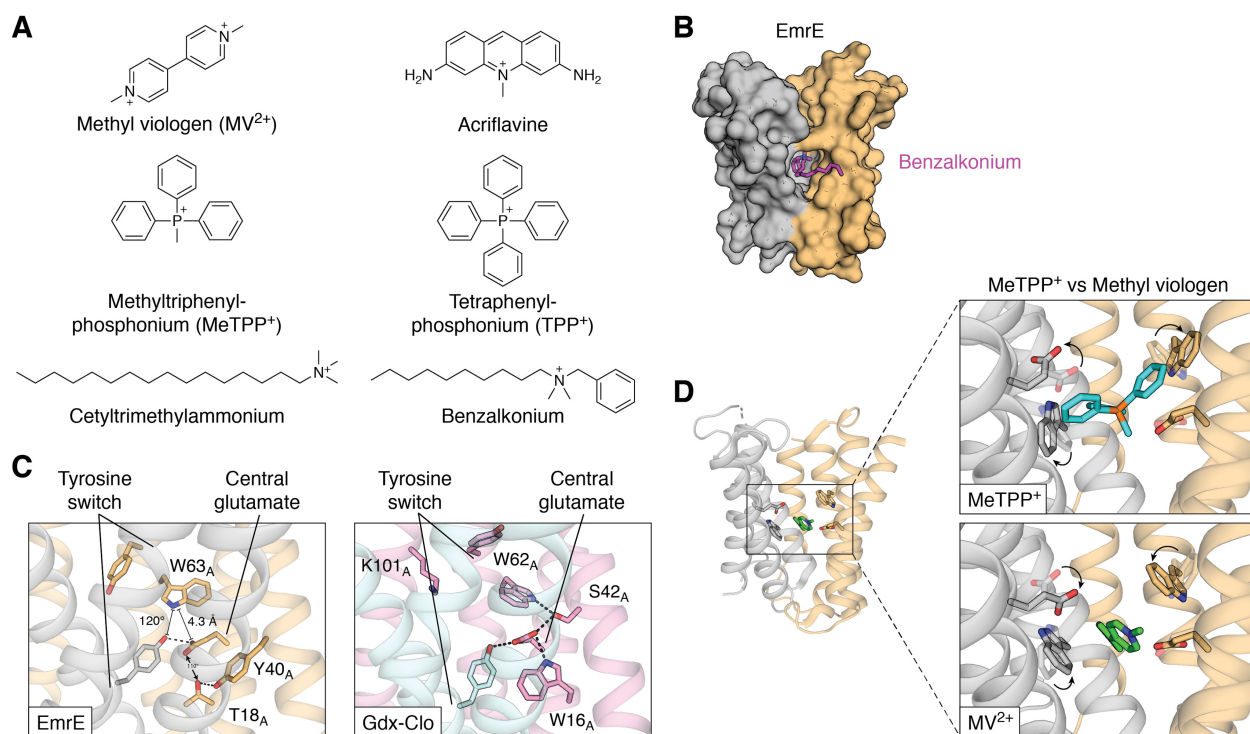


Figure 1.6: Substrate binding to EmrE.

A, examples of planar aromatic, quaternary ammonium/phosphonium, and alkylated substrates transported by EmrE. B, model of EmrE with benzalkonium bound (Kermani et al., 2022). EmrE subunits in gray and tan surface rendering with benzalkonium shown as purple stick representation. The substrate's alkyl tail extends from the binding site into the membrane via the conserved membrane portal. C, comparison of EmrE (PDB:7MH6) and Gdx-Clo (PDB:7SZT) substrate-binding sites with putative hydrogen bonds (distance <3.5 Å, angle 140 – 180°) represented as dashed lines. Distances/angles are shown for potential H-bond partners that possess nonoptimal geometry. D, EmrE substrate-binding site with different substrates bound. Top panel: bulky quaternary phosphonium substrate, methyltriphenylphosphonium (PDB:7SSU). Bottom panel: planar aromatic substrate, methyl viologen (PDB:7MGX). To aid comparison, sidechain positions in the presence of the alternative substrate is rendered as lightly colored, transparent sticks. Arrows indicate movements of the central glutamates and binding site tryptophan (W63) that accommodate the differently sized substrates. PDB, Protein Data Bank.

These EmrE crystal structures were determined with structurally diverse substrates bound, including planar polyaromatics and tetrahedral quaternary phosphoniums and ammoniums **Table 1**. The electron density determined using X-ray crystallography shows high correspondence with previous low resolution EM data obtained for EmrE in lipid bilayers (Ubarretxena-Belandia et al., 2003), implying that the crystal structures represent a native, low energy conformation. The molecular details of the crystal structures also agree with specific predictions from spectroscopic and mutagenic studies (Dastvan et al., 2016). Recent models from

the NMR experiments (Shcherbakov et al., 2021; Shcherbakov et al., 2022) exhibit notable structural differences with the crystal structures (Kermani et al., 2020) and with the available computational and EM models (Fleishman et al., 2006; Vermaas et al., 2018). These differences, and potential reasons for the differences, are discussed in depth in (Kermani et al., 2020) and will only be briefly summarized here. In the NMR models, the subunits of the dimer are arranged more parallel with respect to each other, and the loops are unpacked, permitting aqueous access to the binding site from both sides of the membrane. The hydrogen bond network in the binding site is also rearranged with respect to the recent crystal structures. It is possible that the NMR models represent functional intermediate states, such as those suggested by prior EPR experiments (Dastvan et al., 2016), and that the differences between the crystallography or EM density (which largely agree with each other) and NMR models are due to differences in the experimental conditions. However, it is also important to note that the NMR models are based on relatively few experimental measurements of distances between backbone atoms and bound substrate and that the models (including sidechain placement) are generated computationally, based on these distance restraints and backbone chemical shift measurements. Since the recent crystal structures are in the best agreement with the EM maps and are the only structural models of EmrE with experimental electron density that supports sidechain placement without further computational modeling, we will focus our analysis on these crystal structures (Kermani et al., 2020). Note that previous EmrE crystal structures (Chen, 2007) (Protein Data Bank codes 3B5D, 3B62, 3B61) are low quality and incomplete (only C_α atoms are modeled) and experimental maps are unavailable in the Protein Data Bank. These prior structures are broadly considered inadequate for molecular inference (Jurasz et al., 2021; Kermani et al., 2020; Vermaas et al., 2018) and should not be used.

Both the recent structures and the long history of functional data show that EmrE has many major mechanistic features in common with Gdx-Clo, including the central glutamates in the binding pocket that contribute to alternate binding of drug and protons (Li et al., 2021; Morrison et al., 2015; Muth & Schuldiner, 2000), the tyrosine switch engaged in a cross-subunit interaction (Rotem et al., 2006; Vermaas et al., 2018), the GXG fulcrum that kinks TM3 and defines the N- and C-terminal lobes of each subunit (Amadi et al., 2010; Fleishman et al., 2006), and the hydrophobic portal that permits access between the binding pocket and the hydrophobic interior of the inner lipid membrane for hydrophobic substrates and substrate substituents (Dastvan et al., 2016; Vermaas et al., 2018). Based on information from the crystal structures of EmrE with benzyltrimethylammonium (the headgroup of the common household antiseptic benzalkonium) and Gdx-Clo with octylGdm⁺, a model for benzalkonium binding was constructed (Kermani et al., 2020), illustrating how the membrane portals of both the Gdx and Qac subtypes can be exploited to bind substrates with extended alkyl substituents (**Fig. 1.6B**).

Given the structural similarities and high sequence conservation, why then can EmrE bind and transport a much more diverse range of substrates than Gdx-Clo? The crystal structures in complex with substrates suggest that, despite shared sequences, EmrE and Gdx-Clo also have important structural differences. Although many of the binding pocket residues that serve as hydrogen bond donors or acceptors are conserved in EmrE, they do not form as an extensive inter-residue H-bond network as is observed in the binding site of Gdx-Clo. Peripheral binding site residues W63 and S42 are both present, but they do not H-bond with other EmrE sidechains (Kermani et al., 2020). Many potential H-bond interactions have either poor geometry or longer interaction distances than ideal (**Fig. 1.6C**). As a result, the central glutamates and conserved binding site Trp (W63 in EmrE) are comparatively unconstrained and able to adopt different

rotamers in the presence of different substrates. EmrE's central glutamates move closer together or farther apart to accommodate flat planar substrates or bulky quaternary compounds (**Fig. 1.6D**). At the same time, the binding site Trp rotates over $\sim 80^\circ$ to stack against aromatic rings of different substrates bound in different poses (Kermani et al., 2020) (**Fig. 1.6D**). Unlike Gdx-Clo, in which the residues that contribute to the H-bond stack stabilizing the central glutamates cannot be altered (Kermani et al., 2020), many of the analogous mutations to reduce H-bonding capacity are tolerated in EmrE (Kermani et al., 2020). Notably, the binding site Trp, W63, which had been shown to be essential in all previous studies with aromatic substrates (Amadi et al., 2010; Brill et al., 2015; Elbaz et al., 2005; Wu, 2019), is not required for the transport of nonaromatic substrates by EmrE (Kermani et al., 2020; Saleh et al., 2018). These structural observations are in accord with NMR and computational studies that suggest that EmrE possesses an unusual degree of structural plasticity that might contribute to substrate polyspecificity (Jurasz et al., 2021). However, it should be emphasized that the observed structural perturbations are limited to the sidechains. Larger conformational changes involving the backbone are not necessary to explain the binding of diverse substrates to EmrE, and such perturbations are not observed in any of the five drug-bound crystal structures (Kermani et al., 2020). Likewise, the low pH, proton-bound crystal structures of EmrE (as well as Gdx-Clo) do not exhibit major structural differences relative to the substrate-bound structures, with only local changes in the position of the central glutamates (Kermani et al., 2020). However, EPR studies have suggested that upon protonation, the conformational ensemble of EmrE becomes more heterogeneous than in the presence of the drug TPP⁺ (Dastvan et al., 2016). Thus, the doubly protonated state might not exist in a single predominant stable conformation and the reported low pH crystal structure may capture only one species in this ensemble.

Although the crystal structures provided essential insight into the molecular basis for substrate binding by EmrE, the transport of these disparate substrates poses additional problems, requiring the choreography of substrate binding and dissociation, conformational exchange, and proton antiport. Spectroscopic techniques, including NMR and EPR, have been integral to fleshing out a dynamic picture of EmrE. NMR studies in lipid membranes and in bicelles have identified various mutations that slow or eliminate conformational change but preserve substrate binding, isolating residues involved in the first process and not the second (Leninger et al., 2019; Wu, 2019). NMR measurements have shown that the kinetic behavior of EmrE, including the rate of conformational exchange, differs depending on the substrate, demonstrating that different substrates have different affinities for the transition state of the conformational exchange, as they do in the ground state (Morrison & Henzler-Wildman, 2014). Likewise, protonation of the central glutamates accelerates conformational transition (Gayen et al., 2016), perhaps reflecting the same ground state destabilization that leads to heterogeneity in the conformational ensemble upon protonation (Dastvan et al., 2016). The reduction of free energy of conformational transitions upon substrate binding is a classical requirement for coupled substrate antiport (Gayen et al., 2016). However, emerging evidence also suggests that under certain conditions, EmrE violates tenets of classic transport mechanisms. Conformational exchange of the apo (proton- and drug-free) and single proton-bound transporter have been reported (Cho et al., 2014; Gayen et al., 2013; Morrison et al., 2015), as has simultaneous binding of proton with some drugs such that both can be carried across the membrane together (Robinson et al., 2017; Shcherbakov et al., 2021). The conformational exchange rate of EmrE with different substrates is not tightly correlated with the rate of substrate transport, hinting that different substrates, particularly high affinity substrates, might undergo futile cycles and remain bound as the transporter transits

between inward and outward open states (Morrison & Henzler-Wildman, 2014). Kinetic modeling (Hussey et al., 2020) suggests scenarios in which the microscopic rate constants measured for each potential binding event and conformational transition in the transport cycle combine to reduce the stoichiometry noticeably from the 2:1 H⁺:substrate stoichiometry measured for Gdx (Kermani et al., 2020) and many EmrE substrates (Rotem & Schuldiner, 2004b). These studies suggest that, in some cases, the specific proton and substrate gradients and substrate-binding energy may even lead to cycles of substrate import (Hussey et al., 2020; Robinson et al., 2017).

A limitation of such free exchange transport models is that they permit potential pathways for proton leak. If the inward-to-outward facing transition of the unoccupied or singly protonated transporter is not energetically prohibitive in living bacteria, such transport cycles would contribute to the dissipation of the proton motive force. For EmrE, different mechanisms have been proposed to explain the apparent absence of detrimental leak pathways *in vivo* and *in vitro*. In one proposal, the central glutamates are electrostatically independent so that the proton cannot “hop” from the glutamate with the lower pK_a to that with the higher pK_a, preventing proton release from the singly protonated state after the conformational swap (Li et al., 2021). Alternatively, it has been suggested that the proton pathway is gated by a C-terminal histidine residue that is highly conserved among the Qac subtype, which occludes the binding pocket in the absence of drug, preventing proton leak (Thomas et al., 2018). It should also be mentioned that the quaternary phosphonium substrates used in these transport experiments are not encountered by bacteria outside the laboratory. Thus, the transport properties for such anthropogenic chemicals have not been optimized by purifying selection in bacterial populations over evolutionary time. While it is possible that proton slippage

and deviations from ideal stoichiometry are evolved properties of the transporter to handle diverse substrates (Robinson et al., 2017), it is also possible that these mechanistic features reflect nonoptimized transport of non-native compounds and that transport of native substrates (whether the drug-like molecules produced by microbes in competitive niches or yet-unknown metabolites) is more parsimonious. Native substrates of the Qac transporters have yet to be identified, however, so this remains an open question.

1.4.6 Polyamine transporters

Polyamines, such as spermidine, putrescine, and cadaverine, play myriad roles in diverse bacteria (Di Martino et al., 2013; Michael, 2018; Shah & Swiatlo, 2008; Wortham et al., 2007). These small, charged metabolites are synthesized or taken up from the environment by bacteria, where they can be used as synthetic precursors to siderophores (Abergel et al., 2006; Codd et al., 2018) or structural components of the cell wall (Hirao et al., 2000; Takatsuka & Kamio, 2004), contribute to oxidative stress resistance (Barbagallo et al., 2011; Chattopadhyay et al., 2003), or interact with nucleic acids to modulate translation (Shah & Swiatlo, 2008). Polyamines also serve as signals for the induction of virulence genes (Chan & Chua, 2010; Jelsbak et al., 2012; Ware et al., 2006) and surface behaviors like biofilm formation and swarming (Burrell et al., 2010; Karatan et al., 2005; Kurihara et al., 2009; Patel et al., 2006; Sturgill & Rather, 2004). However, excessive accumulation of polyamines is toxic (Fukuchi et al., 1995). A subset of transporters from the SMR family has been implicated in polyamine export in *E. coli* and *Shigella* (Higashi et al., 2008; Leuzzi et al., 2015). These proteins are proposed to serve as a “safety valve” when intracellular polyamines accumulate to toxic levels (Leuzzi et al., 2015). The transporters form heterodimers, and their genes are always found as pairs,

annotated *mdtI* and *mdtJ* in *E. coli* (Higashi *et al.*, 2008). Native expression is low and is regulated by accumulation of polyamines and bile salts (Higashi *et al.*, 2008; Leuzzi *et al.*, 2015). Although biochemical information is relatively limited for the polyamine transporters, mutagenesis coupled with growth assays has demonstrated that key mechanistic residues for Qac and Gdx function, including the central glutamates, the tyrosine switch, and the binding site tryptophans, are critical for function of the MdtIJ complex (Higashi *et al.*, 2008), implying that the polyamine transporters share mechanistic similarities with the more extensively characterized Gdx and Qac subtypes.

1.4.7 Lipid transport proteins

The most distantly related members of the SMR family are reported to act as glycolipid flippases. The SMR genes (annotated *arnE* and *arnF* in *E. coli* and *Salmonella enterica*) are found in larger biosynthetic operons that contribute to the chemical modification of lipid A in the outer membrane with 4-amino-4-deoxy-l-arabinose (L-Ara4N) (Gunn *et al.*, 1998). This synthetic pathway contributes to polymyxin resistance by reducing the electrostatic interactions of lipid A with the cationic polymyxin antibiotic (Gunn *et al.*, 1998). In *S. enterica* serovar Typhimurium, the role of this SMR subtype is to transport undecaprenyl phosphate aminoarabinose (Yan *et al.*, 2007), a lipid that carries L-Ara4N from its site of synthesis in the cytoplasm to the periplasmic leaflet, on its way toward lipid A in the outer membrane (Trent *et al.*, 2001). Deletion of the SMR transport proteins prevents localization of L-Ara4N to the outer membrane and thus prevents cells from acquiring polymyxin resistance *via* this route (Yan *et al.*, 2007). There are indications that the *S. enterica* flippase has somewhat broader specificity,

including genetic complementation of flippase deletion mutants in synthetic pathways that require transport of different glycolipids (Chin et al., 2021; Larrouy-Maumus et al., 2012).

The lipid SMRs are the only subtype that transports noncationic substrates, and these proteins often possess an asparagine in place of the central glutamate (**Fig. 1.3**). This replacement also suggests that lipid transport may not be proton coupled, since the central glutamates are also responsible for proton binding in proton-coupled SMRs. Because undecaprenyl phosphate aminoarabinose is synthesized in the cytoplasmic leaflet (Trent et al., 2001), transport of the lipid down its concentration gradient to the periplasmic leaflet *via* facilitated diffusion might be sufficient. However, this has not been established experimentally. A membrane portal similar to that observed in structures of EmrE and Gdx-Clo would be an obviously useful feature for lipid transport, permitting the substrate's prenyl tail access to the membrane while the polar headgroup is ensconced within its protein-binding site, a familiar feature of other lipid flippases (Kalienkova et al., 2021). Indeed, in this SMR subtype, the hydrophobic character of the TM2 residues is retained, and we therefore conjecture that the membrane portal is conserved as well. The tyrosine switch is also conserved in the lipid SMRs, despite their overall low sequence similarity with other SMR subtypes.

1.4.8 Structural relationship between the SMR and the SLC35/DMT folds

Although other dual topology transporter families have representatives with inverted repeat topology (Lolkema et al., 2008; Macdonald & Stockbridge, 2017), the SMR family does not possess such internally fused transporters with detectable sequence homology (Lloris-Garcerá et al., 2014). However, structural analysis of transporters from the DMT superfamily, which possess the SLC35/DMT fold, suggests that the SMR fold might nonetheless have been

preserved by evolution as an inverted repeat. Structures of transporters with the SLC35/DMT fold, including a bacterial aromatic amino acid exporter, a protozoan drug exporter, and eukaryotic organellar sugar/nucleotide transporters (Bärland et al., 2022; Kim et al., 2019; Nji et al., 2019; Parker & Newstead, 2017; Tsuchiya, 2016), possess striking structural homology to the bacterial SMRs (TM helix RMSD 2.6–3.8 Å) despite sharing no sequence similarity (**Fig. 1.7**). DMT and SMR transporters have previously been proposed to be evolutionarily related (Jack et al., 2001; Västermark et al., 2011). However, the SLC35/DMT and SMR structures also bear two notable differences. First, each domain of the SLC35 inverted repeat is composed of five TM helices. In the 3D structure, this pair of inserted helices pack against the membrane portal defined by helices 2_A and 2_B of the SMR transporters, sealing the portal and eliminating access to the substrate-binding site from the membrane (Kermani et al., 2020). The second major difference between the SMR and SLC35/DMT structures is in the helix connectivity. Whereas in the SMRs, each transport domain is composed of a single, independently folded monomer; in the SLC35/DMT proteins, the transport domains do not correspond simply to the N- and C-terminal halves of the protein. The first transport domain is composed of helices 1, 2, 8, 9, and 10 and the second is composed of helices 3, 4, 5, 6, and 7 (**Fig. 1.7A**).

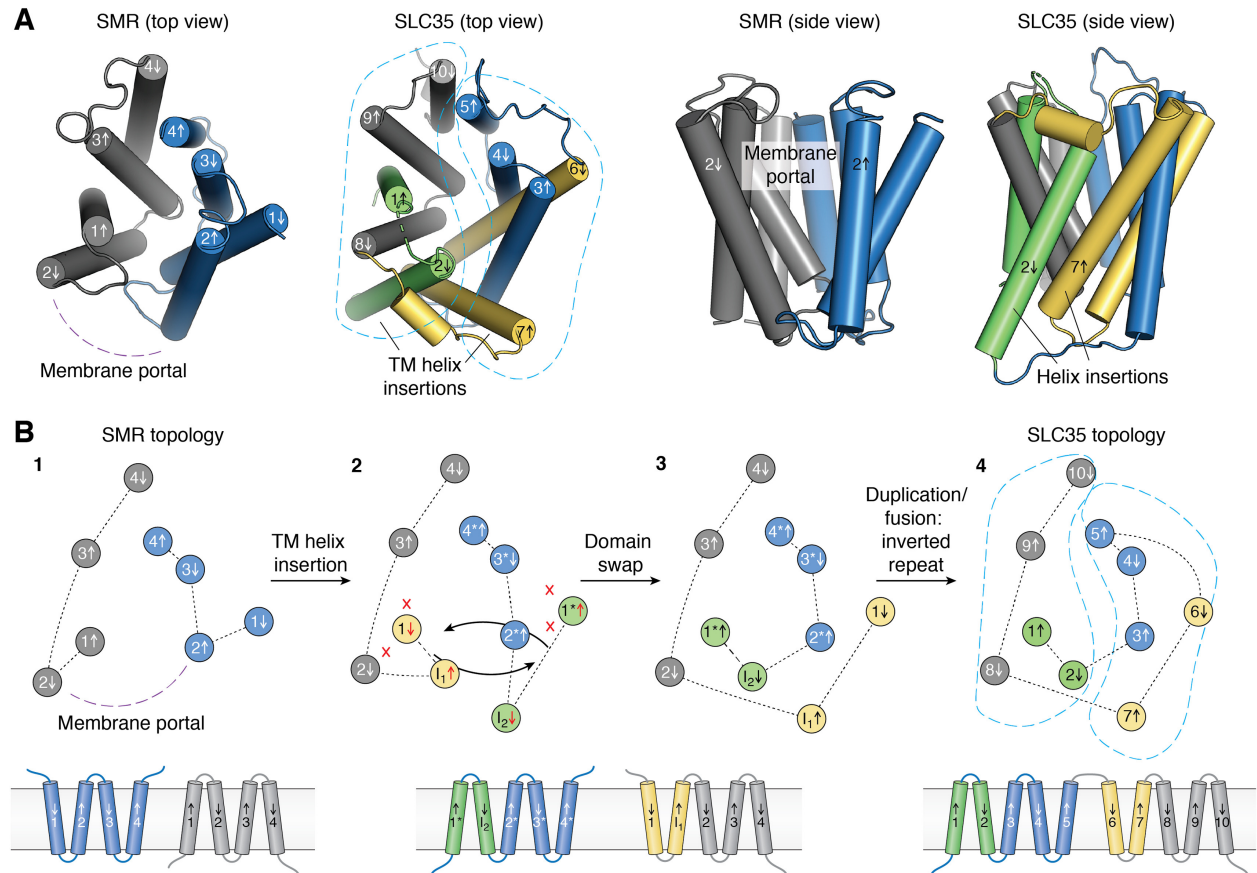


Figure 1.7: Structural relationship between the SMR and SLC35 folds.

A, structural comparisons of Gdx-Clo (PDB:6WK9) (Kermani et al. 2020) and SLC35 CMP-sialic acid transporter (PDB: 6I1R) (Nji et al., 2019). For the SMR fold, the individual subunits are colored in gray and blue and the membrane portal is labeled. For the SLC35/DMT fold, helices involved in a potential domain swap are colored in green (for the blue monomer) and yellow (for the gray monomer). The transmembrane (TM) helix insertions are indicated. The two transport domains are outlined by the dashed turquoise line. B, a potential pathway for divergent evolution of the SMR and SLC35/DMT folds. Topology is shown as a top-down cartoon, as in the left panels of (A). Panel 1: SMR topology with membrane portal indicated. Panel 2: insertion of a TM helix (helices I1 and I2) inverts helix 1 and 1* with respect to the other helices in the subunit, introducing clashes and disrupting helical packing (red x symbols). The original packing can be restored by swapping the positions of the yellow and green helices, indicated by arrows. Panel 3: Dual topology ancestor of the SLC35/DMT fold possesses a helix insertion that seals the membrane portal and a domain swap involving the yellow and green helices that preserves structural homology with the SMR fold. Panel 4: a duplication/fusion event links the C terminus of the first subunit (helix 5) with the N terminus of the second subunit (helix 6) and fixes this topology in the SLC35/DMT lineage. The transport domains are outlined by the turquoise dashed line, as in the SLC35 structure in panel (A). PDB, Protein Data Bank.

Although the possibility that the similarities between the SMR and SLC35/DMT folds arose *via* convergent evolution cannot be ruled out; structural correspondence between proteins with similar functions is usually considered evidence for evolutionary relatedness (Dietmann & Holm, 2001; Theobald & Wuttke, 2005). Moreover, we propose that the topological differences

between the SMRs and SLC35/DMT folds can be plausibly explained by a divergent evolutionary pathway (**Fig. 1.7B**). This model posits an ancestral, dual topology transporter with the SMR fold and a TM helix insertion between TM helices 1 and 2. The introduction of this new TM helix would enforce a reorientation of TM helices 2, 3, and 4 relative to TM helix 1, disrupting the packing between TM helix 1 and the other TM helices in that monomer (**Fig. 1.7B**, panel 2). The 3D SMR fold could be preserved, however, by a domain swap during dimer assembly (**Fig. 1.7B**, panels 2 and 3) such that the now-inverted TM1 from the first monomer trades positions with TM1* in the opposite subunit and *vice versa* (**Fig. 1.7B**, panel 3). TM helix domain swaps have been observed in other membrane proteins (Brohawn et al., 2013; Singh et al., 2017) and for engineered EmrE concatamers (Lloris-Garcera et al., 2014), and an analogous mechanism has been proposed for the evolution of the structurally similar, but topologically distinct, Pnu vitamin transporters and SWEET sugar exporters (Jaehme et al., 2015). Finally, a subsequent duplication/fusion of the 5-helix, domain-swapped ancestral dual topology transporter would give rise to the extant SLC35/DMT fold (**Fig. 1.7B**, panel 4).

Duplication/fusion of 4-TM dual topology transporters are not unprecedented (Macdonald & Stockbridge, 2017), but this process is more common for dual topology transporters with an odd number of transmembrane helices, since the N and C termini are on the same side of the membrane and connecting them does not require introduction of a transmembrane linker (Lloris-Garcera et al., 2014; Lolkema et al., 2008). Although no sequence homology can be detected between the SLC35/DMT transporters and SMRs or even between symmetry-related helices of the SLC35/DMTs, lack of sequence homology is not uncommon in other families of evolutionary divergent transporters with shared folds (Forrest, 2015; Theobald & Miller, 2010).

1.4.9 Conclusions and perspective

The SMRs have provided a tremendously productive system for studying membrane protein evolution, transport mechanism, and microbial multidrug resistance. After 25 years of such studies, the SMR transporters finally joined the high resolution structural era in 2021. These recent crystal structures representing two of the four known functional subtypes have provided a platform for analyzing prior functional studies and understanding the structural features that contribute to substrate binding and transport for each SMR subtype: Gdm⁺, drugs, polyamines, and glycolipids. Moreover, structural homology with SLC35/DMT transporters suggests that, contrary to the proposal that the SMRs are unusual in having not evolved fused, inverted repeat architecture (Lloris-Garcerá et al., 2014), it is probable that the SMR fold has indeed been preserved through this evolutionary mechanism, albeit with a helical insertion and domain swap along the way. These recent advances in understanding the molecular architecture bolster ongoing efforts to develop antimicrobials that target SMR proteins, either by inhibiting transporter assembly in order to sensitize bacteria to transported compounds (Bellmann-Sickert et al., 2015; Ovchinnikov et al., 2018b; Poulsen & Deber, 2012) or by hijacking the nominal exporters to import antimicrobial compounds instead (Shcherbakov et al., 2022; Spreacker et al., 2022).

In addition, the molecular framework described here opens the door for future integrative functional, structural, and computational studies to understand how the SMR scaffold has been tailored to transport diverse substrates as the family has evolved. Such lines of inquiry are urgently important as bacteria continue to evolve around us. The Qac and Gdx subtypes, in particular, have found new roles in human-impacted environments, conferring bacterial resistance to household antiseptics (Saleh et al., 2018), “dead-end” metabolites from the

degradation of metformin (Tassoulas et al., 2021) and other pharmaceuticals that accumulate in the environment or the human microbiome, and other agricultural and industrial chemicals (Kermani et al., 2020). Genes encoding SMR transporters are currently spreading among bacterial populations, encountering new physiological contexts and substrate transport demands, driving co-selection of co-localized antimicrobial resistance genes in environmental reservoirs (Pal et al., 2015), and influencing microbial population compositions in the human microbiome and human-impacted environments (Maier, 2018; Pal et al., 2015; Zhu, 2017), as the SMR transporters contribute to the ongoing story of natural selection at the human–microbe interface.

1.5 References

- Abergel, R. J., Wilson, M. K., Arceneaux, J. E., Hoette, T. M., Strong, R. K., Byers, B. R., & Raymond, K. N. (2006). Anthrax pathogen evades the mammalian immune system through stealth siderophore production. *Proc Natl Acad Sci U S A*, 103(49), 18499-18503. <https://doi.org/10.1073/pnas.0607055103>
- Abhinav, N., & William, M. A. (2008). A quantitative index of substrate promiscuity. *Biochemistry*. <https://doi.org/10.1021/bi701448p>
- Aller, S. G. (2009). Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*, 323. <https://doi.org/10.1126/science.1168750>
- Amadi, S. T., Koteiche, H. A., Mishra, S., & McHaourab, H. S. (2010). Structure, dynamics, and substrate-induced conformational changes of the multidrug transporter EmrE in liposomes. *J. Biol. Chem.*, 285. <https://doi.org/10.1074/jbc.M110.132621>
- An, X. L., Chen, Q. L., Zhu, D., Zhu, Y. G., Gillings, M. R., & Su, J. Q. (2018). Impact of Wastewater Treatment on the Prevalence of Integrins and the Genetic Diversity of Integron Gene Cassettes. *Appl Environ Microbiol*, 84(9). <https://doi.org/10.1128/aem.02766-17>
- Assur Sanghai, Z., Liu, Q., Clarke, O. B., Belcher-Dufresne, M., Wiriyasermkul, P., Giese, M. H., Leal-Pinto, E., Kloss, B., Tabuso, S., Love, J., Punta, M., Banerjee, S., Rajashankar, K. R., Rost, B., Logothetis, D., Quick, M., Hendrickson, W. A., & Mancina, F. (2018). Structure-based analysis of CysZ-mediated cellular uptake of sulfate. *Elife*, 7, e27829. <https://doi.org/10.7554/eLife.27829>
- Barbagallo, M., Di Martino, M. L., Marcocci, L., Pietrangeli, P., De Carolis, E., Casalino, M., Colonna, B., & Prosseda, G. (2011). A new piece of the Shigella Pathogenicity puzzle: spermidine accumulation by silencing of the speG gene [corrected]. *PLoS ONE*, 6(11), e27226. <https://doi.org/10.1371/journal.pone.0027226>

- Bärland, N., Rueff, A. S., Cebrero, G., Hutter, C. A. J., Seeger, M. A., Veening, J. W., & Perez, C. (2022). Mechanistic basis of choline import involved in teichoic acids and lipopolysaccharide modification. *Sci Adv*, 8(9), eabm1122. <https://doi.org/10.1126/sciadv.abm1122>
- Battaglia, R. A., & Ke, A. (2018). Guanidine-sensing riboswitches: how do they work and what do they regulate? . *Wiley Interdiscip Rev RNA*. <https://doi.org/https://doi.org/10.1002/wrna.1482>
- Bay, D. C., Rommens, K. L., & Turner, R. J. (2008). Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochim Biophys Acta*, 1778(9), 1814-1838. <https://doi.org/10.1016/j.bbamem.2007.08.015>
- Bay, D. C., & Turner, R. J. (2009). Diversity and evolution of the small multidrug resistance protein family. *BMC Evol Biol*, 9, 140. <https://doi.org/10.1186/1471-2148-9-140>
- Bay, D. C., & Turner, R. J. (2011). Spectroscopic analysis of the intrinsic chromophores within small multidrug resistance protein SugE. *Biochim Biophys Acta*, 1808(9), 2233-2244. <https://doi.org/10.1016/j.bbamem.2011.05.005>
- Bellmann-Sickert, K., Stone, T. A., Poulsen, B. E., & Deber, C. M. (2015). Efflux by small multidrug resistance proteins is inhibited by membrane-interactive helix-stapled peptides. *J Biol Chem*, 290(3), 1752-1759. <https://doi.org/10.1074/jbc.M114.616185>
- Blair, B. D., Crago, J. P., Hedman, C. J., & Klaper, R. D. (2013). Pharmaceuticals and personal care products found in the Great Lakes above concentrations of environmental concern. *Chemosphere*, 93. <https://doi.org/10.1016/j.chemosphere.2013.07.057>
- Breaker, R. R., Atilho, R. M., Malkowski, S. N., Nelson, J. W., & Sherlock, M. E. (2017). The biology of free guanidine as revealed by riboswitches. *Biochemistry*, 56. <https://doi.org/10.1021/acs.biochem.6b01269>
- Brill, S., Sade-Falk, O., Elbaz-Alon, Y., & Schuldiner, S. (2015). Specificity determinants in small multidrug transporters. *J. Mol. Biol.*, 427. <https://doi.org/10.1016/j.jmb.2014.11.015>
- Brohawn, S. G., Campbell, E. B., & MacKinnon, R. (2013). Domain-swapped chain connectivity and gated membrane access in a Fab-mediated crystal of the human TRAAK K⁺ channel. *Proc Natl Acad Sci U S A*, 110(6), 2129-2134. <https://doi.org/10.1073/pnas.1218950110>
- Burrell, M., Hanfrey, C. C., Murray, E. J., Stanley-Wall, N. R., & Michael, A. J. (2010). Evolution and multiplicity of arginine decarboxylases in polyamine biosynthesis and essential role in *Bacillus subtilis* biofilm formation. *J Biol Chem*, 285(50), 39224-39238. <https://doi.org/10.1074/jbc.M110.163154>
- Chan, Y. Y., & Chua, K. L. (2010). Growth-related changes in intracellular spermidine and its effect on efflux pump expression and quorum sensing in *Burkholderia pseudomallei*. *Microbiology (Reading)*, 156(Pt 4), 1144-1154. <https://doi.org/10.1099/mic.0.032888-0>
- Chattopadhyay, M. K., Tabor, C. W., & Tabor, H. (2003). Polyamines protect *Escherichia coli* cells from the toxic effect of oxygen. *Proc Natl Acad Sci U S A*, 100(5), 2261-2265. <https://doi.org/10.1073/pnas.2627990100>
- Chen, Y. J. (2007). X-ray structure of EmrE supports dual topology model. *Proc. Natl Acad. Sci. USA*, 104. <https://doi.org/10.1073/pnas.0709387104>
- Chen, Y. J., Pornillos, O., Lieu, S., Ma, C., Chen, A. P., & Chang, G. (2007). X-ray structure of EmrE supports dual topology model. *Proceedings of the National Academy of Sciences of the United States of America*, 104(48), 18999-19004. <https://doi.org/10.1073/pnas.0709387104>

- Chin, C. Y., Zhao, J., Llewellyn, A. C., Golovliov, I., Sjöstedt, A., Zhou, P., & Weiss, D. S. (2021). Francisella FlmX broadly affects lipopolysaccharide modification and virulence. *Cell Rep*, 35(11), 109247. <https://doi.org/10.1016/j.celrep.2021.109247>
- Cho, M. K., Gayen, A., Banigan, J. R., Leninger, M., & Traaseth, N. J. (2014). Intrinsic conformational plasticity of native EmrE provides a pathway for multidrug resistance. *J Am Chem Soc*, 136(22), 8072-8080. <https://doi.org/10.1021/ja503145x>
- Chung, Y. J., & Saier, M. H., Jr. (2002). Overexpression of the Escherichia coli sugE gene confers resistance to a narrow range of quaternary ammonium compounds. *J Bacteriol*, 184(9), 2543-2545. <https://doi.org/10.1128/jb.184.9.2543-2545.2002>
- Codd, R., Soe, C. Z., Pakchung, A. A. H., Sresutharsan, A., Brown, C. J. M., & Tieu, W. (2018). The chemical biology and coordination chemistry of putrebactin, avaroferrin, bisucaberin, and alcaligin. *J Biol Inorg Chem*, 23(7), 969-982. <https://doi.org/10.1007/s00775-018-1585-1>
- Coleman, G. A., Davín, A. A., Mahendrarajah, T. A., Szánthó, L. L., Spang, A., Hugenholtz, P., Szöllösi, G. J., & Williams, T. A. (2021). A rooted phylogeny resolves early bacterial evolution. *Science*, 372(6542). <https://doi.org/10.1126/science.abe0511>
- Dastvan, R., Fischer, A. W., Mishra, S., Meiler, J., & McHaourab, H. S. (2016). Protonation-dependent conformational dynamics of the multidrug transporter EmrE. *Proc. Natl Acad. Sci. USA*, 113. <https://doi.org/10.1073/pnas.1520431113>
- Di Martino, M. L., Campilongo, R., Casalino, M., Micheli, G., Colonna, B., & Prosseda, G. (2013). Polyamines: emerging players in bacteria-host interactions. *Int J Med Microbiol*, 303(8), 484-491. <https://doi.org/10.1016/j.ijmm.2013.06.008>
- Dietmann, S., & Holm, L. (2001). Identification of homology in protein structure classification. *Nat Struct Biol*, 8(11), 953-957. <https://doi.org/10.1038/nsb1101-953>
- Domingues, S., da Silva, G. J., & Nielsen, K. M. (2015). Global dissemination patterns of common gene cassette arrays in class 1 integrons. *Microbiology (Reading)*, 161(7), 1313-1337. <https://doi.org/10.1099/mic.0.000099>
- Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol*, 7(10), e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
- Elbaz, Y., Tayer, N., Steinfels, E., Steiner-Mordoch, S., & Schuldiner, S. (2005). Substrate-induced tryptophan fluorescence changes in EmrE, the smallest ion-coupled multidrug transporter. *Biochemistry*, 44. <https://doi.org/10.1021/bi050356t>
- Fleishman, S. J., Harrington, S. E., Enosh, A., Halperin, D., Tate, C. G., & Ben-Tal, N. (2006). Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *J Mol Biol*, 364(1), 54-67. <https://doi.org/10.1016/j.jmb.2006.08.072>
- Forrest, L. R. (2015). Structural Symmetry in Membrane Proteins. *Annu Rev Biophys*, 44, 311-337. <https://doi.org/10.1146/annurev-biophys-051013-023008>
- Fukuchi, J., Kashiwagi, K., Yamagishi, M., Ishihama, A., & Igarashi, K. (1995). Decrease in cell viability due to the accumulation of spermidine in spermidine acetyltransferase-deficient mutant of Escherichia coli. *J Biol Chem*, 270(32), 18831-18835. <https://doi.org/10.1074/jbc.270.32.18831>
- Funck, D., Sinn, M., Fleming, J. R., Stanoppi, M., Dietrich, J., López-Igual, R., Mayans, O., & Hartig, J. S. (2022). Discovery of a Ni(2+)-dependent guanidine hydrolase in bacteria. *Nature*, 603(7901), 515-521. <https://doi.org/10.1038/s41586-022-04490-x>

- Gayen, A., Banigan, J. R., & Traaseth, N. J. (2013). Ligand-induced conformational changes of the multidrug resistance transporter EmrE probed by oriented solid-state NMR spectroscopy. *Angew Chem Int Ed Engl*, 52(39), 10321-10324. <https://doi.org/10.1002/anie.201303091>
- Gayen, A., Leninger, M., & Traaseth, N. J. (2016). Protonation of a glutamate residue modulates the dynamics of the drug transporter EmrE. *Nat Chem Biol*, 12(3), 141-145. <https://doi.org/10.1038/nchembio.1999>
- Gaze, W. H., Abdouslam, N., Hawkey, P. M., & Wellington, E. M. (2005). Incidence of class 1 integrons in a quaternary ammonium compound-polluted environment. *Antimicrob Agents Chemother*, 49(5), 1802-1807. <https://doi.org/10.1128/aac.49.5.1802-1807.2005>
- Gillings, M., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M., & Stokes, H. W. (2008). The evolution of class 1 integrons and the rise of antibiotic resistance. *J Bacteriol*, 190(14), 5095-5100. <https://doi.org/10.1128/jb.00152-08>
- Gillings, M. R. (2017). Class 1 integrons as invasive species. *Curr. Opin. Microbiol.*, 38. <https://doi.org/10.1016/j.mib.2017.03.002>
- Gillings, M. R., Holley, M. P., & Stokes, H. W. (2009). Evidence for dynamic exchange of qac gene cassettes between class 1 integrons and other integrons in freshwater biofilms. *FEMS Microbiol Lett*, 296(2), 282-288. <https://doi.org/10.1111/j.1574-6968.2009.01646.x>
- Greener, T., Govezensky, D., & Zamir, A. (1993). A novel multicopy suppressor of a groEL mutation includes two nested open reading frames transcribed from different promoters. *Embo j*, 12(3), 889-896. <https://doi.org/10.1002/j.1460-2075.1993.tb05729.x>
- Gunn, J. S., Lim, K. B., Krueger, J., Kim, K., Guo, L., Hackett, M., & Miller, S. I. (1998). PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. *Mol Microbiol*, 27(6), 1171-1182. <https://doi.org/10.1046/j.1365-2958.1998.00757.x>
- Gupta, M. N., & Uversky, V. (2023). Moonlighting enzymes: when cellular context defines specificity. *Cellular and Molecular Life Sciences*. <https://doi.org/10.1007/s00018-023-04781-0>
- He, X. (2010). Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature*, 467. <https://doi.org/10.1038/nature09408>
- Higashi, K., Ishigure, H., Demizu, R., Uemura, T., Nishino, K., Yamaguchi, A., Kashiwagi, K., & Igarashi, K. (2008). Identification of a spermidine excretion protein complex (MdtJI) in Escherichia coli. *J Bacteriol*, 190(3), 872-878. <https://doi.org/10.1128/jb.01505-07>
- Hirao, T., Sato, M., Shirahata, A., & Kamio, Y. (2000). Covalent linkage of polyamines to peptidoglycan in Anaerovibrio lipolytica. *J Bacteriol*, 182(4), 1154-1157. <https://doi.org/10.1128/jb.182.4.1154-1157.2000>
- Hussey, G. A., Thomas, N. E., & Henzler-Wildman, K. A. (2020). Highly coupled transport can be achieved in free-exchange transport models. *J Gen Physiol*, 152(1). <https://doi.org/10.1085/jgp.201912437>
- Jack, D. L., Yang, N. M., & Saier, M. H., Jr. (2001). The drug/metabolite transporter superfamily. *Eur J Biochem*, 268(13), 3620-3639. <https://doi.org/10.1046/j.1432-1327.2001.02265.x>
- Jaehme, M., Guskov, A., & Slotboom, D. J. (2015). The twisted relation between Pnu and SWEET transporters. *Trends Biochem Sci*, 40(4), 183-188. <https://doi.org/10.1016/j.tibs.2015.02.002>

- Jelsbak, L., Thomsen, L. E., Wallrodt, I., Jensen, P. R., & Olsen, J. E. (2012). Polyamines are required for virulence in *Salmonella enterica* serovar Typhimurium. *PLoS ONE*, 7(4), e36149. <https://doi.org/10.1371/journal.pone.0036149>
- Jeong, H., & Nasir, A. (2017). A preliminary list of horizontally transferred genes in prokaryotes determined by tree reconstruction and reconciliation. *Front. Genet.*, 8. <https://doi.org/10.3389/fgene.2017.00112>
- Jurasz, J., Baginski, M., Czub, J., & Wieczor, M. (2021). Molecular mechanism of proton-coupled ligand translocation by the bacterial efflux pump EmrE. *PLoS Comput Biol*, 17(10), e1009454. <https://doi.org/10.1371/journal.pcbi.1009454>
- Kalienkova, V., Clerico Mosina, V., & Paulino, C. (2021). The Groovy TMEM16 Family: Molecular Mechanisms of Lipid Scrambling and Ion Conduction. *J Mol Biol*, 433(16), 166941. <https://doi.org/10.1016/j.jmb.2021.166941>
- Karatan, E., Duncan, T. R., & Watnick, P. I. (2005). NspS, a predicted polyamine sensor, mediates activation of *Vibrio cholerae* biofilm formation by norspermidine. *J Bacteriol*, 187(21), 7434-7443. <https://doi.org/10.1128/jb.187.21.7434-7443.2005>
- Karl, H., & Per, B. (2007). Enzyme promiscuity : mechanism and applications. *Trends in Biotechnology*. <https://doi.org/10.1016/j.tibtech.2007.03.002>
- Keller, R., Ziegler, C., & Schneider, D. (2014). When two turn into one: evolution of membrane transporters from half modules. *Biol Chem*, 395(12), 1379-1388. <https://doi.org/10.1515/hsz-2014-0224>
- Kermani, A. A., Burata, O. E., Koff, B. B., Koide, A., Koide, S., & Stockbridge, R. B. (2022). Crystal structures of bacterial small multidrug resistance transporter EmrE in complex with structurally diverse substrates. *Elife*, 11. <https://doi.org/10.7554/eLife.76766>
- Kermani, A. A., Macdonald, C. B., Burata, O. E., Ben Koff, B., Koide, A., Denbaum, E., Koide, S., & Stockbridge, R. B. (2020). The structural basis of promiscuity in small multidrug resistance transporters. *Nat Commun*, 11(1), 6064. <https://doi.org/10.1038/s41467-020-19820-8>
- Kermani, A. A., Macdonald, C. B., Gundepudi, R., & Stockbridge, R. B. (2018). Guanidinium export is the primal function of SMR family transporters. *Proc. Natl Acad. Sci. USA*, 115. <https://doi.org/10.1073/pnas.1719187115>
- Kermani, A. A., Macdonald, C. B., Gundepudi, R., & Stockbridge, R. B. (2018). Guanidinium export is the primal function of SMR family transporters. *Proceedings of the National Academy of Sciences*, 115(12), 3060. <https://doi.org/10.1073/pnas.1719187115>
- Kim, J., Tan, Y. Z., Wicht, K. J., Erramilli, S. K., Dhingra, S. K., Okombo, J., Vendome, J., Hagenah, L. M., Giacometti, S. I., Warren, A. L., Nosol, K., Roepe, P. D., Potter, C. S., Carragher, B., Kossiakoff, A. A., Quick, M., Fidock, D. A., & Mancina, F. (2019). Structure and drug resistance of the *Plasmodium falciparum* transporter PfCRT. *Nature*, 576(7786), 315-320. <https://doi.org/10.1038/s41586-019-1795-x>
- Koide, A., Wojcik, J., Gilbreth, R. N., Hoey, R. J., & Koide, S. (2012). Teaching an old scaffold new tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. *J. Mol. Biol.*, 415. <https://doi.org/10.1016/j.jmb.2011.12.019>
- Kurihara, S., Suzuki, H., Tsuboi, Y., & Benno, Y. (2009). Dependence of swarming in *Escherichia coli* K-12 on spermidine and the spermidine importer. *FEMS Microbiol Lett*, 294(1), 97-101. <https://doi.org/10.1111/j.1574-6968.2009.01552.x>
- Larrouy-Maumus, G., Škovierová, H., Dhoub, R., Angala, S. K., Zuberogitia, S., Pham, H., Villela, A. D., Mikušová, K., Noguera, A., Gilleron, M., Valentínová, L., Korduláková,

- J., Brennan, P. J., Puzo, G., Nigou, J., & Jackson, M. (2012). A small multidrug resistance-like transporter involved in the arabinosylation of arabinogalactan and lipoarabinomannan in mycobacteria. *J Biol Chem*, 287(47), 39933-39941. <https://doi.org/10.1074/jbc.M112.400986>
- Leninger, M., Sae Her, A., & Traaseth, N. J. (2019). Inducing conformational preference of the membrane protein transporter EmrE through conservative mutations. *Elife*. <https://doi.org/https://doi.org/10.7554/eLife.48909>
- Lenkeit, F., Eckert, I., Hartig, J. S., & Weinberg, Z. (2020). Discovery and characterization of a fourth class of guanine riboswitches. *Nucleic Acids Res*, 48(22), 12889-12899. <https://doi.org/10.1093/nar/gkaa1102>
- Leuzzi, A., Di Martino, M. L., Campilongo, R., Falconi, M., Barbagallo, M., Marcocci, L., Pietrangeli, P., Casalino, M., Grossi, M., Micheli, G., Colonna, B., & Prosseda, G. (2015). Multifactor Regulation of the MdtJI Polyamine Transporter in *Shigella*. *PLoS ONE*, 10(8), e0136744. <https://doi.org/10.1371/journal.pone.0136744>
- Li, J., Sae Her, A., & Traaseth, N. J. (2021). Asymmetric protonation of glutamate residues drives a preferred transport pathway in EmrE. *Proc Natl Acad Sci U S A*, 118(41). <https://doi.org/10.1073/pnas.2110790118>
- Lloris-Garcera, P., Seppälä, S., Slusky, J. S., Rapp, M., & von Heijne, G. (2014). Why have small multidrug resistance proteins not evolved into fused, internally duplicated structures? *J Mol Biol*, 426(11), 2246-2254. <https://doi.org/10.1016/j.jmb.2014.03.012>
- Lolkema, J. S., Dobrowolski, A., & Slotboom, D. J. (2008). Evolution of antiparallel two-domain membrane proteins: tracing multiple gene duplication events in the DUF606 family. *J Mol Biol*, 378(3), 596-606. <https://doi.org/10.1016/j.jmb.2008.03.005>
- Macdonald, C. B., & Stockbridge, R. B. (2017). A topologically diverse family of fluoride channels. *Curr Opin Struct Biol*, 45, 142-149. <https://doi.org/10.1016/j.sbi.2017.04.003>
- Maier, L. (2018). Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*, 555. <https://doi.org/10.1038/nature25979>
- Michael, A. J. (2018). Polyamine function in archaea and bacteria. *J Biol Chem*, 293(48), 18693-18701. <https://doi.org/10.1074/jbc.TM118.005670>
- Morrison, E. A., DeKoster, G. T., Dutta, S., Vafabakhsh, R., Clarkson, M. W., Bahl, A., Kern, D., Ha, T., & Henzler-Wildman, K. A. (2011). Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature*, 481(7379), 45-50. <https://doi.org/10.1038/nature10703>
- Morrison, E. A., & Henzler-Wildman, K. A. (2014). Transported substrate determines exchange rate in the multidrug resistance transporter EmrE. *J Biol Chem*, 289(10), 6825-6836. <https://doi.org/10.1074/jbc.M113.535328>
- Morrison, E. A., Robinson, A. E., Liu, Y., & Henzler-Wildman, K. A. (2015). Asymmetric protonation of EmrE. *J Gen Physiol*, 146(6), 445-461. <https://doi.org/10.1085/jgp.201511404>
- Moura, A. (2009). INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics*, 25. <https://doi.org/10.1093/bioinformatics/btp105>
- Mukherjee, S., Seshadri, R., Varghese, N. J., Eloie-Fadros, E. A., Meier-Kolthoff, J. P., Göker, M., Coates, R. C., Hadjithomas, M., Pavlopoulos, G. A., Paez-Espino, D., Yoshikuni, Y., Visel, A., Whitman, W. B., Garrity, G. M., Eisen, J. A., Hugenholtz, P., Pati, A., Ivanova, N. N., Woyke, T., . . . Kyrpides, N. C. (2017). 1,003 reference genomes of bacterial and

- archaeal isolates expand coverage of the tree of life. *Nat Biotechnol*, 35(7), 676-683.
<https://doi.org/10.1038/nbt.3886>
- Muth, T. R., & Schuldiner, S. (2000). A membrane-embedded glutamate is required for ligand binding to the multidrug transporter EmrE. *Embo j*, 19(2), 234-240.
<https://doi.org/10.1093/emboj/19.2.234>
- Nelson, J. W., Atilho, R. M., Sherlock, M. E., Stockbridge, R. B., & Breaker, R. R. (2017). Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Mol. Cell*, 65. <https://doi.org/10.1016/j.molcel.2016.11.019>
- Ninio, S., & Schuldiner, S. (2003). Characterization of an archaeal multidrug transporter with a unique amino acid composition. *J Biol Chem*, 278(14), 12000-12005.
<https://doi.org/10.1074/jbc.M213119200>
- Nishino, K., & Yamaguchi, A. (2001). Analysis of a complete library of putative drug transporter genes in Escherichia coli. *J Bacteriol*, 183(20), 5803-5812.
<https://doi.org/10.1128/jb.183.20.5803-5812.2001>
- Nji, E., Gulati, A., Qureshi, A. A., Coincon, M., & Drew, D. (2019). Structural basis for the delivery of activated sialic acid into Golgi for sialylation. *Nat. Struct. Mol. Biol.*, 26.
<https://doi.org/10.1038/s41594-019-0225-y>
- Ovchinnikov, V., Stone, T. A., Deber, C. M., & Karplus, M. (2018a). Structure of the EmrE multidrug transporter and its use for inhibitor peptide design. *Proc. Natl Acad. Sci. USA*, 115. <https://doi.org/10.1073/pnas.1802177115>
- Ovchinnikov, V., Stone, T. A., Deber, C. M., & Karplus, M. (2018b). Structure of the EmrE multidrug transporter and its use for inhibitor peptide design. *Proc Natl Acad Sci U S A*, 115(34), E7932-e7941. <https://doi.org/10.1073/pnas.1802177115>
- Pal, C., Bengtsson-Palme, J., Kristiansson, E., & Larsson, D. G. (2015). Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics*, 16. <https://doi.org/10.1186/s12864-015-2153-5>
- Parker, J. L., & Newstead, S. (2017). Structural basis of nucleotide sugar transport across the Golgi membrane. *Nature*, 551. <https://doi.org/10.1038/nature24464>
- Partridge, S. R., Tsafnat, G., Coiera, E., & Iredell, J. R. (2009). Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol. Rev.*, 33.
<https://doi.org/10.1111/j.1574-6976.2009.00175.x>
- Patel, C. N., Wortham, B. W., Lines, J. L., Fetherston, J. D., Perry, R. D., & Oliveira, M. A. (2006). Polyamines are essential for the formation of plague biofilm. *J Bacteriol*, 188(7), 2355-2363. <https://doi.org/10.1128/jb.188.7.2355-2363.2006>
- Payandeh, J., Scheuer, T., Zheng, N., & Catterall, W. A. (2011). The crystal structure of a voltage-gated sodium channel. *Nature*, 475. <https://doi.org/10.1038/nature10238>
- Peter, K. T. (2018). Using high-resolution mass spectrometry to identify organic contaminants linked to urban stormwater mortality syndrome in Coho Salmon. *Environ. Sci. Technol.*, 52. <https://doi.org/10.1021/acs.est.8b03287>
- Poulsen, B. E., & Deber, C. M. (2012). Drug efflux by a small multidrug resistance protein is inhibited by a transmembrane peptide. *Antimicrob Agents Chemother*, 56(7), 3911-3916.
<https://doi.org/10.1128/aac.00158-12>
- Qinghu, R., & Ian, T. P. (2007). Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes. *Journal of Molecular Microbiology and Biotechnology*. <https://doi.org/10.1159/000099639>

