Spring 5-15-2016

Structure and Dynamics of a Small Multidrug Resistance Transporter, EmrE

Chao Wu
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Computational and Molecular Biophysics

Dissertation Examination Committee:
Katherine A. Henzler-Wildman, Chair
Gaya K. Amarasinghe
Alexander B. Barnes
Kathleen B. Hall
Peng Yuan

Structure and Dynamics of a Small Multidrug Resistance Transporter, EmrE

by

Chao Wu

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2016
St. Louis, Missouri

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Acknowledgments

I would like to thank my advisor Katie for the strong support and trust I got throughout my graduate studies. I want to thank her for the rather hands-off yet when needed really approachable and inspirational mentoring style. While constantly amazed by her broad scientific expertise and vision, through interactions with her and sometimes just watching, I think I have benefited much more learning how to be a better person. I will forever be grateful for all of these.

I want to thank my thesis committee for the support and guide. I would like to thank my committee chair, Dr. Gaya Amarasinghe, for the constant mentoring throughout the years. Also I would like to thank Dr. Kathleen Hall, Dr. Xander Barnes, and Dr. Peng Yuan for caring and supporting my training.

I have been fortunate to learn from all past and present members of the Henzler-Wildman lab. Dr. Greg DeKoster was instrumental to help me get started with NMR. Dr. Supratik Dutta and Emma Morrison helped me become familiar with experimental techniques. Josh Brettmann and Anne Robinson are great lab mates who offer help on a daily basis. Also I would like to thank my colleagues in Madison. Dr. Marco Tonelli has always been great for pulse sequence development and sharing and data collection. Drs. Eva-Maria Uhlemann and Claudia Cornlilescu have helped a lot with the projects.

I would like to thank the community in the department of Biochemistry and Molecular Biophysics for the great training environment. Especially, I want to thank the Lohman and Hall

lab for equipment sharing and being great neighbors. Also I want to thank Dr. Changguo Tang for magnets maintenance.

I also need to thank my previous advisor, Prof. Niu Huang from National Institute of Biological Sciences, Beijing, for steering me into scientific research.

Last for not the least, I must thank my wife, Jiajia Li. Her support and encouragement were essential for me to get through difficult days. And special thanks to our parents for their support and allowing us to pursue the PhD degree half a world away.

Chao Wu

Washington University in St. Louis

March 2016

Dedicated to my parents.

ABSTRACT OF THE DISSERTATION

Structure and Dynamics of a Small Multidrug Resistance Transporter, EmrE

by

Chao Wu

Doctor of Philosophy in Biology and Biomedical Sciences

Computational and Molecular Biophysics

Washington University in St. Louis, 2016

Professor Katherine A. Henzler-Wildman, Chair

EmrE is a small multidrug resistance transporter in *E. coli*. It effluxes a wide range of antibiotics, thus contributing to the evolving epidemic of drug resistance. Despite its small size, EmrE is a fully functional transporter making it an ideal model system for a comprehensive study of the multidrug transport mechanism. In the transport cycle, EmrE must alternate between outward- and inward-facing conformations upon substrate binding to translocate substrates across the membrane. High-resolution structures of EmrE in complex with substrates facing different sides of the membrane will shed light on the coupling mechanism between substrate binding and transport. However, the conformational plasticity that enables EmrE to transport diverse drugs also makes it a very challenging system for high-resolution structural studies. The conformational dynamics inherent in the transport process require experimental measures of structural transitions to provide the link between static structures and functional transport. This thesis aims to characterize the structure and dynamics of EmrE in atomic detail using NMR, a well-established technique to study structure and dynamics of biomolecules simultaneously under a variety of conditions.

In the case of EmrE, NMR spectroscopy is the best approach for high-resolution structures because the dynamic nature and small size of EmrE hamper X-ray crystallography and cryoEM approaches. I have made significant progress towards a better structure of EmrE using a slow-dynamics mutant and have achieved a near complete backbone and side chain ILV methyl assignment for this highly challenging helical membrane protein system. I have also collected a large data set of distance and orientational restraints. I have also used NMR and functional assays to characterize a series of mutants located near the transmembrane helix 3 (TM3) kink and have demonstrated the important role of TM3 kink formation for the global conformational interconversion required for alternating-access. My NMR data also suggest that hydration within the transport pore may be an important property fine-tuning the rates of conformational interconversion. My NMR pH titrations show that the slow-dynamics mutant also has elevated pK_a values for E14, the critical residue for proton-coupling in EmrE. This provides the first experimental evidence of the physicochemical link between proton and substrate binding and alternating-access necessary for achieving coupled transport. By correlating high-resolution structural and dynamic data with functional transport assays, this thesis provides key insights into the multidrug transport mechanism of EmrE. The principles learned for EmrE set the stage for understanding even more challenging transporters.

Chapter 1: Introduction

"If we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of jigglings and wigglings of atoms."

--- Richard Feynman, The Feynman Lectures on Physics

One of the emerging principles in biochemistry and molecular biology is that the functions of biological molecules is ultimately rooted in the physical motions of atoms composing these molecules[1]. Understanding the nature of these motions at different time and length scales is deeply linked to comprehending molecular mechanisms. With the great success of structural biology in the past decades, we now have the luxury of researching biology at the atomic level. To move beyond structure-function relationships and construct the new paradigm: structure-dynamics-function, both spatial and temporal resolutions are essential. Despite recent exciting progress in biophysical techniques, e.g. spectroscopy and modeling, the field of biomolecular dynamics remains immature due to the limitations of current theories, the tremendous technical difficulties, and the challenging interpretations of dynamic information from measured physical properties.

In my doctoral thesis, I focus on a small multidrug resistance transporter, EmrE, to study its structure and dynamics by using NMR spectroscopy and other biophysical techniques in order to understand the multidrug transport mechanism. Hopefully the knowledge gained will ultimately be useful for combatting the antibiotic resistance problem.

Structure-dynamics-function: the new paradigm

Structural biology has played an instrumental role in modern biology. It has generated beautiful structures for many biomolecules, which have greatly advanced our understandings of cellular processes and facilitated the development of modern medicine[2]. By solving structures of biomolecules, one is provided the opportunity to explain sometimes puzzling phenomena and generate structure-based hypotheses to guide future experiments. This type of structure-function relationship study is widely acknowledged for its usefulness. The most famous case of using structure to gain mechanistic insight is the structure of DNA double helix[3] which immediately suggests the semiconservative replication mechanism[4] by which DNA is replicated in all known cells, one of the most fundamental questions in biology. Even today, it is common a new structure may provide novel insight into questions accumulated over 20 years of biochemistry.

However, life is in motion. These static structures can also be deceptive when they subtly suggest that these molecules are solid rocks, albeit with complicated texture. In fact, biological molecules are defined as soft materials in physics since they can be easily deformed by thermal energy at room temperature. These molecules are not static at all; they sample ensembles of conformations with only some captured in structural biology research. To truly understand the workings of biomolecules it is important to capture different conformational states. Although more difficult, understanding how biomolecules transition between different conformational states is also important since the different modes that a specific molecule samples are subjected to evolution along with the structures. With this concept in mind, biomolecules should better be seen as nano-machines with defined motional modes that can extract order from the environment.

There are already exciting studies on biomolecular dynamics, which signal the importance of this field. For example, systematic studies of the whole catalytic cycle of DHFR have identified key substates along with characterization of interconversion rates between different substates[5-7]. Armed with these beautiful structures, populations (thermodynamics), and rates of interconversion (kinetics), a movie can be constructed to watch biomolecules work in action! Another example where protein dynamics play a key role is allosteric signaling. Since proteins exist in an ensemble of conformations, large scale domain motions provide the physical basis for signal transduction through protein-protein interactions in the nanoscale[8]. In the case of RNA, which is known to be highly dynamic and whose dynamics are encoded by intrinsic properties of RNA structures and external environmental factors, it is not surprising to learn that RNA molecules can both carry genetic information and perform a wide variety of functions[9]. Even for DNA, they can sample many more conformations than simply being just straight double helices as generally assumed[10,11]. The emerging theme for biomolecules is that much of the functional complexity lies not only in the intricate details of the 3D structures but also in their ability to sample distinct conformations.

Transporters are inherently dynamic.

Most progress in the field of protein dynamics has occurred in studies of soluble proteins. Despite the key role of membrane proteins in controlling the information and material flow between cells and the environment and their importance as drug targets, membrane protein dynamics research is limited. This is in part because it is still technically challenging to work with membrane proteins. Due to the hydrophobic nature, membrane proteins need to be purified in a solubilizing environment, e.g. in detergent, which should maintain their native states. A

great deal of trial-and-error is still needed to screen for conditions that can generate useful research materials in the first place. Only in the last decade has structural biology of membrane proteins exploded. Exciting new structures of membrane receptors, ion channels, and transporters are being published, which is paving the way for dynamic studies to provide the next level of insights into understandings of mechanism and function of these important biomolecules.

Membrane transporters are an important family of membrane proteins that are inherently dynamic. As membranes are needed so life forms can separate self from environment, this necessitates biological machineries that reside in the membrane to relay signals and selectively import nutrients and export wastes. As a result, there are many membrane transporters in cells to move molecules in and out, and malfunctions of transporters are implicated in various diseases[12,13]. Membrane transporters, like enzymes, are inherently dynamic entities as required by their functions. In order to transport substrates across the membrane, transporters need to bind substrates and switch between at least two conformations, one outward-facing and one inward-facing. So in the case of transporters, the linkage between structure and dynamics is critical for proper function. They cannot be completely understood solely on the basis of static structures, even if these structures are of high resolution. Therefore, membrane transporters are an exciting new frontier to study structure-dynamics-function relationships.

Alternating access in transport

Initial attempts to understand transport processes proposed a "moving carrier" model[14]. In this model, carriers bind substrates and then move across the membrane before releasing the substrates on the other side. Now, we know that the moving carrier model may account for

transport processes carried out by ionophores like valinomycin. But for protein transporters which are integral membrane proteins, our current understandings of membrane and protein structures suggest that such a model would have energetic barriers that are prohibitively high.

Our mechanistic understandings of transport cycles now are defined by the alternating access model. In this model, transporters have at least two conformations, one outward-facing and one inward-facing, and alternate between these two conformations to expose the substrate binding site to one side of the membrane and then the other[15-18]. In the presence of transporters, the permeability barrier for substrates imposed by the membrane is no longer fixed. Instead, as the transporter alternates conformations, the permeability barrier moves shifting to the closed side of the transporter. As a result, the alternating access model was also called the "moving barrier" model in early literature. This model has been so successful to describe biochemical observations of transport processes that it dominates the field today.

The alternating access model is conceptually compelling and useful since diverse transport processes can be categorized based on minor modifications of the basic concept involved. Transport processes can be divided into active transport and passive transport (known as facilitated diffusion) based on whether an energy source is needed or not. In facilitated diffusion, no energy is consumed and thus the direction of transport is always down the concentration gradient of the transported substrate(s). Uniporters are involved in facilitated diffusion. To complete the transport cycle, uniporters will need to alternate the binding site from one side of membrane to the other both in the presence and absence of substrates. One example of uniporters is the family of glucose transporters, where significant breakthroughs in structural understanding were recently achieved[19-21].

In active transport, substrates are transported against their concentration gradients using some source of energy. In primary active transport, as in the large family of ABC transporters, ATP molecules are consumed to fuel the conformational changes required for the alternating access and to drive substrates against their concentration gradients. Thus metabolic energy stored in ATP is transformed into the chemical potential of the transported substrates. Secondary active transport, also known as cotransport or coupled transport, instead relies on the energy stored in an electrochemical potential to drive transport. The "downhill" movement of one substrate, usually H^+ or Na^+, is used to fuel the "uphill" movement of another substrate. Secondary active transporters can be further classified into symporters and antiporters, depending on whether the transported substrates move in the same or opposite directions (Fig. 1.1). In symport, both substrates cross the membrane in the same direction. Mechanistically, symport is explained through alternating access of the substrate-binding site on the transporters only when either all coupled substrates are bound or when the binding site is empty. The alternating access of the binding site cannot happen when only one substrate is present; otherwise the cotransport would be decoupled. In contrast, in antiport, the substrates move across the membrane in opposite directions, and the coupling is traditionally explained by (i) requiring strict competition between the substrates for a single binding site so only one substrate can bind at a time and (ii) allowing alternating access only when one substrate is bound but not when the antiporter is in the *apo* state.

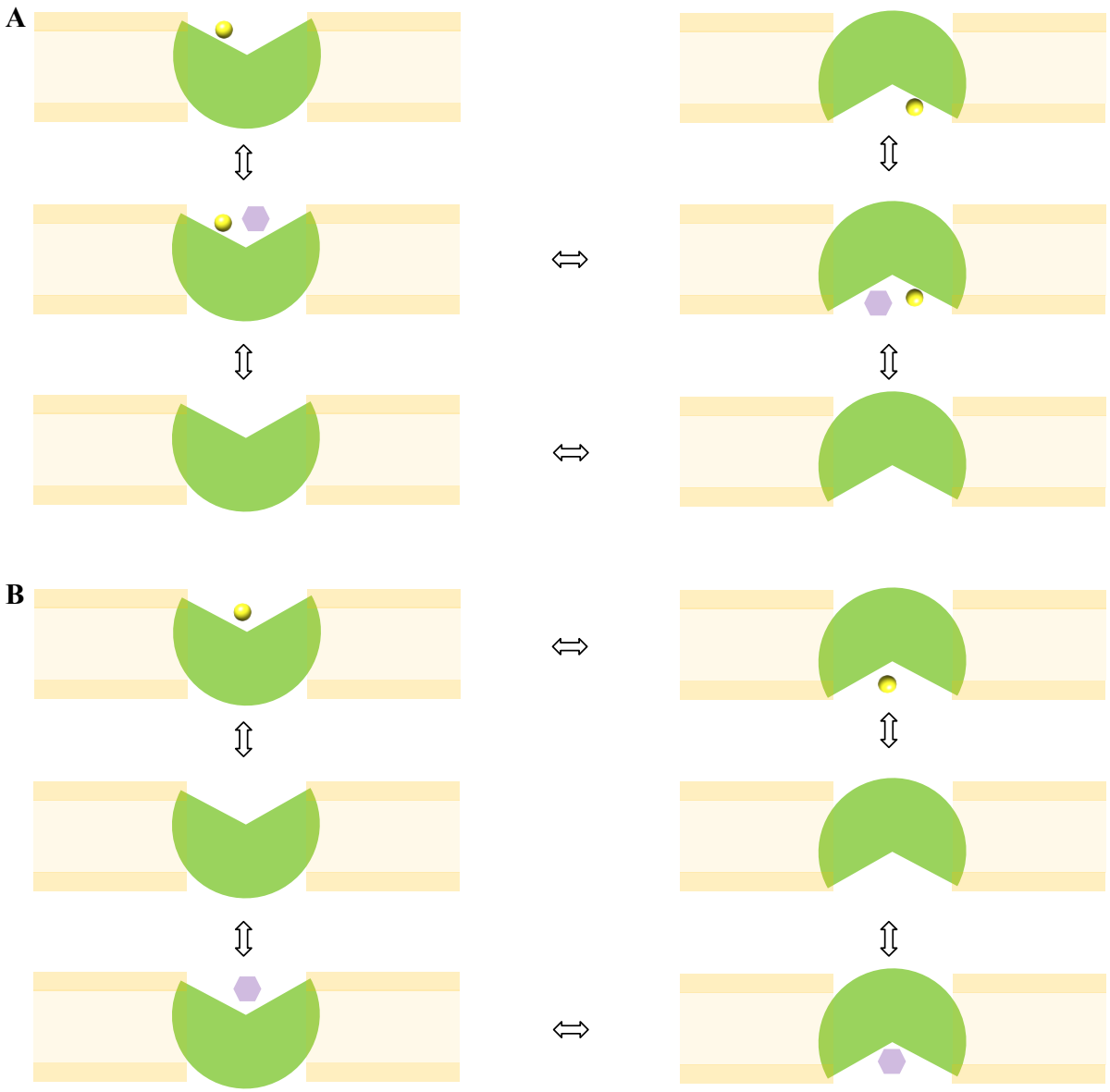


Fig. 1.1 Secondary active transporters can be classified into symporters (A) and antiporters (B). The transporters are represented in green and two different substrates are represented as a yellow sphere and a lavender hexagon.

In the alternating access model, the conformational interconversion, during which transporters switch between outward-facing and inward-facing conformations, is central to the transport cycle. It is at this step that transporters reorient substrate binding sites from facing one side of membrane to the other and substrates are translocated across the major permeability barrier.

Understanding how the binding of substrates to transporters is linked to transport is deeply related to understanding how the conformational interconversion occurs at an atomic level.

As a result of the large scale conformational changes involved in the conformational interconversion process, the movement of substrates across the membrane by transporters is much slower compared to channels. Ion movements through channels can be as fast as diffusion since channels can be viewed as holes in the membrane that have complex regulations of their opening and closing. The topological constraints set by transporters in active transport are necessary for them to transport substrates against their concentration gradients using an energy source. This could never be realized by a pore-like structure as in channels.

As discussed, the transport process is an inherently dynamic process. In this thesis, I focused on the conformational interconversion process in transport, in the hope of providing deeper understanding of the physical basis of the transport mechanism.

EmrE: a model system

EmrE is a small multidrug resistance transporter in *E. coli* that utilizes the proton-motive force across the inner membrane to efflux polyaromatic cationic antibiotics. Due to its small size, EmrE has been extensively studied as a model system for membrane protein folding and insertion[22], membrane protein topology[23,24], membrane protein evolution[25,26], and proton-coupled antiport[27]. In addition, it is implicated in bacterial antibiotic resistance and biofilm formation[28,29], making it potentially clinically relevant as well. Yet, after more than 20 years of studies using a variety of methods ranging from molecular biology to biophysics, our mechanistic understanding of EmrE is still rather immature[30].

One example of how this seemingly "simple" model system turns out to be rather complicated is illustrated in the controversy related to the structure and topology of EmrE[27,31]. The sequence of EmrE has 110 amino acids, and the hydropathy analysis of EmrE immediately (and correctly) suggested four transmembrane helices[32]. However, it took several lines of studies to establish that the functional unit of EmrE is an antiparallel dimer[33-37]. Looking back, several layers of complexity made the topology of EmrE controversial for a long time. In essence, all the major obstacles are a result of our rather primitive understandings about membrane proteins. The first major obstacle lies in the biochemistry involved in purifying EmrE. Earlier studies of EmrE used organic solvents which are now known to disrupt the native state of EmrE as manifested in a weaker affinity to TPP^{+}[38,39], a commonly used EmrE ligand. Also some bulky tags used for purification of recombinant EmrE turned out to affect the membrane insertion and topology[25,40], further complicating data interpretation. Finally, EmrE is among the earliest examples of dual topology, where a membrane protein with no particular charge bias in the loop regions can insert into membrane without preference of a particular orientation. This was totally unexpected given the understanding of membrane protein synthesis and insertion at the time and led to great skepticism of initial reports of antiparallel topology, despite their appeal due to homology with larger transporters[23-25,33,34,40].

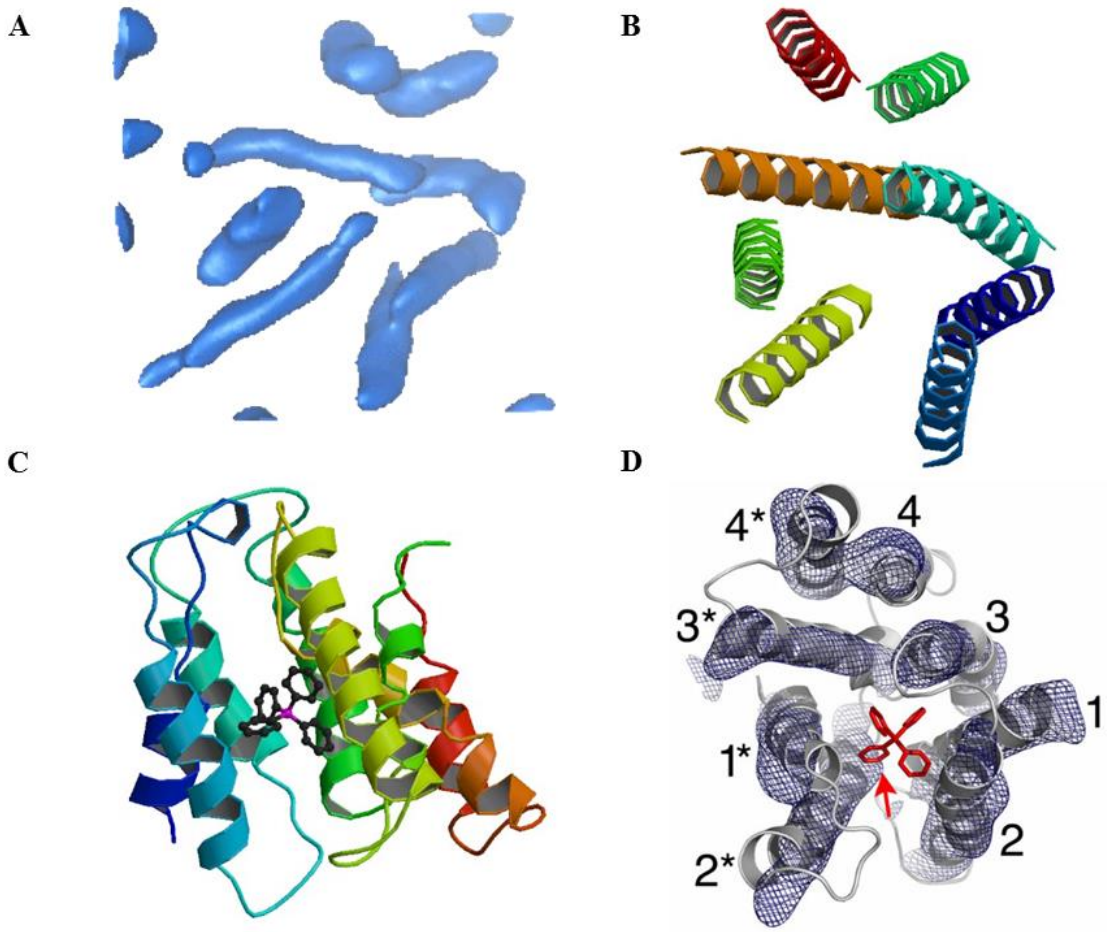


Fig. 1.2 Available experimental structures for TPP^+-bound EmrE. A. A cryoEM electron density map from 2D crystallography (7.5 Å resolution) at a contour level of 59.7. Image from EMDataBank (EMD-1087). EmE was purified in DDM and crystallized in DMPC bilayers. B. A model based on cryoEM map and evolutionary conservation. Image from PDB (2I68). TPP^+ is not shown. C. The X-ray crystal structure (PDB: 3B5D). Image from Chen *et al.*, 2007. EmrE was purified and crystallized in N-nonyl-β-D-glucoside (NG). To properly reflect the low-resolution nature, only Cα atoms were deposited. D. Ribbon representation of the EmrE-TPP^+ X-ray structure docked into the EM density map, contoured at 1.2 σ. The TM helices are labeled and the two monomers are distinguished by asterisks. Image from Chen *et al.*, 2007.

