Complexity in the Membrane: The Fluc family of Fluoride Channels and Small Multidrug Resistance Family of Transporters as Models for Understanding Membrane Protein Structural and Functional Evolution

by

Christian Bernard Macdonald

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in the University of Michigan
2021

Doctoral Committee:

Assistant Professor Randy Stockbridge, Chair
Assistant Professor Aaron T. Frank
Professor Ayyalusamy Ramamoorthy
Associate Professor Stephen A. Smith

Christian Bernard Macdonald

cbmacdo@umich.edu

ORCID iD: 0000-0002-0201-8832

DEDICATION

To Roberta Helmer Stalberg and Mary Ellen Hennessy.

ACKNOWLEDGEMENTS

I have many more people that I should thank than I actually can. The focus on publications as the objects of scientific research gives the impression that the only people who matter are the list of authors. In reality, science is a process, just like getting a PhD. Without support along the way and receptive people to develop ideas with, none of this would be possible. For every name, there are dozens who are essential but unnamed.

First, I must thank Randy for being a wonderful mentor and advisor during my time at Michigan, and for introducing me to membrane proteins. I feel extremely lucky to have had such an exceptional model of a scientist here. In her rigor, creativity, and kindness, I have found few her equal. Her scientific and personal generosity as well are rare. I can't wait to see what comes next, but I know it will be great.

I also want to thank Rams for providing me a first scientific home at Michigan, and a great place to start doing science here. His support and mentorship over the past years and your scientific guidance have been key for completing my PhD. I also thank the other members of my committee, Dr. Aaron Frank and Dr. Stephen Smith, for providing me with consistently useful feedback and advice, and critically helping to shape this dissertation.

Every member of the staff I've met at Michigan has been exceptionally generous and kind, as well, even when I'm confused or tardy or making some mistake or another. In the Biophysics office, Sara Grosky and Sandra Moing have been great for making sure I'm on track and completing what needs to be done, and probably too patient when I inevitably am behind. In MCDB, my contacts have been fewer, but I deeply appreciate the patience of Jacqueline Glebe whenever I was confused about some ordering process, Robert Heller and IT whenever something went wrong with a computer, and Mark Matusko when something went wrong with something that wasn't a computer. Joe Schauerte and Kathleen Wisser were always helpful with instruments and science. Without these efforts, none of this would have been possible, so I give my deep thanks.

Even before I arrived at Michigan, my mentors at ASU were exceptionally generous with their time and energy for me. Dr. George Pettit let me into his lab, where I worked with a talented postdoc Dr. Pablo Arce and Dr. Noeleen Melody on several projects. I will always remember his office and that laboratory. Dr. Xu Wang was remarkably generous with his time by spending a few hours a week helping me figure out density matrices and NMR spin dynamics. I am profoundly lucky that some of my first contacts were as kind and thoughtful as they were.

I would not have been able to complete this process without the support of my colleagues and friends in labs. The Stockbridge lab has been a great place to learn and think and do, and this is largely because of the people in it: Dr. Ali Kermani and Dr. Ben

McIlwain have served as excellent mentors and sounding boards over the years, and Dr. Jason Devlin has been a great and thoughtful addition. My fellow graduate students have helped make all this work too: Olive Burata, Rachael Lucero, Kamirah Demouchet, Chia-Yu Kang, and Trevor Yeh. Roja Gundepudi and Ben Koff have been great techs and colleagues as well, and I will be and was sad to leave them. Fox Baudelaire has been great to work with, even if for a relatively short amount of time.

I have been lucky to work with a series of talented undergraduates, too. Paul and Ali in the Rams lab were very kind to put up with me when I knew even less than I do now. Vivek Parikh was a great assistant over a few years, and I really regret that covid meant the last year didn't take place. Troy Cao was a seemingly tireless and insightful peer, and his work was essential for getting my final project off the ground. I haven't worked as closely with the rest of the undergraduates in the Stockbridge lab, but Jesse Mangat, Alain Sullivan, Alex Davis and Michal Ruprecht have been great to have around.

The best thing about my time here has been the friends I've made, and the worst thing has been that everyone (including me) ends up leaving. David, Jeremy, Leslie, Hannah, Alex, Emily, Katy, Domenic, Guy, and Emma: I will never be able to repay you for the kindness and friendship you have shown me.

My colleagues at Michigan have been extremely excellent. My cohort was a great group that helped me make it through the first few years and beyond: Hayden, Maral, Chu, and Desmond: thank you. My friends in the Ramamoorthy lab have been the best I could ask for. Sam Kotler was truly an excellent mentor during my first experiences with research in biology, and I've tried to model myself after his patience and positivity. Kyle Korshavn was a kind and thoughtful human being whose productivity and intelligence I could only hope to approach. Sarah Cox was an endlessly positive and energetic worker whose diligence helped us all. Katie Gentry helped me through many a difficult experiment with her experience and willingness to gossip. Amit, Nate, Giacomo, Kian, Josh, and Niru, my fellow PhD students, helped make the work bearable. Elke, Vojc, Mukesh, Thiru, Tomo and Bikash were excellent examples of great scientists and workers, and I appreciated their intelligence and energy.

I have been lucky enough to find a second family in the last few years: the Hennessy/Tilka clan. I am so lucky to have found such kind, generous people who have welcomed me into their lives and made me feel at home. Susan, Michael, Bridget, Nora, Kevin, Angela, Catherine, William, and Malcolm: thank you for your love and support.

I wish my Uncle Sidney and Aunt Georgia were here for this, but their examples and love helped me along my way. Thank you Wayne for your support and example. The rest of the Helmer family has always shown that generosity, kindness, and hard work can pay off. The Stalbergs have given me their love and helped my journey.

Thank you Dad, for your continuing love and support. I hope I can make you proud and live up to your example. I love you. Mom, I wish you were here for this day, but I hope I can keep your spirit of goodness and love alive in the world. I hope Yuki and Ignatz are with you now. Finally, the most important person: Mary. There is no way I would have been able to get anywhere near completing this without your tireless love and presence. There is nobody else I would rather have completed a PhD in a global pandemic with, and nobody else I could imagine spending a life with. I love you.

TABLE OF CONTENTS

List of Figures

FIGURE

List of Appendices

ABSTRACT

The cellular membrane comprises the barrier between self and environment, the development of which laid the foundation for the individual organism. The membranes of organisms today are packed full of molecular machines which mediate communication, metabolism, and interaction between the homeostatic internal environment and the uncontrolled environments they find themselves in. Understanding the molecular bases for their function, as well as the evolutionary origins of their roles, is an important task for understanding modern microbial biology.

This thesis examines two families of membrane proteins in their functional and evolutionary context. The first, the Fluc family of fluoride channels, is an exquisitely selective ion channel which confers resistance to environmental Fluoride via electrodiffusive transport. The structure of this family has been shaped by gene duplication, as phylogenetic analysis reveals. This reveals that duplication occurs often but that, once duplicated, reversion is rare. Additionally, functional redundancy leads to one (of two) pores becoming non-functional. Gene duplication has a key role in protein evolution, and in membrane transporters in particular has been involved in most transporters that are known today. The Flucs are an appealing model system for studying this, as they exist in several evolutionary states, and their function depends on asymmetric interactions between dimers which can be used to probe drift after duplication.

This interaction can be harnessed to develop an experimental system for studying the fitness effects of gene duplication. Several models exist which propose mechanisms for the retention of gene duplicates before adaptive or non-adaptive forces fix them, but any experimental test has been difficult to obtain. The asymmetric interactions of a key motif in the Flucs allows constructs which differentiate between unduplicated, duplicated, and heterodimeric (fixed duplicated) evolutionary states of the same protein to be created. Deep mutational scanning of these constructs and comparison between the contexts then allows insight into the natural trajectories duplicates took in the Flucs.

Small Multidrug Resistance family of prokaryotic proton-coupled transporters has long been thought to be mainly involved in antibiotic resistance. A more careful analysis reveals that the major role of these is in fact guanidinium efflux. Using a combination of

phylogenetics and in vitro assays, a variety of homologs are shown to be specific and well-coupled electrogenic guanidinium/proton antiporters. To understand the relationship between subtypes, a novel electrophysiological technique, solid-supported membrane electrophysiology, was used to study the comparative substrate capacities of the guanidinium-exporting subtype (called Gdxs) and the multispecific drug-exporting subtype (called Qacs). Surprisingly, a chemical region of shared recognition between the subtypes was discovered: aromatic or hydrophobic singly-substituted guanidiniums are transported by both subtypes. To understand the structural basis for this, the first high-resolution crystal structure of an SMR was solved. Both subtypes share aromatic binding pockets, but loss of important H-bonds in the Qac subtype likely introduces polyspecificity.

These combined results reveal how structure, function, and topology interact to shape the evolution of membrane proteins, and provide important insights and tools for understanding how extant membrane machinery arose.

CHAPTER 1

Introduction

All forms of life, from the tiniest microbe to the tallest tree, rely on protein machinery. Most cellular active processes are largely conducted by a series of complex protein assemblies: metabolism, reproduction, communication and self-recognition are all examples. As products of evolution, their current variety must be understandable as springing from a long series of prior forms. The nature of this history shapes the nature of its products, though, and we can learn a great deal about present-day protein biology and biochemistry by examining evolutionary histories. Membrane proteins are products of the same processes, but the addition of a specific membrane-relative orientation provides additional constraints and opportunities for this process. In this thesis, two model families of bacterial membrane proteins are used to understand how structure and function are linked via evolution in the membrane, and mechanisms for how complexity arises.

1.1 Membrane proteins

The membrane is a privileged environment in microbes: the thin lipid line separating the homeostatic interior of the cell from the chaos of the environment beyond. It serves variously to exclude noxious materials from entry, to contain nutrients and machinery, and to preserve the identity of the organism. Controlled movement across its boundaries allows communication, sensing, energy generation and more, and is accomplished through a dizzying variety of complex protein machines.

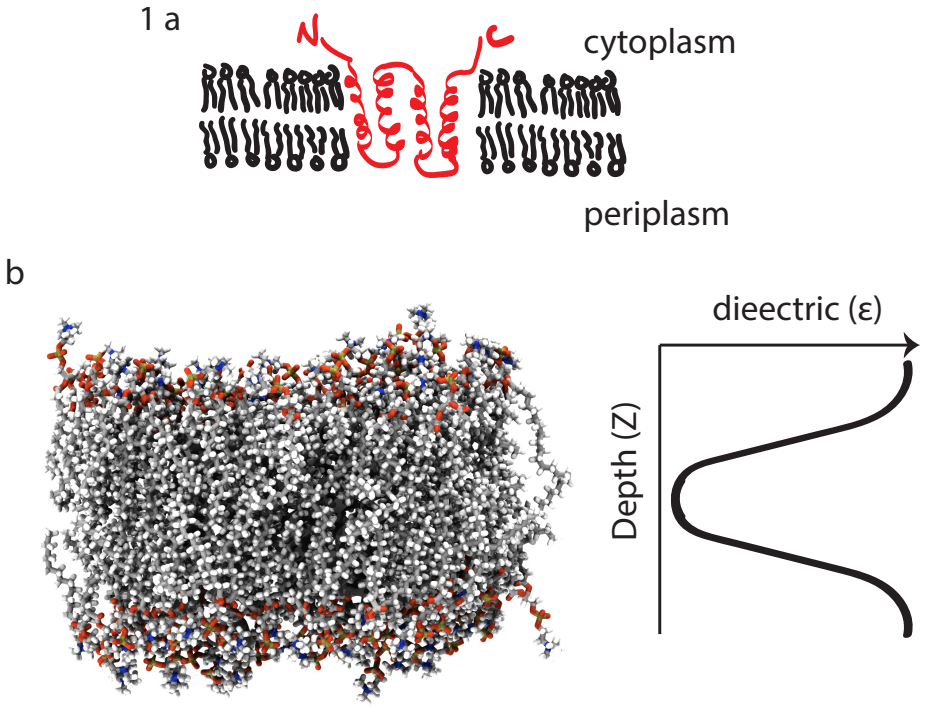

Figure 1.1: Inner membrane protein features. A) Cartoon depiction of an inner membrane protein. Specific features including alpha-helical nature, trans-bilayer embedding, and non-symmetric topology can be observed. B) Left, atomistic depiction of a lipid bilayer[1]. Right, dielectric constant through the plane of the membrane.

Inner membrane proteins are deceptively simple: they consist essentially of alpha-helices strung back and forth across the membrane and packed against each other in bundles (figure 1.1 a). This simplicity is deceiving, as the nature of the membrane-embedding process introduces several important new complexities. Membrane proteins are the products of a highly-regulated expression and insertion process, and are located in an anisotropic heterogeneous environment[2–5]. The solvent for membrane proteins, the lipids themselves, comes in a variety of lengths, shapes, chemistries, each of which will vary the local environment of the membrane and potentially larger-scale membrane features, such as rafts and domains[6]. Interaction with lipids both specifically or transiently can change enzymatic, binding, and oligomeric properties, and is an important point of regulation.

In addition, being embedded in the membrane introduces an essential structural difference to soluble proteins: there is an asymmetry with respect to orientation across the membrane. To some extent, a protein embedded in one direction is not the same as one embedded in the other direction: functionally, they may have very different effects. A receptor must be oriented properly to recognize its signal, and a pump should be oriented so as to export toxins or import nutrients. The specific directions the protein chain adopts with respect to the membrane is called its topology, and the process whereby it is adopted during expression is called topogenesis.

Beyond this fundamental broken symmetry, the bilayer itself is heterogenous as one moves across it. An atomistic view shows the shifting nature of the chemical environment (figure 1.1 b): at the ends, the charged and polar portions of the lipids and the surrounding counterions maintain a bulk-like hydrophilicity and dielectric. As one moves further into the bilayer, the aliphatic tails of the lipids produce an increasingly hydrophobic environment, culminating in the very low dielectric environment at the middle of the bilayer. This gives way in reverse to the aqueous environment again as passage is complete. This heterogeneity introduces a new dimension to protein function, as they must adapt to the multifarious environments they find themselves in to function: the stability of a mem-

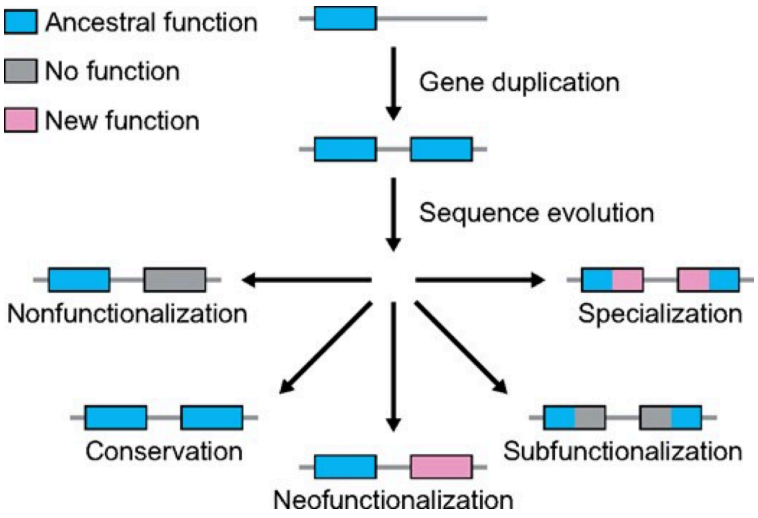

Figure 1.2: Gene duplication for generating protein novelty. After gene duplication, the ancestral functions of a gene can be distributed in a number of ways. The specific distributions result in subfunctionalization, neofunctionalization, or other situations (taken from[9]).

brane protein is differentially affected by different substitutions across the membrane, as experiment has revealed[7,8].

1.2 Biological complexity

Complexity is a slippery concept, covering both the simple idea of a biological complex as well as broader senses of heterogeneity and even the (questionable) idea of a hierarchy of development. Focusing on the molecular level thankfully avoids many of these difficulties, as complexity here is simpler to define as the structural, functional, and regulatory variety adopted by single molecular machine in a cell. For example, hemoglobin is more complex than myoglobin as a multimer rather than a monomer; as a heteromer, it is more complex than its homomeric predecessors. An enzyme which acts on a number of substrates is more complex than one with only a single target, as well. Importantly, complexity here is not teleological: it may be lost or gained along a lineage.

1.3 Gene duplication

Gene duplication is sometimes overlooked as a key element in protein evolution, but it is a central one[10–13]. Both entire genomes and small regions may be duplicated during replication through a variety of mechanisms. These duplicates have a number of fates. They may simply be retained as a pair of duplicates, potentially increasing the dosage only. One or the other may be lost through pseudogenization and deletion. More interestingly, they may be retained with a new distribution of functions (figure 1.2)[9]. In subfunctionalization, the ancestral function is partitioned between the two duplicates. Thus both are required to fulfill the role of the ancestor. In neofunctionalization, one of the pair adopts a useful new function that was not present in the ancestor, potentially expanding the repertoire of the organism. One may imagine that the duplication relaxed purifying selection on the pair through redundancy, allowing this novel function to be reached. Gene duplication thus likely provides a key route to functional complexity in living systems.

In the case of an ancestral oligomer, duplication introduces an additional wrinkle. Immediately after duplication, the paralogs are able to interact with themselves but also each other[14]. This introduces a functional dependency between the two, as the heterodimeric state links the two genes. Any deleterious mutations in one copy then may reduce the overall fitness via a dominant-negative phenotype. This paralog interference restricts the possibilities for developing new functions and will control the dynamics after duplication. If heterodimerization is retained, the fates of the pair are strictly retaining the ancestral function, coevolving improved heterodimeric fitness, or potentially negative dominance becoming a competitive inhibitor. Paralog interference can be avoided by preventing oligomerization. Regulatory evolution that separates their expression in time or space can serve this purpose, or mutations in the interface that disrupt the heterodimeric state but preserve homodimers. One paralog may also accumulate mutations that make it unable to oligomerize at all, thus serving a similar purpose.

All of these possibilities require that gene duplicates stay long enough for selection or drift to produce novelty. Although several models exist for how duplicates are retained, this question is difficult to study experimentally[9,15]. A model system to investigate the fitness landscapes in the immediately duplicated state would thus be useful to understand this key period.

1.4 Complexity in the membrane: oligomerization and function

Membrane proteins are prone to forming complexes. The fundamental reasons for this remain under investigation, but the ease of encounter formation due to dimensional reduction, as well as the orientational constraints pre-arranging dimerization conformations, may play a role. Variations in complex stoichiometry and constituents can mediate signalling, transport, and motility[16–18]. Oligomerization introduces opportunities for new ways of regulation, including cooperativity and allostery, and is one of the central means for the creation of complexity in biochemistry. A mechanistic understanding of how oligomerization is generated and preserved by evolution is key to understand the expansion of life itself.

Many complexes appear to be formed from homologous partners. In membrane proteins, this takes a particular form, the "inverted repeat"[19–21]. This is best understand as the observation that many transporter structures appeared to be formed from two domains which shared a similar fold but were oriented oppositely to each other across the membrane. These domains often share little sequence similarity, with the relationship only detectable via structure.

This suggests an intriguing evolutionary trajectory (figure 1.3 a): the formation of multi-domain inverted repeats likely arose via the gene fusion of two independent proteins, with a dimeric complex becoming a monomer in the process. The orientation of the subunits are opposite, therefore. But where did these paired genes come from, then? Compared to cytosolic proteins, a protein with a fixed topology in a membrane is not the

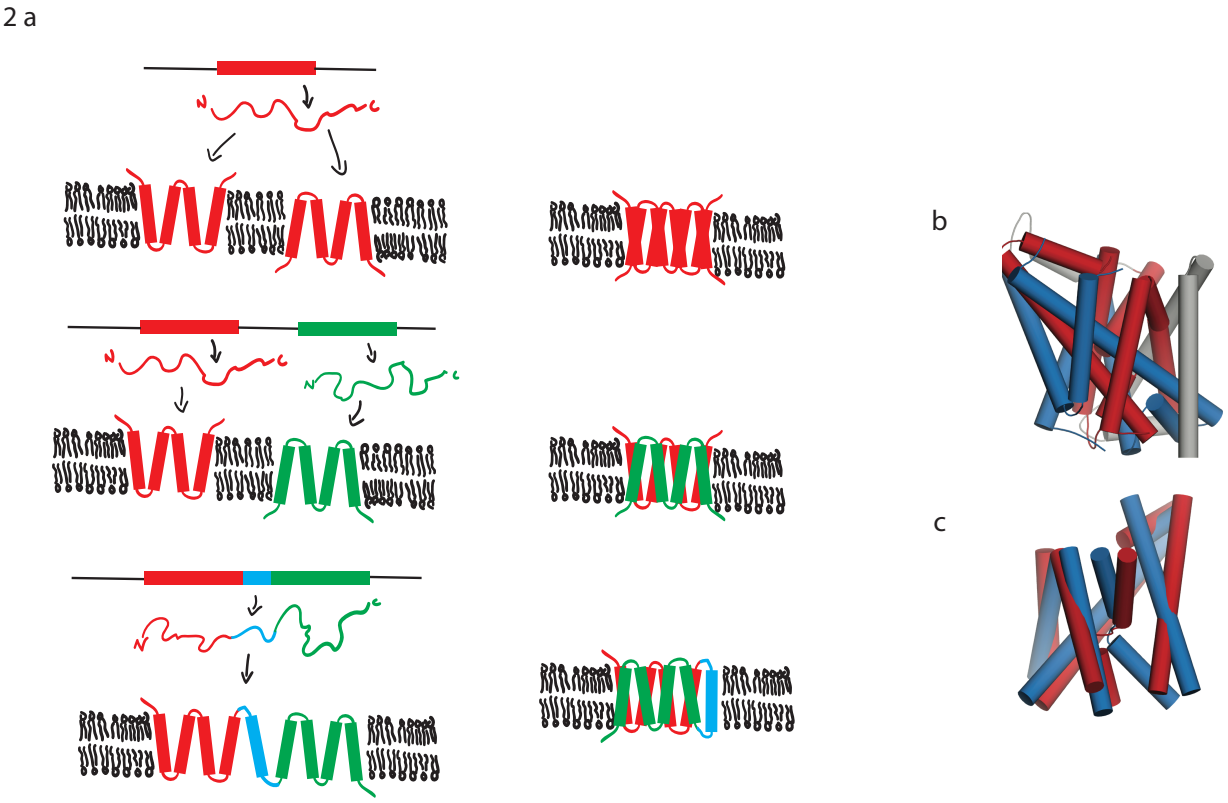

Figure 1.3: Membrane protein evolution through gene duplication and fusion. A) Top: a dual-topology protein, coded by a single gene, produces proteins that insert in both directions in the membrane and assemble into an antiparallel dimer. Middle: A duplicated gene where each copy produces a single protein that inserts with a fixed topology in the membrane, which assemble together into a heterodimer. Bottom: A fused gene pair, with the addition of a linker helix, produces a single protein with two homologous domains that interact with each other while folding. B) Crystal structure of LeuT, a bacterial leucine transporter, with a prototypical inverted repeat: symmetric domains colored in red and blue (from[20]). C) Structural alignment of homologous domains in the LeuT fold.

same as one with another topology. This implies that any unduplicated precursor to such a heterodimeric pair must itself have been able to adopt *both* orientations in the membrane: a so-called "dual-topology" protein.

This was an understandably controversial proposition, as it could be understood to violate the sense that a proteins fold (extended here to include its topology in the membrane) should be determined by its sequence. A wealth of evidence now has shown that such proteins exists, but are exceedingly rare[22–24]. Only a handful of families are suspected to exist where this dual-topology state is extant. Of these, fewer have any structural evidence or biochemical characterization. The work in this thesis examines the sole two families where this is the case: the Fluc family of fluoride channels and the Small Multidrug Resistance (SMR) family of transporters.

1.5 Outline of thesis

This thesis covers a number of ways I have attempted to ask how protein complexity is created in the membrane, and how the specific environment of the membrane constrains those trajectories.

In chapter 2, we examine the first model family I used, the Flucs. Here we study their evolution, with an eye to their oligomeric order and topology within the membrane, and how these influence conservation of specific functional motifs.

In chapter 3, we shift to the Small Multidrug Resistance (SMR) family, the second model family I use. Here the question becomes one of functional novelty and diversification rather than structural complexity. In this chapter, we present the first biochemical characterization of a previously unexplored major subfamily of the SMRs, and we show they are specific guanidinium/proton antiporters which we call the Gdxs.

In chapter 4, we expand upon our previous characterization by the addition of an electrophysiological technique, solid-supported membrane electrophysiology, to better understand the limits of substrate recognition in the Gdxs. In addition, we solve the first

high-resolution structure of an SMR, in this case a Gdx homolog.

In chapter 5, we shift to the *E. coli* Gdx homolog, and conduct a further biophysical characterization to support work with this important model organism.

In chapter 6, we return to the Flucs. To understand how gene duplicates are retained shortly after duplication, we use the specific architecture of the Flucs to develop an assay that distinguishes between different genetic architectures, and combine this with a deep mutational scanning methodology to work towards a high-resolution fitness landscape of duplicate retention in the Flucs.

In chapter 7, we conclude with a summary of the findings and potential future directions.

References

(1) Newport, T. D.; Sansom, M. S. P.; Stansfeld, P. J. The MemProtMD Database: A Resource for Membrane-Embedded Protein Structures and Their Lipid Interactions. *Nucleic Acids Research* **2019**, *47* (D1), D390–D397. https://doi.org/10.1093/nar/gky1047.

(2) Robertson, J. L. The Lipid Bilayer Membrane and Its Protein Constituents. *Journal of General Physiology* **2018**, *150* (11), 1472–1483. https://doi.org/10.1085/jgp.201812153.

(3) von Heijne, G. Formation of Transmembrane Helices In Vivo—Is Hydrophobicity All That Matters? *Journal of General Physiology* **2007**, *129* (5), 353–356. https://doi.org/10.1085/jgp.200709740.

(4) White, S.; Vonheijne, G. Transmembrane Helices Before, During, and After Insertion. *Current Opinion in Structural Biology* **2005**, *15* (4), 378–386. https://doi.org/10.1016/j.sbi.2005.07.004.

(5) Marx, D. C.; Fleming, K. G. Membrane Proteins Enter the Fold. *Current Opinion in Structural Biology* **2021**, *69*, 124–130. https://doi.org/10.1016/j.sbi.2021.03.006.

(6) Bloom, M.; Evans, E.; Mouritsen, O. G. Physical Properties of the Fluid Lipid-Bilayer Component of Cell Membranes: A Perspective. *Quart. Rev. Biophys.* **1991**, *24* (3), 293–397. https://doi.org/10.1017/S0033583500003735.

(7) Marx, D. C.; Fleming, K. G. Local Bilayer Hydrophobicity Modulates Membrane Protein Stability. *J. Am. Chem. Soc.* **2021**, *143* (2), 764–772. https://doi.org/10.1021/jacs.0c09412.

(8) Hessa, T.; Meindl-Beinker, N. M.; Bernsel, A.; Kim, H.; Sato, Y.; Lerch-Bader, M.; Nilsson, I.; White, S. H.; von Heijne, G. Molecular Code for Transmembrane-Helix Recognition by the Sec61 Translocon. *Nature* **2007**, *450* (7172), 1026–1030. https://doi.org/10.1038/nature06387.

(9) DeGiorgio, M.; Assis, R. Learning Retention Mechanisms and Evolutionary Parameters of Duplicate Genes from Their Expression Data. *Molecular Biology and Evolution* **2021**, *38* (3), 1209–1224. https://doi.org/10.1093/molbev/msaa267.

(10) Ohno, S. *Evolution by Gene Duplication*; Springer Berlin Heidelberg: Berlin, Heidelberg, 1970. https://doi.org/10.1007/978-3-642-86659-3.

(11) Lynch, M. Gene Duplication and Evolution. *Science* **2002**, *297* (5583), 945–947. https://doi.org/10.1126/science.1075472.

(12) Zhang, J. Evolution by Gene Duplication: An Update. *Trends in Ecology & Evolution* **2003**, *18* (6), 292–298. https://doi.org/10.1016/S0169-5347(03)00033-8.

(13) Copley, S. D. Evolution of New Enzymes by Gene Duplication and Divergence. *The FEBS Journal* **2020**, 22.

(14) Kaltenegger, E.; Ober, D. Paralogue Interference Affects the Dynamics After Gene Duplication. *Trends in Plant Science* **2015**, *20* (12), 814–821. https://doi.org/10.1016/j.tplants.2015.10.003.

(15) Force, A.; Lynch, M.; Pickett, F. B.; Amores, A.; Yan, Y. L.; Postlethwait, J. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **1999**, *151* (4), 1531–1545.

(16) Cecchetti, C.; Pyle, E.; Byrne, B. Transporter Oligomerisation: Roles in Structure and Function. *Biochemical Society Transactions* **2019**, *47* (1), 433–440. https://doi.org/10.1042/BST20180316.

(17) Zhang, H.; Pan, Y.; Hu, L.; Hudson, M. A.; Hofstetter, K. S.; Xu, Z.; Rong, M.; Wang, Z.; Prasad, B. V. V.; Lockless, S. W.; Chiu, W.; Zhou, M. TrkA Undergoes a Tetramer-to-Dimer Conversion to Open TrkH Which Enables Changes in Membrane Potential. *Nat Commun* **2020**, *11* (1), 547. https://doi.org/10.1038/s41467-019-14240-9.

(18) Alvarez-Sieiro, P.; Sikkema, H. R.; Poolman, B. Heterodimer Formation of the Homodimeric ABC Transporter OpuA. *IJMS* **2021**, *22* (11), 5912. https://doi.org/10.3390/ijms22115912.

(19) Forrest, L. R.; Rudnick, G. The Rocking Bundle: A Mechanism for Ion-Coupled Solute Flux by Symmetrical Transporters. *Physiology* **2009**, *24* (6), 377–386. https://doi.org/10.1152/physiol.00030.2009.

(20) Khafizov, K.; Staritzbichler, R.; Stamm, M.; Forrest, L. R. A Study of the Evolution of Inverted-Topology Repeats from LeuT-Fold Transporters Using AlignMe. *Biochemistry* **2010**, *49* (50), 10702–10713. https://doi.org/10.1021/bi101256x.

(21) Keller, R.; Ziegler, C.; Schneider, D. When Two Turn into One: Evolution of Membrane Transporters from Half Modules. *Biological Chemistry* **2014**, *395* (12), 1379–1388. https://doi.org/10.1515/hsz-2014-0224.

(22) Rapp, M.; Granseth, E.; Seppälä, S.; von Heijne, G. Identification and Evolution of Dual-Topology Membrane Proteins. *Nat Struct Mol Biol* **2006**, *13* (2), 112–116. https://doi.org/10.1038/nsmb1057.

(23) Seppala, S.; Slusky, J. S.; Lloris-Garcera, P.; Rapp, M.; von Heijne, G. Control of Membrane Protein Topology by a Single C-Terminal Residue. *Science* **2010**, *328* (5986), 1698–1700. https://doi.org/10.1126/science.1188950.

(24) Stockbridge, R. B.; Robertson, J. L.; Kolmakova-Partensky, L.; Miller, C. A Family of Fluoride-Specific Ion Channels with Dual-Topology Architecture. *Elife* **2013**, *2*, e01084. https://doi.org/10.7554/eLife.01084.

CHAPTER 2

A Topologically Diverse Family of Fluoride Channels

This chapter is adapted from the following published article:
Macdonald, C. B. & Stockbridge, R. B. A topologically diverse family of fluoride channels. Current Opinion in Structural Biology 45, 142–149 (2017).
C.B.M. and R.B.S wrote the text and performed the analysis.

2.1 Introduction

The advent of the structural era in membrane protein biology revealed that many membrane transport proteins, built on common folds in classes as diverse as LeuT, CLC, MFS, aquaporin, NhaA, and GltPh, are comprised of two homologous domains, arranged antiparallel with respect to each other, in an arrangement known as an inverted repeat (discussed extensively elsewhere, for example, see:[1–3] .) This architecture is the vestige of a gene duplication and fusion so long ago that the structurally homologous domains no longer bear any discernable sequence identity. In almost all cases, that ancestral single domain gene has been lost to time. But dual-topology membrane proteins are rare examples in which that primitive architecture persists; they represent modern day, single-domain building blocks that assemble in the membrane with inverted architecture (Figure 2.1, A).