- Rapp, M., Granseth, E., Seppälä, S., & von Heijne, G. (2006). Identification and evolution of dual-topology membrane proteins. *Nat Struct Mol Biol*, 13(2), 112-116.
<https://doi.org/10.1038/nsmb1057>
- Roberts, T. R., & Hutson, D. H. (1999). Metabolic Pathways of Agrochemicals Part Two: Insecticides and Fungicides. *Royal Society of Chemistry*.
- Robinson, A. E., Thomas, N. E., Morrison, E. A., Balthazor, B. M., & Henzler-Wildman, K. A. (2017). New free-exchange model of EmrE transport. *Proc. Natl Acad. Sci. USA*, 114.
<https://doi.org/10.1073/pnas.1708671114>
- Rotem, D., & Schuldiner, S. (2004a). EmrE, a multidrug transporter from *Escherichia coli*, transports monovalent and divalent substrates with the same stoichiometry. *J. Biol. Chem.*, 279. <https://doi.org/10.1074/jbc.M408187200>
- Rotem, D., & Schuldiner, S. (2004b). EmrE, a multidrug transporter from *Escherichia coli*, transports monovalent and divalent substrates with the same stoichiometry. *J Biol Chem*, 279(47), 48787-48793. <https://doi.org/10.1074/jbc.M408187200>
- Rotem, D., Steiner-Mordoch, S., & Schuldiner, S. (2006). Identification of tyrosine residues critical for the function of an ion-coupled multidrug transporter. *J. Biol. Chem.*, 281.
<https://doi.org/10.1074/jbc.M602088200>
- Russell, A. D. (2002). Introduction of biocides into clinical practice and the impact on antibiotic-resistant bacteria. *J. Appl. Microbiol*, 92 Suppl. <https://doi.org/10.1046/j.1365-2672.92.5s1.12.x>
- Saleh, M., Bay, D. C., & Turner, R. J. (2018). Few Conserved Amino Acids in the Small Multidrug Resistance Transporter EmrE Influence Drug Polyselectivity. *Antimicrob Agents Chemother*, 62(8). <https://doi.org/10.1128/aac.00461-18>
- Salvail, H., Balaji, A., Yu, D., Roth, A., & Breaker, R. R. (2020). Biochemical Validation of a Fourth Guanidine Riboswitch Class in Bacteria. *Biochemistry*, 59(49), 4654-4662.
<https://doi.org/10.1021/acs.biochem.0c00793>
- Sarti, E., Aleksandrova, A. A., Ganta, S. K., Yavatkar, A. S., & Forrest, L. R. (2019). EncoMPASS: an online database for analyzing structure and symmetry in membrane proteins. *Nucleic Acids Res*, 47(D1), D315-d321. <https://doi.org/10.1093/nar/gky952>
- Scheurer, M., Michel, A., Brauch, H. J., Ruck, W., & Sacher, F. (2012). Occurrence and fate of the antidiabetic drug metformin and its metabolite guanyurea in the environment and during drinking water treatment. *Water Res.*, 46.
<https://doi.org/10.1016/j.watres.2012.06.019>
- Schneider, N. O., Tassoulas, L. J., Zeng, D., Laseke, A. J., Reiter, N. J., Wackett, L. P., & Maurice, M. S. (2020). Solving the Conundrum: Widespread Proteins Annotated for Urea Metabolism in Bacteria Are Carboxyguanidine Deiminases Mediating Nitrogen Assimilation from Guanidine. *Biochemistry*, 59(35), 3258-3270.
<https://doi.org/10.1021/acs.biochem.0c00537>
- Schuldiner, S. (2009). EmrE, a model for studying evolution and mechanism of ion-coupled transporters. *Biochim Biophys Acta*, 1794(5), 748-762.
<https://doi.org/10.1016/j.bbapap.2008.12.018>
- Shah, P., & Swiatlo, E. (2008). A multifaceted role for polyamines in bacterial pathogens. *Mol Microbiol*, 68(1), 4-16. <https://doi.org/10.1111/j.1365-2958.2008.06126.x>
- Shcherbakov, A. A., Hisao, G., Mandala, V. S., Thomas, N. E., Soltani, M., Salter, E. A., Davis, J. H., Jr., Henzler-Wildman, K. A., & Hong, M. (2021). Structure and dynamics of the

- drug-bound bacterial transporter EmrE in lipid bilayers. *Nat Commun*, 12(1), 172.
<https://doi.org/10.1038/s41467-020-20468-7>
- Shcherbakov, A. A., Spreacker, P. J., Dregni, A. J., Henzler-Wildman, K. A., & Hong, M. (2022). High-pH structure of EmrE reveals the mechanism of proton-coupled substrate transport. *Nat Commun*, 13(1), 991. <https://doi.org/10.1038/s41467-022-28556-6>
- Sherlock, M. E., & Breaker, R. R. (2017). Biochemical Validation of a Third Guanidine Riboswitch Class in Bacteria. *Biochemistry*, 56(2), 359-363.
<https://doi.org/10.1021/acs.biochem.6b01271>
- Sherlock, M. E., Malkowski, S. N., & Breaker, R. R. (2017). Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria. *Biochemistry*, 56(2), 352-358.
<https://doi.org/10.1021/acs.biochem.6b01270>
- Sikora, C. W., & Turner, R. J. (2005). SMR proteins SugE and EmrE bind ligand with similar affinity and stoichiometry. *Biochem Biophys Res Commun*, 335(1), 105-111.
<https://doi.org/10.1016/j.bbrc.2005.07.051>
- Singh, A. K., Saotome, K., & Sobolevsky, A. I. (2017). Swapping of transmembrane domains in the epithelial calcium channel TRPV6. *Sci Rep*, 7(1), 10669.
<https://doi.org/10.1038/s41598-017-10993-9>
- Sinn, M., Hauth, F., Lenkeit, F., Weinberg, Z., & Hartig, J. S. (2021). Widespread bacterial utilization of guanidine as nitrogen source. *Mol Microbiol*, 116(1), 200-210.
<https://doi.org/10.1111/mmi.14702>
- Slipski, C. J. (2020). Plasmid transmitted small multidrug resistant (SMR) efflux pumps differ in gene regulation and enhance tolerance to quaternary ammonium compounds (QAC) when grown as biofilms. . <https://doi.org/bioRxivhttps://doi.org/10.1101/768630>
- Slipski, C. J., Jamieson, T. R., Lam, A., Shing, V. L., Bell, K., Zhanel, G. G., & Bay, D. C. (2020). Plasmid transmitted small multidrug resistant (SMR) efflux pumps differ in gene regulation and enhance tolerance to quaternary ammonium compounds (QAC) when grown as biofilms. *bioRxiv*, 768630. <https://doi.org/10.1101/768630>
- Slipski, C. J., Jamieson, T. R., Zhanel, G. G., & Bay, D. C. (2020). Riboswitch-Associated Guanidinium-Selective Efflux Pumps Frequently Transmitted on Proteobacterial Plasmids Increase Escherichia coli Biofilm Tolerance to Disinfectants. *J Bacteriol*, 202(23). <https://doi.org/10.1128/jb.00104-20>
- Spreacker, P. J., Thomas, N. E., Beeninga, W. F., Brousseau, M., Porter, C. J., Hibbs, K. M., & Henzler-Wildman, K. A. (2022). Activating alternative transport modes in a multidrug resistance efflux pump to confer chemical susceptibility. *Nat Commun*, 13(1), 7655.
<https://doi.org/10.1038/s41467-022-35410-2>
- Stockbridge, R. B., Kolmakova-Partensky, L., Shane, T., Koide, A., Koide, S., Miller, C., & Newstead, S. (2015). Crystal structures of a double-barrelled fluoride ion channel. *Nature*, 525(7570), 548-551. <https://doi.org/10.1038/nature14981>
- Stockbridge, R. B., Robertson, J. L., Kolmakova-Partensky, L., & Miller, C. (2013). A family of fluoride-specific ion channels with dual-topology architecture. *Elife*, 2, e01084.
<https://doi.org/10.7554/eLife.01084>
- Sturgill, G., & Rather, P. N. (2004). Evidence that putrescine acts as an extracellular signal required for swarming in *Proteus mirabilis*. *Mol Microbiol*, 51(2), 437-446.
<https://doi.org/10.1046/j.1365-2958.2003.03835.x>
- Takatsuka, Y., & Kamio, Y. (2004). Molecular dissection of the *Selenomonas ruminantium* cell envelope and lysine decarboxylase involved in the biosynthesis of a polyamine

- covalently linked to the cell wall peptidoglycan layer. *Biosci Biotechnol Biochem*, 68(1), 1-19. <https://doi.org/10.1271/bbb.68.1>
- Tassoulas, L. J., Robinson, A., Martinez-Vaz, B., Aukema, K. G., & Wackett, L. P. (2021). Filling in the Gaps in Metformin Biodegradation: a New Enzyme and a Metabolic Pathway for Guanylurea. *Appl Environ Microbiol*, 87(11). <https://doi.org/10.1128/aem.03003-20>
- Theobald, D. L., & Miller, C. (2010). Membrane transport proteins: surprises in structural sameness. *Nat Struct Mol Biol*, 17(1), 2-3. <https://doi.org/10.1038/nsmb0110-2>
- Theobald, D. L., & Wuttke, D. S. (2005). Divergent evolution within protein superfolds inferred from profile-based phylogenetics. *J Mol Biol*, 354(3), 722-737. <https://doi.org/10.1016/j.jmb.2005.08.071>
- Thomas, N. E., Feng, W., & Henzler-Wildman, K. A. (2021). A solid-supported membrane electrophysiology assay for efficient characterization of ion-coupled transport. *J Biol Chem*, 297(4), 101220. <https://doi.org/10.1016/j.jbc.2021.101220>
- Thomas, N. E., Wu, C., Morrison, E. A., Robinson, A. E., Werner, J. P., & Henzler-Wildman, K. A. (2018). The C terminus of the bacterial multidrug transporter EmrE couples drug binding to proton release. *J Biol Chem*, 293(49), 19137-19147. <https://doi.org/10.1074/jbc.RA118.005430>
- Trent, M. S., Ribeiro, A. A., Lin, S., Cotter, R. J., & Raetz, C. R. (2001). An inner membrane enzyme in Salmonella and Escherichia coli that transfers 4-amino-4-deoxy-L-arabinose to lipid A: induction on polymyxin-resistant mutants and role of a novel lipid-linked donor. *J Biol Chem*, 276(46), 43122-43131. <https://doi.org/10.1074/jbc.M106961200>
- Tsuchiya, H. (2016). Structural basis for amino acid export by DMT superfamily transporter YddG. *Nature*, 534. <https://doi.org/10.1038/nature17991>
- Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S., & Tate, C. G. (2003). Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J.*, 22. <https://doi.org/10.1093/emboj/cdg611>
- Västermark, Å., Almén, M. S., Simmen, M. W., Fredriksson, R., & Schiöth, H. B. (2011). Functional specialization in nucleotide sugar transporters occurred through differentiation of the gene cluster EamA (DUF6) before the radiation of Viridiplantae. *BMC Evol Biol*, 11, 123. <https://doi.org/10.1186/1471-2148-11-123>
- Vaughan, V. C., & Novy, F. G. (1891). Ptomaines, Leucomaines, and Bacterial Proteids: Or the Chemical factors in the causation of disease. *Lea Brothers & Co.*
- Vermaas, J. V., Rempe, S. B., & Tajkhorshid, E. (2018). Electrostatic lock in the transport cycle of the multidrug resistance transporter EmrE. *Proc. Natl Acad. Sci. USA*, 115. <https://doi.org/10.1073/pnas.1722399115>
- Wales, A. D., & Davies, R. H. (2015). Co-selection of resistance to antibiotics, biocides and heavy metals, and its relevance to foodborne pathogens. *Antibiotics (Basel)*, 4. <https://doi.org/10.3390/antibiotics4040567>
- Wang, B., Dong, T., Myrli, A., Gu, L., Zhu, H., Xiong, W., Maness, P., Zhou, R., & Yu, J. (2019). Photosynthetic production of the nitrogen-rich compound guanidine [10.1039/C9GC01003C]. *Green Chemistry*, 21(11), 2928-2937. <https://doi.org/10.1039/C9GC01003C>
- Wang, B., Xu, Y., Wang, X., Yuan, J. S., Johnson, C. H., Young, J. D., & Yu, J. (2021). A guanidine-degrading enzyme controls genomic stability of ethylene-producing cyanobacteria. *Nat Commun*, 12(1), 5150. <https://doi.org/10.1038/s41467-021-25369-x>

- Ware, D., Jiang, Y., Lin, W., & Swiatlo, E. (2006). Involvement of potD in *Streptococcus pneumoniae* polyamine transport and pathogenesis. *Infect Immun*, 74(1), 352-361. <https://doi.org/10.1128/iai.74.1.352-361.2006>
- William, M. A. (2015). Biological messiness vs. biological genius: Mechanistic aspects and roles of protein promiscuity. *The Journal of Steroid Biochemistry and Molecular Biology*. <https://doi.org/10.1016/j.jsbmb.2014.09.010>
- Wortham, B. W., Patel, C. N., & Oliveira, M. A. (2007). Polyamines in bacteria: pleiotropic effects yet specific mechanisms. *Adv Exp Med Biol*, 603, 106-115. https://doi.org/10.1007/978-0-387-72124-8_9
- Wu, C. (2019). Identification of an alternating-access dynamics mutant of EmrE with impaired transport. *J. Mol. Biol.*, 431. <https://doi.org/10.1016/j.jmb.2019.05.035>
- Yan, A., Guan, Z., & Raetz, C. R. (2007). An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in *Escherichia coli*. *J Biol Chem*, 282(49), 36077-36089. <https://doi.org/10.1074/jbc.M706172200>
- Yerushalmi, H., Lebendiker, M., & Schuldiner, S. (1995). EmrE, an *Escherichia coli* 12-kDa multidrug transporter, exchanges toxic cations and H⁺ and is soluble in organic solvents. *J Biol Chem*, 270(12), 6856-6863. <https://doi.org/10.1074/jbc.270.12.6856>
- Yerushalmi, H., & Schuldiner, S. (2000). An essential glutamyl residue in EmrE, a multidrug antiporter from *Escherichia coli*. *J Biol Chem*, 275(8), 5264-5269. <https://doi.org/10.1074/jbc.275.8.5264>
- Zahn, D. (2019). Identification of potentially mobile and persistent transformation products of REACH-registered chemicals and their occurrence in surface waters. *Water Res.*, 150. <https://doi.org/10.1016/j.watres.2018.11.042>
- Zhu, Y. G. (2017). Continental-scale pollution of estuaries with antibiotic resistance genes. *Nat. Microbiol.*, 2. <https://doi.org/10.1038/nmicrobiol.2016.270>

Chapter 2 Structural Basis of Promiscuity in Small Multidrug Transporters

This chapter is adapted from the following published research article:

Kermani, A.A.[#], Macdonald, C.B.[#], Burata, O.E., Koff, B.B., Koide, A., Denbaum, E., Koide, S. & Stockbridge, R.B. The structural basis of promiscuity in small multidrug resistance transporters. Nat Commun, 11, 6064 (2020). <https://doi.org/10.1038/s41467-020-19820-8>

- equal contributions

Conceptualization, A.A.K., C.B.M., and R.B.S.; Methodology, A.A.K., C.B.M., A.K., S.K., and R.B.S.; Validation, O.E.B.; Formal analysis, A.A.K., C.B.M., and R.B.S.; Investigation, A.A.K., C.B.M., O.E.B., B.B.K., A.K., and E.D.; Writing—original draft, A.A.K., C.B.M., and R.B.S.; Writing—review and editing, A.A.K., C.B.M., O.E.B., A.K., S.K., and R.B.S.; Visualization, A.A.K., C.B.M., O.E.B. and R.B.S.; Supervision, R.B.S.; Project administration, R.B.S.; Funding acquisition, S.K. and R.B.S.

2.1 Introduction

2.1.1 Personal contributions to the published researched article

Although a general list of contributions has been listed above in the completion of the research work, I describe additional scientific contributions from me towards the completion of this published body in more detail. The main goals of this publication include the following: introduction of the first side-chain resolution structure of an SMR transporter along with structures in complex with two individual substrates, teasing apart the chemical properties of

SMR substrates, and *in vitro* functional characterization of Gdx-Clo binding pocket residues and its respective mutants. To obtain the structures of Gdx-Clo we utilized a molecular chaperone for successful crystal formation and structure determination. This method brought forth concerns regarding the potential effects of monobody binding upon protein. To address this, I developed a modified technique requiring solid-supported membrane-based electrophysiology (SSME), that allowed us to observe transport and quantify the inhibitory effect of L10 monobody on the transport cycle. This effort allowed us to confidently assess a functional form of the WT Gdx-Clo crystal structures without concern for monobody interference.

Prior to the determination of the first crystal structure of Gdx-Clo, my main goal was to determine which mutation or set of mutations were responsible for the polyspecific behavior exhibited by SMR_{Qacs}. To do so, I relied on rational mutagenesis based on sequence alignments between several SMR homologs from each of the major subtypes to determine which positions contained highly conserved residues that are unique in each independent subtype. I designed mutations in Gdx-Clo and Gdx-Eco (an *E. coli* SMR homolog) and measured their ability to transport various substrates via solid-support membrane electrophysiology (SSME). **Table 2.1** lists the mutations I have designed and tested along with all their respective phenotypes. Gdm⁺, a substrate transported only by SMR_{Gdx} and not SMR_{Qacs}, was used to determine if the mutant maintain wildtype function. Phegdm⁺, an aromatic compound with a guanidinyll moiety transported by both Gdx and Qacs, was used to ensure that the variant maintained overall functionality despite any substrate specificity changes or any massive impairments in function that may not be detected with SSME sensitivity. Other overlapping substrates between Qacs and Gdx may have been used in lieu of Phegdm⁺, which served the same purpose. The only exception to this rule is dimethylGdm⁺, a substrate that bound robustly to Gdx-Clo but not

transported. We hypothesized that testing the transport capabilities of the variants at the cusp of substrate specificity change with this compound, may result in its successful transport.

Tetrapropyl ammonium⁺, tetraethyl ammonium⁺, and tetramethyl ammonium⁺ are substrates robustly transported by SMR_{Qacs} and were used to screen any mutants that may have successfully changed specificities. In my journey to designing these mutants (mostly single point mutations, and several rational combinations of mutations in both Gdx-Clo and Gdx-Eco), none were able to transport any QAC-like substrates. My initial investigation of these mutants, though, led to a synergistic investigation of the conserved residues in the binding pocket of Gdx-Clo that established the foundation of a strong H-bond network. In addition, the SMR literature has been dominated by characterization of SMR_{Qacs}, specifically EmrE, for almost three decades. My overall contributions to this work laid the foundation for the massive mutagenesis campaign described in Chapter 4 and permitted comparison of the mutational studies and knowledge we obtained from each subtype of the SMR family.

2.1.2 General overview

Membrane proteins from the small multidrug resistance (SMR) family are a major driver of the spread of drug resistance genes among bacteria. Genes encoding SMR proteins (variously annotated *emrE*, *sugE*, *smr*, *qac*, *ebr*) are frequently found in mobile drug resistance gene arrays, and provide a broad selective advantage by conferring resistance to ubiquitous environmental pollutants with low-grade toxicity to microbes (Gillings, 2017; Zhu, 2017). The adaptive effects of the SMR proteins lead to co-selection of other genes in the arrays that confer resistance to the more potent drugs in the antimicrobial arsenal, including sulfonamides, β lactams, and aminoglycosides, increasing the frequency of these genes in environmental reservoirs 3,4. Thus, the dispersal of drug resistance genes among bacteria, the transport capabilities of SMR proteins,

and the distribution of SMR substrates in the biosphere are intimately linked. Despite their importance, functional experiments to test the chemical scope of transported compounds have been limited to a narrow range of SMR homologs and drugs, and although the overall fold has been determined⁵, sidechain-resolution structural data have not been reported for any family member. In this study, we have two objectives: (1) determine the chemical characteristics of substrates transported by representative SMR family proteins; and (2) establish the structural basis of substrate binding and transport by SMR transporters. The sequence diversity of bacterial SMR exporters can be visualized using a sequence-similarity network (**Fig. 2.1a and Supplementary Fig A.1**). The SMR family has two major subtypes that share high sequence identity (~40%) and similarity (**Supplementary Fig. A.2**). Both are broadly distributed across bacterial taxa, and many bacteria possess both subtypes. One group contains proteins that provide resistance against quaternary ammonium cations, including structurally diverse polyaromatic cations such as ethidium and methyl viologen. This group, the Qac cluster, includes EmrE, an *Escherichia coli* homolog and the best-studied member of the SMR family. The other group was characterized more recently, and encompasses guanidinium (Gdm^+)/ H^+ antiporters (Gdx proteins; *E. coli* gene name *sugE*) (Kermani et al., 2020). Gdm^+ is an endogenously produced, nitrogen-rich metabolite that is transformed or exported by genes organized in Gdm^+ -related operons. These operons are often controlled by riboswitches that are selectively responsive to Gdm^+ binding (Battaglia & Ke, 2018; Breaker et al., 2017) (**Fig. 2.1a**). Initial experiments suggested that the Qac and Gdx subtypes fulfill discrete functional roles, since EmrE does not transport Gdm^+ , and the Gdx proteins do not transport canonical EmrE drugs (Kermani et al., 2020). Of the two roles, export of quaternary ammonium ions is most readily associated with multidrug resistance, since these compounds have been used as antiseptics for

almost a century (Russell, 2002). But genes from the Gdx cluster also commonly colocalize with horizontal gene transfer elements (**Fig. 2.1a**) (Jeong & Nasir, 2017; Slipski, 2020), and have been explicitly identified in mobile multidrug resistance gene arrays^{12,13} (**Fig. 2.1**). Is the functional dichotomy between the Qac and Gdx subtypes as strict as early experiments suggested? Or do proteins in the SMR family share transport capabilities that make them broadly adaptive in human-impacted environments? Here we show that SMR proteins from both the Qac and Gdx subtypes engage in promiscuous transport of hydrophobic substituted cations. Both subtypes transport a variety of hydrophobic guanidinyll compounds, and proteins belonging to the Qac subtype additionally transport substituted ammonium compounds and polyaromatic cations. X-ray crystal structures of Gdx-Clo in complex with substituted guanidinyll substrates reveal a cleft between two helices that provides accommodation in the membrane for the hydrophobic substituents of transported drug-like cations.

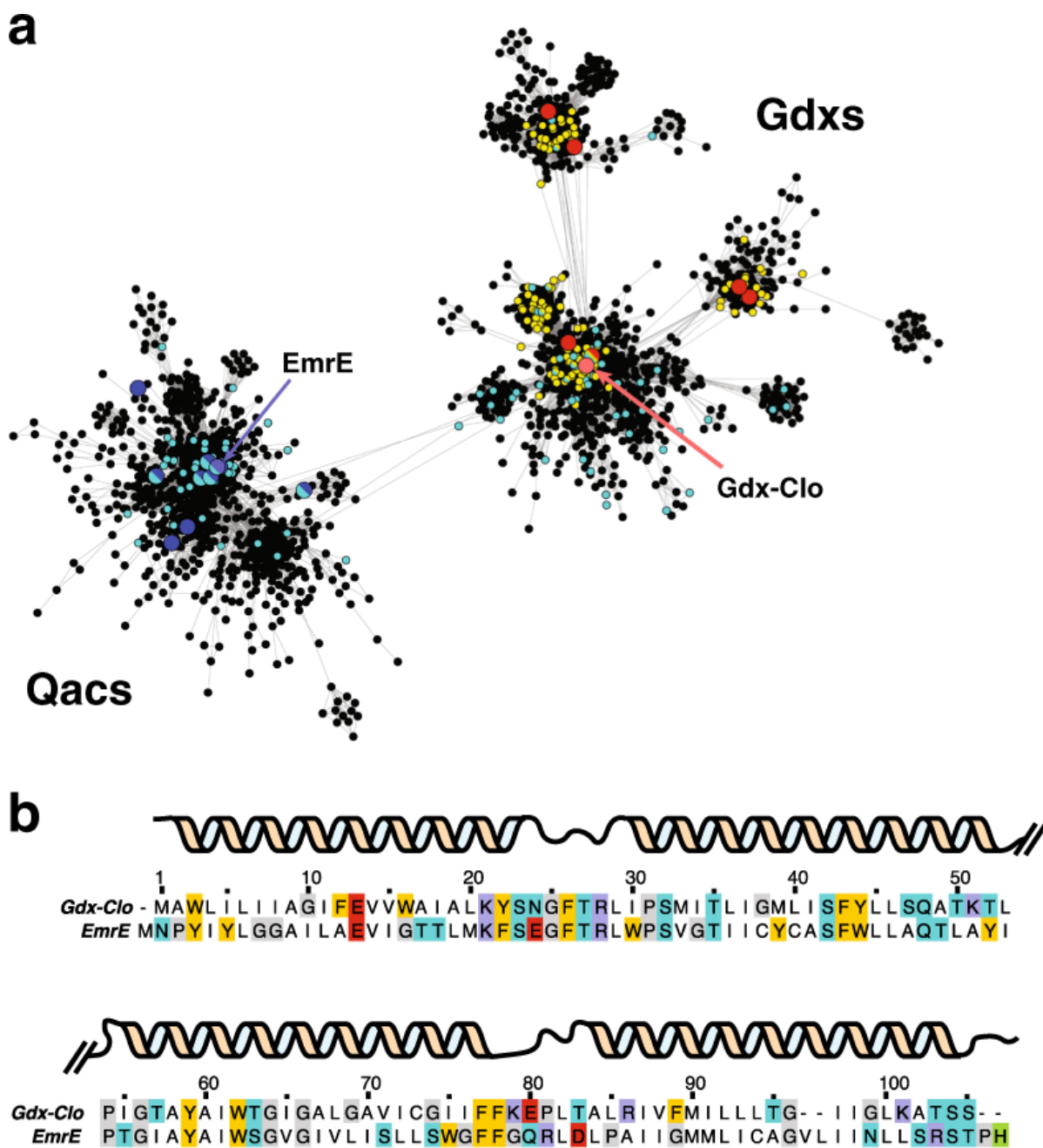


Figure 2.1: Colocalization of SMR genes with guanidine riboswitches and horizontal gene transfer elements.

a Major clusters (>10% of total set) from the sequence-similarity network of the SMR family. Each point corresponds to a cluster of sequences with >50% sequence identity. Edges between points correspond to a pairwise E value of at least 10^{-20} . Biochemically characterized proteins are shown as enlarged red (Gdx) or blue (Qac) points. SMR sequences associated with a guanidine riboswitch are colored in yellow. SMR sequences found on plasmids or genetically colocalized with integron/integrase sequences are colored cyan. Full sequence-similarity network shown in **Supplementary Fig. 1**. **b** Sequence alignment of EmrE and Gdx-Clo with positions of α -helices indicated.

2.2 Results

2.2.1 Overlapping, promiscuous substrate transport by *Qac* and *Gdx* subtypes

In order to probe the chemical characteristics of transported substrates, we performed transport experiments using exemplars of both the *Qac* and *Gdx* subtypes (**Fig. 2.1b**): the polyaromatic cation exporter EmrE, and Gdx-Clo, a functionally characterized Gdm⁺ transporter from Clostridiales oral taxon 876 (Kermani et al., 2020; Kermani et al., 2018; Nelson et al., 2017). Radioactive uptake assays confirm that EmrE transports methyl viologen, but not Gdm⁺; that Gdx-Clo transports Gdm⁺, but not methyl viologen; and that both proteins discriminate against a substituted guanidinyll metabolite, agmatine (**Fig. 2.2a**). To expand the repertoire of substrates tested, we used solid supported membrane (SSM) electrophysiology. These experiments are feasible because the transport cycle of SMR proteins is electrogenic: the *Gdx* proteins couple import of two H⁺ with export of one Gdm⁺ ion (Kermani et al., 2018), and EmrE, though it has been shown to stray slightly from strict 2:1 stoichiometry, transports monovalent substrates in an electrogenic manner as well (Rotem & Schuldiner, 2004; Ubarretxena-Belandia et al., 2003). In SSM electrophysiology, proteoliposomes are capacitively coupled to a gold electrode by adsorption to a lipid monolayer. When the liposomes containing SMR proteins are perfused with substrate, initiating electrogenic transport, transient capacitive currents are evoked (**Fig. 2.2b**). The peak currents are negative, consistent with a net-negative transport cycle expected for 2 H⁺:1 substrate⁺ exchange. (In contrast, translocation of a positively charged substrate, without concomitant proton antiport, would be expected to evoke a positive current.) The amplitudes of the currents are proportional to the initial rate of transport, but decay rapidly to baseline as a membrane potential builds up in the liposomes and the system achieves electrochemical equilibrium. Subsequent replacement of the substrate-containing solution with a

substrate-free solution yields a transient current of the opposite polarity, reflecting efflux of the accumulated substrate from the liposomes, and a return to the starting condition (**Fig. 2.2b**). We tested substrates in the following categories: Gdm^+ , guanidinylated metabolites, hydrophobic substituted guanidinium ions, and hydrophobic substituted amines. For all of these, analogous experiments with protein-free liposomes exhibit no currents (**Supplementary Fig. A.3**). (In contrast, polyaromatic molecules like ethidium and tetraphenylphosphonium produced currents due to nonspecific partition into the membrane and were therefore not analyzed here; **Supplementary Fig. A.4**.) Because an unexpected shift in stoichiometry to 1 H^+ :1 substrate antiport would be electrically silent, all negative results were validated using a second method, exchange with radiolabeled substrate (**Supplementary Fig. A.5** or (**Kermani et al., 2018**). We observed no discrepancies between the electrophysiological results and the radioactive uptake experiments. Our electrophysiology experiments (**Fig. 2.2c**) recapitulate prior observations for metabolites: EmrE does not transport Gdm^+ , and both proteins are strongly selective against substituted guanidinium metabolites like arginine, agmatine, and creatine⁶. However, many of the non-natural compounds we tested were readily transported by both subtypes. Gdx-Clo transported guanidinyll compounds with hydrophobic single substitutions, including the bulky phenyl Gdm^+ . Currents decreased for doubly substituted guanidinyll compounds and were absent for tetramethyl Gdm^+ . Compared with Gdx-Clo, EmrE required additional hydrophobicity and bulk in its substrates. In agreement with the radiolabeled Gdm^+ uptake experiments, Gdm^+ was not transported by EmrE. However, methyl-, ethyl-, and phenyl Gdm^+ evoked increasingly larger currents. In contrast to Gdx-Clo, EmrE also accommodated substrates with reduced or no H-bonding capacity, tetramethyl Gdm^+ and tetramethylammonium, respectively. These experiments show that polyaromaticity is not a requirement for transport by EmrE. Moreover, these

experiments make clear that functional promiscuity is a general trait of the SMR family. The relative transport specificities are summarized in **Supplementary Fig. A.6**.

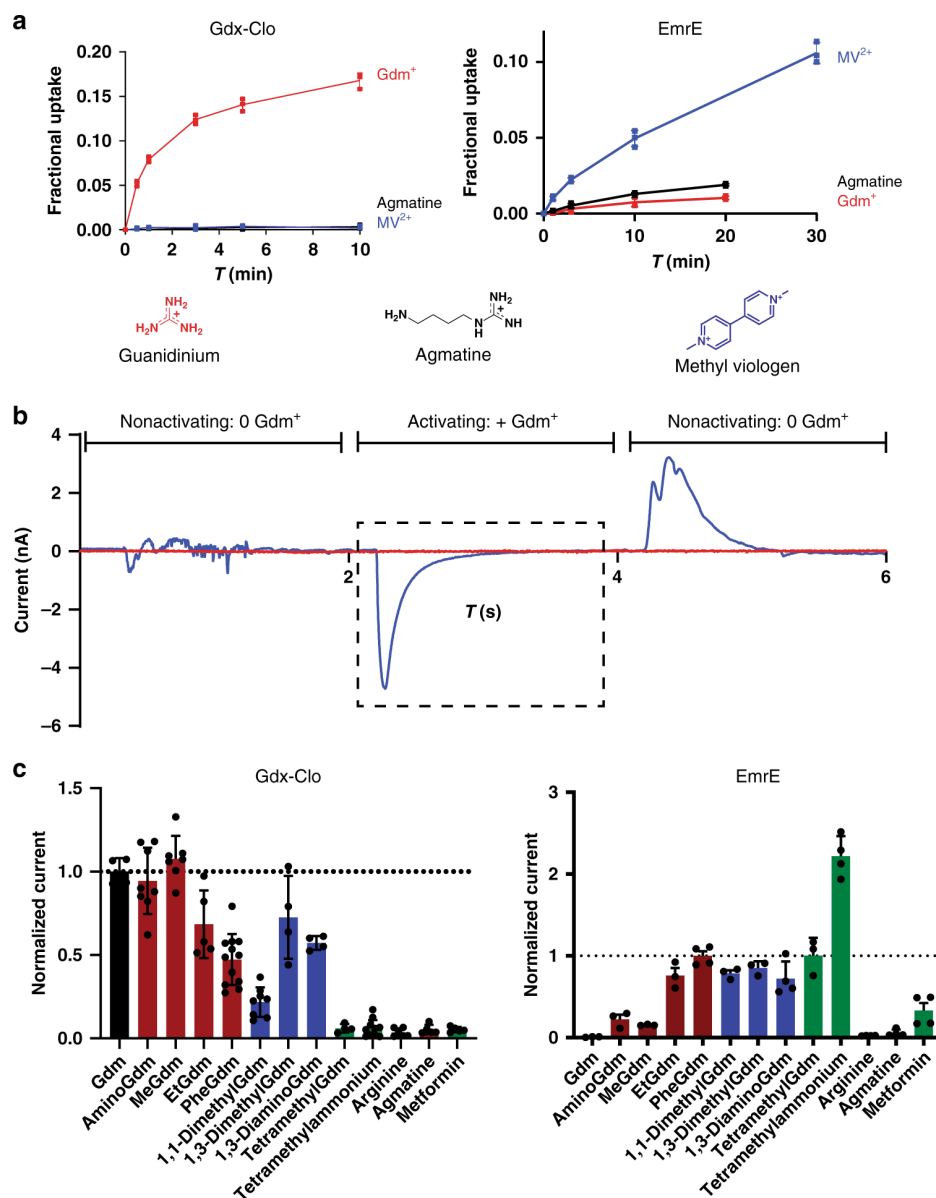


Figure 2.2: Substrate transport of Gdx-Clo and EmrE.

a Radioactive exchange assays. For Gdx-Clo and EmrE, uptake of ^{14}C -Gdm $^+$ or ^{14}C -methyl viologen, respectively, monitored in exchange for the indicated substrate. Points represent individual replicates; error bars represent the mean \pm SEM from three independent experiments. **b** Typical SSM electrical recording of Gdx-Clo proteoliposomes perfused with indicated solutions. The area in the dashed box is used to determine initial rate kinetics. **c** Initial rate of substrate transport (peak currents) by Gdx-Clo (normalized to Gdm $^+$ currents, left panel) and EmrE (normalized to phenylGdm $^+$ currents, right panel). Singly substituted guanidiny compounds are shown as maroon bars, doubly substituted guanidiny compounds are shown as blue bars, with all other compounds shown as green bars. Data were collected from 3 to 4 independent sensor preparations, which were in turn prepared from 2 to 4 independent protein preparations. Individual measurements are shown as points, and error bars represent \pm SEM.

2.2.2 Crystal structure of Gdx-Clo

What molecular features of SMR proteins enable these promiscuous transport functions, while simultaneously prohibiting export of endogenous substituted guanidinium metabolites? Even though this family has proved endlessly intriguing to biochemists, as one of just a few idiosyncratic examples of primitive dual topology antiparallel dimers, the only structural models available include a 7 Å electron microscopy structure of EmrE (Ubarretxena-Belandia et al., 2003), and an X-ray crystal structure of EmrE (Chen, 2007) that has notable deficiencies: it is presented as a C α model without sidechains, and helices that are not long enough to span the membrane and have flawed helical geometry. Computational models of EmrE that build on the low-resolution structural data have also been put forth (Ovchinnikov et al., 2018; Vermaas et al., 2018). In order to rectify the gap in structural information for the SMR family, we focused our crystallography efforts on Gdx-Clo. Though this protein crystallized readily, the crystals diffracted poorly. To improve diffraction, we generated monobodies, synthetic binding proteins based on the human fibronectin type III domain, to use as crystallization chaperones (Koide et al., 2012). Upon optimization, we obtained crystals of Gdx-Clo in complex with one of the monobodies, termed Clo-L10, that diffracted to 3.5 Å Bragg spacing, and we solved the structure with phases determined by single-wavelength anomalous diffraction (SAD) of selenomethionine-substituted samples (**Supplementary Table 3 and Supplementary Fig. A.7**). Ellipsoidal truncation of the anisotropic datasets and addition of substituted Gdm⁺ substrates further improved resolution, ultimately to 2.3 Å. The asymmetric unit contains one Gdx-Clo dimer and two Clo-L10 monobodies, one bound to each subunit. The monobodies primarily use residues diversified in the library to bind to residues 24–32 from loop 1 of each Gdx-Clo subunit in slightly different orientations, each forming a ~400 Å² interface (**Fig. 2.3a and Supplementary**

Fig. A.8). In electrophysiology experiments, Gdm^+ currents mediated by Gdx-Clo decreased upon addition of Clo-L10, but fractional inhibition saturated at ~40%, suggesting that monobody complexation is not incompatible with the transport cycle (**Supplementary Fig. A.9**).

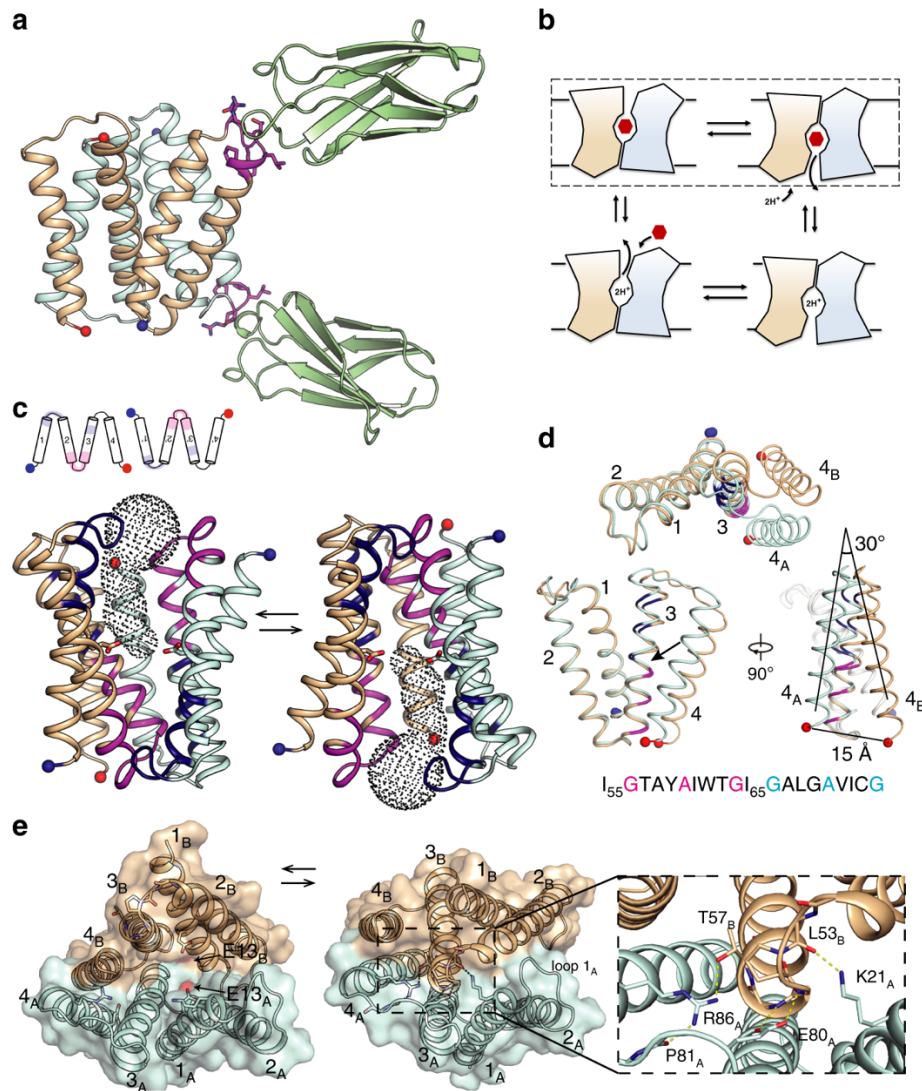


Figure 2.3: Substrate transport of Gdx-Clo and EmrE.

a Structure of Gdx-Clo/monobody complex. Clo-L10 monobodies are shown in green. Transporter shown with subunit A in light blue and subunit B in tan. The N- and C-termini for each subunit are shown as blue and red spheres, respectively. Transporter residues that comprise the monobody binding interface are shown in magenta. **b** Cartoon schematic showing transport cycle for an antiparallel homodimer. The dashed box indicates the conformational exchange step highlighted in panel (c). **c** Changes in accessibility during conformational exchange. For both the upper diagram, and the lower three-dimensional structure, regions that alternate in solvent accessibility are shown in magenta (TM2, loop 2, and the first GxxxAxxxG motif of TM3) and dark blue (TM1, loop 1, and second GxxxAxxxG motif of TM3). The N- and C-termini are shown as blue and red spheres. In the three-dimensional structure, E13 sidechains shown as sticks and solvent-accessible vestibule indicated with dots. **d** Overlay of Gdx-Clo A and B subunits aligned over C α 1–61. The sequence of TM3 is shown with GxxxAxxxG packing motifs colored in magenta and dark blue in structures and sequence. An arrow indicates I₆₅. Three views are shown (counterclockwise from top): top-down view, view through the plane of the membrane (with I₆₅G indicated with an arrow), and rotated 90°. **e** Conformational exchange viewed from top down. E13 sidechains shown in red as surface representation and indicated with arrows. Sidechains that make polar contacts on the closed side of the transporter are shown as sticks.

2.2.3 The structural basis for conformational exchange.

The two 4- TM helix subunits of Gdx-Clo are arranged antiparallel with respect to each other in non-equivalent “A” and “B” conformations. The overall fold agrees with previous low-resolution structural models of EmrE (Ubarretxena-Belandia et al., 2003), and our designation of A and B subunits matches that used for EmrE. A large aqueous chamber is open to one side of the membrane, with the strictly conserved substrate- and proton-binding glutamates, E13_A and E13_B, accessible at the bottom. Positive density is visible between the E13 sidechains but cannot be definitively assigned as Gdm⁺ at this resolution (**Supplementary Fig. A.10**). Transport by the antiparallel SMR proteins involves a conformational swap between the two structurally distinct monomers, which seals the substrate binding site on one side of the membrane while opening an identical site on the opposite side (**Fig. 2.3b, c**). Because of the antiparallel homodimeric architecture, there is no structural difference between the inward-open and outward-open conformations: they are structurally identical and related by twofold symmetry about an axis parallel to the plane of the membrane. To visualize conformational exchange, we have rendered this structure in both the inward- and outward-facing directions (**Fig. 2.3c**). The crux of the conformational exchange is helix 3 (G₅₆xxxAxxTG₆₄IGxxxAxxxG), which possesses two GxxxAxxxG helical packing motifs offset from each other by two amino acids, or just over 180°. The G₆₄IG sequence at the helical midpoint is the fulcrum between an N-terminal domain (TM1, TM2, and the first half of TM3) and a C-terminal domain (the second half of TM3 and TM4). Comparing subunit A and subunit B, the domains possess near structural identity (RMSD 0.5 Å for Cα 1–62), but are offset by a rigid body rotation of about 30° (**Fig. 2.3d**). In agreement with our observations, the analogous G₆₄VG sequence in EmrE has been identified in EPR studies as a “kink” about which the conformational change occurs (Amadi et al., 2010). Inspection of the

regions that change in accessibility during the transport cycle shows that, for each TM3, only one of the two GxxxAxxxG packing motifs is buried at one time, and that burial alternates with conformational exchange (**Fig. 2.3c and Supplementary Fig. A.11**). We posit that competition between the two halves of TM3 to pack against structurally complementary regions of the protein contributes to the structural frustration and conformational exchange in the Gdx transporters. In addition, T63, which immediately precedes the GIG sequence, is in a position to backbond to the mainchain and further perturb the helical geometry. Mutation of the analogous serine at this position in EmrE interferes with the dynamics of the conformational exchange (Wu, 2019). The well-ordered extramembrane loops also exhibit major differences in packing on the open and closed sides of the transporter (**Fig. 2.3e**). On the open side of the transporter, several charged amino acids, K21_A from loop 1_A and E80_A and R86_A from loop 3_A, are solvent-exposed in the aqueous chamber. Upon conformational exchange, K21_A, E80_A, and R86_A, converge on loop 2_B and the N-terminal end of helix 3_B, forming cross-subunit H-bond interactions with the backbone and sidechain atoms of L53_B-T57_B. The hydrophobic loop 1_A also contributes to sealing the binding pocket on the closed side of the transporter, where it is wedged between the antiparallel helices 2_B and 2_A. Thus, the extramembrane loops, which are the least well-resolved features of previous structural models of SMR proteins, likely play an important role in the energetics of subunit packing. The involvement of loop 3 in conformational exchange has also been proposed for EmrE based on spectroscopic studies (Dastvan et al., 2016; Leninger et al., 2019).

2.2.4 The substrate binding site.

In order to visualize substrate coordination, we solved a structure of the Gdx-Clo/L10 monobody complex together with a non-natural transported substrate, phenylGdm⁺, since this

compound's bulky phenyl group would aid modeling of the substrate. Fortuitously, this also improved the resolution to 2.5 Å. We observed conspicuous density near the glutamates, to which we fit one phenylGdm⁺ molecule (**Fig. 2.4a**). Neutralization of these glutamates has previously been shown to abolish substrate transport in Gdx-Clo (Kermani et al., 2018). The substrate's guanidinyll group is coordinated by E13_B, whose position is in turn stabilized by a stack of conserved H-bond donors and acceptors, including W62_B, S42_B, and W16_B. W62 and S42 are highly conserved among SMRs, and have been previously implicated in substrate specificity and transport (Brill et al., 2015; Elbaz et al., 2008). In Gdx-Clo, mutations that remove H-bond potential, S42A and W62F, reduced or eliminated Gdm⁺ exchange, respectively (**Fig. 2.4b**). Conspicuously, W16 is conserved among Gdx proteins, but conserved as a glycine or alanine among the Qac subtype. In Gdx-Clo, the W16G mutant reduces, but does not eliminate Gdm⁺ exchange (**Fig. 2.4b**). The guanidinyll group of phenylGdm⁺ is also in close proximity to the opposite E13_A sidechain. However, E13_A is deflected downward by a cross-subunit interaction with Y59_B, so that the angle between the nitrogen, hydrogen (coplanar with the guanidinyll group), and oxygen atoms is not optimal for Hbond formation. Y59 is absolutely conserved among SMR proteins and the capacity to hydrogen bond has been identified as mechanistically essential at this position (Rotem et al., 2006; Vermaas et al., 2018). Based on our Gdx-Clo structure, we propose that Y59_B and the guanidinium group compete for E13_A, and that displacement of Y59_B by the guanidinyll group initiates the transport motion (**Fig. 2.4c**). Of all the amino acids, Y59 undergoes one of the largest changes in conformation, swinging out away from the binding site and into the aqueous pocket when the subunits swap conformations. Y59F, which cannot form a hydrogen bond with the E13 carboxylate, is not competent for substrate exchange (**Fig. 2.4b**), in accord with the requirement for an H-bond at this position. It is also

notable that E13 only forms a single hydrogen bond with the Gdm^+ ion. This contrasts with the lowest energy coplanar, bidentate coordination of the guanidinium/glutamate complex in solution (Peng et al., 2010), and also draws a contrast to Gdm^+ coordination by the guanine riboswitches, which provide hydrogen bond partners for most or all of the substrate's five hydrogen bond donors (Battaglia et al., 2017; Huang et al., 2017; Reiss & Strobel, 2017; Reiss et al., 2017). The more minimal coordination by the transporter explains its permissiveness towards guanidinium ions with methyl substitutions in one or two positions.

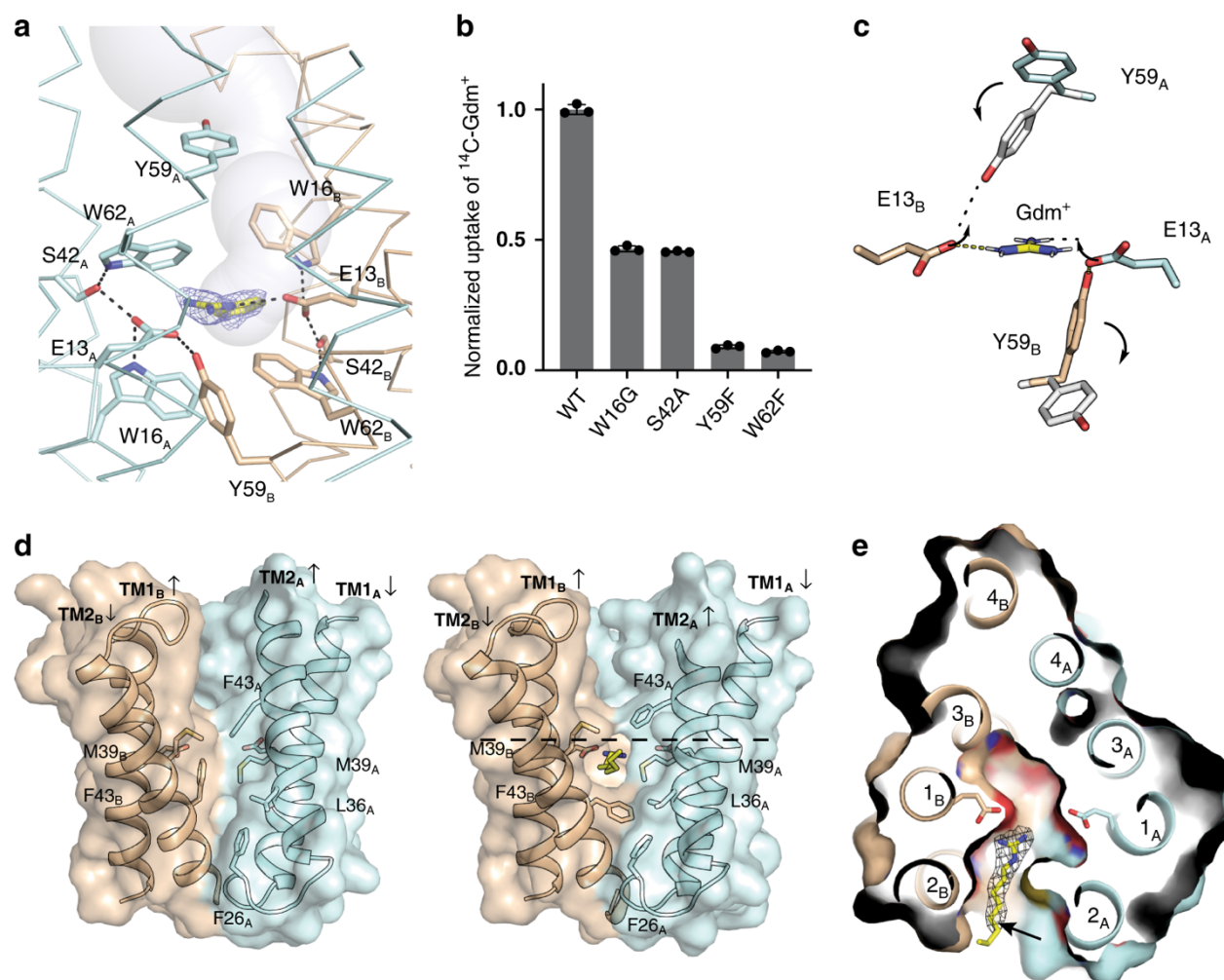


Figure 2.4: Substrate binding by Gdx-Clo.

a PhenylGdm⁺ binding site. Subunits colored in light blue and tan as in Fig. 2.3. The aqueous accessible vestibule is shown as a gray surface rendering. Sidechains that coordinate substrate or E13 shown as sticks, and interactions with appropriate distance and geometry for hydrogen bonds are shown as dashed lines. The electron density assigned to phenylGdm⁺ (2Fo-Fc map contoured at 1.3σ) is shown as blue mesh. **b** Uptake of ¹⁴C-Gdm⁺ into proteoliposomes mediated by the indicated mutant. Total uptake is normalized relative to uptake by WT at 10 min. Error bars represent the SEM for three independent replicates. **c** Illustration of the proposed conformational transition around the transported Gdm⁺. Colored sidechain sticks are in the positions observed in the structure; white sidechain sticks and arrows show proposed conformational change. **d** Membrane portal. The structure from Fig. 2.3 is shown at left, and the octylGdm⁺ bound structure is shown at right. Cartoon is shown with helices 3 and 4 removed for clarity. Sidechains lining the portal, and E13 sidechains, are shown as sticks. OctylGdm⁺ is shown as stick representation, with octyl tail extending toward the viewer. Dashed line indicates the level at which the protein is sliced in panel (e). **e** Top-down view of Gdx-Clo surface and helices with octylGdm⁺, sliced at approximately the midpoint of the membrane. E13 sidechains are shown as sticks. Experimental 2Fo-Fc density for the ligand, contoured at 1.8σ, is shown as gray mesh. The arrow indicates C₅ of the octyl substituent. Agmatine and arginine bear amino groups at this position.