Structural evidence supporting the antiparallel dimer structure came from cryoEM[33,37,41] and X-ray[34] studies (Fig. 1.2), although both are of moderate resolution. The cryoEM structure of EmrE in complex with TPP^+ showed for the first time that the functional unit of EmrE is an asymmetric dimer with six transmembrane helices forming the ligand binding chamber[37]. This discovery was independently supported by X-ray structures from another group[34]. Follow-up studies from the

Tate group in collaboration with Nir Ben-Tal[33] considering the evolutionary conservation of each helix and the pseudo two-fold symmetry between the two monomers insightfully proposed that the antiparallel dimer could switch between outward- and inward-facing conformations if the two monomers simply swap conformations. This predicts the two structures of EmrE open to opposite sides of the membrane are structurally the same. Immediately, this model could reconcile the structural asymmetry in the dimer with functional symmetry suggested by bulk biochemical data. Further dynamics studies (Fig. 1.3) using high-resolution solution NMR, single-molecule FRET, and cross-linking studies lent overriding support for this elegant model[35]. Additional *in vivo* experiments confirmed the antiparallel dimer topology in the native *E. coli* environment[36,42], and the mechanism of antiparallel dimer formation has recently been established[22].

These structures are instrumental to establish the topology of EmrE. However, due to the moderate resolution and the dynamic nature of EmrE, only Cα atoms coordinates were deposited for both the cryoEM and X-ray structures. How is the TPP^+ ligand coordinated in the binding pocket? How could substrates binding lead to the conformational interconversion that lies at the heart of energetic coupling? In order to progress towards answering these mechanistic questions, higher resolution structures are needed.

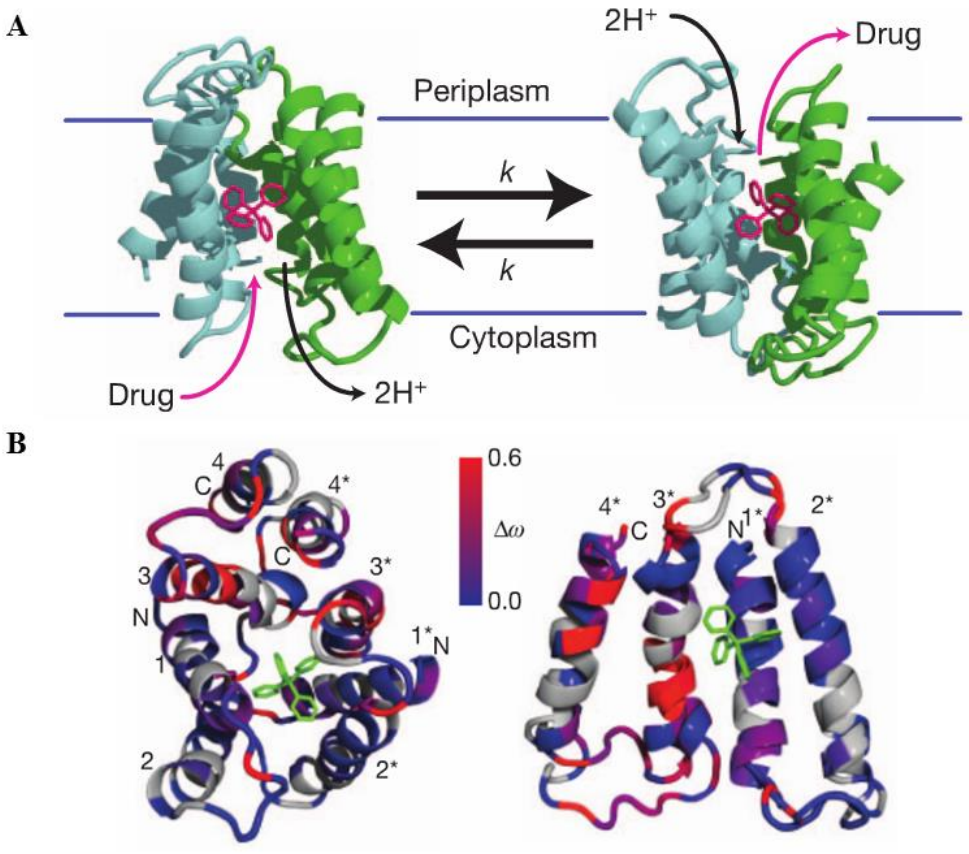


Fig. 1.3 Conformational interconversion in EmrE. A. Conformational interconversion for antiparallel, asymmetric EmrE: the two monomers swap conformations and the outward- and inward-facing states have identical structures except for which aqueous compartment is accessible to the transport pore. The two structures are related by a 180˚ rotation about an axis that is parallel to the bilayer and perpendicular to the page. The measured interconversion rate, k_{conf}, is ~5 s^{-1} at 45˚C. B. Chemical shift difference between the two monomer conformations plotted onto the antiparallel dimer model (left) or an overlay of the two monomers with transmembrane helices 1-3 aligned to compare the two monomers within the antiparallel dimer (right). Figure from Morrsion *et al.*, 2012.

NMR is an ideal tool for determining a better structure of EmrE.

High-resolution structures of transporters facing either side of the membrane provide a great deal of insights into the transport mechanism. These structures will directly reveal how substrates are coordinated in the binding pocket, and which interactions need to be disrupted in inward-facing conformation and then rebuilt in outward-facing conformation as the transporter interconverts

between states. Such high-resolution structures of transporters in different states are challenging to obtain. Our previous NMR studies revealed that drug-bound EmrE has the same structure in outward- and inward-facing conformations. So by determining the structure of drug-bound EmrE, we can get both outward- and inward-facing states. This structure will identify substrate coordinating interactions and shed light on the coupling mechanism between substrate binding and conformational exchange. This long-sought high-resolution structure of EmrE will serve a defining role in integrating all the available data to understand the mechanism and guide future experimental design in the field.

However, the conformational plasticity that enables EmrE to transport diverse drugs also makes it a very challenging system for high-resolution structural studies. This dynamic nature of EmrE and lack of soluble domains hamper the formation of well-packed crystals for X-ray and also prevents high-resolution structure determination by cryoEM. The same considerations lead to the failure of alternative methods for crystalizing membrane proteins by inserting soluble domains or using crystallization antibodies. Thus, NMR spectroscopy is the best approach for EmrE structure determination. Structure determination by NMR is one major area of research (and Kurt Wuthrich was awarded the 2002 Nobel chemistry prize for developing NMR techniques for structure determination of biomolecules). However, although NMR can accommodate some dynamics, it needs a relatively static structure for NOE measurement (long mixing time). Inspired by the success of thermal stabilization by point mutations in GPCR crystallography[43,44], we collaborated with Chris Tate at Cambridge to screen and select EmrE mutants with preserved ligand binding but loss of function with hope that these will reflect loss of dynamics.

NMR provides a dynamic description of the EmrE alternating access process.

As transport is inherently a dynamic process, a dynamic view is needed to provide a necessary link between static structures and functional transport. NMR spectroscopy is a well-established technique to study protein structure and dynamics simultaneously under a variety of conditions. In fact, the most unparalleled application of NMR is its power to study dynamics.

There are a collection of NMR dynamics experiments to study motions at different timescales (Fig. 1.4). The dynamic information provided by these experiments is derived from relaxation time measurements[45,46]. In NMR, relaxation means the return of a spin system to its equilibrium following a perturbation. The rate of return depends on the time-dependent magnetic fields the spin is experiencing. Since local magnetic field fluctuates as a result of molecular motions, the relaxation of spin systems can be linked to dynamic information.

Macromolecules typically tumble in the nanosecond timescale, so motions that are much faster or slower than tumbling time can be detected. Motions faster than nanosecond timescale are usually studied using ^{15}N T_1, ^{15}N T_2, and ^{1}H-^{15}N NOE experiments. Motions in the microsecond-to-millisecond timescale could be studied using $T_{1\rho}$ and CPMG experiments while motions in the millisecond-to-second timescale could be studied using ZZ-exchange, NOESY, and real-time NMR experiments[47].

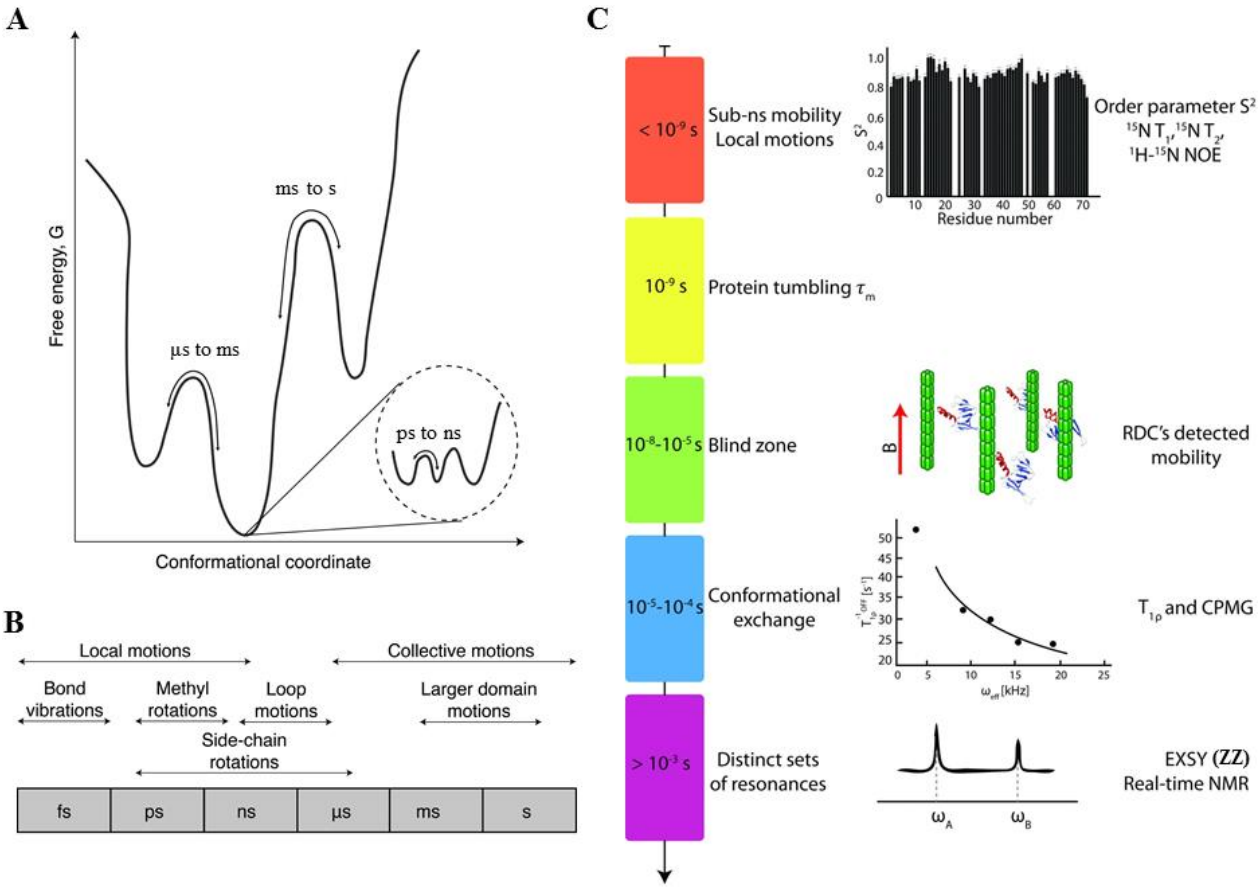


Fig. 1.4 Protein dynamics and NMR dynamics experiments. A. Free energy as a function of an arbitrary conformational coordinate. The free energy landscapes of proteins are very rugged with multiple minima. Fluctuations between different local minima have barriers of varying heights that translate into a wide range of timescales. B. Typical timescale ranges for different types of protein motions. C. NMR experiments that can be used to study dynamics on different timescales. Figure adapted from Ferella, *et al.* 2012.

Chemical shift is one very sensitive NMR observable that could report on dynamics of molecules. When a nucleus is exposed to different chemical environments as the molecule samples different conformations, chemical shifts of that nucleus in the resulting NMR spectrum could provide information about both population (thermodynamics) and exchange rates (kinetics). Based on comparisons between exchange rates and chemical shift differences, this is usually divided into slow-exchange regime, intermediate-exchange regime, and fast-exchange regime. The resulting spectra in different categories have different appearances (Fig. 1.5).

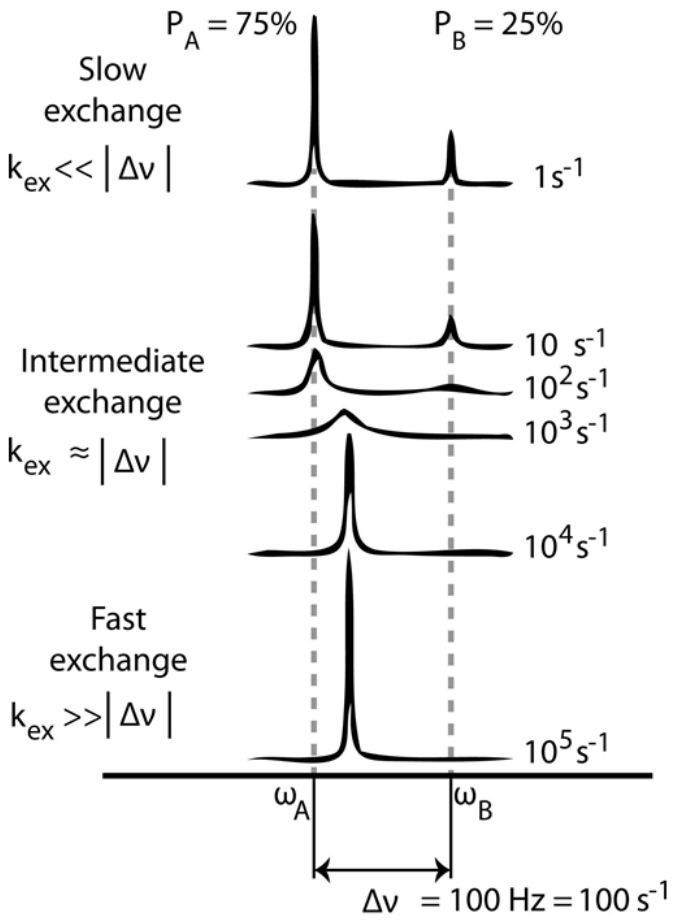


Fig. 1.5 The time regime of chemical exchange in NMR is defined by the relative magnitudes of k_{ex} and $\Delta\nu$. The effects of varying k_{ex} are demonstrated on the one-dimensional NMR spectra with P_A = 75%, P_B = 25%, and $\Delta\nu$ = 100 Hz. In the slow exchange regime ($k_{ex} << \Delta\nu = \Delta\omega_{AB}/2\pi$), two separate peaks are observed for the two states reflecting their distinct chemical shifts, intensities, and linewidths. In the fast exchange regime ($k_{ex} >> \Delta\nu = \Delta\omega_{AB}/2\pi$), only one signal is observed with population-weighted averages of chemical shift, intensities, and linewidths. At intermediate exchange regimes, the linewidth is increased due to exchange broadening. Figure from Ferella, *et al.* 2012.

Our dynamics studies of EmrE focus on the conformational interconversion process, which occurs in the millisecond-second timescale. So ZZ-exchange experiments are the main dynamics experiment used to characterize different mutants of EmrE and EmrE with different substrates. ZZ-exchange is used to characterize the kinetics of interconversion between different conformational states in the slow exchange regime[48]. It is based on multidimensional chemical shift correlation experiments with the addition of a mixing time between indirect and direct

dimension frequency detection. During the mixing time, the magnetization is aligned along the Z axis, parallel to the magnetic field. If conformational exchange happens during the mixing time, there will be additional cross-peaks from exchange in the spectrum. The kinetic exchange rates could be extracted by recording a series of ZZ-exchange experiments with varying mixing time. Homonuclear NOESY experiment could be seen as the simplest ZZ-exchange experiment, and exchange cross-peaks are also observed in NOESY spectra of WT EmrE[35].

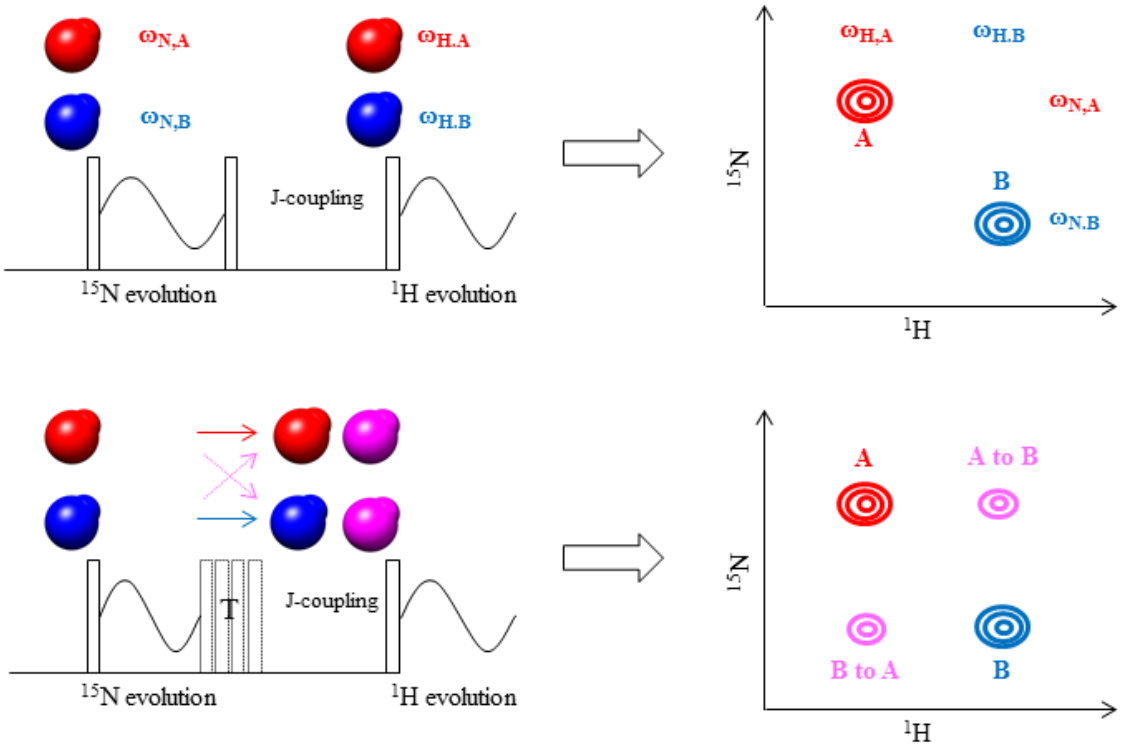


Fig. 1.7 A schematic representation of ZZ-exchange spectroscopy for a ^{1}H-^{15}N spin pair exchanging between two states. ZZ-exchange is useful for studying dynamics in the slow exchange regime where state A (red) and B (blue) give rise to separate peaks in NMR spectra. During the mixing time T, magnetization is aligned along the Z-axis, parallel to the static magnetic field. When T = 0, this experiment gives rise to two auto-peaks at ($\omega_{N,A}$, $\omega_{H,A}$) and ($\omega_{N,B}$, $\omega_{H,B}$) for the two states. When T is long enough for measurable exchange to occur, A to B and B to A conversions during the mixing time lead to two additional cross-peaks (magenta) at ($\omega_{N,A}$, $\omega_{H,B}$) and ($\omega_{N,B}$, $\omega_{H,A}$). The dependences of the peak intensities on mixing time can then be analyzed to yield chemical exchange rate constants.