The orientation of integral membrane proteins in bacteria can be predicted by the "positive inside rule," originally articulated by von Heijne[4], which describes a nearly universal feature of membrane proteins – a large excess of arginines and lysines on the cytoplasmic loops. In 2006, von Heijne described a rare exception to the rule: a class of small, unusually

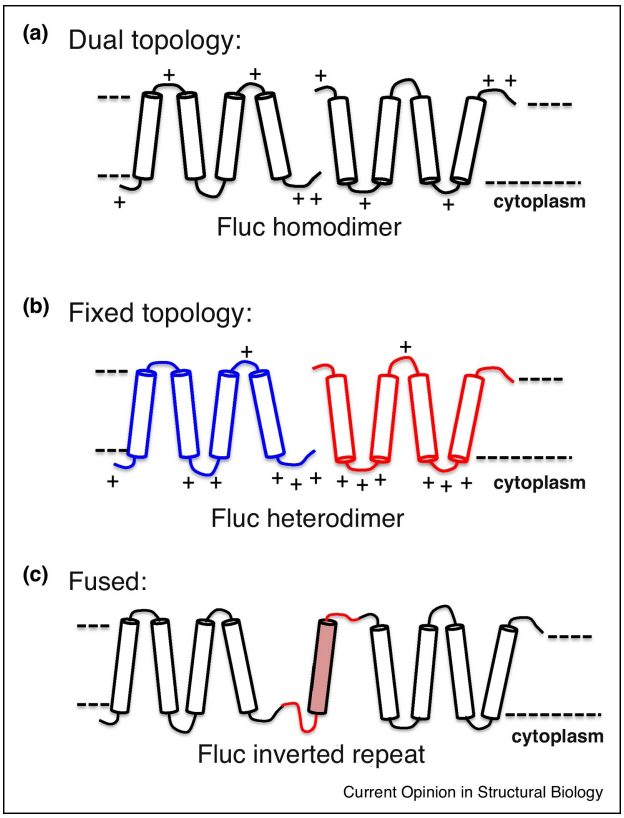

Figure 2.1: Schematic illustrating subunit topologies found in the Fluc family of fluoride channels. (+) symbols indicate arginine and lysine residues of representative bacterial Fluc homologues. A. Subunits of dual topology homodimers are inserted into the membrane in both inward- and outward-facing orientations, and can be identified by balanced (+) charge bias, as for this representative example from Bordetella pertussis. B. For Fluc heterodimers, subunit pairs are inserted into the membrane with preferred, fixed orientations that can be predicted by the "positive inside" rule[4], as in this representative heterodimer from Lactobacillus acidophilus. The termini of one subunit are always inward-facing (blue, positive charge bias on the termini and loop 2) and the termini of the other are always outward-facing (red, with positive charge bias on loops 1 and 3). C. Cartoon of a fused Fluc, with a transmembrane "inversion helix" (pink) enforcing the antiparallel domain architecture. Inverted repeat Flucs are found exclusively in eukaryotes, and have been experimentally determined to insert into the plasma membrane with the N-terminus facing the cytoplasm in Saccharomyces cerevisiae[5]. The factors that determine topology in eukaryotic membrane proteins are more complex than in bacteria[6]; thus, the charge distribution is not shown for this homologue.

hydrophobic proteins that he called dual-topology proteins[7]. Monomers of dual-topology proteins are inserted into cell membranes with no inward/outward-facing bias and can dimerize in an antiparallel orientation in vivo[8]. As the original controversy regarding this unusual pattern of membrane insertion has receded, dual-topology proteins have been recognized as likely evolutionary antecedents to inverted repeat[2,9,10].

2.2 Topological diversity in Fluc family fluoride channels

Using sequence analysis, a number of putative dual-topology protein families have been proposed[7,11,12], but only two have been characterized in great biochemical depth: the multidrug efflux pump EmrE[13,14], and a class of fluoride channels called Flucs (also known as crcB or FEX), which protect against toxic environmental fluoride ion (F−) in weakly acidic conditions[15–17]. Flucs are ubiquitous in bacteria, and also found in archaea and free-living eukaryotes (where they are known as FEX proteins), including yeasts and protozoa, plants, and even simple marine animals such as sponges, sea anemones, and tunicates[18]. Fluoride export function has been established by genetic knockout strains of *Escherichia coli*, *Candida albicans*, *Saccharomyces cerevisiae*, and *Neurospora crassa*, which are all rendered hypersensitive to fluoride at environmental concentrations[16,18]. With this many biological test kitchens, the Flucs have evolved into a topologically diverse family that it is not constrained only to dual-topology homodimers. Modern Flucs sample each evolutionary way station: they are present in genomes as single genes that code for antiparallel homodimers[19] (Figure 2.1, A), oppositely inserted pairs (~30% sequence identity) that transport fluoride as obligate heterodimers[17] (Figure 2.1, B), and in eukaryotes, fused two-Fluc proteins linked by a TM helix that forces the conjoined subunits into an antiparallel orientation: an inverted repeat[18] (Figure 2.1, C). This full assortment of topological states is rare, and unique among proteins with known function.

A phylogenetic tree constructed with ~500 representative bacterial and archaeal Fluc sequences shows that gene duplications leading to probable heterodimers have occurred

at least four times in bacterial lineages: in Firmicutes, in Actinomycetes, in Cyanobacteria, and in a handful of Alphaproteobacteria (Figure 2.2). A fifth possible duplication occurred among a clade of archaea nestled among the Actinomycetes. Archaeal lineages spring from bacterial clades here and elsewhere in the Fluc phylogeny, perhaps indicating that these proteins were transmitted via lateral gene transfer.

All eukaryotic Flucs are two-domain inverted repeats, and the eukaryotic clade is the sole example of a fusion event. Topological inspection reveals a simple explanation for the paucity of fused two-domain Flucs: to retain antiparallel topology for a protein with an even number of helices, the fusion event must also introduce a membrane spanning helix. Fusion events are more common for antiparallel proteins with an odd number of helices, as evidenced by the number of independent duplication/fusion events in the DUF606 (domain of unknown function) family[24]. Sequence analysis suggests that the DUF606 family is topologically diverse, with dual topology, duplicated heterodimers, and fused representatives. The monomer possesses five transmembrane helices, and of the nine independent duplication events observed in bacteria, four have yielded fused two-domain proteins[24].

2.3 Conservation of biological role and function

In the Fluc family, these diverse topologies all serve a common biological purpose. Fluoride ion interferes with some of the most fundamental metabolic activities, and its export is therefore necessary for growth[25]. Isosteric with hydroxide, fluoride ion closely resembles the activated water molecule transition state that occurs during reactions catalyzed by essential metabolic enzymes, including enolase and pyrophosphatase[26,27]. Present at basal levels of 20–100 μM in typical marine and terrestrial environments[28], fluoride can enter and become trapped within cells via a weak acid accumulation pathway[15]. In the absence of a fluoride export system, cytoplasmic fluoride increases to levels as high as mM at a modestly acidic pH value of 5.5, a level well above the ~100 μM inhibition constants for metabolic enzymes[15,18]. The weak acid accumulation phenomenon accounts for

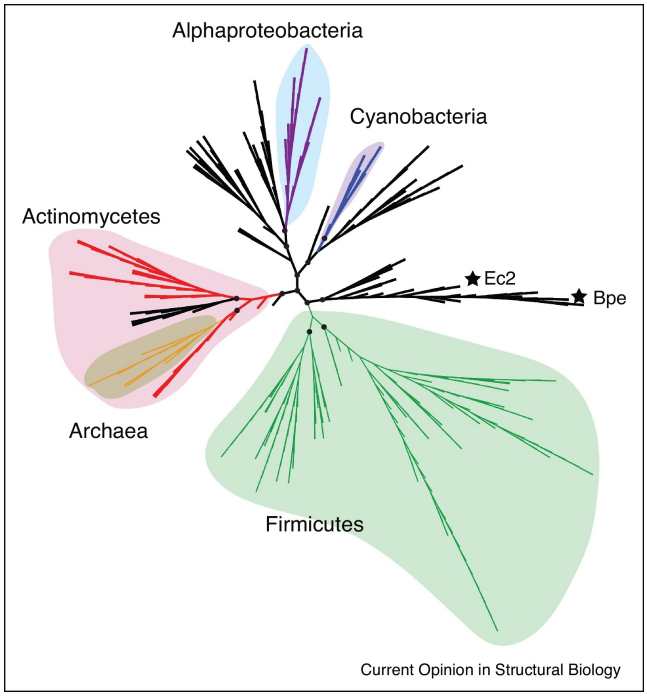

Figure 2.2: Unrooted phylogenetic tree constructed from ~500 representative Fluc proteins using maximum likelihood methods[20–22], with proposed duplication events indicated in color: Green: Firmicutes duplication; red: Actinomycetes duplication; blue: alphaproteobacteria duplication; violet: cyanobacteria duplication; yellow: archaeal duplication. Gene pairs were identified as likely heterodimers based on two criteria: the start of the second coding sequence falls within 100 nucleotides of the end of the first, and the two proteins have opposite loop charge biases, in accordance with the positive inside rule for membrane protein orientation, assessed by consensus prediction of membrane protein topology[23]. 92% of the pairs that met the first criterion also met the second. The remainder of pairs showed little positive charge bias, and thus prediction of the direction of insertion was uncertain. Amino acids likely to encode aqueous-exposed regions, inferred by structural alignment to the homodimeric Fluc structure, were trimmed from the alignment. This maneuver was essential to determine a phylogeny unbiased by the convergent evolution of positive charges on the inward-facing loops of heterodimers from divergent lineages. Nodes with >90% clade confidence according to approximate likelihood ratio test (aLRT[20]) are indicated by circles. Two structurally characterized homodimers, from Bordetella pertussis and a virulence plasmid isolated from Escherichia coli, are indicated by stars.

the counterintuitive notion of protection from an environmental toxin by an energetically passive channel. In addition, metabolizing cells tend to maintain a negative electric potential across the membrane, which provides an additional driving force for anion expulsion. (Though the most widespread, Fluc family channels are not the only fluoride export strategy that has evolved; some bacteria instead use fluoride selective proteins from the CLC family of anion channels and transporters, coupling fluoride export to proton import[29]).

Available evidence implies that the electrodiffusive (i.e. passive flow of fluoride down its electrochemical gradient) mechanism of Fluc channels is conserved across topologies. Single channel recordings of purified bacterial homodimers in planar lipid bilayers show 5– 10 pS currents for the constitutively open single channels, equivalent to a throughput of ~106 ions/s[17]. Estimates of single molecule turnovers for fused *S. cerevisiae* proteins, inferred from patch clamp measurements of many-protein currents, yield a similar value of ~106 ions/s. This high turnover is consistent with a channel rather than a transporter mechanism[5]. Heterodimeric representatives have not been as well characterized, but bulk liposome flux assays imply a high rate of transport of >105 ions/sec for the *Lactobacillus acidophilus* heterodimer[17].

The bacterial homodimers are ideally selective for fluoride over cations, and enormously (>104-fold) selective for fluoride over other abundant cytoplasmic ions like chloride, a closely related halide that only differs in dehydrated radius by about half an Ångström (1.33 Å for fluoride versus 1.81 Å for chloride)[17]. This selectivity is both biologically necessary to avoid catastrophic membrane depolarization and unparalleled among ion channels for close ionic analogs. Such discrimination could probably not be achieved without dehydration of the fluoride ions in the pore[30]. Given the unusually high enthalpy of desolvation for fluoride ion (ΔH ~ 125 kcal/mol, versus ~90 kcal/mol for chloride), selective recognition of fluoride ions is a chemically interesting puzzle in its own right, and the selectivity achieved by these proteins surpasses human attempts to develop host-guest compounds that selectively recognize fluoride over chloride[31]. The two-domain *S. cere-*

visiae proteins also demonstrate fluoride/chloride selectivity at the limit of resolution of the patch-clamp assay (greater than 10-fold), although further measurements will be required to determine whether their selectivity matches their bacterial counterparts[5].

2.4 Symmetrical, antiparallel homodimer structure

Recent 2.1 Å and 2.6 Å crystal structures of homodimers from *Bordetella pertussis* (PDB:5FXB) and an *E. coli* virulence plasmid (PDB:5A43) revealed an elegant and unlikely architecture[32] (Figure 2.3). The previously unseen fold is striking for its perfect (non-crystallographic) two-fold symmetry about an axis parallel to the plane of the membrane, so that the channel presents identical faces to both the cytoplasmic and periplasmic sides of the membrane. The hourglass shaped dimer possesses two deep aqueous vestibules lined with conserved polar side chains (Figure 2.3, A). Within each vestibule, a conserved arginine (Arg23) from the first transmembrane helix contributes to the overall electropositive environment. TM3 has a mid-membrane break of 5 residues, and the crossover of these breaks contributed by the two subunits defines the deepest reaches of the vestibules. The vestibules do not form a continuous aqueous pathway through the protein; instead a plug of protein ~10 Å thick separates the aqueous chambers (Figure 2.3, B). At the center of this plug, at the heart of the protein, resides a deeply buried sodium ion (Na+) located precisely on the two-fold symmetry axis (Figure 2.3, A, 2.3, C). Fulfilling the symmetry of the overall structure, the central sodium ion is present in an unusual tetrahedrally-coordinated conformation, where it is liganded by two symmetry-related backbone carbonyl oxygens from each subunit (Gly77 and Thr80), contributed by the mid-membrane breaks of TM3 (Figure 2.3, D). Although sodium ions are typically coordinated by 5–6 ligands, coordination by four is not unprecedented[33,34]. The role of this central sodium ion has not been elucidated, but it certainly contributes to the general electropositivity of the protein core, and its location at the dimer interface suggests a possible role in homodimer affinity or specificity.

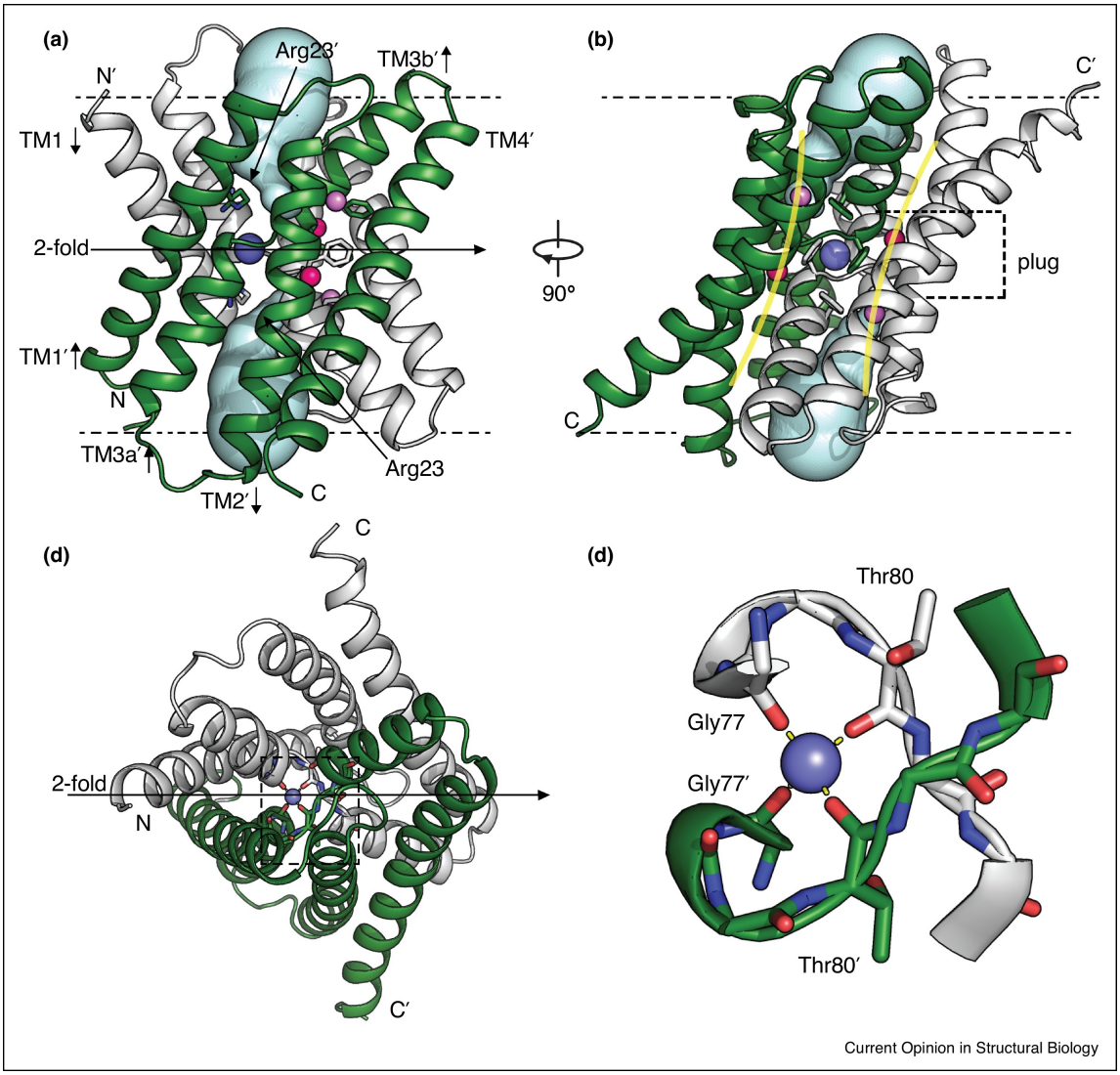

Figure 2.3: Structure of dual-topology Fluc homodimer (PDB: 5FXB). A. Through membrane view of the channel. Volume of the aqueous vestibules are colored in cyan, fluoride ions in pink and violet, central Na+ in blue, and conserved vestibule Arg23 and Phe box residues shown as sticks. Helices are labeled with arrows indicating polarity. The membrane boundaries are indicated by a dashed line. The non-crystallographic 2-fold symmetry axis is indicated by an arrow. Note that because of the channel's perfect 2-fold symmetry, the cytoplasmic and periplasmic sides of the channel are identical. B. View of channel in A rotated 90°. This view emphasizes the double-barrelled architecture; putative pores are highlighted by yellow lines. This view also shows that the aqueous vestibules do not connect; the span of protein separating the vestibules is referred to as the plug, and is indicated by dashed lines. C. Channel viewed perpendicular to the plane of the membrane. The two-fold symmetry axis is indicated by an arrow. Residues from the transmembrane helix 3 (TM3) break are shown as sticks. D. Detail from panel C showing coordination of the central Na+ at the dimer interface.

2.5 Two pore construction of homodimers

The two-fold structural symmetry of the channel begets functional redundancy: the cytoplasmic and periplasmic sides of the channel are identical, and it harbors two pores[32] extending antiparallel with respect to each other across the membrane (Figure 2.3, B). The double-barreled structure is a surprise given the protein's unusually small size, just 128 residues (15 kDa) per subunit in the dimer. (Note that although the pores are related by 2-fold symmetry, and thus have opposite polarity, fluoride uses both conduits to diffuse passively down its gradient.) The pores are defined by four bound fluoride ions, arranged two-by-two, parallel to each other and perpendicular to the membrane plane (Figure 2.3, B). The side chains that coordinate these ions include an absolutely conserved asparagine in TM2 (Asn 43) and a sequence of polar residues lining one face of TM4, spaced every fourth residue, called the TM4 polar track (Figure 2.4, A). Although the absolute identity of the residues lining the TM4 polar track is not conserved among even close Fluc family members, the polar character is. Perhaps the most prominent feature of the fluoride ion coordination sphere is the "phenylalanine box," a symmetrical motif whose corners define the two pores of this double-barreled channel, with the electropositive edges of each phenylalanine coordinating a fluoride ion (Figure 2.4, B). Such "anion-quadrupole" interactions have been observed in proteins before[35,36], notably predicted in the bestrophin family of anion channels[37]. Each pore is not confined entirely to an individual monomer, but is lined by side chains contributed from both. Mutation of these pore-lining residues decreases fluoride throughput, although a mutation has not yet been identified that relaxes discrimination against chloride[32,38].

Outside of these deeply buried crystallographic fluoride sites, the details of the permeation pathway are unclear. Although the observed fluoride binding sites are occluded from the vestibules (Figure 2.3, A), it is tempting to imagine the vestibules' electropositive character drawing in and preliminarily desolvating fluoride ions. Mutation of the

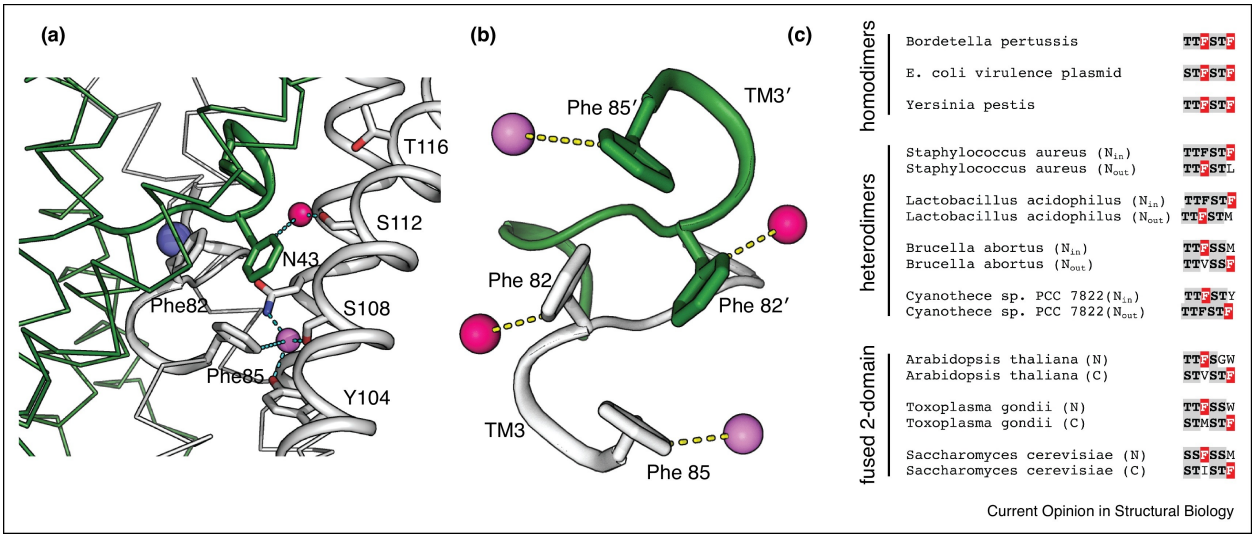

Figure 2.4: Pore construction in Fluc proteins. A. Fluoride coordination sphere in Bordetella pertussis Fluc homodimer (PDB: 5FXB). Two subunits are shown in green and gray, with the "central Na+" shown in blue. Fluoride ions are shown as pink and violet spheres, and coordinating residues Asn43, Phe box residues Phe82 and Phe85, and TM4 polar track residues Tyr104, Ser108, Ser112, and Thr116 shown as sticks. For clarity, only one of the two pores rendered as a cartoon with fluoride ions shown. B. Fluoride coordination by the phenylalanine box demarcates the two pores. Note that each subunit contributes a phenylalanine to the pore. Symmetry-related fluoride ions are colored identically. C. Alignments of Phe box region of Fluc homodimers, heterodimers, and fused 2-domain constructs. Conserved residues are bolded, and Phe residues contributing to the expected functional pore are highlighted red. For heterodimers, the inward-and outward facing subunits are indicated, predicted based on positive charge bias. For fused two-domain proteins, the N-terminal (inward-facing) and C- terminal (outward-facing) domains are indicated.

arginines and a conserved vestibule serine have deleterious effects on function in yeast proteins, supporting the notion that the vestibules are involved in ion transport[5]. Additional evidence that the vestibules provide part of the ion pathway comes from recent electrophysiological assays exploiting small, engineered proteins called "monobodies." These experiments showed that the soluble, toxin-like monobodies inhibit Fluc channels by physically blocking the pores, and that a negatively charged glutamate that acts as an electrostatic surrogate for fluoride is responsible for the pore occlusion[39,40]. Structures of Fluc monobody-complexes show that the loop on which this glutamate resides extends ~9 Å into the aqueous vestibules[32], providing compelling evidence that the vestibules contribute part of the permeation pathway.

2.6 Drift of two-pore construction after duplication

The double-barreled architecture raises the question: does Fluc need both pores? Fluoride flows through both pores in the homodimer, and mutagenic ablation of one cuts the channel's conductance neatly in half[38]. Analysis of sequences of fused eukaryotic FEX proteins and Fluc heterodimers suggests that both pores are not required: the residues that define one pore are well-conserved, whereas those that define the second have drifted substantially[32]. This pore degradation is particularly evident in the phenylalanine box motif, for which each subunit contributes one phenylalanine per pore (Figure 2.4, B, 2.4, C). The fused eukaryotic proteins always retain the first Phe in the N-terminal domain, and the second Phe in the C-terminal domain, which combine to form one complete pore. The heterodimers also tend to conserve opposite phenylalanines in the two domains, although which pair is conserved differs among heterodimeric lineages (Figure 2.4, C). Among pore-lining residues that are conserved in both domains of the S. cerevisiae FEX protein, including the TM2 asparagine, mutation of the one located in the degraded pore is less detrimental for fluoride resistance, confirming experimentally that the two homologous domains in FEX proteins have evolved substantial functional asymmetry[5]. It is not known whether

the degraded pore retains the ability to bind anions, or still plays any functional role.

2.7 Future Directions

As a model of a dual-topology to inverted repeat evolutionary trajectory, key questions for the Fluc family include: Is the functioning homodimer truly symmetrical, as in the crystal structure, or does it adopt an asymmetric functional form, like fellow dual-topology homodimer EmrE[14]? How much of a constraint on function and folding is dual topology, and how are functions elaborated and improved upon once that limitation is lifted?

One could imagine obvious advantages of duplication. The homodimeric constraint enforces 2-fold symmetry and disallows much structurally extraneous evolutionary drift, even among functionally redundant features – exemplified in Flucs by the double-barreled construction. After duplication, some of the structural constraints relax, allowing more sampling in sequence space, a process that is archived in the degradation of the second pore in duplicated Flucs. Furthermore, the two subunits can adopt more specialized roles after duplication, once insertion bias develops and the subunits are fixed in their orientations, as in Figure 2.1, B. This allows asymmetrical presentation to the periplasm and cytoplasm, and the development of side-specific features. There is clear value in developing asymmetry: perhaps the asymmetric protein displays a broader, electropositive vestibule to the inside only, funneling the dilute anionic toxin in the direction of export. Whether such imagined embellishments actually occur remains to be seen, and will require further structural and functional characterization of heterodimeric and inverted-repeat Flucs.

Further analysis of the central Na+ is also likely to be revealing. In some respects, the binding site, located at breaks in TM helices, is reminiscent of Na+ binding sites in Na+-coupled transporters. In other respects, such as the tetrahedral coordination sphere comprised by backbone carbonyls alone, the Na+ binding site in the Flucs is singular. Understanding whether this binding site has broader implications requires determining

a structural or functional rationale for the central Na+ in the homodimers, determining whether the ion is conserved in heterodimers and inverted repeat proteins, and if so, whether asymmetric relaxation has provided additional ligands to an energetically favorable 5- or 6-coordinated form.

Other goals for the dual-topology protein field include identifying molecular functions for, and structurally characterizing other single domain building blocks, like the dual-topology DUF606 family, which, like the Flucs, is particularly rich in topologies including inverted repeats[24], or the DedA family, which has a hydropathy profile intriguingly similar to LeuT proteins[11]. An analogous evolutionary case study with parallel single-domain and fused multi-domain counterparts is underway, with a number of recent crystal structures of two-domain SWEET and single-domain "semi-SWEET" sugar transporters, and the Pnu vitamin transporter, which is unrelated in sequence but similar in fold[10,41–45]. Like the Flucs, the fused SWEETS and single-domain semi-SWEETS are segregated in eukaryotic and prokaryotic genomes, respectively, suggesting some selective advantage for fusion in eukaryotes that is not as strong in prokaryotic membranes.

Via dual-topology Flucs and other modern day "building block" membrane proteins, we can glimpse folds that, for whatever reason, remained in evolutionary infancy. As the number of newly discovered membrane protein folds dwindles, and interest increases in engineering new functions onto existing folds, dual-topology proteins and their kin provide a unique opportunity to understand how nature's fold prototypes changed in time.

2.8 Acknowledgments

We are grateful to Nicholas Last for careful reading of this manuscript. This work was supported by the National Institutes of Health (NIGMS R00GM111767).

References

(1) Forrest, L. R.; Kraemer, R.; Ziegler, C. The Structural Basis of Secondary Active Transport Mechanisms. *Biochimica Et Biophysica Acta-Bioenergetics* **2011**, *1807* (2), 167–188. https://doi.org/10.1016/j.bbabio.2010.10.014.

(2) Drew, D.; Boudker, O. Shared Molecular Mechanisms of Membrane Transporters. In *Annual Review of Biochemistry, Vol 85*; Kornberg, R. D., Ed.; 2016; Vol. 85, pp 543–572. https://doi.org/10.1146/annurev-biochem-060815-014520.

(3) Forrest, L. R. Structural Symmetry in Membrane Proteins. In *Annual Review of Biophysics, Vol 44*; Dill, K. A., Ed.; 2015; Vol. 44, pp 311–337. https://doi.org/10.1146/annurev-biophys-051013-023008.

(4) von Heijne, G. Control of Topology and Mode of Assembly of a Polytopic Membrane Protein by Positively Charged Residues. *Nature* **1989**, *341* (6241), 456–458. https://doi.org/10.1038/341456a0.

(5) Smith, K. D.; Gordon, P. B.; Rivetta, A.; Allen, K. E.; Berbasova, T.; Slayman, C.; Strobel, S. A. Yeast Fex1p Is a Constitutively Expressed Fluoride Channel with Functional Asymmetry of Its Two Homologous Domains. *Journal of Biological Chemistry* **2015**, *290* (32), 19874–19887. https://doi.org/10.1074/jbc.M115.651976.

(6) Gafvelin, G.; Sakaguchi, M.; Andersson, H.; vonHeijne, G. Topological Rules for Membrane Protein Assembly in Eukaryotic Cells. *Journal of Biological Chemistry* **1997**, *272* (10), 6119–6127. https://doi.org/10.1074/jbc.272.10.6119.

(7) Rapp, M.; Granseth, E.; Seppala, S.; von Heijne, G. Identification and Evolution of Dual-Topology Membrane Proteins. *Nature Structural & Molecular Biology* **2006**, *13* (2), 112–116. https://doi.org/10.1038/nsmb1057.

(8) Woodall, N. B.; Yin, Y.; Bowie, J. U. Dual-Topology Insertion of a Dual-Topology Membrane Protein. *Nature Communications* **2015**, *6*, 8099. https://doi.org/10.1038/ncomms9099.

(9) Forrest, L. R.; Rudnick, G. The Rocking Bundle: A Mechanism for Ion-Coupled Solute Flux by Symmetrical Transporters. *Physiology* **2009**, *24* (6), 377–386. https://doi.org/10.1152/physiol.00030.2009.

(10) Keller, R.; Ziegler, C.; Schneider, D. When Two Turn into One: Evolution of Membrane Transporters from Half Modules. *Biological Chemistry* **2014**, *395* (12), 1379–1388. https://doi.org/10.1515/hsz-2014-0224.

(11) Khafizov, K.; Staritzbichler, R.; Stamm, M.; Forrest, L. R. A Study of the Evolution of Inverted-Topology Repeats from LeuT-Fold Transporters Using AlignMe. *Biochemistry* **2010**, *49* (50), 10702–10713. https://doi.org/10.1021/bi101256x.

(12) Diaz-Perez, C.; Cervantes, C.; Campos-Garcia, J.; Julian-Sanchez, A.; Riveros-Rosas, H. Phylogenetic Analysis of the Chromate Ion Transporter (CHR) Superfamily. *Febs Journal* **2007**, *274* (23), 6215–6227. https://doi.org/10.1111/j.1742-4658.2007.06141.x.

(13) Schuldiner, S. EmrE, a Model for Studying Evolution and Mechanism of Ion-Coupled Transporters. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2009**, *1794* (5), 748–762. https://doi.org/10.1016/j.bbapap.2008.12.018.

(14) Morrison, E. A.; DeKoster, G. T.; Dutta, S.; Vafabakhsh, R.; Clarkson, M. W.; Bahl, A.; Kern, D.; Ha, T.; Henzler-Wildman, K. A. Antiparallel EmrE Exports Drugs by Exchanging Between Asymmetric Structures. *Nature* **2012**, *481* (7379), 45–U50. https://doi.org/10.1038/nature10703.

(15) Ji, C.; Stockbridge, R. B.; Miller, C. Bacterial Fluoride Resistance, Fluc Channels, and the Weak Acid Accumulation Effect. *Journal of General Physiology* **2014**, *144* (3), 257–261. https://doi.org/10.1085/jgp.201411243.

(16) Baker, J. L.; Sudarsan, N.; Weinberg, Z.; Roth, A.; Stockbridge, R. B.; Breaker, R. R. Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride. *Science* **2012**, *335* (6065), 233–235. https://doi.org/10.1126/science.1215063.

(17) Stockbridge, R. B.; Robertson, J. L.; Kolmakova-Partensky, L.; Miller, C. A Family of Fluoride-Specific Ion Channels with Dual-Topology Architecture. *Elife* **2013**, *2*, e01084. https://doi.org/10.7554/eLife.01084.

(18) Li, S.; Smith, K. D.; Davis, J. H.; Gordon, P. B.; Breaker, R. R.; Strobel, S. A. Eukaryotic Resistance to Fluoride Toxicity Mediated by a Widespread Family of Fluoride Export Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (47), 19018–19023. https://doi.org/10.1073/pnas.1310439110.

(19) Stockbridge, R. B.; Koide, A.; Miller, C.; Koide, S. Proof of Dual-Topology Architecture of Fluc F- Channels with Monobody Blockers. *Nature Communications* **2014**, *5*, 5120. https://doi.org/10.1038/ncomms6120.