2.2.5 A membrane portal accommodates hydrophobic substrate substituents.

In the case of phenylGdm⁺, the substituent is packed between TM2_A and TM2_B. At this point, the antiparallel TM2 helices splay apart, delimiting a portal from the membrane to the substrate binding site (**Fig. 2.4d**). To interrogate this feature, we solved a structure of Gdx-Clo in complex with octylGdm⁺, a cationic detergent with a Gdm⁺ head group and an eight-carbon tail. The guanidinyll group sits in the same binding pocket as phenylGdm⁺, near E13_B, and the aliphatic tail protrudes from the protein and into the detergent micelle (**Fig. 2.4d, e**). The tail is accommodated by rotameric rearrangements of the hydrophobic amino acids lining TM2 including M39 and F43 (**Fig. 2.4e**). Similar portals have been observed in other drug binding membrane proteins and are thought to provide binding site access for hydrophobic substrates that partition into the membrane (Aller, 2009; He, 2010; Payandeh et al., 2011). Spectroscopic studies and molecular modeling have provided evidence for a similar portal between the TM2 helices of EmrE (Dastvan et al., 2016; Vermaas et al., 2018). This membrane portal could be exploited by hydrophobic compounds to gain access to the binding site. We propose that this portal is also advantageous in the transporter's physiological context. Although this portal allows hydrophobic substituents accommodation by the membrane, metabolites like arginine, creatine, and agmatine all have polar groups on the tails for which insertion into the hydrophobic membrane environment would introduce a high energetic penalty, making the portal a convenient means for selecting against major guanidinyllated metabolites, and rationalizing the conservation of this feature. The SLC35 solute transporters (Nji et al., 2019; Parker & Newstead, 2017; Tsuchiya, 2016) provide a notable point of contrast. SLC35 proteins are assembled as two-domain inverted repeat transporters in which each domain is homologous to the SMR fold but

have an additional two-helix insertion that seals off the portal so that the binding site is only accessible from aqueous solution (**Supplementary Fig. A.12**).

2.3 Results

In summary, our transport experiments show that a representative of the SMR family's Gdx subtype, like the better characterized Qac SMRs, promiscuously transports a series of hydrophobic non-natural compounds, and that functional promiscuity is thus a general feature of the SMR family. Although Gdx-Clo's physiological role is transport of the metabolite Gdm⁺, it is not exquisitely selective for Gdm⁺, and whereas there is a biological imperative to prevent export of valuable guanidinylated metabolites like arginine or agmatine, there is no selective pressure to be discerning towards non-native compounds. Promiscuous functions, those that are not under direct selection, provide a rich source of cryptic variation that can be harnessed to provide evolutionary novelty (Copley, 2015)—perhaps rationalizing the broad distribution of both the Qac and Gdx subtypes with horizontal gene transfer elements. Changing environmental pressures, which could include various human-introduced biocides, may have made these latent functions adaptive. Indeed, environmental contamination by hydrophobic quaternary amines is associated with antiseptic use (Russell, 2002) and substituted guanidinium ions and biguanides have also been identified as widespread, long-lived, environmentally disruptive contaminants that enter the biosphere as agricultural or industrial chemicals (Peter, 2018; Roberts & Hutson, 1999; Zahn, 2019) or pharmaceuticals that impact the human microbiome and that are excreted in waste water (Blair et al., 2013; Maier, 2018; Pryor, 2019; Scheurer et al., 2012).

Structural analysis of Gdx-Clo reveals numerous features that correspond to biochemical or spectroscopic observations made for EmrE, indicating a high degree of mechanistic conservation between the Qac and Gdx subtypes. The Qac and Gdx subtypes also share

multiaromatic binding pockets, which have also been implicated in polyspecificity in several other systems, including QacR transcriptional regulators and P-glycoprotein (Aller, 2009; Murray et al., 2004; Peters, 2011). The structure also identifies other features that contribute to promiscuous substrate transport in the SMR family, including minimal coordination of the substrate and direct access from the membrane to the binding site. We conjecture that SMR proteins have enjoyed such evolutionary success in the modern world because this portal, a conserved selectivity mechanism against major physiological metabolites, proved to be extremely adaptive for the binding and export of hydrophobic, human-introduced chemicals.

2.4 Methods

2.4.1 Sequence-similarity network.

A sequence-similarity network was generated using the EFI-EST webserver(Zallot et al., 2019) starting from PFAM family PF00893 (Multi_Drug_Res), with an alignment score of 20, and visualized with 50% similarity in Cytoscape using the prefuse force-directed layout(Shannon, 2003). A genome neighborhood network was then generated with the EFI-GNT tool, using a neighborhood size of 10. The coordinates of the Guanidine-I, Guanidine-II, and Guanidine-III riboswitches were retrieved from RFAM and used to annotate the SMR PFAM members if they occurred within 100 bp of an RFAM member. A set of plasmid-encoded SMRs was generated from Uniprot using the keyword plasmid. The GNN was used to annotate integron-integrase neighbors using the PFAM domains Phage_int_SAM_4 (PF13495) and Phage_integrase (PF00589).

2.4.2 Transport expression, purification, and proteoliposome reconstitution.

Lipids were from Avanti, detergents from Anatrace. Proteins were expressed and purified as previously described⁶. Briefly, Gdx-Clo bore a C-terminal hexahistidine affinity tag and a LysC recognition site and were cloned into a pET-21c expression vector, and transformed into C41 (DE3) *E. coli*. When cultures reached an OD₆₀₀ of 1.0, protein expression was induced with 0.2 mM Isopropyl β -D-1- thiogalactopyranoside (IPTG) for 3 h at 37 °C. Cell lysate was extracted with 2% (w/v) decyl- β -D-maltoside (DM), and the soluble fraction was purified over cobalt affinity column, washed with 100 mM NaCl, 20 mM imidazole, and then eluted with 400 mM imidazole. The affinity tag was cleaved by incubation with LysC (200 ng per mg protein, 2 h at room temperature), before a final size exclusion purification step using a Superdex 200 gel-filtration column equilibrated in 100 mM NaCl, 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic (HEPES)-NaOH, 5 mM DM, pH 8.1. EmrE was expressed and purified similarly, but the construct bore an N-terminal hexahistidine sequence with a thrombin recognition site. After induction with IPTG, protein was expressed overnight at 16 °C.

E. coli polar lipids dissolved in chloroform were dried under a nitrogen stream and residual chloroform was removed by washing and drying three times with pentane. Lipids were solubilized with reconstitution buffer (100 mM KCl, 100 mM KPO₄, pH 7.5) containing 35 mM 3-((3-cholamidopropyl) dimethylammonio)-1- propanesulfonate (CHAPS). For SSM electrophysiology experiments, proteoliposomes were prepared with 20 mg *E. coli* polar lipid per ml, and a 1:25 protein:lipid mass ratio. For radioactive flux assays and H⁺ transport assays, proteoliposomes were prepared with 10 mg *E. coli* polar lipid per ml, and a 1:5000 protein:lipid mass ratio. The protein/detergent/lipid solution was dialyzed against a 1000-fold excess of

reconstitution buffer, with three buffer changes over 2 days. After the final round of dialysis, proteoliposomes were aliquoted and stored at -80°C until use.

2.4.3 *Radioactive flux assays.*

After reconstitution, proteoliposomes were loaded with test substrate and subjected to three freeze/thaw cycles before extrusion 21 times through a 400 nm membrane. To remove unencapsulated substrate, external solution was exchanged by passing liposomes over a Sephadex G-50 column preequilibrated with reaction buffer (25 mM HEPES, 400 mM sorbitol, pH 7.5). Recovered proteoliposomes were diluted twofold into reaction buffer, and the substrate transport reaction was initiated by addition of ^{14}C -labeled compound (20 μM ^{14}C -Gdm⁺ for Gdx or 7 μM ^{14}C -methyl viologen for EmrE; American Radiolabelled Chemicals, Inc., St. Louis, MO). At time points, 100 μl of reaction mixture was passed over a 1.6 ml Dowex ion exchange resin column (N-methyl-D-glucamine form), then suspended in scintillation fluid (Ultima Gold; Perkin-Elmer) for liquid scintillation counting.

2.4.4 *SSM electrophysiology.*

SSM electrophysiology was conducted using a SURFE2R N1 instrument (Nanon Technologies, Munich, Germany) according to published protocols (Bazzone et al., 2017). SSM sensors were first alkylated by adding 50 μl thiol solution (0.5 mM 1-octadecanethiol in isopropanol) to a clean sensor's well, then incubating for 1 h at room temperature in a closed dish. Afterwards, the sensor was rinsed three times with ethanol and three times with water and dried by tapping on a paper towel. 1.5 μl of lipid solution (7.5 $\mu\text{g}/\mu\text{l}$ 1,2-diphytanoyl-sn-glycero-3-phosphocholine in n-decane) was painted on the gold electrode surface using a pipette tip, followed immediately by addition of 50 μl of nonactivating buffer (100 mM KCl, 100 mM

KPO₄, pH 7.5). Proteoliposomes were diluted 25-fold in buffer and sonicated 30–60 s before addition to the sensor surface and centrifugation at $2500 \times g$ for 30 min. Before experiments, sensors were checked for conductance and capacitance using SURFE2R software protocols. Sensors for which capacitance and conductance measurements were outside an acceptable range (10–40 nF capacitance, 1–5 nS conductance) were not used for experiments. Sensors were periodically rechecked for quality during the course of an experiment. Each substrate was tested for transport at a concentration of 1 mM in buffer containing 100 mM KCl, 100 mM KPO₄, pH 7.5. For measurements in the presence of monobody, recording buffers contained 50 µg bovine serum albumin/ml. To compare measurements recorded on different sensors, currents were normalized relative to a reference compound, as described in the text. Currents elicited by the reference compound were measured both at the outset of the experiment and after collecting data on test compounds. If currents for the first and last perfusions of reference compound differed by more than 10%, this indicated that the amount of reconstituted protein had not remained stable over the course of the experiment, and data collected in this series were not used for further analysis. Data were collected from 3 to 4 independent sensor preparations, which were in turn prepared from 2 to 4 independent protein preparations. Reported data are for peak currents, which represent the initial rate of substrate transport before a membrane potential builds up and inhibits further electrogenic transport (Bazzone et al., 2017) .

2.4.5 Monobody development.

Monobody selection was performed following previously published methods (Koide et al., 2012; Sha et al., 2017; Stockbridge et al., 2014). Four rounds of phage selection with target concentrations of 100, 100, 50 and 20 nM was performed in 10 mM Hepes pH 7.5, 200 mM NaCl, 20 mM GdmCl, 4 mM DM, then sorted pools were subcloned into a yeast-display library

following recombination of 5' and 3' fragments to increase library diversity (Koide et al., 2012). Three rounds of yeast library sorting were performed: the first round for clones binding to 50 nM target, second round for clones exhibiting no binding to 10 μ M streptavidin (negative sorting), and the third round for binding with 5 nM target. Isolated clones were validated for target binding using a yeast display binding assay, as described in detail (Koide et al., 2012; Sha, 2013)

2.4.6 Monobody expression and purification.

Monobody proteins were expressed in *E. coli* (BL21-DE3) grown in Studier's autoinduction media (Studier, 2005) 15–18 h at 37 °C. After harvesting by centrifugation, cell pellets were frozen at –80 °C for 30–45 min prior to being resuspended in breaking buffer (20 mM Tris-Cl pH 8.0, 500 mM NaCl) supplemented with 400 μ g DNase, 2 mM MgCl₂, 1 mM PMSF, 1 mg/ml lysozyme, 25 μ g pepstatin, and 500 μ g leupeptin and lysed by sonication prior to centrifugation (27,000 \times g for 15 min). Inclusion bodies were isolated by addition of Triton X-100 to a final concentration of 1% w/v (Burgess, 2009), incubation of the lysate on ice, and centrifugation (27,000 \times g for 15 min). The pellet containing the L10 inclusion bodies was resuspended in denaturing buffer (20 mM Tris-Cl pH 8.0, 6 M GdmCl) and incubated at room temperature with rotation for 1 h. Debris were removed by centrifugation (17,500 \times g/45 min), and the supernatant was loaded onto a cobalt affinity column (Takara; 3 ml resin/l culture) for on-column refolding (Oganesyan et al., 2005). The column with bound monobody was washed with 10 CV of denaturing buffer, 10 CV of denaturing buffer supplemented with 10 mM imidazole, 10 CV of wash buffer (0.1% (w/v) Triton X-100, 20 mM Tris-Cl pH 8.0, 500 mM NaCl), 10 CV of refolding buffer (5 mM β -cyclodextrin, 20 mM tris-Cl pH 8.0, 500 mM NaCl), and finally, 10 CV of breaking buffer. The resin, with bound, refolded monobody, was incubated with TEV protease (0.03 mg/ml cobalt affinity resin) overnight to cleave the His₆ tag, and

digested monobody was eluted with breaking buffer. A final size exclusion purification step was performed using a Superdex 75 gel-filtration column equilibrated in 10 mM HEPES pH 7.5, 10 mM NaCl.

2.4.7 Crystal preparation.

For X-ray crystallography, Gdx-Clo and monobody Clo-L10 were purified as described above. For the Clo purification, the size exclusion buffer contained 200 mM NaCl, 10 mM HEPES pH 8.1, and 10 mM Gdm⁺ or 20 mM phenylGdm⁺. Proteins were concentrated to 10 mg/ml, Clo-L10 was supplemented with 4 mM DM, and monobody and Gdx-Clo dimer were mixed in a 2.1:1 ratio. The protein solution was then mixed with an equal volume of crystallization solution (0.3 μ l in 96-well plates). Initial hits grew in 200 mM CaCl₂, 0.1 M Tris/HCl pH 8.0 and 32.5% PEG 600. Crystals were subsequently improved by addition of charged detergents lauryldimethylamine-N-Oxide (LDAO; final concentration 6.6 mM), dimethyldodecylphosphine oxide (Apo12; final concentration 2 mM), or octylGdm⁺ (final concentration 3.3 mM) to the protein solution prior to admixture with the crystallization solution (0.45 μ l protein/detergent mixture together with 0.3 μ l crystallization solution). Optimized crystals typically grew to their maximum size in 14 days in a wide range of salt and pH conditions with 32–36% PEG 600. For selenomethionine-incorporated protein, the best diffracting crystals were obtained with Apo12 supplementation, and crystallization solution 0.1 M LiNO₃, 0.1 M N-(2-Acetamido)iminodiacetic acid (ADA) pH 6.8, and 35% PEG 600. For phenylGdm⁺ bound protein, the best diffracting crystals were obtained with Apo12 supplementation, and crystallization solution 0.1 M LiNaSO₄, 100 mM Tris pH 8.8 and 34% PEG 600. For the octylGdm⁺ bound structure, octylGdm⁺ was used as the detergent additive, and crystallization solution contained 0.1 M calcium acetate, 0.1 M HEPES pH 7.5 and 33%

PEG 600. Crystals were frozen in liquid nitrogen before data collection at the Life Sciences Collaborative Access Team beamline 21-ID-D at the Advanced Photon Source, Argonne National Laboratory.

2.4.8 Structure determination.

Diffraction data were collected at an X-ray wavelength of 0.978 Å for selenomethionine-labeled crystals. Diffraction data were processed and scaled using DIALS (Winter, 2018). The space group for the initial crystals was determined to be C121 with one Clo dimer and two monobodies per asymmetric unit. Eight selenium sites were located using SAD implemented in SHELX (Sheldrick, 2008). The positions were refined, and initial phases were calculated using SHARP (Bricogne et al., 2003) with solvent flattening with SOLOMON (Abrahams & Leslie, 1996). A model for the transporter was built into experimental electron density maps using Coot (Emsley et al., 2010). The L10 monobodies were modeled based on a previously determined structure of a loop-library monobody (PDB code: 5NKQ [<https://www.rcsb.org/structure/5NKQ>]) (Stockbridge et al., 2015). Variable loops were not included in the monobody model. These models were placed into the experimental electron density maps using Phaser-MR (McCoy, 2007). Partial models were cycled back into SHARP for phase calculation to improve the initial solvent envelope. Density from both the sidechains and the monobody loops was clearly visible in the electron density maps, and loops and the transporter's amino acid sidechains were built using the Se sites to ensure the correct register, with iterative rounds of refinement in Refmac (Murshudov, 2011) with prior phase information incorporated as Hendrickson–Lattman coefficients. Model validation was carried out using the Molprobity server (Williams, 2018). Diffraction resolution was improved in subsequent datasets upon the addition of phenyl- or octylGdm⁺. With phenylguanidinium as the substrate, proteins

crystallized in C121 as before, and with octylGdm⁺ as the substrate, proteins crystallized in P1. The arrangement of proteins in the crystal lattice was highly similar to the C121 crystal form, but with two transporters and four monobodies per asymmetric unit. Crystals diffracted anisotropically, and electron density maps were improved by anisotropic truncation of the unmerged data using the Staraniso webserver (Tickle, 2018) with a cutoff level of 1.8 for the local $I/\sigma(I)$. Models were built into experimental density maps calculated from Phaser, with the initial models of GdxClo and L10 monobody determined previously, with iterative rounds of refinement in Phenix and Refmac. The structural model was revised in real space in Coot. Solvent-accessible vestibules were visualized with CAVER (Jurcik, 2018).

2.5 Acknowledgments

We are grateful to Andre Bazzone for technical advice about SSM electrophysiology and to the beamline staff at LS-CAT for assistance with crystallography data collection. We thank Ming Li, Chris Miller, and Shimon Schuldiner for helpful comments on the manuscript. This work was supported by a SURFE²R N1 research grant (Nanion Technologies) and National Institutes of Health grants R35 GM128768 to R.B.S. and R01 CA194864 to S.K. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). R.B.S. is a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Disease.

2.6 Competing Interests

A.K. and S.K. are listed as inventors for patents (US9512199 B2 and related patents and applications) covering aspects of the monobody technology filed by the University of Chicago and Novartis. A.A.K., C.B.M., O.E.B., B.B.K., E.D., and R.B.S. declare no competing interests.

2.7 References

- Abrahams, J. P., & Leslie, A. G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D Biol. Crystallogr.*, 52. <https://doi.org/10.1107/S0907444995008754>
- Aller, S. G. (2009). Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*, 323. <https://doi.org/10.1126/science.1168750>
- Amadi, S. T., Koteiche, H. A., Mishra, S., & McHaourab, H. S. (2010). Structure, dynamics, and substrate-induced conformational changes of the multidrug transporter EmrE in liposomes. *J. Biol. Chem.*, 285. <https://doi.org/10.1074/jbc.M110.132621>
- Battaglia, R. A., & Ke, A. (2018). Guanidine-sensing riboswitches: how do they work and what do they regulate? . *Wiley Interdiscip. Rev. RNA*. <https://doi.org/https://doi.org/10.1002/wrna.1482>
- Battaglia, R. A., Price, I. R., & Ke, A. (2017). Structural basis for guanidine sensing by the ykkC family of riboswitches. *RNA*, 23. <https://doi.org/10.1261/rna.060186.116>
- Bazzone, A., Barthmes, M., & Fendler, K. (2017). SSM-based electrophysiology for transporter research. *Methods Enzymol.*, 594. <https://doi.org/10.1016/bs.mie.2017.05.008>
- Blair, B. D., Crago, J. P., Hedman, C. J., & Klaper, R. D. (2013). Pharmaceuticals and personal care products found in the Great Lakes above concentrations of environmental concern. *Chemosphere*, 93. <https://doi.org/10.1016/j.chemosphere.2013.07.057>
- Breaker, R. R., Atilho, R. M., Malkowski, S. N., Nelson, J. W., & Sherlock, M. E. (2017). The biology of free guanidine as revealed by riboswitches. *Biochemistry*, 56. <https://doi.org/10.1021/acs.biochem.6b01269>
- Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M., & Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.*, 59. <https://doi.org/10.1107/S0907444903017694>
- Brill, S., Sade-Falk, O., Elbaz-Alon, Y., & Schuldiner, S. (2015). Specificity determinants in small multidrug transporters. *J. Mol. Biol.*, 427. <https://doi.org/10.1016/j.jmb.2014.11.015>
- Burgess, R. R. (2009). Refolding solubilized inclusion body proteins. *Methods Enzymol.*, 463. [https://doi.org/10.1016/S0076-6879\(09\)63017-2](https://doi.org/10.1016/S0076-6879(09)63017-2)
- Chen, Y. J. (2007). X-ray structure of EmrE supports dual topology model. *Proc. Natl Acad. Sci. USA*, 104. <https://doi.org/10.1073/pnas.0709387104>

- Copley, S. D. (2015). An evolutionary biochemist's perspective on promiscuity. *Trends Biochem. Sci.*, 40. <https://doi.org/10.1016/j.tibs.2014.12.004>
- Dastvan, R., Fischer, A. W., Mishra, S., Meiler, J., & McHaourab, H. S. (2016). Protonation-dependent conformational dynamics of the multidrug transporter EmrE. *Proc. Natl Acad. Sci. USA*, 113. <https://doi.org/10.1073/pnas.1520431113>
- Elbaz, Y., Salomon, T., & Schuldiner, S. (2008). Identification of a glycine motif required for packing in EmrE, a multidrug transporter from Escherichia coli. *J Biol Chem*, 283(18), 12276-12283. <https://doi.org/10.1074/jbc.M710338200>
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, 66. <https://doi.org/10.1107/S0907444910007493>
- Gillings, M. R. (2017). Class 1 integrons as invasive species. *Curr. Opin. Microbiol.*, 38. <https://doi.org/10.1016/j.mib.2017.03.002>
- He, X. (2010). Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature*, 467. <https://doi.org/10.1038/nature09408>
- Huang, L., Wang, J., & Lilley, D. M. J. (2017). The structure of the guanidine-II riboswitch. *Cell Chem. Biol.*, 24. <https://doi.org/10.1016/j.chembiol.2017.05.014>
- Jeong, H., & Nasir, A. (2017). A preliminary list of horizontally transferred genes in prokaryotes determined by tree reconstruction and reconciliation. *Front. Genet.*, 8. <https://doi.org/10.3389/fgene.2017.00112>
- Jurcik, A. (2018). CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics*, 34. <https://doi.org/10.1093/bioinformatics/bty386>
- Kermani, A. A., Macdonald, C. B., Burata, O. E., Ben Koff, B., Koide, A., Denbaum, E., Koide, S., & Stockbridge, R. B. (2020). The structural basis of promiscuity in small multidrug resistance transporters. *Nat Commun*, 11(1), 6064. <https://doi.org/10.1038/s41467-020-19820-8>
- Kermani, A. A., Macdonald, C. B., Gundepudi, R., & Stockbridge, R. B. (2018). Guanidinium export is the primal function of SMR family transporters. *Proc. Natl Acad. Sci. USA*, 115. <https://doi.org/10.1073/pnas.1719187115>
- Koide, A., Wojcik, J., Gilbreth, R. N., Hoey, R. J., & Koide, S. (2012). Teaching an old scaffold new tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. *J. Mol. Biol.*, 415. <https://doi.org/10.1016/j.jmb.2011.12.019>
- Leninger, M., Sae Her, A., & Traaseth, N. J. (2019). Inducing conformational preference of the membrane protein transporter EmrE through conservative mutations. *Elife*. <https://doi.org/https://doi.org/10.7554/eLife.48909>
- Maier, L. (2018). Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*, 555. <https://doi.org/10.1038/nature25979>
- McCoy, A. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.*, 40. <https://doi.org/10.1107/S0021889807021206>
- Murray, D. S., Schumacher, M. A., & Brennan, R. G. (2004). Crystal structures of QacR-diamidine complexes reveal additional multidrug-binding modes and a novel mechanism of drug charge neutralization. *J. Biol. Chem.*, 279. <https://doi.org/10.1074/jbc.M313870200>
- Murshudov, G. N. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, 67. <https://doi.org/10.1107/S0907444911001314>

- Nelson, J. W., Atilho, R. M., Sherlock, M. E., Stockbridge, R. B., & Breaker, R. R. (2017). Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Mol. Cell*, 65. <https://doi.org/10.1016/j.molcel.2016.11.019>
- Nji, E., Gulati, A., Qureshi, A. A., Coincon, M., & Drew, D. (2019). Structural basis for the delivery of activated sialic acid into Golgi for sialylation. *Nat. Struct. Mol. Biol.*, 26. <https://doi.org/10.1038/s41594-019-0225-y>
- Oganesyan, N., Kim, S. H., & Kim, R. (2005). On-column protein refolding for crystallization. *J. Struct. Funct. Genomics*, 6. <https://doi.org/10.1007/s10969-005-2827-3>
- Ovchinnikov, V., Stone, T. A., Deber, C. M., & Karplus, M. (2018). Structure of the EmrE multidrug transporter and its use for inhibitor peptide design. *Proc Natl Acad Sci U S A*, 115(34), E7932-e7941. <https://doi.org/10.1073/pnas.1802177115>
- Parker, J. L., & Newstead, S. (2017). Structural basis of nucleotide sugar transport across the Golgi membrane. *Nature*, 551. <https://doi.org/10.1038/nature24464>
- Payandeh, J., Scheuer, T., Zheng, N., & Catterall, W. A. (2011). The crystal structure of a voltage-gated sodium channel. *Nature*, 475. <https://doi.org/10.1038/nature10238>
- Peng, B., Peng, Q., Zhou, W., & Zhou, Z. (2010). Guanidinium l-glutamate. *Acta Crystallogr. Sect. E Struct. Rep. Online*, 66. <https://doi.org/10.1107/S1600536810036354>
- Peter, K. T. (2018). Using high-resolution mass spectrometry to identify organic contaminants linked to urban stormwater mortality syndrome in Coho Salmon. *Environ. Sci. Technol.*, 52. <https://doi.org/10.1021/acs.est.8b03287>
- Peters, K. M. (2011). A single acidic residue can guide binding site selection but does not govern QacR cationic-drug affinity. *PLoS ONE*, 6. <https://doi.org/10.1371/journal.pone.0015974>
- Pryor, R. (2019). Host-microbe-drug-nutrient screen identifies bacterial effectors of metformin therapy. *Cell*, 178. <https://doi.org/10.1016/j.cell.2019.08.003>
- Reiss, C. W., & Strobel, S. A. (2017). Structural basis for ligand binding to the guanidine-II riboswitch. *RNA*. <https://doi.org/10.1261/rna.061804.117>
- Reiss, C. W., Xiong, Y., & Strobel, S. A. (2017). Structural basis for ligand binding to the guanidine-I riboswitch. *Structure*, 25. <https://doi.org/10.1016/j.str.2016.11.020>
- Roberts, T. R., & Hutson, D. H. (1999). Metabolic Pathways of Agrochemicals Part Two: Insecticides and Fungicides. *Royal Society of Chemistry*.
- Rotem, D., & Schuldiner, S. (2004). EmrE, a multidrug transporter from Escherichia coli, transports monovalent and divalent substrates with the same stoichiometry. *J Biol Chem*, 279(47), 48787-48793. <https://doi.org/10.1074/jbc.M408187200>
- Rotem, D., Steiner-Mordoch, S., & Schuldiner, S. (2006). Identification of tyrosine residues critical for the function of an ion-coupled multidrug transporter. *J. Biol. Chem.*, 281. <https://doi.org/10.1074/jbc.M602088200>
- Russell, A. D. (2002). Introduction of biocides into clinical practice and the impact on antibiotic-resistant bacteria. *J. Appl. Microbiol*, 92 Suppl. <https://doi.org/10.1046/j.1365-2672.92.5s1.12.x>
- Scheurer, M., Michel, A., Brauch, H. J., Ruck, W., & Sacher, F. (2012). Occurrence and fate of the antidiabetic drug metformin and its metabolite guanyurea in the environment and during drinking water treatment. *Water Res.*, 46. <https://doi.org/10.1016/j.watres.2012.06.019>
- Sha, F. (2013). Dissection of the BCR-ABL signaling network using highly specific monoclonal inhibitors to the SHP2 SH2 domains. *Proc. Natl Acad. Sci. USA*, 110. <https://doi.org/10.1073/pnas.1303640110>

- Sha, F., Salzman, G., Gupta, A., & Koide, S. (2017). Monobodies and other synthetic binding proteins for expanding protein science. *Protein Sci.*, 26. <https://doi.org/10.1002/pro.3148>
- Shannon, P. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13. <https://doi.org/10.1101/gr.1239303>
- Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallogr. A*, 64. <https://doi.org/10.1107/S0108767307043930>
- Slipski, C. J. (2020). Plasmid transmitted small multidrug resistant (SMR) efflux pumps differ in gene regulation and enhance tolerance to quaternary ammonium compounds (QAC) when grown as biofilms. . <https://doi.org/bioRxivhttps://doi.org/10.1101/768630>
- Stockbridge, R. B., Koide, A., Miller, C., & Koide, S. (2014). Proof of dual-topology architecture of Fluc F- channels with monobody blockers. *Nat. Commun.*, 5. <https://doi.org/10.1038/ncomms6120>
- Stockbridge, R. B., Kolmakova-Partensky, L., Shane, T., Koide, A., Koide, S., Miller, C., & Newstead, S. (2015). Crystal structures of a double-barrelled fluoride ion channel. *Nature*, 525(7570), 548-551. <https://doi.org/10.1038/nature14981>
- Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.*, 41. <https://doi.org/10.1016/j.pep.2005.01.016>
- Tickle, I. J. (2018). STARANISO <http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi> <http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>
- Tsuchiya, H. (2016). Structural basis for amino acid export by DMT superfamily transporter YddG. *Nature*, 534. <https://doi.org/10.1038/nature17991>
- Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S., & Tate, C. G. (2003). Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J.*, 22. <https://doi.org/10.1093/emboj/cdg611>
- Vermaas, J. V., Rempe, S. B., & Tajkhorshid, E. (2018). Electrostatic lock in the transport cycle of the multidrug resistance transporter EmrE. *Proc. Natl Acad. Sci. USA*, 115. <https://doi.org/10.1073/pnas.1722399115>
- Williams, C. J. (2018). MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.*, 27. <https://doi.org/10.1002/pro.3330>
- Winter, G. (2018). DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D Struct. Biol.*, 74. <https://doi.org/10.1107/S2059798317017235>
- Wu, C. (2019). Identification of an alternating-access dynamics mutant of EmrE with impaired transport. *J. Mol. Biol.*, 431. <https://doi.org/10.1016/j.jmb.2019.05.035>
- Zahn, D. (2019). Identification of potentially mobile and persistent transformation products of REACH-registered chemicals and their occurrence in surface waters. *Water Res.*, 150. <https://doi.org/10.1016/j.watres.2018.11.042>
- Zallot, R., Oberg, N., & Gerlt, J. A. (2019). The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry*, 58. <https://doi.org/10.1021/acs.biochem.9b00735>
- Zhu, Y. G. (2017). Continental-scale pollution of estuaries with antibiotic resistance genes. *Nat. Microbiol.*, 2. <https://doi.org/10.1038/nmicrobiol.2016.270>

Chapter 3 Crystal Structures of Bacterial Small Multidrug Resistance Transporter EmrE in Complex With Structurally Diverse Substrates

This chapter is adapted from the following published research article:

*Kermani, A.A.[#], Burata, O.E.[#], Koff, B.B., Koide, A., Koide, S. & Stockbridge, R.B. Crystal structures of bacterial small multidrug resistance transporter EmrE in complex with structurally diverse substrates. *eLife* 11:e76766, <https://doi.org/10.7554/eLife.76766>*

- equal contributions

Ali A Kermani, Conceptualization, Formal analysis, Investigation, Methodology, Writing – review and editing; Olive E Burata, Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – review and editing; B Ben Koff, Investigation; Akiko Koide, Investigation, Methodology; Shohei Koide, Funding acquisition, Methodology, Writing – review and editing; Randy B Stockbridge, Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Visualization, Writing - original draft

This published work brings to light the first ever high resolution (side-chain resolution) X-ray crystal structure of EmrE, the most well-studied quaternary ammonium compound (QAC) transporter of the small multidrug resistance (SMR) family. For almost three decades, this multidrug transporter has eluded complete molecular investigation through structural biology, and its characterization has been abundantly dependent on *in vivo* work such as cell-based

resistance assays or low-throughput *in vitro* radioactive-based exchange/uptake assays. This publication rectifies that gap in knowledge, allowing us to finally peek into the molecular interactions governing EmrE's promiscuous phenotype. My role in this work was to functionally investigate these interactions through the lens of not just the proton-bound structure, but also through several structures of EmrE in complex with different substrates, highlighting its dynamic ability to interact with diverse ligands. To obtain the structures, we took advantage of EmrE's high sequence conservation of its Loop 1 region with Gdx-Clo WT and engineered the same exact loop from Gdx-Clo to design EmrE₃. This allowed us to repurpose the same molecular chaperone, L10 monobody, used to obtain Gdx-Clo's structures, for successful crystal formation and structure determination of EmrE₃. This same method brought forth the same concerns regarding the potential effects of monobody binding in addition to the changes introduced upon EmrE's loop 1 region, but I was able to show through a transport assay using SSME, that both WT-EmrE and EmrE₃ showed comparable transport kinetics, owing to the unhindered transport capability by the mutations. I also applied the same modified SSME technique I developed and described in **Section 2.1**, showing that monobody did not interfere with EmrE₃ function. Additionally, the versatility of this technique, especially since molecular chaperones have been increasingly utilized to address the lack of membrane protein structures, will be of immense help to characterizing more of these 'difficult' macromolecules, specifically those involving protein-protein interactions.

I led the exploration of the binding pocket residues in EmrE₃, by designing the mutations disrupting any H-bonding potentials required to potentially stabilize substrate binding. This effort led us to the realization that EmrE₃ requires very little intricacies when it comes to binding site stability. In fact, EmrE₃ function is highly tolerant of H-bonding impairment. In combination

with the data we obtained through structural biology, this work had two major impacts: 1) three decades of cumulative characterization of EmrE in combination with our findings have advanced our understanding of the basis of substrate promiscuity in QACs for subsequent investigations towards functional inhibition and drug development. 2) provided us the proper tools and knowledge to investigate the molecular basis of promiscuity in the SMR family. The latter will be instrumental in the next steps of this thesis as we explore the molecular basis that distinguishes the selective SMR_{Gdxs} and the promiscuous SMR_{Qacs}.

3.1 Introduction

3.1.1 Personal contributions to the published research article

This published work brings to light the first ever high resolution (side-chain resolution) X-ray crystal structure of EmrE, the most well-studied quaternary ammonium compound (QAC) transporter of the small multidrug resistance (SMR) family. For almost three decades, this multidrug transporter has eluded complete molecular investigation through structural biology, and its characterization has been abundantly dependent on *in vivo* work such as cell-based resistance assays or low-throughput *in vitro* radioactive-based exchange/uptake assays. This publication rectifies that gap in knowledge, allowing us to finally peek into the molecular interactions governing EmrE's promiscuous phenotype. My role in this work was to functionally investigate these interactions through the lens of not just the proton-bound structure, but also through several structures of EmrE in complex with different substrates, highlighting its dynamic ability to interact with diverse ligands. To obtain the structures, we took advantage of EmrE's high sequence conservation of its Loop 1 region with Gdx-Clo WT and engineered the same exact loop from Gdx-Clo to design EmrE₃. This allowed us to repurpose the same molecular

chaperone, L10 monobody, used to obtain Gdx-Clo's structures, for successful crystal formation and structure determination of EmrE₃. This same method brought forth the same concerns regarding the potential effects of monobody binding in addition to the changes introduced upon EmrE's loop 1 region, but I was able to show through a transport assay using SSME, that both WT-EmrE and EmrE₃ showed comparable transport kinetics, owing to the unhindered transport capability by the mutations. I also applied the same modified SSME technique I developed and described in **Section 2.1**, showing that monobody did not interfere with EmrE₃ function. Additionally, the versatility of this technique, especially since molecular chaperones have been increasingly utilized to address the lack of membrane protein structures, will be of immense help to characterizing more of these 'difficult' macromolecules, specifically those involving protein-protein interactions.

I led the exploration of the binding pocket residues in EmrE₃, by designing the mutations disrupting any H-bonding potentials required to potentially stabilize substrate binding. This effort led us to the realization that EmrE₃ requires very little intricacies when it comes to binding site stability. In fact, EmrE₃ function is highly tolerant of H-bonding impairment. In combination with the data we obtained through structural biology, this work had two major impacts: 1) three decades of cumulative characterization of EmrE in combination with our findings have advanced our understanding of the basis of substrate promiscuity in QACs for subsequent investigations towards functional inhibition and drug development. 2) provided us the proper tools and knowledge to investigate the molecular basis of promiscuity in the SMR family. The latter will be instrumental in the next steps of this thesis as we explore the molecular basis that distinguishes the selective SMR_{Gdxs} and the promiscuous SMR_{Qacs}.

3.1.2 General overview

The small multidrug resistance (SMR) family of microbial membrane proteins is a well-studied family composed of primitive dual-topology proton-coupled transporters. The SMR family has two major physiological subtypes that can be distinguished based on sequence (Kermani et al., 2020). Representatives of the ‘Gdx’ (guanidinium export) subtype export a bacterial metabolite, guanidinium ion (Gdm^+), in exchange for two protons (Kermani et al., 2018; Nelson et al., 2017). Representatives of the ‘Qac’ (quaternary ammonium compound) subtype are proton-coupled exchangers of quaternary ammoniums and other hydrophobic, cationic compounds. Since the first quaternary ammonium antiseptics were introduced approximately one hundred years ago, proteins from the Qac cluster have been closely associated with the spread of multidrug resistance elements (Gillings, 2017; Pal et al., 2015; Russell, 2002; Zhu, 2017).

Many bacteria possess SMR proteins belonging to both subtypes. Transporters from the Qac and Gdx clusters do not overlap in terms of physiological role: the Qac proteins do not transport Gdm^+ and require additional hydrophobicity in transported substrates, whereas the Gdx transporters require substrates to have a guanidinyl moiety and cannot export quaternary ammoniums or other cations (Kermani et al., 2020). However, the two subtypes transport an overlapping subset of hydrophobic substituted guanidinium ions and share high sequence conservation (~35% sequence identity), strongly suggesting conservation of the overall fold.

The best-studied of the Qac proteins is the *E. coli* member, EmrE. The substrate repertoire of EmrE includes planar, conjugated aromatic ring systems, quaternary ammoniums and phosphoniums (with or without aromatic substituents), and substituted guanidiniums. EmrE also provides resistance to biocides from these substrate classes with long alkyl tails, such as

benzalkonium and cetyltrimethylammonium, which are found in common household antiseptics. Mechanisms to explain the transport promiscuity have been proposed, typically focusing on protein dynamics as a feature that allows it to transport many different substrates (Jurasz et al., 2021; Robinson et al., 2017). However, the structural basis for substrate binding is unknown, and for many years, structural information was limited to low-resolution models without loops or sidechains (Fleishman et al., 2006; Ubarretxena-Belandia et al., 2003), impeding a full description of the molecular mechanism. A previous crystal structure of EmrE was unreliable for molecular analysis, with no sidechains modeled, poor helical geometry, and helices too short to span the membrane (Chen, 2007). Computational models constrained by the low-resolution data have also been proposed (Ovchinnikov et al., 2018; Vermaas et al., 2018). Recently, high-resolution structural information for the SMR family has begun to emerge. First, crystal structures of a Gdx homologue from Clostridiales, Gdx-Clo, were resolved in complex with substituted guanidinium compounds including octylguanidinium (Kermani et al., 2020). In addition to revealing the binding mode of the guanidinyll headgroup, the structure of Gdx-Clo with octylguanidinium showed that hydrophobic repacking of residues lining one side of the binding pocket opens a portal from the substrate binding site to the membrane interior, accommodating the substrate's long alkyl tail. In addition, a model of an EmrE mutant with reduced conformational exchange dynamics, S64V, computed from extensive NMR measurements, was also reported recently (Shcherbakov et al., 2021)

Here, we report several crystal structures of EmrE, including a low-pH (proton-bound) structure and five structures in complex with structurally diverse quaternary phosphonium, quaternary ammonium, and planar aromatic substrates. Structure determination was facilitated by repurposing a monobody crystallization chaperone that we originally developed for Gdx-Clo

(Kermani et al., 2020). The EmrE structure reported here has high structural similarity to Gdx-Clo, but with notable differences in the hydrogen bond network of the substrate-binding site. The various substrates are accommodated by EmrE with minimal changes in the backbone structure. Instead, binding site tryptophan and glutamate sidechains adopt different rotamers to accommodate different drugs. These sidechain motions expand or reduce the binding pocket and provide ring-stacking interactions for structurally disparate substrates. We propose that, compared with the closely related but more selective SMR, Gdx-Clo, a reduced network of hydrogen bond interactions in the EmrE binding site allows sidechain flexibility to accommodate polyaromatics, substituted guanidinyll compounds, and quaternary ammoniums and phosphoniums without requiring substantial alteration of EmrE's backbone configuration.

3.2 Results

3.2.1 Engineering of EmrE to introduce a monobody binding site

We recently solved a crystal structure of a metabolic Gdm⁺ exporter from the SMR family, Gdx-Clo (Kermani et al., 2020). For this effort, we selected monobody crystallization chaperones from large combinatorial libraries (Koide et al., 2012; Sha et al., 2017), which aided in crystallization of the transporter. Upon structure determination, we noticed that the interface between Gdx-Clo and monobody L10 is limited to a nine-residue stretch of loop one that is relatively well-conserved among SMR proteins (**Figure 3.1A**). Moreover, crystal contacts are mediated almost entirely by the monobody, whereas contacts between the transporter and a symmetry mate are limited to five hydrophobic residues contributed by TM4_A and TM4_B (**Supplementary Fig. B.1.1**). These observations suggested that conservative mutagenesis of EmrE loop one to introduce the Gdx-Clo residues might permit monobody L10 binding in order

to facilitate crystallization of EmrE. We therefore designed a triple mutant, E25N, W31I, V34M, which we call EmrE₃. Previous studies showed minimal functional perturbation upon mutation of E25 and W31 to Ala or Cys (Elbaz et al., 2005; Yerushalmi & Schuldiner, 2000). All three residues are located at a distance from the substrate-binding site, and none of the three are conserved in the SMR family.

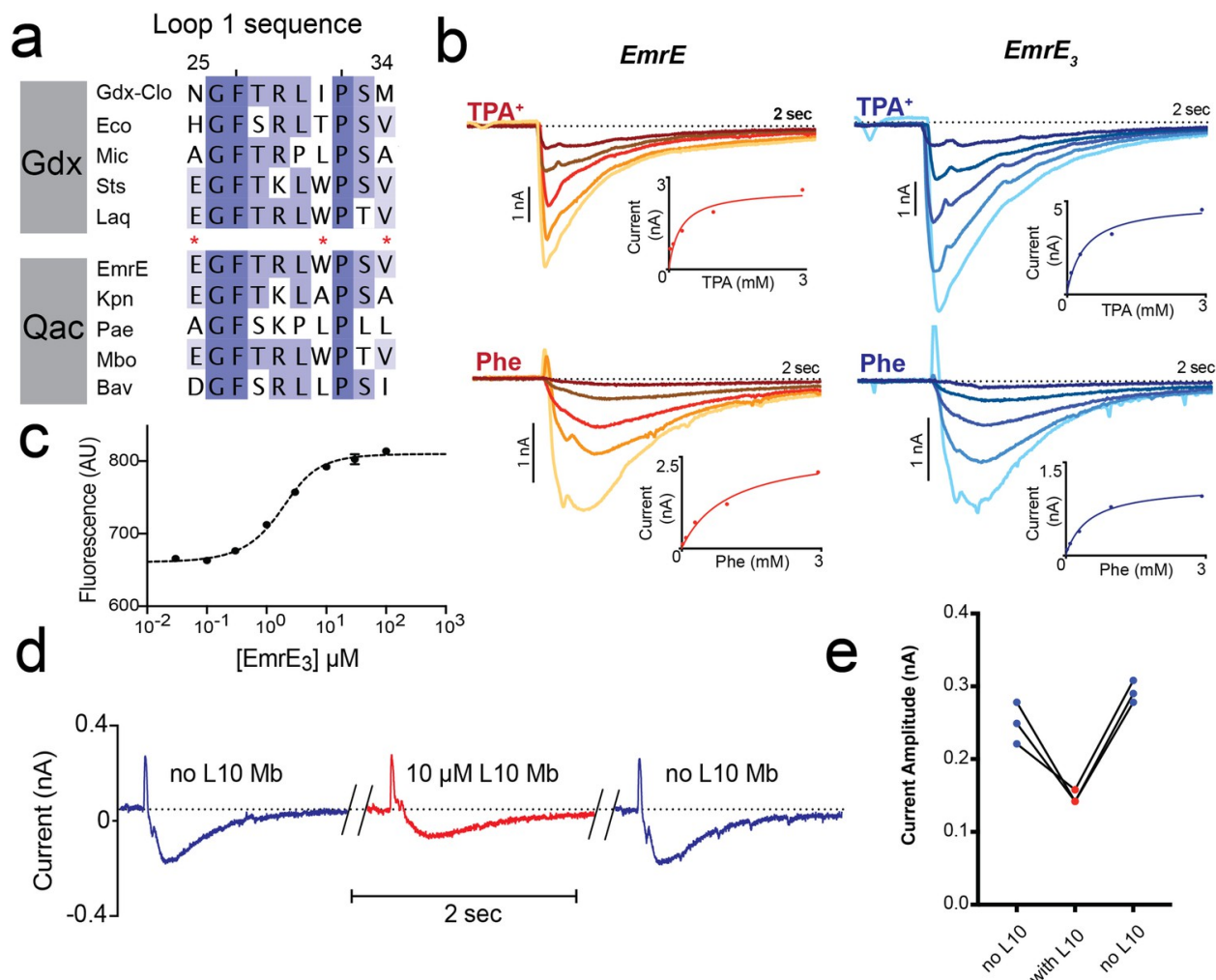


Figure 3.1: Introduction of monobody binding epitope to EmrE.

(A) Sequence alignment for loop 1 of selected SMR proteins, numbered according to EmrE sequence. From top to bottom: representative Gdx sequences (*Clostridiales* bacterium oral taxon 876, *Escherichia coli*, *Micromonospora*, *Streptomyces tsukubensis*, and *Leifsonia aquatica*) and representative Qac sequences (*Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Mycobacterium bovis*, and *Bordetella avium*). Positions mutated in the EmrE₃ construct (E25N, W31I, V34M) are indicated with red asterisks. Sequence conservation analysis for this loop is shown in **Figure 3.4**. (B) Representative currents evoked by WT EmrE or EmrE₃ sensors (shades of red and blue, respectively) with 30 μM – 3 mM TPA⁺ (top panels) or PheGdm⁺ (Phe, lower panels). Insets show plot of peak current amplitude as a function of substrate concentration for a representative titration performed using a single sensor. Solid lines represent fit of datapoints from a single titration series to the Michaelis-Menten equation. K_m values for independent replicates are reported in **Figure 1—figure supplement 2**. (C) Microscale thermophoresis measurement of EmrE₃ binding to monobody L10. Points and error bars represent mean and SEM of three independently prepared samples. Where not visible, error bars are smaller than the diameter of the point. Dashed line represents fit to **Equation 1** with K_d = 850 nM. Representative raw data trace is shown in **Figure 1—figure supplement 3**. (D) EmrE₃ currents evoked by 1 mM PheGdm⁺. Sensors were incubated for 10 min in the presence (red traces) or absence (blue traces) 10 μM monobody L10 prior to initiating transport by perfusion with PheGdm⁺. Currents shown are from a representative experimental series using a single sensor preparation. (E) Peak currents measured for three independent perfusion series performed as in panel D. Peak currents decreased an average of 40% ± 1.5% in the presence of monobody.

In accord with these observations, solid supported membrane (SSM) electrophysiology experiments showed that EmrE₃ mutant is active and transports representative substrates tetrapropylammonium (TPA⁺) and phenylguanidinium (PheGdm⁺). Upon perfusion with substrate, negative capacitive currents are evoked, indicating an electrogenic transport cycle, with substrate transport coupled to the antiport of ~2 H⁺, as has been previously reported for these (Kermani et al., 2020) and other substrates (Adam et al., 2007; Rotem & Schuldiner, 2004; Soskine et al., 2004). In SSM experiments, the peak capacitive current corresponds to the initial rate of substrate transport (Bazzone et al., 2017). The SSM electrophysiology traces are very similar for WT EmrE and EmrE₃ (**Figure 3.1B**). Measurements of peak currents as a function of substrate concentration were fit to the MichaelisMenten equation, yielding Km values within twofold of those measured for WT EmrE (**Figure 3.1B, Supplementary Fig. B.1.2**). Microscale thermophoresis experiments show that EmrE₃ binds monoclonal antibody L10 with a K_d of 850 nM (**Figure 3.1C, Supplementary Fig. B.1.3**), indicating that these small modifications at surface exposed residues were sufficient to create a monoclonal antibody binding site. Similar to our observation for Gdx-Clo (Kermani et al., 2020), addition of saturating L10 monoclonal antibody (10 μM) depresses transport currents mediated by EmrE₃ by about 40% but does not altogether inhibit substrate transport (**Figure 3.1D and E**). Currents are fully restored upon subsequent incubation with monoclonal antibody-free solution. Thus, EmrE₃ is functionally equivalent to WT EmrE, is capable of binding monoclonal antibody L10, and retains function when this monoclonal antibody is bound.

3.2.2 Structure of EmrE3 without ligand at pH 5.2

When combined with monoclonal antibody L10, EmrE₃ crystallized and diffracted to a maximum resolution of 2.9 Å. The crystallization conditions differed from those used for the Gdx-

Clo/monobody complex, but the space group, C121, and approximate dimensions of the unit cell were the same (Kermani et al., 2020). We solved the structure using molecular replacement, with the L10 monobodies and the first three helices of each Gdx-Clo monomer as search models. After phasing, loop 3 and helix 4 were built into the experimental density followed by iterative rounds of refinement (**Figure 3.2A, Table 4, Supplementary Fig. B.2.1.A and B**). The model was validated by preparing a composite omit map in which 5% of the atoms in the model were removed at a time (Terwilliger et al., 2008) (**Supplementary Fig. B.2.1.C and D**). Our EmrE₃ model corresponds well with the composite omit maps, suggesting that model bias introduced by using Gdx-Clo as a molecular replacement search model does not unduly influence our model of EmrE₃.

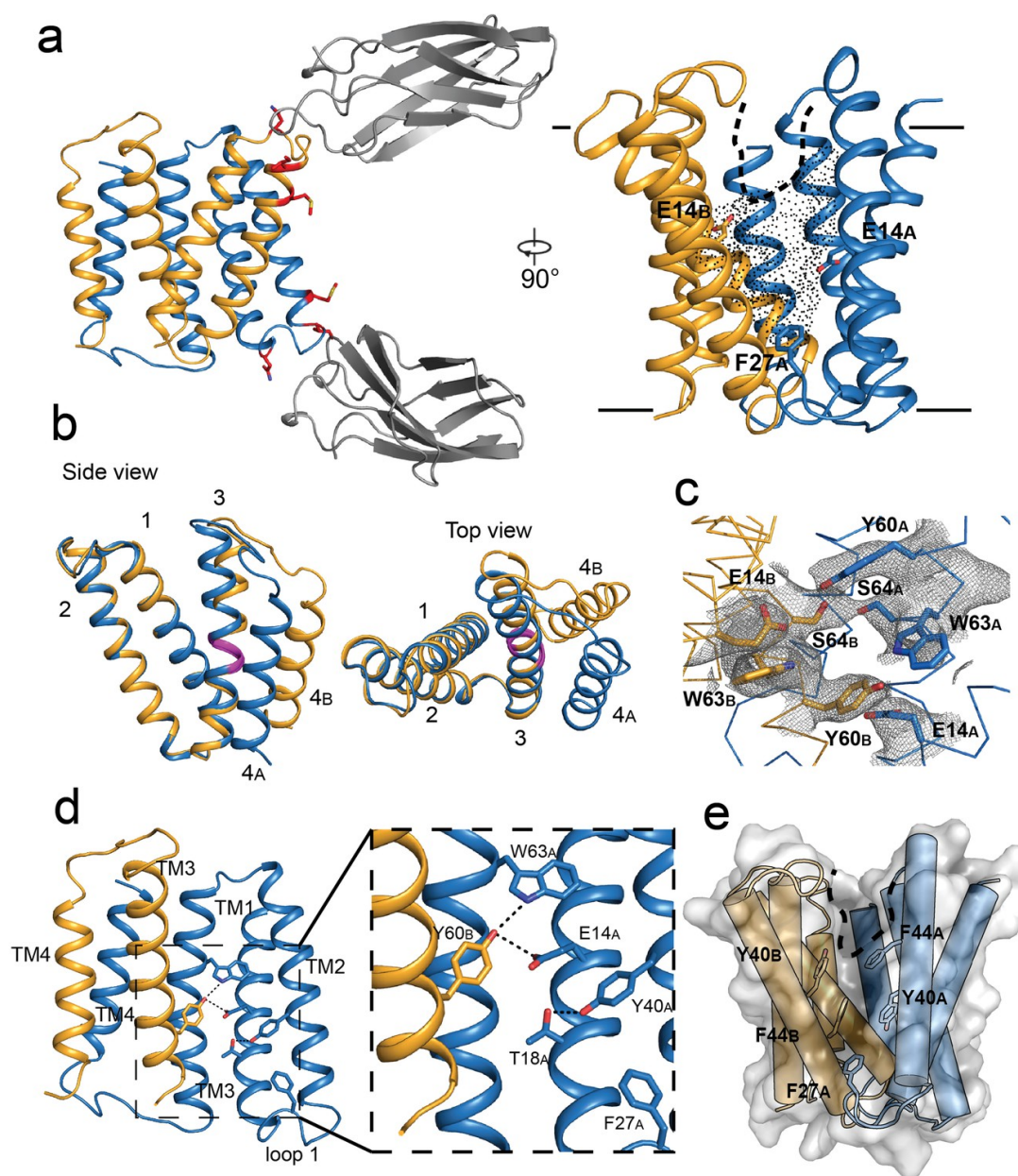


Figure 3.2: Crystal structure of EmrE₃.

(A) Subunits A and B are shown in blue and orange, respectively, and monobody L10 is shown in gray. In the left panel, mutated residues E25N, W31I, V34M are shown in red with sidechain sticks. In the right panel, the monobodies are removed for clarity. E14_A, E14_B, and F27_A are shown as sticks, and the aqueous accessible region of the transporter is indicated with dots. Approximate membrane boundaries are shown as solid lines, and the boundary of the membrane portal is shown as a dashed line. (B) A (blue) and B (orange) subunits of EmrE₃, aligned over residues 1–63. The GVG fulcrum sequence in TM3 is colored in magenta. (C) S64 and surrounding sidechains with 2F_o-F_c density shown as gray mesh (contoured at 1.0 σ within 2 Å of selected residues). (D) Y60_B hydrogen bonding network. EmrE dimers are shown with TM1 and TM2 of subunit B (orange) removed for clarity. Lower panels show zoomed in view. In each view, interactions within hydrogen bonding distance and geometry are shown as dashed lines. E. Surface rendering of EmrE₃. TM2 sidechains that line the portal are shown as sticks.

The structure of the EmrE₃/L10 complex (**Figure 3.2A**) shows an antiparallel EmrE₃ dimer bound to two monobodies in slightly different orientations via the loop one residues. The crystal packing is similar to Gdx-Clo, with the majority of contacts mediated by monobody. The introduced E25N sidechain of EmrE₃ is within hydrogen bonding distance of a tyrosine sidechain contributed by the monobody, and W31I contributes to a hydrophobic patch of the transporter/monobody interface. These interactions are homologous to those observed for the Gdx-Clo/L10 complex. The third mutant sidechain of EmrE₃, V34M, does not interact with monobody in this structure, and therefore might not be necessary for monobody binding to EmrE₃.