Molecular dynamics (MD) simulation is another major technique that can reveal a great deal of dynamic information about biomolecules. These "computational microscopes"[49] could deliver unparallelled all-atom descriptions about the dynamics of biomolecules. However, these simulations are based on imperfect force fields, and the timescale involved here is several orders of magnitude longer than could be afforded by currently available computing resources. So a combination of both experiment and modeling is needed. Through collaboration with the lab of Emad Tajkorshid (University of Illinois at Urbana-Champaign), we have also integrated molecular modeling and molecular dynamics (MD) simulations to complement our solution NMR studies of EmrE structure and dynamics.

Scope of Thesis

This thesis uses mainly solution NMR to characterize the structure and dynamics of EmrE in atomic detail. Chapter 2 summarizes efforts to identify and characterize EmrE dynamics mutants suitable for structure determination. Chapter 3 describes efforts toward a high-resolution structure of EmrE in complex with drug using solution NMR spectroscopy. Chapter 4 investigates a series of mutants with different ligands to elucidate the role of helix kinking in conformational interconversion. Chapter 5 reports the asymmetric features of the EmrE C-termini that have largely been overlooked to date. By correlating high-resolution structural and dynamic data with data from functional transport assays, this thesis investigates the multidrug transport mechanism of EmrE. The new insights enrich our understanding of the key properties controlling proton-driven drug efflux in EmrE and set the stage for understanding even more challenging transporters.

References

1 Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-972, doi:10.1038/nature06522 (2007).

2 Shi, Y. A glimpse of structural biology through X-ray crystallography. *Cell* **159**, 995-1014, doi:10.1016/j.cell.2014.10.051 (2014).

3 Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737-738 (1953).

4 Watson, J. D. & Crick, F. H. C. Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* **171**, 964-967 (1953).

5 Schnell, J. R., Dyson, H. J. & Wright, P. E. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annual review of biophysics and biomolecular structure* **33**, 119-140, doi:10.1146/annurev.biophys.33.110502.133613 (2004).

6 Boehr, D. D., Dyson, H. J. & Wright, P. E. An NMR perspective on enzyme dynamics. *Chemical reviews* **106**, 3055-3079, doi:10.1021/cr050312q (2006).

7 Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science (New York, N.Y.)* **313**, 1638-1642, doi:10.1126/science.1130258 (2006).

8 Bu, Z. & Callaway, D. J. Proteins move! Protein dynamics and long-range allostery in cell signaling. *Advances in protein chemistry and structural biology* **83**, 163-221, doi:10.1016/B978-0-12-381262-9.00005-7 (2011).

9 Mustoe, A. M., Brooks, C. L. & Al-Hashimi, H. M. Hierarchy of RNA functional dynamics. *Annual review of biochemistry* **83**, 441-466, doi:10.1146/annurev-biochem-060713-035524 (2014).

10 Han, D. *et al.* DNA origami with complex curvatures in three-dimensional space. *Science (New York, N.Y.)* **332**, 342-346, doi:10.1126/science.1202998 (2011).

11 Nikolova, E. N. *et al.* Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498-502, doi:10.1038/nature09775 (2011).

12 Aperia, A. Membrane transport proteins in health and disease. *Journal of Internal Medicine* **261**, 2-4, doi:10.1111/j.1365-2796.2006.01756.x (2007).

13 Lin, L., Yee, S. W., Kim, R. B. & Giacomini, K. M. SLC transporters as therapeutic targets: emerging opportunities. *Nature reviews. Drug discovery* **14**, 543-560, doi:10.1038/nrd4626 (2015).

14 Crane, R. K., Forstner, G. & Eichholz, A. Studies on the mechanism of the intestinal absorption of sugars. X. An effect of Na+ concentration on the apparent Michaelis constants for intestinal sugar transport, in vitro. *Biochimica et biophysica acta* **109**, 467-477 (1965).

15 Mitchell, P. A General Theory of Membrane Transport From Studies of Bacteria. *Nature* **180**, 134-136 (1957).

16 Patlak, C. S. Contributions to the theory of active transport: II. The gate type non-carrier mechanism and generalizations concerning tracer flow, efficiency, and measurement of energy expenditure. *The bulletin of mathematical biophysics* **19**, 209-235 (1957).

17 Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969-970, doi:10.1038/211969a0 (1966).

18 Mitchell, P. Osmochemistry of solute translocation. *Research in microbiology* (1990).

19 Deng, D. *et al.* Crystal structure of the human glucose transporter GLUT1. *Nature* **510**, 121-125, doi:10.1038/nature13306 (2014).

20 Deng, D. *et al.* Molecular basis of ligand recognition and transport by glucose transporters. *Nature* **526**, 391-396, doi:10.1038/nature14655 (2015).

21 Nomura, N. *et al.* Structure and mechanism of the mammalian fructose transporter GLUT5. *Nature* **526**, 397-401, doi:10.1038/nature14909 (2015).

22 Woodall, N. B., Yin, Y. & Bowie, J. U. Dual-topology insertion of a dual-topology membrane protein. *Nature communications* **6**, 8099, doi:10.1038/ncomms9099 (2015).

23 von Heijne, G. Membrane-protein topology. *Nature Reviews Molecular Cell Biology* **7**, 909-918, doi:10.1038/nrm2063 (2006).

24 Seppälä, S., Slusky, J. S., Lloris-Garcerá, P., Rapp, M. & von Heijne, G. Control of membrane protein topology by a single C-terminal residue. *Science (New York, N.Y.)* **328**, 1698-1700, doi:10.1126/science.1188950 (2010).

25 Rapp, M., Seppälä, S., Granseth, E. & von Heijne, G. Emulating membrane protein evolution by rational design. *Science (New York, N.Y.)* **315**, 1282-1284, doi:10.1126/science.1135406 (2007).

26 Bay, D. C. & Turner, R. J. Diversity and evolution of the small multidrug resistance protein family. *BMC evolutionary biology* **9**, 140, doi:10.1186/1471-2148-9-140 (2009).

27 Schuldiner, S. EmrE, a model for studying evolution and mechanism of ion-coupled transporters. *Biochimica et biophysica acta* **1794**, 748-762, doi:10.1016/j.bbapap.2008.12.018 (2009).

28 Matsumura, K., Furukawa, S., Ogihara, H. & Morinaga, Y. Roles of multidrug efflux pumps on the biofilm formation of Escherichia coli K-12. *Biocontrol science* **16**, 69-72 (2011).

29 Nasie, I., Steiner-Mordoch, S. & Schuldiner, S. New substrates on the block: clinically relevant resistances for EmrE and homologues. *Journal of bacteriology* **194**, 6766-6770, doi:10.1128/JB.01318-12 (2012).

30 Sanders, C. R. Perplexing new insight into the dynamics of the EmrE transporter. *The Journal of general physiology* **146**, 441-444, doi:10.1085/jgp.201511523 (2015).

31 Korkhov, V. M. & Tate, C. G. An emerging consensus for the structure of EmrE. *Acta Crystallographica Section D: Biological Crystallography* **65**, 186-192, doi:10.1107/S0907444908036640 (2009).

32 Arkin, I. T., Russ, W. P., Lebendiker, M. & Schuldiner, S. Determining the secondary structure and orientation of EmrE, a multi-drug transporter, indicates a transmembrane four-helix bundle. *Biochemistry* **35**, 7233-7238, doi:10.1021/bi960094i (1996).

33 Fleishman, S. *et al.* Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *Journal of molecular biology* **364**, 54-67, doi:10.1016/j.jmb.2006.08.072 (2006).

34 Chen, Y.-J. *et al.* X-ray structure of EmrE supports dual topology model. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18999-19004, doi:10.1073/pnas.0709387104 (2007).

35 Morrison, E. *et al.* Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature* **481**, 45-50, doi:10.1038/nature10703 (2012).

36 Lloris-Garcerá, P. *et al.* Antiparallel dimers of the small multidrug resistance protein EmrE are more stable than parallel dimers. *The Journal of biological chemistry* **287**, 26052-26059, doi:10.1074/jbc.M112.357590 (2012).

37 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. *The EMBO journal* **22**, 6175-6181, doi:10.1093/emboj/cdg611 (2003).

38 Sikora, C. W. & Turner, R. J. Investigation of ligand binding to the multidrug resistance protein EmrE by isothermal titration calorimetry. *Biophysical journal* **88**, 475-482, doi:10.1529/biophysj.104.049247 (2005).

39 Winstone, T. L. *et al.* Organic solvent extracted EmrE solubilized in dodecyl maltoside is monomeric and binds drug ligand. *Biochemical and biophysical research communications* **327**, 437-445, doi:10.1016/j.bbrc.2004.11.164 (2005).

40 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* **285**, 15234-15244, doi:10.1074/jbc.M110.108746 (2010).

41 Tate, C. G., Ubarretxena-Belandia, I. & Baldwin, J. M. Conformational changes in the multidrug transporter EmrE associated with substrate binding. *Journal of molecular biology* **332**, 229-242, doi:10.1016/S0022-2836(03)00895-7 (2003).

42 Lloris-Garcerá, P., Seppälä, S., Slusky, J. S., Rapp, M. & von Heijne, G. Why have small multidrug resistance proteins not evolved into fused, internally duplicated structures? *Journal of molecular biology* **426**, 2246-2254, doi:10.1016/j.jmb.2014.03.012 (2014).

43 Vaidehi, N., Grisshammer, R. & Tate, C. G. How Can Mutations Thermostabilize G-Protein-Coupled Receptors? *Trends in pharmacological sciences* **37**, 37-46, doi:10.1016/j.tips.2015.09.005 (2016).

44 Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. Conformational thermostabilization of the 1-adrenergic receptor in a detergent-resistant form. *Proceedings of the National Academy of Sciences* **105**, 877-882, doi:10.1073/pnas.0711253105 (2008).

45 Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. (Wiley, 2001).

46 Cavanagh, J., Fairbrother, W. J., Palmer Iii, A. G., Rance, M. & Skelton, N. J. in *Protein NMR Spectroscopy (Second Edition)* 333-404 (Academic Press, 2007).

47 Mittermaier, A. K. & Kay, L. E. Observing biological dynamics at atomic resolution using NMR. *Trends in biochemical sciences* **34**, 601-611, doi:10.1016/j.tibs.2009.07.004 (2009).

48 Li, Y. & Palmer, A. G. TROSY-selected ZZ-exchange experiment for characterizing slow chemical exchange in large proteins. *Journal of biomolecular NMR* **45**, 357-360, doi:10.1007/s10858-009-9385-0 (2009).

49 Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. & Shaw, D. E. Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics* **41**, 429-452, doi:10.1146/annurev-biophys-042910-155245 (2012).

Chapter 2: Search for putative dynamics mutants

Introduction

The transport cycle of a transporter requires several structural states. EmrE is a proton-coupled antiporter; therefore, there are at least three structural states in terms of ligand binding: drug-bound, proton-bound, and the true *apo* state. In our structural studies of EmrE, we focus on the drug-bound state. This drug-bound state is the most studied; and there are two low-resolution structures available[1,2], significantly lowering barriers involved in solving a membrane protein structure using NMR. Also, with the drug occupying the binding site, there is reduced amide exchange[3] useful for long NMR experiments. For the ligand we focus on TPP^+, which is the mostly commonly used ligand for EmrE. The affinity of EmrE for TPP^+ is in the nanomolar-to-micromolar range[4], making a fully ligand-saturated state easy to achieve. This avoids potential complications from ligand binding equilibrium with a low-affinity binder.

The constant conformational interconversion of EmrE poses a significant challenge for NMR structure determination of EmrE. This is a problem for any NMR experiments long enough to allow cross-peaks from conformational interconversion to build up. For example, NOESY experiments, to measure NOEs, the major class of structural restraints, necessitate the use of really long mixing times (up to a few hundred milliseconds). But for TPP^+-bound WT EmrE, the interconversion rate is ~5 s^{-1} in our NMR conditions[4]. Thus, cross-peaks from conformational interconversion build up long before NOEs peaks appear, seriously complicating data interpretation. This was experimentally verified[4] – exchange cross-peaks were visible with NOESY mixing times as short as 25 ms for TPP^+-bound WT EmrE at pH 7, and mixing times

long enough to measure NOEs resulted in identical NOE observations for both monomers due to the extensive exchange occurring during the mixing time effectively averaging the two states.

Inspired by the success of thermal stabilization by point mutations in GPCR crystallography[5,6], we collaborated with the lab of Chris Tate (MRC Laboratory of Molecular Biology, Cambridge, UK) to screen and select for putative dynamics mutants of EmrE. The Tate lab constructed a library of EmrE mutants containing V, A, or G at each position and then screened these mutants using dilution plate assays[7] and [^{3}H]-TPP$^+$ binding assays[8]. Mutants that displayed good ligand binding but loss of function were selected with hope that these mutants would reflect loss of dynamics.

I went on to further characterize these putative dynamics mutants identified in the screen. I used an in-cell drug export assay[9] to evaluate the ethidium bromide export capabilities of cells expressing EmrE mutants. This assay provides a more nuanced readout on mutant transport activity by monitoring the kinetics of ethidium efflux over short periods of time, whereas dilution plating simply reflects whether drug efflux is sufficient to enable survival over 24 hours. For mutants that indeed have impaired drug export, I then purified isotopically labeled protein to record NMR TROSY-HSQC spectra. The ideal mutants suitable for structural studies should maintain the overall asymmetric dimer peak pattern with favorable lineshape properties, reflecting a stable and properly folded structure. I then performed ZZ-exchange experiments to quantify the conformational interconversion rates[4]. I wanted mutants that have slower interconversion rates than WT EmrE, since our initial goal was structure determination. I also performed isothermal titration calorimetry (ITC) to evaluate the drug binding capabilities of these EmrE mutants[9].

Materials and methods

Sample preparation WT and mutant EmrE was expressed, purified, and reconstituted into DLPC/DHPC isotropic bicelles as previously described[4,9,10]. Single point mutants were constructed using QuickChange (Stratagene).

In-cell assay These assays were carried out using the BL21(DE3) strains of E. coli transformed with empty pET15b vectors or pET15b-EmrE. The cells were grown in M9 minimal media with 100 μg/ml of ampicillin at 37 ˚C until the OD_{600} reached 0.4. Then cells were induced with 0.33 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) at the same temperature for 30 minutes followed by incubation with 2.5 μM ethidium bromide and 40 μM carbonyl cyanide p-chlorophenylhydrazone (CCCP) for an hour to load ethidium bromide into the cells. The cell cultures were then stored on ice until assays were complete. For each experiment, 2 ml of cell culture was spun down and immediately resuspended in 1 ml fresh M9 media with 2.5 μM ethidium bromide. Fluorescence of ethidium bromide was monitored with an excitation wavelength at 545 nm and emission wavelength at 610 nm. The time course of fluorescence was plotted after normalization to the initial value of each run.

NMR spectroscopy NMR data were collected on samples containing 0.8-1.5 mM $^{2}H/^{15}N$ EmrE in 20 mM potassium phosphate, 20 mM NaCl, 8-10% D_2O, pH 7 at 45 ˚C on a Varian 700 MHz spectrometer with a room temperature probe. For TPP^+-bound EmrE, 1-2 mM TPP^+ was added to saturate EmrE. 2D $^{1}H/^{15}N$ TROSY-HSQC and TROSY-selected ZZ-exchange experiments[11] with a lipid flip-back pulse[4] were carried out with a recycle delay of 2 s and 128-144 increments. The conformational interconversion rate, k_{conf}, was determined from the ZZ-exchange data as

previously described[12] using the composite peak ratio method with an 11.1 ms time offset to account for the back-transfer time in the pulse sequence[9]. All NMR spectra were processed with NMRPipe[13] and analyzed in CcpNmr Analysis[14].

Isothermal titration calorimetry All isothermal titration calorimetry (ITC) experiments were performed with EmrE reconstituted into DLPC/DHPC (q = 0.33) isotropic bicelles in 20 mM potassium phosphate, 20 mM NaCl, pH 7, 45 ºC. All ITC experiments were performed on a TA instruments LV Nano calorimeter. Data were fit simultaneously to a model of ligand binding to n independent and identical sites plus a constant which is used to represent the baseline of bicelle mixing[9].

Results and discussion

Putative dynamics mutants from screening Chris Tate's lab identified and provided us with the first batch of seven putative dynamics mutants from the comprehensive screening (Table 2.1): M21G, A59L, S64V, G90V, I101G, N102A, and N102V.

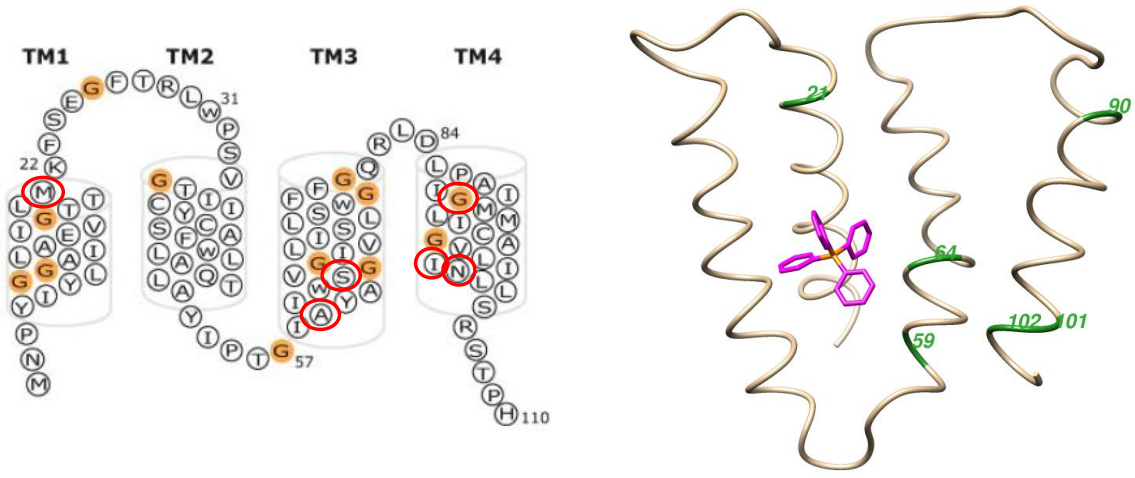


Figure 2.1 Locations of putative dynamics mutations in the secondary structure (left) and in monomer A of the X-ray structure (right). The secondary structure figure was adapted from Elbaz *et al.*, 2008. Glycine residues are colored in orange and the seven mutants are in red circles.

The locations of these seven residues (Fig. 2.1) can be categorized into three regions that may be important in conformational interconversion. Residue M21 lies at the end of TM1, with one M21 at the bottom of the binding pocket while the other is above the binding pocket due to the asymmetric dimer architecture of EmrE. This residue is located in the same TM helix as E14, the residue responsible for proton binding. A59 and S64 are close to the ^{65}GVG^{67} TM3 kink region, which is important for conformation interconversion and multidrug recognition from cryoEM and NMR studies[1,4,9,15,16]. Like M21, A59 lies at one end of a TM helix where it transitions into the loop region. G90, I101, and N102 are located in TM4, the dimerization domain of EmrE[17,18]. In TM4 of EmrE, there is a knobs-into-holes motif where glycine residues (90/97) from one monomer pack against Leu/Ile residues (93/94) from the other monomer.

Table 2.1 Growth phenotype summary from dilution plate assays for V/A/G mutants of EmrE. Residues in each transmembrane helix (TM) are labeled with the helix number in the TM column. Three different substrates and two different concentrations (low and high) of drug substrates were screened: methyl viologen (50 and 200 μM), acriflavine (116 and 463 μM), and ethidium bromide (EtBr, 400 and 1500 μM). Symbols: ++, similar to WT EmrE; +, growth on higher concentrations of substrates, but weak; +/-, only growth on lower concentration; -, no growth. Another set of dilution plate assays[7] by Mchaourab lab for a library of single Cys mutants is also shown for comparison in the last column. The column labeled K_D lists the TPP^+ affinities of EmrE mutants tested.