(20) Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **2010**, *59* (3), 307–321. https://doi.org/10.1093/sysbio/syq010.

(21) Edgar, R. C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Research* **2004**, *32* (5), 1792–1797. https://doi.org/10.1093/nar/gkh340.

(22) Pei, J.; Kim, B.-H.; Grishin, N. V. Promals3d: A Tool for Multiple Protein Sequence and Structure Alignments. *Nucleic Acids Research* **2008**, *36* (7), 2295–2300. https://doi.org/10.1093/nar/gkn072.

(23) Tsirigos, K. D.; Peters, C.; Shu, N.; Kall, L.; Elofsson, A. The TOPCONS Web Server for Consensus Prediction of Membrane Protein Topology and Signal Peptides. *Nucleic Acids Research* **2015**, *43* (W1), W401–W407. https://doi.org/10.1093/nar/gkv485.

(24) Lolkema, J. S.; Dobrowolski, A.; Slotboom, D.-J. Evolution of Antiparallel Two-Domain Membrane Proteins: Tracing Multiple Gene Duplication Events in the Duf606 Family. *Journal of Molecular Biology* **2008**, *378* (3), 596–606. https://doi.org/10.1016/j.jmb.2008.03.005.

(25) Marquis, R. E.; Clock, S. A.; Mota-Meira, M. Fluoride and Organic Weak Acids as Modulators of Microbial Physiology. *Fems Microbiology Reviews* **2003**, *26* (5), 493–510. https://doi.org/10.1016/S0168-6445(02)00143-2.

(26) Qin, R.; Chai, G. Q.; Brewer, J. M.; Lovelace, L. L.; Lebioda, L. Fluoride Inhibition of Enolase: Crystal Structure and Thermodynamics. *Biochemistry* **2006**, *45* (3), 793–800. https://doi.org/10.1021/bi051558s.

(27) Samygina, V. R.; Moiseev, V. M.; Rodina, E. V.; Vorobyeva, N. N.; Porov, A. N.; Kurilova, S. A.; Nazarova, T. I.; Avaeva, S. M.; Bartunik, H. D. Reversible Inhibition of Escherichia Coli Inorganic Pyrophosphatase by Fluoride: Trapped Catalytic Intermediates in Cryo-Crystallographic Studies. *Journal of Molecular Biology* **2007**, *366* (4), 1305–1317. https://doi.org/10.1016/j.jmb.2006.11.082.

(28) Weinstein, L. H.; Davison, A. *Fluorides in the Environment: Effects on Plants and Animals*; Cabi, 2004.

(29) Stockbridge, R. B.; Lim, H.-H.; Otten, R.; Williams, C.; Shane, T.; Weinberg, Z.; Miller, C. Fluoride Resistance and Transport by Riboswitch-Controlled CLC Antiporters. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109* (38), 15289–15294. https://doi.org/10.1073/pnas.1210896109.

(30) Hille, B. *Ion Channels of Excitable Membranes*, 3rd ed.; Sinauer: Sunderland, Mass, 2001.

(31) Cametti, M.; Rissanen, K. Highlights on Contemporary Recognition and Sensing of Fluoride Anion in Solution and in the Solid State. *Chemical Society Reviews* **2013**, *42* (5), 2016–2038. https://doi.org/10.1039/c2cs35439j.

(32) Stockbridge, R. B.; Kolmakova-Partensky, L.; Shane, T.; Koide, A.; Koide, S.; Miller, C.; Newstead, S. Crystal Structures of a Double-Barrelled Fluoride Ion Channel. *Nature* **2015**, *525* (7570), 548–+. https://doi.org/10.1038/nature14981.

(33) Harding, M. M. Metal-Ligand Geometry Relevant to Proteins and in Proteins: Sodium and Potassium. *Acta Crystallographica Section D-Biological Crystallography* **2002**, *58* (5), 872–874. https://doi.org/10.1107/S0907444902003712.

(34) Torrance, G. M.; Leader, D. P.; Gilbert, D. R.; Milner-White, E. J. A Novel Main Chain Motif in Proteins Bridged by Cationic Groups: The Niche. *Journal of Molecular Biology* **2009**, *385* (4), 1076–1086. https://doi.org/10.1016/j.jmb.2008.11.007.

(35) Philip, V.; Harris, J.; Adams, R.; Nguyen, D.; Spiers, J.; Baudry, J.; Howell, E. E.; Hinde, R. J. A Survey of Aspartate-Phenylalanine and Glutamate-Phenylalanine Interactions in the Protein Data Bank: Searching for Anion-Pi Pairs. *Biochemistry* **2011**, *50* (14), 2939–2950. https://doi.org/10.1021/bi200066k.

(36) Jackson, M. R.; Beahm, R.; Duvvuru, S.; Narasimhan, C.; Wu, J.; Wang, H.-N.; Philip, V. M.; Hinde, R. J.; Howell, E. E. A Preference for Edgewise Interactions Between Aromatic Rings and Carboxylate Anions: The Biological Relevance of Anion-Quadrupole Interactions. *Journal of Physical Chemistry B* **2007**, *111* (28), 8242–8249. https://doi.org/10.1021/jp0661995.

(37) Dickson, V. K.; Pedi, L.; Long, S. B. Structure and Insights into the Function of a Ca2+-Activated Cl- Channel. *Nature* **2014**, *516* (7530), 213–+. https://doi.org/10.1038/nature13913.

(38) Last, N. B.; Kolmakova-Partensky, L.; Shane, T.; Miller, C. Mechanistic Signs of Double-Barreled Structure in a Fluoride Ion Channel. *Elife* **2016**, *5*, e18767. https://doi.org/10.7554/eLife.18767.

(39) Turman, D. L.; Nathanson, J. T.; Stockbridge, R. B.; Street, T. O.; Miller, C. Two-Sided Block of a Dual-Topology F- Channel. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112* (18), 5697–5701. https://doi.org/10.1073/pnas.1505301112.

(40) Turman, D. L.; Stockbridge, R. B. Mechanism of Single- and Double-Sided Inhibition of Dual Topology Fluoride Channels by Synthetic Monobodies. *Journal of General Physiology* **2017**, *149* (4), 511–522. https://doi.org/10.1085/jgp.201611747.

(41) Jaehme, M.; Guskov, A.; Slotboom, D. J. Crystal Structure of the Vitamin B-3 Transporter PnuC, a Full-Length SWEET Homolog. *Nature Structural & Molecular Biology* **2014**, *21* (11), 1013–1015. https://doi.org/10.1038/nsmb.2909.

(42) Lee, Y.; Nishizawa, T.; Yamashita, K.; Ishitani, R.; Nureki, O. Structural Basis for the Facilitative Diffusion Mechanism by SemiSWEET Transporter. *Nature Communications* **2015**, *6*, 6112. https://doi.org/10.1038/ncomms7112.

(43) Tao, Y.; Cheung, L. S.; Li, S.; Eom, J.-S.; Chen, L.-Q.; Xu, Y.; Perry, K.; Frommer, W. B.; Feng, L. Structure of a Eukaryotic SWEET Transporter in a Homotrimeric Complex. *Nature* **2015**, *527* (7577), 259–+. https://doi.org/10.1038/nature15391.

(44) Xu, Y.; Tao, Y.; Cheung, L. S.; Fan, C.; Chen, L.-Q.; Xu, S.; Perry, K.; Frommer, W. B.; Feng, L. Structures of Bacterial Homologues of SWEET Transporters in Two Distinct Conformations. *Nature* **2014**, *515* (7527), 448–+. https://doi.org/10.1038/nature13670.

(45) Jaehme, M.; Guskov, A.; Slotboom, D. J. Pnu Transporters: Ain't They SWEET? *Trends in Biochemical Sciences* **2016**, *41* (2), 117–118. https://doi.org/10.1016/j.tibs.2015.11.013.

CHAPTER 3

Guanidinium Export Is the Primal Function of SMR Family Transporters

This chapter is adapted from the following published article:

Kermani, A. A., Macdonald, C. B., Gundepudi, R. & Stockbridge, R. B. Guanidinium export is the primal function of SMR family transporters. Proc Natl Acad Sci USA 115, 3060–3065 (2018).

Author contributions: A.A.K., C.B.M., and R.B.S. designed research; A.A.K., C.B.M., R.G., and R.B.S. performed research; A.A.K., C.B.M., and R.B.S. analyzed data; and A.A.K., C.B.M., and R.B.S. wrote the paper.

3.1 Introduction

The small multidrug resistance (SMR) family comprises small four-transmembrane (TM) helical proteins that first became prominent because of their role in antibiotic and antiseptic resistance and their rare "dual topology" architecture[1–4]. At ~100 residues, they are among the smallest known membrane transport proteins, and their small size makes them amenable to many biophysical analyses. The best studied member of the SMR family is the multidrug efflux pump EmrE, which protects bacteria from an array of hydrophobic, cationic antimicrobials including ethidium and methyl viologen[1]. SMRs that are phenotypically similar to EmrE have been disseminated widely among bacteria through horizontal gene transfer[5]. But many bacterial genomes harbor SMR proteins that are highly similar to the multidrug pumps but that do not transport drugs (variously annotated qacE, sugE,

ebrA, ebrB, or emrE). For example, in Escherichia coli, EmrE and its non–drug-exporting SMR homolog share 33% sequence identity and 61% similarity, including a conspicuous midmembrane glutamate that is essential for function in EmrE[6].

The function of these "other" SMRs has remained enigmatic, from their initial erroneous assignment as chaperone-like molecules[7], stemming from a PCR error that interfered with an adjacent chaperone[8,9] (to this day the genes are often annotated "sugE," a misnomer that stands for "Suppresor of GroEL"), to the unsuccessful attempts to measure transport or conformational change in the presence of drugs[9–11]. Moreover, no clear sequence-based criteria have been identified to distinguish SMR proteins that export drugs and those that do not. The determination of whether an SMR contributes to a pathogen's drug resistance is typically determined using case-by-case bacterial fitness assays[5]. And without understanding the native role of SMR proteins, the evolutionary history that spawned this particular variety of multidrug resistance in bacteria is indecipherable.

A solid lead on the native role of the SMRs was provided by the recent discovery of bacterial operons dedicated to guanidinium (Gdm+) metabolism[12–15]. These operons are controlled by Gdm+-binding riboswitches, which frequently up-regulate either enzymes that chemically modify Gdm+ or membrane proteins from the SMR family[15]. These riboswitches are widespread in both Gram-negative and Gram-positive bacteria, including firmicutes, cyanobacteria, actinobacteria, proteobacteria, and others. There exist three structurally distinct Gdm+-responsive riboswitches, termed guanidine-I (or ykkC)[12,16,17], guanidine-II (or mini-ykkC)[14,18,19], and guanidine-III (or ykkC-III)[13]. These riboswitches are mechanistically variable as well, with guanidine-I exerting transcriptional control[12], whereas guanidine-II and -III are translational regulators[13,14]. Their common feature is the genes whose expression they control, including, for all three riboswitch classes, genes that are annotated as SMRs[15]. That the nondrug transporting SMRs might instead transport Gdm+ is an appealing hypothesis. Despite the obvious difference in size between Gdm+ and the bulky substrates transported by EmrE, the substrates share a chemical

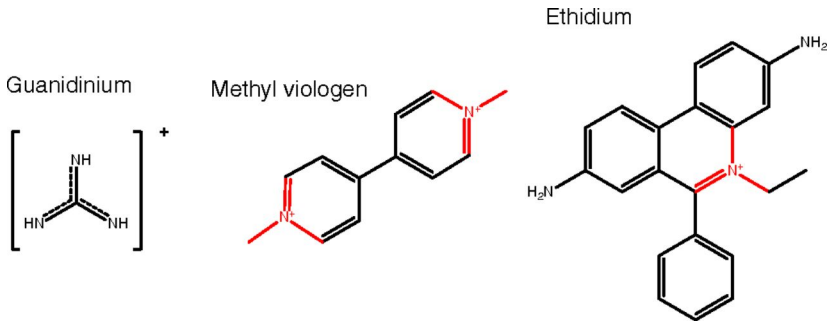

Figure 3.1: Scheme 1. Structural comparison of Gdm+ and some canonical EmrE substrates.

essence that alludes to the familial relationship: cationic aromaticity. Gdm+ is a chemically unusual "Y-aromatic"—a planar molecule with electrons delocalized in pi orbitals over both faces[20,21]. It is easy to imagine how a Gdm+ exporter might share common traits with a promiscuous exporter of larger substrates with shared chemical attributes (Scheme 1 3.1). Indeed, experiments have shown that a representative riboswitch-controlled SMR binds Gdm+[12], but there is no indication of whether Gdm+ is the transported substrate or merely an effector, whether drug- and Gdm+-related functions overlap, and how broadly Gdm+ exporters are distributed in the SMR family. Here we demonstrate that Gdm+ export is the main functional role of the SMR family, whereas the experimentally validated multidrug exporters[5] are all confined to a single clade of the phylogenetic tree. We characterize several proteins dispersed throughout the phylogeny, showing that they couple Gdm+ export to the proton motive force, exchanging one Gdm+ for two protons. These Gdm+-exporting SMRs, which we call Gdx (guanidinium exporter), are mechanistically distinct from the multidrug exporters in two important ways: They are highly selective for their substrate, and they couple its export tightly with the import of two protons.

3.2 Results

3.2.1 Most SMR Proteins Are Associated with Guanidine Riboswitches

Over half of guanidine-riboswitch controlled genes are annotated as members of the SMR family, but of the SMRs with validated drug-export functions, none are associated with a riboswitch. [The E. coli riboswitch-controlled SMR, also called SugE, has been shown not to transport drugs[10].] To examine the evolutionary relationships between the riboswitch-controlled proteins and multidrug exporters from the SMR family, we constructed sequence alignments and a phylogeny from a representative sample of all known SMRs. The SMR family comprises predicted homo-oligomers, encoded by a single gene, and predicted heterodimers, encoded by adjacent genes in the genome that arose from a gene duplication event. Analysis of the sequences shows that, in the single gene versions, the solvent-accessible loops possess few charged residues, suggesting dual topology, the propensity for subunits to insert into the membrane in both inward- and outward-facing orientations, setting up the possibility for antiparallel homodimers to occur[3,22]. In contrast, the subunits of putative heterodimers possess oppositely directed charge biases (figure 3.2 A), circumstantial evidence that the subunits oligomerize with anti-parallel orientation. In dual topology proteins, duplication events are typically followed by the relatively rapid evolution of positive charge bias (lysines and arginines) on the cytoplasmic loops, fixing the topology of the heterodimer subunits[3,23,24]. Our preliminary analyses suggested that several different duplication events have occurred in different SMR lineages, each yielding pairs with oppositely biased charge distributions. This example of convergent evolution, the evolution of similarly biased loops among several independent duplication events, is in conflict with the assumption of divergent evolution necessary for phylogenetic reconstruction. Therefore, we estimated solvent-exposed sequences based on structural homology to EmrE and trimmed these sequences from the alignment, a necessary maneuver to avoid bias by convergent evolution.

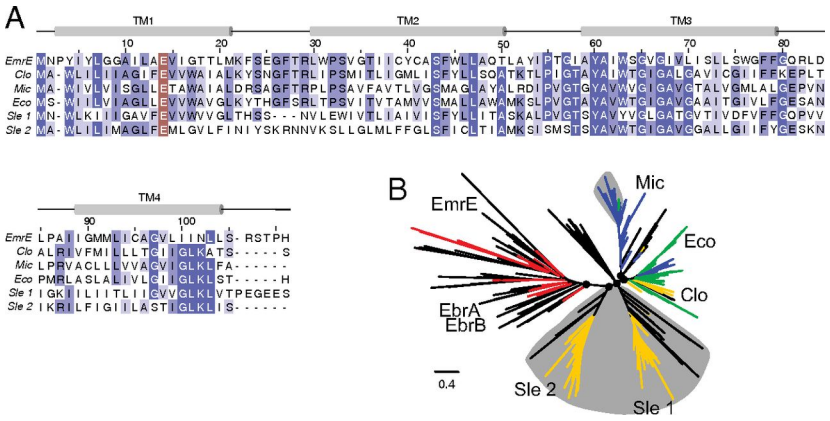

Figure 3.2: Phylogeny and sequence alignment of the SMR family. (A) Sequence alignment of representa- tive Gdx proteins in this study (Table 1) and multi- drug exporter EmrE. (B) Phylogeny of representative SMR proteins constructed using maximum likelihood methods. Major nodes were statistically significant (support > 0.9), indicated with a black circle. Tree is annotated by colored branches: red, experimentally implicated in multidrug resistance; yellow, associated with guanidine-I riboswitch (12); green, associated with guanidine-II riboswitch (14); blue, associated with guanidine-III riboswitch (13). Duplication events leading to probable heterodimers are indicated by gray shading. Functionally validated proteins are labeled with the protein name. EbrA/EbrB is a Bacillus subtilis drug exporter that functions as an obligate heterodimer (45).

Our maximum likelihood phylogeny (figure 3.2 B) provides strong support for five major distinct protein clades: one each comprised of proteins controlled by the guanidine-II and guanidine-III riboswitches, two clades of proteins associated with the guanidine-I riboswitches, and a fifth branch in which experimentally validated multidrug exporters (reviewed in[5]) cluster. Notably, most of the proteins from Gdx branches are associated with riboswitches, whereas many of the proteins from the multidrug resistance branch are associated with transposons or plasmids[5]. The SMRs controlled by guanidine-I riboswitches are distributed across two deeply diverged clades. Closer inspection of these SMRs reveals that most of them likely form heterodimeric pairs, with one partner from each clade. The genes encoding these pairs are almost always located adjacent to each other in genomes, and the paired proteins themselves possess opposite charge biases, implying that their topologies became fixed after an ancient gene duplication event. Other more recent duplication events yielding probable heterodimers, notably in the guandine-

III clade, are also evident.

3.2.2 Riboswitch-Controlled SMRs Are Proton-Coupled Gdm+ Transporters

To identify the molecular function of riboswitch-controlled SMRs, we identified, heterologously expressed, and purified representatives controlled by each riboswitch subtype, including both a homodimer and a heterodimer from the guanidine-I family (Table 3.1 and supplementary figure A.1). Proteins were reconstituted into liposomes to assess Gdm+ transport. Proteoliposomes were initially loaded with 10 mM unlabeled Gdm+ and then diluted into an external solution containing 20 µM 14C-labeled Gdm+. In the presence of a Gdm+ transport pathway, the outward-directed Gdm+ gradient drives uptake of radiolabeled Gdm+ into the liposomes. All four proteins catalyzed robust uptake of 14C-Gdm+. Gdm+ exchange was abolished by mutation of a midmembrane glutamate, E13 (figure 3.3 A), which is completely conserved among SMRs and known to be essential for substrate binding and transport in EmrE[6]. Gdm+ exchange is limited to riboswitch-controlled SMRs; despite its promiscuity among drugs, EmrE fails to bind or transport Gdm+ (figure 3.3 A and supplementary figure A.2).

Table 3.1: Gdx proteins examined in this study

Protein name	Source	Sequence ID, NCBI	Riboswitch
Gdx-Eco	E. coli	WP_032226439.1	Guanidine-II
Gdx-Clo	Clostridiales bacterium oral taxon 876	WP_021653285.1	Guanidine-I
Gdx-Mic	Micromonospora	WP_013284696.1	Guanidine-III
Gdx-Sle 1	Staphylococcus lentus (subunit 1)	WP_016999238.1	Guanidine-I
Gdx-Sle 2	S. lentus (subunit 2)	WP_016999239.1	Guanidine-I

Based on the conservation of the TM1 glutamate, we speculated that Gdx, like EmrE, might couple substrate exchange to a proton gradient. To test this, we examined whether

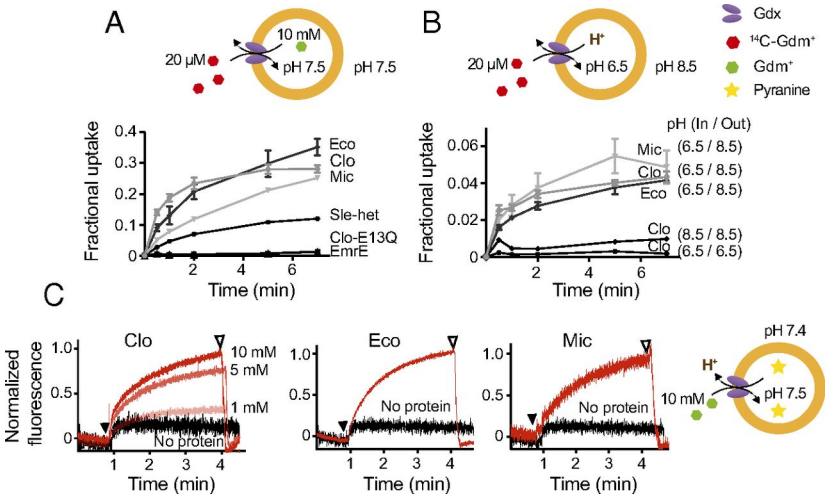

Figure 3.3: Gdx-mediated Gdm+ transport. For A–C, liposome cartoons illustrate experimental conditions. (A) Time course of 14C-Gdm+ uptake into liposomes reconstituted with different SMR proteins. Uptake is reported as a fraction of total 14C counts added at the outset of the experiment. Individual data points are the mean and SEM of at least three independent experiments from at least two independent protein preps. Where not visible, error bars are smaller than the diameter of the point. (B) Time course of 14C-Gdm+ uptake into Gdx-containing liposomes holding the indicated internal/external pH gradients. Values are the mean and SEM of at least three independent experiments. (C) Fluorescence of pyranine trapped inside liposomes. Addition of guanidinium to the ex- ternal buffer is indicated by solid triangle. After reaction reached steady state, FCCP was added to collapse the pH gradient (open triangle).

an H+ gradient could drive 14C-Gdm+ accumulation in liposomes reconstituted with each of the three Gdx homodimers. Proteoliposomes were prepared with an internal pH of 6.5 and diluted into an otherwise identical buffer with a pH of 8.5. This outward-directed H+ gradient induced 14C-Gdm+ uptake into the proteoliposomes (figure 3.3 B). Radioactive uptake was not observed in the absence of a pH gradient, at either pH value. We also monitored the ability of Gdm+ to drive H+ translocation. Using a pH-sensitive fluorophore, pyranine, we monitored pH changes in the liposome interior upon addition of external Gdm+. The fluorescence intensity increased when Gdm+ was added to the external solution, indicating H+ efflux from the liposomes (figure 3.3 C). The steady-state magnitude of H+ efflux increased in tandem with added Gdm+. H+ transport was not observed in liposomes without protein, and addition of urea (an uncharged structural analog of Gdm+) or arginine to Gdx-Clo proteoliposomes also failed to reproduce transport (supplementary figure A.3).

3.2.3 Gdx Proteins Are Selective for Gdm+.

A hallmark of multidrug exporting SMRs, like EmrE, is their ability to export diverse cationic substrates that range in size and in valence[25,26]. To determine whether Gdx is similarly nonselective, we monitored 14C-Gdm+ uptake in the presence of a number of different substrates, including urea, which is a neutral, isosteric analog of Gdm+; common biological compounds with guanidinyl or quaternary amine moieties (arginine, agmatine, creatine, betaine, choline, and spermidine); EmrE substrates (ethidium and TPP+); and N-substituted guanidinyl analogs (methyl- and aminoguanidinium) (Fig. 3). Among these, only aminoguanidinium and methylguanidinium were comparable to cold Gdm+ in competition with 14C-Gdm+ accumulation for all three homologs (figure 3.4 A). To distinguish between block and permeation, counterflow assays, in which liposomes were preloaded with amino- or methyl-Gdm+, rather than Gdm+, were performed (figure 3.4 B). Both N-substituted analogs were exchanged for radiolabeled Gdm+, albeit more slowly. Gdm+

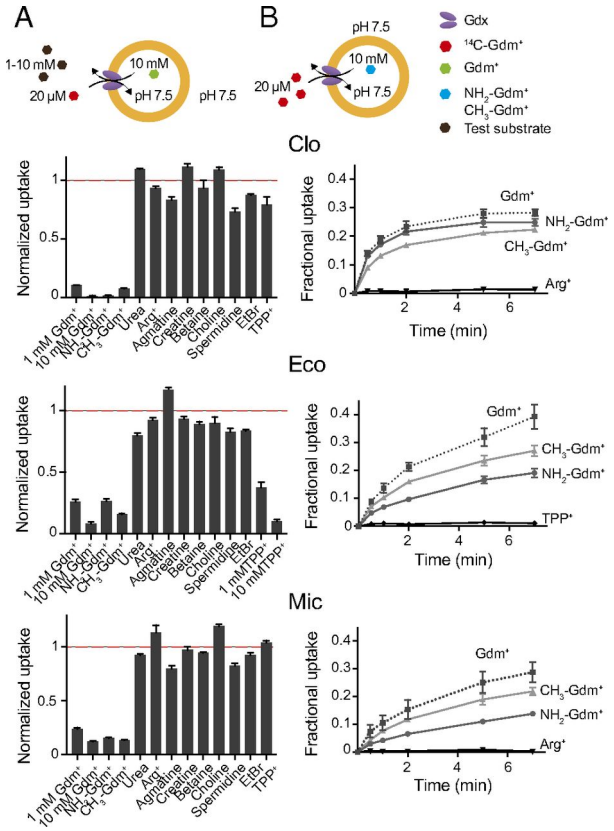

Figure 3.4: Substrate selectivity of Gdx proteins. (A) Effect of external test sub- strates on 14C-Gdm+ uptake. 14C-Gdm+ uptake was measured after 3 min, in the presence of 1 mM (or 10 mM, where indicated) test substrate. Counts were normalized with respect to 14C uptake in the absence of an external test substrate (control bar). (B) Time course of 14C-Gdm+ exchange with competi- tive substrates for Gdx-Clo, Gdx-Eco, and Gdx-Mic. Liposomes were loaded with 10 mM test substrate to monitor exchange with external 14C-Gdm+. Data points represent the mean and SEM of three independent measurements. Where not visible, error bars are smaller than the diameter of the point.

uptake in Gdx-Eco was also competitively inhibited by TPP+, in accord with previous reports that it binds this substrate[10], but TPP+ was not transported in our experiments (figure 3.4 B).

3.2.4 H+/Gdm+ Antiport Occurs with 2:1 Stoichiometry

Recent experiments with EmrE have fashioned a model of a transporter in which substrate promiscuity and transport stoichiometry are intimately linked. Depending on the substrate, EmrE undergoes some combination of 2:1 and 1:1 proton:drug exchange[27,28].

Given the selective nature of the Gdxs, we sought to determine whether they behave more like classical transporters, with tightly coupled exchange, or whether they share flexible stoichiometry with their multidrug-transporting family members. To this end, we loaded Gdx-Eco proteoliposomes with pyranine and diluted these into an otherwise identical buffer. A membrane potential was set using the potassium ionophore valinomycin in the presence of a K+ gradient, which effectively "clamps" the voltage across the membrane at the Nernst potential for K+. When the K+ concentration is equal on both sides of the membrane, no substrate movement occurs because all ions in the system are at chemical and electrical equilibrium (figure 3.5 A, middle trace). When we apply an inwardly directed potassium gradient (positive inside voltage), we observe an increase in fluorescence, indicating basification of the liposome interior and net proton export (figure 3.5 A, top trace). When the K+ gradient is reversed, so that the liposomes are negative inside, we observe net proton import (figure 3.5 A, bottom trace). These experi- ments show that membrane potential alone can drive substrate translocation, indicating an electrogenic transport cycle with >1 H+ per Gdm+. Electrogenic transport was also tested and observed for a second homolog, Gdx-Clo (supplementary figure A.4). To quantify the stoichiometry with greater precision for Gdx-Eco, we introduced a 10-fold Gdm+ gradient across the membrane and then interrogated H+ movement in the presence of several different K gradients to determine the point at which substrate translocation achieved electrochemical equilibrium (Erev), with no net H+ movement[29] (Fig. 4 B and C). Under the conditions of this experiment, the observed Erev value of −60 mV corresponds to a 2 H+:1 Gdm+ stoichiometry.

If present, simultaneous 1:1 transport cycles would contribute to H+ efflux (net increase in fluorescence) at all voltages due to the 10-fold Gdm+ gradient applied in these experiments, and net proton movement would reflect the sum of 2:1 and 1:1 transport cycles. Since we do not see evidence of proton uptake at −60 mV (where proton movement by 2:1 transport cycles is null), the most straightforward interpretation of our data

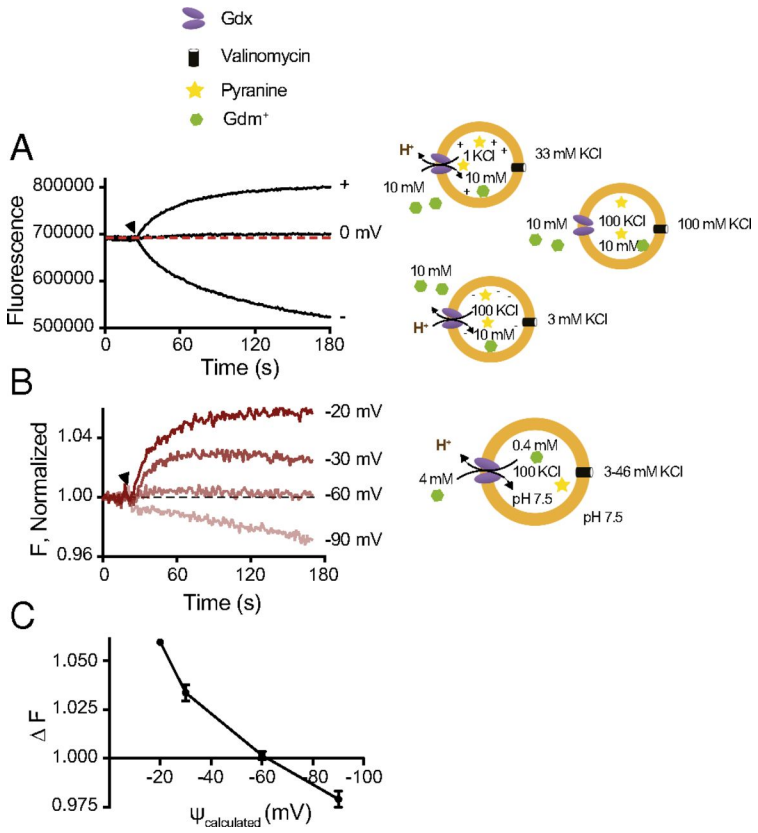

Figure 3.5: Stoichiometry of Gdm+/H+ exchange. (A) Proton movement across Eco-Gdx proteoliposomes monitored using pyranine fluorescence. Proteoli- posomes were reconstituted with equal internal and external pH and Gdm+ concentration. Membrane potentials (approximately +90, 0, and −90 mV from top to bottom) were imposed by addition of valinomycin to liposomes reconstituted with different K+ gradients, indicated by the liposome cartoon immediately to the right of each trace. (B) Proteoliposomes reconstituted with equal pH and a 10-fold outward-directed Gdm+ gradient. Proton move- ment was monitored with pyranine fluorescence in response to different membrane potentials imposed using a potassium gradient and valinomycin. Valinomycin addition is indicated by the triangle. Data were normalized by dividing by the baseline measurement (20 s running average before vali- nomycin addition). (C) Change in fluorescence as a function of membrane potential, measured as in B at 3 min. Error bars represent the mean and SEM of three independent measurements.

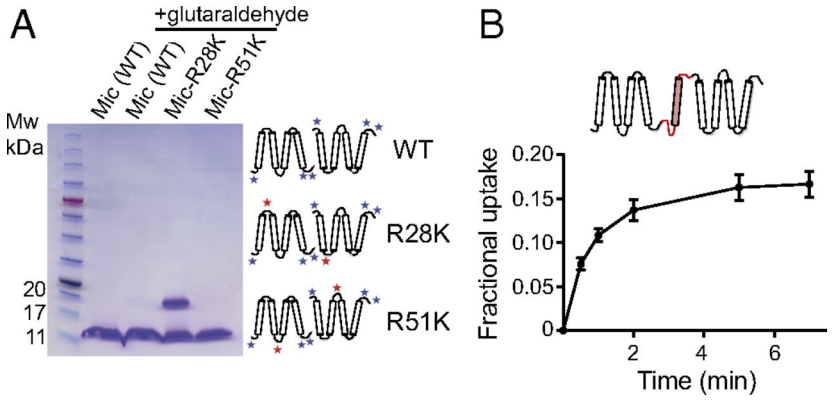

Figure 3.6: Architecture of Gdx proteins. (A) Glutaraldehyde cross-linking of Gdx-Mic. Cartoon indicates the distribution of primary amines in Gdx-Mic and Gdx-Mic mutants, with blue stars indicating native primary amines and red stars representing introduced lysine residues. (B) 14C-Gdm+ uptake into liposomes by antiparallel Gdx-Eco concatamer.

is that Gdx-Eco described by Schuldiner and coworkers[30] and later used to correctly predict the topology of Fluc family fluoride channels[31]. We constructed a fused homodimeric construct linked by a nondimerizeable glycophorin A TM helix to allow the subunits to adopt an antiparallel topology[31]. This construct expressed robustly in E. coli (supplementary figure A.5 B), and activity assays showed that the antiparallel concatamer transports Gdm+ at low protein:lipid ratios (figure 3.6 B). We also attempted to design a concatamer with short, charged linker to force a parallel construction. However, we were unable to express and purify this protein.