In our EmrE₃ model, the positions of the helices agree with those observed in existing low-resolution electron microscopy maps of EmrE (Ubarretxena-Belandia et al., 2003 (Ubarretxena-Belandia et al., 2003) (**Supplementary Fig. B.2.2A**). Compared with a previous MD model based on that EM data (Vermaas et al., 2018), our current EmrE₃ crystal structure has a C α RMSD of 2.5 Å, with close correspondence of residues that contribute to the substrate-binding pocket (**Supplementary Fig. B.2.2B**). Although EmrE₃ has high structural similarity to Gdx-Clo (C α RMSD 1.2 Å for the dimer), the structures display clear differences in subunit packing. Relative to Gdx-Clo, in EmrE₃ helices 1–3 of the A subunit, which line the binding pocket, are each displaced by 1.5–2.5 Å (**Supplementary Fig. B.2.2C**). These shifts slightly expand the aqueous cavity of EmrE₃ relative to Gdx-Clo.

As in Gdx-Clo, the two monomers adopt different structures. Monomers A and B differ from each other in the relative orientation of their two lobes (residues 1–66 and 67–103) about a fulcrum at the conserved GVG motif in helix 3 (residues 65–67; **Figure 3.2B**). The angle of the bend in TM3 at the GVG sequence is somewhat more pronounced in monomer A (17°) than in

monomer B (9°). The observed architecture is in accord with the proposed conformational swap of two structurally distinct monomers (Morrison et al., 2011).

The residue S64 is positioned immediately before the GVG fulcrum, at the boundary of lobe 1 and lobe 2 for each EmrE₃ subunit. In the crystal structure, the S64 sidechains contributed by the two subunits are within hydrogen bonding distance and geometry, with strong contiguous electron density between them (**Figure 3.2C**). Due to the antiparallel architecture, the outward- and inward-facing conformations of the transporter are expected to be structurally identical and related by twofold symmetry about an axis parallel to the plane of the membrane (Fleishman et al., 2006). Thus, the S64 interaction should be preserved when the transporter is open to the opposite side of the membrane; we therefore imagine that the S64 sidechains remain hydrogen bonded to each other during the entire transport cycle, forming the pivot point around which the conformational change occurs.

In the absence of ligand, EmrE₃ possesses a deep, spacious aqueous pocket that is accessible from one side of the membrane (**Figure 3.2A**). The E14 sidechains contributed by both subunits define the edges of this binding pocket. E14 is invariant in the SMR family and essential for binding both substrate and protons (Yerushalmi & Schuldiner, 2000). The present crystals formed at pH 5.2, at which both E14 sidechains are expected to be protonated (Li et al., 2021; Morrison et al., 2015). There is a small, spherical density in the vestibule between W63_B and E14_A that is consistent with a water molecule, although no other ordered water molecules are visible at this resolution (**Supplementary Fig. B.2.3**). The cross-subunit interaction between Y60_B and E14_A proposed by Vermaas et al. is observed (**Figure 3.2D**). A conserved hydrogen bond acceptor, T18_A, is located one helical turn down from E14_A and engaged in an intrasubunit interaction with Y40_A (**Figure 3.2D**).

As in Gdx-Clo, the TM2 helices splay apart on the open side of the transporter, defining a portal from the membrane to the substrate binding site that is lined with hydrophobic sidechains (**Figure 3.2E**). This portal may play a dual role, rearranging to allow alkyl substituents to reside in the membrane during the transport cycle, as well as providing the opportunity for hydrophobic drugs to diffuse laterally from the membrane into the substrate binding site. Aromatic residues contributed by loop 1_A, including the highly conserved F27 sidechain, are wedged between the hydrophobic sidechains lining helices 2_A and 2_B, sealing the closed side of the transporter (**Figure 3.2E**).

3.2.3 Structures of substrate-bound *EmrE*₃

To understand how different substrates interact with EmrE, we screened a variety of transported compounds in crystallization trials. We were able to obtain diffracting crystals in the presence of five structurally diverse compounds transported by EmrE: monovalent planar aromatic harmane (3.8 Å), divalent planar aromatic methyl viologen (3.1 Å), quaternary phosphoniums tetraphenylphosphonium (TPP⁺; 3.4 Å) and methyltriphenylphosphonium (MeTPP⁺; 3.2 Å), and quaternary ammonium benzyltrimethylammonium (3.9 Å) (**Table 4**). We were unable to generate crystals that diffracted to high resolution in the presence of metformin, benzalkonium, cetyltrimethylammonium, or octylguanidinium. Phases of the EmrE₃/substrate/L10 monobody complexes were determined using molecular replacement with the pH 5.2 structure as a search model. Although the crystallization conditions varied for each substrate, the TPP⁺-, MeTPP⁺-, benzyltrimethylammonium-, and harmane-bound proteins crystallized in the same unit cell as proton-bound EmrE₃, with one copy of the EmrE₃/L10 complex in the asymmetric unit. The methyl viologen-bound protein crystallized in P1 with two

pseudosymmetric copies of the EmrE₃/L10 complex in the asymmetric unit, organized in the same relative orientation as individual complexes in the C121 crystal form.

Since Gdx-Clo and EmrE₃ were both accommodated in this crystal lattice despite differences in the tilt and packing of helices 1, 2, and 3, we expect that small 1–2 Å substrate-dependent movements in the backbone of EmrE₃ would also be tolerated within this crystal lattice. However, in all four substrate-bound structures, the transmembrane helices, and loops 1 and 2 conform almost perfectly to the pH 5.2 structure (C α RMSD = 0.5–0.65 Å), suggesting that the observed backbone conformation is the lowest energy state for both the substrate- and proton-bound transporter. Loop three is poorly ordered and adopts a different conformation in each of the structures in which it is resolved well enough to model.

For all substrate-bound structures, the maps show positive densities between the substrate binding E14 residues, including a four-lobed density for TPP⁺, a three-lobed density for MeTPP⁺, and oblong densities for the harmane and the methyl viologen structures. We modeled the corresponding substrates into each of these densities (**Figure 3.3A and B**). All five drugs are bound at the bottom of the aqueous cavity, in overlapping positions at the midpoint of the membrane. In the two copies of the methyl viologen-bound transporter, the drug is bound in different (but overlapping) positions (**Supplementary Fig. B.3.1**). For all substrates, the center of mass is poised midway between the E14 residues. To different extents, the substrates also interact with the protein's aromatic residues via ring stacking, especially Y60 and W63.

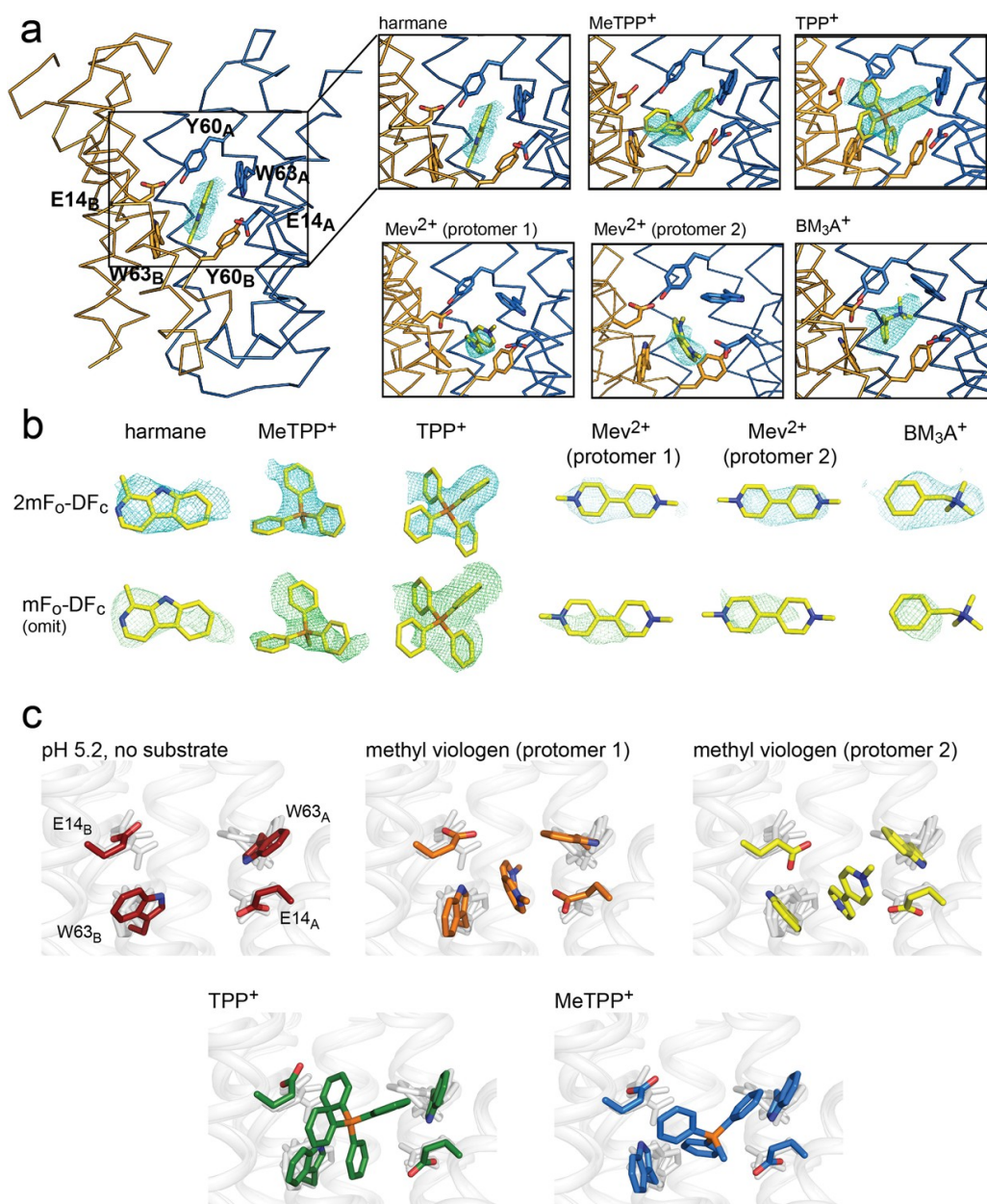


Figure 3.3: Substrate binding to EmrE3.

(A) Structures are shown in ribbon representation, with sidechains E14, W63, and Y60 shown as sticks. All panels are zoomed and oriented the same. 2mF_o-DF_c maps (carved 2 Å around each substrate) are shown as cyan mesh. Maps are contoured at 1σ for harmane and 1.2σ for MeTPP⁺, TPP⁺, methylviologen, and benzyltrimethylammonium (BM₃A⁺). (B) Top row: Substrate structures and 2mF_o-DF_c maps from the panels in A, individually rotated to view each substrate. Bottom row: mF_o-DF_c substrate omit maps shown as green mesh. Omit maps are contoured at 1.8σ for harmane and 2σ for MeTPP⁺, TPP⁺, methylviologen, and BM₃A⁺. (C) Comparison of E14 and W63 positions in each substrate-bound structure. Individual panels show substrate, E14, and W63 from indicated structure in color aligned with the other four structures, which are rendered in light gray.

Comparison of these structures permitted evaluation of the specific orientations of the sidechains that line the substrate binding site (**Figure 3.3C**). The harmane- and benzyltrimethylammonium-bound structure was excluded from this analysis because, at 3.8–3.9 Å resolution, we were not as confident about interpreting subtle changes in sidechain orientation. For the other substrates (methyl viologen, TPP⁺, and MeTPP⁺), this comparison showed that binding site sidechains, especially E14 and W63, adopt different rotamers, thus accommodating the differently sized substrates. For example, the carboxylate of E14_B is displaced by 2.5 Å when the bulky quaternary phosphonium TPP⁺ is bound, compared to its position when the planar methyl viologen occupies the binding site. Likewise, the position of the W63_A indole ring rotates over approximately 80° depending on the substrate that occupies the binding site. To validate these observations, we performed refinements with models in which the position of the W63_A or E14_B sidechain was adjusted to match its position in the presence of a dissimilar substrate; the resulting difference density demonstrates that these substrate-dependent changes in sidechain rotamer are not due to model bias during the refinement (**Supplementary Fig. B.3.2 and B.3.3**). Thus, these structures provide a first suggestion of how rotameric movements of EmrE's charged and aromatic sidechains can change the dimensions of the binding pocket and interact favorably with diverse substrates.

3.2.4 Structure of Gdx-Clo at pH 5 and comparison to the substrate binding site of EmrE

The overall fold and many of the binding site sidechains are shared between EmrE and Gdx-Clo, yet the two proteins have markedly different substrate selectivity profiles. We therefore sought to analyze how molecular interactions among binding site residues might explain the different substrate selectivity for EmrE and Gdx-Clo. Previous structures of Gdx-Clo were solved at pH ≥ 7.5 in complex with substituted guanidinyll compounds (Kermani et al., 2020). To

compare the substrate binding sites of Gdx-Clo and EmrE₃ in equivalent states, we solved a new structure of Gdx-Clo at pH 5.0, which is close to the value for the present low pH EmrE₃ structure, pH 5.2 (**Table 4, Supplementary Fig. B.4.1A**). Both transporters are likely proton-bound at this pH, minimizing differences in sidechain positioning that might stem from interactions with bound substrate. This new structure of proton-bound Gdx-Clo, which is resolved to 2.3 Å, is highly similar to the structure of substrate-bound Gdx-Clo (PDB: 6WK8), with only a local change in the rotamer of the substrate-binding glutamate E13_B (**Supplementary Fig. B.4.1B**).

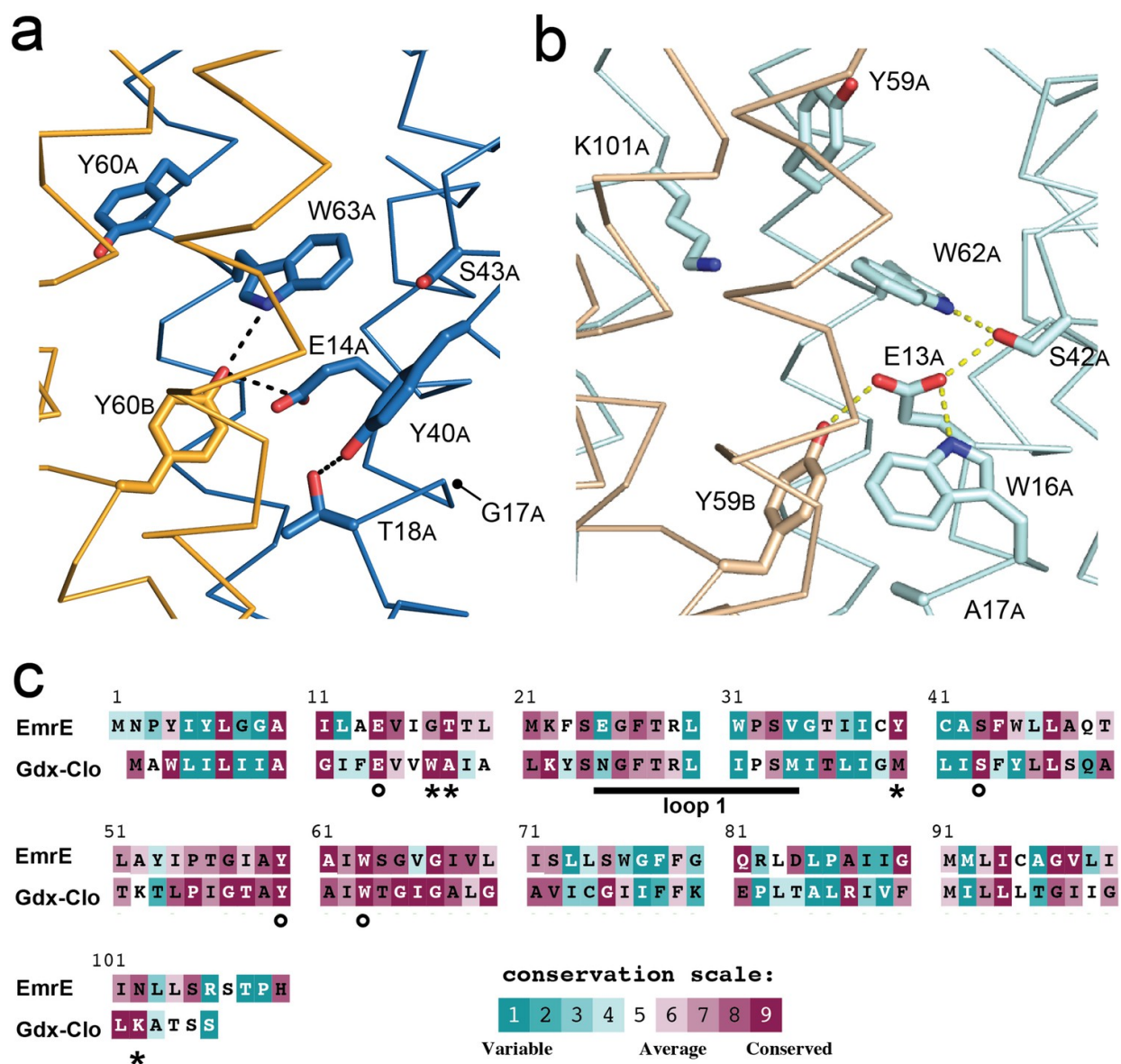


Figure 3.4: Structure and sequence conservation of substrate binding site residues in Qac and Gdx subtypes.

(A) Substrate-binding site in EmrE, with subunit B in orange and subunit A in blue. (B) Substrate-binding site in Gdx-Clo, with subunit B in wheat and subunit A in pale cyan (PDB: 6WK8). For panels A and B, the proteins are shown in the same orientation. Note that residue numbering is offset by one in Gdx-Clo. Potential hydrogen bonds are shown as dashed lines. (C) Amino acid conservation analysis for the Qac and Gdx subtypes overlaid on exemplar sequences of EmrE and Gdx-Clo, respectively. Analysis was performed using ConSurf (Ashkenazy et al., 2016; Berezin et al., 2004). Residues that contribute to the binding pocket and that are conserved between the Qac and Gdx subtypes are indicated with an asterisks. Residues that contribute to the binding pocket and that differ between the Qac and Gdx subtypes are indicated with a circle. The monobody binding loop 1 is indicated by the solid line. Alignments of representative sequences are shown in **Supplementary Fig. B.4.2**.

A comparison of the low-pH EmrE₃ and Gdx-Clo structures reveals conspicuous differences in the hydrogen bond network within the binding cavity (**Figure 3.4A and B**), despite the conservation of many key residues. In Gdx-Clo, Ser42 participates in the stack of alternating hydrogen bond donors and acceptors (W16_{Clo}/E13_{Clo}/S42_{Clo}/W62_{Clo}) that fixes the position of the central Glu, E13. Although the analogous serine (S43_{EmrE}) is present in EmrE, it is not playing an analogous role. A 1.5 Å displacement in helix two has distanced this Ser from the other sidechains in the binding pocket, beyond hydrogen bonding distance with W63_{EmrE}. Instead, S43_{EmrE} is rotated away from the aqueous cavity and the central E14_{EmrE} residues. Despite strict conservation of this serine among the Gdx subtype, mutation to alanine occurs in ~30% of homodimeric Qacs (**Supplementary Fig. B.4.2**). In lieu of an interaction with S43_{EmrE}, both W63_{EmrE} sidechains in EmrE adopt different rotamers compared to their counterparts in Gdx-Clo. W63_{A, EmrE} is oriented so that its indole NH is within H-bonding distance of Y60_{B, EmrE}, although the angle between the H-bond donor and acceptor is ~30° off normal.

The fourth residue from Gdx-Clo's H-bond stack, W16_{Clo}, is universally conserved in Gdx proteins, but replaced with a glycine or alanine in the Qacs (G17 in EmrE). There is no equivalent H-bond donor to the central Glu in EmrE. Instead, the sidechain Y40_{EmrE} occupies this space, but interacts with T18_{EmrE} located one helical turn away from E14_{EmrE}. This pair, Y40_{EmrE} and T18_{EmrE}, are highly conserved among the Qacs, and variable and typically hydrophobic in Gdx proteins. In Gdx-Clo, the corresponding positions are M39_{Clo} and A17_{Clo}. This trio of correlated positions (W16_{Clo}/G17_{EmrE}, A17_{Clo}/T18_{EmrE}, and M39_{Clo}/Y40_{EmrE}) in the substrate-binding site are among the main features that differentiate the Gdx and Qac subtypes in sequence alignments (**Figure 3.4C**).

Y60_A, EmrE also adopts a different orientation in EmrE relative to the position of the analogous Tyr, Y59_{Clo} in Gdx-Clo. Rather than extending out of the binding pocket toward the exterior solution, as it does in Gdx-Clo, Y60_A, EmrE is pointed down toward the S64_{EmrE} diad. This rotamer would not be possible in Gdx-Clo, since this space is occupied by K101_{Clo} instead, which extends from the C-terminal end of helix 4 and points down into the substrate-binding pocket toward the glutamates. K101_{Clo} is completely conserved in the Gdx subtype.

The overall picture that emerges from this comparison of the Gdx-Clo and EmrE structures is that the two proteins share many binding site residues but differ in the relative organization of these residues. In Gdx-Clo, E13_{Clo}, S42_{Clo}, Y59_{Clo}, and W62_{Clo} are constrained in a highly organized H-bond network. In EmrE, residues peripheral to the binding site have encroached on these positions, disrupting the network and reducing the number of protein hydrogen bond partners for each of these conserved sidechains.

3.2.5 EmrE is tolerant of mutations that eliminate hydrogen bonding in the binding pocket

Based on structural comparison of the Gdx-Clo and EmrE-binding pockets, we hypothesize that even for conserved residues in the binding pocket, the importance of hydrogen bonding is diminished in EmrE relative to Gdx-Clo. To probe this, we performed a head-to-head comparison of SSM currents mediated by EmrE and Gdx-Clo proteins with mutations at three conserved positions adjacent to the functionally essential central Glu: Y59F_{Clo}/Y60F_{EmrE}, S42A_{Clo}/S43A_{EmrE}, and W62F_{Clo}/W63F_{EmrE} (**Figure 3.5, Table 2**). All six mutant transporters were expressed at near-WT levels and monodisperse by size exclusion chromatography. For EmrE mutants, we tested transport of 2 mM PheGdm⁺ or 2 mM TPA⁺, and for Gdx-Clo, we tested transport of its native substrate, 1 mM Gdm⁺. For all experiments, substrate concentration was ~4 fold higher than the transport K_m.

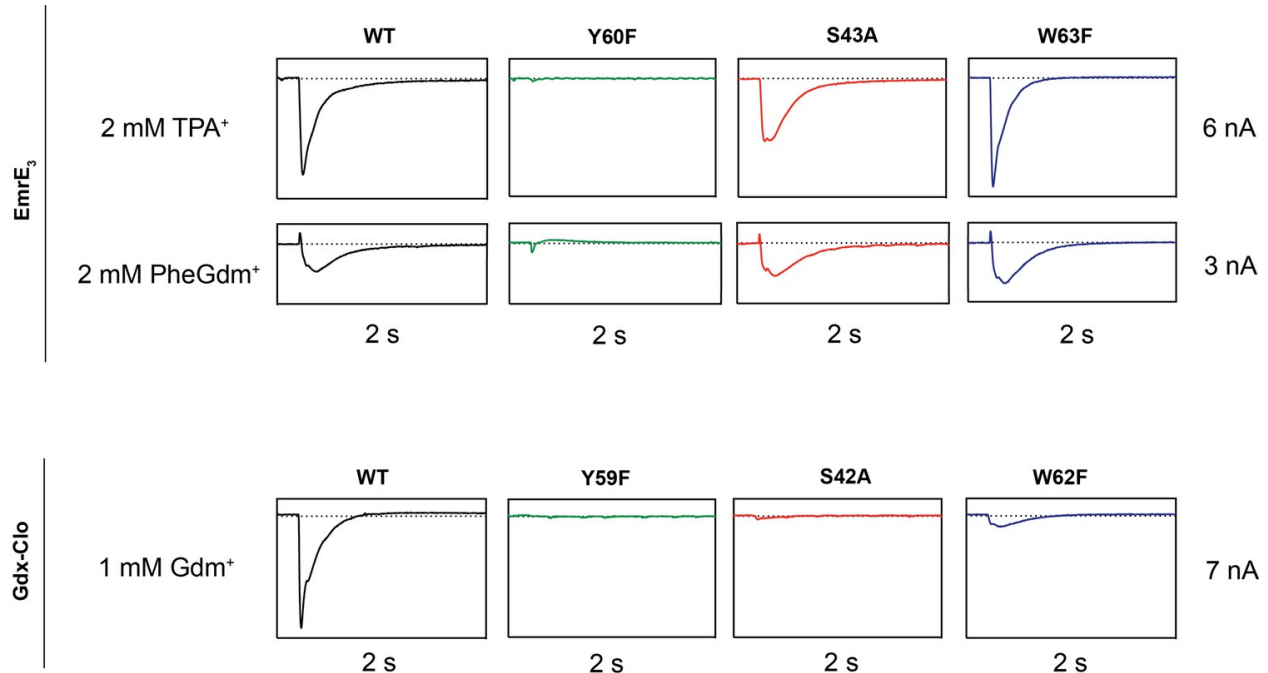


Figure 3.5: Representative SSM electrophysiology recordings for EmrE₃ and Gdx-Clo mutants.

For EmrE₃, PheGdm⁺ and TPA⁺ traces are from the same sensor and shown on the same scale. Vertical box edges are 3 nA for PheGdm⁺ traces, and 6 nA for TPA⁺ traces. For Gdx-Clo, vertical box edges are 7 nA. Horizontal box edges are 2 s for all traces. Dashed line represents the zero-current level. Traces are representative of currents from three independently prepared sensors and two independent biochemical preparations. Peak current values for all replicates are reported in **Table 2**. Note that because there is some sensor-to-sensor variation in liposome fusion, comparisons of current amplitude among the mutants are qualitative.

In line with its proposed role as a conformational switch (Kermani et al., 2020; Vermaas et al., 2018), no currents were observed when the binding site Tyr (Y59_{Clo}/Y60_{EmrE}) was mutated in either protein. This result recapitulates results from prior radioactive uptake studies of both mutants (Kermani et al., 2020; Rotem et al., 2006). It also establishes a dead-transporter control for our SSM electrophysiology assays. We likewise find that Gdx-Clo does not tolerate perturbation to its hydrogen bond stack. Although neither S42A_{Clo} nor W62F_{Clo} directly bind Gdm⁺, both mutations eliminate Gdm⁺ currents in SSM electrophysiology assays. In contrast, EmrE₃ was relatively indifferent to the S43A_{EmrE} and W63F_{EmrE} mutations, with robust currents evoked by both TPA⁺ and PheGdm⁺.

This result for S43_{EmrE} reinforces the structural suggestion that the serine's functional role in the Gdx transporters is not conserved in the Qac subtype and is also in agreement with prior transport and resistance assays that showed that S43_{EmrE} modulates substrate specificity in EmrE, but is not required for transport function (Brill et al., 2015; Wu, 2019). The observation of robust transport by W63F_{EmrE} is more surprising, since this mutant has been shown to reduce TPP⁺ binding by two orders of magnitude, and abolish methyl viologen transport and bacterial resistance to TPP⁺, methyl viologen, and acriflavine (Elbaz et al., 2005). Other mutations to W63 (to C, A, or V) also fail to provide resistance against polyaromatic substrates (Amadi et al., 2010; Elbaz et al., 2005; Wu, 2019).

3.3 Discussion

In this work, we describe substrate- and proton-bound crystal structures of the *E. coli* SMR transporter EmrE, which is wildtype except for three functionally neutral mutations that enable monobody binding, and thus, crystallization. Functional assays show that the engineered protein, EmrE₃ behaves like wildtype, and that the transporter remains functional in the presence of monobody. Below, we discuss the crystallization strategy, we evaluate differences between our crystal structures and a recent NMR-derived model of EmrE (Shcherbakov et al., 2021), and discuss the implications of our structures for understanding substrate polyspecificity by EmrE.

3.3.1 The application of multipurpose chaperones for crystallization

The minimal monobody binding interface permitted a crystallization chaperone developed for Gdx-Clo to be repurposed for binding and crystallization of a new target with structural homology, but only 35% sequence identity to the original, streamlining the structural

characterization process. Given the similarity of this loop among diverse SMR proteins, we think that this approach would likely facilitate the structural characterization of any target within the SMR family. Such general adapters and chaperones to facilitate structural biology have been described before for various targets (Dutka et al., 2019; Koldobskaya et al., 2011; McIlwain et al., 2021; Mukherjee et al., 2020). Although identification of a general SMR monobody was not the original intent of the monobody selection, in cases where multiple homologous targets have been identified, variants with identical or near-identical epitopes could be generated, and binders with broad utility could presumably be selected. Especially in the case of bacterial proteins, in which there are many clinically relevant homologues from many diverse species, such general structural biology approaches hold promise to facilitate molecular characterization of membrane protein targets.

The monobody chaperones mediate most of the crystal contacts, permitting Gdx-Clo and EmrE to crystallize in a nearly identical unit cell, despite some structural differences, including 1–2 Å displacements of helices that contribute to the binding pocket. Although it is a misconception that crystallization chaperones can ‘force’ the transporter into a non-native, high-energy conformation (Koide, 2009), it is plausible that the monobody chaperones recognize a less-prevalent conformation, and kinetically trap the transporter in a minority state within the native conformational ensemble. Because these monobodies were not selected against EmrE, but against a different homologue from the SMR family, this is a possibility that should be considered. However, two lines of evidence disfavor the possibility that the monobody-bound state is aberrant. First, we showed that monobody binding has only a minor effect on transport function, and second, our model corresponds closely to the helix density in the EM dataset, which was obtained without exogenous binding proteins (UbarretxenaBelandia et al., 2003).

Although local perturbations at the monobody-binding interface of loops 1_A and 1_B cannot be ruled out, the position of loop 1_A is consistent with prior spectroscopic data, which predicted that in the major solution conformation, F27_A packs against the B subunit with its sidechain oriented toward the substrate-binding site (Dastvan et al., 2016). Loop 1_B is located on the open side of the transporter and does not form any intra-transporter contacts. Therefore, even if monobody does stabilize a less-prevalent conformation of loop 1_B, this would not change the major interpretations of the present structures.

3.3.2 Comparison to the NMR model of EmrE S64V

An NMR-based model of the ‘slow-exchanging’ EmrE mutant S64V was recently published (Shcherbakov et al., 2021). S64V binds substrate with similar affinity as wildtype, but the rate of conformational exchange is about an order of magnitude slower (Wu et al., 2019). This model was computed based on chemical shift measurements and distance restraints between the protein backbone and the fluorinated substrate tetrafluorophenyl phosphonium (F-TPP⁺). Although our present crystal structures agree with the NMR model in general aspects, such as the antiparallel topology, there are also notable differences in the global conformation, with an overall RMSD of 2.3 Å for the two models. Relative to other models of EmrE, including the computational models (Ovchinnikov et al., 2018; Vermaas et al., 2018), the EM α -helical model (Ubarretxena-Belandia et al., 2003), and the present crystal structures, in the NMR model the first lobe of the A subunit is shifted down in a direction perpendicular to the membrane with respect to the B subunit (**Figure 3.6A**). Note that chain A of the NMR structure is more structurally homologous to chain B of the crystal structure and vice versa. Our designation of chains A and B in the present crystal structure correspond to the A and B chains in previous literature, including SMR family homologue Gdx-Clo (Kermani et al., 2020), the low-resolution

EmrE structures of EmrE (Chen et al., 2007; Fleishman et al., 2006), and theoretical EmrE models (Ovchinnikov et al., 2018; Vermaas et al., 2018) This difference in subunit packing is accompanied by subtle differences in the tilts of the helices (**Figure 3.6B**). In the NMR structure, helix 2_A and 2_B become more parallel, and the gap between them is narrowed, reducing membrane access to the binding site via the portal.

3.3.3 Comparison to the NMR model of EmrE S64V

The difference in global conformation of the NMR and crystallography models is supported by a reorganization of the hydrogen bonding network in the substrate binding site (**Figure 6C**). The heart of this change is a rotameric switch by Y60: In the crystal structures, Y60_B participates in a pair of cross-subunit interactions, within coordination distance and geometry of E14_A and W63_A in the opposite subunit. In the NMR model, the same Y60_B sidechain is assigned a different rotamer, its hydroxyl moving 6 Å along helix 1, so that it is now coordinating T18_A, one helical turn away from E14_A. The interaction with Y60_B has displaced Y40_A from its interaction with T18_A. Helix 2_A slides in a direction perpendicular to the membrane so that Y40_A now encroaches on the position of F27_A at the tip of loop 1, which is packed between helices 2_A and 2_B in the crystal structure. In the NMR ensemble, the displaced loop one is flexible and adopts various conformations. The helix density observed in the low-resolution EM dataset corresponds closely to the present crystallography models (Real space correlation coefficient (RSCC) = 0.67), and is less consistent with the NMR model (RSCC = 0.51; Figure 6—figure supplement 1; Ubarretxena-Belandia et al., 2003).

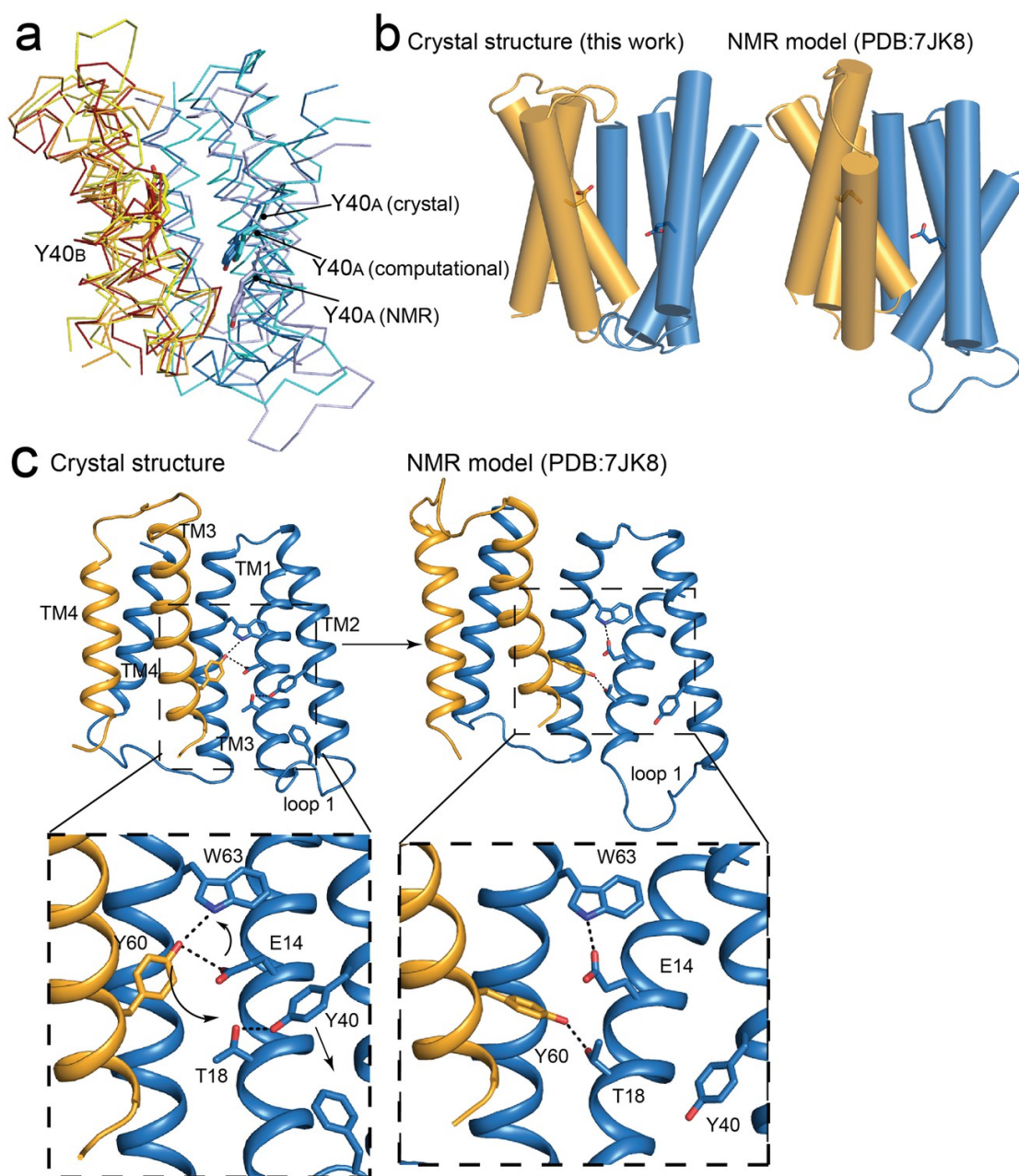


Figure 3.6: Comparisons of NMR and crystallography models of EmrE.

(A) Overlay of crystallography (orange/blue), computational (yellow/cyan; Vermaas et al., 2018) and NMR (dark red/pale blue; Shcherbakov et al., 2021) models, aligned over the B subunit. Y40 sidechain sticks are shown as landmarks. (B) Side-by-side comparisons of the crystallography and NMR models, with A subunit in blue and B subunit in orange. E14 sidechains shown as landmarks. (C) Comparison of Y60_B hydrogen bonding network in the crystal structure (left) and NMR structure (right). EmrE dimers are shown with TM 1 and 2 of subunit B (orange) removed for clarity. Lower panels show zoomed in view. In each view, interactions within hydrogen bonding distance and geometry are shown as dashed lines. Arrows are shown to help visualize sidechain rearrangements between the two structures.

The differences in conformation between the crystallography/EM datasets and the NMR model are unlikely to be due to membrane mimetic (which is shared for the EM and NMR datasets), the presence of monobodies (the EM data was collected without monobodies), or the S64V mutation used for NMR studies (NMR experiments showed little change in backbone configuration for this mutant (Wu et al., 2019)). The elevated temperature of the NMR experiments (45 ° C, compared to 20 ° C for crystallization) may favor different states in a conformational ensemble. Previous EPR measurements may lend support to this possibility (Dastvan et al., 2016). Those experiments showed that at pH 8, with TPP⁺ bound, EmrE adopts a major conformation consistent with our current crystallography model. But when substrate is removed and the pH dropped to 5.5, EmrE's conformational ensemble becomes more heterogeneous. The loops disengage and become more flexible, and a population emerges in which the two subunits have adopted a more symmetric conformation. Perhaps the NMR experiments, which were performed at pH 5.5 (albeit with substrate) reflect that second conformation from the ensemble. Nevertheless, it is also worth noting that our crystallography model is not inconsistent with the backbone chemical shifts measured in bicelles based on structure-trained predictions of chemical shift (Frank et al., 2015; Xie et al., 2020).

3.3.4 movements accommodate diverse substrates

In addition to substantiating prior EmrE experiments, our structures also provide new molecular insights into the binding of structurally diverse substrates by EmrE. Methyl viologen, harmane, Me-TPP⁺, TPP⁺, and benzyltrimethylammonium have considerable structural differences, but are all accommodated in the EmrE binding site with only sidechain rearrangements. The closely related, but substantially more selective SMR family member, Gdx-Clo, provides a useful point of comparison to understand why EmrE can interact with this

chemically diverse range of compounds. In Gdx-Clo, the substrate-binding glutamate sidechains are constrained by a polarized stack of hydrogen bond donors and acceptors that also includes W16_{Clo}, S42_{Clo}, and W62_{Clo}. This hydrogen bonded network would be disrupted by the rotamerization of either E13_{Clo} or W62_{Clo}. We show that in Gdx-Clo, mutations to sidechains that contribute to the hydrogen bond stack seriously impair transport activity.

In contrast, in EmrE, the corresponding residues E14_{EmrE} and W63_{EmrE} are not constrained by such a stack of H-bond donors and acceptors. The current structures and SSM electrophysiology experiments both suggest that, in contrast to Gdx-Clo, a rigid H-bond network is not essential for substrate transport by EmrE, which remains functional when hydrogen bond capacity is eliminated at S43_{EmrE} or W63_{EmrE}. Without the stricter geometric constraints imposed by a polarized stack of sidechain hydrogen bond partners, both E14_{EmrE} and W63_{EmrE} have more flexibility to adopt different rotamers. Like a pair of calipers, the E14_{EmrE} sidechains can move farther apart to accommodate large substrates such as quaternary ammoniums, or closer together for flat, aromatic substrates or substrates with small headgroups, like harmane and methyl viologen or singly substituted guanidinyll compounds. Similarly, W63_{EmrE} has the space and flexibility to rotamerize, which can expand or narrow the binding pocket or allow W63_{EmrE} to pack against the aromatic groups of bound substrates. These structural observations are in agreement with numerous prior studies that have demonstrated an important role for W63_{EmrE} in transport of polyaromatic substrates (Amadi et al., 2010; Elbaz et al., 2005; Saleh et al., 2018; Wu, 2019). We note that although W63_{A, EmrE} does change position in order to conform to different substrates, we did not always observe optimal pi stacking geometry between the substrate and the protein's aromatic residues. Instead, substrate positioning appeared to optimize

electrostatic interactions first, with all substrates situated directly between E14_{A,EmrE} and E14_{B,EmrE}.

Likewise, many EmrE substrates lack the capacity to donate strong hydrogen bonds, reducing the geometric constraints for protein-substrate interactions. Prior MD simulations suggested a dynamic interaction between TPP⁺ and the EmrE-binding pocket (Vermaas et al., 2018), and we expect that many compounds transported by EmrE have some mobility within the binding pocket. In the present structural experiments, we observe this explicitly for methyl viologen, which we identified in different but overlapping positions in the two transporters in the asymmetric unit.

While our experiments indicate that altering sidechain configuration is important to accommodate diverse substrates, backbone conformational changes do not need to be invoked to explain polyspecificity. Indeed, we do not see perturbations in EmrE's main chain structure in the six different EmrE crystal structures resolved here. In addition, the general correspondence of the structures of EmrE and Gdx-Clo indicates that same tertiary architecture can also accommodate substrates with guanidinyll headgroups and/or alkyl tails. These observations also concur with observations from cryo-EM, which showed only minor differences in helix orientation and packing for the apo and TPP⁺-bound structures (Tate et al., 2003). Thus, the crystallized conformation can accommodate substrates from major classes, including quaternary ammoniums, quaternary phosphoniums, planar polyaromatics, and substituted guanidiniums without substantial backbone rearrangement.

3.3.5 Binding of benzalkonium⁺ and other substrates with alkyl chains

Because benzalkonium is especially relevant as a common household and hospital antiseptic to which the Qac proteins provide resistance, we sought to visualize how this

quaternary ammonium compound might interact with EmrE. Although we were unable to generate diffracting crystals of EmrE₃ in the presence of substrates with long alkyl tails, our current structure of EmrE₃ with benzyltrimethylammonium bound (a chemical homologue of benzalkonium with a methyl group in place of the alkyl chain), combined with our previous Gdx-Clo structure, provides a strong indication of how benzalkonium or other detergent-like substrates might bind.

In Gdx-Clo, octylGdm⁺ binds such that its alkyl tail extends out of the aqueous binding pocket and into the membrane. To accommodate the alkyl tail, hydrophobic sidechains lining GdxClo's TM2 portal, including M39_{Clo} and F43_{Clo}, adopted alternative rotamers (Kermani et al., 2020). Although all the substrates in the present EmrE₃ structures were contained within the aqueous pocket, we similarly observe rotameric rearrangements of the TM2 sidechains in different structures, including Y40_{EmrE} and F44_{EmrE} (equivalent to Gdx-Clo's M39_{Clo} and F43_{Clo}) in the harmane and methyl viologen structures. These observations suggest that, as in Gdx-Clo, in EmrE the sidechain packing at the TM2 interface is malleable, and that movements of these residues may remodel the TM2 portal to permit binding of substrates with detergent-like alkyl chains.

Indeed, when the quatery ammonium headgroup of benzalkonium is superposed onto the experimentally determined position of benzyltrimethylammonium in the EmrE₃ binding pocket, the alkyl tail of benzalkonium extends towards the portal defined by the TM2 helices. Although the extended alkyl chain would clash with F44_{B, EmrE}, positioning this sidechain in the 'down' rotamer (analogous to that adopted by F43_{B, Clo} in Gdx-Clo) alleviates all clashes between the substrate and protein and provides unobstructed access for the alkyl tail to the membrane interior. **Figure 3.7** shows a proposed model of benzalkonium binding to EmrE

prepared by aligning its headgroup with benzyltrimethylammonium followed by energy minimization of the complex using MMTK (Hinsen, 2000).

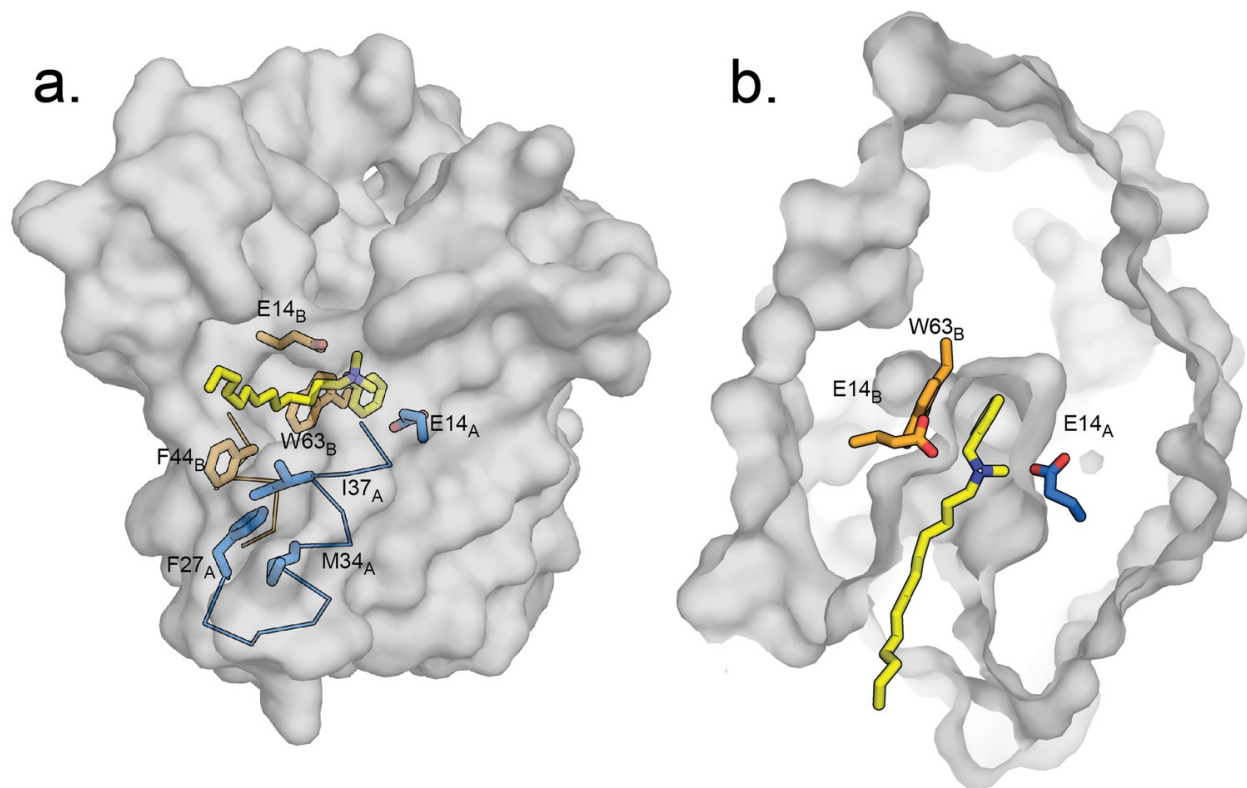


Figure 3.7: Hypothetical model of benzalkonium binding to EmrE.

(A) Benzalkonium is shown in yellow stick representation. Sidechains from the A and B subunits are colored as before. The mainchain for helices lining the TM2 portal is shown in ribbon format, with the portal-lining sidechains shown as sticks. (B) Top-down view of binding site with benzalkonium. EmrE is sliced at the midpoint of the membrane. Comparisons of this model to the experimental models of EmrE in complex with benzyltrimethylammonium (PDB:7T00) and Gdx-Clo in complex with octylguanidinium (PDB:6WK9) are shown in Figure 7—figure supplement 1.

Thus, we propose that sidechain rearrangements along the membrane portal also contribute to substrate polyspecificity by allowing hydrophobic substituents to extend out of the substrate-binding site and access the membrane interior. Similarly, we imagine that dipartite drugs transported by EmrE, such as propidium (a planar polyaromatic group linked to a tetraethyl ammonium) and dequalinium (two aromatic groups with a 10-carbon linker) may also

utilize the portal for transport, with the protein mediated transport of one moiety dragging its tethered lipophilic partner across the membrane.

3.4 Conclusion

In summary, we have developed a multipurpose crystallization chaperone for SMR proteins and used this tool to resolve the first sidechain-resolution crystal structures of the bacterial SMR transporter, EmrE. In order to establish the structural basis of substrate polyspecificity, we resolved structures with five different substrates bound, including quaternary phosphoniums, planar aromatics, and a quaternary ammonium compound. We propose that, compared with more selective representatives of the SMR family, a relatively sparse hydrogen bond network among binding site residues in EmrE permits sidechain flexibility to conform to structurally diverse substrates.

3.5 Materials and Methods

3.5.1 Bioinformatics and sequence analysis

Multiple sequence alignment was performed using MUSCLE (Edgar, 2004). ConSurf was used for sequence conservation analysis (Ashkenazy et al., 2016; Berezin et al., 2004). For this analysis, SMR sequences from GEBA bacterial reference genomes (Mukherjee et al., 2017) that were identified as probable homodimers based on genetic context (those encoded by a single gene in an operon) were further sorted into either Qac or Gdx subclasses using profile Hidden Markov Models built from the corresponding sequence clusters of the functionally annotated sequence similarity network described in Kermani et al., 2020. Representative sequences were

selected for the alignments in **Figures 3.1 and 3.4** because (1) proteins have been characterized in transport or resistance assays and (2) sequences are distributed among different major clades of the phylogenetic tree.

3.5.2 Protein purification and crystallization

L10 monobody was purified from inclusion bodies exactly as described in detail previously (Kermani et al., 2020). pET15b plasmids bearing the EmrE₃ coding sequence with an N-terminal hexahistidine tag and a thrombin cut site were transformed into *E. coli* C41 and grown overnight (15–18 hr) in Studier's autoinduction media at 37 °C. Pellets were resuspended in breaking buffer (50 mM tris-Cl pH 8.0, 100 mM NaCl, 10 mM tris(2-carboxyethyl)phosphine (TCEP)) with 400 µg DNase, 2 mM MgCl₂, 1 mM PMSF, 1 mg/mL lysozyme, 25 µg pepstatin, and 500 µg leupeptin. Resuspended pellets were lysed by sonication and extracted with 2% n-Decyl-β-D-Maltopyranoside (DM) (Anatrace) for 2 hr at room temperature. Extract was clarified by centrifugation (16,000 rpm, 4 °C, 45 min), and loaded onto TALON cobalt resin equilibrated with wash buffer (20 mM tris-Cl pH 8.0, 100 mM NaCl, 5 mM DM) supplemented with 5 mM TCEP. Column was washed with wash buffer, and wash buffer supplemented with 10 mM imidazole before elution of EmrE₃ with wash buffer supplemented with 400 mM imidazole. After exchange into wash buffer using PD-10 desalting columns (GE Healthcare) His tags were cleaved with thrombin (1 U/mg EmrE₃) overnight at room temperature (21 °C) prior to a final size exclusion purification step using a Superdex 200 column equilibrated with 10 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) pH 7.5, 100 mM NaCl, 4 mM DM.

For functional measurements, protein was reconstituted by dialysis as previously described in Kermani et al., 2020. For SSM electrophysiology experiments, proteoliposomes

were prepared with 20 mg EPL per ml, and a 1:20 protein:lipid mass ratio. Proteoliposomes were aliquoted and stored at -80°C until use. For crystallography of EmrE₃, monobody L10 and EmrE₃ were each concentrated to 10 mg/mL, and the L10 protein solution was supplemented with 4 mM DM. EmrE₃ and L10 were combined in a 2.1:1 molar ratio and supplemented with lauryldimethylamine oxide (LDAO, final concentration of 6.6 mM). The protein solution was mixed with an equal volume of crystallization solution (0.3 μL in 96-well plates). Crystals formed after approximately 4 weeks, and were frozen in liquid nitrogen before data collection. For crystallization with substrate, the EmrE³/monobody/LDAO solution was prepared as before, and substrate was added from a stock solution immediately before setting crystal trays (final concentrations of 1 mM for methyl viologen, 500 μM for harmane, 300 μM for benzyltrimethylammonium, 100 μM for TPP⁺, or 300 μM for MeTPP⁺). The low pH EmrE₃ crystals grew in 200 mM NaCl, 100 mM sodium cacodylate, pH 5.2, 34% PEG 600. The substrate-bound EmrE₃ crystals grew in 100 mM LiNO₃ or 100 mM NH₄SO₄, 100 mM ADA, pH 6.5 or 100 mM HEPES, pH 7.1–7.3, and 30–35% PEG 600. Gdx-Clo protein and crystals were prepared exactly as described previously (Kermani et al., 2020). Crystals grew in 100 mM calcium acetate, 100 mM sodium acetate, pH 5.0, 40% PEG600.

3.5.3 Structure determination and analysis

Crystallography data was collected at the Life Sciences Collaborative Access Team beamline 21-ID-D at the Advanced Photon Source, Argonne National Laboratory. Diffraction data were processed and scaled using Mosflm 7.3 (Battye et al., 2011) or DIALS (Winter, 2018). Crystals diffracted anisotropically, and electron density maps were improved by anisotropic truncation of the unmerged data using the Staraniso webserver (Tickle et al., 2018; Tickle, 2018) with a cutoff level of 1.2–1.8 for the local $I/\sigma(I)$. For the low pH EmrE₃ dataset, phases were

determined using molecular replacement with Phaser (McCoy, 2007; McCoy et al., 2007), using the first three helices of Gdx-Clo and the L10 monobody structures (PDB:6WK8) as search models. Loop 3, helix 4, and the C-terminal loop were built into the experimental electron density using Coot (Emsley et al., 2010), with iterative rounds of refinement in Phenix (Liebschner et al., 2019) and Refmac (Murshudov et al., 2011). For the low pH Gdx-Clo structure, Gdx-Clo and the L10 monobody structures (PDB:6WK8) were used as molecular replacement search models. Models were validated using Molprobity (Williams et al., 2018) and by preparing composite omit maps in Phenix, omitting 5% of the model at a time (Terwilliger et al., 2008). The substrate-bound structures were phased using molecular replacement with monobody L10 and the A and B subunits of the initial EmrE₃ model as the search models. Proteins typically crystallized in C121, although the methyl viologen-bound EmrE₃ structure and the low pH Gdx-Clo crystallized in P1. For both, the unit cell contained two pseudosymmetric copies of the transportermonobody complex. The angle of the bend in TM3 was analyzed using Kink Finder (Wilman et al., 2014).