Amino acid	TM	Mut	Methyl viologen	Acriflavine	EtBr	K_D (nM)	Mchaourab Cys mutants-growth on 507 μM EtBr
N2		Val	++	++	++		
N2		Ala	++	++	++		
N2		Gly	++	++	++		
P3		Val	++	++	++		Grew at 1/10 diln
P3	1	Ala	++	++	++		
P3		Gly	++	++	++		
Y4		Val	+	+	+		Grew at 1/10 diln
Y4	1	Ala	+	+	+		
Y4		Gly	+	+	+		
I5		Val	++	++	++		
I5	1	Ala	++	++	++		

I5		Gly	++	++	++		
Y6		Val	++	++	++		Grew at 1/1000 diln
Y6	1	Ala	++	++	++		
Y6		Gly	nd	nd	nd		
L7		Val	+	++	++		Grew at 1/10 diln
L7	1	Ala	+	++	++		
L7		Gly	-	+/-	+/-		
G8		Val	++	++	++		Grew at 1/1000 diln
G8	1	Ala	++	++	++		
G8		Gly	++	++	++		
G9		Val	++	++	++		
G9	1	Ala	++	++	++		
G9		Gly	++	++	++		
A10		Val	-	-	-		No growth for Cys mutant
A10	1	Ala	++	++	++		
A10		Gly	++	++	++		
I11		Val	++	++	++		
I11	1	Ala	++	++	++		
I11		Gly	+/-	-	-		
L12		Val	++	++	++		
L12	1	Ala	++	++	++		
L12		Gly	++	++	++		
A13		Val	++	++	++		
A13	1	Ala	++	++	++		
A13		Gly	++	++	++		
E14		Val	-	-	-		No growth for Cys mutant
E14	1	Ala	nd	nd	nd		
E14		Gly	-	-	-	Negative control	
V15		Val	++	++	++		
V15	1	Ala	+/-	+	++		
V15		Gly	+/-	+	++		
I16		Val	++	++	++		Grew at 1/1000 diln
I16	1	Ala	++	++	++		
I16		Gly	++	++	++		
G17		Val	-	-	-		No growth for Cys mutant
G17	1	Ala	+/-	+/-	+/-		
G17		Gly	++	++	++		
T18		Val	+/-	-	-		No growth for Cys mutant
T18	1	Ala	-	-	-		
T18		Gly	-	-	-		
T19		Val	++	++	++		Grew at 1/1000 diln
T19	1	Ala	++	++	++		
T19		Gly	++	+/-	++		
L20		Val	++	++	++		
L20	1	Ala	++	++	++		
L20		Gly	++	++	++		
M21		Val	+/-	+	+		Grew at 1/1000 diln
M21	1	Ala	-	+	+		
M21		Gly	-	+	+		

K22		Val	++	++	+/-		Grew at 1/10 diln
K22	1	Ala	++	++	+		
K22		Gly	++	++	+		
F23		Val	++	++	+/-		
F23	1	Ala	++	++	++		
F23		Gly	++	++	++		
S24		Val	++	++	++		Grew at 1/1000 diln
S24		Ala	++	++	++		
S24		Gly	++	++	++		
E25		Val	++	++	++		
E25		Ala	++	++	++		
E25		Gly	++	++	++		
G26		Val	++	++	++		Grew at 1/1000 diln
G26		Ala	++	++	++		
G26		Gly	++	++	++		
F27		Val	++	++	++		Grew at 1/1000 diln
F27		Ala	++	++	++		
F27		Gly	++	++	++		
T28		Val	++	++	++		No growth for Cys mutant
T28		Ala	++	++	++		
T28		Gly	++	++	++		
R29		Val	++	++	++		
R29		Ala	++	++	++		
R29		Gly	++	++	++		
L30		Val	++	++	++		
L30	2	Ala	++	++	++		
L30		Gly	++	++	++		
W31		Val	++	++	++		
W31	2	Ala	++	++	++		
W31		Gly	++	++	++		
P32		Val	++	++	++		
P32	2	Ala	++	++	++		
P32		Gly	++	++	++		
S33		Val	++	++	++		
S33	2	Ala	++	++	++		
S33		Gly	++	++	++		
V34		Val	++	++	++		Grew at 1/1000 diln
V34	2	Ala	++	++	++		
V34		Gly	++	++	++		
G35		Val	++	++	++		
G35	2	Ala	++	++	++		
G35		Gly	++	++	++		
T36		Val	+	++	++		
T36	2	Ala	++	++	++		
T36		Gly	+	++	++		
I37		Val	++	++	++		
I37	2	Ala	++	++	++		
I37		Gly	++	++	++		
I38		Val	++	++	++		
I38	2	Ala	++	++	++		
I38		Gly	++	++	++		
C39		Val	++	++	++		
C39	2	Ala	++	++	++		

C39		Gly	++	++	++		
Y40		Val	+	+	+		Grew at 1/1000 diln
Y40	2	Ala	+/-	-	+/-		
Y40		Gly	+/-	+/-	+/-		
C41		Val	++	++	++		
C41	2	Ala	++	++	++		
C41		Gly	++	++	++		
A42		Val	++	++	++		
A42	2	Ala	++	++	++		
A42		Gly	++	++	++		
S43		Val	-	++	++		
S43	2	Ala	++	++	++		
S43		Gly	+/-	++	++		
F44		Val	+/-	++	++		
F44	2	Ala	nd	nd	nd		
F44		Gly	-	++	++		
W45		Val	++	++	++		
W45	2	Ala	++	++	++		
W45		Gly	++	++	++		
L46		Val	++	++	++		
L46	2	Ala	++	++	++		
L46		Gly	++	++	++		
L47		Val	+/-	++	++		Grew at 1/10 diln
L47	2	Ala	-	-	-		
L47		Gly	-	-	-		
A48		Val	-	+	+/-		
A48	2	Ala	++	++	++		
A48		Gly	++	++	++		
Q49		Val	++	++	++		
Q49	2	Ala	++	++	++		
Q49		Gly	++	++	++		
T50		Val	++	++	++		
T50	2	Ala	++	++	++		
T50		Gly	++	++	++		
L51		Val	+	++	++		
L51	2	Ala	+	++	+		
L51		Gly	+	+	+		
A52		Val	++	++	++		
A52	2	Ala	++	++	++		
A52		Gly	++	++	++		
Y53		Val	++	++	++		
Y53	2	Ala	++	++	++		
Y53		Gly	++	++	++		
I54		Val	++	++	++		Grew at 1/1000 diln
I54	2	Ala	+/- (++)	+	++		
I54		Gly	-	-	+/-		
P55		Val	++	++	+		Grew at 1/10 diln
P55		Ala	++ (+/-)	++	++		
P55		Gly	++	++	++		
T56		Val	++	++	++		
T56		Ala	++ (+)	++	++		
T56		Gly	++	++	++		
G57		Val	-	-	-		Grew at 1/10 diln
G57		Ala	+/-	++	++		

G57		Gly	++	++	++		
I58		Val	++	++	++		No growth for Cys mutant
I58	3	Ala	++	++	++		
I58		Gly	+	++	++		
A59		Leu	-	-	-		
A59		Val	++	++	++		
A59	3	Ala	++	++	++		
A59		Gly	++	++	++		
Y60		Val	-	-	-		No growth for Cys mutant
Y60	3	Ala	++	+	+		
Y60		Gly	++	+	+		
A61		Val	++	++	++		
A61	3	Ala	++	++	++		
A61		Gly	++	++	++		
I62		Val	++	++	++		
I62	3	Ala	++	++	++		
I62		Gly	+	++	++		
W63		Val	nd	nd	nd		No growth for Cys mutant
W63	3	Ala	-	-	-		
W63		Gly	-	-	-		
S64		Val	+/-	-	-	1.44±0.71	
S64	3	Ala	++	++	++		
S64		Gly	++	++	++		
G65		Val	++	++	++		
G65	3	Ala	++	++	++		
G65		Gly	++	++	++		
V66		Val	++	++	++		
V66	3	Ala	++	++	++		
V66		Gly	++	++	++		
G67		Val	-	-	-		No growth for Cys mutant
G67	3	Ala	-	-	-	Low expression	
G67		Gly	++	++	++		
I68		Val	++	++	++		
I68	3	Ala	++ (+)	++	++ (+)		
I68		Gly	++	++	++		
V69		Val	++	++	++		
V69	3	Ala	++	++	++		
V69		Gly	++	++	++		
L70		Val	++	++	++		
L70	3	Ala	++	++	++		
L70		Gly	++	++	++		
I71		Val	++	++	++		
I71	3	Ala	++	++	++		
I71		Gly	+/-	++	++		
S72		Val	++	++	++		
S72	3	Ala	++	++	++		
S72		Gly	++	++	++		
L73		Val	++	++	++		
L73	3	Ala	++	++	++		

L73		Gly	++	++	++		
L74		Val	++	++	++		
L74	3	Ala	++	++	++		
L74		Gly	++	++	++		
S75		Val	++	++	++		
S75	3	Ala	++	++	++		
S75		Gly	++	++	++		
W76		Val	++	++	++		
W76	3	Ala	++	++	++		
W76		Gly	++	++	++		
G77		Val	++	++	++		
G77	3	Ala	++	++	++		
G77		Gly	++	++	++		
F78		Val	++	++	++		
F78	3	Ala	++	++	++		
F78		Gly	++	++	++		
F79		Val	++	++	++		
F79	3	Ala	++	++	++		
F79		Gly	nd	nd	nd		
G80		Val	++	++	++		
G80	3	Ala	++	++	++		
G80		Gly	++	++	++		
Q81		Val	++	++	++		
Q81		Ala	++	++	++		
Q81		Gly	++	++	++		
R82		Val	++	++	++		
R82		Ala	++	++	++		
R82		Gly	++	++	++		
L83		Val	+	++	++		
L83		Ala	++	++	++		
L83		Gly	++	++	++		
D84		Val	+	++	+		
D84		Ala	nd	nd	nd		
D84		Gly	+	++	++		
L85		Val	++	++	++		
L85	4	Ala	++	++	++		
L85		Gly	++	++	++		
P86		Val	++	++	++		
P86	4	Ala	++	++	++		
P86		Gly	++	++	++		
A87		Leu	+/-	+	-		Grew at 1/1000 diln
A87		Val	++	++	++		
A87	4	Ala	++	++	++		
A87		Gly	++	++	++		
I88		Val	++	++	++		
I88	4	Ala	++	++	++		
I88		Gly	nd	nd	nd		
I89		Val	++	++	++		
I89	4	Ala	++	++	++		
I89		Gly	++	++	++		
G90		Val	-	-	-	5.28±0.60	No growth for Cys mutant
G90	4	Ala	++	++	++		
G90		Gly	++	++	++		

M91		Val	++	++	++		Grew at 1/10 diln
M91	4	Ala	++	++	++		
M91		Gly	++	++	++		
M92		Val	++	++	++		Grew at 1/10 diln
M92	4	Ala	++	++	++		
M92		Gly	++	++	++		
L93		Val	+/-	++	++		
L93	4	Ala	++	++	++		
L93		Gly	++	++	++		
I94		Val	++	++	++		No growth for Cys mutant
I94	4	Ala	++	++	+		
I94		Gly	-	-	-	6.12±0.70	
C95		Val	nd	nd	nd		
C95	4	Ala	++	++	++		
C95		Gly	++	++	++		
A96		Val	++	++	++		
A96	4	Ala	++	++	++		
A96		Gly	++	++	++		
G97		Val	-	-	-	Low expression	No growth for Cys mutant
G97	4	Ala	++	++	++		
G97		Gly	++	++	++		
V98		Val	++	++	++		
V98	4	Ala	++	++	++		
V98		Gly	++	++	++		
L99		Val	++	++	++		Grew at 1/1000 diln
L99	4	Ala	++	++	++		
L99		Gly	++	++	++		
I100		Val	++	++	++		Grew at 1/1000 diln
I100	4	Ala	++	++	++		
I100		Gly	++	++	++		
I101		Val	++	++	++		No growth for Cys mutant
I101	4	Ala	++	++	++		
I101		Gly	+	++	+		
N102		Val	-	++	+/-	1.25±0.25	No growth for Cys mutant
N102	4	Ala	-	++	+/-	2.30±0.37	
N102		Gly	+	++	+		
L103		Val	++	++	++		
L103	4	Ala	++	++	++		
L103		Gly	++	++	++		
L104		Val	++	++	++		
L104	4	Ala	++	++	++		
L104		Gly	++	++	++		
S105		Val	++	++	++		
S105	4	Ala	++	++	++		
S105		Gly	++	++	++		
R106		Val	++	++	++		
R106		Ala	++	++	++		
R106		Gly	++	++	++		
S107		Val	++	++	++		
S107		Ala	++	++	++		

S107		Gly	++	++	++		
T108		Val	++	++	++		
T108		Ala	++	++	++		
T108		Gly	++	++	++		
P109		Val	++	++	++		
P109		Ala	++	++	++		
P109		Gly	++	++	++		
H110		Val	++	++	++		
H110		Ala	++	++	++		
H110		Gly	++	++	++		

Characterization of putative dynamics mutants I went on to characterize these putative dynamics mutants, and evaluated whether they are indeed dynamics mutants suitable for our structural studies. Unexpectedly, most of the seven mutants identified have *faster* interconversion rates. For example, the mutation M21G indeed reduces the rate of drug export by EmrE as assessed from our in-cell export assay (Fig. 2.2). However, the TROSY-HSQC spectrum (Fig. 2.3) for TPP^+-bound EmrE-M21G suggests that conformational exchange between open-in and open-out in EmrE is actually faster for this mutant. Although there are still two sets of peaks corresponding to the asymmetric dimer, there is noticeable line-broadening and some pairs of amide resonances are collapsing together. This is confirmed by the ZZ exchange experiment and the rate is around 10-15 s^{-1}. This is also true for the EmrE-A59L mutant since the broadening of peaks due to exchange is more pronounced suggesting even faster interconversion. Some of the observed line-broadening may also arise from structural heterogeneity and the peak positions are slightly shifted, suggesting that there may be reduced asymmetry in these mutants. Together these results suggest that these mutants may destabilize packing of the loops needed to define the asymmetric closed-on-one-side structure of EmrE. This would explain enhanced exchange and greater structural heterogeneity. Incomplete closure of the transport pore would also lead to substrate leakage and prevent active efflux.

The mutants in TM4 (I101G, N102A, and N102V) have faster interconversion as inferred from the TROSY-HSQC spectra, possibly through disruption of the interface between the TM4 dimerization motif and the TM3 kink region important for open-in and open-out exchange. This might be true for G90V as well, although most likely this mutation disrupts the dimerization of EmrE in the first place[17], as shown by the reduced number of peaks visible in the spectrum.

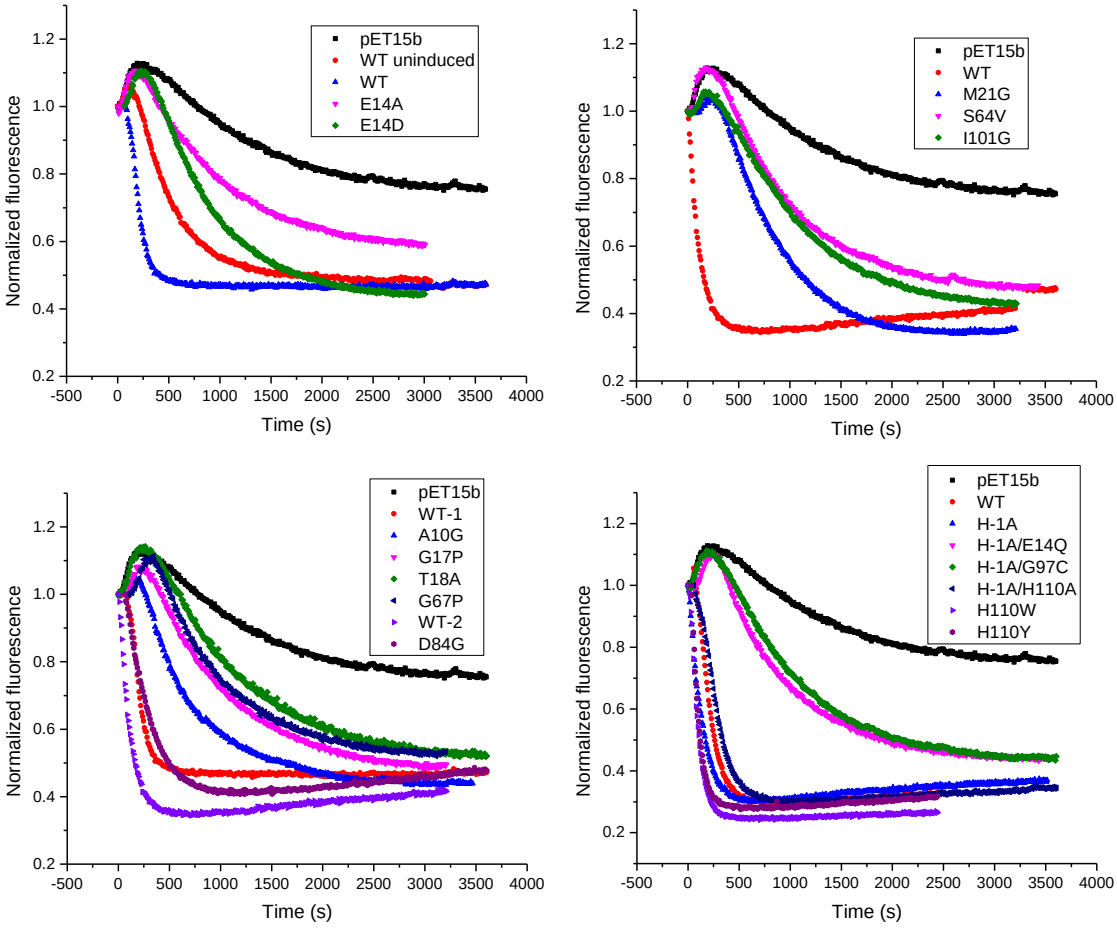


Fig. 2.2 Ethidium export properties for cells expressing EmrE mutants from in-cell assay. In each subfigure, the WT control was performed the same day. There is an additional control in lower left subfigure for D84G (WT-2).

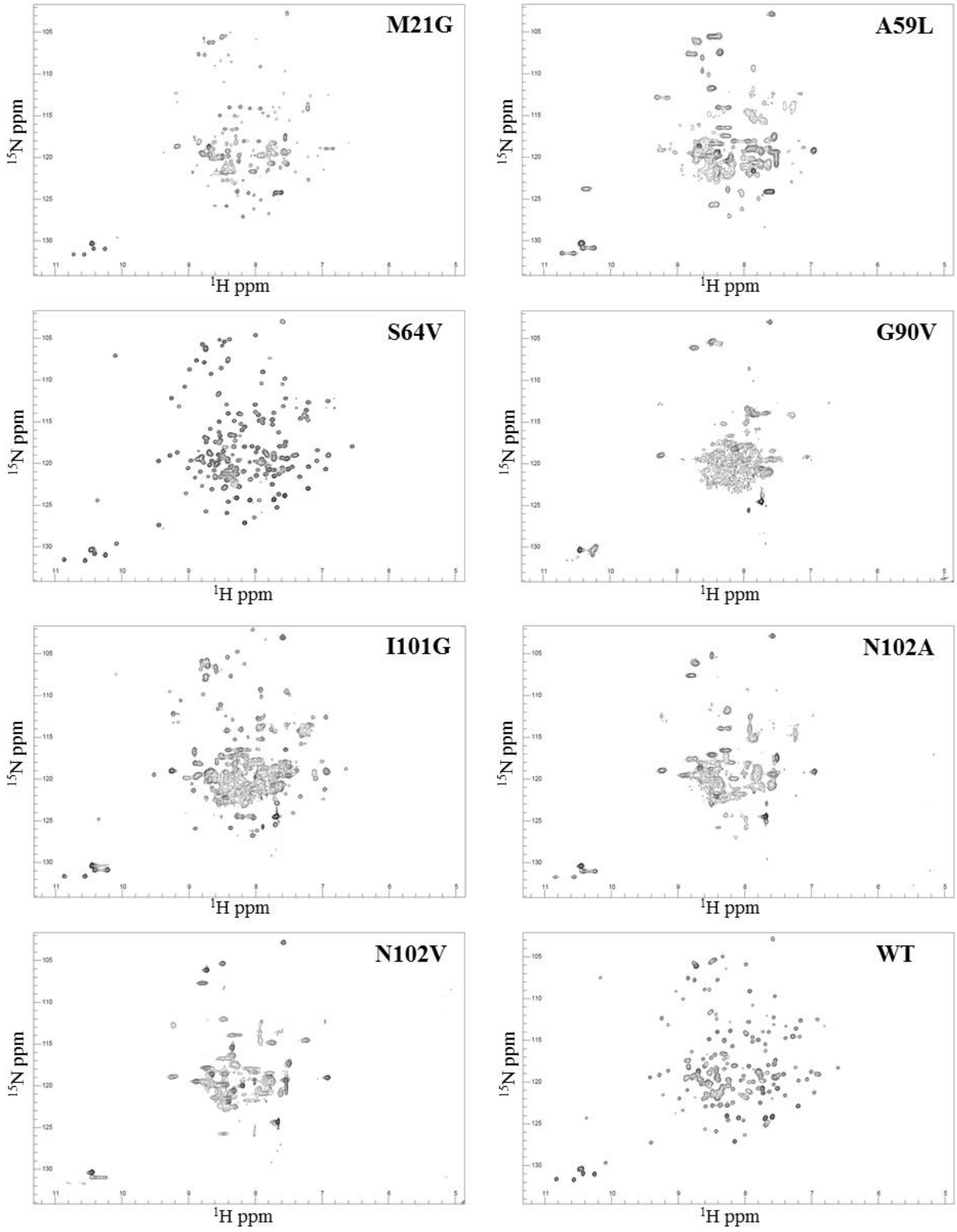


Fig. 2.3 TROSY-HSQC spectra for TPP^{+}-bound EmrE in DLPC/DHPC (q = 0.33) bicelles. Spectra were collected at 45˚C on a Varian 700 MHz spectrometer with a room temperature cryoprobe.

Identification of EmrE-S64V as a slow dynamics mutant One mutant, EmrE-S64V, had the desired behavior with reduced conformational interconversion and consequently better NMR spectral quality (Fig. 2.4). Using ITC, I confirmed the TPP^+ binding affinity of EmrE-S64V is largely the same as WT. So the reduced drug export from the in-cell assay for this mutant is not due to impaired drug binding and likely arises as a result of slow transport across the membrane.