3.3 Discussion

These experiments establish that diverse SMR proteins are Gdm+/H+ antiporters, exporting Gdm+ from the cytoplasm by coupling its translocation to the proton motive force. These results resolve a decades-long mystery about the function of the non–drug-exporting SMRs. The diverse homologs selected here are all notable for their selectivity for Gdm+ over other common biological molecules with guanidinyl or primary amine moieties, and we expect that these traits are general among SMR proteins in riboswitch-associated clades. This draws a sharp contrast with the promiscuous multidrug exporters

within the SMR family, including EmrE.

Several factors point to Gdm+ export as the primal role of the SMR family. In both number and diversity, SMRs associated with guanidine riboswitches exceed those that are highly similar to multidrug exporters. Put another way, multidrug exporters are a minority clade in a family of Gdm+ exporters. Gdx proteins are in well-structured sequences within the genomic DNA of both Gram-negative and Gram-positive bacteria, typically associated with regulatory RNA elements and often adjacent to additional genes involved in Gdm+ remediation, suggesting that Gdm+ export function has been useful to microbes over evolutionary time. In contrast, SMR proteins associated with multidrug resistance are typically found on plasmids or associated with horizontal gene transfer elements, suggesting that these genes have spread among microbes in response to relatively more recent stressors[5,12]. Our phylogeny shows that proteins in Gdx clades are genetically diverse, despite performing highly similar functions, and that there are examples of gene duplications that lead to deeply diverged heterodimer subunits that nonetheless assemble to perform Gdm+ export function. The multidrug resistance clade, in contrast, is less deeply diverged, even though its functions are more variable[25,27]. The most parsimonious interpretation of our data is that multidrug export evolved from an ancestral Gdm+ exporter, which is also the forebear of homodimeric and heterodimeric Gdx proteins.

The evolutionary relationships among SMR family members are especially noteworthy in light of the structural similarities and mechanistic differences between EmrE and the Gdx proteins. In terms of architecture, Gdx and EmrE are overwhelmingly similar. They possess ~60% sequence similarity; antiparallel, four-TM helical construction; similarities in hydropathy; and a mechanistically essential glutamate residue in TM1. What, then, is the origin of the differences in both substrate specificity and mechanism? EmrE is mechanistically sloppy, exchanging substrates with different stoichiometries and undergoing uncoupled transport and substrate-free conformational exchange[27,28] . This mechanism has been proposed to facilitate its handling of a broad array of substrates[27]. In contrast,

the Gdx proteins hew closely to 2:1 coupled stoichiometry, and we do not observe 1:1 transport cycles or uncoupled H+ movement by Gdx proteins in our experiments. These two SMR subtypes provide an excellent platform to probe the structural factors that contribute to promiscuity versus specificity and coupling efficiency. Gdx might even realize the original hope for EmrE: serving as a stripped-down model system to understand nature's minimalist solution to coupled, alternating access transport.

A final mystery of these proteins, and the other recently established components of Gdm+-responsive operons, is what role Gdm+ plays in bacteria. These proteins are widespread, conserved, and evidently fulfilling an ancient function. Gdm+ has been recognized as a byproduct of bacterial creatine metabolism since the late 1800s, when high concentrations of Gdm+ were found in spoiled meats[32], but there is scant reference to Gdm+ in the microbial literature before the discovery of the guanidine riboswitches. Recent results show that Gdm+ is produced by bacteria when nutrients are scarce[12], but neither the metabolic pathway that leads to accumulation nor the motivation for export is known. It is certainly possible that Gdm+ is simply a broad-spectrum toxin, to be recognized and removed, like fluoride ion[31,33]. Since a number of metabolites, including arginine, agmatine, and creatine, possess guanidinyl groups, it seems likely that Gdm+ has interfered with chemical biology since early evolution. While biological levels of Gdm+ are probably much less than the ~molar concentrations required to unfold proteins, Gdm+ has been shown to inhibit a number of cellular processes at low-millimeter concentrations. Gdm+ competitively inhibits arginine-binding metabolic enzymes like arginase[34], is a pore blocker of potassium channels[35], and electrostatically inhibits an ATPase involved in prion formation in yeast[36,37]. Further research will be required to determine whether Gdm+ is a general toxin or whether it has some heretofore unknown dedicated role in bacterial life.

3.4 Materials and Methods

3.4.1 Phylogenetic Tree

A representative set of ~400 genes from the Multi_Drug_Res PFAM family (PF00893) were aligned together with a set of genes associated with multidrug resistance in the literature using MAFFT[38]. Based on a structural alignment to EmrE, loop regions were removed from the alignment. A maximum-likelihood phylogenetic tree was constructed using IQ-TREE with evolutionary model JTT+G4, selected by ModelFinder on the basis of the Bayesian information criterion[39,40]. Branch support was calculated using UFBoot to provide rapid bootstrap approximation[41].

3.4.2 Protein Purification and Liposome Reconstitution

Protein was expressed in E. coli C41 (DE3) cells from pET21c plasmids with synthetic Gdx genes (Genscript) with a C-terminal protease recognition site and hexahistidine tag. Expression was induced with 0.2 mM IPTG at an OD600 of 1.0, for 3 h at 37 °C. Cell membranes were extracted with 2% (wt/vol) decyl-β-D-maltoside (DM; Anatrace), and the detergent-solubilized fraction was purified with cobalt affinity beads, washed with 100 mM NaCl and 20 mM imidazole, and eluted with 400 mM imidazole. After histidine tag removal, protein was further purified on a Superdex200 gel-filtration column in 10 mM Hepes, pH 8.1, 200 mM NaCl, and 5 mM DM. Proteoliposomes were prepared from a 3:1 mixture of POPE (1-palmitoyl, 2-oleoylphosphatidylethanolamine) and POPG (1-palmitoyl, 2-oleoylphosphatidylglycerol), with 0.2 µg protein per milligram of lipid (0–1 Gdx dimers per liposome) by dialysis against the desired intraliposomal solution, typically 200 mM NaCl and 20 mM Hepes, pH 7.5. Proteoliposomes were stored in aliquots at −80 °C until use and extruded 21 times through a 400-nm membrane filter before functional assays. Heterodimeric and concatameric Gdx proteins were prepared with slight alterations to this protocol. EmrE was expressed and purified as described previously[42].

3.4.3 Radiolabeled Guanidinium Transport Assays

Liposomes preloaded with 1– 10 mM test ion (typically Gdm+) were passed over a Sephadex G-50 column pre-equilibrated with reaction buffer. Proteoliposomes were diluted two-fold into transport buffer (25 mM Hepes, pH 7.5, and 400 mM sorbitol). Transport was initiated by addition of 20 µM 14C radiolabeled Gdm+ (American Radiolabeled Chemicals, Inc.). At time intervals, 100 µL of reaction mix was passed over a 1.6-mL Dowex ion exchange resin column (N-methyl D-glucamine form) to remove external 14C-Gdm+ and resuspended in liquid scintillation mixture (Ultima Gold; Perkin-Elmer) for scintillation counting. For competition assays, 1–10-mM test substrate was added immediately before initiation with 14C-Gdm+.

3.4.4 H+ Transport and Stoichiometry Assay

Liposomes were preloaded with 1 mM pyranine (trisodium 8-hydroxypyrene-1, 3, 6-trisulfonate; Sigma-Aldrich) and external buffer exchanged using Sephadex G-50, as described above. Proteoliposomes were diluted 200-fold into buffer (25 mM Hepes, pH 7.4, and 400 mM sorbitol). Transport was initiated by addition of 10 mM Gdm+, and fluorescence was monitored (excitation, 460 nm; emission, 510 nm). At the end of the assay, the proton gradient was collapsed by addition of 1 µM FCCP (carbonyl cyanide p-trifluoromethoxyphenyl hydrazone). For assays monitoring the effect of membrane potential, liposomes were prepared with 3–100 mM KCl, Gdm-Cl (10 mM for electrogenicity experiments and 0.4 mM for stoichiometry experiments), 25 mM Hepes 7.5, and 1 mM pyranine. NaCl was added to maintain ionic strength at 200 mM. External buffer contained 25 mM Hepes, pH 7.5, and 3–100 mM KCl to establish a potassium gradient (as indicated in the text/figures), NaCl to maintain 200 mM ionic strength, 10 mM Gdm+ (for experiments to determine the electrogenicity) or 4 mM Gdm+ (for experiments to determine stoichiometry), and valinomycin (20 µg/mL) to set the membrane potential. Membrane potential was calculated from the K+ gradient according to the Nernst potential. The

stoichiometry was determined using the method described by Mindell and coworkers[29,43], from the experimentally determined reversal potential, Erev. Pyranine fluorescence was monitored as a proxy for proton transport as before.

References

(1) Yerushalmi, H.; Lebendiker, M.; Schuldiner, S. EmrE, an Escherichia Coli 12-kDa Multidrug Transporter, Exchanges Toxic Cations and H+ and Is Soluble in Organic Solvents. *Journal of Biological Chemistry* **1995**, *270* (12), 6856–6863. https://doi.org/10.1074/jbc.270.12.6856.

(2) Paulsen, I. T.; Skurray, R. A.; Tam, R.; Saler, M. H.; Turner, R. J.; Weiner, J. H.; Goldberg, E. B.; Grinius, L. L. The SMR Family: A Novel Family of Multidrug Efflux Proteins Involved with the Efflux of Lipophilic Drugs. *Molecular Microbiology* **1996**, *19* (6), 1167–1175. https://doi.org/10.1111/j.1365-2958.1996.tb02462.x.

(3) Rapp, M.; Granseth, E.; Seppala, S.; von Heijne, G. Identification and Evolution of Dual-Topology Membrane Proteins. *Nature Structural & Molecular Biology* **2006**, *13* (2), 112–116. https://doi.org/10.1038/nsmb1057.

(4) Morrison, E. A.; DeKoster, G. T.; Dutta, S.; Vafabakhsh, R.; Clarkson, M. W.; Bahl, A.; Kern, D.; Ha, T.; Henzler-Wildman, K. A. Antiparallel EmrE Exports Drugs by Exchanging Between Asymmetric Structures. *Nature* **2012**, *481* (7379), 45–U50. https://doi.org/10.1038/nature10703.

(5) Bay, D. C.; Rommens, K. L.; Turner, R. J. Small Multidrug Resistance Proteins: A Multidrug Transporter Family That Continues to Grow. *Biochimica Et Biophysica Acta-Biomembranes* **2008**, *1778* (9), 1814–1838. https://doi.org/10.1016/j.bbamem.2007.08.015.

(6) Muth, T. R.; Schuldiner, S. A Membrane-Embedded Glutamate Is Required for Ligand Binding to the Multidrug Transporter EmrE. *Embo Journal* **2000**, *19* (2), 234–240. https://doi.org/10.1093/emboj/19.2.234.

(7) Greener, T.; Govezensky, D.; Zamir, A. A Novel Multicopy Suppressor of a groEL Mutation Includes Two Nested Open Reading Frames Transcribed from Different Promoters. *The EMBO Journal* **1993**, *12* (3), 889–896. https://doi.org/10.1002/j.1460-2075.1993.tb05729.x.

(8) Bay, D. C.; Turner, R. J. Small Multidrug Resistance Efflux Pumps. In *Efflux-Mediated Antimicrobial Resistance in Bacteria: Mechanisms, Regulation and Clinical Implications*; Li, X.-Z., Elkins, C. A., Zgurskaya, H. I., Eds.; Springer International Publishing: Cham, 2016; pp 45–71. https://doi.org/10.1007/978-3-319-39658-3_3.

(9) Chung, Y. J.; Saier, M. H. Overexpression of the Escherichia Coli sugE Gene Confers Resistance to a Narrow Range of Quaternary Ammonium Compounds. *Journal of Bacteriology* **2002**, *184* (9), 2543–2545. https://doi.org/10.1128/JB.184.9.2543-2545.2002.

(10) Sikora, C. W.; Turner, R. J. SMR Proteins SugE and EmrE Bind Ligand with Similar Affinity and Stoichiometry. *Biochemical and Biophysical Research Communications* **2005**, *335* (1), 105–111. https://doi.org/10.1016/j.bbrc.2005.07.051.

(11) Son, M. S.; Del Castilho, C.; Duncalf, K. A.; Carney, D.; Weiner, J. H.; Turner, R. J. Mutagenesis of SugE, a Small Multidrug Resistance Protein. *Biochemical and Biophysical Research Communications* **2003**, *312* (4), 914–921. https://doi.org/10.1016/j.bbrc.2003.11.018.

(12) Nelson, J. W.; Atilho, R. M.; Sherlock, M. E.; Stockbridge, R. B.; Breaker, R. R. Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. *Molecular Cell* **2017**, *65* (2), 220–230. https://doi.org/10.1016/j.molcel.2016.11.019.

(13) Sherlock, M. E.; Breaker, R. R. Biochemical Validation of a Third Guanidine Riboswitch Class in Bacteria. *Biochemistry* **2017**, *56* (2), 359–363. https://doi.org/10.1021/acs.biochem.6b01271.

(14) Sherlock, M. E.; Malkowski, S. N.; Breaker, R. R. Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria. *Biochemistry* **2017**, *56* (2), 352–358. https://doi.org/10.1021/acs.biochem.6b01270.

(15) Breaker, R. R.; Atilho, R. M.; Malkowski, S. N.; Nelson, J. W.; Sherlock, M. E. The Biology of Free Guanidine As Revealed by Riboswitches. *Biochemistry* **2017**, *56* (2), 345–347. https://doi.org/10.1021/acs.biochem.6b01269.

(16) Reiss, C. W.; Strobel, S. A. Structural Basis for Ligand Binding to the Guanidine-II Riboswitch. *Rna* **2017**, *23* (9), 1338–1343. https://doi.org/10.1261/rna.061804.117.

(17) Battaglia, R. A.; Price, I. R.; Ke, A. Structural Basis for Guanidine Sensing by the ykkC Family of Riboswitches. *Rna* **2017**, *23* (4), 578–585. https://doi.org/10.1261/rna.060186.116.

(18) Reiss, C. W.; Xiong, Y.; Strobel, S. A. Structural Basis for Ligand Binding to the Guanidine-I Riboswitch. *Structure* **2017**, *25* (1), 195–202. https://doi.org/10.1016/j.str.2016.11.020.

(19) Huang, L.; Wang, J.; Lilley, D. M. J. The Structure of the Guanidine-II Riboswitch. *Cell Chemical Biology* **2017**, *24* (6), 695–+. https://doi.org/10.1016/j.chembiol.2017.05.014.

(20) Trujillo, C.; Rodriguez-Sanz, A. A.; Rozas, I. Aromatic Amino Acids-Guanidinium Complexes Through Cation-Pi Interactions. *Molecules* **2015**, *20* (5), 9214–9228. https://doi.org/10.3390/molecules20059214.

(21) Gund, P. Guanidine, Trimethylenemethane, and "Y-Delocalization." Can Acyclic Compounds Have "Aromatic" Stability? *Journal of Chemical Education* **1972**, *49* (2), 100–103. https://doi.org/10.1021/ed049p100.

(22) von Heijne, G. Control of Topology and Mode of Assembly of a Polytopic Membrane Protein by Positively Charged Residues. *Nature* **1989**, *341* (6241), 456–458. https://doi.org/10.1038/341456a0.

(23) Lolkema, J. S.; Dobrowolski, A.; Slotboom, D.-J. Evolution of Antiparallel Two-Domain Membrane Proteins: Tracing Multiple Gene Duplication Events in the Duf606 Family. *Journal of Molecular Biology* **2008**, *378* (3), 596–606. https://doi.org/10.1016/j.jmb.2008.03.005.

(24) Macdonald, C. B.; Stockbridge, R. B. A Topologically Diverse Family of Fluoride Channels. *Current Opinion in Structural Biology* **2017**, *45*, 142–149. https://doi.org/10.1016/j.sbi.2017.04.003.

(25) Rotem, D.; Schuldiner, S. EmrE, a Multidrug Transporter from Escherichia Coli, Transports Monovalent and Divalent Substrates with the Same Stoichiometry. *Journal of Biological Chemistry* **2004**, *279* (47), 48787–48793. https://doi.org/10.1074/jbc.M408187200.

(26) Morrison, E. A.; Henzler-Wildman, K. A. Transported Substrate Determines Exchange Rate in the Multidrug Resistance Transporter EmrE. *Journal of Biological Chemistry* **2014**, *289* (10), 6825–6836. https://doi.org/10.1074/jbc.M113.535328.

(27) Robinson, A. E.; Thomas, N. E.; Morrison, E. A.; Balthazor, B. M.; Henzler-Wildman, K. A. New Free-Exchange Model of EmrE Transport. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (47), E10083–E10091. https://doi.org/10.1073/pnas.1708671114.

(28) Cho, M.-K.; Gayen, A.; Banigan, J. R.; Leninger, M.; Traaseth, N. J. Intrinsic Conformational Plasticity of Native EmrE Provides a Pathway for Multidrug Resistance. *Journal of the American Chemical Society* **2014**, *136* (22), 8072–8080. https://doi.org/10.1021/ja503145x.

(29) Fitzgerald, G. A.; Mulligan, C.; Mindell, J. A. A General Method for Determining Secondary Active Transporter Substrate Stoichiometry. *Elife* **2017**, *6*, e21016. https://doi.org/10.7554/eLife.21016.001.

(30) Nasie, I.; Steiner-Mordoch, S.; Gold, A.; Schuldiner, S. Topologically Random Insertion of EmrE Supports a Pathway for Evolution of Inverted Repeats in Ion-Coupled Transporters. *Journal of Biological Chemistry* **2010**, *285* (20), 15234–15244. https://doi.org/10.1074/jbc.M110.108746.

(31) Stockbridge, R. B.; Robertson, J. L.; Kolmakova-Partensky, L.; Miller, C. A Family of Fluoride-Specific Ion Channels with Dual-Topology Architecture. *Elife* **2013**, *2*, e01084. https://doi.org/10.7554/eLife.01084.

(32) Vaughan, V. C.; Novy, F. G. *Ptomaïnes and Leucomaïnes, and Bacterial Proteids: Or the Chemical Factors in the Causation of Disease*; Lea brothers & Company, 1891.

(33) Baker, J. L.; Sudarsan, N.; Weinberg, Z.; Roth, A.; Stockbridge, R. B.; Breaker, R. R. Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride. *Science* **2012**, *335* (6065), 233–235. https://doi.org/10.1126/science.1215063.

(34) Bewley, M. C.; Jeffrey, P. D.; Patchett, M. L.; Kanyo, Z. F.; Baker, E. N. Crystal Structures of Bacillus Caldovelox Arginase in Complex with Substrate and Inhibitors Reveal New Insights into Activation, Inhibition and Catalysis in the Arginase Superfamily. *Structure with Folding & Design* **1999**, *7* (4), 435–448. https://doi.org/10.1016/S0969-2126(99)80056-2.

(35) Kalia, J.; Swartz, K. J. Elucidating the Molecular Basis of Action of a Classic Drug: Guanidine Compounds As Inhibitors of Voltage-Gated Potassium Channels. *Molecular Pharmacology* **2011**, *80* (6), 1085–1095. https://doi.org/10.1124/mol.111.074989.

(36) Zeymer, C.; Werbeck, N. D.; Schlichting, I.; Reinstein, J. The Molecular Mechanism of Hsp100 Chaperone Inhibition by the Prion Curing Agent Guanidinium Chloride. *Journal of Biological Chemistry* **2013**, *288* (10), 7065–7076. https://doi.org/10.1074/jbc.M112.432583.

(37) Grimminger, V.; Richter, K.; Imhof, A.; Buchner, J.; Walter, S. The Prion Curing Agent Guanidinium Chloride Specifically Inhibits ATP Hydrolysis by Hsp104. *Journal of Biological Chemistry* **2004**, *279* (9), 7378–7383. https://doi.org/10.1074/jbc.M312403200.

(38) Katoh, K.; Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **2013**, *30* (4), 772–780. https://doi.org/10.1093/molbev/mst010.

(39) Kalyaanamoorthy, S.; Minh, B. Q.; Wong, T. K. F.; von Haeseler, A.; Jermiin, L. S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nature Methods* **2017**, *14* (6), 587–+. https://doi.org/10.1038/NMETH.4285.

(40) Nguyen, L.-T.; Schmidt, H. A.; von Haeseler, A.; Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **2015**, *32* (1), 268–274. https://doi.org/10.1093/molbev/msu300.

(41) Minh, B. Q.; Nguyen, M. A. T.; von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution* **2013**, *30* (5), 1188–1195. https://doi.org/10.1093/molbev/mst024.

(42) Ma, C.; Chang, G. Structure of the Multidrug Resistance Efflux Transporter EmrE from Escherichia Coli (Retraction of Vol 101, Pg 2852, 2004). *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104* (9), 3668–3668. https://doi.org/10.1073/pnas.0700711104.

(43) Parker, J. L.; Mindell, J. A.; Newstead, S. Thermodynamic Evidence for a Dual Transport Mechanism in a POT Peptide Transporter. *Elife* **2014**, *3*, e04273. https://doi.org/10.7554/eLife.04273.

CHAPTER 4

The Structural Basis of Promiscuity in Small Multidrug Resistance Transporters

This chapter is adapted from the following published 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).

† - 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.

4.1 Abstract

By providing broad resistance to environmental biocides, transporters from the small multidrug resistance (SMR) family drive the spread of multidrug resistance cassettes among bacterial populations. A fundamental understanding of substrate selectivity by SMR transporters is needed to identify the types of selective pressures that contribute to this process. Using solid-supported membrane electrophysiology, we find that promiscuous transport

of hydrophobic substituted cations is a general feature of SMR transporters. To understand the molecular basis for promiscuity, we solved X-ray crystal structures of a SMR transporter Gdx-Clo in complex with substrates to a maximum resolution of 2.3 Å. These structures confirm the family's extremely rare dual topology architecture and reveal a cleft between two helices that provides accommodation in the membrane for the hydrophobic substituents of transported drug-like cations.

4.2 Introduction

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[1,2]. 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 determined[5], 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 (figure 4.1 a and supplementary figure B.1). The SMR family has two

major subtypes that share high sequence identity (~40%) and similarity (supplementary figure B.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 guandinium (Gdm)/H antiporters (Gdx proteins; E. coli gene name sugE)[6]. 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[7,8] (4.1 a).

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[6]. 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[9]. But genes from the Gdx cluster also commonly colocalize with horizontal gene transfer elements (4.1 a)[10,11], and have been explicitly identified in mobile multidrug resistance gene arrays[12,13] (4.1). 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 guanidinyl 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 guanidinyl substrates reveal a cleft between two helices that provides accommodation in the membrane for the

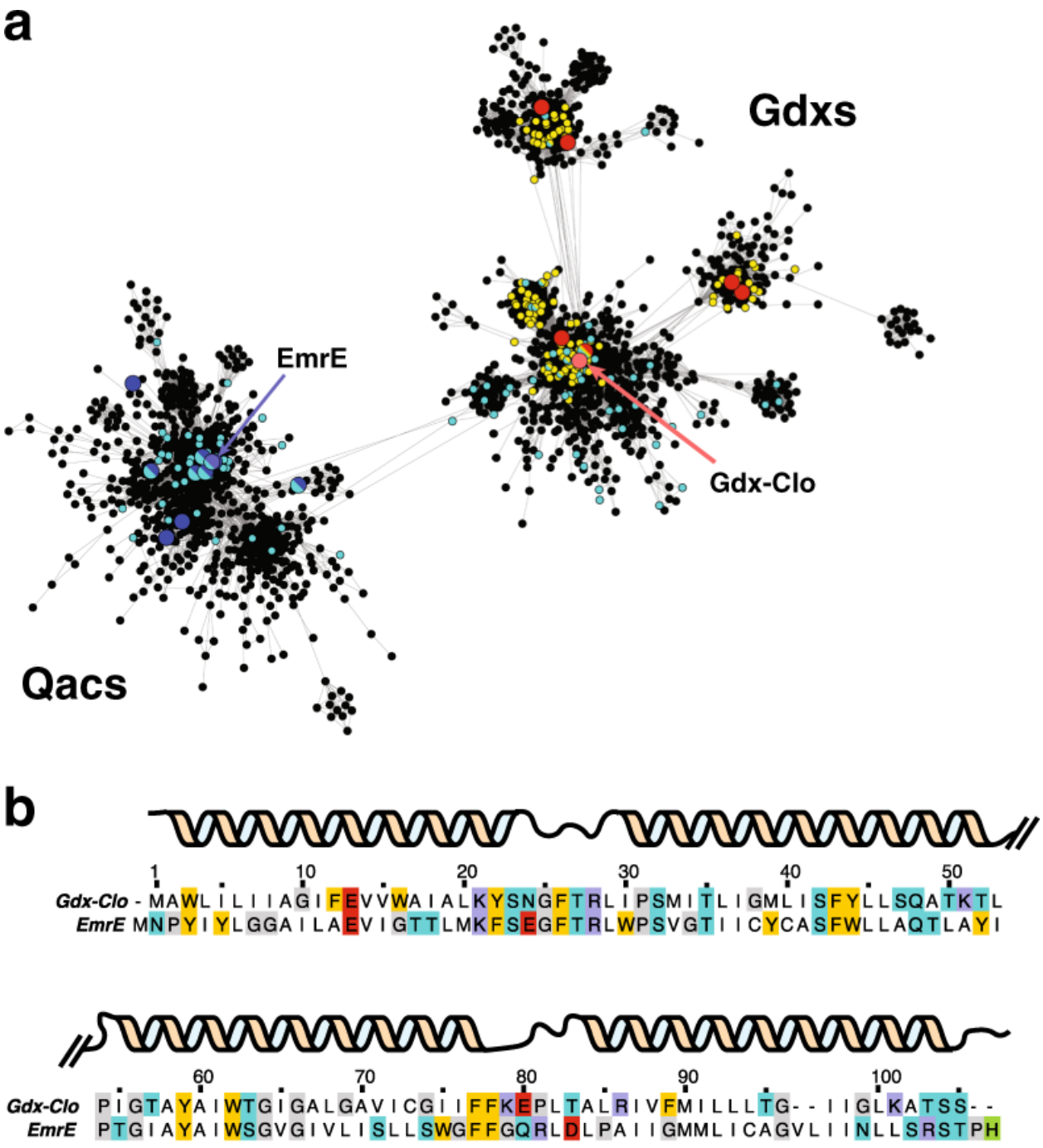

Figure 4.1: Colocalization of SMR genes with guanidine riboswitches and horizontal gene transfer elements. a Major clusters (>10% of total set) from the sequence-similarity network of the SMR family. Each point corresponds to a cluster of sequences with >50% sequence identity. Edges between points correspond 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 B.1 b. Sequence alignment of EmrE and Gdx-Clo with positions of α-helices indicated.

hydrophobic substituents of transported drug-like cations.

4.3 Results

4.3.1 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 (4.1 b): the polyaromatic cation exporter EmrE, and Gdx-Clo, a functionally characterized Gdm+ transporter from Clostridales oral taxon 876[6,14]. 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 guanidinyl metabolite, agmatine (figure 4.2 a).

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+ ion6, and EmrE, though it has been shown to stray slightly from strict 2:1 stoichiometry, transports monovalent substrates in an electrogenic manner as well[15,16]. 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 (figure 4.2 b). 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,

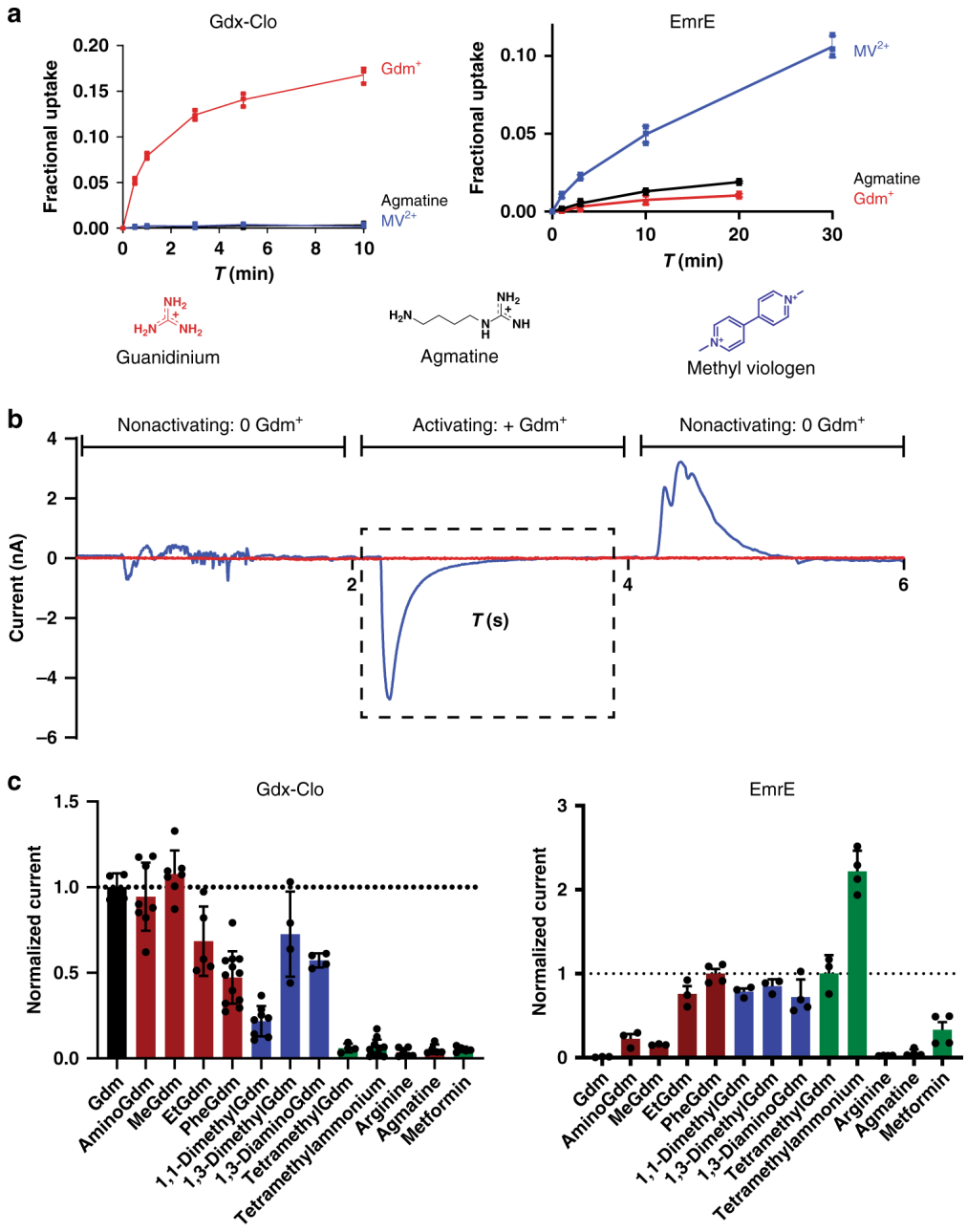

Figure 4.2: Substrate transport by Gdx-Clo and EmrE. a Radioactive exchange assays. For Gdx-Clo and EmrE, uptake of 14C-Gdm+ or 14C-methyl viologen, respectively, monitored in exchange for the indicated substrate. Points represent individual replicates; error bars represent the mean ± SEM from three 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 guanidinyl compounds are shown as maroon bars, doubly substituted guanidinyl 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 ±SEM.

and a return to the starting condition (figure 4.2 b).

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 figure B.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 figure B.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 figure B.5 or ref. 6). We observed no discrepancies between the electrophysiological results and the radioactive uptake experiments.

Our electrophysiology experiments (figure 4.2 c) recapitulate prior observations for metabolites: EmrE does not transport Gdm+, and both proteins are strongly selective against substituted guanidinium metabolites like arginine, agmatine, and creatine[6]. However, many of the non-natural compounds we tested were readily transported by both subtypes. Gdx-Clo transported guanidinyl compounds with hydrophobic single substitutions, including the bulky phenylGdm+. Currents decreased for doubly substituted guanidinyl compounds, and were absent for tetramethylGdm+. 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 phenylGdm+ evoked increasingly larger currents. In contrast to Gdx-Clo, EmrE also accommodated substrates with reduced or no H-bonding capacity, tetramethylGdm+ 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 figure B.6.

4.3.2 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 EmrE5, and an X-ray crystal structure of EmrE[17] that has notable deficiencies: it is presented as a Cα 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[19].

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[20]. 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 figure B.13 and supplementary figure B.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 (figure 4.3 a and supplementary figure B.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 figure B.9).