3.5.4 Microscale thermophoresis

Monobody L10 was labeled at a unique, introduced cysteine, A13C, with fluorescein maleimide. Binding to EmrE₃ was measured using microscale thermophoresis (Nanotemper, Munich, Germany). For these experiments, labeled monobody was held constant at 2 μ M, and the concentration of EmrE₃ was varied from 30 nM to 100 μ M. Buffer contained 100 mM NaCl, 10 mM HEPES, pH 7, 4 mM DM, and 50 μ g/mL bovine serum albumin. Samples were incubated at least 30 min prior to measurement of binding interactions. Experiments were performed using three independent sample preparations and fit to a one site binding equilibrium with total L10 as the experimental variable:

$$MST([EmrE]) = MST_0 + \frac{(MST_f - MST_0)}{2} \left(1 + \frac{[EmrE]}{[L10]} + \frac{K_D}{[L10]} \right) \left[1 - \sqrt{1 - \frac{4 \frac{[EmrE]}{[L10]}}{\left(1 + \frac{[EmrE]}{[L10]} + \frac{K_D}{[L10]} \right)^2}} \right]$$

where $MST([EmrE])$ is the MST signal as a function of total EmrE added to a fixed concentration of labelled L10 monobody, and MST_0 and MST_f are the arbitrary initial and final MST fluorescence signals.

3.5.5 SSM electrophysiology

SSM electrophysiology was conducted using a SURFE2 R N1 instrument (Nanon Technologies, Munich, Germany) according to published protocols (Bazzone and Barthmes, 2020; Bazzone et al., 2017). The sensor was alkylated and painted with lipid solution (7.5 µg/µl 1,2-diphytanoyl-sn-glycero-3-phos phocholine in n-decane), followed immediately by addition of recording buffer (100 mM KCl, 100 mM KPO4, pH 7.5). For measurements in the presence of monobody, buffers also contained 50 µg bovine serum albumin/mL. Proteoliposomes were applied to the sensor surface and centrifuged at 2500 x g for 30 min. Before experiments, sensors were checked for conductance and capacitance using SURFE2 R software protocols. Sensors for which capacitance and conductance measurements were outside an acceptable range (10–40 nF capacitance, 1–5 nS conductance) were not used for experiments. Sensors were periodically rechecked for quality during the course of an experiment. When multiple measurements were performed on a single sensor, currents elicited by a reference compound were measured at the outset of the experiment and again after collecting data on test compounds. If currents differed by more than 10% between the first and last perfusions, this indicated that the proteoliposomes associated with the sensor had not remained stable over the course of the experiment, and data collected in this series was discarded. Between measurements, sensors were perfused with

substrate-free solution for 2 s; observation of capacitive currents with opposite polarity indicated substrate efflux from the proteoliposomes and a return to the resting condition.

3.5.6 NMR chemical shift prediction

The chemical shifts of the Ca atoms of the NMR ensemble and the unliganded crystallography model were predicted using LARMORCa (Frank et al., 2015) as implemented with PyShifts (Xie et al., 2020).

3.6 Acknowledgements

We thank the Stockbridge lab for comments on the project and manuscript, and we are grateful to Aaron Frank (University of Michigan) for helpful conversations about chemical shift-based comparisons of the structures. Funding: This work was supported by NSF CAREER award 1845012 to RBS and R01 CA194864 to SK. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). RBS is a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Disease.

3.7 Competing Interests

Akiko Koide: is listed as inventor for patents (US9512199 B2 and related patents and applications) covering aspects of the monobody technology filed by the University of Chicago and Novartis. Shohei Koide: is listed as inventor for patents (US9512199 B2 and related patents

and applications) covering aspects of the monobody technology filed by the University of Chicago and Novartis. Is a scientific advisory board member and holds equity in and receives consulting fees from Black Diamond Therapeutics; receives research funding from Puretech Health and Argenx BVBA. Randy B Stockbridge: Reviewing editor, eLife. The other authors declare that no competing interests exist.

3.8 References

- Adam, Y., Tayer, N., Rotem, D., Schreiber, G., & Schuldiner, S. (2007). The fast release of sticky protons: kinetics of substrate binding and proton release in a multidrug transporter. *Proc Natl Acad Sci U S A*, 104(46), 17989-17994. <https://doi.org/10.1073/pnas.0704425104>
- Amadi, S. T., Koteiche, H. A., Mishra, S., & McHaourab, H. S. (2010). Structure, dynamics, and substrate-induced conformational changes of the multidrug transporter EmrE in liposomes. *J. Biol. Chem.*, 285. <https://doi.org/10.1074/jbc.M110.132621>
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*, 44(W1), W344-350. <https://doi.org/10.1093/nar/gkw408>
- Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr*, 67(Pt 4), 271-281. <https://doi.org/10.1107/s0907444910048675>
- Bazzone, A., Barthmes, M., & Fendler, K. (2017). SSM-based electrophysiology for transporter research. *Methods Enzymol.*, 594. <https://doi.org/10.1016/bs.mie.2017.05.008>
- Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., Casadio, R., & Ben-Tal, N. (2004). ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics*, 20(8), 1322-1324. <https://doi.org/10.1093/bioinformatics/bth070>
- Brill, S., Sade-Falk, O., Elbaz-Alon, Y., & Schuldiner, S. (2015). Specificity determinants in small multidrug transporters. *J. Mol. Biol.*, 427. <https://doi.org/10.1016/j.jmb.2014.11.015>
- Chen, Y. J. (2007). X-ray structure of EmrE supports dual topology model. *Proc. Natl Acad. Sci. USA*, 104. <https://doi.org/10.1073/pnas.0709387104>
- Dastvan, R., Fischer, A. W., Mishra, S., Meiler, J., & McHaourab, H. S. (2016). Protonation-dependent conformational dynamics of the multidrug transporter EmrE. *Proc. Natl Acad. Sci. USA*, 113. <https://doi.org/10.1073/pnas.1520431113>
- Dutka, P., Mukherjee, S., Gao, X., Kang, Y., de Waal, P. W., Wang, L., Zhuang, Y., Melcher, K., Zhang, C., Xu, H. E., & Kossiakoff, A. A. (2019). Development of "Plug and Play"

- Fiducial Marks for Structural Studies of GPCR Signaling Complexes by Single-Particle Cryo-EM. *Structure*, 27(12), 1862-1874.e1867. <https://doi.org/10.1016/j.str.2019.10.004>
- Elbaz, Y., Tayer, N., Steinfeld, E., Steiner-Mordoch, S., & Schuldiner, S. (2005). Substrate-induced tryptophan fluorescence changes in EmrE, the smallest ion-coupled multidrug transporter. *Biochemistry*, 44. <https://doi.org/10.1021/bi050356t>
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, 66. <https://doi.org/10.1107/S0907444910007493>
- Fleishman, S. J., Harrington, S. E., Enosh, A., Halperin, D., Tate, C. G., & Ben-Tal, N. (2006). Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *J Mol Biol*, 364(1), 54-67. <https://doi.org/10.1016/j.jmb.2006.08.072>
- Frank, A. T., Law, S. M., Ahlstrom, L. S., & Brooks, C. L., 3rd. (2015). Predicting protein backbone chemical shifts from C α coordinates: extracting high resolution experimental observables from low resolution models. *J Chem Theory Comput*, 11(1), 325-331. <https://doi.org/10.1021/ct5009125>
- Gillings, M. R. (2017). Class 1 integrons as invasive species. *Curr. Opin. Microbiol.*, 38. <https://doi.org/10.1016/j.mib.2017.03.002>
- Hinsen, K. (2000). The molecular modeling toolkit: a new approach to molecular simulations. *Journal of Computational Chemistry*, 21(2), 79-85.
- Jurasz, J., Baginski, M., Czub, J., & Wiczor, M. (2021). Molecular mechanism of proton-coupled ligand translocation by the bacterial efflux pump EmrE. *PLoS Comput Biol*, 17(10), e1009454. <https://doi.org/10.1371/journal.pcbi.1009454>
- Kermani, A. A., Macdonald, C. B., Burata, O. E., Ben Koff, B., Koide, A., Denbaum, E., Koide, S., & Stockbridge, R. B. (2020). The structural basis of promiscuity in small multidrug resistance transporters. *Nat Commun*, 11(1), 6064. <https://doi.org/10.1038/s41467-020-19820-8>
- Kermani, A. A., Macdonald, C. B., Gundepudi, R., & Stockbridge, R. B. (2018). Guanidinium export is the primal function of SMR family transporters. *Proc. Natl Acad. Sci. USA*, 115. <https://doi.org/10.1073/pnas.1719187115>
- Koide, A., Wojcik, J., Gilbreth, R. N., Hoey, R. J., & Koide, S. (2012). Teaching an old scaffold new tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. *J. Mol. Biol.*, 415. <https://doi.org/10.1016/j.jmb.2011.12.019>
- Koide, S. (2009). Engineering of recombinant crystallization chaperones. *Curr Opin Struct Biol*, 19(4), 449-457. <https://doi.org/10.1016/j.sbi.2009.04.008>
- Koldobskaya, Y., Duguid, E. M., Shechner, D. M., Suslov, N. B., Ye, J., Sidhu, S. S., Bartel, D. P., Koide, S., Kossiakoff, A. A., & Piccirilli, J. A. (2011). A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nat Struct Mol Biol*, 18(1), 100-106. <https://doi.org/10.1038/nsmb.1945>
- Li, J., Sae Her, A., & Traaseth, N. J. (2021). Asymmetric protonation of glutamate residues drives a preferred transport pathway in EmrE. *Proc Natl Acad Sci U S A*, 118(41). <https://doi.org/10.1073/pnas.2110790118>
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., . . . Adams, P. D. (2019). Macromolecular structure determination using

- X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*, 75(Pt 10), 861-877. <https://doi.org/10.1107/s2059798319011471>
- McCoy, A. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.*, 40. <https://doi.org/10.1107/S0021889807021206>
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr*, 40(Pt 4), 658-674. <https://doi.org/10.1107/s0021889807021206>
- McIlwain, B. C., Erwin, A. L., Davis, A. R., Ben Koff, B., Chang, L., Bylund, T., Chuang, G. Y., Kwong, P. D., Ohi, M. D., Lai, Y. T., & Stockbridge, R. B. (2021). N-terminal Transmembrane-Helix Epitope Tag for X-ray Crystallography and Electron Microscopy of Small Membrane Proteins. *J Mol Biol*, 433(16), 166909. <https://doi.org/10.1016/j.jmb.2021.166909>
- Morrison, E. A., DeKoster, G. T., Dutta, S., Vafabakhsh, R., Clarkson, M. W., Bahl, A., Kern, D., Ha, T., & Henzler-Wildman, K. A. (2011). Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature*, 481(7379), 45-50. <https://doi.org/10.1038/nature10703>
- Morrison, E. A., Robinson, A. E., Liu, Y., & Henzler-Wildman, K. A. (2015). Asymmetric protonation of EmrE. *J Gen Physiol*, 146(6), 445-461. <https://doi.org/10.1085/jgp.201511404>
- Mukherjee, S., Erramilli, S. K., Ammirati, M., Alvarez, F. J. D., Fennell, K. F., Purdy, M. D., Skrobek, B. M., Radziwon, K., Coukos, J., Kang, Y., Dutka, P., Gao, X., Qiu, X., Yeager, M., Eric Xu, H., Han, S., & Kossiakoff, A. A. (2020). Synthetic antibodies against BRIL as universal fiducial marks for single-particle cryoEM structure determination of membrane proteins. *Nat Commun*, 11(1), 1598. <https://doi.org/10.1038/s41467-020-15363-0>
- Mukherjee, S., Seshadri, R., Varghese, N. J., Eloë-Fadrosch, E. A., Meier-Kolthoff, J. P., Göker, M., Coates, R. C., Hadjithomas, M., Pavlopoulos, G. A., Paez-Espino, D., Yoshikuni, Y., Visel, A., Whitman, W. B., Garrity, G. M., Eisen, J. A., Hugenholtz, P., Pati, A., Ivanova, N. N., Woyke, T., . . . Kyrpides, N. C. (2017). 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat Biotechnol*, 35(7), 676-683. <https://doi.org/10.1038/nbt.3886>
- Nelson, J. W., Atilho, R. M., Sherlock, M. E., Stockbridge, R. B., & Breaker, R. R. (2017). Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Mol. Cell*, 65. <https://doi.org/10.1016/j.molcel.2016.11.019>
- Ovchinnikov, V., Stone, T. A., Deber, C. M., & Karplus, M. (2018). Structure of the EmrE multidrug transporter and its use for inhibitor peptide design. *Proc. Natl Acad. Sci. USA*, 115. <https://doi.org/10.1073/pnas.1802177115>
- Pal, C., Bengtsson-Palme, J., Kristiansson, E., & Larsson, D. G. (2015). Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics*, 16. <https://doi.org/10.1186/s12864-015-2153-5>
- Robinson, A. E., Thomas, N. E., Morrison, E. A., Balthazor, B. M., & Henzler-Wildman, K. A. (2017). New free-exchange model of EmrE transport. *Proc. Natl Acad. Sci. USA*, 114. <https://doi.org/10.1073/pnas.1708671114>
- Rotem, D., & Schuldiner, S. (2004). EmrE, a multidrug transporter from *Escherichia coli*, transports monovalent and divalent substrates with the same stoichiometry. *J Biol Chem*, 279(47), 48787-48793. <https://doi.org/10.1074/jbc.M408187200>

- Rotem, D., Steiner-Mordoch, S., & Schuldiner, S. (2006). Identification of tyrosine residues critical for the function of an ion-coupled multidrug transporter. *J. Biol. Chem.*, 281. <https://doi.org/10.1074/jbc.M602088200>
- Russell, A. D. (2002). Introduction of biocides into clinical practice and the impact on antibiotic-resistant bacteria. *J. Appl. Microbiol*, 92 Suppl. <https://doi.org/10.1046/j.1365-2672.92.5s1.12.x>
- Saleh, M., Bay, D. C., & Turner, R. J. (2018). Few Conserved Amino Acids in the Small Multidrug Resistance Transporter EmrE Influence Drug Polyselectivity. *Antimicrob Agents Chemother*, 62(8). <https://doi.org/10.1128/aac.00461-18>
- Sha, F., Salzman, G., Gupta, A., & Koide, S. (2017). Monobodies and other synthetic binding proteins for expanding protein science. *Protein Sci.*, 26. <https://doi.org/10.1002/pro.3148>
- Shcherbakov, A. A., Hisao, G., Mandala, V. S., Thomas, N. E., Soltani, M., Salter, E. A., Davis, J. H., Jr., Henzler-Wildman, K. A., & Hong, M. (2021). Structure and dynamics of the drug-bound bacterial transporter EmrE in lipid bilayers. *Nat Commun*, 12(1), 172. <https://doi.org/10.1038/s41467-020-20468-7>
- Soskine, M., Adam, Y., & Schuldiner, S. (2004). Direct evidence for substrate-induced proton release in detergent-solubilized EmrE, a multidrug transporter. *J Biol Chem*, 279(11), 9951-9955. <https://doi.org/10.1074/jbc.M312853200>
- Tate, C. G., Ubarretxena-Belandia, I., & Baldwin, J. M. (2003). Conformational changes in the multidrug transporter EmrE associated with substrate binding. *J Mol Biol*, 332(1), 229-242. [https://doi.org/10.1016/s0022-2836\(03\)00895-7](https://doi.org/10.1016/s0022-2836(03)00895-7)
- Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Adams, P. D., Read, R. J., Zwart, P. H., & Hung, L. W. (2008). Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr D Biol Crystallogr*, 64(Pt 5), 515-524. <https://doi.org/10.1107/s0907444908004319>
- Tickle, I., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonnrhein, C., & Bricogne, G. (2018). STARANISO Global Phasing Ltd. *Cambridge, United Kingdom*.
- Tickle, I. J. (2018). STARANISO <http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi> <http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>
- Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S., & Tate, C. G. (2003). Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J.*, 22. <https://doi.org/10.1093/emboj/cdg611>
- Vermaas, J. V., Rempe, S. B., & Tajkhorshid, E. (2018). Electrostatic lock in the transport cycle of the multidrug resistance transporter EmrE. *Proc. Natl Acad. Sci. USA*, 115. <https://doi.org/10.1073/pnas.1722399115>
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., 3rd, Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., & Richardson, D. C. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci*, 27(1), 293-315. <https://doi.org/10.1002/pro.3330>
- Wilman, H. R., Shi, J., & Deane, C. M. (2014). Helix kinks are equally prevalent in soluble and membrane proteins. *Proteins*, 82(9), 1960-1970. <https://doi.org/10.1002/prot.24550>
- Winter, G. (2018). DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D Struct. Biol.*, 74. <https://doi.org/10.1107/S2059798317017235>

- Wu, C. (2019). Identification of an alternating-access dynamics mutant of EmrE with impaired transport. *J. Mol. Biol.*, 431. <https://doi.org/10.1016/j.jmb.2019.05.035>
- Xie, J., Zhang, K., & Frank, A. T. (2020). PyShifts: A PyMOL Plugin for Chemical Shift-Based Analysis of Biomolecular Ensembles. *J Chem Inf Model*, 60(3), 1073-1078. <https://doi.org/10.1021/acs.jcim.9b01039>
- Yerushalmi, H., & Schuldiner, S. (2000). An essential glutamyl residue in EmrE, a multidrug antiporter from Escherichia coli. *J Biol Chem*, 275(8), 5264-5269. <https://doi.org/10.1074/jbc.275.8.5264>
- Zhu, Y. G. (2017). Continental-scale pollution of estuaries with antibiotic resistance genes. *Nat. Microbiol.*, 2. <https://doi.org/10.1038/nmicrobiol.2016.270>

Chapter 4 Necessary and Sufficient Molecular Determinants of Quaternary Ammonium Antiseptic Export by Small Multidrug Resistance Transporters

This chapter is a manuscript in progress. Current contributors to this work are: Olive E. Burata, Victoria E. O'Donnell, Jeonghoon Hyun, Junius E. Thomas, Nolan A. Carney, and Randy B. Stockbridge.

4.1 Introduction

Transport proteins are complex, but essential membrane-embedded macromolecules required for microbial homeostasis and overall survival. Although their native roles within a cell can be diverse, and their structural features and mechanisms can vary, the movement of solutes in and out of the cell through these proteins affects drug transport, nutrient acquisition, metabolism, cell signaling, and waste disposal (Adrián et al., 2015; César-Razquin et al., 2015; Davidson et al., 2008; Martinez et al., 2009; Pao et al., 1998). In addition to cellular housekeeping, transport proteins provide the first line of defense against toxic molecules ranging from antibiotics, antiseptics, solvents, heavy metals, and detergents (Martinez et al., 2009). Although these roles are evidence of the significance of membrane protein transporters to microbial health, it's essential to note that not all transporters are designed equally. Some are carefully calibrated to specific molecules, while others, recognize a wider array of different substrates (Kobayashi et al., 2001). This distinction can be important for survival in various environments, for example, selective transporters are useful in environments where specific

nutrients are limited and toxins are in abundance (Argüello, 2003; Fotiadis et al., 2013; Mueckler & Thorens, 2013), whereas promiscuous transporters have the evolutionary advantage to respond to unpredictable and changing environments (Alvarez-Ortega et al., 2013; Martinez et al., 2009; Piddock, 2006). This intrinsic adaptability brings forth medical challenges, since promiscuous transporters can efflux a wide range of toxic compounds, thus undermining our ability to treat microbial infections. The versatility of promiscuous transporters comes with its unique challenges and complexities. On several fronts having a promiscuous transporter could easily skew the balance between pumping out toxic molecules and keeping vital compounds (Sun et al., 2014) or introduce inefficient energy costs suggestive of an evolutionary trade-off, which would not otherwise be conserved and maintained in selective transporter systems (Rosenburg 2000). Additionally, promiscuous transporters may also introduce further intricacies in the molecular dynamics and mechanics of ligand interactions with the binding sites. These complex interactions were underscored very recently through structural analysis of exemplars from the major families of efflux pumps through a comprehensive review by Kim et al. (Kim et al., 2021), highlighting diverse mechanisms of polyspecificity within families. How do microbes manage such complexities to their advantage? Understanding the evolutionary processes behind polyspecificity and characterizing its molecular mechanisms can be insightful especially as these promiscuous transport systems exist across different organisms. Although other promiscuous transporter systems exist, the Small Multidrug Resistance family provides an efficient and simple model for promiscuous transport since the family itself harbors varied specificities that can be investigated at a fundamental level (Burata et al., 2022), and its evolutionary significance is continually investigated as the family continues to grow (Bay et al., 2008).

Small multidrug resistance (SMR) proteins are a family of bacterial antiporters that play an important role in expelling toxic compounds out of bacterial cells. It has been established that there are two major functional subtypes within the SMR family - guanidinium exporters (SMR_{Gdx}) and quaternary ammonium compound exporters (SMR_{Qac}). While SMR_{Gdx} selectively transport the bacterial metabolite guanidinium physiologically, SMR_{Qac} can transport a broad range of anthropogenic cationic compounds with disparate structures and binding affinities, such as quaternary ammonium antiseptics or polyaromatic antimicrobials. The best characterized representatives from each subtype are the SMR_{Qac} from *E. coli*, EmrE and the SMR_{Gdx} from *Clostridiales* oral taxon 876, Gdx-Clo.

Structural studies of these representatives show a highly similar three-dimensional architecture, including an antiparallel topology, the GXG motif that defines a molecular hinge, and the conserved central glutamates required for substrate and proton binding (Fleishman et al., 2006; Morrison et al., 2011; Muth & Schuldiner, 2000; Ubarretxena-Belandia et al., 2003). Differences that support the functional differences between EmrE and Gdx-Clo are more profound in the binding pocket, where conserved residues found in both Qacs and Gdxs are playing non-analogous roles. For example, Gdx-Clo has an exacting network of hydrogen bonds stabilizing the central glutamates established by W16, W62, and S42 (Kermani et al., 2020), and mutagenic removal of H-bonding potential has been shown to negate or severely impair overall function. The same cannot be said for EmrE's binding pocket. Most of the similar mutations made to undermine H-bonding potential are still allowed functionally. This unrestrained pocket has resulted in a bigger binding site to allow bulkier substrates and free reign rotation for the conserved residues E13 and W63, helping stabilize a diverse range of substrates (Kermani et al., 2022). Additionally, Gdx-Clo consistently followed a strict 2:1 H⁺:Gdm⁺ stoichiometry

(Kermani et al., 2018b) whereas previous studies have shown EmrE to be capable of deviating from this ideal stoichiometry which can lead to inefficient transport cycles or proton leakage (Gayen et al., 2016; Hussey et al., 2020; Morrison et al., 2011; Robinson et al., 2017).

Despite the many mutational and functional studies, we still don't have a clear understanding of the critical residues responsible for the polyspecific phenotype of the SMR_{Qacs}, or why their selectivity differs from the SMR_{Gdx} despite high overall sequence & structural conservation, especially conservation of the binding site residues. Being the most well-studied member of the SMR family, a number of scanning mutagenesis studies have been performed on EmrE. A majority of these investigations have tested standard substrates that are bulky, hydrophobic, and polyaromatic, and most have been assessed *in vivo* (Amadi et al., 2010; Elbaz et al., 2008; Elbaz et al., 2005; Gutman et al., 2003; Mordoch et al., 1999; Rotem et al., 2006; Saleh et al., 2018; Wu, 2019; Yerushalmi & Schuldiner, 2000). These studies have consistently flagged E13 as a crucial residue for overall drug resistance, unable to tolerate any other changes at this position without diminished function. Other positions have had hints of importance regardless of their proximity to the substrate binding pocket, impairment to resistance to a specific of hydrophobic, polyaromatic compounds but are not necessarily the entire class. Although the comprehensive look into EmrE's mutagenic landscape offers invaluable data, a general agreement of the blueprint behind substrate polyspecificity remains elusive. Many of these studies have focused on single site mutations, ignoring the likelihood that a concert of mutations could be responsible for the expanded specificity. There was only ever one attempt that addressed this potential concern, which was a study by Brill *et. al.*, 2012 where random mutations were introduced on EmrE resulting in the design of a polyamine importer through a single mutation (Brill et al., 2012). Nonetheless, the generated library was only screened to non-

canonical EmrE substrates with analogous chemical properties, which fail to provide a holistic overview of the molecular determinants of promiscuity in the SMR family.

In this chapter, we address the gap in knowledge by utilizing both rational and directed evolution to engineer a Gdx-Clo variant capable of transporting structurally distinct quaternary ammonium antiseptics. In addition to assessing the drug resistance capabilities of this new variant, we also aim to provide a biochemical and biophysical approach to quantifying its function through novel techniques, a layer of characterization that has long eluded earlier studies on membrane transporters with hydrophobic substrates. Gdx-Clo and EmrE will be guiding blueprints as we explore the combination of mutations that are necessary switches required to design a promiscuous transporter in the SMR family.

4.2 Results

4.2.1 Engineering Gdx-Clo for Resistance Against Quaternary Ammonium Antiseptics

Analysis of sequence alignments of the SMR_{Qac} and SMR_{Gdx} transporters (Burata et al., 2022) revealed six residues that are highly conserved in each subtype, but that differ between the subtypes. These residues were, for Gdx-Clo and EmrE, respectively, Gly10/Ile11, Trp16/Gly17, Ala17/Thr18, Met39/Tyr40, Ala67/Ile68, and Lys101/Asn102. We reasoned that these residues might be mainly responsible for the differing substrate selectivity of the SMR_{Qac} and SMR_{Gdx} subtypes. We therefore introduced these mutations to all six positions in Gdx-Clo (G10I, W16G, A17T, M39Y, A67I, and K101N), but the resulting transporter failed to confer bacterial resistance to quaternary ammonium antiseptics.

To identify additional mutations that might furnish quaternary ammonium resistance and substrate polyselectivity, we further subjected this mutated Gdx-Clo construct to random

mutagenesis followed by selection for bacterial growth in the presence of two different quaternary ammoniums, cetrimonium and tetrapropylammonium. These substrates were chosen to select for promiscuous transporters. The substrates differ structurally in the bulkiness of the quaternary ammonium group, and only cetrimonium possesses a long alkyl tail. Moreover, these two quaternary ammoniums also exhibit widely divergent inhibitory concentrations ($\sim 200\ \mu\text{M}$ and $\sim 20\ \text{mM}$, respectively), demanding export efficiency over ~ 2 orders of magnitude concentration. Under these selection conditions, one round of directed evolution yielded a construct that possessed the six original rationally designed mutations and an additional 7th mutation, A60T, that supported robust growth of ΔemrE *E. coli* in the presence of cetrimonium and tetrapropylammonium. The resistance conferred by this construct, dubbed Gdx-Clo-7X, is comparable to that conferred by the native EmrE introduced on a rescue plasmid (**Figure 4.1A**). All seven individual mutants, including the mutation identified via directed evolution, A60T, were, by themselves, insufficient for quaternary ammonium antiseptic resistance (**Supplementary Figure C.2**). Moreover, directed evolution on a wildtype Gdx-Clo background also did not yield variants with quaternary ammonium resistance (data not shown).

The seven mutations of Gdx-Clo-7X identified are distributed across the protein in three structurally co-localized clusters (**Figure 4.1B**). The first, G10I (red), occurs at a helical packing motif between helices 1 and 3. These helices contribute several residues to the binding site hydrogen bond network, suggesting that G10I mutation will alter the packing of these two helices and disrupt the hydrogen bond network, which has previously been proposed to contribute to substrate polyspecificity. The second mutant cluster is comprised of W16G, A17T, and M40Y (orange) adjacent to the substrate binding site. W16 is the only one of the seven mutant residues that directly participates in the WT Gdx-Clo hydrogen bond network, where it

directly coordinates the substrate-binding central glutamate, E13. A17 and M40 sandwich W16 in place. A60T, A67I, K101N (green) comprise the third cluster at the periphery of the aqueous binding pocket. These residues undergo large changes in position during the inward- to outward-facing conformational change, and interpolation of the inward- and outward-facing structures suggests that they pass in close proximity.

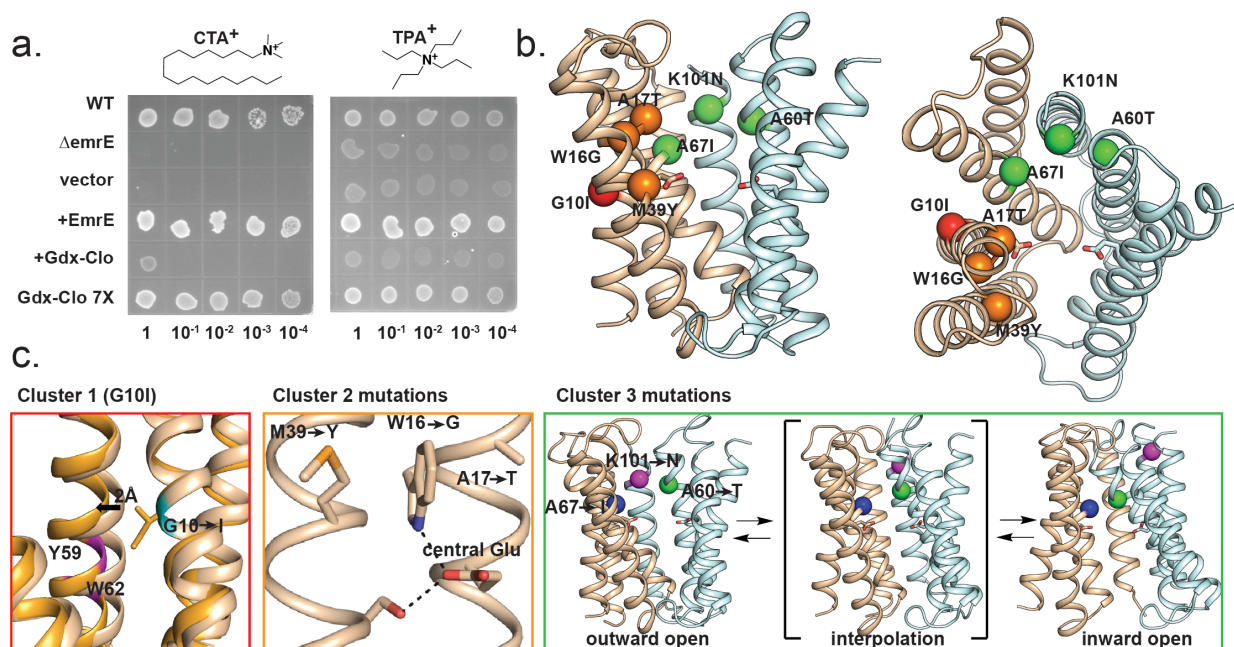


Figure 4.1: An engineered variant of Gdx-Clo confers bacterial resistance to quaternary ammonium antiseptics.

A Gdx-Clo 7X in pBAD24 vector exhibiting resistance to 120 μ M CTA⁺ and 18 mM TPA⁺. **B** Mutations in Gdx-Clo 7X mapped out on Gdx-Clo WT (PDB: 6WK8). Mutations are separated into three clusters: clusters 1 (G10I, red orbs), cluster 2 (W16G, A17T, and M39Y, orange orbs), cluster 3 (A60T, A67I, and K101N, green orbs). **C** schematics of potential role of the individual clusters.

4.2.2 Gdx-Clo-7X gains the ability to bind and transport quaternary ammoniums

To quantify the binding and transport of quaternary ammonium antiseptics by the Gdx-Clo-7X variant, we overexpressed, purified, and reconstituted this transporter into proteoliposomes (**Supplementary Fig. C.3**). We assessed binding by exploiting the proteins' intrinsic tryptophan fluorescence, which increases upon substrate binding (Nelson et al., 2017)

(Figure 4.2). For TPA^+ and Gdm^+ , the seven mutations invert substrate preference: Gdx-Clo WT binds Gdm^+ (K_d 800 μM) but not TPA^+ , whereas Gdx-Clo-7X exhibits TPA^+ binding (K_d 15 mM, in line with the MIC in bacterial culture) but greatly impaired Gdm^+ binding (>1 mM). Both substrates bound to cetrimonium with comparable affinities (7.2 μM for WT Gdx-Clo and 4.7 μM for Gdx-Clo-7X). We determined a crystal structure of cetrimonium bound to WT Gdx-Clo, which showed the small methylated headgroup located near the central glutamates in the binding site, with the alkyl tail extending out of the binding site into the membrane, in the manner observed for other such substrates with long alkyl tails.

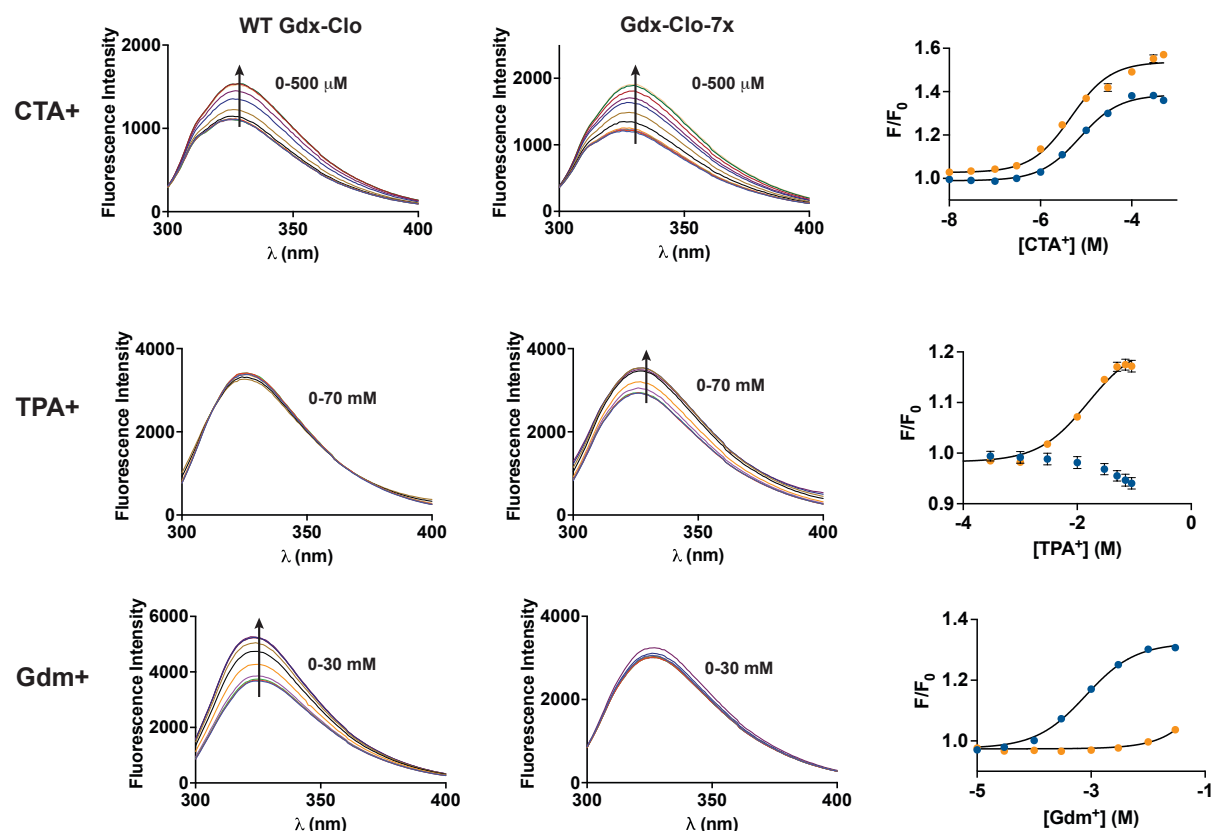


Figure 4.2: In vitro analysis shows that the substrate binding profile has changed for Gdx-Clo 7X.

Trp fluorescence measured from titration of Gdm^+ , TPA^+ , or CTA^+ with Gdx-Clo WT and Gdx-Clo 7X in micelles. Apparent K_d values were obtained by fitting into single-site binding isotherm equation found in Materials and Methods section.

We also assessed transport of TPA^+ by Gdx-Clo-7X using solid-supported membrane electrophysiology (SSME). Due to its hydrophobicity, TPA^+ elicits large, positive currents that reflect protein-independent interactions between the cationic substrate and the membrane. However, for Gdx-Clo-7X, titration with mM concentrations of TPA^+ also yielded small negative capacitive currents that evolve more slowly than the TPA^+ binding currents. These currents are consistent with electrogenic proton/ TPA^+ antiport (**Figure 4.3A**). Peak currents fit to the Michaelis-Menten equation with a K_m value of 2 mM (**Figure 4.3B**). For comparison, EmrE transports TPA^+ with a K_m value of $\sim 800 \mu\text{M}$. Gdm^+ currents were absent in the Gdx-Clo-7X mutant, but, as in EmrE, increasingly hydrophobic substitutions of the guanidinium moiety restored increasingly robust transport currents (**Figure 4.3C**).

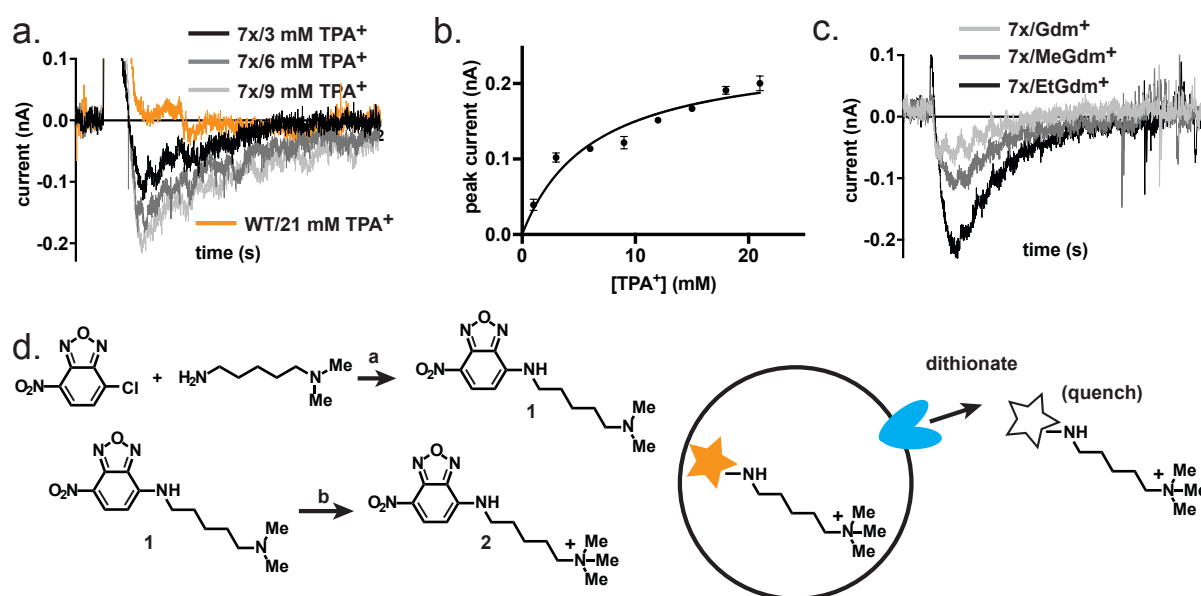


Figure 4.3: Gdx-Clo 7X transports quaternary ammonium compounds.

A SSME current traces of Gdx-Clo 7X when titrated with TPA^+ , and Gdx-Clo WT with 21 mM TPA^+ over 2 seconds. **B** Current amplitudes from titration of TPA^+ in panel A fitted in a Michaelis-Menten curve ($K_m \sim 2\text{--}3 \text{ mM TPA}^+$). **C** SSME current traces of Gdx-Clo 7X screened with 5 mM Gdm^+ , methyl Gdm^+ , and ethyl Gdm^+ over 2 seconds. **D** (left) Synthetic scheme of NBD-CTA⁺(5) synthesis, (right) schematic of transport assay using NBD-CTA⁺(5). 500 μM NBD-CTA⁺(5) is internalized into proteoliposomes by gradual freeze/thaw and subsequently extruded into a 40 mm membrane. Transport is initiated upon 200-fold dilution into assay buffer containing 5 mM sodium dithionite, 100 mM KCl, 100 mM K_2HPO_4 , pH 7.52. Reaction is ran for 200 seconds and terminated using 0.1% TritonX-100. **E** NBD-CTA⁺(5) fluorescence trace measured for EmrE, Gdx-Clo WT, empty and Gdx-Clo 7X proteoliposomes.

We were unable to use SSME to evaluate cetrimonium transport due to its hydrophobicity. To assess cetrimonium transport, we leveraged our structural observation that substrates with alkyl tails bind the SMR transporters such that the alkyl tail extends out through a lateral portal in the protein and into the membrane (Kermani et al., 2022; Kermani et al., 2020). This allowed us to adapt an approach used to study lipid transport by flippases, which bind substrate lipids in a similar manner (Ploier & Menon, 2016). We chemically synthesized a novel substrate analog that has a fluorophore, nitrobenzoxadiazole (NBD), conjugated to the aliphatic 5-carbon tail with a trimethylammonium headgroup (**Figure 4.3D, Supplementary Figure C.6 and C.7**). We internalized NBD-CTA⁺(5) in Gdx-Clo-7X proteoliposomes and initiated transport by diluting the liposomes into an external buffer with dithionate to quench NBD fluorescence. For positive control EmrE, NBD-CTA⁺(5) fluorescence decreases by 80% over ~200 seconds, reflecting protein-mediated export and quenching of NBD-CTA⁺. Addition of 20% triton solubilizes remaining lipid vesicles, permitting complete quenching of encapsulated NBD. In contrast, protein-free liposomes retain a steady fluorescence signal until addition of triton, as NBD-CTA⁺ remains protected from quenching. As anticipated from the resistance assays, WT Gdx-Clo does not transport NBD-CTA⁺. However, Gdx-Clo-7X exhibits a slow transport rate that increases with increasing reconstituted protein concentrations. Transport of the same substrate is accomplished by Gdx-Clo 7X, but at a slower rate than EmrE (**Figure 4.3E**). The present experiments show that Gdx-Clo-7X contributes to bacterial resistance to TPA⁺ and cetrimonium by facilitating active, proton-coupled export of these substrates, in contrast with the guanidinium-selective transport behavior of WT Gdx-Clo.

4.2.3 Screening a combinatorial library to identify necessary and sufficient mutations for quaternary ammonium resistance

To assess whether all seven mutations were essential for the engineered quaternary ammonium resistance activity, or whether a subset of these mutations would be sufficient for this activity, we constructed a combinatorial library of all 128 possible variants of the seven mutations using a modified Gibson assembly method (**Figure 4.4A, Table 8**). We verified the presence of all variants at approximately equimolar ratios using next generation sequencing (NGS). This library was transformed into $\Delta emrE$ cells (at 16.7 pg pDNA:1 cell ratio that ensure that each cell possesses one or fewer plasmids), followed by selection with cetrimonium or TPA⁺, and identification of surviving variants using NGS. The selection was not performed as a competitive growth in liquid culture; rather we plated the bacteria to isolate any colonies that showed resistance to cetrimonium.

In contrast to the starting library shown in **Figure 4.4B**, selection with TPA⁺ or CTA⁺ eliminates growth of most variants. The surviving variants possess most of the originally identified seven mutations. The most frequent variant isolated from the cetrimonium plates retains 6 of the 7 mutations (only A67I is absent), and the most frequent variant from the TPA⁺ screen possesses all 7. For TPA⁺ resistance (**Figure 4.4C**), variants lacking A67I, K101N, or both A67I and K101N were also identified at levels >30-fold higher than the original library. For CTA⁺ resistance (**Figure 4.4D**), variants that lacked any one, two, or even all three cluster 3 mutations were identified at levels >100-fold more abundant than in the original library. Thus, these experiments suggest that all the cluster 1 and 2 mutations are essential to confer quaternary ammonium resistance, and that the cluster 3 mutations are, in general, less critical.

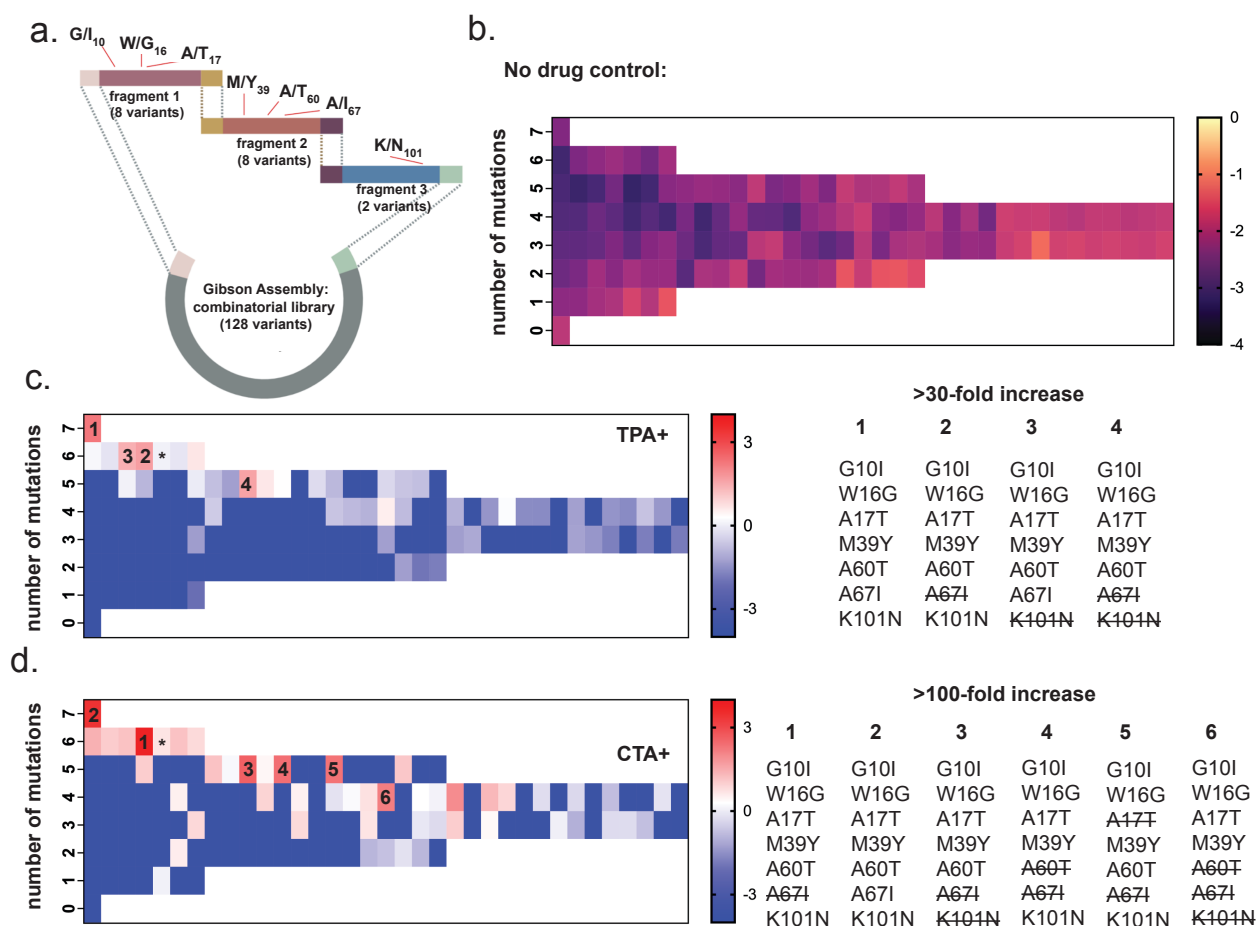


Figure 4.4: Necessary and sufficient mutants identified from combinatorial library.

A Design of oligo fragments inserted in a pBAD24 vector using a modified Gibson assembly method of all 128 variants of Gdx-Clo 7X combinatorial library. **B** NGS analysis of Gdx-Clo 7X combinatorial library screen for growth in no drug LB agar. **C,D** Gdx-Clo 7X library screened for growth in 120 μ M CTA⁺ and 18 mM TPA⁺. Variants from each screen listed out depending on how well they exhibited resistance (>30-fold for TPA⁺, >100-fold for CTA⁺)

4.2.4 Cluster 1 and 2 mutations are essential for TPA⁺ binding and quaternary ammonium transport, whereas Cluster 3 mutations contribute to faster export kinetics

To dissect the contribution of the three mutational clusters to substrate binding and export, we introduced mutations from these clusters alone and in combination, guided by our NGS analysis. Tryptophan fluorescence quenching experiments showed that Gdm⁺ binding was

severely impaired by either the cluster 1 (unobservable) or cluster 2 ($K_d \sim 10$ mM) mutations. The cluster 3 mutations resulted in a >10 -fold decrease in Gdm^+ K_m compared to Gdx-Clo (**Figure 4.5A**), perhaps due to the loss of electrostatic repulsion from K101. However, tight substrate binding is not necessarily desirable for a transporter, since proton antiport involves well-calibrated competition between substrate and proton binding to the same central glutamates. Indeed, electrogenic Gdm^+ transport is lost for all three clusters (**Figure 4.5B**). The cluster 2 transporter retains the ability to transport methyl-, ethyl- and phenyl Gdm^+ , recapitulating the behavior of Gdx-Clo-7X with the full complement of mutants, and of EmrE.

None of the clusters, introduced individually, permitted TPA^+ binding. However, cluster 1 and 2, introduced together, were sufficient for TPA^+ binding (**Figure 5C**). Both clusters impact the binding pocket size and hydrogen bond network. We were unable to measure TPA^+ transport in our SSME assay by mutational combinations other than Gdx-Clo-7X. The NGS analysis suggests that this is the best variant for TPA^+ export, and TPA^+ transport is only just above the threshold for detection by SSME. We expect that other surviving variants might contribute to low levels of TPA^+ resistance yet not meet the threshold for detection.

In contrast, CTA^+ , with a small headgroup, is able to bind even to WT Gdx-Clo. However, our NGS experiments show that this same complement of cluster 1 and cluster 2 mutations are essential for CTA^+ resistance. Reconstituted transport assays recapitulate the NGS results: individually the clusters do not permit NBD- CTA^+ transport. However, transport is detected for cluster 1 and 2, and the further addition of cluster 3 mutations (Gdx-Clo-5X, Gdx-Clo-6X, and Gdx-Clo-7X) increases the rate of NBD- CTA^+ export. Combining the results of the NGS analysis and the biochemical analysis, we find that cluster 1 and cluster 2 mutations are sufficient to introduce polyselective quaternary ammonium transport to the selective Gdx-Clo

WT background, but that the introduction of additional mutations from cluster 3 increases the rate of quaternary ammonium export, contributing to increased fitness of the bacteria.

We screened additional substrates (phenylGdm⁺ and TPA⁺), none of which are transported by any of the clusters with the exception of cluster 2, which showed transport currents for ethylGdm⁺ and increasing currents for a more hydrophobic substrate, phenylGdm⁺ (**Figure 4.5B**). Although the clusters failed to elicit any resistance to CTA⁺ *in vivo* nor transport NBD-CTA⁺(5), all were able to still bind to CTA⁺ at WT levels, except for cluster 1, like with Gdm⁺, lost its ability to bind to the Qac as well. However, when clusters 1 and 2 (named Gdx-Clo 4X) were combined, transport of NBD-CTA⁺(5) was rescued. TPA⁺ binding with the individual clusters were expected to be nonexistent, just based on lack of resistance against the drug in resistance assays. However, the same Gdx-Clo 4X was endowed with the ability to bind TPA⁺, but we were unable to measure ability to transport TPA⁺ via SSME.

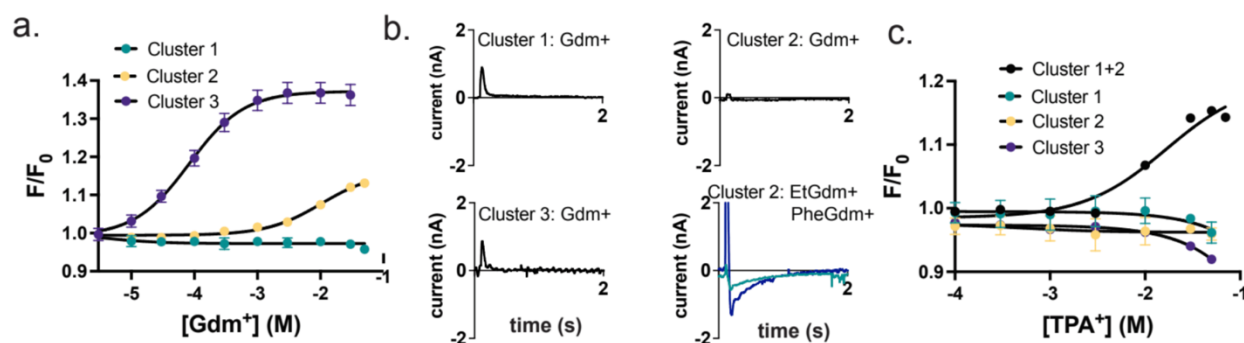


Figure 4.5: Functional characterization of the three variant clusters.

A Trp fluorescence spectra of individual clusters measured with titrated Gdm⁺. **B** (top, and bottom left panels) SSME current traces of individual clusters when perfused with Gdm⁺, (bottom right panel) and cluster 2 when perfused with ethylGdm⁺ and PheGdm⁺. **C** Trp fluorescence spectra of individual clusters measured with titrated TPA⁺.

4.3 Discussion

Promiscuous transporters have consistently been the leading edge of recent scientific inquiry, specifically in the SMR family, due to their unique ability transport a broad range of substrates and the challenges they present in the medical field. To better understand how promiscuous transporters are designed, we engineered Gdx-Clo 7X through unbiased means, guided by structural data from Gdx-Clo and EmrE, exemplars from of the two dominant subtypes in the SMR family. Gdx-Clo 7X which mostly has a rudimentary structural feature of a Gdx, has been functionally characterized to transport quaternary ammonium compounds, substrates only Qacs can transport. Four out of the seven mutations (G10I, W16G, A17T, M39Y, A60T, A67I, and K101N) have contributed in ways that were initially predicted after comparisons between the crystal structures of Gdx-Clo and EmrE: an expanded binding pocket (G10I) and destabilization of the H-bond network (W16G, A17T, and M39Y). The enhancement of the transport rate for Qacs made by possible by cluster 3 (A60T, A67I, and K101N) provided a novel piece of molecular insight that completed this specificity change. This further shows that hardly a single mutation was adequate to facilitate a change and that despite the small size of the SMRs, a concert of mutations was necessary to change specificity. Additionally, these 7 mutations only provided half the story as they were unable to provide resistance to polyaromatic substrates typically characterized with EmrE. These missing molecular details might also account for the suboptimal activity of Gdx-Clo 7X relative to EmrE, and that once these additional key residues are discovered is when we can observe further expansion of the substrate specificity to match that of EmrE's capabilities.