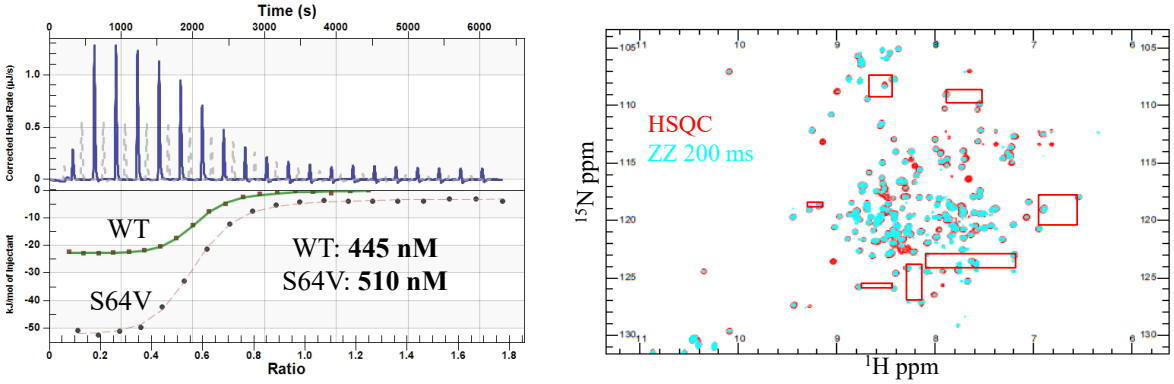


Fig. 2.4 EmrE-S64V is a slow dynamics mutant. The binding affinity from ITC for TPP^+ for EmrE-S64V is similar to WT. The conformational interconversion rate is 1.2 ± 0.2 s^{-1} for EmrE-S64V in complex with TPP^+ in DLPC/DHPC (q=0.33) bicelles at 45˚C.

A more complete characterization of EmrE-S64V is demonstrated in Chapters 3 and 4. The conformational interconversion rate of EmrE-S64V in complex with TPP^+ (0.6 ± 0.1 s^{-1} in DMPC/DHPC bicelles and 1.2 ± 0.2 s^{-1} in DLPC/DHPC bicelles) makes NMR structural studies of EmrE possible.

As an alternative approach, we also tried to pursue the structure of WT EmrE in complex with $DPhTPP^+$ (0.4 s^{-1} in DLPC/DHPC bicelles)[9], a substrate that slows the open-in and open-out exchange rate of EmrE. However, EmrE-S64V is structurally more rigid or homogenous as evidenced from the sharp peaks in TROSY spectra. This results in higher quality NMR spectra

and more feasible structure determination for S64V-EmrE bound to TPP^+ than WT-EmrE bound to $DPhTPP^+$ even though both systems have the same rate of open-in to open-out exchange. We therefore pursued NMR structure determination using the S64V mutant, and further explored the effects of S64 mutation to better understand the factors controlling the dynamics of EmrE critical to its transport activity.

Conclusion

A better structure of EmrE is needed to understand the multidrug transport process of EmrE. However, dynamics due to the conformational interconversion between open-in and open-out make high-resolution structures of EmrE challenging for all structural approaches. The conformational interconversion needed to be reduced in order to solve the structure of EmrE using solution NMR. Through a collaborative comprehensive screening, we have identified EmrE-S64V as a slow dynamics mutant suitable for structural studies. We have also explored the importance of S64 in EmrE to gain insight into the mechanism coupling substrate binding and conformational exchange between open-in and open-out to better understand how EmrE performs active drug efflux.

Author contributions

This work was collaboration between Katherine A. Henzler-Wildman lab and Chris G. Tate lab. Samantha Wynne from Tate lab collected the dilution plate assay data and provided the seven putative dynamics mutants. Katherine Henzler-Wildman did the mutagenesis work. Chao Wu

further characterized these mutants and identified EmrE-S64V as a slow dynamics mutant suitable for structural studies.

References

1 Fleishman, S. *et al.* Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *Journal of molecular biology* **364**, 54-67, doi:10.1016/j.jmb.2006.08.072 (2006).

2 Chen, Y.-J. *et al.* X-ray structure of EmrE supports dual topology model. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18999-19004, doi:10.1073/pnas.0709387104 (2007).

3 Rule, G. S. & Hitchens, T. K. *Fundamentals of Protein NMR Spectroscopy*. (Springer Netherlands, 2005).

4 Morrison, E. *et al.* Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature* **481**, 45-50, doi:10.1038/nature10703 (2012).

5 Vaidehi, N., Grisshammer, R. & Tate, C. G. How Can Mutations Thermostabilize G-Protein-Coupled Receptors? *Trends in pharmacological sciences* **37**, 37-46, doi:10.1016/j.tips.2015.09.005 (2016).

6 Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. Conformational thermostabilization of the 1-adrenergic receptor in a detergent-resistant form. *Proceedings of the National Academy of Sciences* **105**, 877-882, doi:10.1073/pnas.0711253105 (2008).

7 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. *The Journal of biological chemistry* **285**, 26710-26718, doi:10.1074/jbc.M110.132621 (2010).

8 Tate, C. G., Ubarretxena-Belandia, I. & Baldwin, J. M. Conformational changes in the multidrug transporter EmrE associated with substrate binding. *Journal of molecular biology* **332**, 229-242, doi:10.1016/S0022-2836(03)00895-7 (2003).

9 Morrison, E. A. & Henzler-Wildman, K. A. Transported substrate determines exchange rate in the multidrug resistance transporter EmrE. *The Journal of biological chemistry* **289**, 6825-6836, doi:10.1074/jbc.M113.535328 (2014).

10 Morrison, E. & Henzler-Wildman, K. Reconstitution of integral membrane proteins into isotropic bicelles with improved sample stability and expanded lipid composition profile. *Biochimica et biophysica acta* **1818**, 814-820, doi:10.1016/j.bbamem.2011.12.020 (2012).

11 Li, Y. & Palmer, A. G. TROSY-selected ZZ-exchange experiment for characterizing slow chemical exchange in large proteins. *Journal of biomolecular NMR* **45**, 357-360, doi:10.1007/s10858-009-9385-0 (2009).

12 Miloushev, V. Z. *et al.* Dynamic properties of a type II cadherin adhesive domain: implications for the mechanism of strand-swapping of classical cadherins. *Structure (London, England : 1993)* **16**, 1195-1205, doi:10.1016/j.str.2008.05.009 (2008).

13 Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277-293, doi:10.1007/bf00197809 (1995).

14 Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696, doi:10.1002/prot.20449 (2005).

15 Korkhov, V. M. & Tate, C. G. Electron crystallography reveals plasticity within the drug binding site of the small multidrug transporter EmrE. *Journal of molecular biology* **377**, 1094-1103, doi:10.1016/j.jmb.2008.01.056 (2008).

16 Gayen, A., Banigan, J. R. & Traaseth, N. J. Ligand-induced conformational changes of the multidrug resistance transporter EmrE probed by oriented solid-state NMR spectroscopy. *Angewandte Chemie (International ed. in English)* **52**, 10321-10324, doi:10.1002/anie.201303091 (2013).

17 Elbaz, Y., Salomon, T. & Schuldiner, S. Identification of a glycine motif required for packing in EmrE, a multidrug transporter from Escherichia coli. *The Journal of biological chemistry* **283**, 12276-12283, doi:10.1074/jbc.M710338200 (2008).

18 Dutta, S., Morrison, E. A. & Henzler-Wildman, K. A. EmrE dimerization depends on membrane environment. *Biochimica et biophysica acta* **1838**, 1817-1822, doi:10.1016/j.bbamem.2014.03.013 (2014).

Chapter 3: Structure determination of EmrE by NMR

Introduction

Structure determination by NMR spectroscopy relies on multidimensional NMR experiments performed on samples labeled with ^{13}C and ^{15}N stable isotopes (plus ^{2}H for large systems) to determine resonance frequencies of nuclei in the protein[1]. This resonance assignment step is necessary for further studies of the structure (and dynamics). To determine the 3D structure, conformational restraints are then collected which usually consist mainly of proton-proton distance restraints and dihedral angles restraints. These conformational restraints are integrated with physicochemical knowledge of protein structures in the force field to calculate final structures.

Resonance assignment

NMR assignment experiments can be grouped into two major categories: through-bond experiments and through-space experiments. In through-bond experiments, connectivities are established between nuclei linked by covalent bonds. For example, triple-resonance ($^{1}H/^{13}C/^{15}N$) backbone-walk experiments correlate nuclei in neighboring residues through scalar couplings and are the most commonly used experiments for backbone resonance assignment. The sensitivities of different experiments depend on the transfer efficiencies, which are related to the scalar coupling constants (Fig. 3.1). These backbone-walk experiments greatly simplify spectral analyses due to their simplicity since often only one or two peaks are associated with a root resonance. Highly deuterated proteins can be used for backbone-walk experiments, making them suitable for backbone assignment of large systems with the constraint that only experiments

beginning on HN can be used because the other protons have been replaced by deuterium. Thus, the out-and-back HNCACB must be used instead of CBCANH.

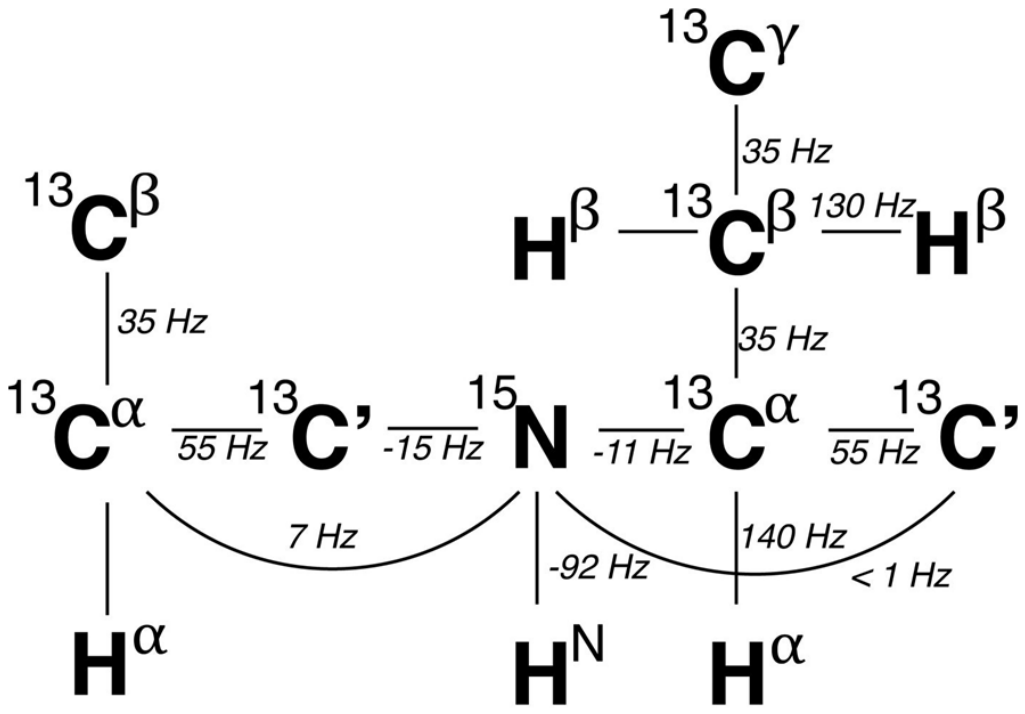


Fig. 3.1 Scalar coupling constants in proteins. Figure from Ferella *et al*., 2012.

COSY (Correlation spectroscopy) and TOCSY (Total correlation spectroscopy) experiments, which are used mainly for side chain assignment of small proteins, belong to another major class of through-bond experiments. In COSY experiments, cross-peaks between pairs of vicinal protons arise due to ^{2}J and ^{3}J scalar couplings (protons that are connected by more than three bonds give no cross-peaks since ^{4}J and above are close to 0). In TOCSY experiments, magnetization is transferred to the entire spin system of an amino acid by successive scalar couplings. So in addition to cross-peaks that appear in COSY spectra, there will be cross-peaks corresponding to interactions of protons connected by more than three bonds. The challenge for using these experiments in large systems lies at the tricky balance between the need for

deuteration and detectable protons. The fast relaxation of signals in highly protonated systems prevents the use of many coherence transfer steps, and the use of deuteration eliminates detectable protons.

Assignments can also be aided by through-space experiments. In through-space experiments, dipolar couplings of nuclei occur as a result of spatial proximity and are independent of the presence of chemical bonds between them. In nuclear Overhauser effect spectroscopy (NOESY) experiments, cross-peaks for pairs of protons within 5-6 Å are observed. These connectivities in space provide tertiary restraints, but may also help in sequential resonance assignment due to the short-range nature of NOE peaks. This is especially useful in regions containing secondary structures[2]. NOESY data is used more often to aim resonance assignment in large systems to resolve ambiguity due to peak overlap, and assist in regions where backbone-walk experiments have lower sensitivity since the coherence transfer experiments based on scalar couplings suffer from shorter transverse relaxation times in high molecular-weight systems. In NOESY experiments, however, magnetization is stored along the Z axis therefore the slower longitudinal relaxation is in effect and good quality spectra can be obtained for larger systems. However, care must be taken since NOE spectra contain peaks due to not only short-range intra and inter-residue but also medium- and long-range contacts.

Structural restraints

The major sources of structural restraints for NMR structure determination are distance restraints derived from NOEs between nearby hydrogen atoms. NOE distance restraints are crucial to define the overall protein fold. However, as the NOE only reports information about local structural features, a large number of NOEs are needed to reconstruct the overall structure[3]. Unfortunately, it is more difficult to collect enough NOE restraints to determine the structure for

larger systems due to ambiguity arising from resonance overlap, inability to fully assign side chain resonances, and simple lack of side chain data due to deuteration.

Additional NMR conformational restraints are often used in addition to NOEs to assemble the 3D structure. Backbone dihedral angles, useful structural restraints to define secondary structure, are determined from the backbone resonance assignment using amides, CA, CB, and CO chemical shifts by comparing the chemical shifts to those of short stretches of peptides in a database using TALOS-N[4]. Residual dipolar couplings (RDCs) rely on the reintroduction of a residual alignment (~0.1%), so the molecules still tumble quickly to obtain high-resolution solution NMR spectra, but are aligned just enough to measure residual anisotropic information, such as dipolar coupling, to obtain orientational restraints. The key advantage of RDCs is that these orientational restraints are measured relative to the common molecular frame and are thus good complements to NOE distance restraints which may accumulate errors in the process of assembling local structural features[5].

Early NMR structure determinations of small proteins used distance geometry (based on a large amount of NOEs) to calculate structures. For large systems, NMR data alone would not be enough to determine all the atomic coordinates in the molecule. During structural calculation, usually a primitive force field is used where the ideal covalent structure is maintained and only the steric repulsion is active. Then all the NMR restraints are converted into energetic potentials to guide the structure calculation towards energetic minimums. NMR structure determination from resonance assignment to structural calculation is a highly iterative process as restraints are incorporated and structures refined.

Methyl labeling – an approach to overcome the challenges of large proteins

NMR is one of the three major techniques, the other two being X-ray and cryoEM, that could lead to high-resolution structures of biomolecules. In fact, NMR-derived structures account for 10% of the PDB archive. However, an NMR-derived structure is not an image of the molecule in the sense that an X-ray structure is. Rather, the process of structure determination using NMR is to assemble a lot of indirect structural evidences into 3D structures[3]. So as the molecule gets larger and more complicated, NMR structure determination becomes significantly more challenging. This is one major reason why most of the structures solved by NMR tend to be small (Fig. 3.2). Besides, for solution NMR, large molecules tumble slowly in solution. This leads to a fast decay of NMR signals therefore low signal-to-noise ratio and low resolution of resulting peaks. Meanwhile, the large number of signals present severe peak overlap problem in NMR spectra. So structure determination by NMR is challenging for systems larger than 25 kDa.

One strategy for overcoming the challenges of studying large systems by NMR is the use of methyl labeling. Isotopic labeling of methyl groups of Ile, Leu, and Val residues provides a nice opportunity for side chain resonance assignment and restraints collection for high molecular-weight systems by introducing a limited set of protons to provide distance restraints within an otherwise highly deuterated molecule. Due to the fact that the three protons of each methyl group are equivalent as a result of the rapid rotation around the threefold axis, NMR experiments on methyl groups deliver high sensitivity and favorable relaxation properties even for really high molecular-weight systems[6]. Also methyl groups tend to localize to the hydrophobic cores of proteins so they are excellent reporters of structures and dynamics, especially for soluble proteins. The linearized labeling scheme, where only Ile-δ1 and one of the two methyl groups of Leu and Val are $^{13}CH_3$-labeled using α-ketoacids, paved the way for NMR assignment

experiments designed to link methyl groups to Cα/Cβ/C' or even to NH, which usually are already assigned from backbone-walk experiments[7].

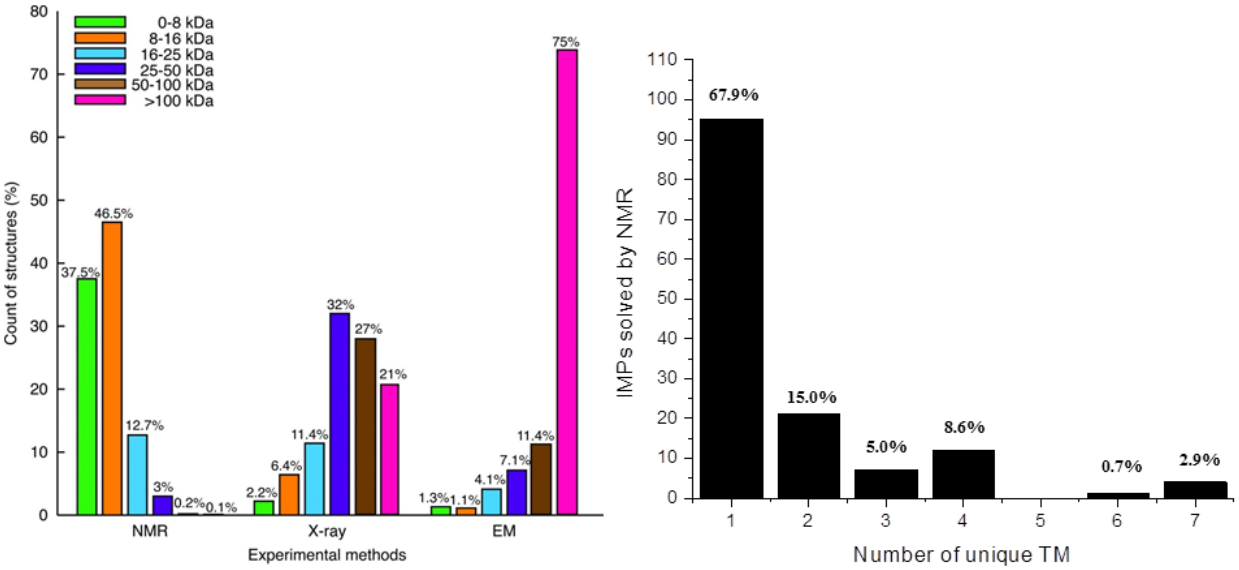


Fig. 3.2 NMR structure determination is challenging for membrane protein systems. Left, the molecular weight distributions of structures solved by NMR, X-ray, and cryoEM. Figure from Ferella *et al.*, 2012. Right, the distributions of integral membrane protein structures solved by NMR (both solid-state and solution NMR) according to the number of unique transmembrane helices (TM). The statistics was based on drorlist on membrane protein of known structure by NMR as of Apr. 15th, 2015.

Challenges of NMR structure determination of membrane proteins

Currently, there are fewer than 200 membrane protein structures that have been solved by NMR, as a result of the following factors. First of all, membrane protein biochemistry is difficult, which is why there has only been a burst of membrane protein structures determined in last decade by any structural method. Second, membrane proteins need a membrane-mimetic environment to be folded and functional. Thus, even a small membrane protein behaves as a large one due to the overall size of the protein-micelle/bicelle/nanodisc/lipodisq complex, and all the challenges involved in solution NMR studies of large systems apply. Alpha helical integral membrane proteins belong to an even more challenging subclass, because the chemical shift dispersion is

generally limited and there are fewer long-range tertiary interactions available to define the 3D structure. Solid-state NMR spectroscopy has made great advances in the last decade and can be used to determine the backbone fold of membrane proteins in a lipid bilayer[8,9]. However, peak overlap and limited side chain data still present challenges, as in solution NMR approaches.

Challenges involved in NMR structure determination of EmrE

There are additional layers of difficulty in the NMR structure determination of EmrE. Although EmrE has only 110 amino acids in the primary sequence, the functional unit is an antiparallel dimer. So in the structure of EmrE, the number of unique transmembrane (TM) helices is eight which is daunting (Fig. 3.2) considering only five integral membrane protein structures with 6 or 7 TM helices have been solved thus far using NMR approaches[10-14]. What is more, although the two monomers assume different conformations in the asymmetric dimer of EmrE, the two sets of TM helices give similar NMR signals in many places, complicating resonance assignment. While these increase the difficulty, there is a final major hurdle that makes NOE-based structure determination of WT EmrE in complex with TPP^+ not feasible: the constant conformational interconversion of EmrE between open-in and open-out states. This conformational interconversion process, occurs at 5 s^{-1} in our NMR conditions, so cross-peaks from conformational interconversion will build up in the NOESY spectra with a much shorter mixing time than necessary for NOE build up and measurement (up to a few hundred ms). Therefore, we focused on solving the structure of the S64V mutant of EmrE where it is possible to get useful NOE distance restraints due to its reduced rate of conformational exchange.