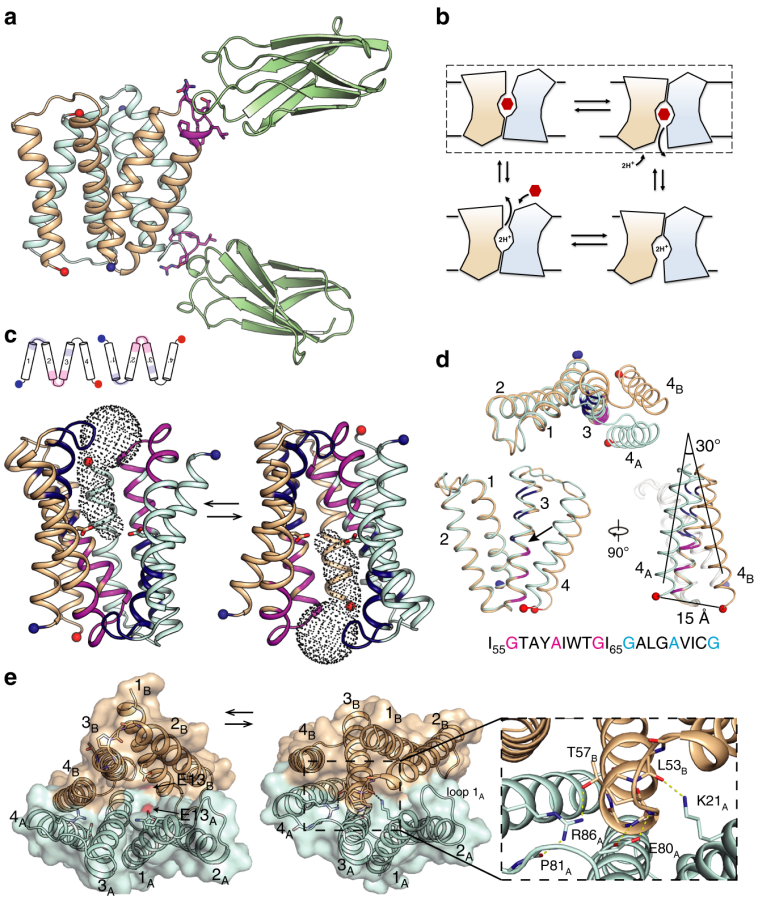

Figure 4.3: Gdx-Clo structure and conformational exchange. (a) Structure of Gdx-Clo/monobody complex. Clo-L10 monobodies are shown in green. Transporter shown with subunit A in light blue and subunit B in tan. The N- and C-termini for each subunit are shown as blue and red spheres, respectively. Transporter 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 Cα 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 GI65G 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.

4.3.3 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 EmrE5, 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, E13A and E13B, accessible at the bottom. Positive density is visible between the E13 sidechains, but cannot be definitively assigned as Gdm+ at this resolution (B.10). Transport by the anti-parallel 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 (figure 4.3 b, c). As a consequence 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 (figure 4.3 c).

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° (figure 4.3 d). 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[21]. 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 ex-

change (figure 4.3 c and B.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[22].

The well-ordered extramembrane loops also exhibit major differences in packing on the open and closed sides of the transporter (figure 4.3 e). On the open side of the transporter, several charged amino acids, K21A from loop 1A and E80A and R86A from loop 3A, are solvent-exposed in the aqueous chamber. Upon conformational exchange, K21A, E80A, and R86A, converge on loop 2B and the N-terminal end of helix 3B, forming cross-subunit H-bond interactions with the backbone and sidechain atoms of L53B-T57B. The hydrophobic loop 1A also contributes to sealing the binding pocket on the closed side of the transporter, where it is wedged between the antiparallel helices 2B and 2A. 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[23,24].

4.3.4 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 (figure 4.4 a). Neutralization of these glutamates has previously been shown to abolish substrate transport in Gdx-Clo[6]. The substrate's guanidinyl group is coordinated by E13B, whose position is in turn stabilized

by a stack of conserved H-bond donors and acceptors, including W62B, S42B, and W16B. W62 and S42 are highly conserved among SMRs, and have been previously implicated in substrate specificity and transport[25,26]. In Gdx-Clo, mutations that remove H-bond potential, S42A and W62F, reduced or eliminated Gdm+ exchange, respectively (figure 4.4 b). 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 (figure 4.4 b).

The guanidinyl group of phenylGdm+ is also in close proximity to the opposite E13A sidechain. However, E13A is deflected downward by a cross-subunit interaction with Y59B, so that the angle between the nitrogen, hydrogen (coplanar with the guanidinyl group), and oxygen atoms is not optimal for H-bond formation. Y59 is absolutely conserved among SMR proteins and the capacity to hydrogen bond has been identified as mechanistically essential at this position[18,27]. Based on our Gdx-Clo structure, we propose that Y59B and the guanidinium group compete for E13A, and that displacement of Y59B by the guanidinyl group initiates the transport motion (figure 4.4 c). 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 (figure 4.4 b), 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[28], and also draws a contrast to Gdm+ coordination by the guanidine riboswitches, which provide hydrogen bond partners for most or all of the substrate's five hydrogen bond donors[29–32]. The more minimal coordination by the transporter explains its permissiveness towards guanidinium ions with methyl substitutions in one or two positions.

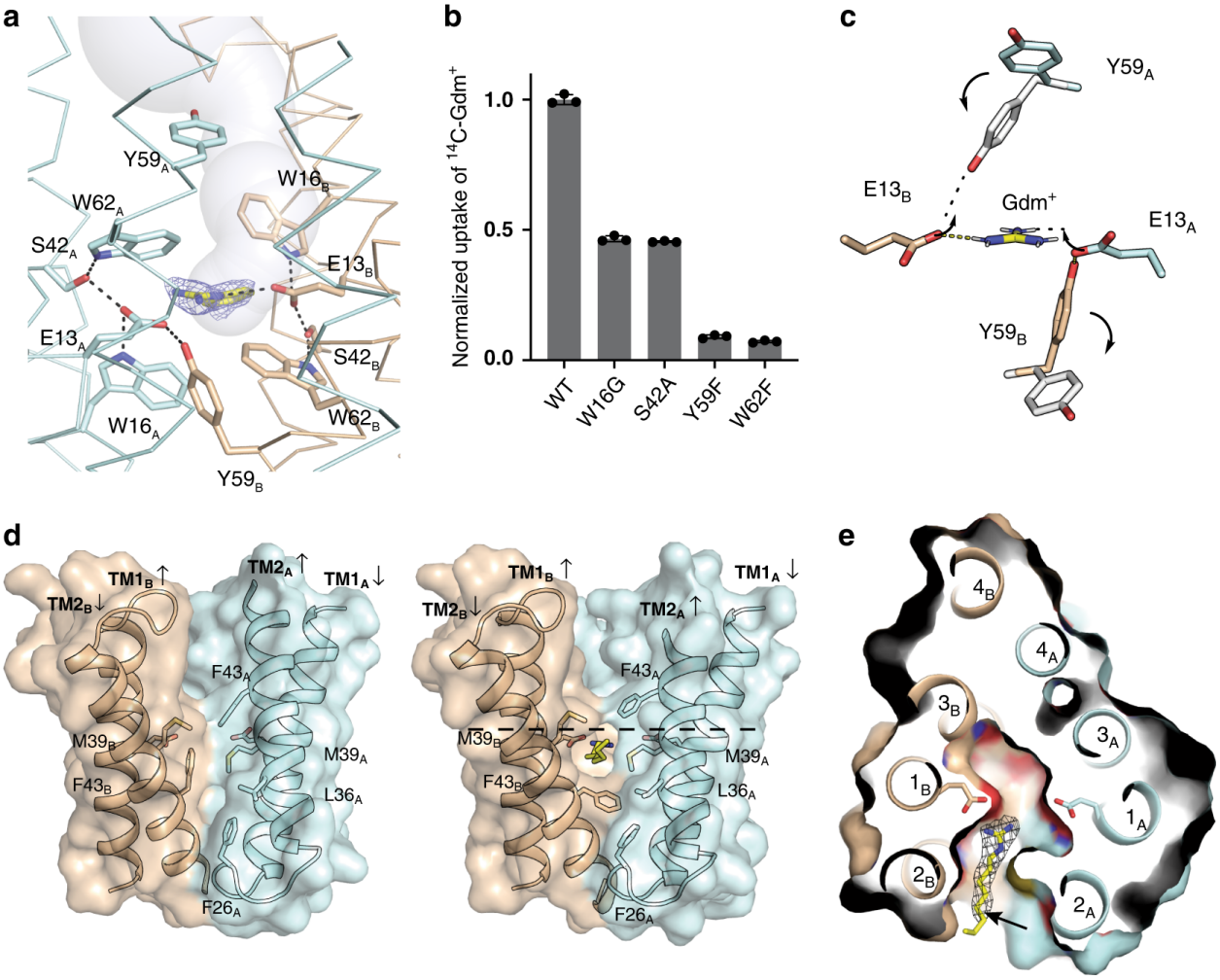

Figure 4.4: Substrate binding by Gdx-Clo. a PhenylGdm+ binding site. Subunits colored in light blue and tan as in figure 4.3. The aqueous accessible vestibule is shown as a gray surface rendering. Sidechains that coordinate substrate or E13 shown as sticks, and interactions with appropriate distance 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 figure 4.3 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.

4.3.5 A membrane portal accommodates hydrophobic substrate substituents

In the case of phenylGdm+, the substituent is packed between TM2A and TM2B. At this point, the antiparallel TM2 helices splay apart, delimiting a portal from the membrane to the substrate binding site (figure 4.4 d). In order 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 guanidinyl group sits in the same binding pocket as phenylGdm+, near E13B, and the aliphatic tail protrudes from the protein and into the detergent micelle (figure 4.4 d, e). The tail is accommodated by rotameric rearrangements of the hydrophobic amino acids lining TM2 including M39 and F43 (figure 4.4 e). 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[33–35]. Spectroscopic studies and molecular modeling have provided evidence for a similar portal between the TM2 helices of EmrE[18,23].

It is clear that 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 guanidinylated metabolites, and rationalizing the conservation of this feature. The SLC35 solute transporters[36–38] 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 figure B.12).

4.4 Discussion

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[39]—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 use9, 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[40–42] or pharmaceuticals that impact the human microbiome and that are excreted in waste water[43–46].

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[33,47,48]. 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.

4.5 Methods

4.5.1 Sequence-similarity network

A sequence-similarity network was generated using the EFI-EST webserver[49] 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[50]. 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).

4.5.2 Transporter expression, purification, and proteoliposome reconstitution

Lipids were from Avanti, detergents from Anatrace. Proteins were expressed and purified as previously described[6]. 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. 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 pre-equilibrated 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 14C-labeled compound (20 µM 14C-Gdm+ for Gdx or 7 µM 14C-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.

4.5.3 SSM electrophysiology

SSM electrophysiology was conducted using a SURFE2R N1 instrument (Nanion Technologies, Munich, Germany) according to published protocols[51]. 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 µg/µl 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 × 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[51].

4.5.4 Monobody development

Monobody selection was performed following previously published methods[20,52,53]. 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[20]. 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 µM 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[20,54].

4.5.5 Monobody expression and purification

Monobody proteins were expressed in E. coli (BL21-DE3) grown in Studier's autoinduction media[55] 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 µg DNase, 2 mM MgCl2, 1 mM PMSF, 1 mg/ml lysozyme, 25 µg pepstatin, and 500 µg leupeptin and lysed by sonication prior to centrifugation (27,000 × g for 15 min). Inclusion bodies were isolated by addition of Triton X-100 to a final concentration of 1% w/v[56], incubation of the lysate on ice, and centrifugation (27,000 × 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 × g/45 min), and the supernatant was loaded onto a cobalt affinity column (Takara; 3 ml resin/l culture) for on-column refolding[57]. 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.

4.5.6 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 µl 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 µl protein/detergent mixture together with 0.3 µl 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.

4.5.7 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[58]. 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[59]. The positions were refined and initial phases were calculated using SHARP[60] with solvent flattening with SOLOMON[61]. A model for the transporter was built into experimental electron density maps using Coot[62]. 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])[63]. Variable loops were not included in the monobody model. These models were placed into the experimental electron density maps using Phaser-MR[64]. 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[65] with prior phase information incorporated as Hendrickson–Lattman coefficients. Model validation was carried out using the Molprobity server[66].

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 arrange-

ment 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[67] with a cutoff level of 1.8 for the local $I/\sigma\langle I\rangle$. Models were built into experimental density maps calculated from Phaser, with the initial models of Gdx-Clo 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[68].

4.6 Acknowledgements

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.

4.7 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.

References

(1) Zhu, Y.-G.; Zhao, Y.; Li, B.; Huang, C.-L.; Zhang, S.-Y.; Yu, S.; Chen, Y.-S.; Zhang, T.; Gillings, M. R.; Su, J.-Q. Continental-Scale Pollution of Estuaries with Antibiotic Resistance Genes. *Nature Microbiology* **2017**, *2* (4), 16270. https://doi.org/10.1038/nmicrobiol.2016.270.

(2) Gillings, M. R. Class 1 Integrons as Invasive Species. *Current Opinion in Microbiology* **2017**, *38*, 10–15. https://doi.org/10.1016/j.mib.2017.03.002.

(3) Pal, C.; Bengtsson-Palme, J.; Kristiansson, E.; Larsson, D. G. J. Co-Occurrence of Resistance Genes to Antibiotics, Biocides and Metals Reveals Novel Insights into Their Co-Selection Potential. *Bmc Genomics* **2015**, *16*, 964. https://doi.org/10.1186/s12864-015-2153-5.

(4) Wales, A. D.; Davies, R. H. Co-Selection of Resistance to Antibiotics, Biocides and Heavy Metals, and Its Relevance to Foodborne Pathogens. *Antibiotics-Basel* **2015**, *4* (4), 567–604. https://doi.org/10.3390/antibiotics4040567.

(5) Ubarretxena-Belandia, I.; Baldwin, J. M.; Schuldiner, S.; Tate, C. G. Three-Dimensional Structure of the Bacterial Multidrug Transporter EmrE Shows It Is an Asymmetric Homodimer. *Embo Journal* **2003**, *22* (23), 6175–6181. https://doi.org/10.1093/emboj/cdg611.

(6) Kermani, A. A.; Macdonald, C. B.; Gundepudi, R.; Stockbridge, R. B. Guanidinium Export Is the Primal Function of SMR Family Transporters. *Proc Natl Acad Sci USA* **2018**, *115* (12), 3060–3065. https://doi.org/10.1073/pnas.1719187115.

(7) Breaker, R. R.; Atilho, R. M.; Malkowski, S. N.; Nelson, J. W.; Sherlock, M. E. The Biology of Free Guanidine As Revealed by Riboswitches. *Biochemistry* **2017**, *56* (2), 345–347. https://doi.org/10.1021/acs.biochem.6b01269.

(8) Battaglia, R. A.; Ke, A. Guanidine-Sensing Riboswitches: How Do They Work and What Do They Regulate? *Wiley Interdisciplinary Reviews-Rna* **2018**, *9* (5), e1482. https://doi.org/10.1002/wrna.1482.

(9) Russell, A. D. Introduction of Biocides into Clinical Practice and the Impact on Antibiotic-Resistant Bacteria. *Journal of Applied Microbiology* **2002**, *92*, 121S–135S. https://doi.org/10.1046/j.1365-2672.92.5s1.12.x.

(10) Jeong, H.; Nasir, A. A Preliminary List of Horizontally Transferred Genes in Prokaryotes Determined by Tree Reconstruction and Reconciliation. *Frontiers in Genetics* **2017**, *8*, 112. https://doi.org/10.3389/fgene.2017.00112.

(11) Slipski, C. J.; Jamieson, T. R.; Lam, A.; Shing, V. L.; Bell, K.; Zhanel, G. G.; Bay, D. C. 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* **2019**, 768630. https://doi.org/10.1101/768630.

(12) Partridge, S. R.; Tsafnat, G.; Coiera, E.; Iredell, J. R. Gene Cassettes and Cassette Arrays in Mobile Resistance Integrons. *Fems Microbiology Reviews* **2009**, *33* (4), 757–784. https://doi.org/10.1111/j.1574-6976.2009.00175.x.

(13) Moura, A.; Soares, M.; Pereira, C.; Leitao, N.; Henriques, I.; Correia, A. INTEGRALL: A Database and Search Engine for Integrons, Integrases and Gene Cassettes. *Bioinformatics* **2009**, *25* (8), 1096–1098. https://doi.org/10.1093/bioinformatics/btp105.

(14) Nelson, J. W.; Atilho, R. M.; Sherlock, M. E.; Stockbridge, R. B.; Breaker, R. R. Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. *Molecular Cell* **2017**, *65* (2), 220–230. https://doi.org/10.1016/j.molcel.2016.11.019.

(15) Robinson, A. E.; Thomas, N. E.; Morrison, E. A.; Balthazor, B. M.; Henzler-Wildman, K. A. New Free-Exchange Model of EmrE Transport. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (47), E10083–E10091. https://doi.org/10.1073/pnas.1708671114.

(16) Rotem, D.; Schuldiner, S. EmrE, a Multidrug Transporter from Escherichia Coli, Transports Monovalent and Divalent Substrates with the Same Stoichiometry. *Journal of Biological Chemistry* **2004**, *279* (47), 48787–48793. https://doi.org/10.1074/jbc.M408187200.

(17) Chen, Y.-J.; Pornillos, O.; Lieu, S.; Ma, C.; Chen, A. P.; Chang, G. X-Ray Structure of EmrE Supports Dual Topology Model. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104* (48), 18999–19004. https://doi.org/10.1073/pnas.0709387104.

(18) Vermaas, J. V.; Rempe, S. B.; Tajkhorshid, E. Electrostatic Lock in the Transport Cycle of the Multidrug Resistance Transporter EmrE. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115* (32), E7502–E7511. https://doi.org/10.1073/pnas.1722399115.

(19) Ovchinnikov, V.; Stone, T. A.; Deber, C. M.; Karplus, M. Structure of the EmrE Multidrug Transporter and Its Use for Inhibitor Peptide Design. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115* (34), E7932–E7941. https://doi.org/10.1073/pnas.1802177115.

(20) Koide, A.; Wojcik, J.; Gilbreth, R. N.; Hoey, R. J.; Koide, S. Teaching an Old Scaffold New Tricks: Monobodies Constructed Using Alternative Surfaces of the Fn3 Scaffold. *Journal of Molecular Biology* **2012**, *415* (2), 393–405. https://doi.org/10.1016/j.jmb.2011.12.019.

(21) Amadi, S. T.; Koteiche, H. A.; Mishra, S.; Mchaourab, H. S. Structure, Dynamics, and Substrate-Induced Conformational Changes of the Multidrug Transporter EmrE in Liposomes. *Journal of Biological Chemistry* **2010**, *285* (34), 26710–26718. https://doi.org/10.1074/jbc.M110.132621.

(22) Wu, C.; Wynne, S. A.; Thomas, N. E.; Uhlemann, E.-M.; Tate, C. G.; Henzier-Wildman, K. A. Identification of an Alternating-Access Dynamics Mutant of EmrE with Impaired Transport. *Journal of Molecular Biology* **2019**, *431* (15), 2777–2789. https://doi.org/10.1016/j.jmb.2019.05.035.

(23) Dastvan, R.; Fischer, A. W.; Mishra, S.; Meiler, J.; Mchaourab, H. S. Protonation-Dependent Conformational Dynamics of the Multidrug Transporter EmrE. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113* (5), 1220–1225. https://doi.org/10.1073/pnas.1520431113.

(24) Leninger, M.; Her, A. S.; Traaseth, N. J. Inducing Conformational Preference of the Membrane Protein Transporter EmrE Through Conservative Mutations. *Elife* **2019**, *8*, e48909. https://doi.org/10.7554/eLife.48909.

(25) Elbaz, Y.; Tayer, N.; Steinfels, E.; Steiner-Mordoch, S.; Schuldiner, S. Substrate-Induced Tryptophan Fluorescence Changes in EmrE, the Smallest Ion-Coupled Multidrug Transporter. *Biochemistry* **2005**, *44* (19), 7369–7377. https://doi.org/10.1021/bi050356t.

(26) Brill, S.; Sade-Falk, O.; Elbaz-Alon, Y.; Schuldiner, S. Specificity Determinants in Small Multidrug Transporters. *Journal of Molecular Biology* **2015**, *427* (2), 468–477. https://doi.org/10.1016/j.jmb.2014.11.015.

(27) Rotem, D.; Steiner-Mordoch, S.; Schuldiner, S. Identification of Tyrosine Residues Critical for the Function of an Ion-Coupled Multidrug Transporter. *Journal of Biological Chemistry* **2006**, *281* (27), 18715–18722. https://doi.org/10.1074/jbc.M602088200.

(28) Peng, B.; Peng, Q.; Zhou, W.; Zhou, Z. Guanidinium L-Glutamate. *Acta Crystallographica Section E-Crystallographic Communications* **2010**, *66* (10), O2679–U1523. https://doi.org/10.1107/S1600536810036354.

(29) Reiss, C. W.; Strobel, S. A. Structural Basis for Ligand Binding to the Guanidine-II Riboswitch. *Rna* **2017**, *23* (9), 1338–1343. https://doi.org/10.1261/rna.061804.117.

(30) Reiss, C. W.; Xiong, Y.; Strobel, S. A. Structural Basis for Ligand Binding to the Guanidine-I Riboswitch. *Structure* **2017**, *25* (1), 195–202. https://doi.org/10.1016/j.str.2016.11.020.

(31) Battaglia, R. A.; Price, I. R.; Ke, A. Structural Basis for Guanidine Sensing by the ykkC Family of Riboswitches. *Rna* **2017**, *23* (4), 578–585. https://doi.org/10.1261/rna.060186.116.

(32) Huang, L.; Wang, J.; Lilley, D. M. J. The Structure of the Guanidine-II Riboswitch. *Cell Chemical Biology* **2017**, *24* (6), 695–+. https://doi.org/10.1016/j.chembiol.2017.05.014.

(33) Aller, S. G.; Yu, J.; Ward, A.; Weng, Y.; Chittaboina, S.; Zhuo, R.; Harrell, P. M.; Trinh, Y. T.; Zhang, Q.; Urbatsch, I. L.; Chang, G. Structure of P-Glycoprotein Reveals a Molecular Basis for Poly-Specific Drug Binding. *Science* **2009**, *323* (5922), 1718–1722. https://doi.org/10.1126/science.1168750.

(34) He, X.; Szewczyk, P.; Karyakin, A.; Evin, M.; Hong, W.-X.; Zhang, Q.; Chang, G. Structure of a Cation-Bound Multidrug and Toxic Compound Extrusion Transporter. *Nature* **2010**, *467* (7318), 991–U139. https://doi.org/10.1038/nature09408.

(35) Payandeh, J.; Scheuer, T.; Zheng, N.; Catterall, W. A. The Crystal Structure of a Voltage-Gated Sodium Channel. *Nature* **2011**, *475* (7356), 353–U104. https://doi.org/10.1038/nature10238.

(36) Parker, J. L.; Newstead, S. Structural Basis of Nucleotide Sugar Transport Across the Golgi Membrane. *Nature* **2017**, *551* (7681), 521–+. https://doi.org/10.1038/nature24464.

(37) Nji, E.; Gulati, A.; Qureshi, A. A.; Coincon, M.; Drew, D. Structural Basis for the Delivery of Activated Sialic Acid into Golgi for Sialyation. *Nature Structural & Molecular Biology* **2019**, *26* (6), 415–+. https://doi.org/10.1038/s41594-019-0225-y.

(38) Tsuchiya, H.; Doki, S.; Takemoto, M.; Ikuta, T.; Higuchi, T.; Fukui, K.; Usuda, Y.; Tabuchi, E.; Nagatoishi, S.; Tsumoto, K.; Nishizawa, T.; Ito, K.; Dohmae, N.; Ishitani, R.; Nureki, O. Structural Basis for Amino Acid Export by DMT Superfamily Transporter YddG. *Nature* **2016**, *534* (7607), 417–+. https://doi.org/10.1038/nature17991.

(39) Copley, S. D. An Evolutionary Biochemist's Perspective on Promiscuity. *Trends in Biochemical Sciences* **2015**, *40* (2), 72–78. https://doi.org/10.1016/j.tibs.2014.12.004.

(40) Organophosphorus Insecticides. In *Metabolic Pathways of Agrochemicals: Part 2: Insecticides and Fungicides*; Roberts, T. R., Hutson, D. H., Eds.; The Royal Society of Chemistry, 1999; pp 187–522. https://doi.org/10.1039/9781847551375-00187.

(41) Zahn, D.; Mucha, P.; Zilles, V.; Touffet, A.; Gallard, H.; Knepper, T. P.; Froemel, T. Identification of Potentially Mobile and Persistent Transformation Products of REACH-Registered Chemicals and Their Occurrence in Surface Waters. *Water Research* **2019**, *150*, 86–96. https://doi.org/10.1016/j.watres.2018.11.042.

(42) Peter, K. T.; Tian, Z.; Wu, C.; Lin, P.; White, S.; Du, B.; McIntyre, J. K.; Scholz, N. L.; Kolodziej, E. P. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. *Environmental Science & Technology* **2018**, *52* (18), 10317–10327. https://doi.org/10.1021/acs.est.8b03287.

(43) Blair, B. D.; Crago, J. P.; Hedman, C. J.; Klaper, R. D. Pharmaceuticals and Personal Care Products Found in the Great Lakes Above Concentrations of Environmental Concern. *Chemosphere* **2013**, *93* (9), 2116–2123. https://doi.org/10.1016/j.chemosphere.2013.07.057.

(44) Scheurer, M.; Michel, A.; Brauch, H.-J.; Ruck, W.; Sacher, F. Occurrence and Fate of the Antidiabetic Drug Metformin and Its Metabolite Guanylurea in the Environment and During Drinking Water Treatment. *Water Research* **2012**, *46* (15), 4790–4802. https://doi.org/10.1016/j.watres.2012.06.019.

(45) Maier, L.; Pruteanu, M.; Kuhn, M.; Zeller, G.; Telzerow, A.; Anderson, E. E.; Brochado, A. R.; Fernandez, K. C.; Dose, H.; Mori, H.; Patil, K. R.; Bork, P.; Typas, A. Extensive Impact of Non-Antibiotic Drugs on Human Gut Bacteria. *Nature* **2018**, *555* (7698), 623–+. https://doi.org/10.1038/nature25979.

(46) Pryor, R.; Norvaisas, P.; Marinos, G.; Best, L.; Thingholm, L. B.; Quintaneiro, L. M.; De Haes, W.; Esser, D.; Waschina, S.; Lujan, C.; Smith, R. L.; Scott, T. A.; Martinez-Martinez, D.; Woodward, O.; Bryson, K.; Laudes, M.; Lieb, W.; Houtkooper, R. H.; Franke, A.; Temmerman, L.; Bjedov, I.; Cocheme, H. M.; Kaleta, C.; Cabreiro, F. Host-Microbe-Drug-Nutrient Screen Identifies Bacterial Effectors of Metformin Therapy. *Cell* **2019**, *178* (6), 1299–+. https://doi.org/10.1016/j.cell.2019.08.003.

(47) Murray, D. S.; Schumacher, M. A.; Brennan, R. G. Crystal Structures of QacR-Diamidine Complexes Reveal Additional Multidrug-Binding Modes and a Novel Mechanism of Drug Charge Neutralization. *Journal of Biological Chemistry* **2004**, *279* (14), 14365–14371. https://doi.org/10.1074/jbc.M313870200.

(48) Peters, K. M.; Brooks, B. E.; Schumacher, M. A.; Skurray, R. A.; Brennan, R. G.; Brown, M. H. A Single Acidic Residue Can Guide Binding Site Selection but Does Not Govern QacR Cationic-Drug Affinity. *Plos One* **2011**, *6* (1), e15974. https://doi.org/10.1371/journal.pone.0015974.

(49) Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58* (41), 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735.

(50) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **2003**, *13* (11), 2498–2504. https://doi.org/10.1101/gr.1239303.

(51) Bazzone, A.; Barthmes, M.; Fendler, K. SSM-Based Electrophysiology for Transporter Research. In *Structure-Function Toolbox for Membrane Transporter and Channels*; Ziegler, C., Ed.; 2017; Vol. 594, pp 31–83. https://doi.org/10.1016/bs.mie.2017.05.008.

(52) Stockbridge, R. B.; Koide, A.; Miller, C.; Koide, S. Proof of Dual-Topology Architecture of Fluc F- Channels with Monobody Blockers. *Nature Communications* **2014**, *5*, 5120. https://doi.org/10.1038/ncomms6120.

(53) Sha, F.; Salzman, G.; Gupta, A.; Koide, S. Monobodies and Other Synthetic Binding Proteins for Expanding Protein Science. *Protein Science* **2017**, *26* (5), 910–924. https://doi.org/10.1002/pro.3148.

(54) Sha, F.; Gencer, E. B.; Georgeon, S.; Koide, A.; Yasui, N.; Koide, S.; Hantschel, O. Dissection of the BCR-ABL Signaling Network Using Highly Specific Monobody Inhibitors to the Shp2 Sh2 Domains. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (37), 14924–14929. https://doi.org/10.1073/pnas.1303640110.

(55) Studier, F. W. Protein Production by Auto-Induction in High-Density Shaking Cultures. *Protein Expression and Purification* **2005**, *41* (1), 207–234. https://doi.org/10.1016/j.pep.2005.01.016.

(56) Burgess, R. R. Chapter 17 Refolding Solubilized Inclusion Body Proteins. In *Methods in Enzymology*; Elsevier, 2009; Vol. 463, pp 259–282. https://doi.org/10.1016/S0076-6879(09)63017-2.

(57) Oganesyan, N.; Kim, S.-H.; Kim, R. On-Column Protein Refolding for Crystallization. *Journal of Structural and Functional Genomics* **2005**, *6* (2-3), 177–182. https://doi.org/10.1007/s10969-005-2827-3.

(58) Winter, G.; Waterman, D. G.; Parkhurst, J. M.; Brewster, A. S.; Gildea, R. J.; Gerstel, M.; Fuentes-Montero, L.; Vollmar, M.; Michels-Clark, T.; Young, I. D.; Sauter, N. K.; Evans, G. DIALS: Implementation and Evaluation of a New Integration Package. *Acta Crystallographica Section D-Structural Biology* **2018**, *74* (2), 85–97. https://doi.org/10.1107/S2059798317017235.

(59) Sheldrick, G. M. A Short History of SHELX. *Acta Crystallographica a-Foundation and Advances* **2008**, *64* (1), 112–122. https://doi.org/10.1107/S0108767307043930.

(60) Bricogne, G.; Vonrhein, C.; Flensburg, C.; Schiltz, M.; Paciorek, W. Generation, Representation and Flow of Phase Information in Structure Determination: Recent Developments in and Around SHARP 2.0. *Acta Crystallographica Section D-Biological Crystallography* **2003**, *59* (11), 2023–2030. https://doi.org/10.1107/S0907444903017694.

(61) Abrahams, J. P.; Leslie, A. G. W. Methods Used in the Structure Determination of Bovine Mitochondrial F-1 ATPase. *Acta Crystallographica Section D-Biological Crystallography* **1996**, *52* (1), 30–42. https://doi.org/10.1107/S0907444995008754.

(62) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **2010**, *66* (4), 486–501. https://doi.org/10.1107/S0907444910007493.

(63) Stockbridge, R. B.; Kolmakova-Partensky, L.; Shane, T.; Koide, A.; Koide, S.; Miller, C.; Newstead, S. Crystal Structures of a Double-Barrelled Fluoride Ion Channel. *Nature* **2015**, *525* (7570), 548–+. https://doi.org/10.1038/nature14981.

(64) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *Journal of Applied Crystallography* **2007**, *40* (4), 658–674. https://doi.org/10.1107/S0021889807021206.

(65) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. Refmac5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallographica Section D-Structural Biology* **2011**, *67* (4), 355–367. https://doi.org/10.1107/S0907444911001314.

(66) 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.; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Science* **2018**, *27* (1), 293–315. https://doi.org/10.1002/pro.3330.

(67) Tickle, I. J.; Flensburg, C.; Keller, P.; Paciorek, W.; Sharff, A.; Vonrhein, C.; Bricogne, G. *STARANISO*; Global Phasing Ltd: Cambridge, United Kingdom, 2018.

(68) Jurcik, A.; Bednar, D.; Byska, J.; Marques, S. M.; Furmanova, K.; Daniel, L.; Kokkonen, P.; Brezovsky, J.; Strnad, O.; Stourac, J.; Pavelka, A.; Manak, M.; Damborsky, J.; Kozlikova, B. CAVER Analyst 2.0: Analysis and Visualization of Channels and Tunnels in Protein Structures and Molecular Dynamics Trajectories. *Bioinformatics* **2018**, *34* (20), 3586–3588. https://doi.org/10.1093/bioinformatics/bty386.

CHAPTER 5

Biochemical Characterization of a Guanidinium Exporter from the Small Multidrug Resistance Family

Portions of this work were performed by Olive Burata (alternative substrate K_M measurements).

5.1 Introduction

The Small Multidrug Resistance (SMR) family of transporters has been a source of interest to the membrane transport, folding, function, and evolution fields alike, due to its biochemical tractability, facile expression, and intriguing "dual-topology" nature, where the amino acid sequence does not determine the orientation of the protein in the membrane. This feature allows for the assembly of antiparallel dimeric transporters, with each monomer oriented in the opposite direction.