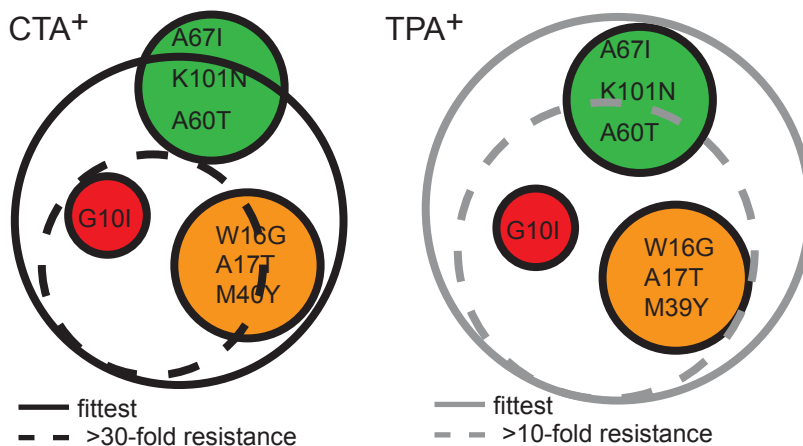


Figure 4.6: Fitness schematic of Gdx-Clo 7X combinatorial variants in the presence of QACs.

Resistance profiles of Gdx-Clo 7X separated by their respective cluster orbs. Solid circle indicates the clusters with the highest resistance, dash circle indicates required level of resistance (>30-fold for CTA⁺, and >10-fold for TPA⁺).

Most of these residues (Gly10/Ile11, Trp16/Gly17, Ala17/Thr18, Met39/Tyr40) are localized in the binding pocket and are known to be implicated in previous binding and resistance assays (Gutman et al., 2003; Mordoch et al., 1999; Sharoni et al., 2005; Wu, 2019). Although Gly10 and Ile11 are positioned in the binding pocket of Gdx-Clo and EmrE, respectively, they are not involved in the H-bond network in either protein. Gdx-Clo Ala67 and EmrE Ile68 may simply play a role in the integrity of helical packing as mutations in EmrE Ile68 to smaller non-polar residues (Val, Ala, and Gly) did not result in any impairment of resistance to toxic polyaromatic compounds such as TPP⁺, Methyl Viologen, Ethidium, or Acriflavine (Wu, 2019). It is important to note that Ala67/Ile68 residues are within the vicinity of the GVG fulcrum, a highly conserved region across the SMRs that acts as a hinge during exchanges between inward and outward conformations. The only set of polar residues not involved in the first shell H-bonding network of the binding pocket in either protein is the Lys101/Asn102 pair. Based on current high-resolution structural data for both Gdx-Clo and EmrE, this pair exhibits similar behavior in either protein, that it never occupies the same space as Tyr60 (Tyr59 in Gdx-

Clo). This tyrosine is conserved across the SMR proteins and is known to stabilize the binding site and regulate the conformational changes during the transport cycle (Kermani et al., 2022; Rotem & Schuldiner, 2004; Rotem et al., 2006; Wu, 2019). Although the role of the Lys101/Asn102 pair is unclear in their respective proteins, mutations at this site in EmrE to small, nonpolar residues does negatively impact resistance to some toxic polyaromatic compounds (Wu, 2019). What is clear based on the NGS analysis from **Figure 4.4**, is that a fundamental incorporation of cluster 1 and 2 is required for proper transport and binding to Qacs. The intermittent presence of cluster 3 mutations isolates their secondary importance in increasing the likelihood of efficiently completing the transport cycle, but not entirely negate the possibility of transport (**Figure 4.6**). Discovering this interplay between the 7X residues at the interface of a selective and promiscuous transporter is a pivotal breakthrough upon the nuances of the two dominant subtypes in the SMR family. It underscores the importance of the molecular determinants and their roles in substrate preference, but it also provides the groundwork for further protein engineering beyond this family of simple transporters.

Understanding how polyspecific transport is possible in the SMR family of transporters is important for several reasons. Firstly, Qacs play a key role in bacterial resistance to antibiotics, which continues to be a major global health concern. The ability of these transporters to efflux a wide range of structurally diverse substrates greatly contributes to the ability of bacterial cells to resist a variety of biocides, leading to multidrug resistance. Understanding how Qacs can recognize and transport such a diverse range of compounds is therefore pivotal in the development of new strategies to combat antimicrobial resistance. Secondly, the mechanisms underlying polyspecific transport in Qacs may also have broader implications for understanding the evolution of substrate specificity in other transporter families. By gaining insight into the

structural and biochemical features that enable Qacs to transport such a diverse range of substrates, we may be able to identify similar features in other transporters that exhibit similar substrate promiscuity, and therefore gain a better understanding of the evolution of transporter specificity. Finally, understanding the mechanisms of polyspecific transport in SMRs has implications for biotechnological applications. SMRs have the potential to be used as biotechnological tools for the removal of toxic compounds from industrial waste and environmental pollutants. However, to do so, we need to understand how to engineer SMRs to recognize and transport specific substrates. By understanding the underlying mechanisms of substrate recognition and transport in SMRs, we can develop more effective strategies for engineering these transporters for bioremediation and bioaugmentation applications.

4.4 Materials and Methods

4.4.1 Quaternary Ammonium Resistance Assays

Genes encoding SMR transporters were inserted into the pBAD24 vector. LB agar plates were prepared with the addition of 70 mM K₂HPO₄, pH 7.2 for buffering, 0.2% arabinose to induce protein expression, 50 µg/mL kanamycin, and 100 µg/mL carbenicillin. Overnight cultures (12-16 hrs) of $\Delta emrE$ *E. coli* (Keio collection, Coli Genetic Stock Center, New Haven, CT) bearing the appropriate plasmid were diluted to OD₆₀₀ of 0.05 with LB media containing 13.3 mM arabinose and antibiotic, then grown to OD₆₀₀ 0.5 – 0.8 (37°C, 240 rpm). 10-fold serial dilutions were spotted onto plates containing 120 µM cetrimonium bromide, 18 mM tetrapropylammonium chloride, or no quaternary ammonium and examined for growth after 24-48 hours, as indicated in figure legends.

4.4.2 Directed evolution

Directed evolution was performed on a background construct of Gdx-Clo with G10I, W16G, A17T, M39Y, A67I, K101N mutations using the GeneMorph II EZClone Domain Mutagenesis Kit (Stratagene) with the manufacturer's protocol and the following primers:

Forward Primer: 5'-CAGGAGGAATTCACCATGGCGTGGCTGATC-3'

Reverse Primer: 5- ACAGCCAAGCTTATTAGCTGCTGGTCGCTTT-3'

For the synthesis of the library megaprimers, 400 ng of plasmid DNA and 125 ng of each primer were used for the reaction. 250 ng of the megaprimer reaction and 50 ng of the template plasmid in pBAD24 vector was used for the EZ-Clone reaction. The library was then transformed into high-efficiency, electrocompetent DH5 α cells and the transformants were collected to prepare a stock of the library plasmid DNA. For screenings, the purified library was then transformed into BW25113 $\Delta emrE$ cell lines via electroporation and half of the recovery culture was used to inoculate a 25 mL overnight growth. A 5 mL culture with a starting OD₆₀₀ of 0.05 was made from the overnight growth and expression of the library was induced with 0.2% arabinose until OD of 0.5. Serial dilutions were plated on buffered LB plates containing 200 μ M-300 μ M CTA⁺, or 20 mM-30 mM TPA⁺, or no drug. Individual colonies were harvested, and plasmid DNA was isolated and sequenced. For resistance assays, pDNA was transformed into chemically competent $\Delta emrE$ *E. coli* for resistance assays as described earlier.

4.4.3 Combinatorial Library Construction, Selection Assays, and Illumina Sequencing

Gene blocks (Supplementary Table 1) were synthesized by Azenta (Chelmsford, MA) and prepared as a pooled master mix (10 ng/ μ L for each fragment). Fragments were assembled into a pBAD24 vector using HiFi DNA Assembly Kit (New England Biolabs, Ipswich, MA) with 1-hour, 50° C incubation and 0.2 pmol total fragments. After transformation

(electrocompetent NEB 10-beta cells), half of the recovery culture was plated to evaluate transformation efficiency (LB agar plates with carbenicillin), and the remaining half used to inoculate a 25-mL overnight culture (LB with carbenicillin) for purification of the combinatorial plasmid library (QIAprep Spin Miniprep Kit, Qiagen, Germantown, MD). The assembled combinatorial library was transformed into $\Delta emrE$ *E. coli* via electroporation. Half of the recovery culture was plated to evaluate transformation efficiency, and the other half was prepared as for resistance assays described above. 100 μ L aliquots of each serial dilution were plated on LB plates with 120 μ M cetrimonium bromide or “no drug” plates, without additional antibiotics. The 1:1000 dilution yielded plates with a maximum number of isolated colonies (~500 colonies). The entire plate was resuspended, and plasmids were mini-prepped. Illumina adapters were added via PCR using the following primers:

Forward Illumina Adapter: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTXXXX-3'

Reverse Illumina Adapter: 5'-GACTGGAGTTCAGACGTGTGCTCTTCCGATCTXXXX-3'

Next generation (Illumina) sequencing was performed by Azenta (Amplicon-EZ). Analysis of reads processed by Azenta included removal of reads <10 to avoid non-specific readings. All possible mutagenic combinations in the library were accounted for manually as the highest set of reads. Total reads were determined from all the mutagenic combinations only. Three biological replicates were used to account for all the variants in the “no drug” plates and three biological replicates were used for the 120 μ M cetrimonium bromide and 18 mM tetrapropyl screenings.

4.4.4 Protein Purification and Liposome Reconstitution

Proteins were expressed and purified as previously described (Kermani et al., 2020; Kermani et al., 2018a). Briefly, hexahistidine-tagged proteins were overexpressed in *E. coli*,

extracted from *E. coli* membranes with 2% n-decyl- β -D-maltoside (DM) and purified using cobalt affinity resin and size exclusion chromatography. Purified protein was reconstituted into proteoliposomes (10 mg *E. coli* polar lipid extract, Avanti Polar Lipids, Alabaster, AL) by dialysis in 100 mM KCl, 100 mM K₂HPO₄, pH 7.52 as described (Kermani et al., 2020; Kermani et al., 2018a). For the NBD-CTA⁺ transport assay, proteoliposomes were prepared with a protein:lipid ratio of 0.5 μ g /mg. For SSM electrophysiology experiments, proteoliposomes were prepared with a protein: lipid ratio of 40 μ g/mg. Proteoliposomes were stored at -80° C until use.

4.4.5 Synthesis of NBD-CTA⁺(5)

All commercial reagents and solvents were used as supplied without further purification. Proton nuclear magnetic resonance (¹H NMR) and carbon (¹³C) NMR spectroscopy were performed on Bruker Advance 400 NMR spectrometers. ¹H NMR spectra are reported in parts per million (ppm) downfield from tetramethylsilane (TMS). All ¹³C NMR spectra are reported in ppm and obtained with ¹H decoupling. In the spectral data reported, the format (δ) chemical shift (multiplicity, J values in Hz, integration) was used with the following abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet. Electrospray ionization (ESI) mass spectral (MS) analysis was performed on a Thermo Scientific LCQ Fleet mass spectrometer. The final products were purified by reverse phase HPLC (RP-HPLC) with solvent A (0.1% of TFA in water) and solvent B (0.1% of TFA in CH₃CN) as eluents with a flow rate of 45 mL/min. All final compounds have purity \geq 95% as determined by Waters ACQUITY ultra-performance liquid chromatograph (UPLC) using reverse phase column (SunFire, C18, 5 μ m, 4.6 \times 150 mm²) and a gradient of solvent A (H₂O with 0.1% of TFA) and solvent B (CH₃CN with 0.1% of TFA).

N¹,N¹-dimethylpentane-1,5-diamine (1): A solution of NBD-Cl (100mg, 0.5 mmol) in 0.5ml of anhydrous DMF was added dropwise in 5 min under a nitrogen atmosphere at room

temperature, to a stirred solution of (5-aminopentyl) dimethylamine dihydrochloride (0.5 mmol) and triethylamine (154 μ l, 1 mmol) in 1ml of anhydrous DMF. The reaction mixture was purified by reverse HPLC using the gradient of 10%_50% B in 40 min_60ml. Product eluted at 26% B and the fractions were combined and lyophilized to produce compound 1 (70mg) as a mahogany-colored solid. ^1H NMR (400 MHz, DMSO- d_6) δ 9.55 (s, 1H), 8.51 (d, J = 9.0 Hz, 1H), 6.44 (d, J = 9.0 Hz, 1H), 3.49 (d, J = 7.6 Hz, 2H), 3.03 – 2.96 (m, 2H), 2.72 (d, J = 4.9 Hz, 6H), 1.76 – 1.64 (m, 4H), 1.39 (p, J = 7.7 Hz, 2H). ESI-MS m/z (M+H) $^+$ = 294.16.

N,N,N-trimethyl-5-((7-nitrobenzo[*c*][1,2,5]oxadiazol-4-yl)amino)pentan-1-aminium (2). To a stirred solution of compound 1 (50mg, 0.17 mmol) in 3ml of DCM was added triethylamine (2.3 μ l, 0.02 mmol). Methyl iodide (53 μ l, 0.85 mmol) was then added and the reaction mixture turned a reddish-brown color. The reaction mixture was purified by reverse HPLC using the gradient of 10%_50% B in 40 min_60ml. Product eluted at 25% B and the fractions were combined and lyophilized to produce compound 1 (52mg) as a mahogany-colored semi-solid. ^1H NMR (400 MHz, DMSO- d_6) δ 9.54 (s, 1H), 8.53 (d, J = 9.0 Hz, 1H), 6.44 (d, J = 9.0 Hz, 1H), 3.51 (d, J = 7.6 Hz, 2H), 3.28 – 3.23 (m, 2H), 3.03 (s, 9H), 1.73 (h, J = 7.1 Hz, 4H), 1.38 (q, J = 7.7 Hz, 2H). ESI-MS m/z (M+H) $^+$ = 308.24.

4.4.6 NBD-CTA $^+$ Transport Assay

Proteoliposomes were preloaded with 500 μ M NBD-CTA $^+$ (5), subjected to 5 freeze/thaw cycles, and extruded 25 times through a 400-nm membrane filter. Liposomes were diluted 200-fold with assay buffer (100 mM KCl, 100 mM K₂HPO₄, pH 7.5) containing 5 mM freshly prepared sodium hydrosulfite to quench external NBD-CTA $^+$. NBD-CTA $^+$ fluorescence was monitored (λ_{ex} = 470 nm, λ_{em} = 540 nm) using a fluorimeter (FP-8300, Jasco, Easton, MD) for 200 seconds prior to addition of 0.1% Triton X-100 to release remaining encapsulated NBD-

CTA⁺. Experiments were done in triplicate for each of two independent biochemical purifications.

4.4.7 Tryptophan fluorescence

Assay buffer was matched to the size exclusion chromatography buffer (100 mM NaCl, 10 mM HEPES, pH 8.0, 5 mM DM). Changes in tryptophan fluorescence upon substrate addition were monitored using a Jasco FP8300 fluorimeter ($\lambda_{\text{ex}} = 280 \text{ nm}$, $\lambda_{\text{em}} = 300\text{-}400 \text{ nm}$) and fit into a single-site binding isotherm:

$$\Delta F = \left(\frac{F_{\text{max}}[S]}{K_d + [S]} \right)$$

4.4.8 Solid supported membrane (SSM) electrophysiology

SSM electrophysiology measurements were accomplished using the SURFE²R N1 Instrument (Nanion Technologies, Munich, Germany). Reconstitution and sensor preparation were performed exactly as described in Kermani & Macdonald *et al.* 2020. At least 3 perfusions were collected for each sensor. Sensors were prepared from at least two independent protein purifications.

4.5 References

- Adrián, C.-R., Berend, S., Tristan, F.-B., Ruth, I., Gergely, G., Xiaoyun, B., Reinhart, A. F. R., David, H., Matthias, A. H., Aled, E., Aled, M. E., & Giulio, S.-F. (2015). A Call for Systematic Research on Solute Carriers. *Cell*. <https://doi.org/10.1016/j.cell.2015.07.022>
- Alvarez-Ortega, C., Olivares, J., & Martínez, J. L. (2013). RND multidrug efflux pumps: what are they good for? *Front Microbiol*, 4, 7. <https://doi.org/10.3389/fmicb.2013.00007>
- Amadi, S. T., Koteiche, H. A., Mishra, S., & McHaourab, H. S. (2010). Structure, dynamics, and substrate-induced conformational changes of the multidrug transporter EmrE in liposomes. *J. Biol. Chem.*, 285. <https://doi.org/10.1074/jbc.M110.132621>

- Argüello, J. M. (2003). Identification of ion-selectivity determinants in heavy-metal transport P1B-type ATPases. *J Membr Biol*, 195(2), 93-108. <https://doi.org/10.1007/s00232-003-2048-2>
- Bay, D. C., Rommens, K. L., & Turner, R. J. (2008). Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochim Biophys Acta*, 1778(9), 1814-1838. <https://doi.org/10.1016/j.bbamem.2007.08.015>
- Brill, S., Falk, O. S., & Schuldiner, S. (2012). Transforming a drug/H⁺ antiporter into a polyamine importer by a single mutation. *Proc Natl Acad Sci U S A*, 109(42), 16894-16899. <https://doi.org/10.1073/pnas.1211831109>
- Burata, O. E., Yeh, T. J., Macdonald, C. B., & Stockbridge, R. B. (2022). Still rocking in the structural era: A molecular overview of the small multidrug resistance (SMR) transporter family. *J Biol Chem*, 298(10), 102482. <https://doi.org/10.1016/j.jbc.2022.102482>
- César-Razquin, A., Snijder, B., Frappier-Brinton, T., Isserlin, R., Gyimesi, G., Bai, X., Reithmeier, R. A., Hepworth, D., Hediger, M. A., Edwards, A. M., & Superti-Furga, G. (2015). A Call for Systematic Research on Solute Carriers. *Cell*, 162(3), 478-487. <https://doi.org/10.1016/j.cell.2015.07.022>
- Davidson, A. L., Dassa, E., Orelle, C., & Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev*, 72(2), 317-364, table of contents. <https://doi.org/10.1128/mmbr.00031-07>
- Elbaz, Y., Salomon, T., & Schuldiner, S. (2008). Identification of a glycine motif required for packing in EmrE, a multidrug transporter from Escherichia coli. *J Biol Chem*, 283(18), 12276-12283. <https://doi.org/10.1074/jbc.M710338200>
- Elbaz, Y., Tayer, N., Steinfels, E., Steiner-Mordoch, S., & Schuldiner, S. (2005). Substrate-induced tryptophan fluorescence changes in EmrE, the smallest ion-coupled multidrug transporter. *Biochemistry*, 44. <https://doi.org/10.1021/bi050356t>
- Fleishman, S. J., Harrington, S. E., Enosh, A., Halperin, D., Tate, C. G., & Ben-Tal, N. (2006). Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *J Mol Biol*, 364(1), 54-67. <https://doi.org/10.1016/j.jmb.2006.08.072>
- Fotiadis, D., Kanai, Y., & Palacín, M. (2013). The SLC3 and SLC7 families of amino acid transporters. *Mol Aspects Med*, 34(2-3), 139-158. <https://doi.org/10.1016/j.mam.2012.10.007>
- Gayen, A., Leninger, M., & Traaseth, N. J. (2016). Protonation of a glutamate residue modulates the dynamics of the drug transporter EmrE. *Nat Chem Biol*, 12(3), 141-145. <https://doi.org/10.1038/nchembio.1999>
- Gutman, N., Steiner-Mordoch, S., & Schuldiner, S. (2003). An amino acid cluster around the essential Glu-14 is part of the substrate- and proton-binding domain of EmrE, a multidrug transporter from Escherichia coli. *J Biol Chem*, 278(18), 16082-16087. <https://doi.org/10.1074/jbc.M213120200>
- Hussey, G. A., Thomas, N. E., & Henzler-Wildman, K. A. (2020). Highly coupled transport can be achieved in free-exchange transport models. *J Gen Physiol*, 152(1). <https://doi.org/10.1085/jgp.201912437>
- Kermani, A. A., Burata, O. E., Koff, B. B., Koide, A., Koide, S., & Stockbridge, R. B. (2022). Crystal structures of bacterial small multidrug resistance transporter EmrE in complex with structurally diverse substrates. *Elife*, 11. <https://doi.org/10.7554/eLife.76766>

- Kermani, A. A., Macdonald, C. B., Burata, O. E., Ben Koff, B., Koide, A., Denbaum, E., Koide, S., & Stockbridge, R. B. (2020). The structural basis of promiscuity in small multidrug resistance transporters. *Nat Commun*, 11(1), 6064. <https://doi.org/10.1038/s41467-020-19820-8>
- Kermani, A. A., Macdonald, C. B., Gundepudi, R., & Stockbridge, R. B. (2018a). Guanidinium export is the primal function of SMR family transporters. *Proc Natl Acad Sci U S A*, 115(12), 3060-3065. <https://doi.org/10.1073/pnas.1719187115>
- Kermani, A. A., Macdonald, C. B., Gundepudi, R., & Stockbridge, R. B. (2018b). Guanidinium export is the primal function of SMR family transporters. *Proc. Natl Acad. Sci. USA*, 115. <https://doi.org/10.1073/pnas.1719187115>
- Kim, J., Cater, R. J., Choy, B. C., & Mancina, F. (2021). Structural Insights into Transporter-Mediated Drug Resistance in Infectious Diseases. *J Mol Biol*, 433(16), 167005. <https://doi.org/10.1016/j.jmb.2021.167005>
- Kobayashi, N., Nishino, K., & Yamaguchi, A. (2001). Novel macrolide-specific ABC-type efflux transporter in Escherichia coli. *J Bacteriol*, 183(19), 5639-5644. <https://doi.org/10.1128/jb.183.19.5639-5644.2001>
- Martinez, J. L., Sánchez, M. B., Martínez-Solano, L., Hernandez, A., Garmendia, L., Fajardo, A., & Alvarez-Ortega, C. (2009). Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. *FEMS Microbiol Rev*, 33(2), 430-449. <https://doi.org/10.1111/j.1574-6976.2008.00157.x>
- Mordoch, S. S., Granot, D., Lebendiker, M., & Schuldiner, S. (1999). Scanning cysteine accessibility of EmrE, an H⁺-coupled multidrug transporter from Escherichia coli, reveals a hydrophobic pathway for solutes. *J Biol Chem*, 274(27), 19480-19486. <https://doi.org/10.1074/jbc.274.27.19480>
- Morrison, E. A., DeKoster, G. T., Dutta, S., Vafabakhsh, R., Clarkson, M. W., Bahl, A., Kern, D., Ha, T., & Henzler-Wildman, K. A. (2011). Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature*, 481(7379), 45-50. <https://doi.org/10.1038/nature10703>
- Mueckler, M., & Thorens, B. (2013). The SLC2 (GLUT) family of membrane transporters. *Mol Aspects Med*, 34(2-3), 121-138. <https://doi.org/10.1016/j.mam.2012.07.001>
- Muth, T. R., & Schuldiner, S. (2000). A membrane-embedded glutamate is required for ligand binding to the multidrug transporter EmrE. *Embo j*, 19(2), 234-240. <https://doi.org/10.1093/emboj/19.2.234>
- Nelson, J. W., Atilho, R. M., Sherlock, M. E., Stockbridge, R. B., & Breaker, R. R. (2017). Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. *Mol Cell*, 65(2), 220-230. <https://doi.org/10.1016/j.molcel.2016.11.019>
- Pao, S. S., Paulsen, I. T., & Saier, M. H., Jr. (1998). Major facilitator superfamily. *Microbiol Mol Biol Rev*, 62(1), 1-34. <https://doi.org/10.1128/mmbr.62.1.1-34.1998>
- Piddock, L. J. (2006). Multidrug-resistance efflux pumps - not just for resistance. *Nat Rev Microbiol*, 4(8), 629-636. <https://doi.org/10.1038/nrmicro1464>
- Ploier, B., & Menon, A. K. (2016). A Fluorescence-based Assay of Phospholipid Scramblase Activity. *J Vis Exp*(115). <https://doi.org/10.3791/54635>
- Robinson, A. E., Thomas, N. E., Morrison, E. A., Balthazor, B. M., & Henzler-Wildman, K. A. (2017). New free-exchange model of EmrE transport. *Proc. Natl Acad. Sci. USA*, 114. <https://doi.org/10.1073/pnas.1708671114>

- Rotem, D., & Schuldiner, S. (2004). EmrE, a multidrug transporter from *Escherichia coli*, transports monovalent and divalent substrates with the same stoichiometry. *J. Biol. Chem.*, 279. <https://doi.org/10.1074/jbc.M408187200>
- Rotem, D., Steiner-Mordoch, S., & Schuldiner, S. (2006). Identification of tyrosine residues critical for the function of an ion-coupled multidrug transporter. *J. Biol. Chem.*, 281. <https://doi.org/10.1074/jbc.M602088200>
- Saleh, M., Bay, D. C., & Turner, R. J. (2018). Few Conserved Amino Acids in the Small Multidrug Resistance Transporter EmrE Influence Drug Polyselectivity. *Antimicrob Agents Chemother*, 62(8). <https://doi.org/10.1128/aac.00461-18>
- Sharoni, M., Steiner-Mordoch, S., & Schuldiner, S. (2005). Exploring the binding domain of EmrE, the smallest multidrug transporter. *J Biol Chem*, 280(38), 32849-32855. <https://doi.org/10.1074/jbc.M504910200>
- Sun, J., Deng, Z., & Yan, A. (2014). Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. *Biochem Biophys Res Commun*, 453(2), 254-267. <https://doi.org/10.1016/j.bbrc.2014.05.090>
- Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S., & Tate, C. G. (2003). Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J.*, 22. <https://doi.org/10.1093/emboj/cdg611>
- Wu, C. (2019). Identification of an alternating-access dynamics mutant of EmrE with impaired transport. *J. Mol. Biol.*, 431. <https://doi.org/10.1016/j.jmb.2019.05.035>
- Yerushalmi, H., & Schuldiner, S. (2000). An essential glutamyl residue in EmrE, a multidrug antiporter from *Escherichia coli*. *J Biol Chem*, 275(8), 5264-5269. <https://doi.org/10.1074/jbc.275.8.5264>

Chapter 5 Conclusions & Future Works

Portions of this chapter were performed by Mo Siddiq (phylogenetic ancestral reconstruction), Rachael Lucero & Olive Burata (purification of ancestral sequences; preliminary functional characterization), Kemal Demirer (preliminary functional characterization with expanded substrate repertoire) and Randy Stockbridge.

The SMR family is an intriguing model to use to investigate the riddles of substrate promiscuity. They are the smallest transporters currently known to date and are structurally simpler models than other multidrug efflux pumps. In the recent decade, SMR exemplars from both major specificities within the family have been extensively characterized for both their function and structural features, mostly due in part to the advances in biophysical techniques required to study them *in vitro*. Through this, high resolution models of exemplars from both distinct substrate specificities, the Qacs and the Gdxs, have enabled us to discover the diversity in their binding pockets and how they coordinate the type of substrates they can bind to and transport. The differences are subtle rotameric changes in the conserved residues that lead to a wider, less rigid binding site in Qas, but a more organized H-bond network stabilizing the same site in Gdxs.

As we engineered a Gdx variant, 7X, to behave more like a Qac, we observe similar features indicative of QAC-like function: the ability to transport quaternary ammonium compounds, and the preference for bulkier, more hydrophobic guanidinylated compounds. All in

all, the 7 required mutations have led to a slightly dysfunctional H-bond network in the binding pocket, allowing free rotameric reign of the conserved glutamates and tryptophans, disrupting the helical packing to accommodate larger substrates, and inducing a collaborative change in the rates of the subunits as they swivel from inward to outward conformations. The fact that 7 mutations were required to induce the best specificity jump in these conditions indicates that the likelihood of this happening all at once in nature is very low and would likely require several evolutionary events (or several favorable energetic landscapes) to occur. However, with these same mutations, we failed to see transport of bulky, polyaromatic toxic compounds commonly associated as cognate substrate of Qacs. Therefore, despite requiring 7 mutations, the ease with which quaternary compounds were the first to be transported by this Gdx variant could potentially hint that an additional set of mutations may be required for bigger substrates to be accommodated. This could also hint the evolutionary timescale of when different classes of substrates are recognized by Qacs. The possibility of quaternary ammoniums as the first classes of substrates transported by Qacs remains likely especially since Qacs and Gdxs share a membrane portal connecting the binding pocket to the interior of the membrane, a cryptic feature conserved in both subtypes that accommodates long hydrophobic tails mostly found in quaternary ammonium compounds. These features could explain quaternary ammoniums are most likely transported first, before the big leap towards transporting bulkier, hydrophobic compounds commonly associated with antibiotics.

5.1 The Ancestral Reconstruction of the SMR Family

Although the existence of two contrasting specificities in the SMR family has been experimentally useful, it remains an odd phenomenon. Throughout three decades worth of deep

research on the SMR family, there remains a controversy as to which subtype is the true ancestor from which other functions may have diverged. The prevailing theory is that Gdxs are the ancestral function of the family since their native substrate, guanidinium, has been recently found to be essentially linked to microbial metabolism, making this substrate physiological relevant (Martinez-Vaz et al., 2022; Nelson et al., 2017). Substrates transported by Qacs are typically synthetic and non-natural, and in the case of antibiotics, which have been over utilized since its discovery over a century ago, is suggestive of the Qacs as a recent evolutionary adaptation by microbes. However, this theory relies solely on Qac genes as externally acquired resistomes, which is unlikely as SMR genes are also found in bacterial chromosomes indicating intrinsic resistance. Since gene duplication events are a potential driving force for neofunctionalization in the SMR family, the idea that promiscuous transport by Qacs can be the primal function is not entirely farfetched. Substrate promiscuity is a challenging aspect to optimize from, as evolutionary adaptations typically benefit selective enzymes, at least based on enzyme-centric promiscuity models. Therefore, an ancestral reconstruction of the SMR family offers a robust tool in helping us uncover the true ancestral gene. **Figures 5.1a** shows preliminary attempts at constructing a rootless phylogenetic tree using SMR sequences curated from the GEBA genomes.

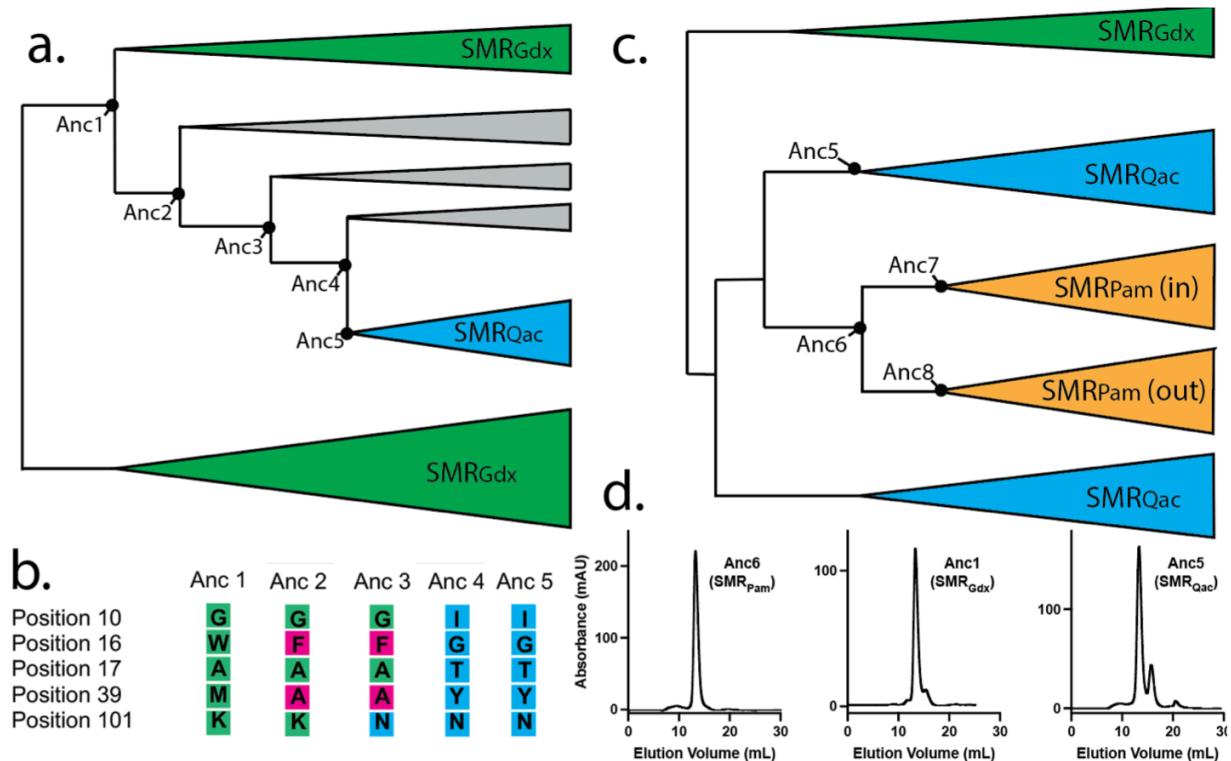


Figure 5.1: Preliminary ancestral reconstruction of the SMRs.

A, C. Schematic of phylogeny from Figure 1 with nodes of interest indicated. **B.** Identities of residues at critical positions for ancestral nodes of SMRGdx (Anc-1), SMRQac (Anc-5), and intermediate clades. Green indicates correspondence with SMRGdx, blue with SMRQac, and magenta indicates a residue that is not present in either subtype. **D.** Size exclusion chromatograms of Anc-1, Anc-5, and Anc-6.

Five nodes were chosen and their residue content at these critical positions are shown in **Fig. 5.1b**. Ancestor 1 from the Gdx-like clade has the residues in this positions that are uniquely conserved in Gdx. On the other hand, Ancestors-4 and -5, both Qac-like sequences, also maintain the same residues at these positions that are only high conserved in Qacs. The intermediary Ancestors-2 and -3, differ from both subtypes but only at two positions: F16 and A39. Both ancestors lost their capabilities of H-bonding at these positions and may have provided an evolutionary stepping-stone from one subtype to the other. Ancestor-6 demonstrated sequences of polyamine transporters from the SMR family are shown in **Fig. 5.1c** which diverged from the Qac-like Ancestor-5. Ancestor-6 further dichotomizes into two sequential nodes, each

representing the conserved sequences representing the “inward” and “outward” individual subunits. To validate the function of these preliminary ancestral sequences, heterologous expression was done for Ancestors-1, -5 and -6 in the same manner as the previously characterized SMRs. Their SEC purification profiles mirrored those we have seen before with other well-characterized SMR models, indicating that these ancestral sequences are beyond theoretical constructs and are poised for downstream *in vitro* characterization.

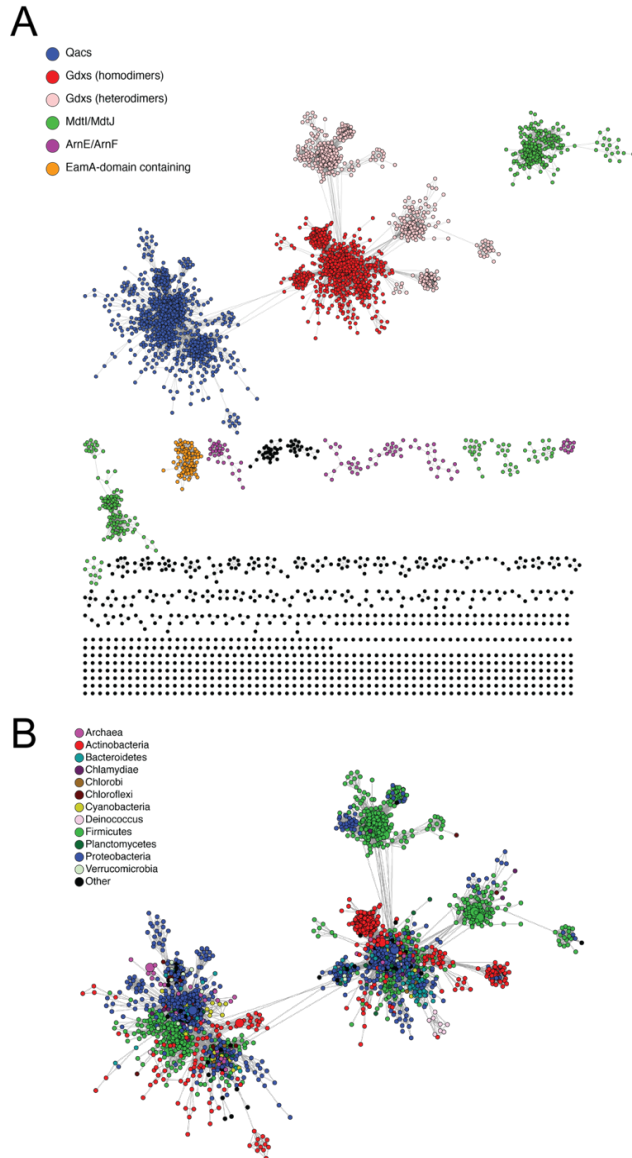
The investigation into the SMR family's substrate promiscuity, highlighted by the structural simplicity and evolutionary adaptability of the smallest known transporters, elucidates key evolutionary developments in microbial resistance mechanisms. The creation of the SMR variant Gdx-Clo 7X, bridging the functional divide between SMR_{Gdx} and SMR_{Qac} subtypes, reveals the intricate molecular collaboration of amino acid permutations necessary for the evolutionary leap in substrate specificity. This adaptive process, necessitating a series of mutations, suggests a complex, multi-prong evolutionary trajectory. Furthermore, the ancestral reconstruction of the SMR family opens new possibilities for understanding the origins of substrate specificity and promiscuity. It motivates a possible ancient lineage of SMR_{Qac}-like transporters, potentially reshaping our understanding of microbial evolutionary history and offering a promising direction for future research in combatting antimicrobial resistance.

5.2 References

- Martinez-Vaz, B. M., Dodge, A. G., Lucero, R. M., Stockbridge, R. B., Robinson, A. A., Tassoulas, L. J., & Wackett, L. P. (2022). Wastewater bacteria remediating the pharmaceutical metformin: Genomes, plasmids and products. *Front Bioeng Biotechnol*, 10, 1086261. <https://doi.org/10.3389/fbioe.2022.1086261>
- Nelson, J. W., Atilho, R. M., Sherlock, M. E., Stockbridge, R. B., & Breaker, R. R. (2017). Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Mol. Cell*, 65. <https://doi.org/10.1016/j.molcel.2016.11.019>

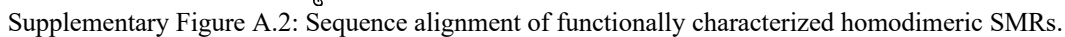
Appendices

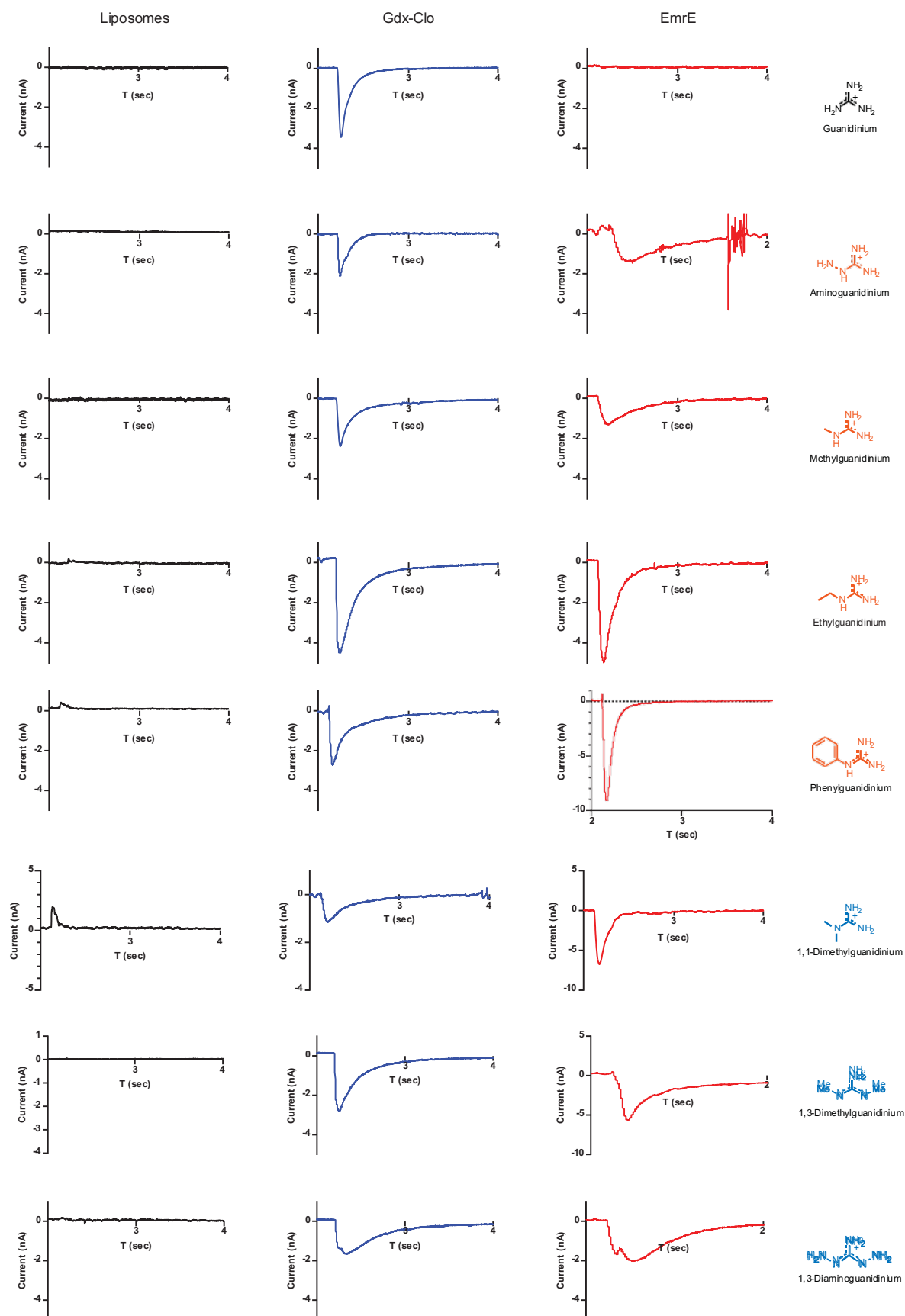
Appendix A: Supporting Information for Chapter 2

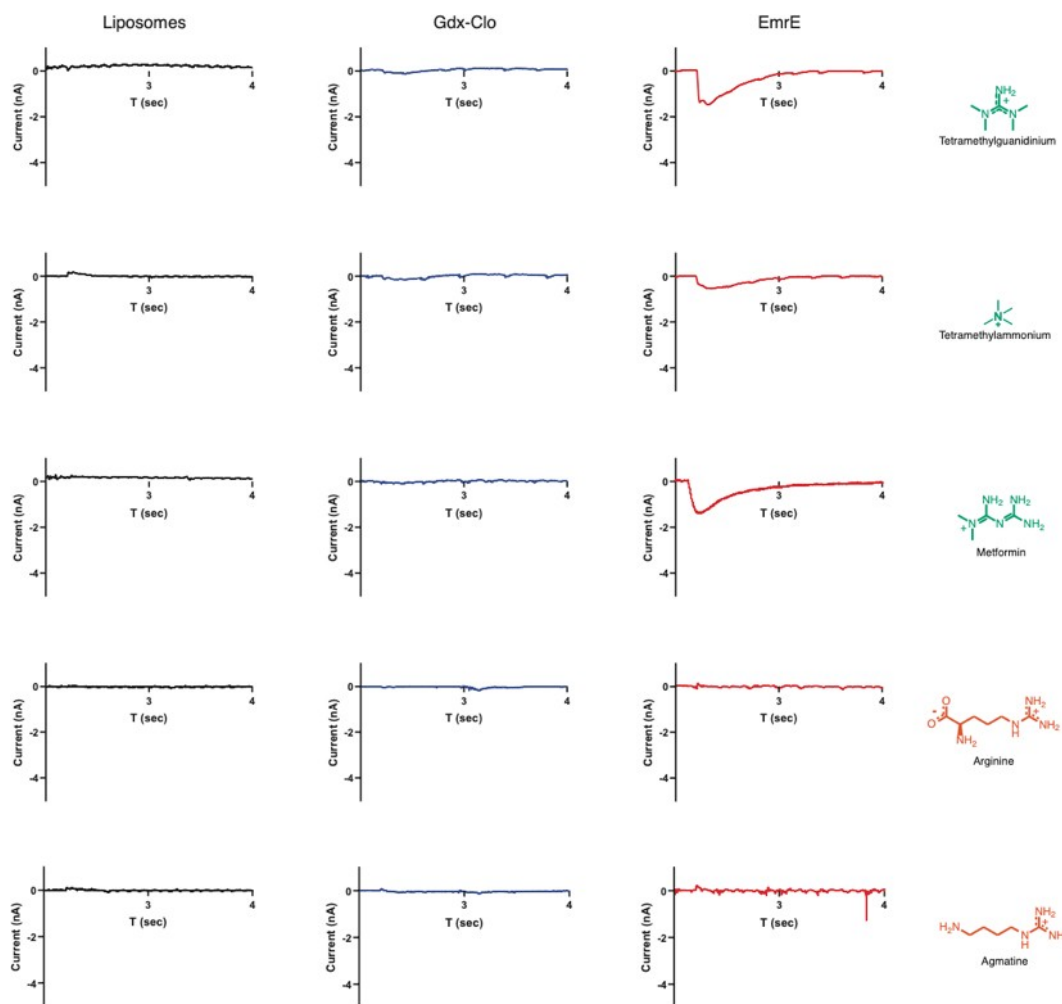


Supplementary Figure A.1: SMR sequence similarity networks with additional annotation.

A. Full sequence similarity network with gene annotation of each cluster, where known. B. Distribution of SMR proteins in archaea and different bacterial taxa. Sequence similarity network as in Figure 1A with coloring according to taxa as indicated.

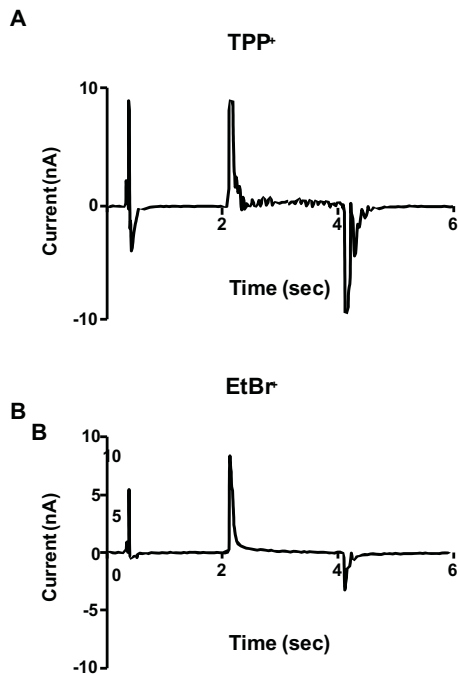




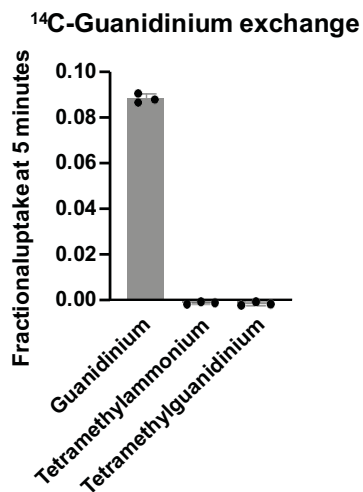


Supplementary Figure A.3: Representative SSM electrophysiology recordings.

Currents elicited after perfusion with substrate are shown for empty liposomes (left; black), Gdx-Clo proteoliposomes (middle; blue) and EmrE proteoliposomes (right; red).

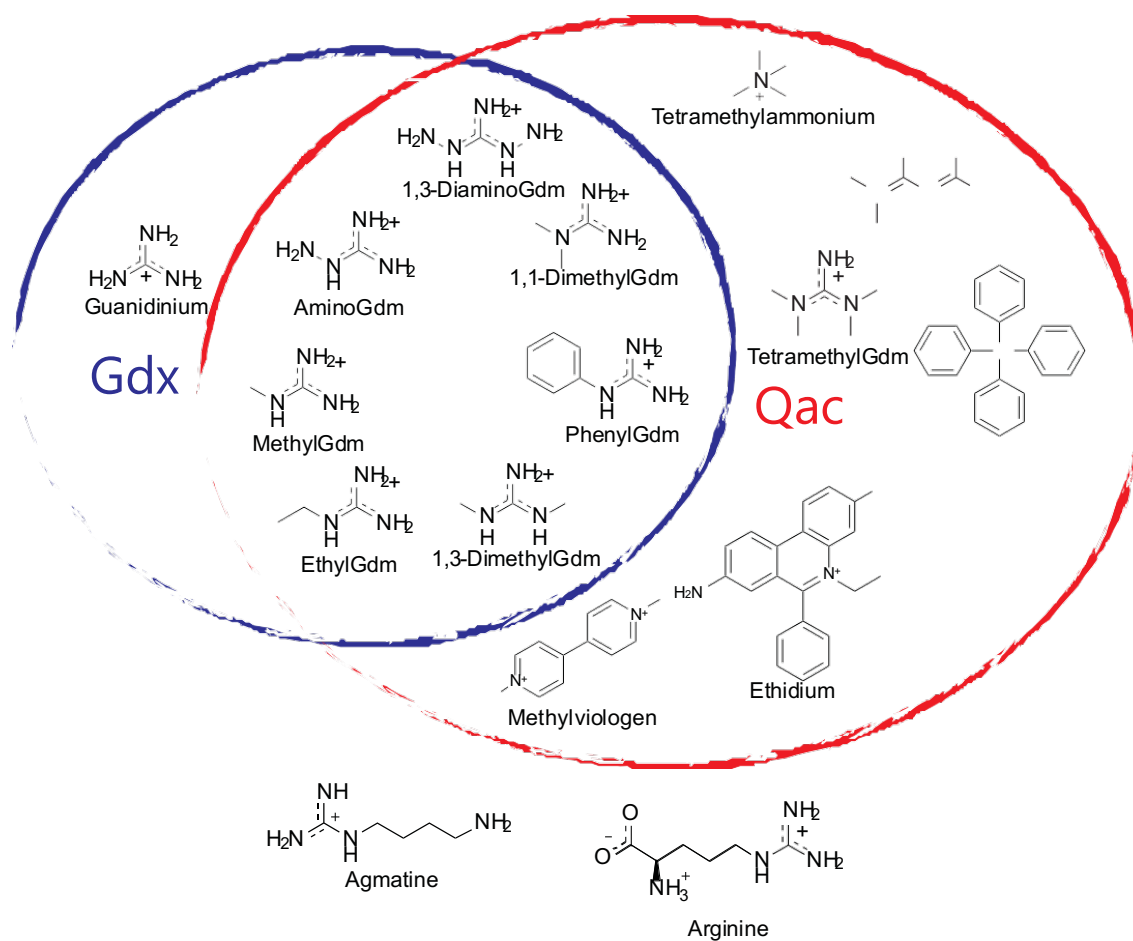


Supplementary Figure A.4: SSM traces for tetraphylphosphonium (TPP⁺, top) and ethidium (EtBr⁺) perfusion of liposomes that do not contain protein.

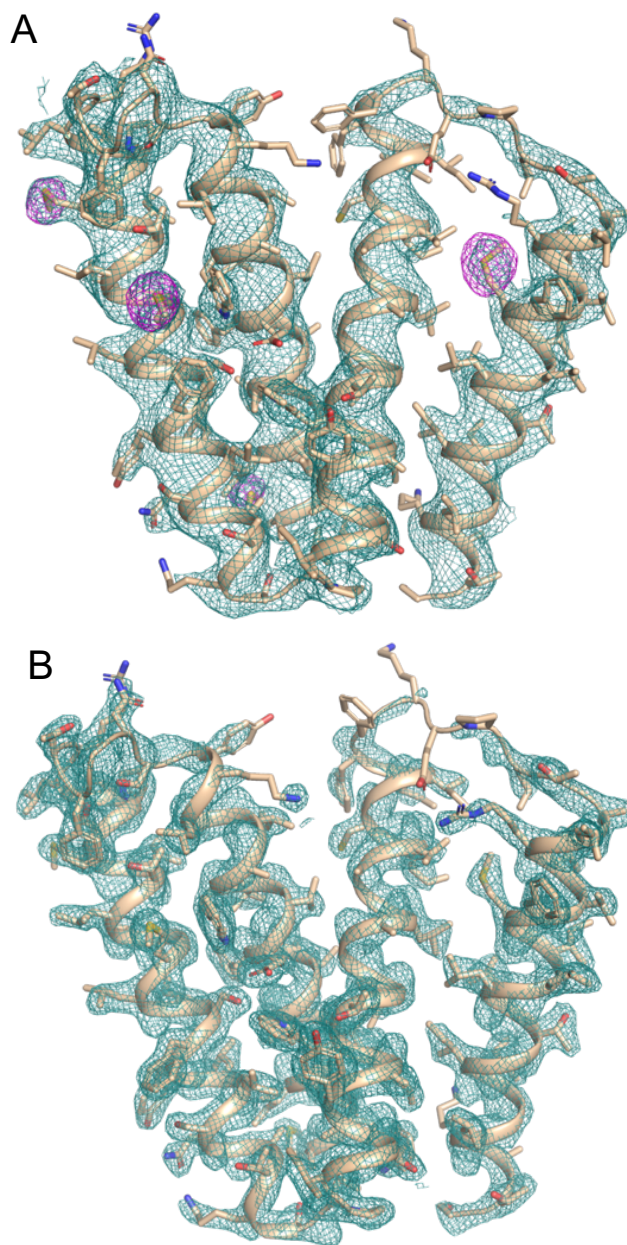


Supplementary Figure A.4: Uptake of ¹⁴C Gdm⁺ into Gdx-Clo proteoliposomes in exchange for the indicated substrate.

Experiment performed as in main text Figure 2, with fractional uptake measured relative to total radioactive counts in reaction mixture. Error bars represent the mean and SEM of three technical replicates.