In what follows, I summarize our efforts towards a better structure of EmrE using the S64V-EmrE bound to TPP^+ in bicelles.

Materials and methods

Protein expression and purification EmrE was expressed, purified, and reconstituted into isotropic bicelles as previously described[15-17] with some modifications. When OD_{600} was close to 0.9-1.0, IPTG was added into cell cultures which were cooled down on ice[18]. This was found to increase EmrE yield. During the detergent extraction stage, the membrane fraction from the first high-speed spin was solubilized with 40 mM decylmaltoside (DM, Anatrace) overnight to increase protein extraction by detergent for high yield samples.

Sample preparation NMR data, unless stated otherwise, were collected on samples containing 0.8-1.5 mM EmrE in DMPC/DHPC bicelles (q = 0.33) with a protein to DMPC molar ratio of 1:50-100 at pH 7, 45 ˚C. Samples also contained 100 mM MOPS, 10-30 mM NaCl, 2 mM TCEP, and 8-10% D_2O. The use of 100 mM MOPS instead of the previously used 20 mM phosphate buffer provided higher pH buffering capacity crucial for long time NMR experiments, and also reduced the conductivity of solution boosting the NMR sensitivity using cryoprobes[19]. For TPP^+-bound EmrE, 1-2 mM TPP^+ was added to saturate EmrE[20].

$^2H,^{15}N$ samples for 2D TROSY-HSQC, ZZ-exchange experiments, H/D exchange, and temperature titration were isotopically labeled by growing cells in D_2O M9 media supplemented with 2 g/L glucose, 1 g/L $^{15}NH_4Cl$, and 0.5 g/L $^2H,^{15}N$ ISOGRO (Sigma-Isotech).

$^2H,^{15}N$ samples for water accessibility experiments had 0, 1, or 5 mM paramagnetic gadobenate dimeglumine[15].

$^2H,^{15}N$ samples for pH titration were prepared similarly but the final buffer system was 20 mM acetate, 50 mM MOPS, and 50 mM bicine to buffer across a broad range of pH[21].

A $^{2}H,^{13}C,^{15}N$ backbone-walk sample was prepared from D_2O M9 media containing 1.5-2 g/L $^{2}H,^{13}C$ glucose, 1 g/L $^{15}NH_4Cl$, and 0.5 g/L $^{2}H,^{13}C,^{15}N$ ISOGRO. This sample was reconstituted into D54-DMPC/D22-DHPC bicelles with extensively deuterated acyl chains.

1-^{13}C amino acid selective labeling samples were produced by growing cells in D_2O M9 media supplemented with 1.5-2 g/L glucose, 1 g/L $^{15}NH_4Cl$, and 0.5-1.0 g/L $^{2}H,^{15}N$ ISOGRO. Before IPTG induction, each 1-^{13}C amino acid was added into the media and incubated for 15-30 minutes. For each amino acid labeling, cell cultures were supplemented at least four fold of the amount of that amino acid present in ISOGRO[22]. That is: glycine (240 mg/L with 1 g/L $^{2}H,^{15}N$ ISOGRO), alanine (260 mg/L with 1 g/L $^{2}H,^{15}N$ ISOGRO), valine (300 mg/L with 1 g/L $^{2}H,^{15}N$ ISOGRO), leucine (104 mg/L with 0.5 g/L $^{2}H,^{15}N$ ISOGRO), isoleucine (200 mg/L with 1 g/L $^{2}H,^{15}N$ ISOGRO), and tyrosine (39 mg/L with 0.5 g/L $^{2}H,^{15}N$ ISOGRO). None of these amino acid labeling samples resulted in significant scrambling.

A 50% randomly deuterated ^{15}N sample was prepared by growing cells in ~50% deuterated M9 media supplemented with 2 g/L glucose, 1 g/L $^{15}NH_4Cl$, and 0.5 g/L $^{2}H,^{15}N$ ISOGRO.

A 50% randomly deuterated $^{13}C,^{15}N$ sample was prepared by growing cells in ~50% deuterated M9 media supplemented with 2 g/L ^{13}C glucose, 1 g/L $^{15}NH_4Cl$, and 0.5 g/L $^{13}C,^{15}N$ ISOGRO. The sample was reconstituted into D54-DMPC/D22-DHPC bicelles.

A $^{1}H,^{13}C,^{15}N$ sample was prepared by growing cells in H_2O M9 media supplemented with 2 g/L $^{1}H.^{13}C$ glucose, 1 g/L $^{15}NH_4Cl$, and 0.5 g/L $^{1}H,^{13}C,^{15}N$ ISOGRO. The sample was reconstituted into D54-DMPC/D22-DHPC bicelles. The final sample was in D_2O based buffer.

A Methyl-CHD_2 sample was prepared by growing cells in H_2O M9 media supplemented with 2 g/L $^{2}H,^{13}C$ glucose[23], 1 g/L $^{15}NH_4Cl$, and 0.5 g/L $^{1}H,^{13}C,^{15}N$ ISOGRO. The sample was reconstituted into D54-DMPC/D22-DHPC bicelles.

A methyl protonated I(δ1 only), L($^{13}CH_3,^{12}CD_3$), V($^{13}CH_3,^{12}CD_3$), U-[$^{2}H,^{13}C,^{15}N$] ILV sample for methyl assignment was prepared by growing cells in D_2O M9 media supplement with 2 g/L $^{2}H,^{13}C$ glucose, 1g/L $^{15}NH_4Cl$, and 0.2 g/L $^{1}H,^{13}C,^{15}N$ ISOGRO. Before the induction, 120 mg/L alpha-ketoisovaleric acid (1, 2, 3, 4-^{13}C4, 99%; 3, 4', 4', 4'-D4, 97-98%) and 70 mg/L alpha-ketobutyric acid (^{13}C4, 98%; 3, 3-D2, 98%) were added into the media[7]. The sample was reconstituted into D54-DMPC/D22-DHPC bicelles. 2 mM deuterated TPP^{+} was added to saturate EmrE.

A 1-^{13}C glucose sample was prepared by growing cells in H_2O M9 media supplement with 2 g/L 1-^{13}C glucose[24], 1g/L $^{15}NH_4Cl$, and 0.5 g/L $^{13}C,^{15}N$ ISOGRO. The sample was reconstituted into D54-DMPC/D22-DHPC bicelles.

A 2-^{13}C pyruvate alternate labeling[25] sample was prepared by growing cell in H_2O M9 media supplemented with 3 g/L 2-^{13}C pyruvate, 1 g/L $NaH^{13}CO_3$, and 1 g/L $^{15}NH_4Cl$. For pyruvate labeling, the pH of media needed to be closely monitored and adjusted before induction. The sample was reconstituted into D54-DMPC/D22-DHPC bicelles.

Samples for measuring residual dipolar couplings under the isotropic condition were prepared similarly and loaded into a Shigemi tube as most samples. For the aligned condition, the same sample was pH adjusted and then soaked into stretched polyacrylamide gels generated using a simple gel press apparatus (New Era)[26]. Briefly, the gel chamber was cleaned using nitrogen gas and parafilmed on the bottom. Then 1 ml of newly prepared 3.8% polyacrylamide gel solution

(2.1 ml 3.8% neutral gel: 1.306 ml H_2O, 0.572 ml 1.25 M pH 6.6 bis-tris buffer, 0.2 ml 40% 19:1 acrylamide/bis solution, 12 μl 10% ammonium persulfate, 10 μl 10% TEMED) was added into the chamber and 0.2 ml isopropanol was placed on top to avoid air contact. The gel was allowed to polymerize overnight before it was washed with water twice, the final buffer twice, and water twice again. Then the gel was cut on a piece of parafilm to a desired length (0.4-0.5 cm for 400 μl sample) and transferred to a petri dish which was coated with 100 μl silicone oil to reduce the contact between the gel and the surface. The gel was carefully rotated several times and let dry for a day. Then the dried gel was placed in the chamber with bottom parafilmed and the sample was added into the chamber (the final protein to long-chain lipid ratio was 1:50 which was critical for successfully soaking the protein-bicelle sample into the gel for alignment). The soaking process took at least one day. After the soaking was done, the gel press apparatus was assembled and the gel was loaded into an open-ended sample tube (connection from bottom to top: sample tube – funnel – chamber – piston (o-ring on top) – cap – piston driver). Then the piston driver was used to push the gel into the tube. After most of the gel was loaded, the chamber was unscrewed and the gel was further pushed into the tube by pipetting with 1 ml tip. At last, the bottom of the tube was sealed with an end-plug and a plunger was inserted on top.

NH RDC sample was prepared by growing cells in D_2O M9 media supplement with 2 g/L glucose, 1g/L $^{15}NH_4Cl$, and 0.5 g/L $^{1}H,^{15}N$ ISOGRO.

NC' RDC sample was prepared by growing cells in D_2O M9 media supplement with 2 g/L $^{2}H,^{13}C$ glucose[24], 1g/L $^{15}NH_4Cl$, and 0.5 g/L $^{2}H,^{13}C,^{15}N$ ISOGRO.

NMR spectroscopy and data analysis NMR data were collected at 45 ˚C on a Varian 700 MHz spectrometer with a room temperature probe unless otherwise noted. All NMR spectra were

processed with NMRPipe[27] and analyzed in CcpNmr Analysis[28]. Non-uniform sampling (NUS) was used to increase both resolution and sensitivity for all 3D experiments[29]. Possion gap sampling[30] was used to generate sampling schedule (10-50% sampling density) for NUS data collection and IST[31] and NESTA[32] were used for spectral reconstruction from NUS data. All NOESY exchange used a mixing time of 150 ms to allow cross-peaks to build up while avoiding peaks from conformational exchange.

For backbone assignment, TROSY-HNCA and TROSY-HNCACB backbone-walk experiments were collected on a $^{2}H,^{13}C,^{15}N$ labeled sample on a Varian 800 MHz spectrometer with a cryoprobe at NMRFAM. A TROSY-selected ZZ-exchange spectrum with a 200 ms mixing time was acquired for a $^{2}H,^{15}N$ EmrE sample in DLPC/DHPC bicelles (k_{conf} is 1.2 s^{-1}) and used to facilitate assignment of each pair of exchange-related amides. For each 1-^{13}C amino acid selective labeling in a $^{2}H,^{15}N$ background, 2D TROSY-HSQC, a ZZ-exchange plane with 200 ms mixing time, and the ^{1}H-^{15}N plane of TROSY-HNCO were taken. A series of TROSY-HSQC spectra were also taken using $^{2}H,^{15}N$ samples for pH titration (4.5-9.0), temperature titration (15-45 ˚C), gadobenate dimeglumine titration (0-5 mM), and H/D exchange.

For methyl assignment, a methyl protonated ILV sample was used to collect a series of 3D experiments to link methyl groups of ILV residues back to assigned Cβ, Cα, C', and NH groups[7]. In addition, a time-shared 3D C, N-HMQC-NOESY-TROSY and 3D C, N-HMQC-NOESY-HMQC were collected. All experiments involving methyl groups here used selective pulses on the methyl proton to reduce lipid signal contamination (mostly from the non-deuterated head groups of the lipids). These experiments were collected on a Varian 800 MHz spectrometer with a cryoprobe at NMRFAM.

For residual dipolar couplings (RDCs) measurement, we used intensity-based methods to measure RDCs for NH and NC' bond vectors since for large systems spectral overlap forbade frequency-based methods. We used ARTSY[33] for NH bond RDCs measurement. For both isotropic and aligned condition, the reference and attenuated spectra were collected in an interleaved manner. The delay T was set to 10.75 ms for the isotropic condition and 9.5 ms for the attenuated condition. For each condition, both reference and attenuated spectra were processed using the same script. The peak list from the reference experiment was used to find corresponding peak locations in the attenuated experiment. The peak intensity ratios of the two spectra were used to calculate coupling constants. RDCs were calculated from the difference between coupling constants of the isotropic and aligned experiments.

NC' RDCs were collected using a TROSY-based J-modulation method[34]. Due to the much smaller NC' J coupling constants, a series of modulation points were needed to more precisely fit the coupling constants. For both isotropic and aligned conditions, the modulation time points 10 ms, 50 ms, 70 ms, and 90 ms were used. For each time point, both reference and attenuated spectra were processed using the same script. The peak list from the reference experiment was used to find corresponding peak locations in the attenuated experiment. The peak intensity ratios of the reference and attenuated experiment at different modulation time points were used to fit coupling constants. The NC' RDCs were calculated as differences between fitted coupling constants from the aligned and isotropic conditions. These experiments were performed on a Varian 900 MHz spectrometer with a cryoprobe at NMRFAM.

The 50% randomly deuterated ^{15}N sample was used to collect a 3D ^{15}N-NOESY-TROSY spectrum. This was collected on a Bruker 900 MHz spectrometer with a cryoprobe at the University of Minnesota.

The ^{1}H, ^{13}C, ^{15}N sample was used to collect a 3D NOESY-HSQC (carbon aliphatics), a 3D NOESY-HSQC (carbon aromatics), a 3D HCCH-TOCSY, and a 3D CCH-TOCSY experiment. These were collected using a Varian 800 MHz spectrometer with a cryoprobe at NMRFAM.

The 50% randomly deuterated ^{13}C, ^{15}N sample was used to collect a 3D NOESY-HSQC (carbon aliphatics), a 3D NOESY-HSQC (carbon aromatics), a 3D HCCH-TOCSY, and a 3D CCH-TOCSY experiment.

The methyl CHD_2 sample was used to collect a 3D NOESY-HSQC (carbon aliphatics), a 3D NOESY-HSQC (carbon aromatics), a 3D NOESY-TROSY, and a 3D CCH-TOCSY experiments.

The 1-^{13}C glucose sample was used to collect a 3D NOESY-HSQC (carbon aromatics) and a 3D filtered/edited NOESY-HSQC for aromatic assignment and protein-ligand NOE collection.

Preliminary Xplor-NIH structure calculations were performed to evaluate the restraints needed for structure calculation. TPP^{+} force field parameters were generated using ACPYPE[35].

Molecular Dynamics Flexible Fitting and Molecular Dynamics Due to the approximately 4 Å resolution in the available crystal structure, only Cα positions are reported. The positions of the Cα atoms form the TPP^{+}-bound crystal structure (PDBID: 3D5D) were used to restrain a model of EmrE generated using a combination of PSFGEN and MODELLER, effectively using the available Cα positions as a template as the side chain and backbone shake out any instabilities. However, the initial structure lacked α-helical secondary structure throughout the hydrophobic transmembrane regions of the protein. Furthermore, the crystal structure placed a number of proline residues in the loop regions between helical domains, rather than as helix terminators

where they are far more commonly found. Thus considerable effort was initiated to use available experimental information and other membrane protein trends to refine the model to better approximate the *in vivo* state. Besides the crystal structure, there is also a cryoEM based model available, as well as the original electron density used to construct the model. This cryoEM map was combined with Molecular Dynamics flexible Fitting (MDFF) [36-38], where additional forces added to molecular simulation attract atoms to areas of electron density, to refine the initial structural model. This was achieved through two broad steps, simulation of the naked protein in a high-dielectric ($\varepsilon = 80$) implicit solvent model, and embedding this model into an atomistic membrane representation for further relaxation with additional electron density restraints. This split was done so that the large structural changes required could be done interactively. After the large structural changes were achieved, further equilibration was carried out in a membrane environment.

Results and discussion

My efforts towards a better structure of EmrE are presented in three main sections: backbone assignment, side chain assignment, and structural restraint collection. But structure determination of macromolecules using NMR is an iterative process, and spectral analyses of all three parts were done simultaneously.

Backbone assignment We have used a non-standard assignment strategy to combine backbone-walk data with ZZ-exchange, amino-acid specific labeling, solvent accessibility, and backbone NOESY data to achieve a near complete backbone assignment. That is, we have assigned 98.1% (206/210) amides, 98.6% (217/220) Cα, 92.8% (182/196) Cβ, 96.8% (213/220) C'.

The backbone assignment was initiated using a perdeutrated ^{13}C, ^{15}N sample to record TROSY-HNCA, TROSY-HN(CO)CA, and TROSY-HNCACB spectra. The use of perdeutration and TROSY experiments for this slow tumbling protein-bicelle system are crucial to suppress spin diffusion and retain favorable lineshape properties of spin systems. In addition, the use of non-uniform sampling[29] was necessary to make all 3D experiments possible considering the long experimental time needed as a result of low sensitivities of NMR experiments for this large protein-bicelle system. The most sensitive pair of backbone-walk experiments, HNCA and HN(CO)CA, were used to connect Cα(i-1) and Cα(i) resonances. The Cβ chemical shifts from HNCACB, especially those of Ala, Ser, and Thr, provided additional clues for spin system assignments. Cβ(i-1) data was used whenever available. The conformational interconversion that caused a lot of problems in our NMR experiments was leveraged here for assignment as the ZZ-exchange peaks help identify exchange-related pairs of amides (Fig. 3.3).

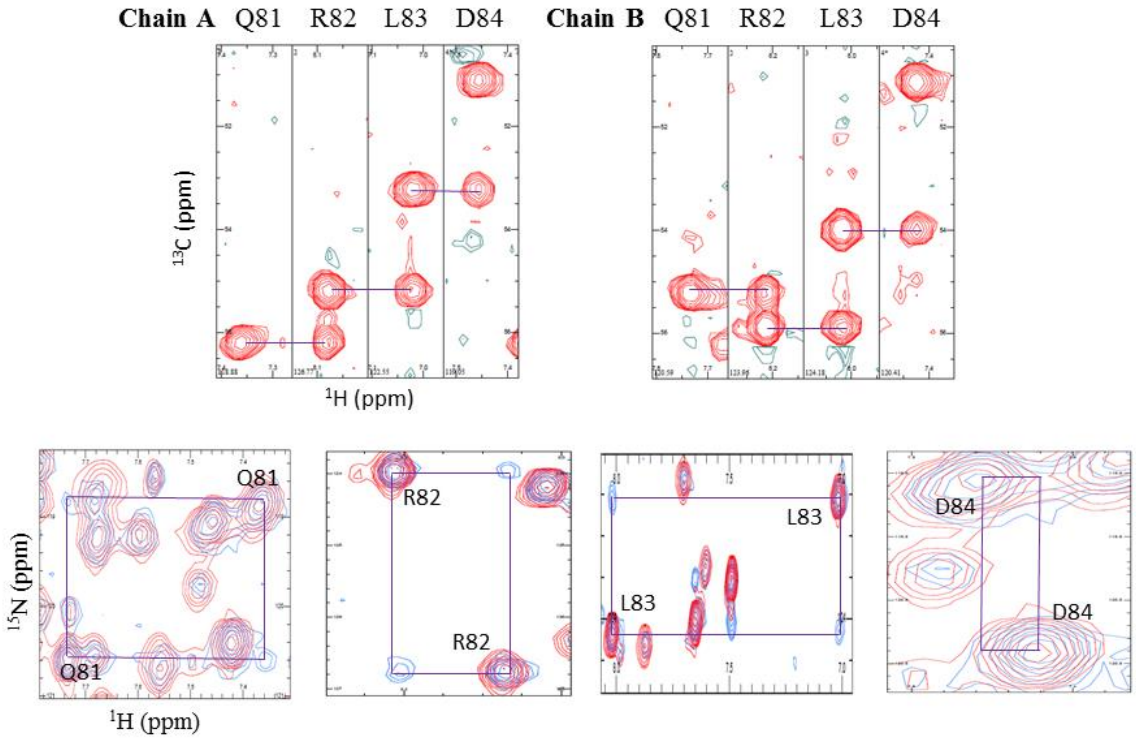


Fig. 3.3 Backbone assignment from backbone-walk and ZZ-exchange data. HNCA strips (top panel) from TROSY-HNCA experiments were shown to illustrate the concept of backbone-walk for a sequence from residue 81 to 84 (QRLD) for both chains of EmrE. In addition to that, the assignment was confirmed using ZZ-exchange data (bottom panel: red, TROSY-HSQC; blue, a ZZ plane with 200 ms mixing time). For each pair of exchange-related amides, the next two residues suggested from backbone-walk should lead to another pair of exchange-related amides.

We also used 1-^{13}C amino acid specific labeling in a ^{12}C, ^{15}N background to generate starting points for assignments[22]. This specific labeling for each amino acid used here (G/A/V/L/I/Y) generated the following information (Fig. 3.4). First, missing peaks in the TRSOY-HSQC spectrum of an amino acid labeled sample indicate the corresponding residue type of that peak is the specifically labeled amino acid. Second, in the amide plane of the HNCO experiment, only peaks from residues immediately following the labeled amino acid will appear. This amino acid specific labeling was particularly helpful in crowded regions, and provided multiple anchor points and information to fill gaps where Cα(i-1) peaks are missing in the HN(CO)CA.