The first characterized homolog, EmrE from *Escherichia coli*, has served variously as a model for proton-coupled antiport[1,2], multi-specific drug export[3,4], membrane protein folding and topogenesis[5,6], evolutionary diversification, and transporter structure[7,8]. In each area, however, it has tended to thwart simplicity: rather than fulfilling the promise of a simple model, experiments have revealed a wealth of subtle and confounding details that militate against "simple" biochemistry: from folding[9] to stoichiometry[10] to coupling[11], EmrE seems more of a special case than a standard example. While this has created productive research programs exploring this novel biochemistry, the original hope for a simple

model system is apparently unfulfilled.

Recently, representative homologs from the other major subtype of SMRs, called the Gdxs, were characterized functionally and structurally[13]. This led to the unexpected discovery that the majority of the SMRs are likely specific guanidinium/proton antiporters, rather than multidrug exporters. Happily, the functional properties appear much simpler than with EmrE: the coupling stoichiometry is tight[12], and substrate recognition is simpler to rationalize structurally. The presence of guanidinium riboswitches also provides an additional method for inferring function (and the stability thereof).

Previous work, based in analogy to EmrE, was not able to clearly define a function for Gdx-Eco (previously called SugE)[14–16]. To clarify this gap in our understanding of this key model organism, and driven by the observations above, we sought to characterize the *E. coli* Gdx homolog using electrophysiology.

5.2 Results

5.2.1 Gdx-Eco is a good model for the Gdxs

To understand Gdx-Eco's place within the Gdx subtype, and ensure that it is a reasonable candidate for generalization, we conducted a sequence analysis. Evolutionary conservation of residues is a good guide to functional conservation, and so we mapped the conserved residues of the Gdx subtype onto the membrane-embedded topology of Gdx-Eco (Figure 5.1, a). Although the helices vary in their degree of conservation, all of them have at least one highly-conserved helical face evident from the plot. We suspected this represents the substrate-binding residues, with the lipid-facing residues having a lower degree of functional constraint.

To examine this hypothesis, we harnessed the recent high-resolution structure of a Gdx homolog from *Clostridia*, Gdx-Clo. Despite their phylogenetic distance, Gdx-Eco and Gdx-Clo share significant sequence similarity, with 53% identity and 78% similarity, which provides further evidence of broadly conserved function across the subfamily. This degree

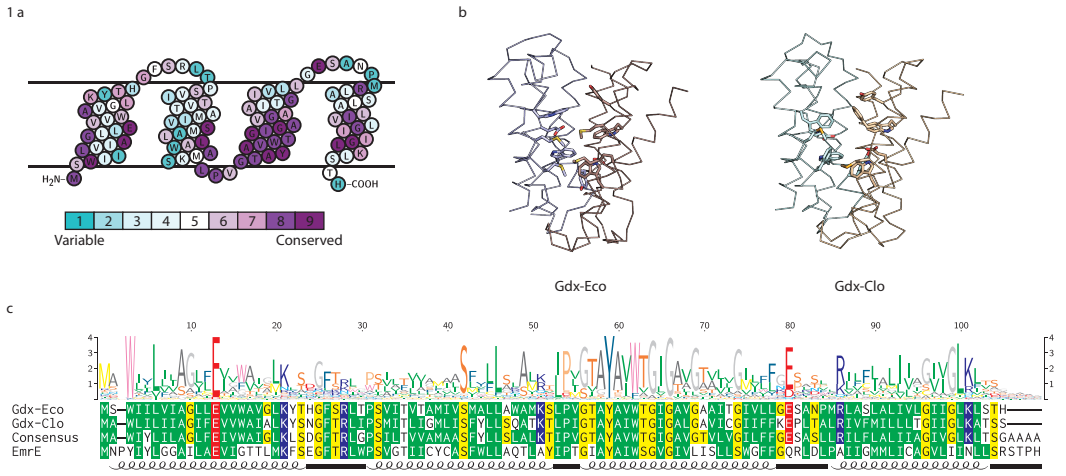

Figure 5.1: Sequence and structural comparison of Gdx-Eco and Gdx-Clo. a) Gdx-Eco topology and domain arrangement, with Gdx subtype conservation coloring. b) Gdx-Clo crystal structure (PDB:6WK5) and Gdx-Eco homology model comparison. Conserved binding-pocket residue sidechains explicitly shown. c) Sequence alignment of Gdx-Eco, Gdx-Clo, Gdx subtype consensus sequence, and EmrE, with sequence logo of Gdx types. Coloring for conserved residues: green, hydrophobic; red, acidic; blue, basic; yellow, polar.

of similarity should support accurate homology modelling, as well. A homology model based on the apo Gdx-Clo structure shows an essentially identical fold, with the conserved positions indeed being oriented towards the interior (Figure 5.1, b). Despite some minor sidechain differences in the model, the binding pocket is much the same.

Finally, the dual-topology generation of EmrE likely proceeds through a transiently misfolded intermediate, driven by marginally hydrophobic helices, which "kinetically anneals" to form stable dimers[9]. In contrast, Gdx helices are more hydrophobic, likely supported simpler folding and topogenesis (SI Figure C.1). In our hands, Gdx subtypes are stabler and more tractable than Qacs, simplifying biochemical analysis.

5.2.2 SSM electrophysiology of Gdx-Eco

We used solid-supported membrane (SSM) electrophysiology to probe Gdx-Eco function. SSM electrophysiology measures capacitively-coupled currents generated from movement across liposomal membranes adsorbed to a sensor chip during a rapid substrate

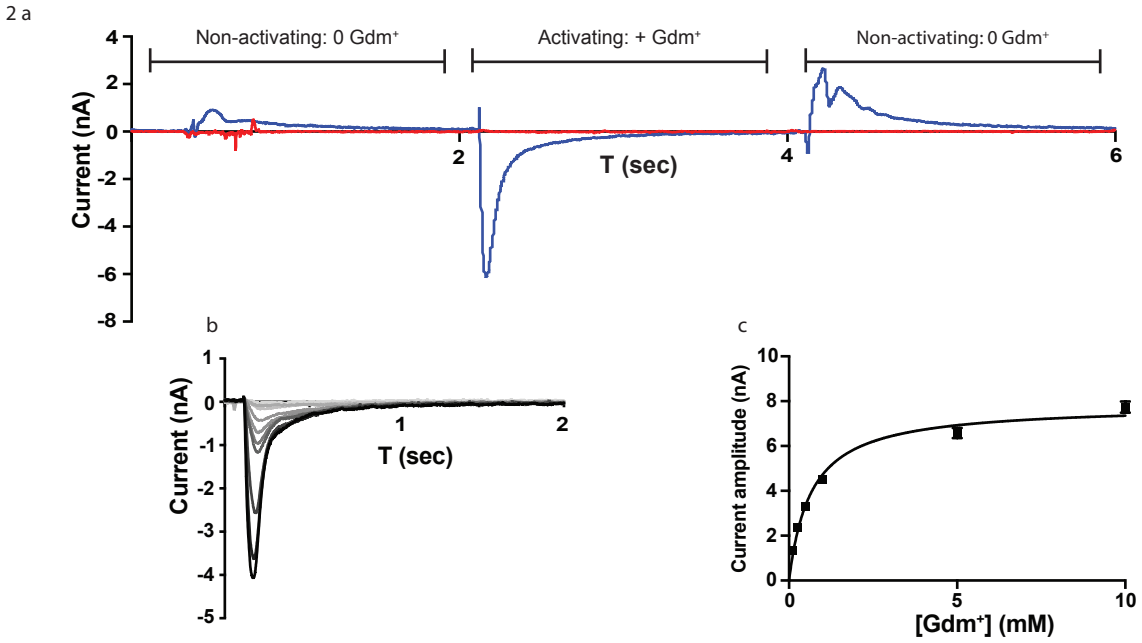

Figure 5.2: Solid-supported membrane electrophysiology of Gdx-Eco. a) Example currents evoked with 1 mM guanidinium titration using protein-free liposomes (red trace) and with Gdx-Eco proteoliposomes (blue trace). b) Guanidinium titration series: substrate perfusion from 10 mM (black) to 100 µM (light grey) guandinium with Gdx-Eco at pH 7. c) Fit of currents to Michaelis-Menten kinetics.

perfusion. It is a sensitive and convenient method for observing electrogenic transport, and it particularly suited for transporters, which are typically intractable via traditional bilayer electrophysiology. The maximum value of the transport-generated current (e.g., 5.2 a, blue trace) is related to turnover kinetics, with the sign indicating the net charge movement for the completely coupled process.

Titration of a single sensor with decreasing levels of guanidinium results in decreasing currents (5.2 b), which can be fit to Michaelis-Menten kinetics to obtain an apparent K_M value (5.2 c). At neutral pH, the K_M of 660 µM is similar to other *E. coli* proton-coupled transporters.

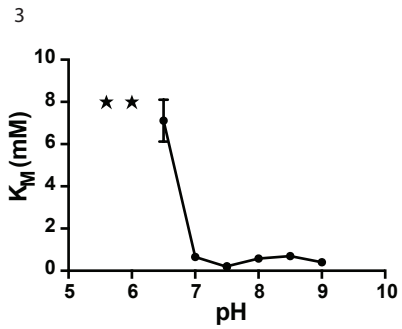

Figure 5.3: pH dependence of Gdx-Eco currents. K_M values obtained from 3 titration series at indicated pH: at pH 6 and below, no titrations were observed and data could not be fit (star).

5.2.3 pH dependence of Gdx-Eco transport

Given the importance of the glutamates for coupling the export of guanidinium to the import of protons, we wanted to examine the sensitivity of apparent K_M to the ambient pH during transport. The pK_a of EmrE has been a subject of debate, but it is clearly elevated several units with respect to free glutamate. We generated a series of K_Ms at different pHs (5.3), which revealed a weak dependence above pH 7.0, but a precipitous dropoff below that. Indeed, at 6.0 or below, we struggled to observe any currents at all, let alone fit them to a titration. This is consistent with the expected physiology, where a lower substrate affinity the lower pH of the periplasm would support directional export. The possibility of an altered non-electrogenic stoichiometry (i.e., one proton and one guanidinium imported for two protons exported) cannot be ruled out with these observations, however.

5.2.4 Transport of substituted guanidinium compounds

The substrate specificity of Gdx-Eco was also probed with SSM, by comparing the currents evoked from individual sensors from different compounds and normalizing them to guanidinium at 1 mM for each compound. This resulted in an affinity series where singly-substituted guanidiniums are transported identically to guanidinium (figure 5.4 a, red bars). Neither arginine nor agmatine, both singly-substituted guanidiniums, are trans-

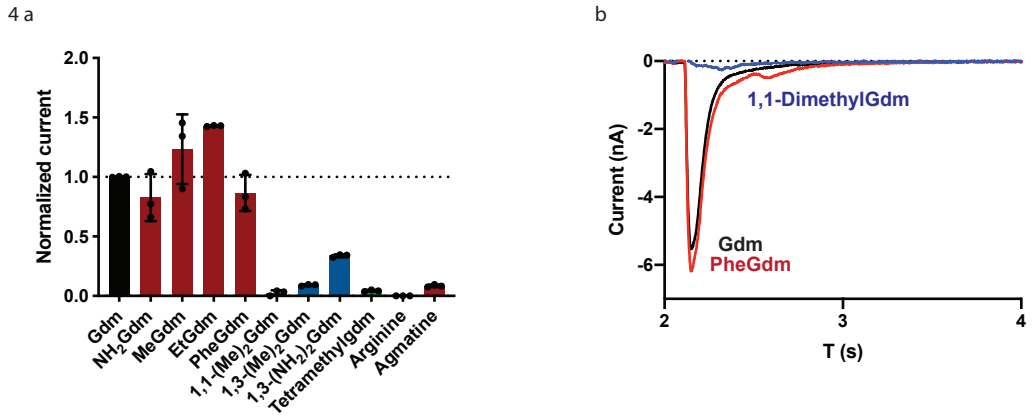

Figure 5.4: Substrate selectivity of Gdx-Eco. a) Normalized relative transports from a series of substituted guanidinium compounds or other analogs at 1 mM. Currents were normalized to the 1 mM currents from 1 mM guanidinium on the same sensor. Red bars, monosubstituted guanidiniums; blue bars, disubstituted guanidiniums; green bars, other. b) Example traces at 1 mM substrate concentration.

ported, which was also observed with Gdx-Clo. Substitutions at two different nitrogens, or twice on the same nitrogen, largely abolish transport (figure 5.4 a, blue bars). As in Gdx-Clo, recognition is likely mediated by bidentate H-bonding to the glutamates, which cannot support double-substituted guanidiniums. One compound, 1,3-diaminoguanidinium, reintroduces a substituent capable of H-bonding on its own, and is observed to support some transport.

The lineshapes of guanidinium and phenylguanidinium transport are similar, reflecting similar transport kinetics (figure 5.4 b); 1,1-dimethylguanidinium displays essentially no transport at 1 mM.

We were curious about this singular distinction with Gdx-Clo, and so we conducted a titration with increasing 1,1-dimethylguanidinium (figure 5.5 a). We observed binding kinetics with a K_M of 3.8 mM, almost an order of magnitude higher than guanidinium. We also conducted a titration series with phenylguanidinium, and found, on the other hand, a much lower K_M of 190 µM (figure 5.5 b). This may represent the favorable interaction of the aromatic phenyl ring with the multi-aromatic binding pocket of Gdx-Eco.

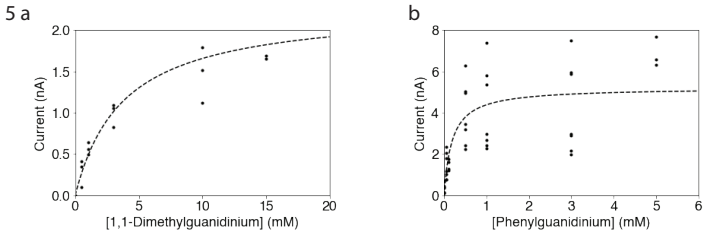

Figure 5.5: Transport kinetics of alternative substrates with Gdx-Eco. Currents evoked by titration with a) 1,1-dimethylguanidinium and b) phenylguanidinium, and Michaelis-Menten fits to titrations.

5.3 Materials and Methods

5.3.1 Bioinformatics

Briefly, a representative set of Gdx sequence was generated by clustering the large subcomponent of Gdx-containing sequences from a sequence-similarity network that was generated previously[13,18]. The sequences in this subcomponent were clustered using CD-HIT with a cutoff of 50% identity[19], then aligned with MAFFT in L-INS-i mode[20]. These sequences were used in the ConSurf webserver, with the Gdx-Clo apo structure[21], with a maximum-likelihood tree generated using the JTT model. The resulting conservation scores were mapped onto the Gdx-Eco residues and displayed with the Protter webserver[22]. The alignment and sequence logo were visualized using TeXshade[23].

The homology model was generated using the Robetta webserver in comparative modeling (CM) mode, with two copies of the monomeric sequence as template[24]. The first model (of 5) was used for analysis, as it had the highest confidence as well as the lowest RMSD to Gdx-Clo.

Insertion ΔG_{app} were predicted using the ΔG prediction server v1.0[25].

5.3.2 Protein expression and reconstitution

Lipids were from Avanti, detergents from Anatrace. Gdx-Eco was expressed with a C-terminal hexahistidine affinity tag and a LysC recognition site, and was cloned into a

pET-21c expression vector. This was transformed into C41 (DE3) E. coli cells for expression. When cultures reached an OD600 of 1.0, 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 a cobalt affinity column, washed with 100 mM NaCl, 20 mM imidazole, and then eluted with 400 mM imidazole while monitoring UV absorbance at 280 nm. 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.

For reconstitution, 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). Proteoliposomes were prepared with 20 mg E. coli polar lipid per ml, and a 1:25 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. Protein-free liposomes were created by following the same procedure with no protein present. After the final round of dialysis, proteoliposomes were aliquoted and stored at −80 °C until use.

5.3.3 SSM electrophysiology

SSM electrophysiology was conducted using a SURFE2R N1 instrument (Nanion Technologies, Munich, Germany) according to published protocols and our previous work[13]. 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 µg/µl 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 × 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[26].

For pH dependence measurements, the liposomes were reconstituted in a mixed buffer, with 33 mM MES, 34 mM HEPES, and 33 mM bicine, and 100 mM KCl.

5.4 Discussion

Our results support the notion that Gdx-Eco is an appropriate model system for coupled transport, with well-behaved and well-coupled transport amenable to electrophysiology and other functional measurements. The small set of transported substrates with clear chemical distinctions also provides an amenable system for understanding the structural basis of substrate recognition, as was accomplished with Gdx-Clo. Further investigation into additional non-natural compounds may clarify the bounds of the binding pocket, as well as the nature of the conserved arginine/agmatine exclusion.

As a central model organism, *E. coli* has a large set of annotated metabolic pathways and transporters. With the precise physiological role of guanidinium export in bacteria unclear, a precise understanding of the function of this transporter is essential to utilizing this large amount of underlying data. Additionally, the specific organism hosting Gdx-Clo is not known, so a holistic understanding of it in its native physiology is difficult to obtain.

Finally, we have found SSM to be a useful tool for further characterization of secondary-coupled transporters, with an expanded set of assays compared to our previous work. The ability to readily measure apparent substrate affinities in different pH environments is a boon to understanding the nature of coupling in these small machines. We hope to continue this further with the addition of mutagenesis to uncover sequence-specific determinants of coupling in the SMRs.

References

(1) Yerushalmi, H.; Schuldiner, S. An Essential Glutamyl Residue in EmrE, a Multidrug Antiporter from Escherichia Coli. *Journal of Biological Chemistry* **2000**, *275* (8), 5264–5269. https://doi.org/10.1074/jbc.275.8.5264.

(2) Morrison, E. A.; Henzler-Wildman, K. A. Transported Substrate Determines Exchange Rate in the Multidrug Resistance Transporter EmrE. *Journal of Biological Chemistry* **2014**, *289* (10), 6825–6836. https://doi.org/10.1074/jbc.M113.535328.

(3) Cho, M.-K.; Gayen, A.; Banigan, J. R.; Leninger, M.; Traaseth, N. J. Intrinsic Conformational Plasticity of Native EmrE Provides a Pathway for Multidrug Resistance. *Journal of the American Chemical Society* **2014**, *136* (22), 8072–8080. https://doi.org/10.1021/ja503145x.

(4) Brill, S.; Sade-Falk, O.; Elbaz-Alon, Y.; Schuldiner, S. Specificity Determinants in Small Multidrug Transporters. *Journal of Molecular Biology* **2015**, *427* (2), 468–477. https://doi.org/10.1016/j.jmb.2014.11.015.

(5) Nasie, I.; Steiner-Mordoch, S.; Gold, A.; Schuldiner, S. Topologically Random Insertion of EmrE Supports a Pathway for Evolution of Inverted Repeats in Ion-Coupled Transporters. *Journal of Biological Chemistry* **2010**, *285* (20), 15234–15244. https://doi.org/10.1074/jbc.M110.108746.

(6) Seppala, S.; Slusky, J. S.; Lloris-Garcera, P.; Rapp, M.; von Heijne, G. Control of Membrane Protein Topology by a Single C-Terminal Residue. *Science* **2010**, *328* (5986), 1698–1700. https://doi.org/10.1126/science.1188950.

(7) Shcherbakov, A. A.; Hisao, G.; Mandala, V. S.; Thomas, N. E.; Soltani, M.; Salter, E. A.; Davis, J. H.; Henzler-Wildman, K. A.; Hong, M. Structure and Dynamics of the Drug-Bound Bacterial Transporter EmrE in Lipid Bilayers. *Nat Commun* **2021**, *12* (1), 172. https://doi.org/10.1038/s41467-020-20468-7.

(8) Ubarretxena-Belandia, I. Three-Dimensional Structure of the Bacterial Multidrug Transporter EmrE Shows It Is an Asymmetric Homodimer. *The EMBO Journal* **2003**, *22* (23), 6175–6181. https://doi.org/10.1093/emboj/cdg611.

(9) Seurig, M.; Ek, M.; von Heijne, G.; Fluman, N. Dynamic Membrane Topology in an Unassembled Membrane Protein. *Nat Chem Biol* **2019**, *15* (10), 945–948. https://doi.org/10.1038/s41589-019-0356-9.

(10) Rotem, D.; Schuldiner, S. EmrE, a Multidrug Transporter from Escherichia Coli, Transports Monovalent and Divalent Substrates with the Same Stoichiometry. *Journal of Biological Chemistry* **2004**, *279* (47), 48787–48793. https://doi.org/10.1074/jbc.M408187200.

(11) Morrison, E. A.; Henzler-Wildman, K. A. Transported Substrate Determines Exchange Rate in the Multidrug Resistance Transporter EmrE. *Journal of Biological Chemistry* **2014**, *289* (10), 6825–6836. https://doi.org/10.1074/jbc.M113.535328.

(12) Kermani, A. A.; Macdonald, C. B.; Gundepudi, R.; Stockbridge, R. B. Guanidinium Export Is the Primal Function of SMR Family Transporters. *Proc Natl Acad Sci USA* **2018**, *115* (12), 3060–3065. https://doi.org/10.1073/pnas.1719187115.

(13) Kermani, A. A.; Macdonald, C. B.; Burata, O. E.; Ben Koff, B.; Koide, A.; Denbaum, E.; Koide, S.; Stockbridge, R. B. The Structural Basis of Promiscuity in Small Multidrug Resistance Transporters. *Nat Commun* **2020**, *11* (1), 6064. https://doi.org/10.1038/s41467-020-19820-8.

(14) Chung, Y. J.; Saier, M. H. Overexpression of the *Escherichia Coli sugE* Gene Confers Resistance to a Narrow Range of Quaternary Ammonium Compounds. *J Bacteriol* **2002**, *184* (9), 2543–2545. https://doi.org/10.1128/JB.184.9.2543-2545.2002.

(15) Sikora, C. W.; Turner, R. J. SMR Proteins SugE and EmrE Bind Ligand with Similar Affinity and Stoichiometry. *Biochemical and Biophysical Research Communications* **2005**, *335* (1), 105–111. https://doi.org/10.1016/j.bbrc.2005.07.051.

(16) Bay, D. C.; Turner, R. J. Spectroscopic Analysis of the Intrinsic Chromophores Within Small Multidrug Resistance Protein SugE. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2011**, *1808* (9), 2233–2244. https://doi.org/10.1016/j.bbamem.2011.05.005.

(17) Van Lehn, R. C.; Zhang, B.; Miller, T. F. Regulation of Multispanning Membrane Protein Topology via Post-Translational Annealing. *eLife* **2015**, *4*, e08697. https://doi.org/10.7554/eLife.08697.

(18) Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58* (41), 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735.

(19) Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the Next-Generation Sequencing Data. *Bioinformatics* **2012**, *28* (23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

(20) Katoh, K.; Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **2013**, *30* (4), 772–780. https://doi.org/10.1093/molbev/mst010.

(21) Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: An Improved Methodology to Estimate and Visualize Evolutionary Conservation in Macromolecules. *Nucleic Acids Res* **2016**, *44* (W1), W344–350. https://doi.org/10.1093/nar/gkw408.

(22) Omasits, U.; Ahrens, C. H.; Müller, S.; Wollscheid, B. Protter: Interactive Protein Feature Visualization and Integration with Experimental Proteomic Data. *Bioinformatics* **2014**, *30* (6), 884–886. https://doi.org/10.1093/bioinformatics/btt607.

(23) Beitz, E. TeXshade: Shading and Labeling of Multiple Sequence Alignments Using LaTeX2e. *Bioinformatics* **2000**, *16* (2), 135–139. https://doi.org/10.1093/bioinformatics/16.2.135.

(24) Song, Y.; DiMaio, F.; Wang, R. Y.-R.; Kim, D.; Miles, C.; Brunette, T.; Thompson, J.; Baker, D. High-Resolution Comparative Modeling with RosettaCM. *Structure* **2013**, *21* (10), 1735–1742. https://doi.org/10.1016/j.str.2013.08.005.

(25) Hessa, T.; Meindl-Beinker, N. M.; Bernsel, A.; Kim, H.; Sato, Y.; Lerch-Bader, M.; Nilsson, I.; White, S. H.; von Heijne, G. Molecular Code for Transmembrane-Helix Recognition by the Sec61 Translocon. *Nature* **2007**, *450* (7172), 1026–1030. https://doi.org/10.1038/nature06387.

(26) Bazzone, A.; Barthmes, M.; Fendler, K. SSM-Based Electrophysiology for Transporter Research. In *Structure-Function Toolbox for Membrane Transporter and Channels*; Ziegler, C., Ed.; 2017; Vol. 594, pp 31–83. https://doi.org/10.1016/bs.mie.2017.05.008.

CHAPTER 6

A Multi-Context Deep Mutational Scan to Uncover the Evolutionary Forces behind Gene Duplicate Retention in the Flucs

The work in this chapter was performed with the help of Troy Cao.

Portions of this work also appeared in Troy Cao's senior honors thesis.

6.1 Introduction

Gene duplication is the primary source of novel genetic material in living systems[1,2]. It plays roles in processes as varied as speciation, xenobiotic tolerance, and antibiotic resistance[3]. Despite its core role in evolution, the key question of how duplicates are retained prior to developing new functions has received little experimental investigation, and what does exist is confined to soluble proteins[4].

Many gene products form complexes with other proteins. In the case of an initially homodimeric unduplicated gene, the creation of a duplicate potentially introduces a new structural complexity. These may suffer a variety of fates after duplication (figure 6.1 a). They may quickly lose function in one or the other, creating a pseudogene which can be lost. They may simply retain their function, effectively altering the gene dosage. Some outcomes might lead to changes in complex assembly, however: changes to the dimerization interface which either necessitate or preclude paralog dimerization will create complexes which are not equivalent to the unduplicated state, and which may introduce new functional and structural complexity. Each case requires the duplicates to remain long

enough to accumulate these beneficial mutations[5–7]. The evolutionary forces which do so are unclear.

One possible mechanism supporting duplicate retention is selective redundancy: that is, two genetic copies of a gene allowing mutational buffering. Other models propose that duplication relaxes functional constraints on multi-functional proteins, allowing duplicates to specialize; still others suppose that a degradation in the function of one paralog that requires complementation by the other may be responsible for duplicate retention[6]. Genomic data clearly reveals the importance of this, with in many species the majority of the genes existing as duplicates[3]. The rate of adaptive mutations is much lower than the duplication rate, however, which raises the question of exactly how duplicates are retained until they strike on a beneficial change.

Duplication in membrane proteins has a specific character as well, where duplicated domains that have fused with opposite orientations with respect to the membrane, or "inverted repeats", being commonly observed in transporter structures (figure 6.1 b). Indeed, the majority of alpha-helical membrane transporters with solved structures have symmetric domains that arose through gene duplication and fusion[8], implying a pervasive and essential role for duplication. Intriguingly, these domains are mostly oriented antiparallel with respect to the membrane: this inverted-repeat architecture would require that the unduplicated precursors be able to insert in both orientations in the membrane, or form dual-topology dimers (figure 6.1 b).

This hypothesis has been confirmed by the recent discovery and structure determination of a family of dual-topology fluoride channels, known as the Flucs. Fluoride, an environmental toxin, can accumulate intracellularly under certain conditions. These channels alleviate this stress by allowing fluoride to flow out of the cell. This broadly-distributed family exists across eukaryotes, archaeans, and bacteria, and also has the unique distinction of existing in all states of the evolutionary trajectory: as dual-topology homodimers, fixed-topology heterodimers, and fused inverted-repeat monomers. This makes them an

obvious candidate to study gene duplication, as phylogenetic evidence suggests has happened independently multiple times across this family[9].

In this work, we present a novel phenotypic screen that we combine with a multi-context deep mutational scan to uncover the logic of duplicate retention in the Flucs.

6.2 Results

6.2.1 Gene duplication has shaped Fluc evolution

The Flucs are a widespread family of electrodiffusive fluoride channels which protect organisms against and fluoride stress, which accumulates intracellularly under acidic conditions. Structural and electrophysiological investigation of them has revealed an intriguing two-pore character, where each functional dimeric unit contains two distinct pores, oriented oppositely to each other, each capable of supporting fluoride export by itself (figure 6.1 c). Mutagenesis has revealed several essential residues involved in fluoride transport, including a central aromatic structure called the "phenylalanine box," composed of two phenylalanines from each monomer. These fall within a conserved helix break in the center of TM 3 (figure 6.1 d). Mutagenesis has demonstrated that single point mutations here (F80I, F83I) are sufficient to abolish function[10].

In homodimeric Flucs, both phenylalanines are conserved (figure 6.1 d). Heterodimeric Flucs have a tendency for mutations in one or the other position, potentially ablating one pore: this is also seen in the fused multi-domain Flucs[9]. This finding suggests that only one pore is necessary for function, whereas the symmetry of the homodimeric structure requires that both pores always be functional.

6.2.2 Construction of an assay linking gene context and phenotype

The challenge for experiment here is constructing an assay that specifically reveals the effects of duplication *qua* duplication, vs gene dosage effects. Based on the above observations, we reasoned that we could differentiate different gene product assemblies

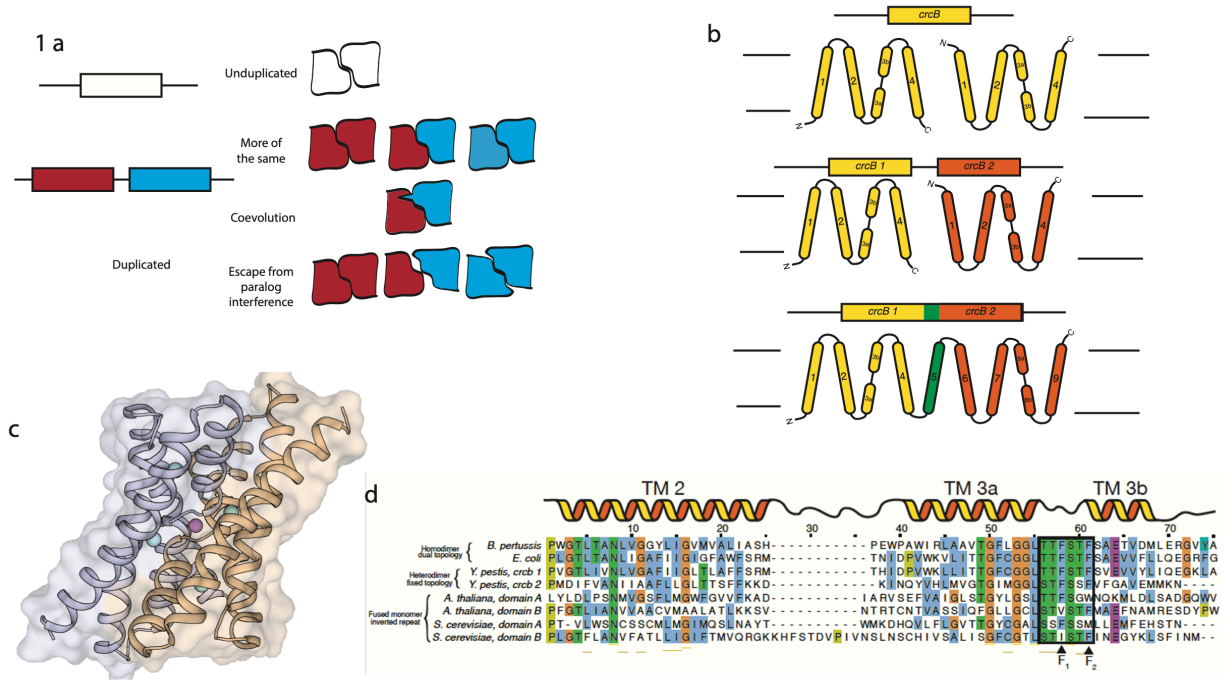

Figure 6.1: Gene duplication plays an important role in membrane protein evolution. A) Fates of duplicated genes which form homodimeric complex. If binding is preserved, the result is maintenance of function with changes in gene dosage. If binding is impaired in one duplicate, the other may accumulate compensatory mutations to achieve a purely heteromeric interaction. Changes in the interface may alternatively partition the two gene products and remove paralog interference. B) Membrane protein trajectories with inverted repeats. An initially unduplicated gene which forms a homodimeric product where each subunit is oriented oppositely in a dual-topology state may, via duplication, lead to an obligate dimeric state composed of two subunits with fixed topologies. A fusion event, with attendant gain of a linker domain, results in a single gene with an internal homology. The Flucs (gene name crcB) contain members in all three states. C) Crystal structure of *E. coli* dual-topology Fluc channel (PDB: 6BX5). Central sodium in purple, and crystallographic fluoride ions in teal. D) Sequence alignment of representative Fluc channels focused on central "phenylalanine box" motif. In duplicated and fused homologs, there is a tendency for a single pore to be retained.