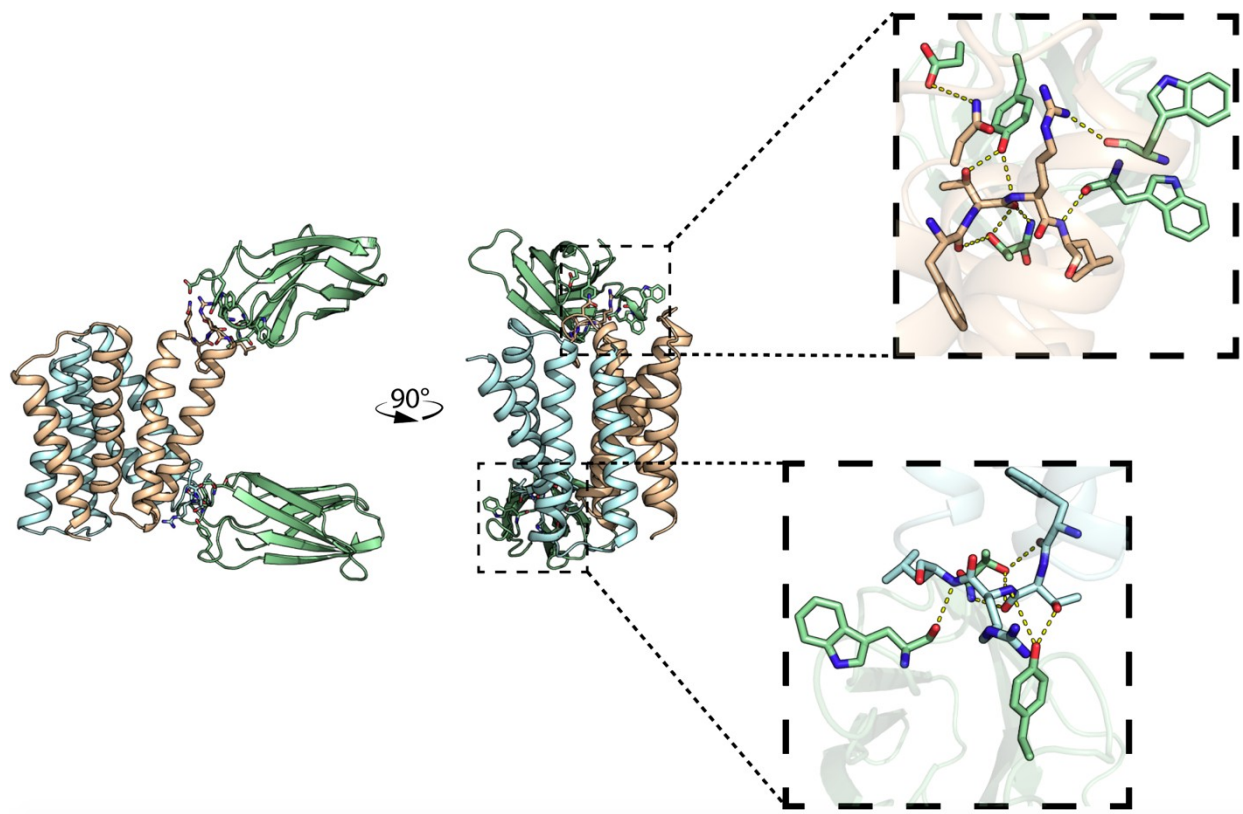


Supplementary Figure A.5: Venn diagram showing overlapping transport specificities of Gdx- Clo and EmrE.



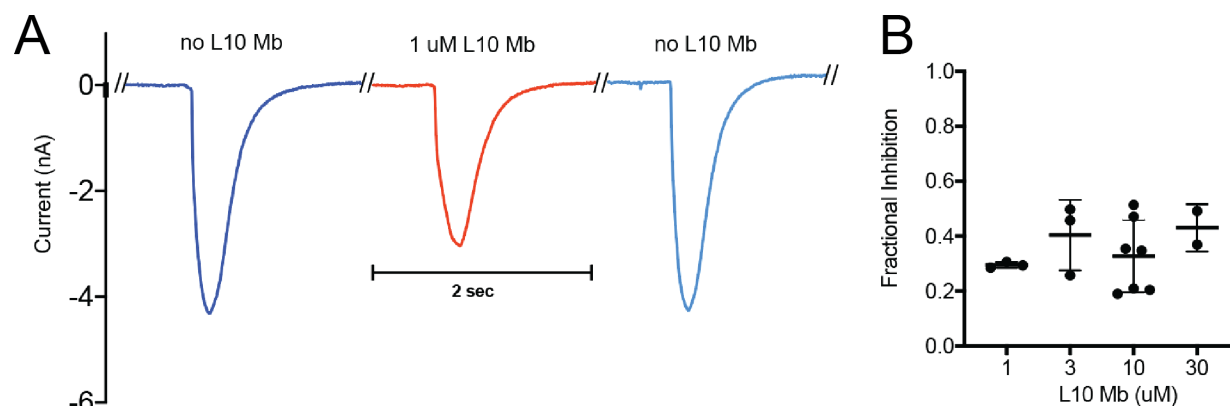
Supplementary Figure A.6: Experimental electron density maps for Gdx-Clo.

A. Cartoon view of one subunit from Gdx-Clo, with the solvent-flattened electron density map calculated from SHARP contoured at 1.3σ (teal), and anomalous difference density from seleno-L- methionine contoured at 5σ (magenta). B. Electron density map for one subunit of Gdx-Clo (octylGdm⁺-bound structure) contoured at 1.8σ .



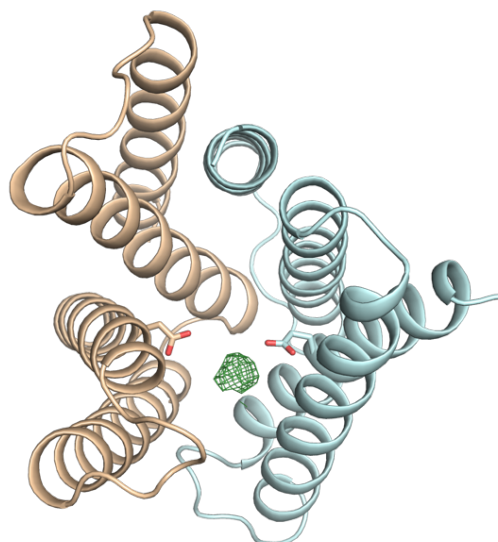
Supplementary Figure A.7: Binding interface between monobody Clo-L10 and Gdx-Clo.

Gdx-Clo shown in tan and cyan; monobody in green. Residues within H-bonding distance are shown as sticks, with H-bond interactions shown as dashed lines.



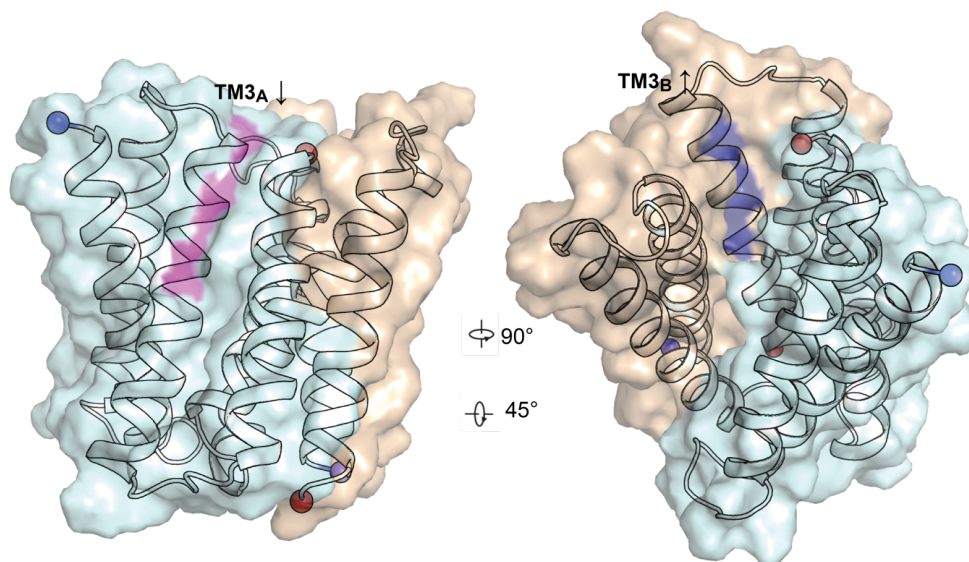
Supplementary Figure A.8: Currents mediated by Gdx-Clo in the presence and absence of monobody L10.

A. Currents elicited by perfusion with 1 mM Gdm⁺. For these experiments, benchmark current values for Gdx-Clo proteoliposomes were measured first (dark blue trace). Second, Gdm⁺ was removed by perfusion and L10 monobody was added and incubated with proteoliposomes for three minutes. Third, proteoliposomes were perfused with 1 mM Gdm⁺ containing L10 monobody (red trace). Fourth, both Gdm⁺ and L10 monobody were perfused away and proteoliposomes were incubated in buffer without L10 monobody for three minutes. Finally, a third recording was collected upon perfusion with 1 mM Gdm⁺ (light blue trace) to ensure that currents returned to the benchmark value. B. Fractional inhibition of Gdm⁺ currents by L10 monobody, added to the indicated concentrations. Error bars represent the mean and SEM. Data collected from at least three independent sensor preparations derived from two independent protein preparations.



Supplementary Figure A.9: Electron density between E13 and E13' in 3.2 Å structure solved with 10 mM Gdm⁺.

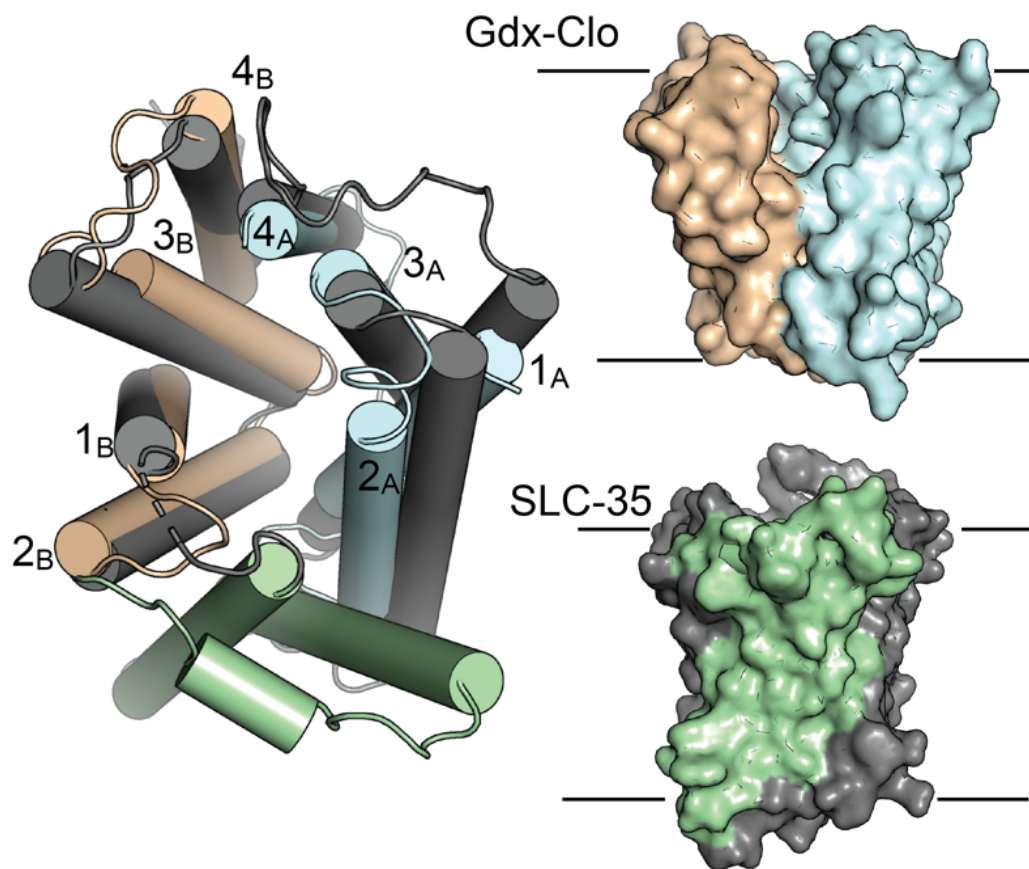
Top-down view of Clo-Gdx with subunits colored tan and light blue and E13 sidechains shown as sticks. Fo-Fc map in the region of the E13 sidechains is contoured at 3.2 a.



I₅₅GTAYAIWTGI₆₅GALGAVICG

Supplementary Figure A.10: Surface rendering of exposed TM3 GxxxG motifs.

Left, surface view of Gdx-Clo viewed through the plane of the membrane. Right, view is rotated 90° and tilted to view the interior of the aqueous-exposed vestibule. Coloring of surface rendering corresponds to TM3 sequence shown below. The first (magenta) GxxxG motif is exposed to the membrane in subunit A, and packed in the protein interior in subunit B. The second (dark blue) GxxxG motif is exposed to the aqueous vestibule in subunit B, but packed in the protein interior in subunit A. Conformational exchange swaps the accessibility of each GxxxG motif.



Supplementary Figure A.11: Structural alignment of Gdx-Clo

(subunits in tan and light blue as in main text) and CMP-sialic acid transporter from the SLC35 family (PDB: 6I1R [<https://www.rcsb.org/structure/6i1r>]; shown in dark gray with helix insertions in green). SLC35 proteins that share this fold have been structurally characterized, including ¹⁵⁻¹⁷. Left panel: top- down view of structural alignment. The helices are numbered for Gdx-Clo. Right panels: surface representation viewed through plane of membrane, with approximate membrane boundaries shown.

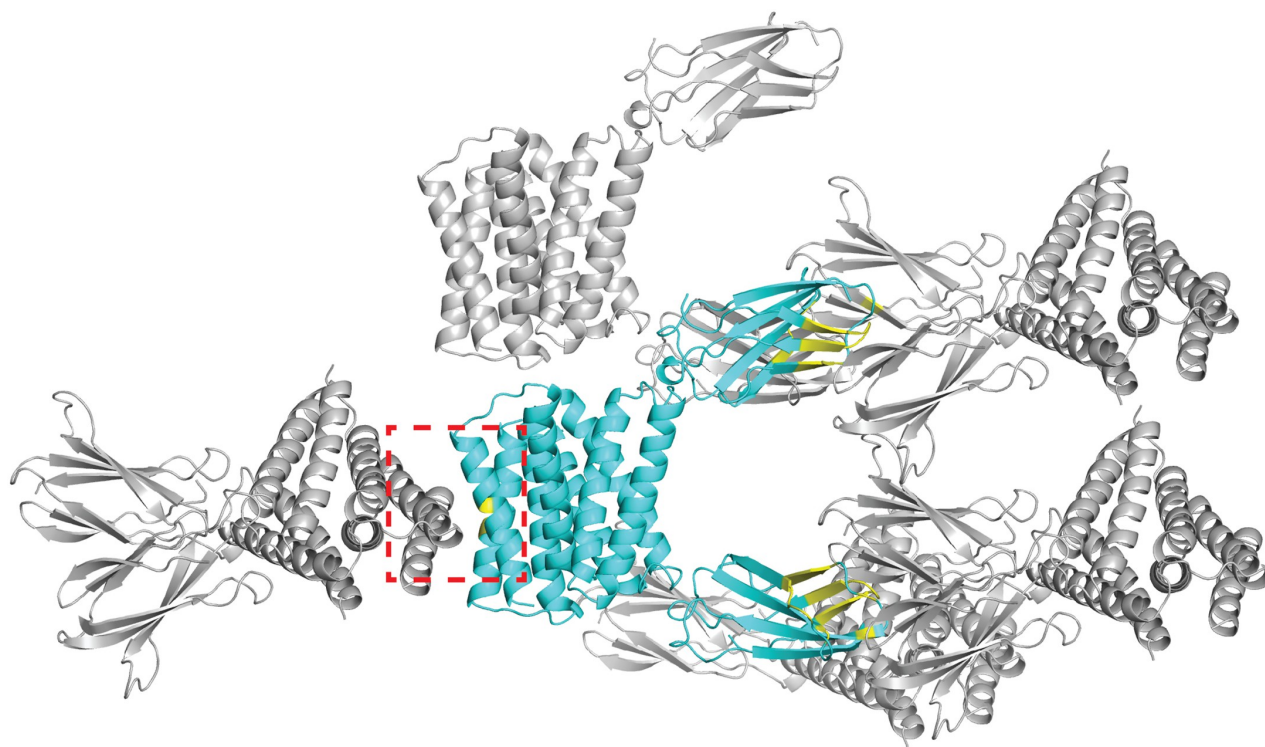
	Clo-L10-Se	Clo-L10-PheGdm ⁺	Clo-L10-OctylGdm ⁺
Data collection			
Space group	C121	C121	P1
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	141.8, 51.09, 108.43	141.82, 50.51, 108.58	51.04, 75.83, 109.12
α , β , γ (°)	90, 93.08, 90	90, 92.18, 90	92.54, 90.04, 109.63
Resolution (Å)	54.14-3.2(3.43-3.2)	58.32-2.53 (2.98-2.53)	109.0-2.32 (2.72-2.32)
Ellipsoidal Resolution Limit (best/worst) ^a	N/A	2.53/4.52	2.32/3.89
% Spherical Data Completeness ^a	99.9 (100)	36.5 (4.7)	37.8 (5.0)
% Ellipsoidal Data Completeness ^a	N/A	85.5 (66.2)	85.3 (54.2)
<i>R</i> _{merge}	0.06 (0.15)	0.132 (0.65)	0.074 (0.38)
<i>R</i> _{meas} ^a	--	0.145 (0.72)	0.086 (0.44)
Mn <i>I</i> / α ^a	17.3 (8.6)	6.4 (2.8)	9.2 (3.0)
Multiplicity ^a	7.5 (7.6)	7.1 (4.9)	3.8 (3.8)
<i>R</i> _{culis}	91.2%	N/A	N/A
Phasing Power ^b	.592	N/A	N/A
Refinement			
Resolution (Å)	44.3-3.5	58.3-2.5	33.0-2.3
No. reflections	10,076	9,018	24,995
<i>R</i> _{work} / <i>R</i> _{free}	25.2 / 27.8	25.7/30.9	24.6/28.6
Ramachandran Favored	93.6	83.6	94.1
Ramachandran Outliers	1.6	2.7	1.7
Clashscore	7.0	13.2	6.9
R.m.s. deviations			
Bond lengths (Å)	0.003	.007	.002
Bond angles (°)	.707	1.53	.585
Coordinates in Protein Databank	6WK5	6WK8	6WK9

^a Where applicable, values reported are for anisotropically truncated data performed using the Staraniso webserver (Global Phasing). See *Methods* for details.

^b Phasing Power = rms (|F_H| / ((F_H + F_P)-(F_{PH})))

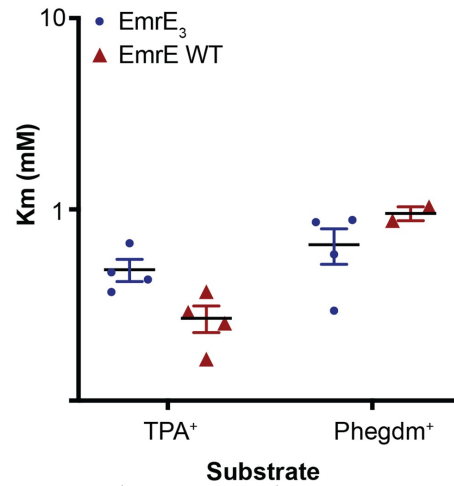
Table 3: Data collection, phasing and refinement statistics for Gdx-Clo complexes

Appendix B: Supporting Information for Chapter 3



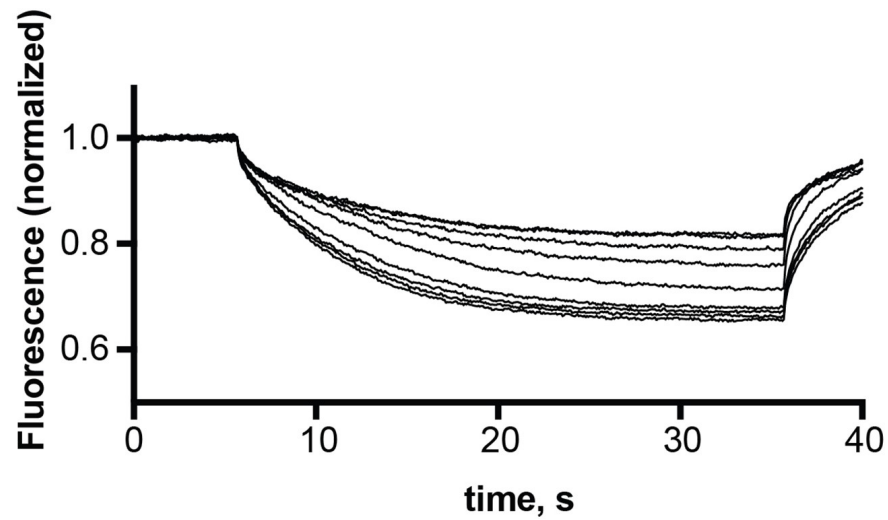
Supplementary Figure B.1: Crystal lattice for Gdx-Clo/L10 monobody complex (PDB: 6WK8).

The asymmetric unit, composed of one Gdx-Clo dimer and two monobodies, is shown in cyan. Symmetry mates are shown in gray. Residues that contribute to an interface between the asymmetric unit and its symmetry mates are colored yellow. Five Gdx-Clo residues are in contact with a symmetry mate: TM4 residues V88B, L92B, T95B, F89A, L92A (dashed red box).

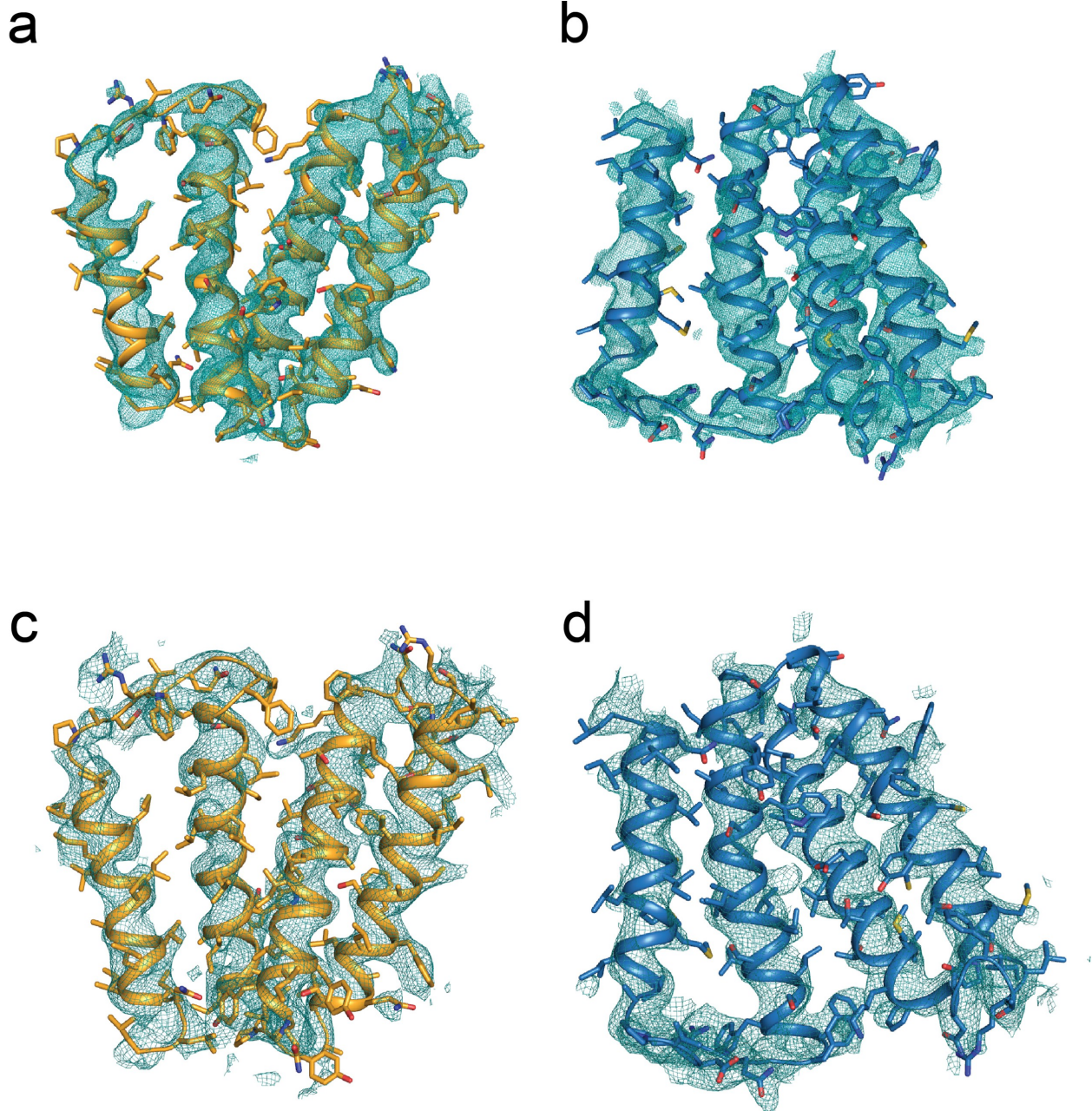


Supplementary Figure B.2: K_m values for TPA⁺ and PheGdm⁺ transport by EmrE₃ (blue) and WT EmrE (red).

Individual points are derived from Michaelis-Menten fits of titration experiments performed on a single sensor. Each K_m value was measured from a full titration series on an independently prepared sensor. Sensors are prepared from two to three independent biochemical purifications.



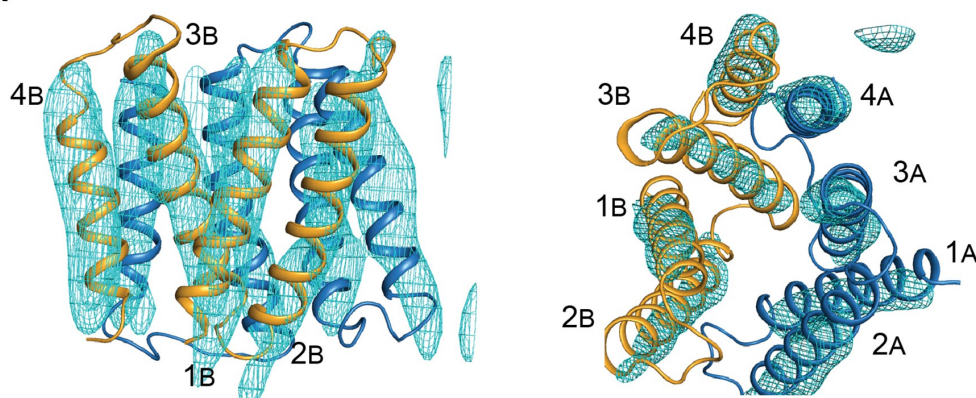
Supplementary Figure B.3: Representative microscale thermophoresis traces for monobody L10 in the presence of 30 nM – 10 μ M EmrE₃.



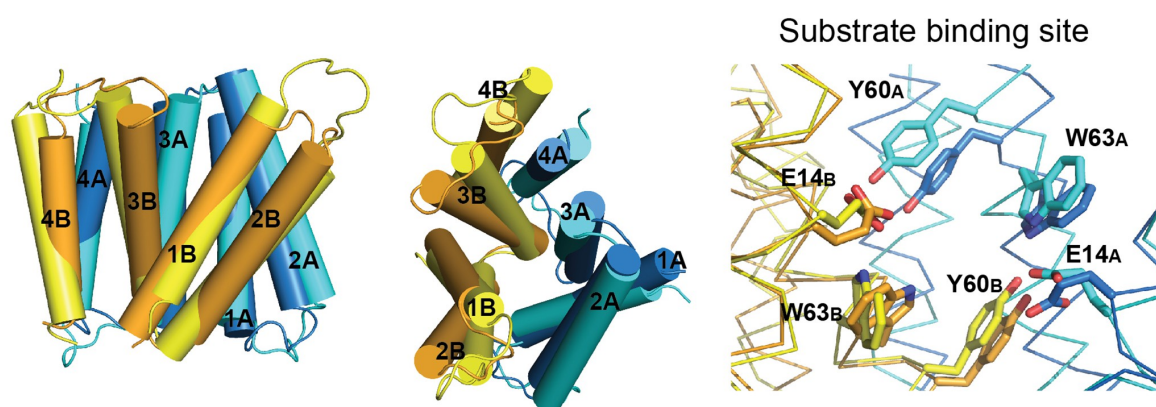
Supplementary Figure B.4: EmrE₃ maps.

Subunits colored as in main text, with subunit B in orange, and subunit A in blue. Panels A and B: 2Fo-Fc maps for EmrE₃, contoured at 1.2σ. Panels C and D: 2Fo-Fc composite omit maps for EmrE₃, contoured at 1.0σ, prepared by omitting 5% of the atoms in the model at a time.

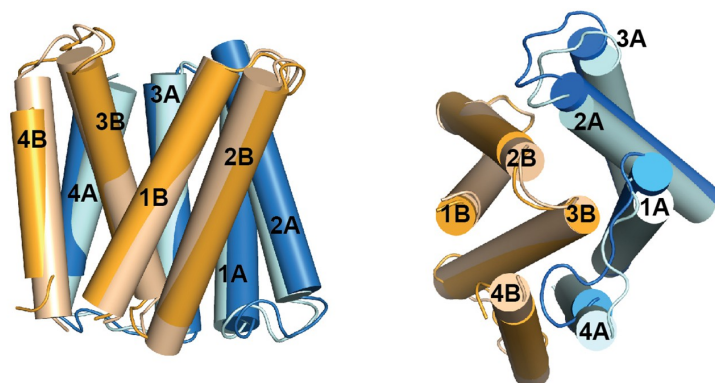
a Crystal structure overlaid with EmrE EM density from Ubarretxena-Belandia et al.(2003).



b Crystal structure overlaid with EmrE computational model from Vermaas et al. (2018).

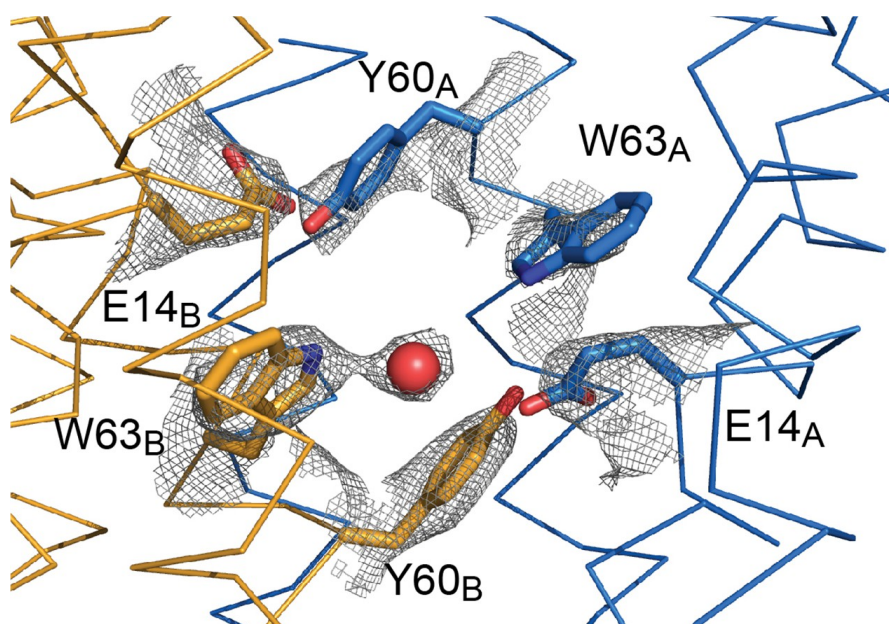


c Crystal structure overlaid with Gdx-Clo crystal structure from Kermani et al. (2020).



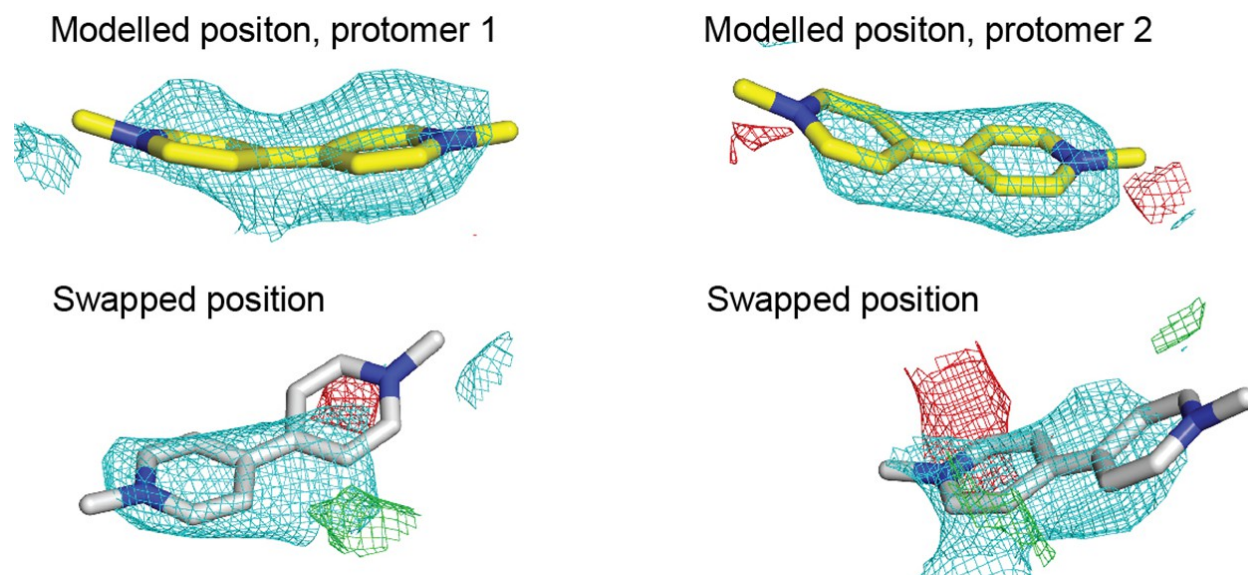
Supplementary Figure B.5: Structural comparison of EmrE₃ crystal structure with electron microscopy maps, theoretical model, and Gdx-Clo.

(A) Crystal structure of EmrE₃ (orange and blue cartoon) overlaid with experimental electron microscopy (EM) density (cyan mesh contoured at 1.5 σ) (Ubarretxena-Belandia et al., 2003). (B) Crystal structure of EmrE₃ (orange and blue) compared to a computational model (yellow and cyan) constrained by EM data (Vermaas et al., 2018). (C) Crystal structure of EmrE₃ (orange and blue) compared to crystal structure of a homologue from the SMR family, Gdx-Clo (wheat and pale cyan) (Kermani et al., 2020). Models are aligned along the B subunit.



Supplementary Figure B.6: Sidechain density in the EmrE₃ binding site.

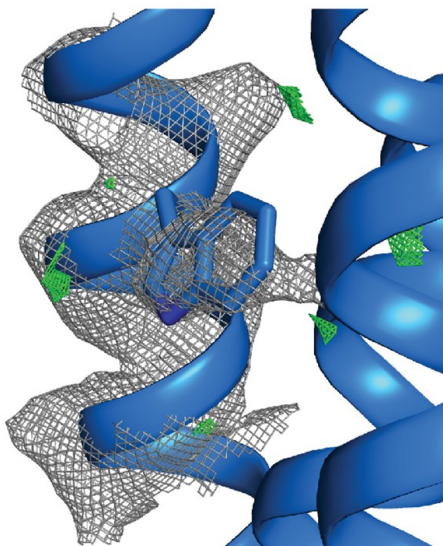
2Fo-Fc map around selected residues is contoured at 1.5 σ . The red sphere represents a water molecule.



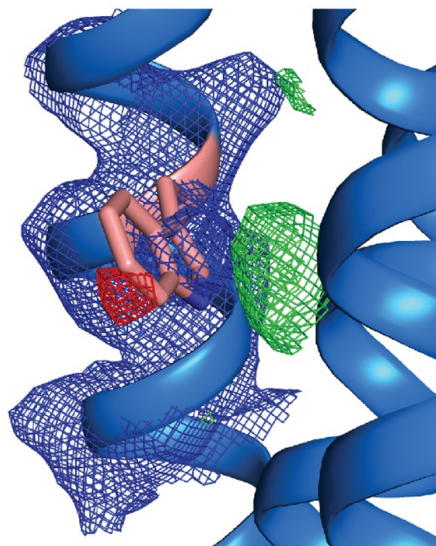
Supplementary Figure B.7: Electron density maps of methyl viologen in different EmrE₃ protomers in the asymmetric unit.

Yellow stick representations (top panels) show the modeled position of methyl viologen in each protomer, with the final refined maps shown as mesh. White stick representations show the methyl viologen position swapped between the two protomers. Maps show a subsequent re-refinement with the substrates in the swapped positions. For all panels, 2Fo-Fc density (cyan) contoured at 1.2 σ and Fo-Fc density (green or red) contoured at 2.5 σ .

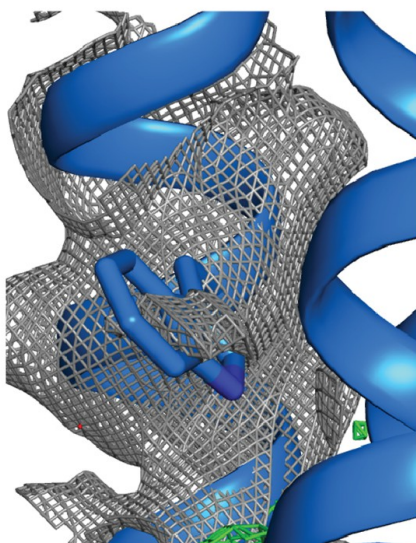
MeTPP⁺ structure with modelled W63_A



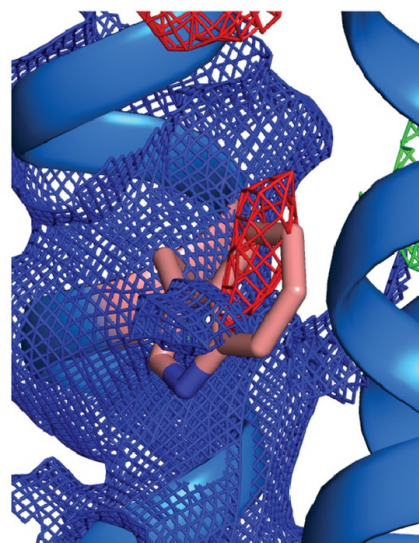
MeTPP⁺ structure with W63_A in same position as methylviologen structure



Methylviologen structure with modelled W63_A



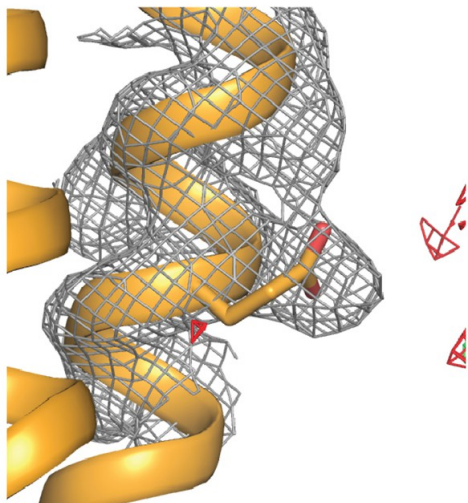
Methylviologen structure with W63_A in same position as MeTPP⁺ structure



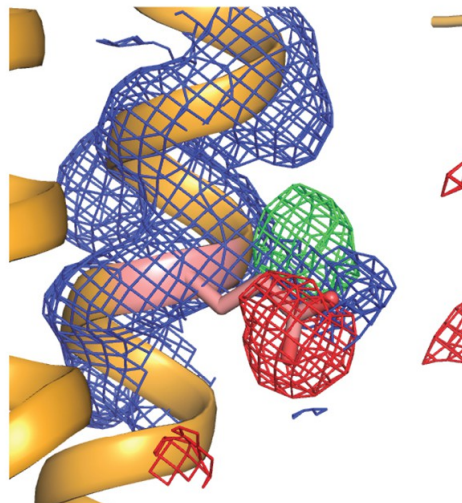
Supplementary Figure B.8: Electron density maps for W63A modeled in different positions.

Top panels: 2Fo-Fc density contoured at 1.8σ and Fo-Fc density contoured at 3σ . Bottom panels: 2Fo-Fc density contoured at 1.2σ and Fo-Fc density contoured at 2.5σ .

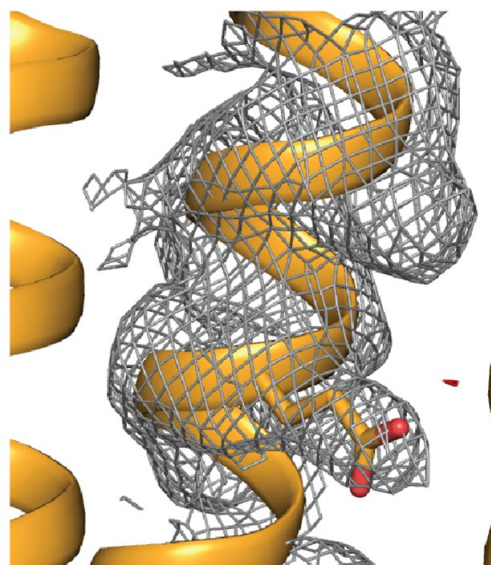
TPP⁺ structure with modelled E14_B



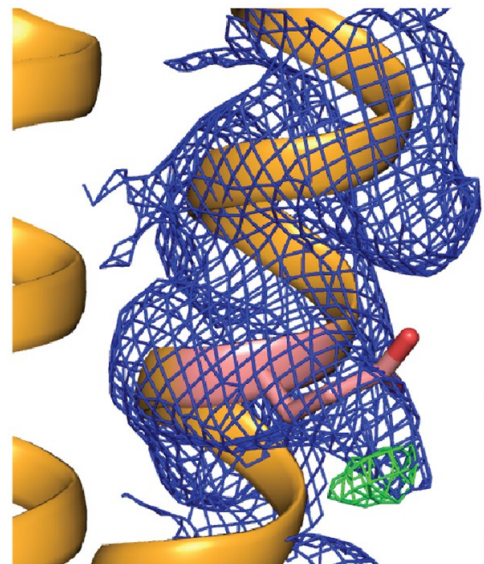
TPP⁺ structure with E14_B in same position as methylviologen structure



Methylviologen structure with modelled E14_B

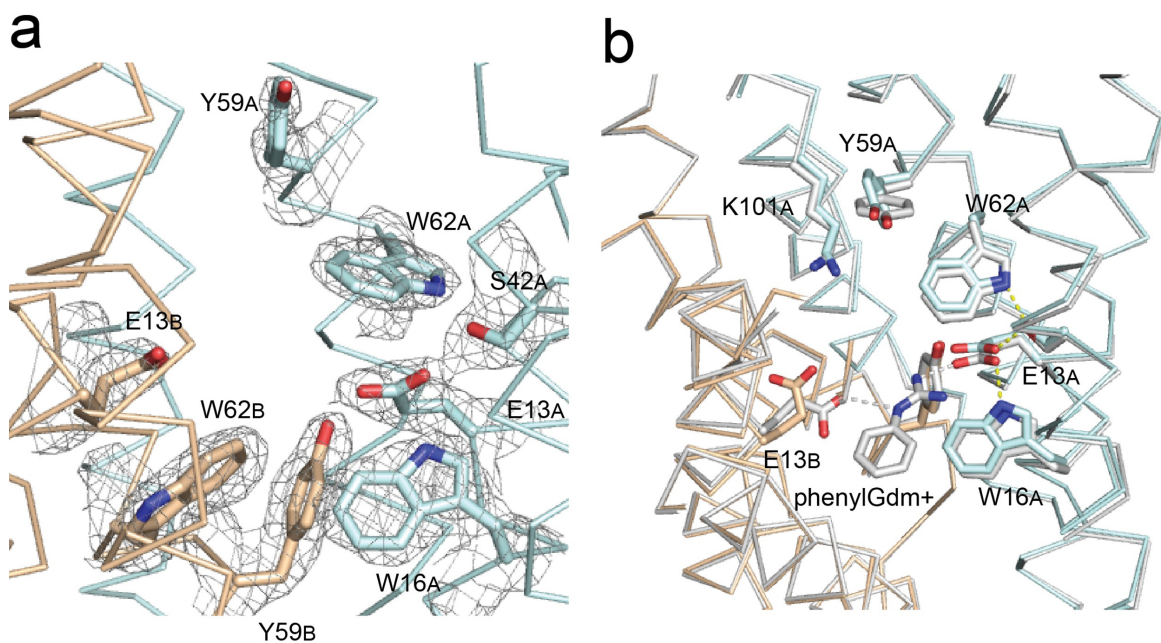


Methylviologen structure with E14_B in same position as TPP⁺ structure



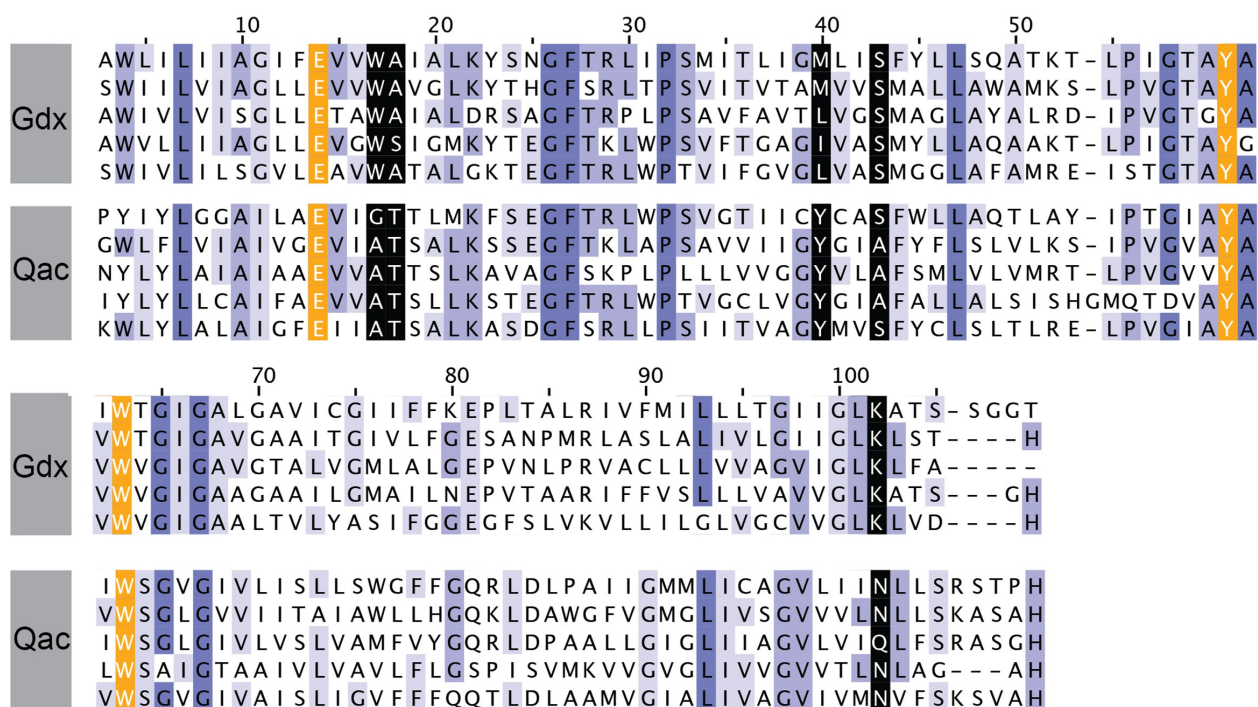
Supplementary Figure B.9: Electron density maps for E14B modeled in different positions.

Top panels: 2Fo-Fc density contoured at 1.8 σ and Fo-Fc density contoured at 3 σ . Bottom panels: 2Fo-Fc density contoured at 1.2 σ and Fo-Fc density contoured at 2.5 σ .



Supplementary Figure B.10: Gdx-Clo and EmrE substrate binding sites.

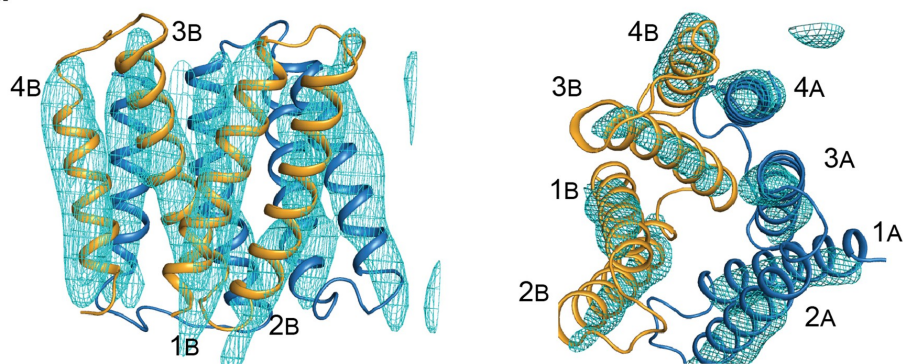
(A) 2Fo-Fc map shown around selected residues in the Gdx-Clo substrate binding site (pH 5.2) contoured at 1.5σ .
 (B) Alignment of Gdx-Clo structures. The present pH 5.2 structure is shown in wheat and cyan with putative H-bond interactions shown as yellow dashed lines. The structure with phenylGdm⁺ bound is shown in light gray with putative H-bonds between the substrate and the E13 residues shown as gray dashed lines.



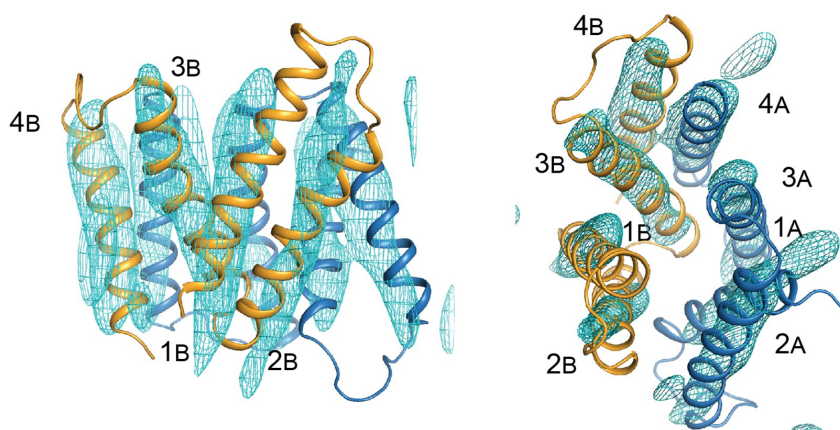
Supplementary Figure B.11: Sequence alignments of five representative Gdx proteins (from top to bottom: Clostridiales bacterium oral taxon 876, *E. coli*, *Micromonospora*, *Streptomyces tsukubensis*, and *Leifsonia aquatica*) and five representative Qac proteins (from top to bottom: *E. coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Mycobacterium bovis*, and *Bordetella avium*).

Sequence numbering corresponds to EmrE. Sequences are colored according to sequence conservation (shades of blue). Residues that contribute to the binding pocket and that are conserved between the Qac and Gdx subtypes are highlighted in orange. Residues that contribute to the binding pocket and that differ in the Qac and Gdx subtypes are highlighted in black.

a Crystal structure overlaid with EmrE EM density from Ubarretxena-Belandia et al.(2003).

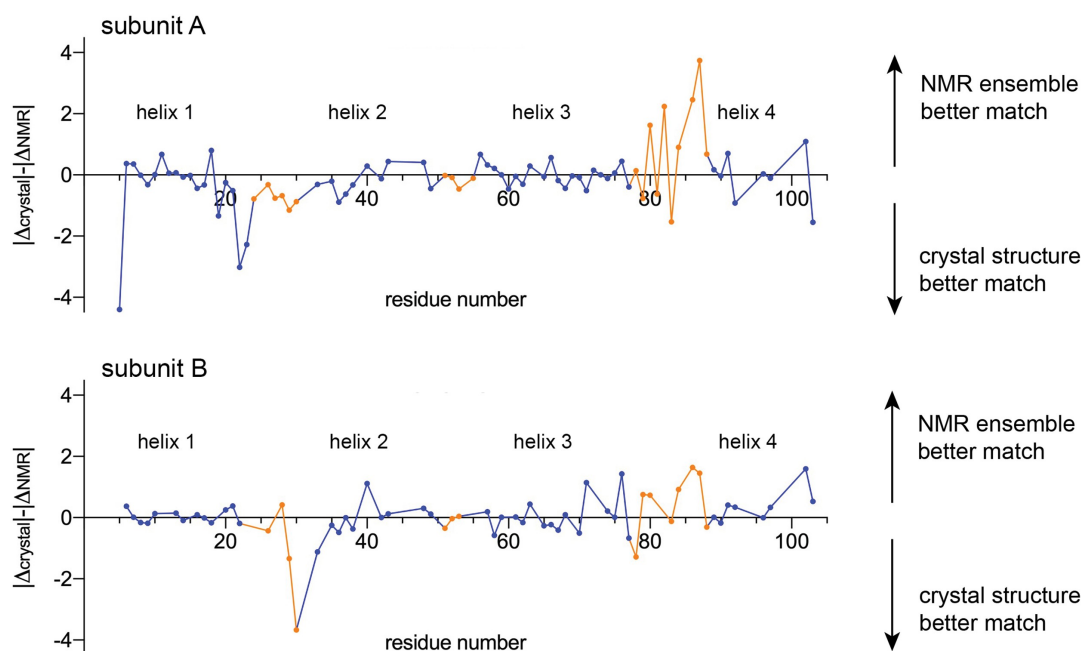


b NMR model (PDB:7JK8; Shcherbakov et al., 2021) overlaid with EM density (Ubarretxena-Belandia et al., 2003)



Supplementary Figure B.12: (A) Crystal structure of EmrE₃ (orange and blue cartoon) overlaid with experimental electron microscopy density (cyan mesh contoured at 1.5 σ) (Ubarretxena-Belandia et al., 2003).

(Panel repeated from Figure 2—figure supplement 2 to aid visual comparison). (B) NMR model of EmrE S64V (orange and blue cartoon) (Shcherbakov et al., 2021) overlaid with experimental electron microscopy density shown in panel A (cyan mesh contoured at 1.5 σ) (Ubarretxena-Belandia et al., 2003).