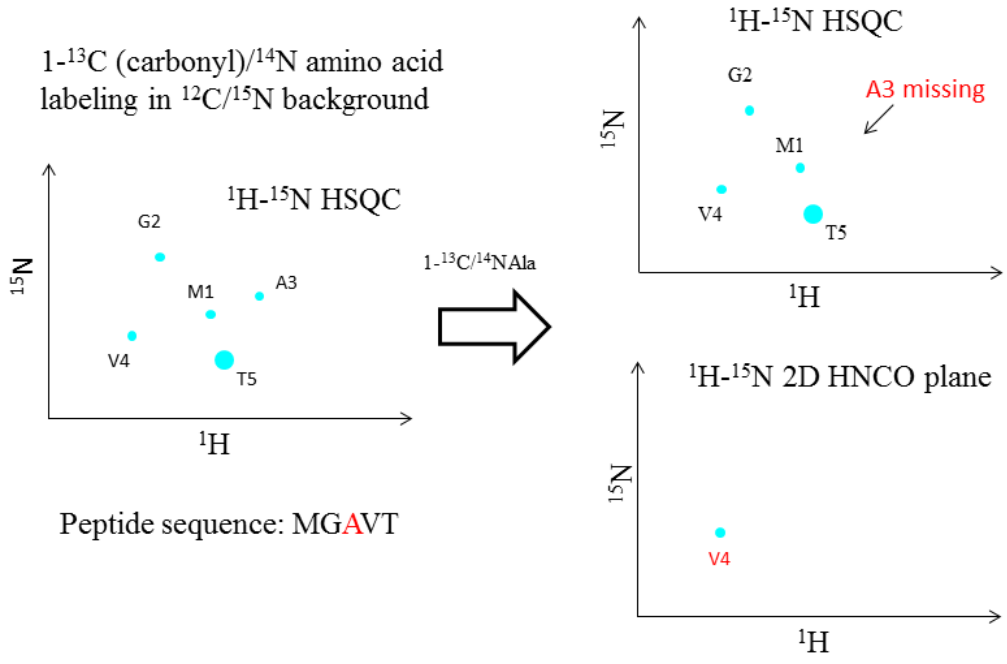


Fig. 3.4 Backbone assignment from 1-^{13}C amino acid selective labeling in ^{12}C, ^{15}N background illustrated using a model peptide system (sequence: MGAVT). The HSQC spectrum for the uniformly ^{15}N labeled peptide was

illustrated in the left. The HSQC and the amide plane of HNCO spectra were shown in right for a ^{15}N labeled sample with 1-^{13}C alanine selective labeling. Because of the use of 1-^{13}C alanine, peaks from alanine residues will disappear from HSQC. Meanwhile, only peaks corresponding to residues that are after alanine in sequence will appear in the amide plane of HNCO experiment.

As TPP^+-bound EmrE has an asymmetric architecture, we need to assign resonances to the correct monomer. Backbone-walk data will have breaks due to Pro residues, lack of signal, or overlap in the Cα chemical shift that provided the primary backbone connectivity. The overlap problem is further exaggerated by the nature of NMR of α-helical membrane proteins, which have reduced chemical shift dispersion. Therefore, ambiguity in chain assignment was a significant problem. In addition to backbone-walk data, another major source of information for chain assignment was from a NOESY walk using the N-N NOESY data. Taking advantage of the α-helical secondary structure, we used NOE peaks connecting neighboring residues as another source of reference when the chain assignment was ambiguous from available backbone-walk Cα and Cβ data. We used water accessibility data to help chain assignment in the loop regions, since the asymmetric EmrE is open to water on one side but not the other[15]. Also, while we know that the binding site of EmrE is asymmetric, our iterative assignment gradually converged to suggest one set of peaks corresponding to a single monomer is more involved in ligand binding and the other set of peaks corresponding to the other monomer is sensitive to pH [21] (and other sample condition changes). Therefore, we have used this information to revise our chain assignment.

The backbone assignment was also facilitated from side chain methyl assignment. In addition, in collaboration with Emad Tajkhorshid lab from UIUC, our assignment efforts were also supported by a MDFF refined EmrE structure using both the X-ray structure and cryoEM electron map (Fig. 3.5). Integrative assignment and modeling efforts were performed to guide

initial resonance assignment based on assignment possibilities and identify possible errors. This refined EmrE structure has a lot of structural features that are supported by our NMR data, and provides a good starting structure for final structure calculation which will dramatically reduce the sampling space.

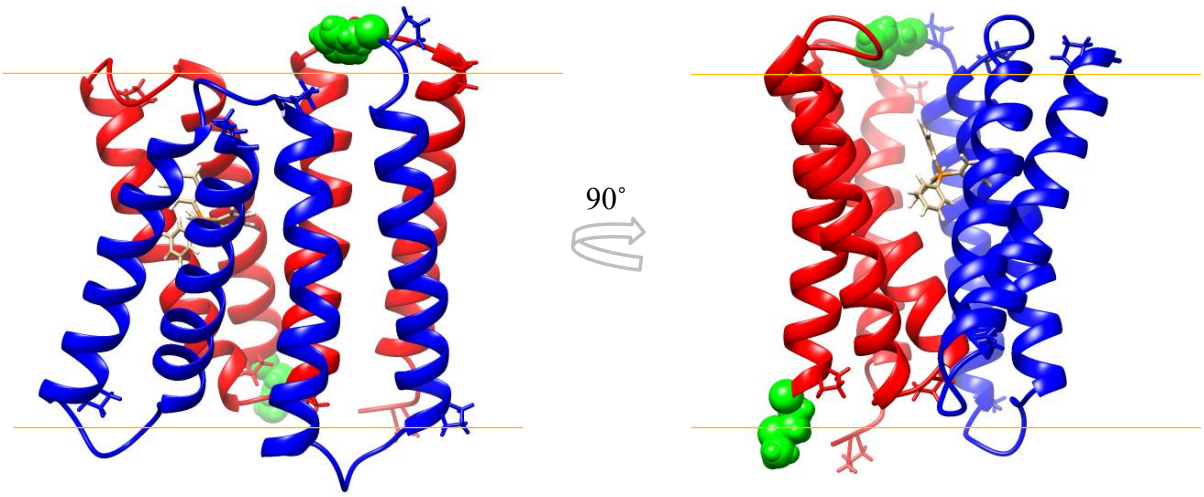


Fig. 3.5 Structural model of EmrE constructed using MDFF. The monomer A is in blue, and the monomer B is in red. A green sphere has been drawn on the C-terminus of each monomer to help identify the loops. Proline residues 3, 32, 55, and 86 are also drawn to highlight their important role in terminating helices in the refined model. Structure modeled by Josh V. Vermaas in Emad Tajkhorshid lab.

Using these strategies, we have achieved a near complete backbone resonance assignment (Fig. 3.6), ensuring high confidence in the final assignment.

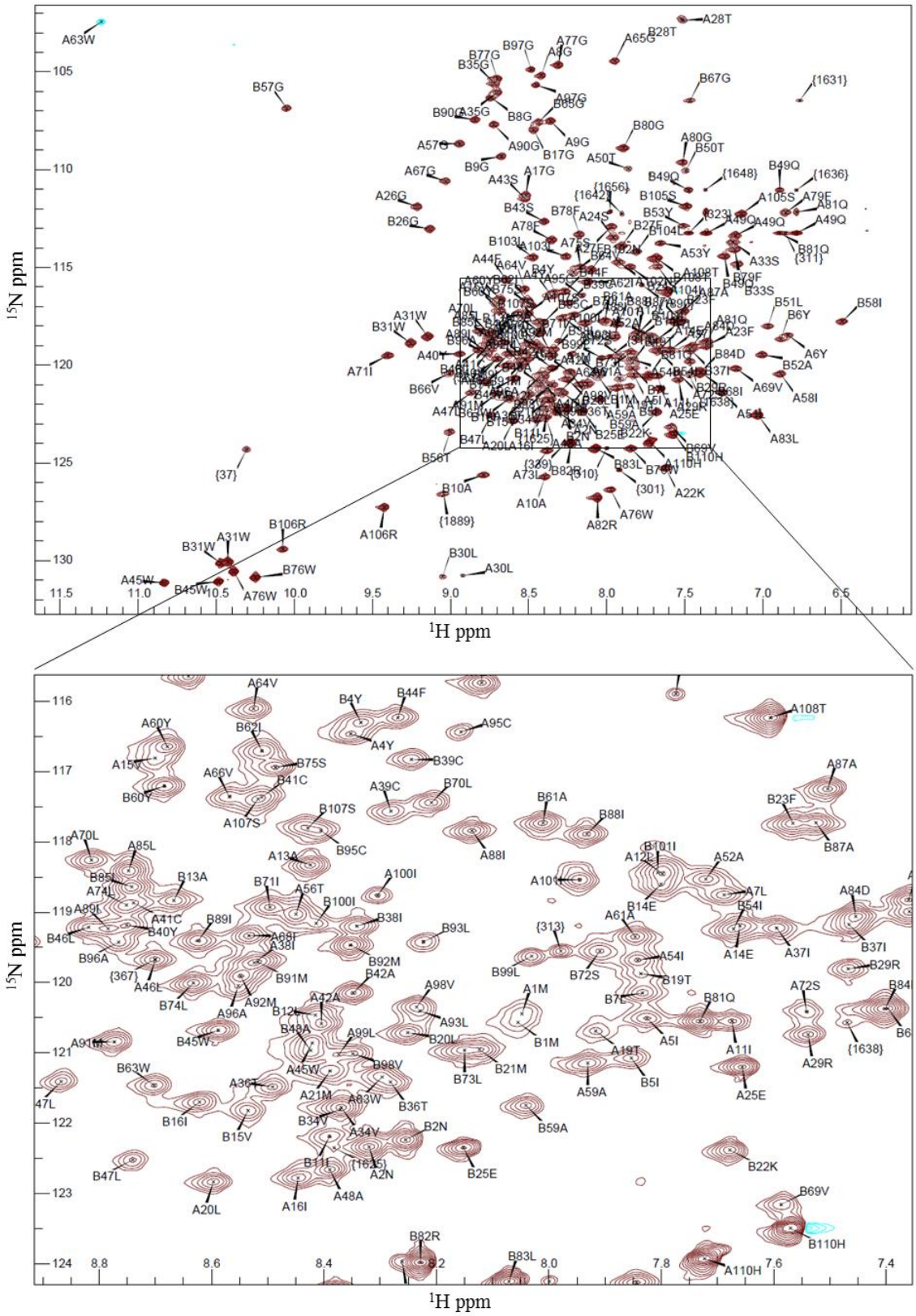


Fig. 3.6 Backbone assignment for EmrE-S64V in complex with TPP^{+}.

As previously mentioned, the chemical shifts from one monomer titrated more as the pH drifts down. From our earlier studies and assignments here (and unpublished molecular dynamics simulation studies which supports a higher pK_a value for E14 in monomer B), we know that monomer B titrates more strongly with pH (Fig. 3.7). Here I have plotted the chemical shift differences between pH 7 and 6 for residues in both monomer A and B. Immediately, there are two main regions in monomer B that respond to pH changes: one is near the E14 residue which becomes protonated and the other is the TM3 kink region. While it makes sense that peaks from residues close to E14B will titrate, the discovery that the largest chemical shift changes are from TM3 kink region in TM3 is very interesting, especially considering the chemical shift changes are much larger (e.g. more than 2 ppm change for G67 in monomer B) than usually seen in the literature. This could be rationalized by ring current effect associated with movements of aromatic residues in the binding site as well as the ligand TPP^+ with its four phenyl rings. This is mechanistically intriguing as it suggests that the protonation of E14B is coupled to the TM3 kink region, a critical region for the conformational interconversion process as will be demonstrated in Chapter 4.

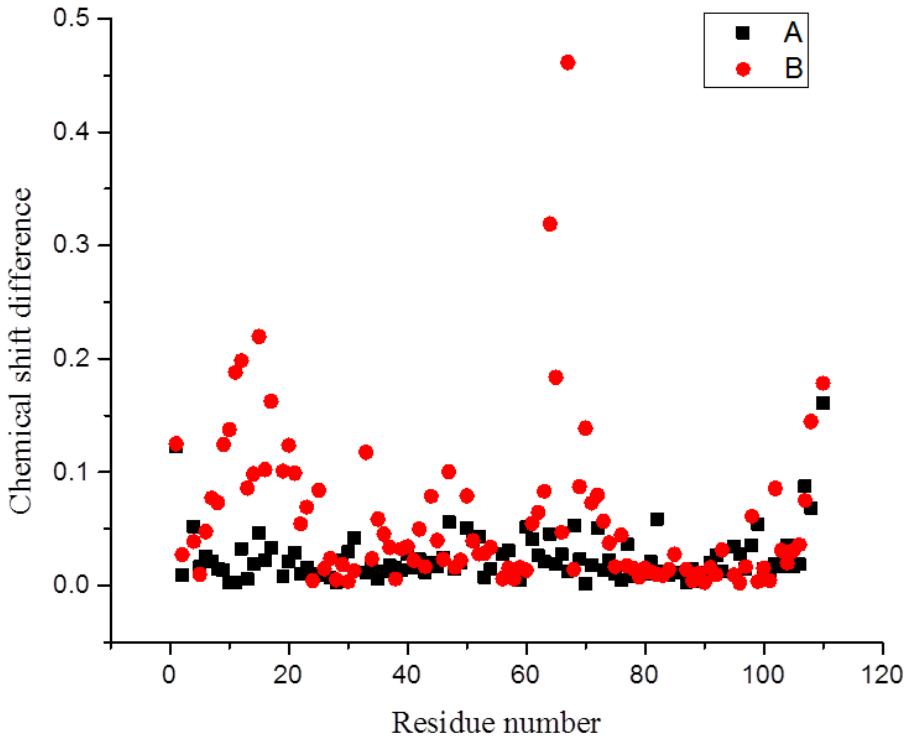


Fig. 3.7 Chemical shift differences between pH 6 and 7 as a function of residue number.

With the near complete backbone resonance assignment, it became clear that the C-terminus from monomer A samples a minor state. This is evidenced by the presence of additional minor peaks (Fig. 3.8) for residues in the C-terminus of monomer A and the loop region connecting TM3 and TM4 in monomer B, which is on the same face of the asymmetric dimer as the monomer A C-terminus. Arg106 from monomer A is very sensitive to the identity of bound ligand[17] as well as sample conditions. It is very clear that there is an additional minor state for Arg106 in monomer A but not in monomer B. This is supported by molecular dynamics simulations showing the C-terminus of monomer A is conformationally much more heterogeneous compared to the other C-terminus of monomer B. Interestingly, one molecular dynamics simulation shows the C-terminus of monomer A is extending into the binding site, interacting with residue E14B. As will be demonstrated in Chapter 5, the C-terminus of

monomer A is undergoing microsecond timescale motions in WT EmrE. This minor state might be a general feature in EmrE that is not seen in WT spectra due to lower signal-to-noise ratio and perhaps the difference in conformational exchange timescale.

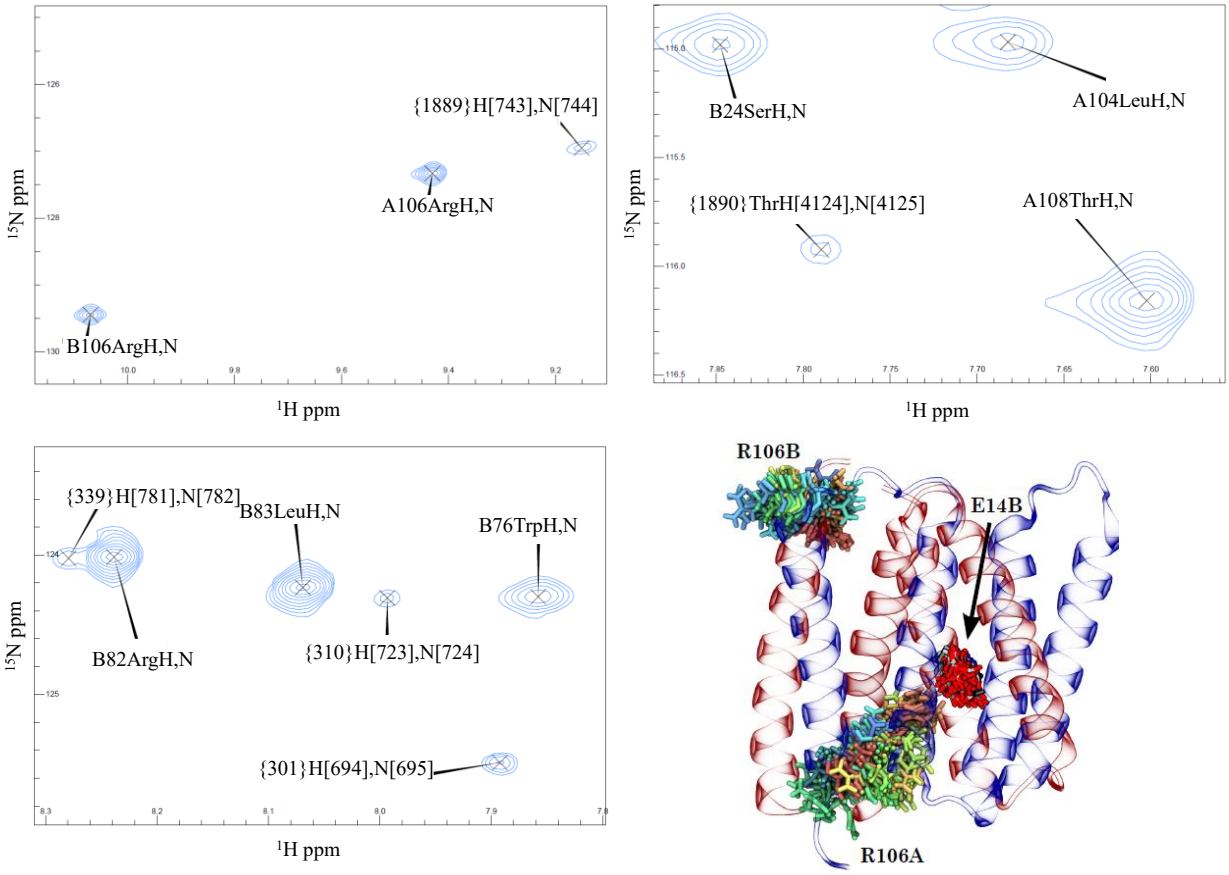


Fig. 3.8 The C-terminus of monomer A of EmrE samples a minor state. The set of minor peaks were shown for residues R106 and Thr108 in the C-terminus of monomer A, as well as residues R82B and L83B in the loop connecting TM3 and TM3 in monomer B. This minor population of the C-terminus in monomer A was also observed in molecular dynamics simulations (work by Josh Vermaas) where R106 from monomer B would protrude into the binding site and is conformationally more heterogeneous than that of the other monomer.

This observation of a minor state in the C-terminus of monomer A is very intriguing since it might have a functional role. EmrE is a proton-coupled transporter, and our recent NMR data demonstrates that it violates the long-standing single-site alternating access model and is able to interconvert between open-in and open-out in multiple protonation states, at least under

symmetric membrane conditions. If true, this would suggest that EmrE should leak protons, however, proton leak measured in liposomes is slower than expected under ΔpH and not detectable with a membrane potential. Since the C-terminus contains Arg106 and His110, it is interesting to speculate whether it might function as a cap on the binding site to prevent proton leakage. Movement of the C-terminus into the transport pore blocking water access would result in these charges effectively moving partway through the membrane potential and could alter the kinetics and thermodynamics related to the conformational exchange process[39].

Side chain assignment Restraints involving side chain groups are important to define the tertiary structure of a protein. But side chain assignment for large systems in general is very difficult. First of all, the complete side chain assignment is impossible as a result of very low sensitivity and severe spectral overlap. For partial side chain assignment, unlike backbone resonance assignment, there is a major consideration for balance between the need for deuteration and detectable protons. Higher levels of deuteration suppress spin diffusion and reduce spin relaxation since deuterium has a much smaller gyromagnetic ratio ($\gamma D/\gamma H \sim 1/6.5$), therefore improving both sensitivity and resolution. But the deuteration also reduces the number of detectable protons, therefore providing fewer restraints. As a result of the high cost and limited life of each sample, we tried a variety of labeling strategies to explore a cost-effective side chain assignment strategy for EmrE.

We first tried a fully prototonated ^{13}C, ^{15}N labeled sample to test sensitivity in TPP^+-bound EmrE-S64V bicelle system. Encouragingly, we collected reasonable 2D carbon spectra for the aliphatic region (Fig. 3.9) even for this fully protonated sample. For this sample, the HCCH/CCH-TOCSY experiments could transfer magnetization from terminal methyl to Cβ

position for ILV residues. However, as can be seen from the spectra, the proton dimension of peaks are broadened as a result of ^{1}H-^{1}H dipolar coupling in the fully protonated sample.

We have also collected TOCSY and NOESY data for a 50% randomly deuterated ^{13}C, ^{15}N labeled sample. The deuteration effectively cut off spin diffusion improving relaxation properties. But additional complications arise from the random proton incorporation into methyl groups resulting in a collection of different isotopomer species[40]. This rendered the already crowded spectra even more problematic, therefore making it hard to initiate side chain assignment from these data.