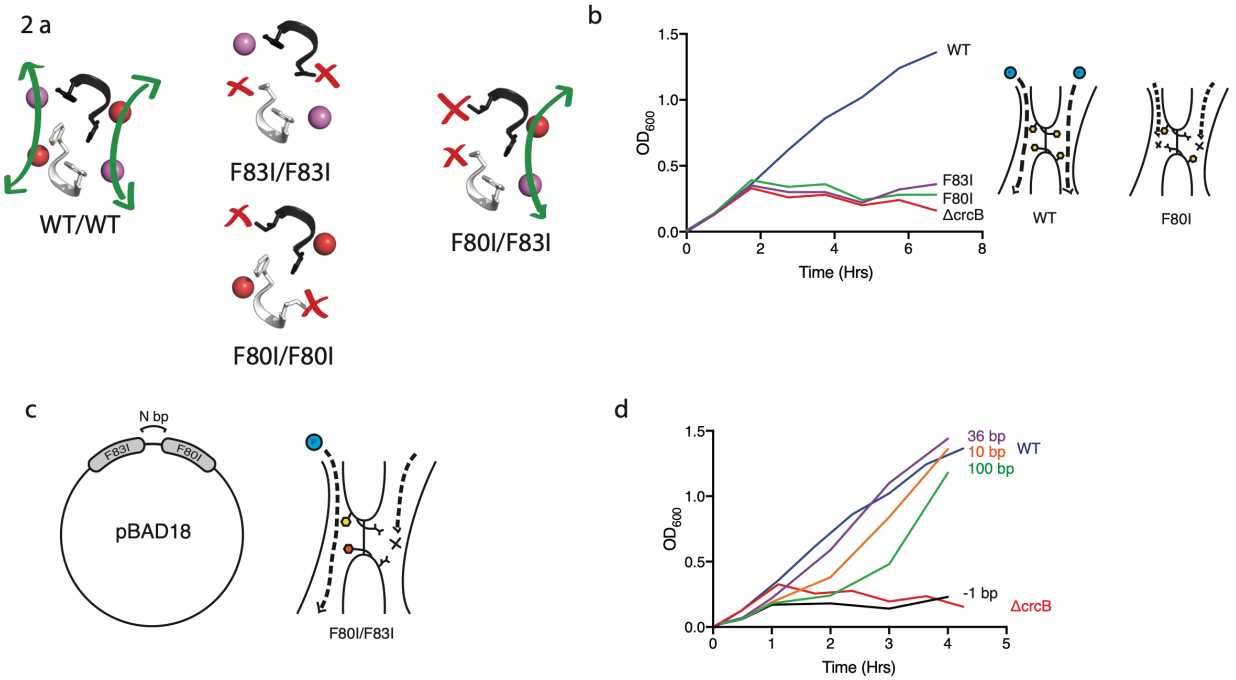

Figure 6.2: Details of heteromeric construct. A) Closeup of the "phenylalanine box" motif which regulates two-pore fluoride traffic. Wild-type homodimers retain both pores, while homodimeric mutants (F80I or F83I) ablate both pores due to the complex's symmetry. A heterodimer, however, retains a single pore. B) Sensitivity of *E. coli* cells to 10 mM NaF in rich media. A wild-type strain shows no growth defect, but a knockout (ΔcrcB) strain is unable to grow. Transformation with plasmid-encoded mutant Fluc protein is unable to rescue growth. C) Details of the heterodimeric expression construct. The two mutants were cloned into a single operon with a variable length between the two. D) Transformation of the sensitive knockout strain with the heterodimeric construct can rescue growth, but shows some sensitivity to the distance between the two genes.

using the Phe-box: while homodimeric wildtype dimers have two functional pores, and both forms of mutant homodimers have no pores, a mixed mutant heterodimer actually recovers a single functional pore (figure 6.2 a, b)!

We created a construct with two copies of *E. coli* Fluc in a single operon, each with a distinct mutation but otherwise identical sequences, and cloned it into a pBAD18 vector (figure 6.2 c). Transforming this into *E. coli* robustly rescues growth, demonstrating that heteromeric assemblies are being made *in vivo* (figure 6.2 d). Intriguingly, there was some slight sensitivity to the distance between the two genes in the operon: in most cases, there was little differences between constructs, but an overlapping pair was unable to rescue

growth.

6.2.3 Construction of a mutational library for a multi-context deep mutational scan

Using this construct, we planned to create a mutational library in three different genetic constructs to investigate the evolutionary forces specific to duplication that are at work immediately after duplication. Deep mutational scanning, coupled to selection with NaF, will allow us to characterize different apparent fitness effects of mutations in different genetic constructs, and answer how the presence of a duplicate changes the overall fitness landscape of a single gene. Beyond being a novel experimental intervention in the field, the application of saturation mutagenesis to a membrane protein itself represents an important theoretical development as well[11].

Following a nicking mutagenesis strategy, we used an oligo pool containing every amino acid substitution at every site in our gene (figure 6.3)[12,13]. We then subclone these into three different selection vectors: by itself, in an operon with WT Fluc, or with an obligate heterodimeric (F80I) Fluc. During subcloning, unique molecular identifiers were added, so that all variants in all subcloned libraries share the same identifiers. These identifiers will be mapped to variants using long-read sequencing.

Armed with an experimental system which provides specific selective conditions, as fluoride stress does for fluoride channels, we can utilize saturation mutagenesis to both quantitatively determine the residue-level determinants of function in our WT protein and also broadly determine the spectra of allowed, forbidden, and neutral mutations in different genetic contexts.

To investigate the hypothesis that duplication provides an adaptive robustness by offering a mutational buffer, we will conduct parallel saturation mutagenesis of singleton wildtype, duplicated wildtype, and heterodimeric (mixed F80I/F83I) constructs. The relative enrichment of mutations between the singleton and duplicated cases will report on mutational buffering and dominant negative properties of the system. The comparison

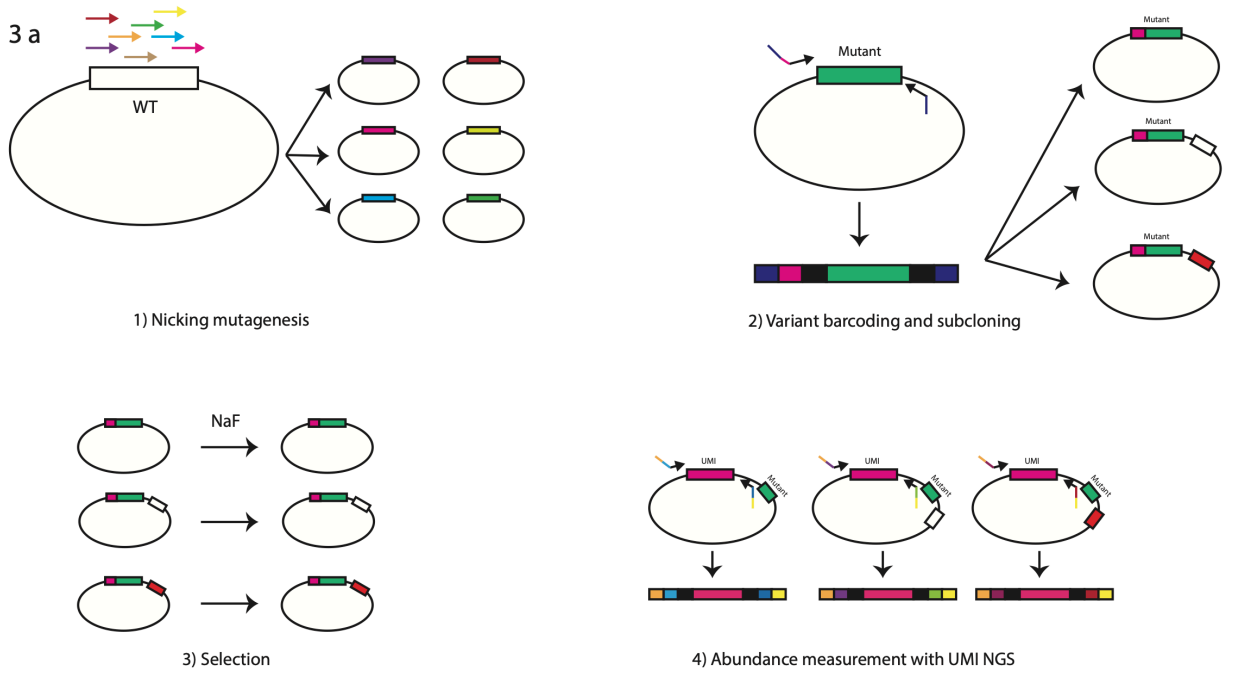

Figure 6.3: Construction of mutational library and DMS overview. Nicking mutagenesis using an oligo pool produces a library of sequence variants. Unique barcodes are introduced during subcloning into three different genetic contexts. Each library is then passed through NaF selection in parallel. After mapping of unique barcodes to genotypes, the abundances of pre- and post-selection libraries are calculated using amplicon sequencing of the barcodes.

between the wildtype duplicates and the mixed heterodimers will indicate whether the introduction of specifically heterodimeric complexes, rather than simply gene dosage, has an adaptive role, as we expect should be the case for this membrane complex.

6.3 Methods

Saturation mutagenesis is a relatively new technique that takes advantage of the rapid developments in next-generation sequencing to comprehensively investigate genotype/phenotype relationships in an unbiased way. In a system with a well-defined map between protein function and overall organismal fitness, population expansion under selective conditions should be proportional to overall protein function. Deep sequencing of the starting and post-selection populations can be used to determine enrichment of strains bearing different mutant proteins, which can be used to calculate the overall energetic impact of a given mutation. Assembly of a mutagenic library: To construct a mutational library, we will use a technique known as nicking mutagenesis[12,13]. This first creates a pool of single-stranded WT plasmid DNA through the combined action of a nicking endonuclease that acts on only one strand and an exonuclease that degrades the nicked strand. A pool of mutagenic oligos is then used to synthesize the complementary strand. The process is repeated with the opposite nicking endonuclease, except a universal primer is used to synthesize the complementary strand. The final library is purified and transformed into E. coli for selection. This process will be conducted with a starting singleton WT construct, a bicistronic WT/WT construct, and a heterodimeric F83I/F80I construct as above. Selection and sequencing: The E. coli populations containing the libraries will be subject to growth on NaF, which provides a clear selective pressure on growth (Figure 2). Samples before and after selection will be taken and subject to deep sequencing in collaboration with the Advanced Genomics Core. To overcome read-length and fidelity limitations, sequencing primers will incorporate unique barcodes[14]. The relative enrichments of mutants will be determined by amounts of reads.

6.4 Methods

6.4.1 Bacterial strains

Cloning and plasmid preparation were performed in E. coli DH5α (NEB). For protein expression and western blotting, E. coli BL21 (DE3) was used. Resistance and growth assays were conducted using Keio collection strains: obtained from Coli Genetic Stock Center: E. coli BW25113 served as WT background, and ΔcrcB::kanR sensitive mutants[15]. In ΔcrcB::kanR cells, the genomic Fluc gene (crcB) is replaced with a gene coding for kanamycin resistance.

6.4.2 Plasmids and cloning.

Fluc Ec2 gene was cloned into a pBAD18 vector in front of an arabinose operon promoter at XbaI and HindIII sites using overlap-extension PCR[16]. The primers were specific for the coding sequence, and removed the affinity purification sequence from the original expression vector. The vector contains the bla gene coding for a β-lactamase which provides resistance to carbenicillin, allowing for transformant selection. F80I and F83I mutants were generated using Q5 site-directed mutagenesis kit (NEB) The coding sequence of the F83I gene was amplified using primers containing overhanging NheI and KpnI sites, then cloned into the F80I plasmid construct by digestion and ligation with T4 ligase. The correct assembly was verified by Sanger sequencing. The resulting bicistronic construct (two-gene system) contained 36 base pairs from the multiple cloning site between the two genes, and is referred to as thus in the main text. Additional two-gene constructs were generated using mutagenesis with the bicistronic construct (deletions and additions of randomly assembled DNA using DNA randomizer programs).

For western blotting, the Fluc Ec2 gene was cloned into a vector incorporating a C-terminal 1D4 tag. An anti-1D4 antibody was used. F80I and F83I Fluc mutants were generated from this 1D4-tagged Fluc Ec2 gene by site-directed mutagenesis as described

above.

6.4.3 Library generation

Starting with the Fluc Ec2 gene above, a mutagenic library was constructed using the method of Medina-Cucurella[12,13]. Briefly, a BbvCI site was added after the Ec2 sequence with site-directed mutagenesis. This starting template was digested with Nt.BbvCI, ExoIII, and ExoI to generate a single-stranded template. The pooled and phosphorylated mutagenic oligos were then added (in a 1:200 dilution), and a mutagenic strand was synthesized using Phusion HF polymerase with Taq ligase. After column purification, the bottom strand was then digested using Nb.BbvCI, ExoIII, and ExoI. The second strand was synthesized using a previously phosphorylated secondary primer that was not specific for the gene using Phusion HF polymerase and Taq ligase. After a final column purification, the library was electroporated into NEB 5-alpha electrocompetent cells and plated on large BioAssay plates with carbenicillin. The next day, the entire growth was scraped using LB and miniprepped. This library was transformed into electrocompetent ΔcrcB cells for selection.

6.4.4 Growth assays.

For growth assays, different linker-length constructs were transformed into chemically-competent ΔcrcB::kanR E. coli cells and plated on lysogeny broth (LB) agar plates containing carbenicillin (100 µg/ml) and kanamycin (50 µg/ml) overnight. Single colonies were then picked and used to grow overnight cultures in 5 ml LB media supplemented with 10 µl of carbenicillin (100 µg/ml) and 5 µl of kanamycin (50 µg/ml). Saturated overnight cultures were diluted 1:100 into LB media supplemented with 50 mM sodium phosphate at pH 7.4 (to buffer pH) to start new growth assays. After 30 minutes, 0.2% sterile arabinose was added to culture. After an additional 30 minutes to allow induction, sterile sodium fluoride (NaF) was added at various concentrations. OD was monitored after addition of

arabinose and NaF and then at one hour intervals throughout growth.

For the competition assay, an overnight culture of each construct was prepared independently with appropriate antibiotics. The next morning, new cultures were created by diluting the overnight culture 1:100 in fresh media. After the cultures had reached exponential phase (by OD), the competition growths were prepared by diluting each culture into a shared tube to a final OD of 0.1 each (0.4 total), with appropriate antibiotics and supplements (arabinose and/or NaF). The next day, 1 mL of cells was pelleted by benchtop centrifuge (13,000 g), then resuspended in 200 uL lysis buffer. PCR was performed with 1 uL of resuspended cells using pBAD primers, then PCR products were run on 1.5% agarose gel. The gel was stained with ethidium bromide and visualized using UV transilluminator.

6.4.5 Western blotting.

To verify expression of Fluc protein and membrane localization, western blots of whole cells and the membrane fraction were performed. For whole-cell blots, cells were transformed with 1D4-tagged constructs and grown to an OD of 1.0, then induced with IPTG. After three hours, 1 mL of culture was harvested by centrifugation at 13,000 G. The pellet was solubilized in lysis buffer, then 10 µL of extract was run on SDS-PAGE gel. The gel was transferred to a nitrocellulose membrane, then washed and stained with mouse anti-1D4 and AP goat anti-mouse secondary antibodies. The membrane was developed using BCIP and NBT staining and then imaged.

For Western blot of the membrane fraction, the cells were first pelleted as above, then resuspended in lysis buffer before the membrane fraction was isolated using high-speed centrifugation (30,000G for an hour). The pellet was then resuspended again and blotted similar to the whole-cell western blot.

6.4.6 Bioinformatics.

Bacterial and archaean protein sequences belonging to crcB family (PFAM PF02537) were retrieved from Uniprot[17]. Three rounds of iterative clustering using CD-HIT was used to reduce sequence redundancy, which resulted in ~500 unique clusters[18]. One representative sequence was taken from each cluster. A multiple sequence alignment was generated using MAFFT[19]. Residues in aqueous-exposed regions, based on structural alignment to solved Fluc structures, were removed the alignment, in order to remove bias caused by convergent evolution based on topological control via the positive-inside rule[20]. A maximum-likelihood tree was then generated using PhyML[21]. Annotation was performed as in[9].

6.5 Future directions.

With the completion of this preliminary work, the selection and analysis of the alternative genetic contexts can proceed. Reconstruction of the fitness landscapes, and observation of how genetic contexts change them, will provide empirical information that may be used to distinguish between duplicate retention models. Additionally, with the recent creation of a pair of oriented mutants that, in combination, form functional dimers, an additional comparison may be made. This would shed light on the relative roles of topology and functional interactions for shaping the dynamics of paralogs after duplication. This may also provide useful information for understanding the observed patterns of duplication, topological fixation and functional drift that are observed in the Fluc phylogeny.

References

(1) Ohno, S. *Evolution by Gene Duplication*; 1970.

(2) Lynch, M. The Evolutionary Fate and Consequences of Duplicate Genes. *Science* **2000**, *290* (5494), 1151–1155. https://doi.org/10.1126/science.290.5494.1151.

(3) Zhang, J. Evolution by Gene Duplication: An Update. *Trends in Ecology & Evolution* **2003**, *18* (6), 292–298. https://doi.org/10.1016/S0169-5347(03)00033-8.

(4) Dhar, R.; Bergmiller, T.; Wagner, A. Increased Gene Dosage Plays a Predominant Role in the Initial Stages of Evolution of Duplicate TEM-1 Beta Lactamase Genes. *Evolution* **2014**, *68* (6), 1775–1791. https://doi.org/10.1111/evo.12373.

(5) Force, A.; Lynch, M.; Pickett, F. B.; Amores, A.; Yan, Y. L.; Postlethwait, J. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **1999**, *151* (4), 1531–1545.

(6) Hahn, M. W. Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates. *Journal of Heredity* **2009**, *100* (5), 605–617. https://doi.org/10.1093/jhered/esp047.

(7) Kaltenegger, E.; Ober, D. Paralogue Interference Affects the Dynamics After Gene Duplication. *Trends in Plant Science* **2015**, *20* (12), 814–821. https://doi.org/10.1016/j.tplants.2015.10.003.

(8) Hennerdal, A.; Falk, J.; Lindahl, E.; Elofsson, A. Internal Duplications in α-Helical Membrane Protein Topologies Are Common but the Nonduplicated Forms Are Rare. *Protein Science* **2010**, *19* (12), 2305–2318. https://doi.org/10.1002/pro.510.

(9) Macdonald, C. B.; Stockbridge, R. B. A Topologically Diverse Family of Fluoride Channels. *Current Opinion in Structural Biology* **2017**, *45*, 142–149. https://doi.org/10.1016/j.sbi.2017.04.003.

(10) Last, N. B.; Kolmakova-Partensky, L.; Shane, T.; Miller, C. Mechanistic Signs of Double-Barreled Structure in a Fluoride Ion Channel. *eLife* **2016**, *5*, e18767. https://doi.org/10.7554/eLife.18767.

(11) Gupta, K.; Varadarajan, R. Insights into Protein Structure, Stability and Function from Saturation Mutagenesis. *Current Opinion in Structural Biology* **2018**, *50*, 117–125. https://doi.org/10.1016/j.sbi.2018.02.006.

(12) Wrenbeck, E. E.; Klesmith, J. R.; Stapleton, J. A.; Adeniran, A.; Tyo, K. E. J.; Whitehead, T. A. Plasmid-Based One-Pot Saturation Mutagenesis. *Nat Methods* **2016**, *13* (11), 928–930. https://doi.org/10.1038/nmeth.4029.

(13) Medina-Cucurella, A. V.; Steiner, P. J.; Faber, M. S.; Beltrán, J.; Borelli, A. N.; Kirby, M. B.; Cutler, S. R.; Whitehead, T. A. User-Defined Single Pot Mutagenesis Using Unamplified Oligo Pools. *Protein Engineering, Design and Selection* **2019**, *32* (1), 41–45. https://doi.org/10.1093/protein/gzz013.

(14) Wrenbeck, E. E.; Faber, M. S.; Whitehead, T. A. Deep Sequencing Methods for Protein Engineering and Design. *Current Opinion in Structural Biology* **2017**, *45*, 36–44. https://doi.org/10.1016/j.sbi.2016.11.001.

(15) Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. Construction of Escherichia Coli K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection. *Mol Syst Biol* **2006**, *2*, 2006.0008. https://doi.org/10.1038/msb4100050.

(16) Green, M. R.; Sambrook, J.; Sambrook, J. *Molecular Cloning: A Laboratory Manual*, 4th ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y, 2012.

(17) El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S. R.; Luciani, A.; Potter, S. C.; Qureshi, M.; Richardson, L. J.; Salazar, G. A.; Smart, A.; Sonnhammer, E. L. L.; Hirsh, L.; Paladin, L.; Piovesan, D.; Tosatto, S. C. E.; Finn, R. D. The Pfam Protein Families Database in 2019. *Nucleic Acids Res* **2019**, *47* (D1), D427–D432. https://doi.org/10.1093/nar/gky995.

(18) Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the Next-Generation Sequencing Data. *Bioinformatics* **2012**, *28* (23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

(19) Katoh, K.; Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **2013**, *30* (4), 772–780. https://doi.org/10.1093/molbev/mst010.

(20) von Heijne, G. Membrane Protein Structure Prediction. *Journal of Molecular Biology* **1992**, *225* (2), 487–494. https://doi.org/10.1016/0022-2836(92)90934-C.

(21) Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* **2010**, *59* (3), 307–321. https://doi.org/10.1093/sysbio/syq010.

CHAPTER 7

Conclusions and Future Directions

7.1 Conclusions

Through these studies of the Fluc and SMR families of membrane proteins, some of the unique features of membrane protein evolution have been elaborated. While the Flucs have conserved a function while varying their structural complexity, the SMRs have adopted new functions while largely maintaining their assembly. In each case, the combination of structural and functional constraints has shaped the nature of the current family. The goal of this thesis has been to understand the mechanisms through which these constraints are effected, and how inferences can be generalized to membrane proteins more broadly.

In chapter 2, the evolution of bacterial Fluc oligomeric state and function was examined to understand how gene duplication operated in this family. The phylogenetic tree suggested that gene duplications took place multiple times independently across the tree, and that reversions to a homodimeric state from a heterodimeric state are exceedingly rare. Linked to this structural change is a functional one, where one of the two redundant pores become degraded. Importantly, the tree cannot tell which process came first: whether, post-duplication, drift degraded one pore before the orientations became established, or whether orientation preceded functional drift.

After this initial study of the Flucs, a second family of dual-topology proteins, the SMRs, were studied in chapter 3. Here, the focus was on functional differentiation rather

than structural. Spurred on by bioinformatic and preliminary biochemical evidence, a detailed characterization of a previously unstudied subgroup of the SMRs was pursued. This revealed that this subgroup, called the Gdxs after, is at least as numerous as the multidrug-exporting subtype, called the Qacs. Further, in vitro reconstituted assays revealed that these are very specific guanidinium/proton antiporters, with an electrogenic 1 guanidinium to 2 proton stoichiometry.

In chapter 4, additional investigation using more flexible assays was undertaken to understand the nature of the functional split between subtypes. Bioinformatic and sequence analysis firmed the understanding of the subfamily architecture and distribution. Solid-supported membrane electrophysiology of Gdx and Qac subtypes surprisingly revealed that, while "canonical" substrate specificity was strictly split between the two, a spectrum of singly- and multiply-substituted guanidinium compounds were transported by both subtypes. Intriguingly, both subtypes failed to transport arginine or agmatine, important cellular metabolites with guanidinyl moities. These results suggested that guanidinium specificity in the Gdx subtype is driven by H-bonding recognition, whereas Qac subtypes recognize simpler physicochemical properties. The structure determination of a Gdx homolog rationalized our findings by revealing the structural determinants of binding and recognition in the Gdx subtype. The aromatic binding pocket is structured by a number of essential conserved H-bonds, which orient the glutamates for proper recognition. Ablation of these diminished or removed function in Gdx. A "portal" open to the lipid interior is speculated to play a role in exclusion of arginine and agmatine transport via the energetic cost of partitioning their polar ends into the membrane. This may have also provided a ready mechanism for the development of drug resistance, with hydrophobic compounds potentially being able to partition into the binding pocket directly from the lipid bilayer. The combined structural and functional investigations emphasize the importance of promiscuous functions for evolving new functions in changed environments.

In chapter 5, this analysis was extended to the E. coli Gdx homolog, and comparison

with a homology model was made. This suggested that the important residues for function are likely conserved broadly across the family, and our inferences in chapter 4 are broadly applicable. The functional characterization also demonstrated that the E. coli homolog is similar in substrate specificity, lending further credence to the importance of our substrate screen in chapter 4.

In chapter 6, the Flucs were used as a model system now to understand how gene duplication reshapes fitness landscapes with an eye towards understanding how duplicates are retained. A system was developed that can specifically differentiate between homodimeric and heterodimeric functionality, and a mutagenic library of single-site mutants was created using nicking mutagenesis. This will be subject to selection in multiple genetic contexts, and then deep sequencing will provide fitness effects of mutations. This DMS approach will provide the first experimental evidence of the fitness effects of gene duplication.

7.2 Future directions and outlooks

7.2.1 The SMRs

The SMRs are a natural model system for understanding functional evolution, and the present work leads naturally to asking what the sequence-level distinctions between the Qac and Gdx subtypes are. Understanding this, in combination with a resolved phylogeny, may allow the reconstruction of the evolutionary pathway that produced the two subtypes. This may also allow one to engineer alternative specificities between subtypes and combinatorially measure their phenotypes. In short, given the wealth of sequence and structural evidence present, a concrete definition of subfamily differentiation may be possible.

The SMRs are also an interesting potential system for understanding how paralog interference may be escaped in the membrane. Despite having 6 SMR homologs, E. coli cells do not appear to accumulate Qac/Gdx heterodimers. While sterics may appear to explain

this somewhat, a fundamental understanding of helix-helix interactions in membrane that could firmly explain this is lacking[1]. Ancestral reconstruction or experimental evolution may be useful methods to examine how helix-helix interaction specificity changed over time and was used to escape paralog interferences in the SMRs.

7.2.2 The Flucs

The present work focuses on the transition from homodimeric dual-topology proteins to heterodimeric oriented ones. The fusion process, where inverted-repeats are generated, is equally important for shaping extant transporter diversity, but is similarly unknown. The Flucs again are an excellent model system to understand this process, as they also exist in fused states with a conserved function. Phylogenetics presents new challenges here, but analysis of the linker helices and functional drift in the pores may provide novel insights. Focused DMS approaches in a linked construct could provide analogous information into how the fitness landscape changes when fusion occurs, and structure determination could suggest how the broken symmetry allows for more flexibly helix-helix interactions.

The Flucs are also a useful system to understand paralog interference and its escape. Based on their phylogeny, structural and functional drift occur together, but it is unclear which comes first. Each allows heteromers to escape paralog interactions, but via different means. A more detailed phylogenetic reconstruction may provide hints of what actually came first; alternatively, experimental evolution or DMS may provide insight into sequence-specific trajectories to alter oligomerization.

References

(1) DeGrado, W. F. How Do Helix-Helix Interactions Help Determine the Folds of Membrane Proteins? Perspectives from the Study of Homo-Oligomeric Helical Bundles. *Protein Science* **2003**, *12* (4), 647–665. https://doi.org/10.1110/ps.0236503.

APPENDIX A

Supporting Information for Chapter 3

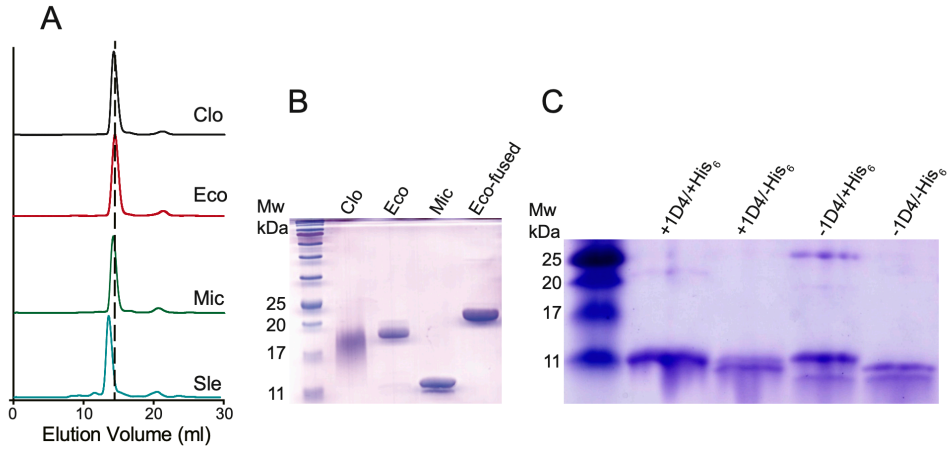

Figure A.1: SDS/PAGE gel of purified Gdx homologs. (A) Size exclusion chromatography elution profiles of Gdx proteins. Dashed line indicates the elution volume of a 30 kDa membrane protein standard, Fluc-Bpe. (B) Coomassie-stained SDS/PAGE of Gdx homologs used in this study. (C) Coomassie-stained SDS/PAGE of Sle-het. With affinity tags, the subunits run at the same molecular mass on the gel. Complete cleavage of tags individually or together reveals two bands in purified Sle-het protein.

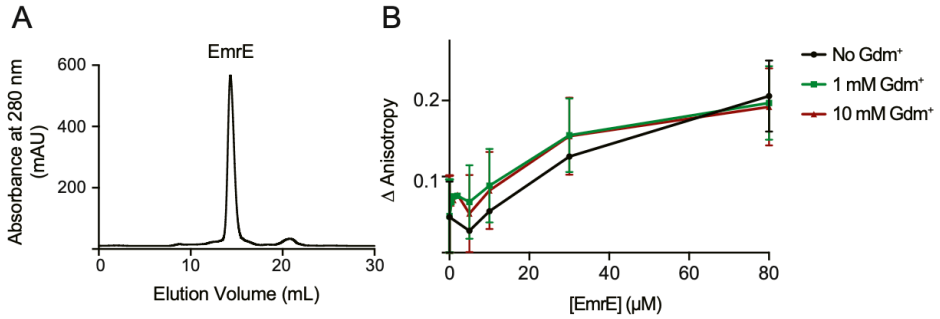

Figure A.2: Purification and substrate binding assays for EmrE. (A) Superdex 200 elution profile of EmrE. (B) Ethidium fluorescence anisotropy binding assay in the presence and absence of guanidinium ion. These results demonstrate that high amounts of Gdm+ do not compete with ethidium binding in EmrE.

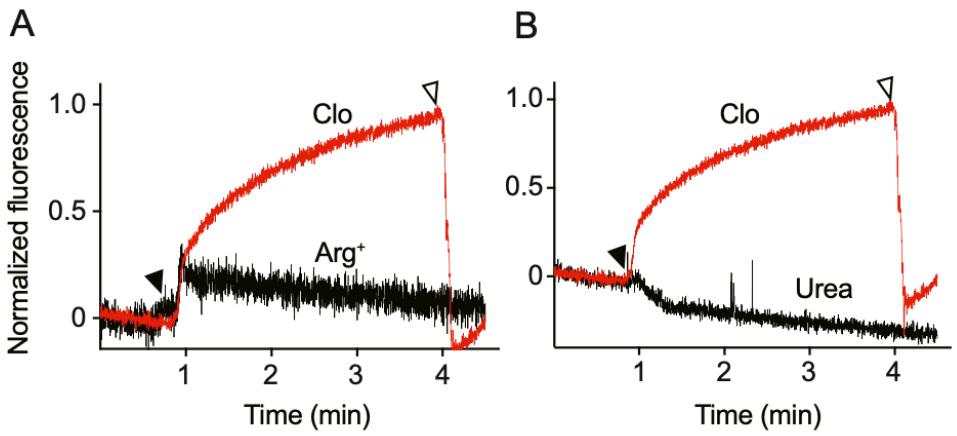

Figure A.3: Proton uptake into liposomes monitored by pyranine fluorescence. As in figure 3.3, test substrates were added to Gdx-Clo proteoliposomes containing pyranine at the closed triangle. Gdm+ addition is shown in red; test substrates arginine (A) and urea (B) are shown in black. After 3 min, FCCP was added to collapse the pH gradient (open arrow).

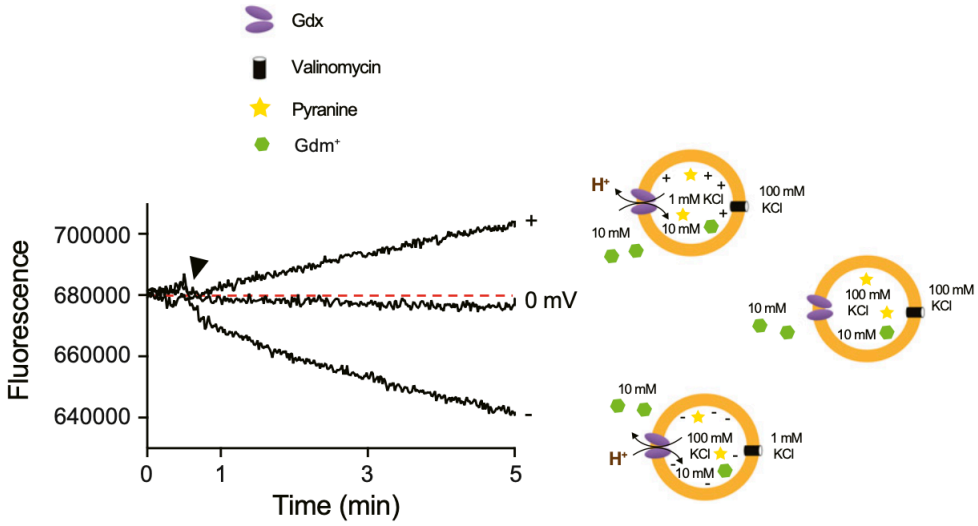

Figure A.4: Electrogenic proton transport by Gdx-Clo. Experiments were performed as in figure 3.5 A.