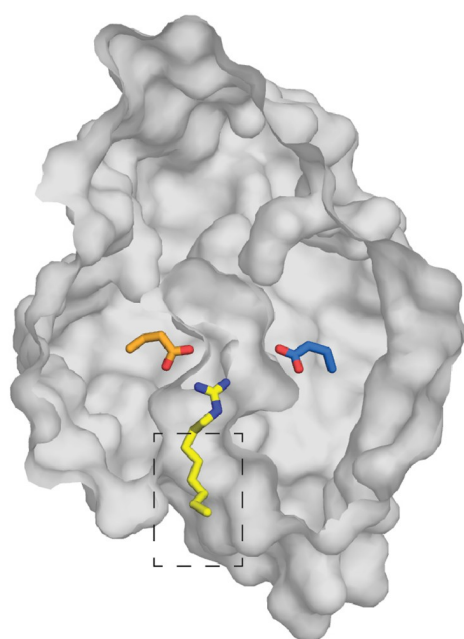
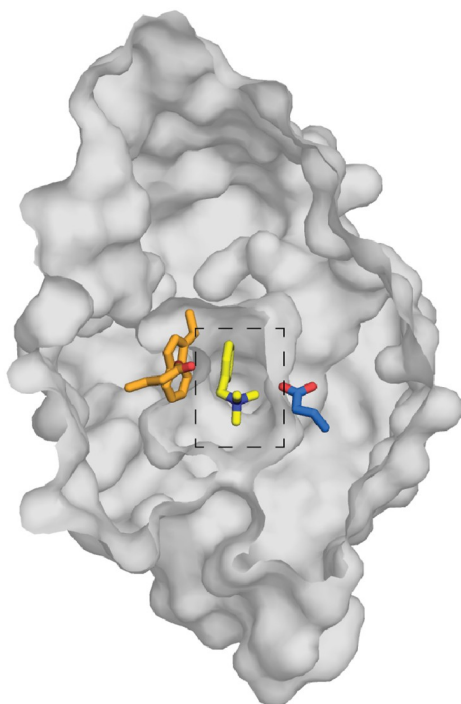


Supplementary Figure B.13: Comparison of experimental chemical shifts for EmrE (BMRB accession number 50411) with chemical shifts predicted from the crystallography model and NMR ensemble using LARMORC α (Frank et al., 2015).

Residue number is plotted along the x-axis. The y-axis compares the relative difference between the experimental chemical shifts and the predicted chemical shifts for the crystallography and NMR models. For each C α position, the difference between the predicted and experimental chemical shifts was calculated ($\delta_{\text{predicted, NMR model}} - \delta_{\text{experimental}} = \Delta_{\text{NMR}}$ and $\delta_{\text{predicted, crystallography model}} - \delta_{\text{experimental}} = \Delta_{\text{crystal}}$), and their relative magnitude compared ($|\Delta_{\text{crystal}}| - |\Delta_{\text{NMR}}|$). Values above the origin line indicate that the experimental chemical shifts are in better agreement with the predicted chemical shifts for the NMR model; values below the origin line indicate that the experimental chemical shifts are in better agreement with the predicted chemical shifts for the crystallography model. Residues in TM helices are shown as blue points, and residues in loop regions are shown as orange points. Residues that were not assigned in the NMR dataset, or that are mutated in either the NMR or crystal structures (E25, W31, V34, S64) are absent from this plot.

EmrE structure: Benzyltrimethylammonium

Gdx-Clo structure: octylguanidinium



Supplementary Figure B.14: Top down structures of EmrE in complex with benzyltrimethylammonium (PDB:7T00; model for benzalkonium headgroup binding) and Gdx-Clo in complex with octylguanidinium (PDB:6WK9; model for alkyl tail positioning).

Structures are sliced at the midpoint of the membrane, as in Figure 7. Dashed boxes indicate the headgroup and alkyl group positions used to prepare the hypothetical model of benzalkonium binding.

	EmrE ₃ /L10/MeTPP [*]	EmrE ₃ /L10/TPP [*]	EmrE ₃ /L10/harmane	EmrE ₃ /L10/methyl viologen	EmrE ₃ /L10, pH 5.2	Gdx-Clo/L10, pH 5.0	EmrE ₃ /L10/BM ₃ A ⁺
Crystallization conditions	0.1 M LiNO ₃ , 0.1 M ADA pH 6.5, 32.8% PEG 600	0.1 M (NH ₄) ₂ SO ₄ , 0.1 M HEPES pH 7.25, 30.8% PEG 600	0.1 M (NH ₄) ₂ SO ₄ , 0.1 M HEPES pH 7.1, 33.8% PEG 600	0.1 M (NH ₄) ₂ SO ₄ , 0.1 M ADA pH 6.3, 34.8% PEG 600	0.2 M NaCl, 0.1 M sodium cacodylate pH 5.2, 34% PEG 600	0.1 M calcium acetate, 0.1 M sodium acetate pH 5.0, 40% PEG 600	0.1 M NH ₄ SO ₄ , 0.1 M HEPES pH 7.25, 33% PEG 600
Data collection							
Space group	C121	C121	C121	P1	C121	P1	C121
Cell dimensions							
a, b, c (Å)	141.17, 50.87, 110.79	140.71, 50.14, 110.28	145.7, 51.83, 114.95	50.91, 75.07, 111.43	140.64, 49.85, 109.83	49.70, 74.32, 107.43	140.18, 50.12, 110.73
α, β, γ (°)	90, 92.69, 90	90, 93.45, 90	90, 92.67, 90	92.03, 90.33, 109.20	90, 93.75, 90	93.56, 89.71, 109.92	90, 92.79, 90
Resolution (Å)	70.5–3.22 (3.42–3.22)	70.2–3.36 (3.62– 3.36)	114.8–3.75 (4.37–3.75)	70.8–3.13 (3.41–3.13)	70.2–2.85 (3.16– 2.85)	107.2–2.32 (2.67–2.32)	70.50–3.22 (3.42–3.22)
Ellipsoidal Resolution Limit (best/worst)*	3.22/4.33	3.36/5.1	3.75/6.34	3.13/4.50	2.85/3.72	2.32/3.55	3.22/4.33
% Spherical Data Completeness*	69.0 (20.9)	54.5 (13.6)	44.0 (10.1)	52.0 (11.1)	62.0 (12.0)	41.9 (6.0)	69.0 (20.9)
% Ellipsoidal Data Completeness*	88.6 (80.1)	84.1 (78.5)	82.7 (65.7)	82.0 (72.3)	87.0 (62.6)	80.3 (45.6)	88.6 (80.1)
R _{merge} [*]	0.152 (0.656)	0.349 (1.053)	0.365 (0.752)	0.123 (0.697)	0.118 (1.85)	0.089 (0.4)	0.152 (0.656)
R _{meas} [*]	0.166 (0.707)	0.384 (1.15)	0.396 (0.817)	0.144 (0.814)	0.129 (1.99)	0.104 (0.465)	0.166 (0.707)
CC _{1/2}	0.967 (0.861)	0.779 (0.610)	0.992 (0.862)	0.939 (0.629)	0.994 (0.366)		0.967 (0.861)
Mn I / σI	10.4 (2.7)	4.0 (1.8)	7.4 (2.3)	7.7 (1.5)	9.5 (1.2)	6.5 (2.8)	10.4 (2.7)
Multiplicity [*]	6.6 (7.1)	5.9 (6.2)	6.7 (6.6)	3.7 (3.8)	6.4 (7.1)	3.8 (3.8)	6.6 (7.1)
Refinement							
Resolution (Å)	55.3–3.22	55.0–3.36	60.2–3.91	32.9–3.13	35.2–2.85	35.5–2.32	55.3–3.91
No. reflections	8,025	6,097	3,347	14,194	11,149	26,026	5,040
R _{work} / R _{free}	29.4 / 33.4	29.0/31.4	34.2/34.4	30.0/33.1	30.7/32.7	25.1/29.5	33.0/36.7
Ramachandran Favored	89.4	89.6	90.9	89.1	91.0	92.9	88.7
Ramachandran Outliers	1.9	1.9	2.4	2.6	1.9	1.5	2.7
Clashscore	11.8	13.6	8.4	16.8	8.6	10.4	14.9
R.m.s. deviations							
Bond lengths (Å)	0.003	0.003	0.003	0.004	0.003	0.004	0.004
Bond angles (°)	0.70	0.68	0.60	0.82	.65	0.70	0.67
Coordinates in Protein Databank	7SSU	7SV9	7SVX	7MGX	7MH6	7SZT	7T00

*Where applicable, values reported are for anisotropically truncated data performed using the Staraniso webserver (Global Phasing). See *Methods* for details.

Table 4: Data collection, phasing, and refinement statistics for EmrE and Gdx-Clo complexes.

	Prep 1/Sensor 1		Prep 1/Sensor 2		Prep 2/Sensor 1	
	TPA ⁺	PheGdm ⁺	TPA ⁺	PheGdm ⁺	TPA ⁺	PheGdm ⁺
No protein	0	0	0	0	0	0
WT	-4.8	-1.4	-4.1	-1.2	-3.9	-1.0
Y60F	-0.14	-0.03	-0.05	-0.07	-0.04	-0.05
S43A	-3.7	-1.6	-3.9	-1.3	-3.2	-1.2
W63F	-5.4	-2.0	-4.6	-1.5	-4.0	-1.0
Gdx-Clo						
	Prep 1/Sensor 1 (Gdm ⁺)		Prep 1/Sensor 2 (Gdm ⁺)		Prep 2/Sensor 1 (Gdm ⁺)	
No protein	0		0		0	
WT	-6.3		-6.7		-6.3	
Y60F	0.04		0.007		0.6	
S43A	-0.15		-0.15		-0.15	
W63F	-0.60		-0.33		-0.30	

Table 5: SSM electrophysiology peak currents (nA) for EmrE3 and Gdx-Clo mutants summarized by experimental replicate.

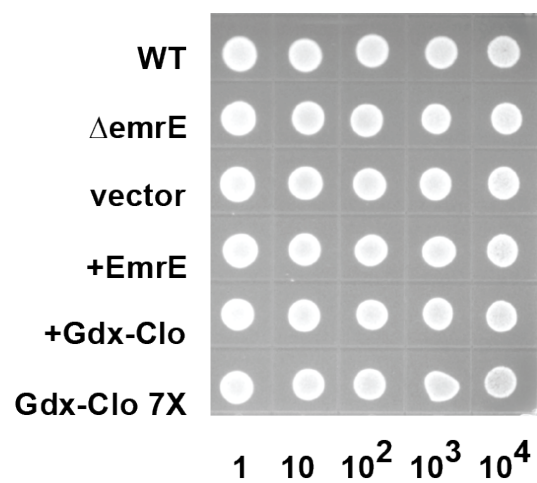
Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Gene (<i>Escherichia coli</i>)	EmrE ₃	Uniprot	P23895	Bears mutations E25N, W31I, V34M to bind monobody (this paper – see Figure 1)
Gene (<i>Clostridiales</i> bacterium oral taxon 876)	Gdx-Clo	GenBank	ERI95081.1	PMID: 33247110
Recombinant DNA reagent	EmrE ₃ in pET15b (plasmid)	This publication		Expression vector for EmrE ₃ . Available upon request.
Recombinant DNA reagent	Gdx-Clo in pET21c (plasmid)	PMID: 33247110		Expression vector for Gdx-Clo. Available upon request.
Chemical compound, drug	<i>E. coli</i> polar lipids	Avanti, Alabaster, AL	#100600 C	
Chemical compound, drug	n-decyl- β -D-maltopyranoside	Anatrace, Maumee, OH	D322	
Recombinant DNA reagent	Monobody L10 in pHBT1 (plasmid)	PMID: 33247110		Expression vector for monobody L10. PMID: 33247110 . Addgene ID: 183,406

Table 6: Key resources table

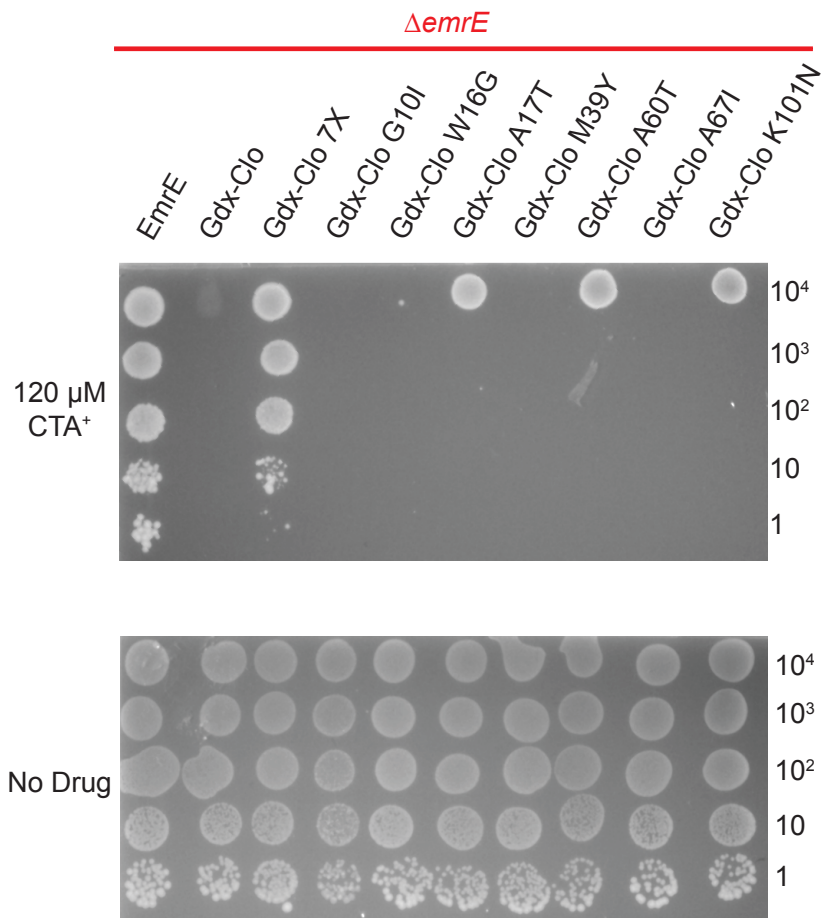
Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Kermani AA, Stockbridge RB	2022	Structure of EmrE-D3 mutant in complex with monobody L10 in low pH (protonated state)	https://www.rcsb.org/structure/7MH6	RCSB Protein Data Bank, 7MH6
Kermani AA, Stockbridge RB	2022	Structure of EmrE-D3 mutant in complex with monobody L10 and methyl viologen	https://www.rcsb.org/structure/7MGX	RCSB Protein Data Bank, 7MGX
Kermani AA, Stockbridge RB	2022	Structure of EmrE-D3 mutant in complex with monobody L10 and harmane	https://www.rcsb.org/structure/7SVX	RCSB Protein Data Bank, 7SVX
Kermani AA, Stockbridge RB	2022	Structure of EmrE-D3 mutant in complex with monobody L10 and methyltriphenylphosphonium	https://www.rcsb.org/structure/7SSU	RCSB Protein Data Bank, 7SSU
Kermani AA, Stockbridge RB	2022	Structure of EmrE-D3 mutant in complex with monobody L10 and TPP	https://www.rcsb.org/structure/7SV9	RCSB Protein Data Bank, 7SV9
Kermani AA, Stockbridge RB	2022	Structure of EmrE-D3 mutant in complex with monobody L10 and benzyltrimethylammonium	https://www.rcsb.org/structure/7T00	RCSB Protein Data Bank, 7T00
Kermani AA, Stockbridge RB, Burata OE	2022	Crystal structure of Gdx-Clo from Small Multidrug Resistance family of transporters in low pH (protonated state)	https://www.rcsb.org/structure/7SZT	RCSB Protein Data Bank, 7SZT

Table 7:Generated datasets

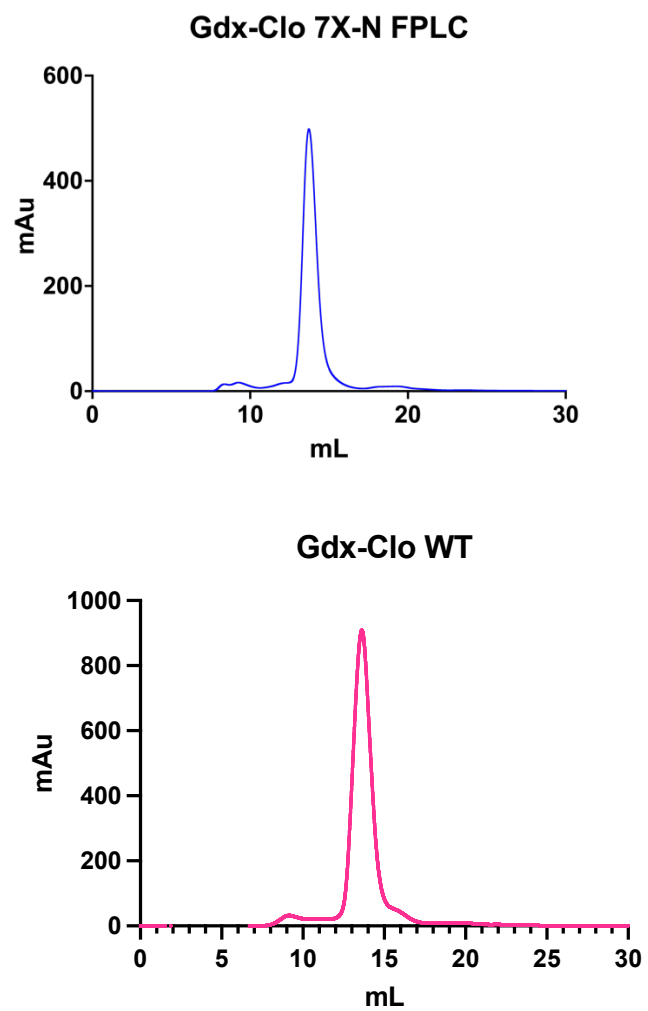
Appendix C: Supporting Information for Chapter 4



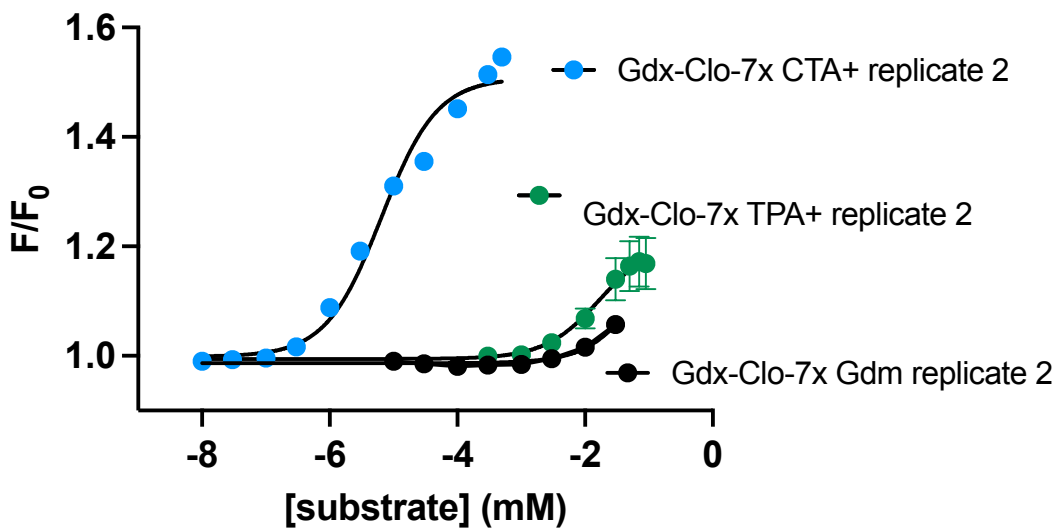
Supplementary Figure C.1: No-drug control for bacterial dilution assays in Figure 4A.



Supplementary Figure C.2: Spotting assay of individual mutations found in Gdx-Clo 7X variant when grown in 120 μ M CTA⁺



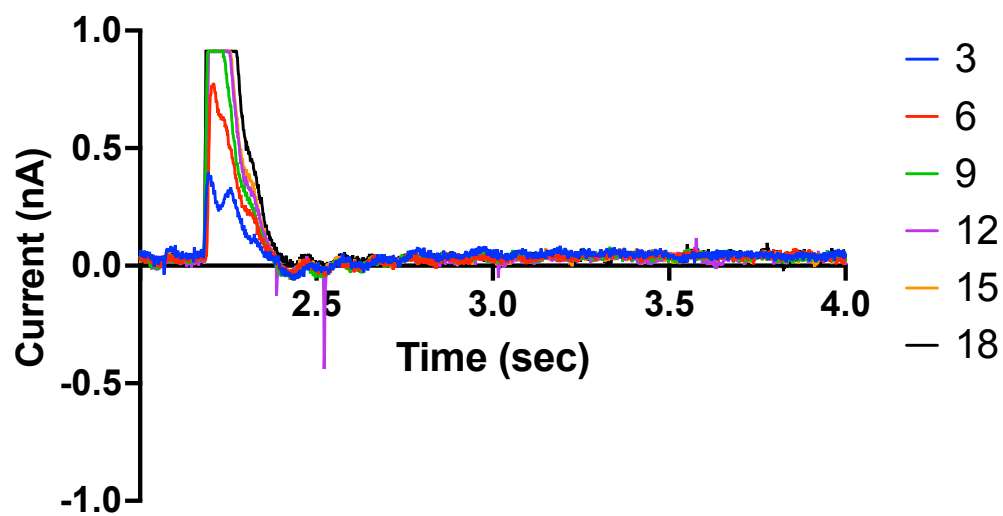
Supplementary Figure C.3: FPLC profiles for Gdx-Clo WT and Gdx-Clo-7X purification



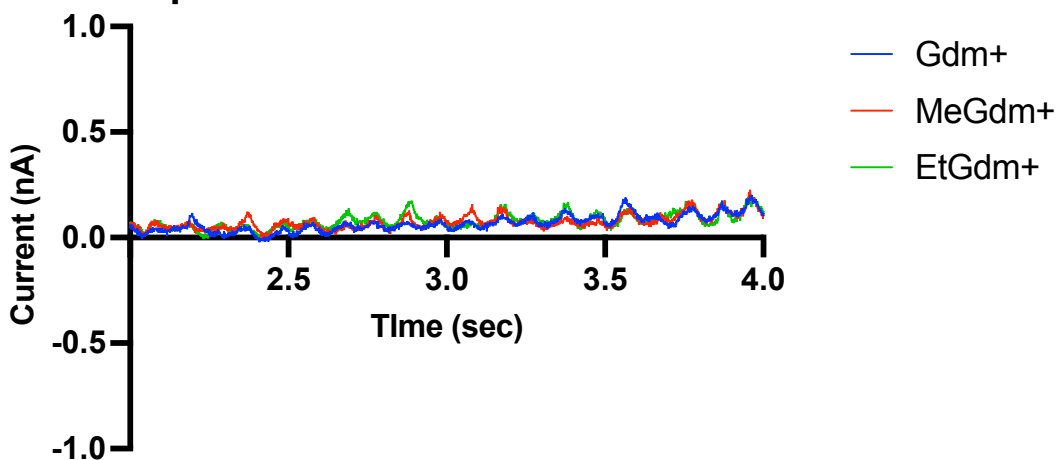
Supplementary Figure C.4: Second independent protein preparation of Gdx-Clo-7X: binding to Gdm⁺, CTA⁺, or TPA⁺, as indicated.

Solid lines show fits to a single site binding isotherm with K_d values of 6.5 μM for CTA⁺ and 17.5 mM for TPA⁺.

No protein control: TPA titration

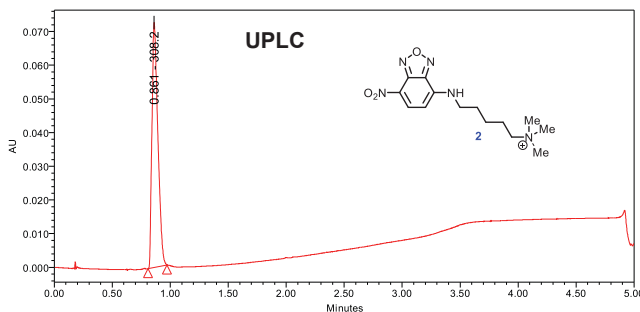
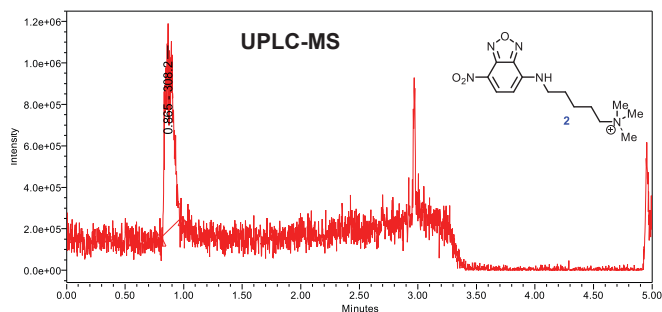
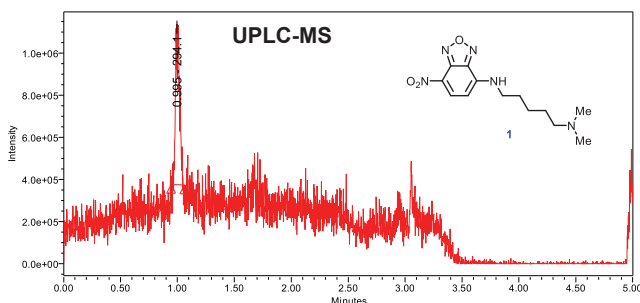
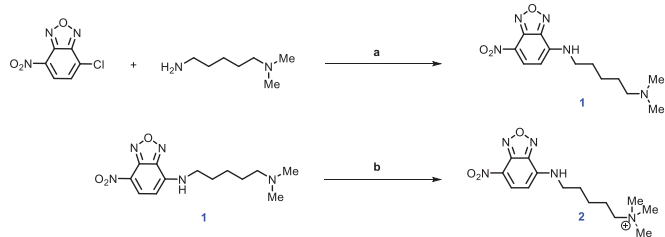


No protein control: Gdm⁺ substrates



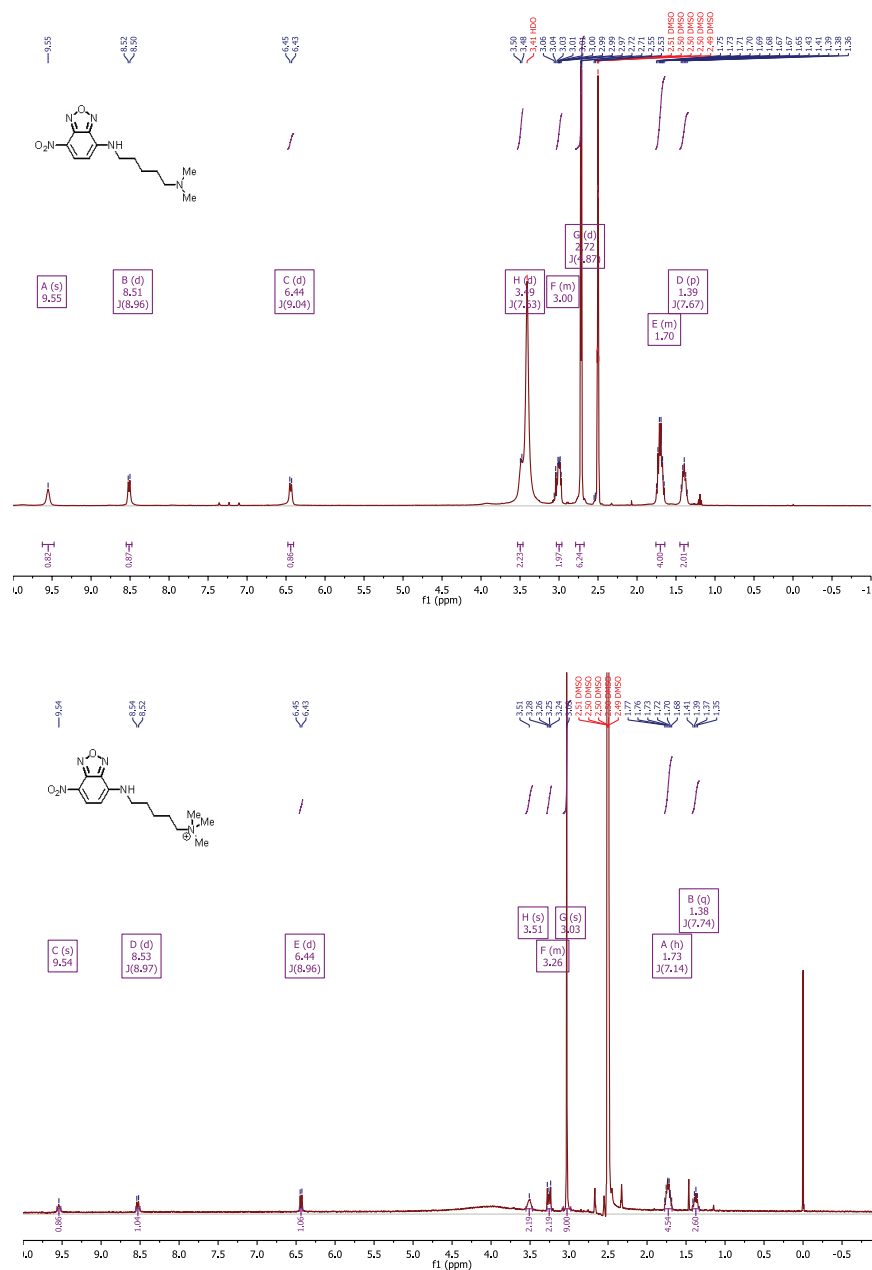
Supplementary Figure C.5: Solid supported membrane electrophysiology: no protein controls.

TPA⁺ was titrated from 3-18 mM as indicated. Gdm⁺ and substituted guanidiniums were perfused at 1 mM.



Supplementary Figure C.6: LCMS analysis of NBD-CTA⁺ synthesis

Reagents and conditions: **(a)** 4-chloro-7-nitrobenzo[c][1,2,5]oxadiazole (1 equiv), N¹,N¹-dimethylpentane-1,5-diamine (1 equiv), Et₃N (2 equiv), Anhydrous DMF, 90°C, 3 hrs, 48 %; **(b)** N¹,N¹-dimethyl-N5-(7-nitrobenzo[c][1,2,5]oxadiazol-4-yl)pentane-1,5-diamine (1 equiv), Et₃N (0.1 equiv), Methyl Iodide (5 equiv), DCM, rt, 1hr, 100%. All final compounds have purity ≥95% as determined by Waters ACQUITY ultra-performance liquid chromatograph (UPLC) using reverse phase column (SunFire, C18, 5 μm, 4.6 × 150 mm²) and a solvent gradient of A (H₂O with 0.1% of TFA) and solvent B (CH₃CN with 0.1% of TFA).

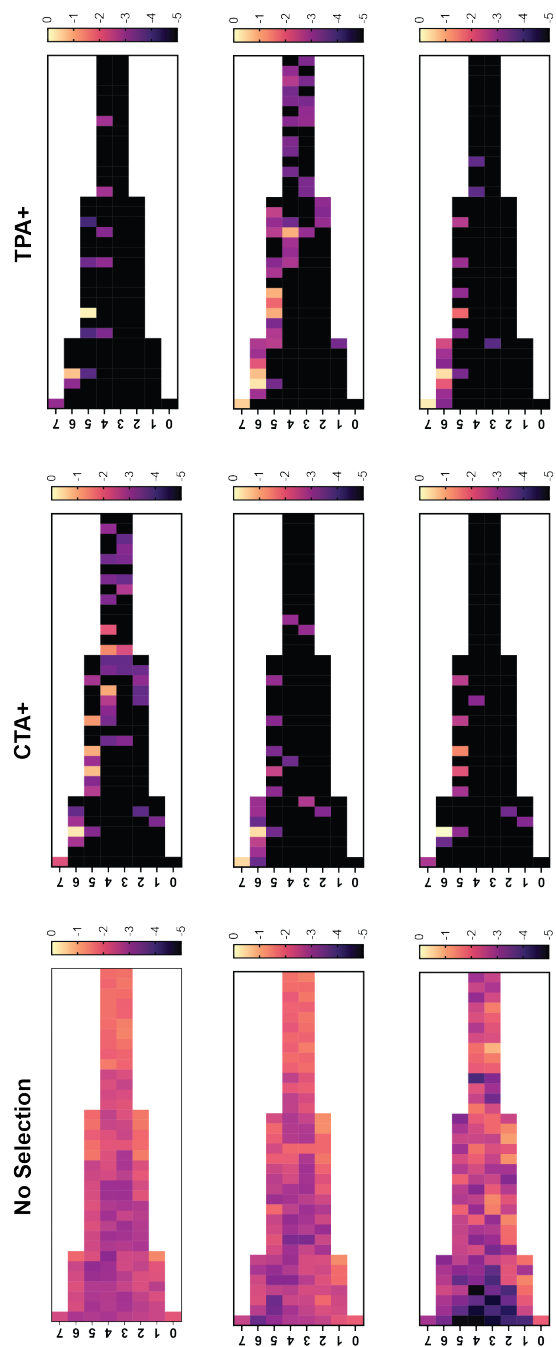


Supplementary Figure C.7: NMR analysis of NBD-CTA⁺ synthesis

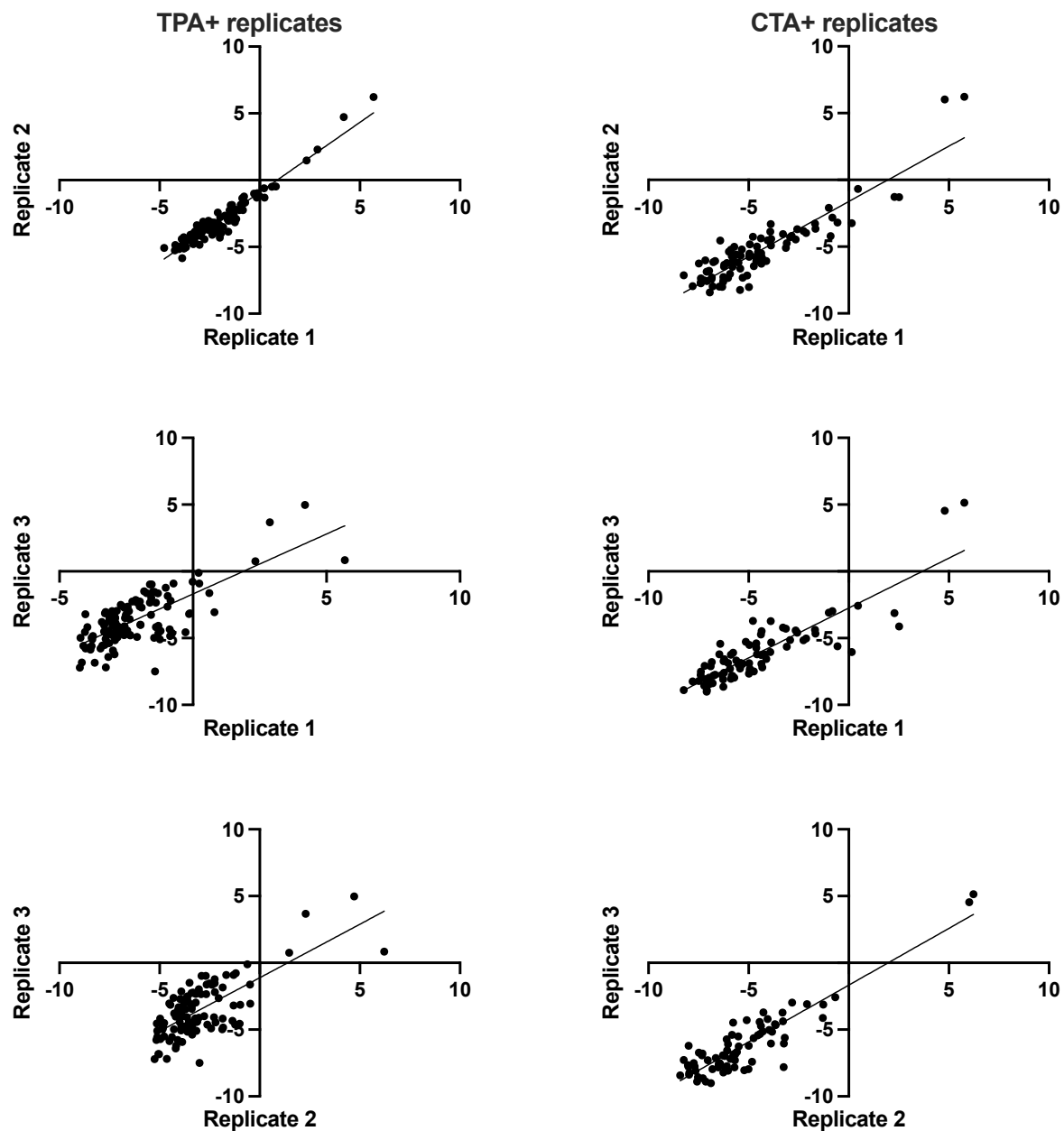
In the spectral data reported, the format (δ) chemical shift (multiplicity, J values in Hz, integration) was used with the following abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet. Electrospray ionization (ESI) mass spectral (MS) analysis was performed on a Thermo Scientific LCQ Fleet mass spectrometer. The final products were purified by reverse phase HPLC (RP-HPLC) with solvent A (0.1% of TFA in water) and solvent B (0.1% of TFA in CH₃ CN) as eluents with a flowrate of 45 mL/min.

Oligo #	Gene Block Sequences	Fragment	Mutations	Tm
1	GGCTAGCAGGAGGAATTCACCATGGCGTGGCTGATCCTGATCATTGCGGGTATTTTCGAGGTGGTTTGGCGATCGC GCTGAAGTACAGCAACGGTTTTACCCGCTCTGATCCCGAGCATGATCACCCGTATTGGC	1	None	56-57°C
2	GGCTAGCAGGAGGAATTCACCATGGCGTGGCTGATCCTGATCATTGCGATCATTTTCGAGGTGGTTTGGCGATCGC GCTGAAGTACAGCAACGGTTTTACCCGCTCTGATCCCGAGCATGATCACCCGTATTGGC	1	G10I	56-57°C
3	GGCTAGCAGGAGGAATTCACCATGGCGTGGCTGATCCTGATCATTGCGGGTATTTTCGAGGTGGTTTGGCGATCGC GCTGAAGTACAGCAACGGTTTTACCCGCTCTGATCCCGAGCATGATCACCCGTATTGGC	1	W16G	56-57°C
4	GGCTAGCAGGAGGAATTCACCATGGCGTGGCTGATCCTGATCATTGCGGGTATTTTCGAGGTGGTTTGGACGATCGC GCTGAAGTACAGCAACGGTTTTACCCGCTCTGATCCCGAGCATGATCACCCGTATTGGC	1	A17T	56-57°C
5	GGCTAGCAGGAGGAATTCACCATGGCGTGGCTGATCCTGATCATTGCGATCATTTTCGAGGTGGTTTGGCGATCGC GCTGAAGTACAGCAACGGTTTTACCCGCTCTGATCCCGAGCATGATCACCCGTATTGGC	1	G10I, W16G	56-57°C
6	GGCTAGCAGGAGGAATTCACCATGGCGTGGCTGATCCTGATCATTGCGATCATTTTCGAGGTGGTTTGGACGATCGC GCTGAAGTACAGCAACGGTTTTACCCGCTCTGATCCCGAGCATGATCACCCGTATTGGC	1	G10I, A17T	56-57°C
7	GGCTAGCAGGAGGAATTCACCATGGCGTGGCTGATCCTGATCATTGCGGGTATTTTCGAGGTGGTTTGGACGATCGC GCTGAAGTACAGCAACGGTTTTACCCGCTCTGATCCCGAGCATGATCACCCGTATTGGC	1	W16G, A17T	56-57°C
8	GGCTAGCAGGAGGAATTCACCATGGCGTGGCTGATCCTGATCATTGCGATCATTTTCGAGGTGGTTTGGACGATCGC GCTGAAGTACAGCAACGGTTTTACCCGCTCTGATCCCGAGCATGATCACCCGTATTGGC	1	G10I, W16G, A17T	56-57°C
9	CCGAGCATGATCACCCGTATTGGCATGCTGATTAGCTTCTACCTGCTGAGCCAAGCGACCAAGACCCCTGCCGATTGG TACCCGCTATGCGATCTGGACCGGTATTGGCGCGCTGGGTGCGGTGATTGGCGGCATCATTTTC	2	None	55-56°C
10	CCGAGCATGATCACCCGTATTGGCATGCTGATTAGCTTCTACCTGCTGAGCCAAGCGACCAAGACCCCTGCCGATTGG TACCCGCTATGCGATCTGGACCGGTATTGGCGCGCTGGGTGCGGTGATTGGCGGCATCATTTTC	2	M39Y	55-56°C
11	CCGAGCATGATCACCCGTATTGGCATGCTGATTAGCTTCTACCTGCTGAGCCAAGCGACCAAGACCCCTGCCGATTGG TACCCGCTATACGATCTGGACCGGTATTGGCGCGCTGGGTGCGGTGATTGGCGGCATCATTTTC	2	A60T	55-56°C
12	CCGAGCATGATCACCCGTATTGGCATGCTGATTAGCTTCTACCTGCTGAGCCAAGCGACCAAGACCCCTGCCGATTGG TACCCGCTATGCGATCTGGACCGGTATTGGCATCTCGGGTGCGGTGATTGGCGGCATCATTTTC	2	A67I	55-56°C
13	CCGAGCATGATCACCCGTATTGGCATGCTGATTAGCTTCTACCTGCTGAGCCAAGCGACCAAGACCCCTGCCGATTGG TACCCGCTATACGATCTGGACCGGTATTGGCGCGCTGGGTGCGGTGATTGGCGGCATCATTTTC	2	M39Y, A60T	55-56°C
14	CCGAGCATGATCACCCGTATTGGCATGCTGATTAGCTTCTACCTGCTGAGCCAAGCGACCAAGACCCCTGCCGATTGG TACCCGCTATGCGATCTGGACCGGTATTGGCATCTCGGGTGCGGTGATTGGCGGCATCATTTTC	2	M39Y, A67I	55-56°C
15	CCGAGCATGATCACCCGTATTGGCATGCTGATTAGCTTCTACCTGCTGAGCCAAGCGACCAAGACCCCTGCCGATTGG TACCCGCTATACGATCTGGACCGGTATTGGCATCTCGGGTGCGGTGATTGGCGGCATCATTTTC	2	A60T, A67I	55-56°C
16	CCGAGCATGATCACCCGTATTGGCATGCTGATTAGCTTCTACCTGCTGAGCCAAGCGACCAAGACCCCTGCCGATTGG TACCCGCTATACGATCTGGACCGGTATTGGCATCTCGGGTGCGGTGATTGGCGGCATCATTTTC	2	M39Y, A60T, A67I	55-56°C
17	GTGATTTGCGGCATCATTTTCTTTAAAGAACCGCTGACCGCGCTGCGTATCGTTTTATGATTCTGCTGCTGACCGGTA TCATTGGCCTGAAAGCGACCAAGCAAGCTAAAGCTTGGCTGTTTGGCGGATGAGAGAAG	3	None	56-57°C
18	GTGATTTGCGGCATCATTTTCTTTAAAGAACCGCTGACCGCGCTGCGTATCGTTTTATGATTCTGCTGCTGACCGGTA TCATTGGCCTGAATGCGACCAAGCAAGCTAAAGCTTGGCTGTTTGGCGGATGAGAGAAG	3	K101N	56-57°C

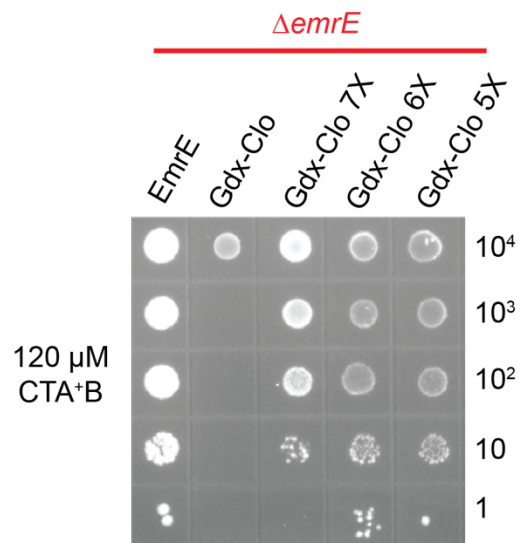
Table 8: Designed oligos for engineering of 128 combinatorial variants of Gdx-Clo 7X



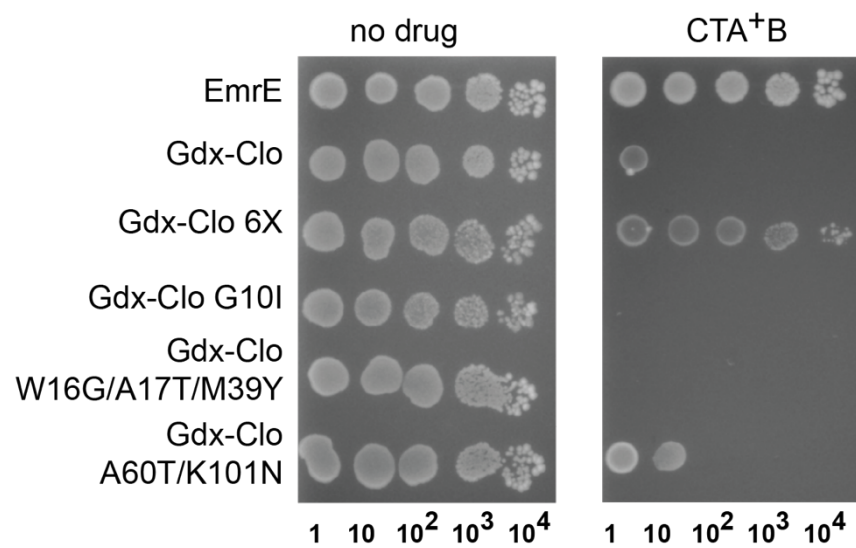
Supplementary Figure C.8: Heat maps showing raw counts for combinatorial library with no drug (also shown in Figure 4), CTA⁺ selection, and TPA⁺ selection.



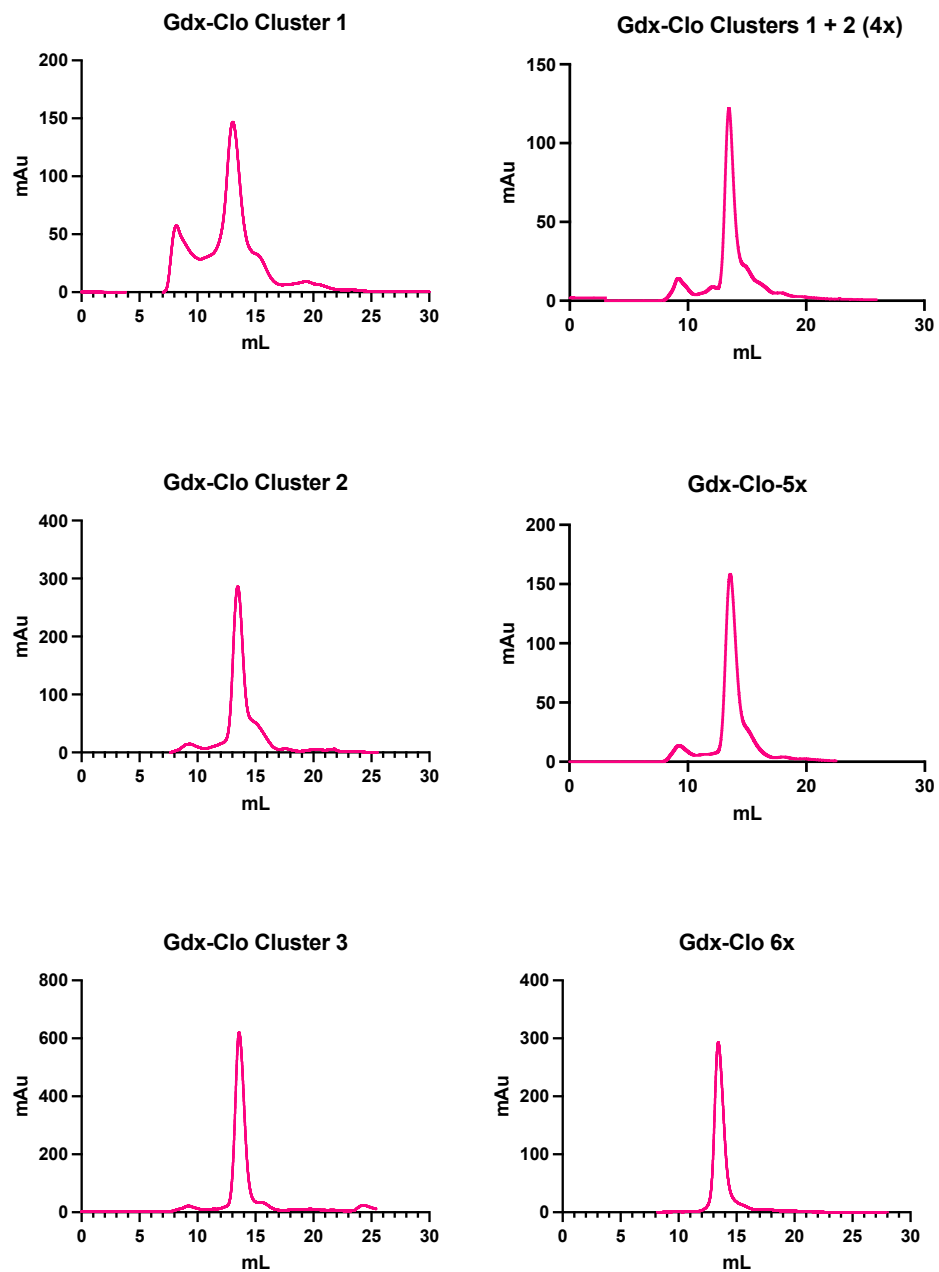
Supplementary Figure C.9: Correlations between replicate selection experiments.



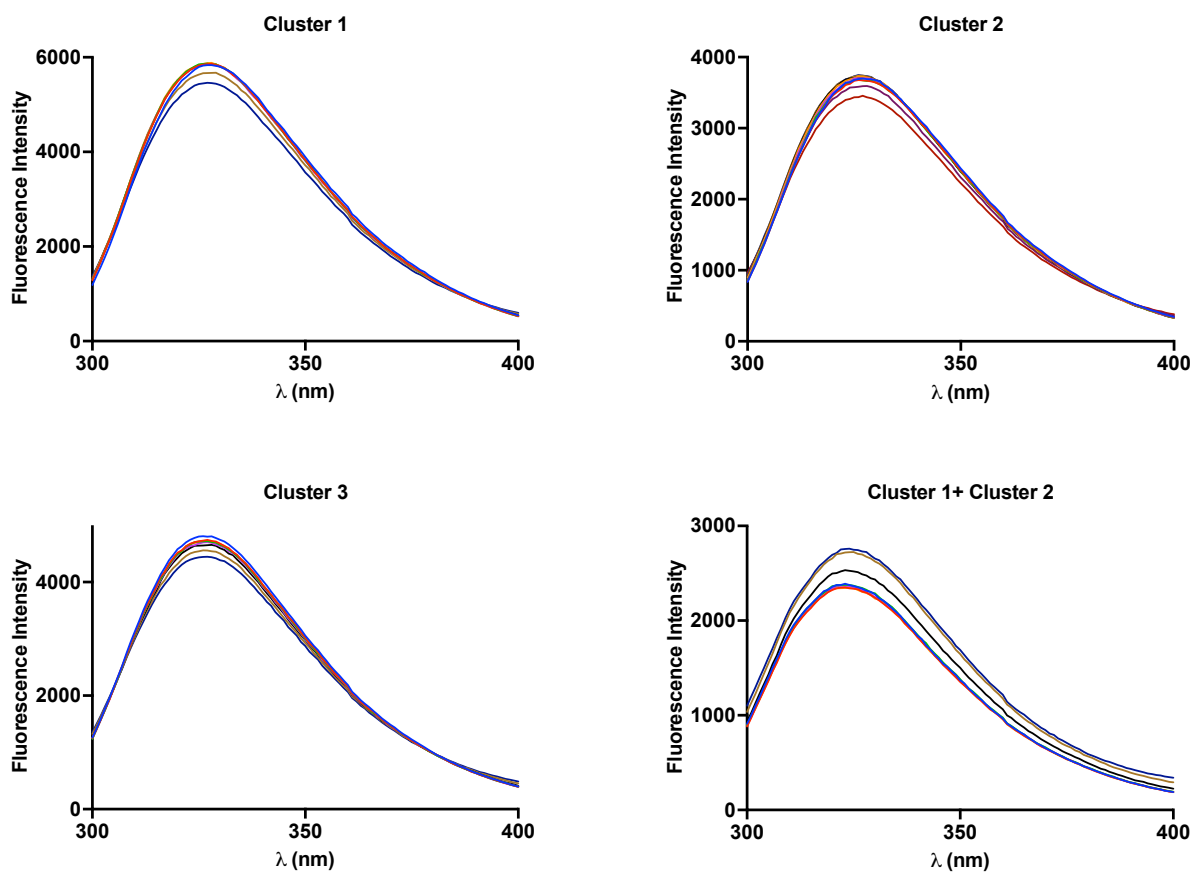
Supplementary Figure C.10: Dilution assays for 5X, 6X, and 7X variants.



Supplementary Figure C.11: Dilution assays of Gdx-Clo cluster 1, cluster 2, and cluster 3 mutations: no-drug control (left) and 120 μ M CTA⁺ (right).



Supplementary Figure C.12: FPLC profiles of mutant Gdx-Clo proteins used for functional analysis: cluster 1, cluster 2, cluster 3, cluster 1+2 (4X), 5X, 6X.



Supplementary Figure C.13: Tryptophan fluorescence spectra as a function of TPA⁺ titration.

Replicate	No selection			TPA ⁺			CTA ⁺		
	1	2	3	1	2	3	1	2	3
Variants observed	128	128	128	128	121	114	106	91	82
Reads	135524	195523	245971	232249	175323	314292	320504	254410	337407
ε (Gdx-Clo)				-2.77	-3.56	-4.55	-7.16	-6.03	-8.08

Table 9: Sequencing statistics for replicate NGS experiments.

Selection	Comparison	R value
TPA ⁺	1 vs 2	0.965
	1 vs 3	0.739
	2 vs 3	0.714
CTA ⁺	1 vs 2	0.904
	1 vs 3	0.866
	2 vs 3	0.906

Table 10: Correlation between enrichment coefficients of replicate selection experiments.

Rank	Fold increase	G10I	W16G	A17T	M39Y	A60T	A67I	K101N
1	6020	•	•	•	•	•		•
2	2560	•	•	•	•	•	•	•
3	410	•	•	•	•	•		
4	250	•	•	•	•			•
5	220	•	•		•	•		•
6	140	•	•	•	•			
7	94	•	•		•	•		
8	26	•		•	•	•	•	•
9	21	•	•		•			•
10	15	•	•	•	•	•	•	
11	15	•	•		•	•	•	•
12	15	•	•	•		•		•
13	12		•	•	•	•		•
14	11	•	•	•		•	•	•
15	11	•	•		•			
16	10	•		•	•	•		•

Table 11: Overrepresented variants (>10-fold) in CTA⁺ selection

Rank	Fold increase	G10I	W16G	A17T	M39Y	A60T	A67I	K101N
1	180	●	●	●	●	●	●	●
2	50	●	●	●	●	●		●
3	42	●	●	●	●	●		
4	28	●	●	●	●	●	●	
5	5		●	●	●	●	●	●
6	4	●	●	●	●		●	
7	4	●	●	●	●			
8	2	●	●	●	●			●

Table 12: Overrepresented variants (>2-fold) in TPA⁺ selection