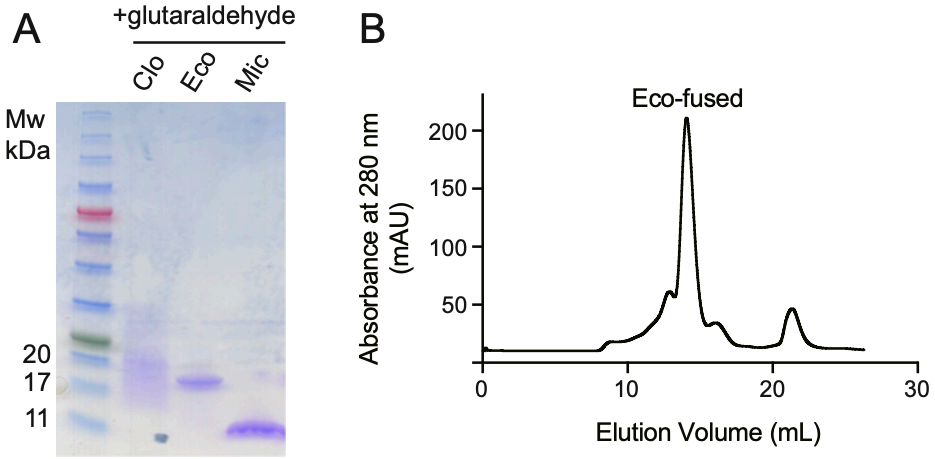

Figure A.5: Oligomerization assays. (A) Glutaraldehyde cross-linking of homodimeric Gdx proteins. All samples are treated with 0.125% glutaraldehyde for 1 h but do not form higher order oligomers and run at the same molecular mass as shown in figure A.1. (B) Superdex 200 elution profile of Eco concatamer, linked by glycophorin A TM helix.

APPENDIX B

Supporting Information for Chapter 4

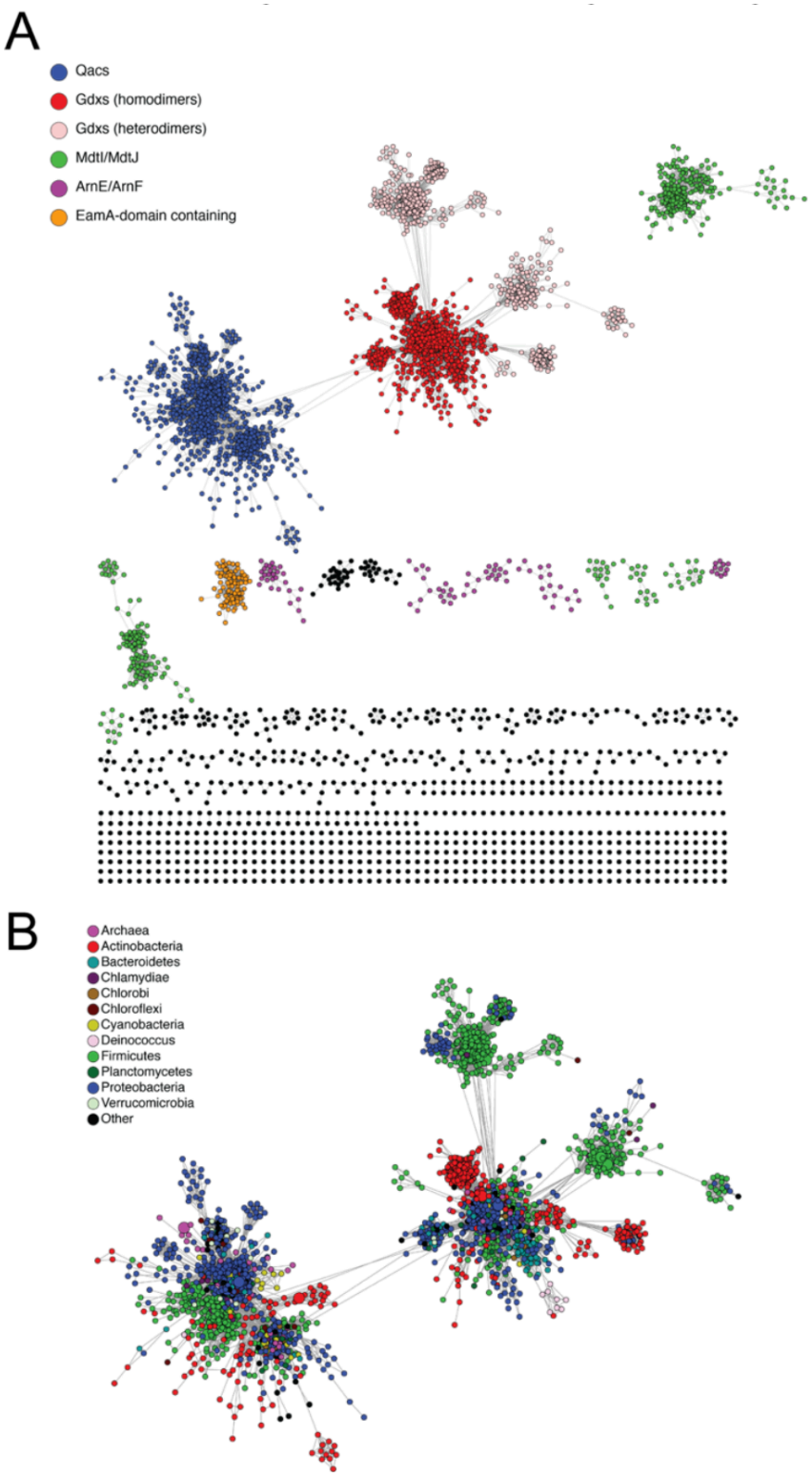

Figure B.1: SMR sequence similarity networks with additional annotation. A. Full sequence similarity network with gene annotation of each cluster, where known. B. Distribution of SMR proteins in archaea and different bacterial taxa. Sequence similarity network as in figure 4.1 A with coloring according to taxa as indicated.

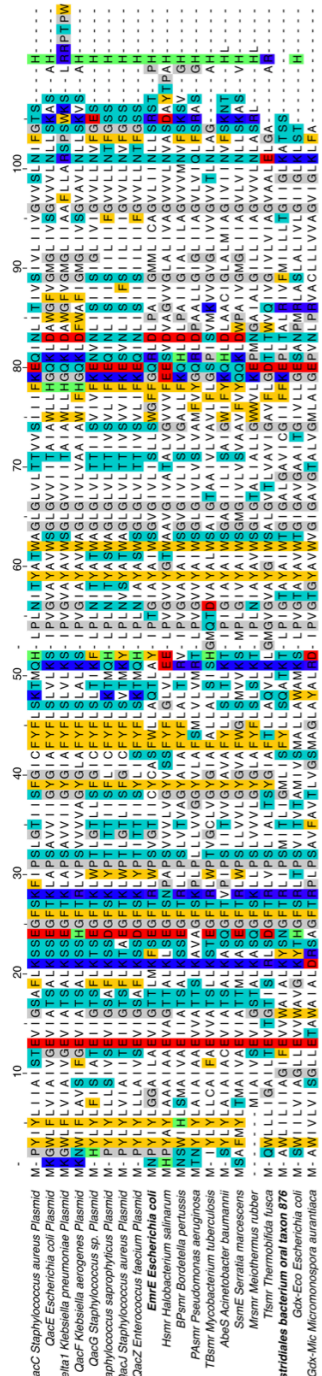

Figure B.2: Sequence alignment of functionally characterized homodimeric SMRs

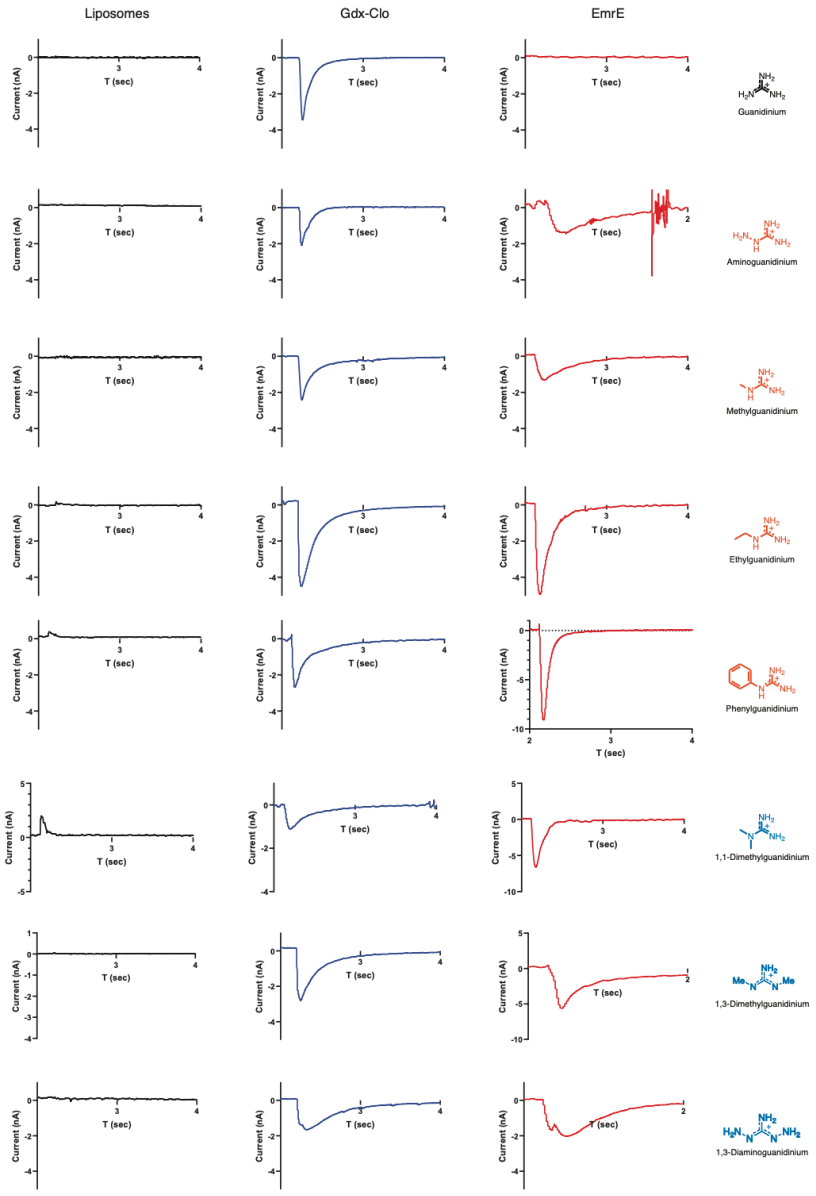

Figure B.3: Representative SSM electrophysiology recordings. Currents elicited after perfusion with substrate are shown for empty liposomes (left; black), Gdx-Clo proteoliposomes (middle; blue) and EmrE proteoliposomes (right; red).

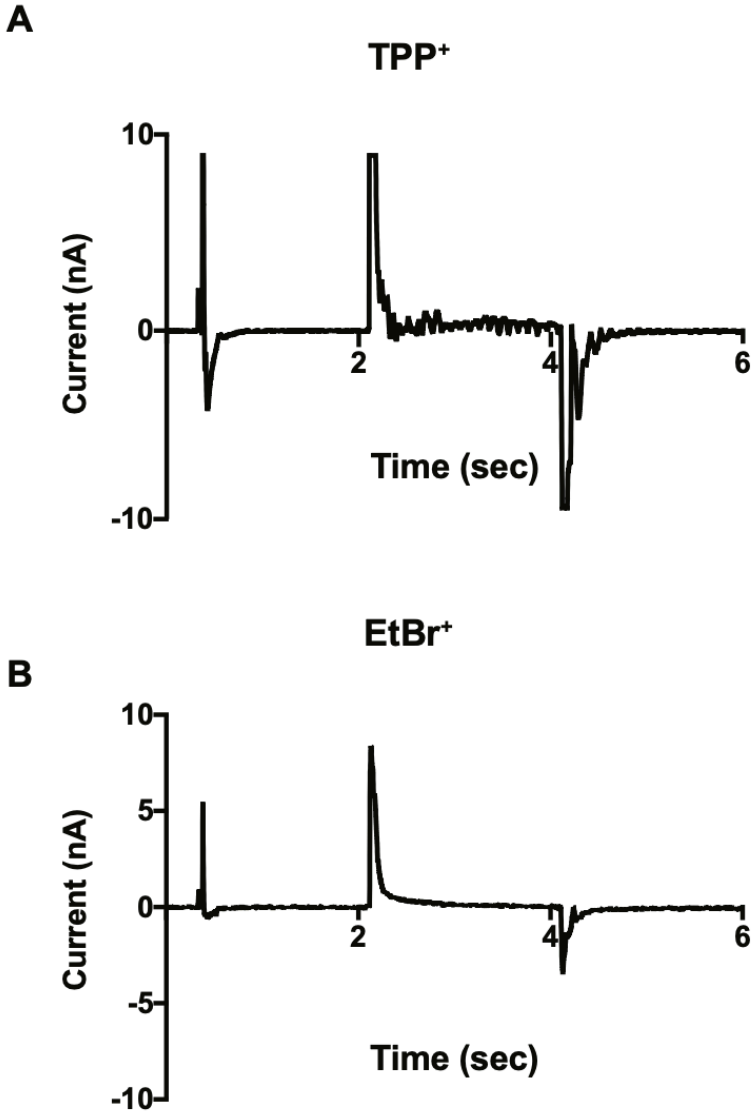

Figure B.4: SSM traces for tetraphylphosphonium (TPP+, top) and ethidium (EtBr+) perfusion of liposomes that do not contain protein

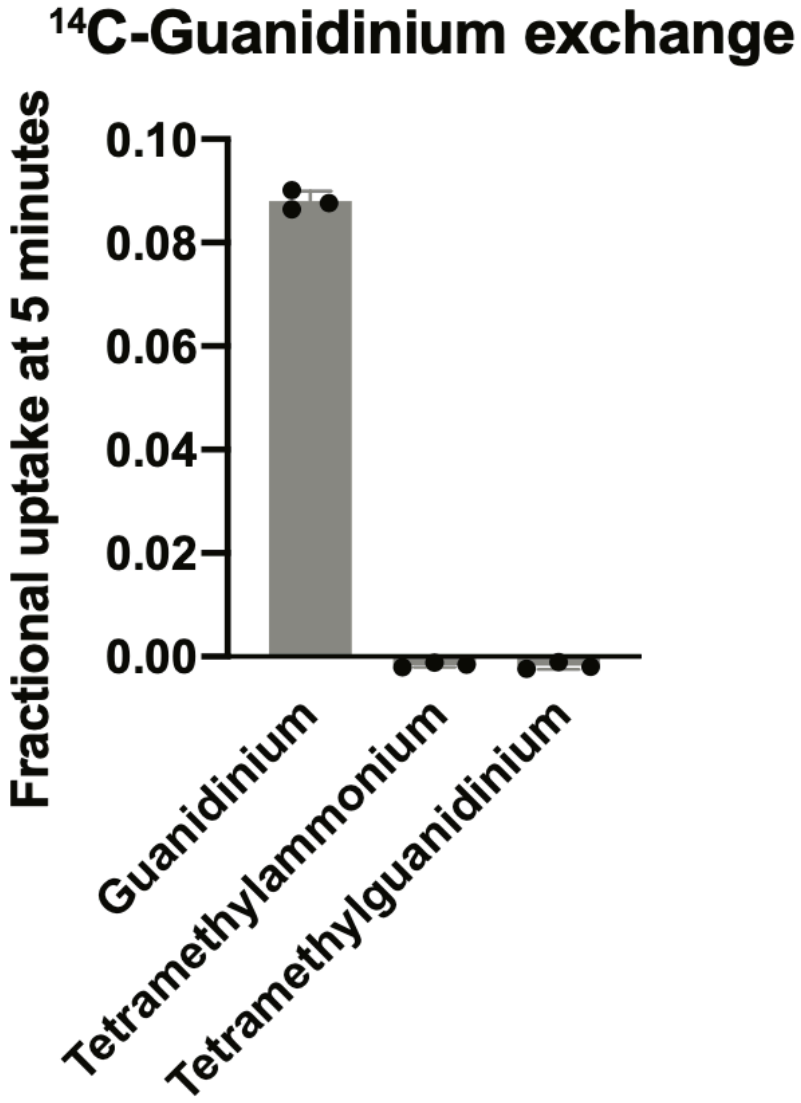

Figure B.5: Uptake of 14C Gdm+ into Gdx-Clo proteoliposomes in exchange for the indicated substrate. Experiment performed as in main text Figure 2, with fractional uptake measured relative to total radioactive counts in reaction mixture. Error bars represent the mean and SEM of three technical replicates

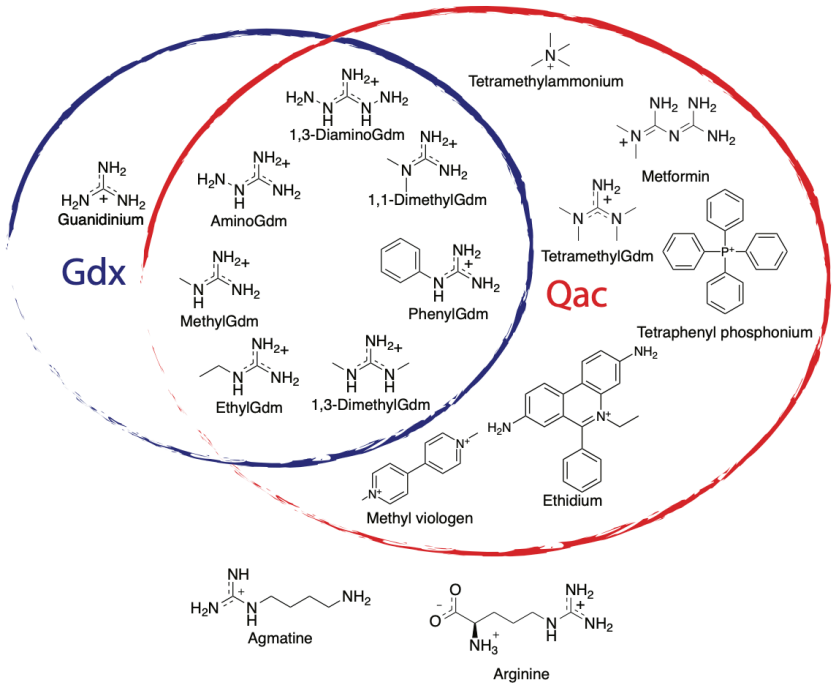

Figure B.6: Venn diagram showing overlapping transport specificities of Gdx-Clo and EmrE.

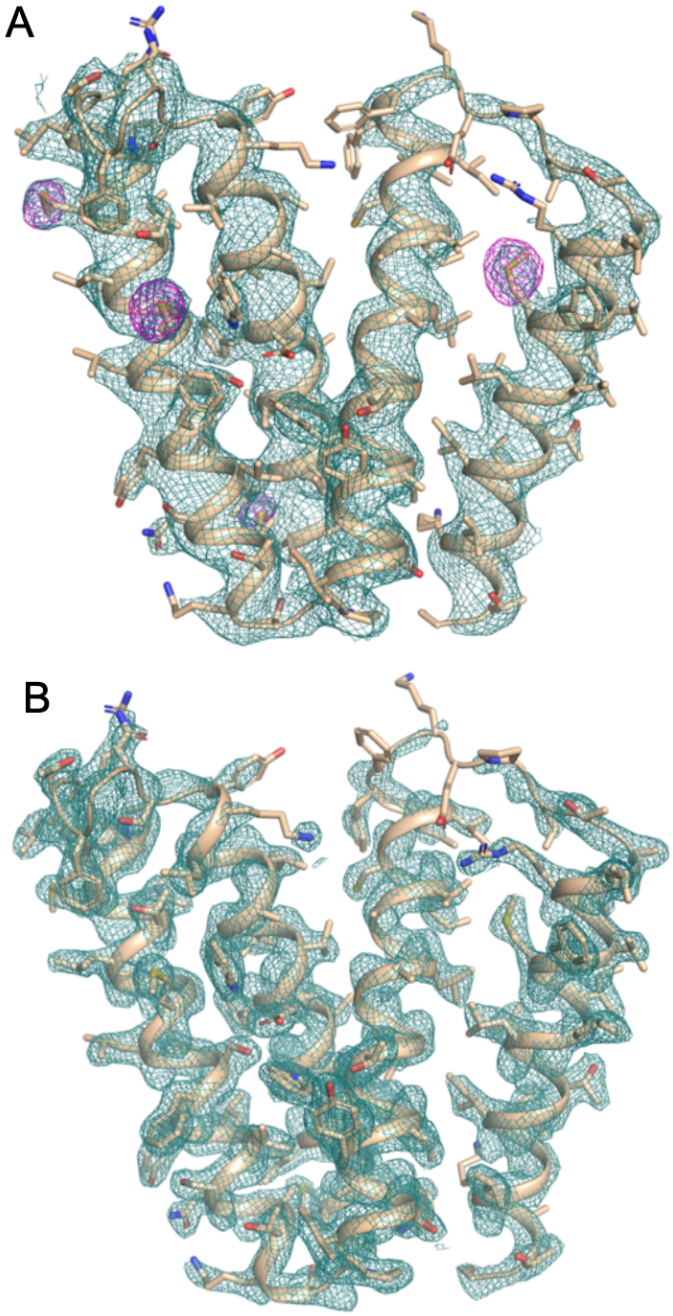

Figure B.7: Experimental electron density maps for Gdx-Clo. A. Cartoon view of one subunit from Gdx-Clo, with the solvent-flattened electron density map calculated from SHARP contoured at 1.3σ (teal), and anomalous difference density from selenomethionine contoured at 5σ (magenta). B. Electron density map for one subunit of Gdx-Clo (octylGdm+-bound structure) contoured at 1.8σ.

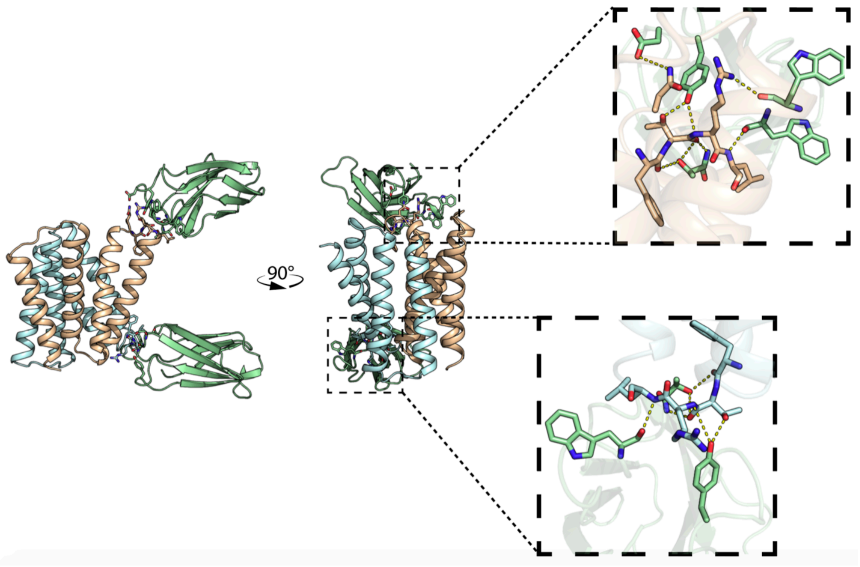

Figure B.8: Binding interface between monobody Clo-L10 and Gdx-Clo. Gdx-Clo shown in tan and cyan; monobody in green. Residues within H-bonding distance are shown as sticks, with H-bond interactions shown as dashed lines.

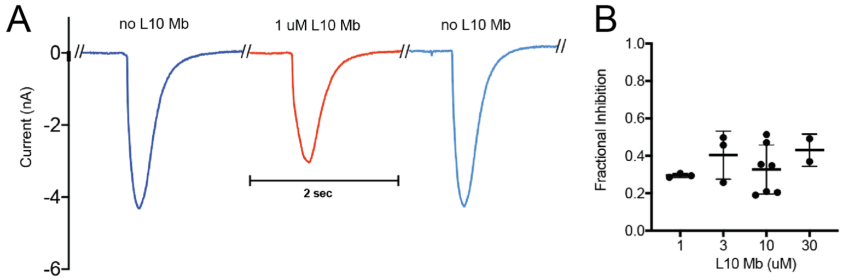

Figure B.9: Currents mediated by Gdx-Clo in the presence and absence of monobody L10. A. Currents elicited by perfusion with 1 mM Gdm+. For these experiments, benchmark current values for Gdx-Clo proteoliposomes were measured first (dark blue trace). Second, Gdm+ was removed by perfusion and L10 monobody 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.

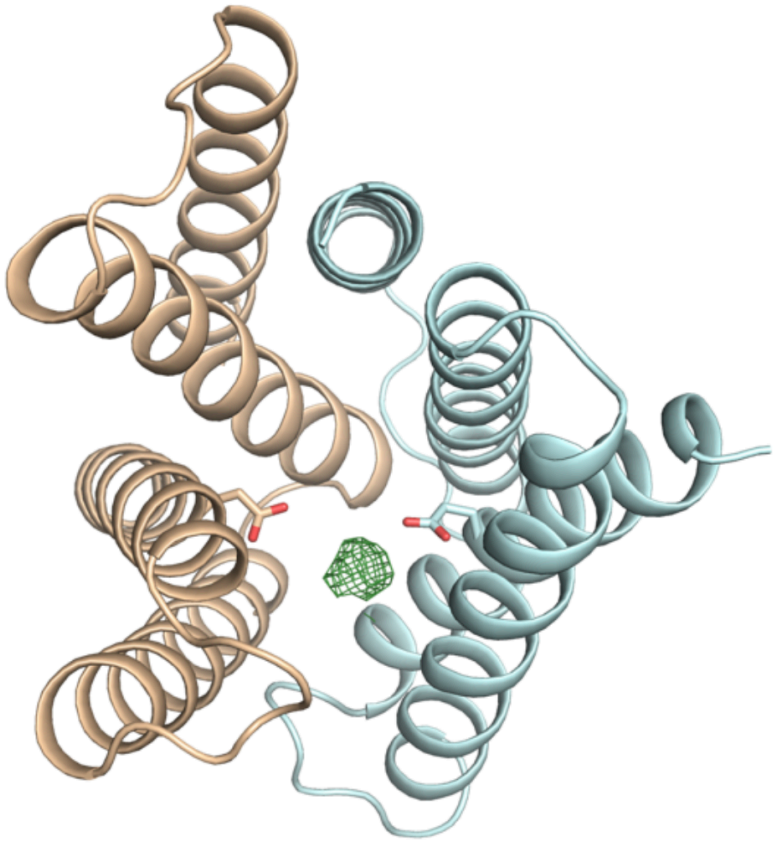

Figure B.10: Electron density between E13 and E13′ in 3.2 Å structure solved with 10 mM Gdm+. Top-down view of Clo-Gdx with subunits colored tan and light blue and E13 sidechains shown as sticks. Fo-Fc map in the region of the E13 sidechains is contoured at 3.2σ.

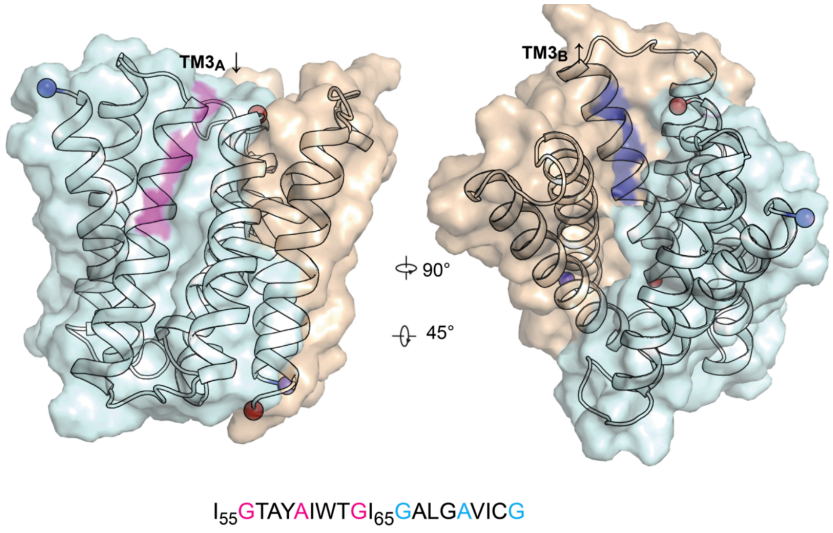

Figure B.11: Surface rendering of exposed TM3 GxxxG motifs. Left, surface view of GdxClo viewed through the plane of the membrane. Right, view is rotated 90° and tilted to view the interior of the aqueous-exposed vestibule. Coloring of surface rendering corresponds to TM3 sequence shown below. The first (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.

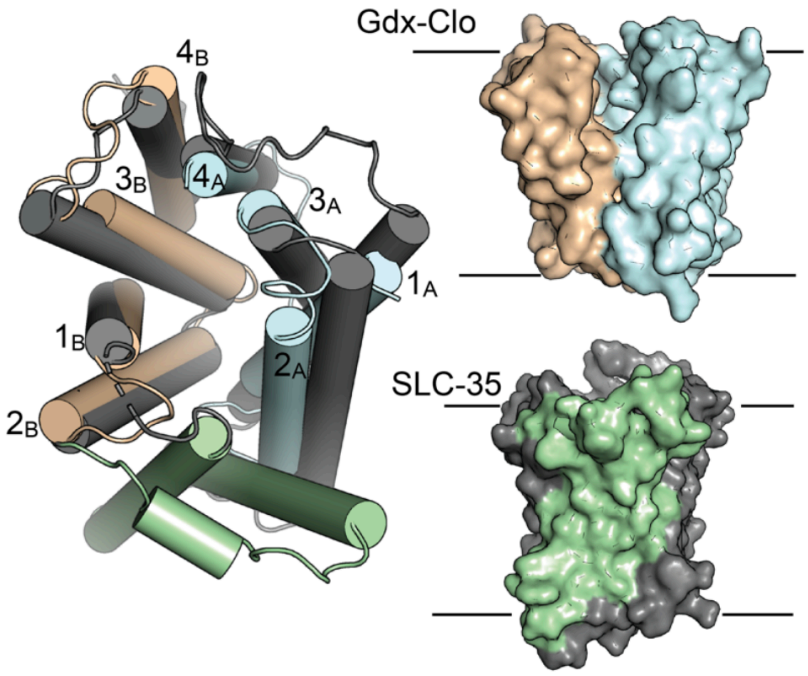

Figure B.12: Structural alignment of Gdx-Clo (subunits in tan and light blue as in main text) and CMP-sialic acid transporter from the SLC35 family (PDB: 6I1R; shown in dark gray with helix insertions in green). SLC35 proteins that share this fold have been structurally characterized, including 15-17. 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			
a, *b*, *c* (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)
R_{merge}[a]	0.06 (0.15)	0.132 (0.65)	0.074 (0.38)
R_{meas}[a]	--	0.145 (0.72)	0.086 (0.44)
Mn I / σI[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)
R_{cullis}	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
R_{work} / R_{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})))

Figure B.13: Data collection, phasing and refinement statistics for Gdx-Clo complexes

APPENDIX C

Supporting Information for Chapter 5

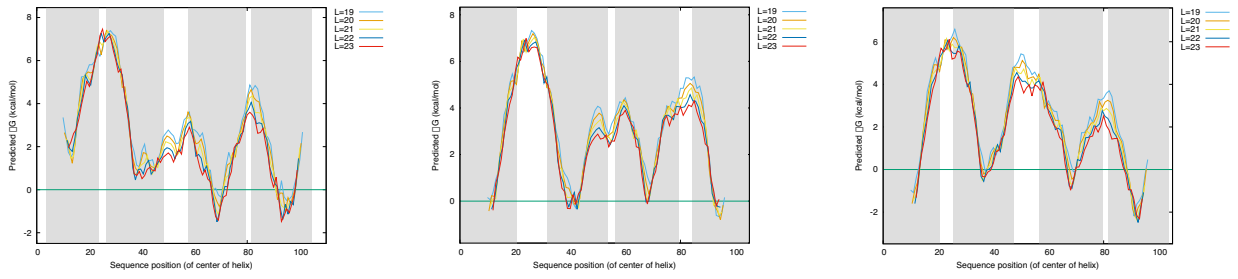

Figure C.1: Predicted ΔG_{app} of insertion for SMR homologs from ΔG predictor webserver.