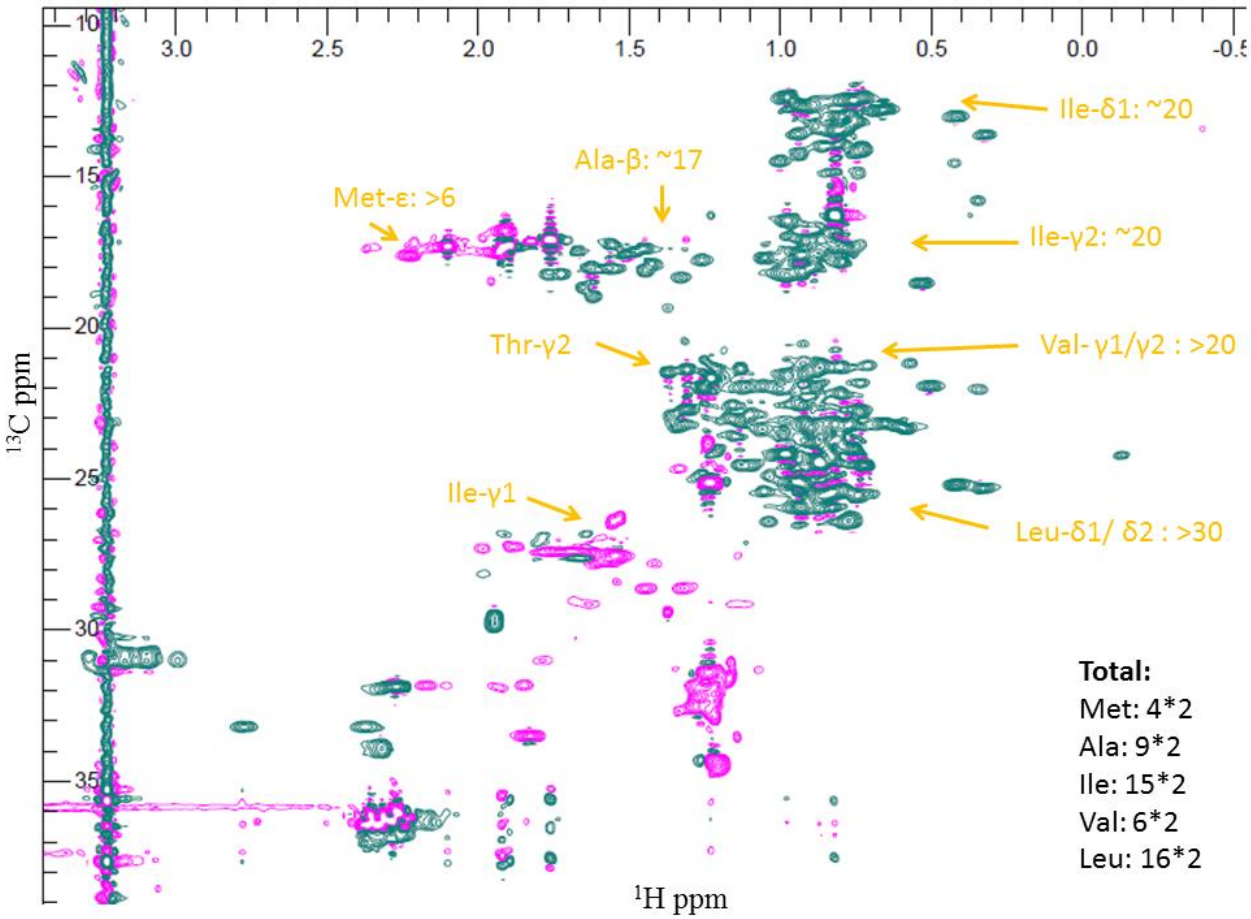


Fig 3.9 Constant-time carbon HSQC of the aliphatic region for ^{1}H, ^{13}C, ^{15}N labeled EmrE-S64V in complex with TPP^{+} at 45˚C. The spectrum was acquired on a Varian 800 MHz spectrometer with a cryoprobe.

We then tried the Otten labeling scheme[23] where proteins are expressed using ^{1}H, ^{13}C glucose in 100% D_2O. This labeling scheme has been demonstrated to generate a cost-effective sample for methyl group assignment of proteins up to 50 kDa. This labeling scheme produces proteins that have highly deuterated Cα and Cβ, and ~50% CHD_2 for terminal methyl groups. As a result of the high level of deuteration at Cα and Cβ positions, magnetization can be transferred along the side chain without much loss to the terminal methyl groups. Only CHD_2 labeled methyl groups are selected for detection so the spectrum looks relatively clean (Fig. 3.10). But the CHD_2 detection in the pulse sequence reduces the sensitivity. This is on top of the already reduced effective concentration of useful methyl isotopomer species. Again, these set of data provide useful restraints but could not be used to initiate side chain assignment.

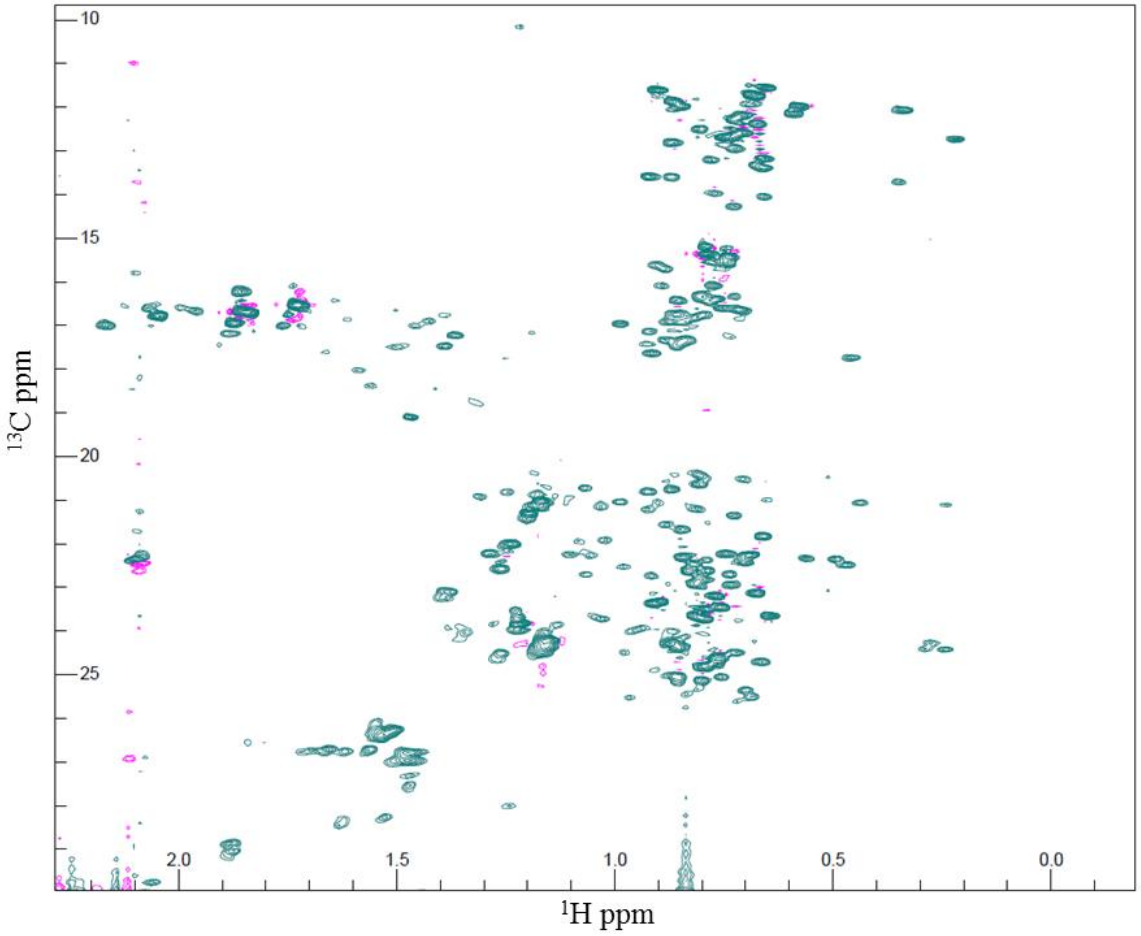


Fig. 3.10 Constant-time carbon HSQC for the methyl CHD_2 labeled EmrE-S64V in complex with TPP^+ at 45˚C. The spectrum was acquired on a Varian 900 MHz spectrometer with a cryoprobe.

Using the above samples, we have collected a lot of NOESY spectra which contain many potential distance restraints. From these efforts, we were encouraged by the signal we got from these rather proton-rich samples of the EmrE-S64V system. However, resonance assignment from these spectra turned out to be rather difficult and we realized that to initiate side chain assignment we need clean specific labeling schemes where high level of deuteration is introduced to both simply the spectra and boost sensitivity. Here I have shown the proton distribution (Fig. 3.11) for two main categories of side chains that we have decide to focused on for side chain assignment and structural restraints collection. One category includes ILV methyl

groups, and the other includes aromatics residues (WYF) important for defining the binding pocket of EmrE based on our collective NMR and modeling analysis.

Fig. 3.11 Proton distribution in EmrE important for NMR structure determination. The modeled structure using MDFF based on cryoEM map was showed in ribbon with the ligand TPP^+ (magenta) shown explicitly. Methyl groups from ILV are shown in black. Residues involved in ligand coordination in the modeled structure are shown as well with aromatics residues in ball-and-stick.

ILV methyl assignment Isotopic labeling of methyl groups of Ile, Leu, and Val residues provides a set of side chain resonance upon which assignment and restraints collection is still amenable to high molecular-weight systems. For membrane proteins, despite the fact that a significant portion of methyl groups face lipids and therefore do not provide useful restraints, there are still a large portion of methyl groups that are very useful to define helix packing and lay the foundation for assignment of other side chain resonances. So we made use of the ILV labeling scheme to initiate side chain assignment.

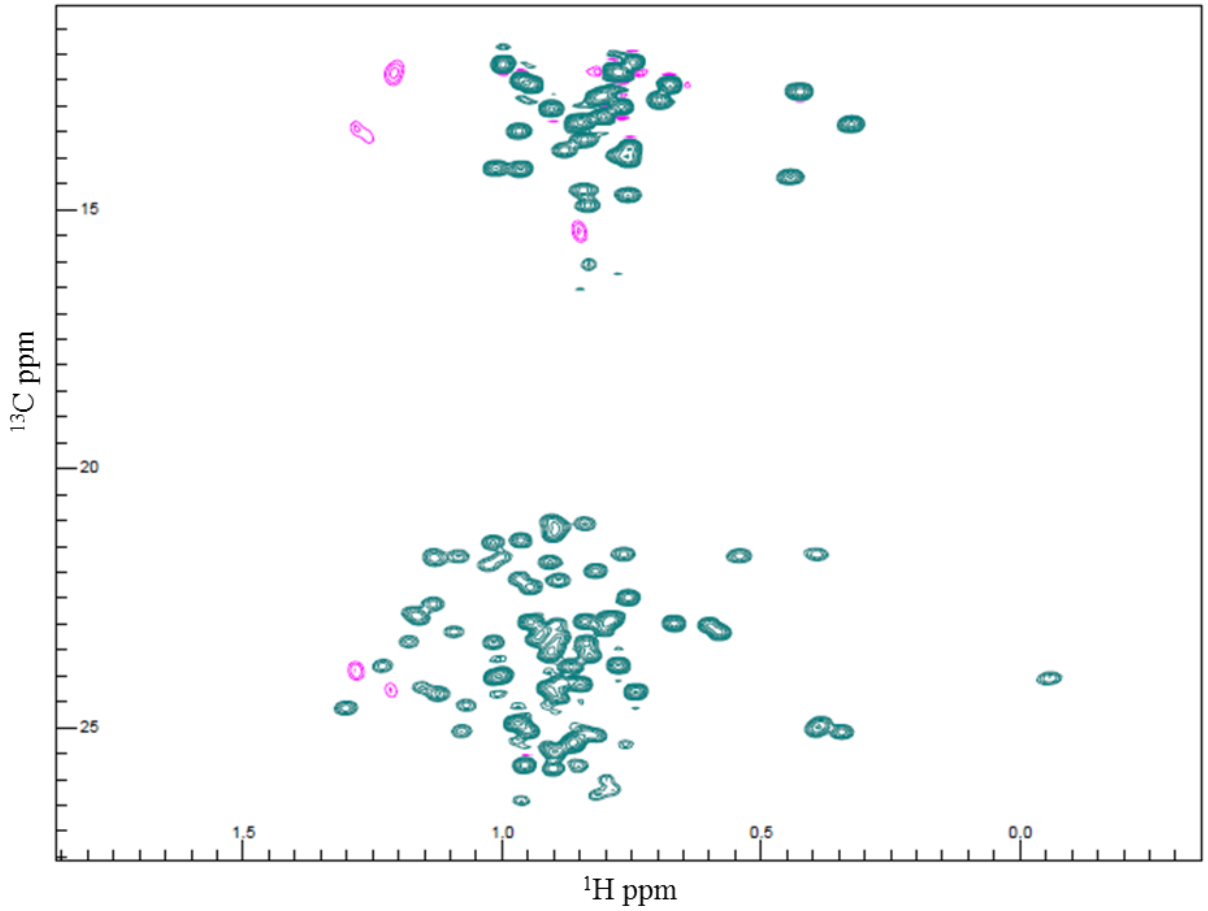


Fig. 3.12 Constant-time carbon HMQC for the selectively protonated ILV sample for EmrE-S64V in complex with TPP^+ at 45˚C. The spectrum was acquired on a Varian 800 MHz spectrometer with a cryoprobe.

A methyl protonated ILV sample was prepared which generated really clean methyl spectra for Ile-δ1, Leu-δ, and Val-γ methyl groups (Fig. 3.12). We collected a series of 3D experiments to link methyl groups of ILV residues back to assigned Cβ, Cα, C', and NH groups[7]. A 3D HMCM[CG]CB/CA experiment was performed twice; one with two C-C COSY transfer steps to reach Cβ for Leu/Ile and Cα for Val, and the other one with three C-C COSY transfer steps to reach Cα for Leu/Ile/Val (opposite sign for Val peaks) (Fig. 3.13 and Fig. 3.14). A 3D Ile/Leu-HMCM(CGCBCA)CO and a 3D Val-HMCM(CBCA)CO were collected to link ILV methyl groups to C'. These methyl out-and-back experiments were found to be rather sensitive even for

this relatively large protein-bicelle system despite the pessimistic impression from limited studies reported in literature.

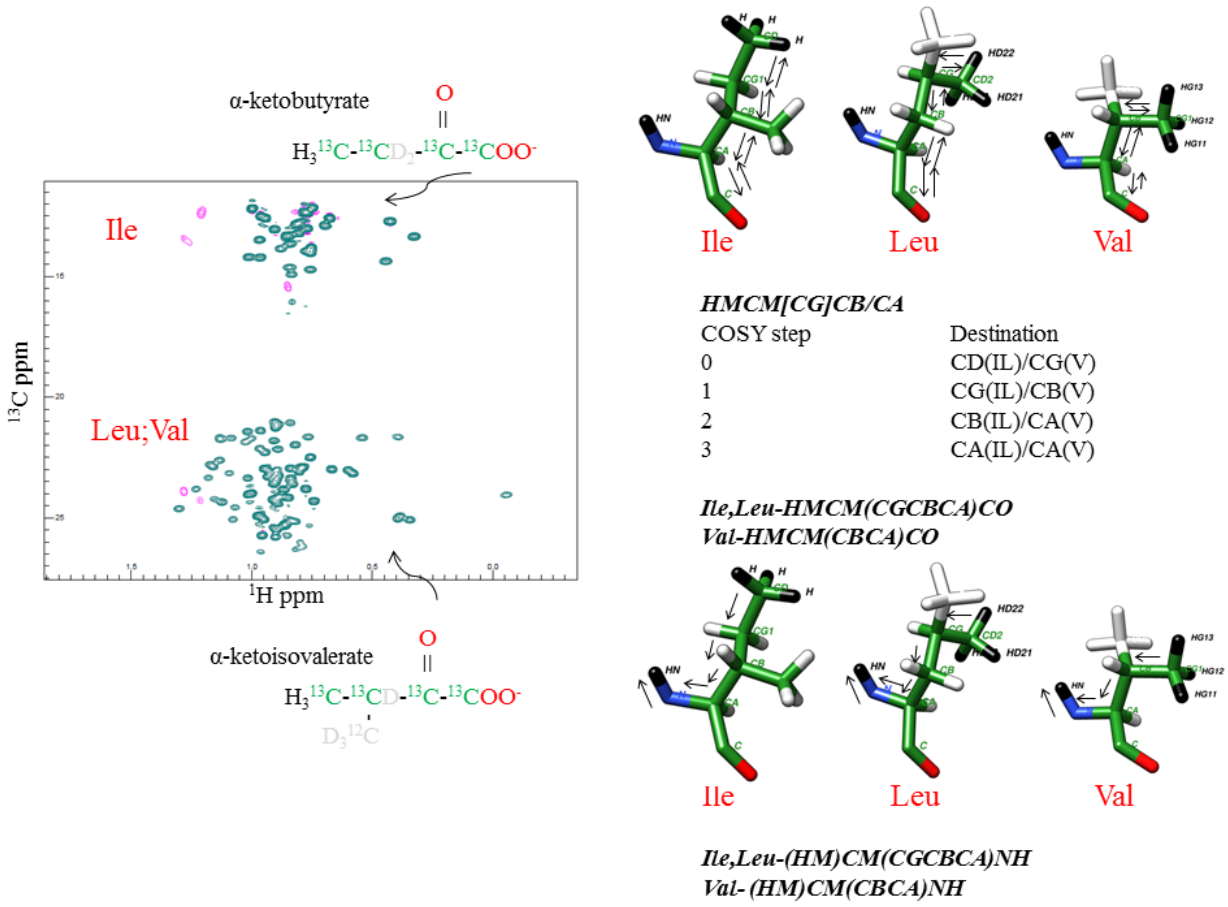


Fig. 3.13 ILV methyl assignment from selective labeling using α-ketoacids. The $^{1}H/^{13}C$ constant-time HMQC of the methyl region of a methyl protonated I(δ1), L($^{13}CH_3$, $^{12}CD_3$), V($^{13}CH_3$, $^{12}CD_3$), U-[^{2}H, ^{13}C, ^{15}N] sample of EmrE-S64V in complex with TPP^{+} at 45 ˚C was collected on a 800 MHz Varian spectrometer with a cryoprobe. The precursors used for labeling are shown: α-ketobutyrate is the precursor for Ile and α-ketoisovalerate is the precursor for Leu and Val. Two types of assignment experiments were used for methyl assignment: methyl-detected out-and-back experiments where methyl groups are linked to Cβ/Cα/C' and amide-detected experiments where methyl groups are linked to NH. Colors for the labeling: green, ^{13}C; light grey, ^{12}C and D; black, ^{1}H; blue, ^{15}N; red, O.

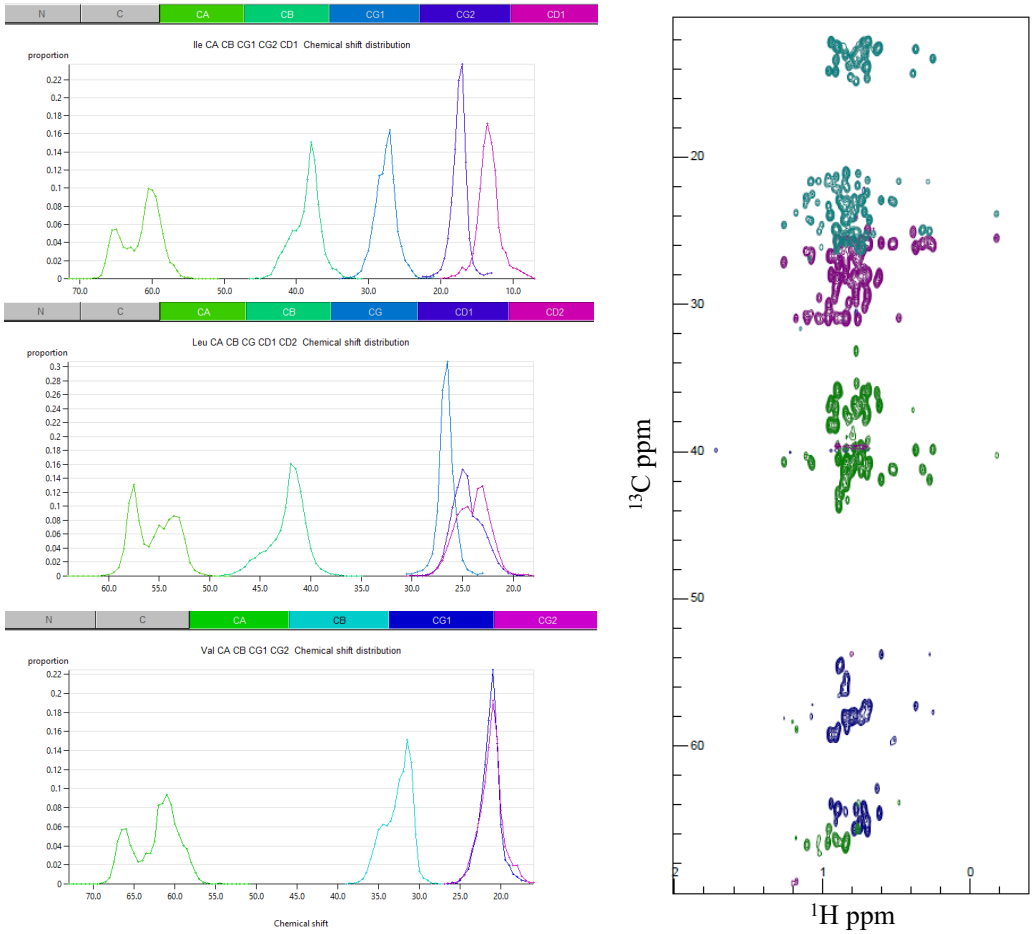


Fig. 3.14 Carbon chemical shift distributions for ILV residues and methyl out-and-back assignment experiments with different COSY transfer steps for the ILV sample of EmrE. The chemical shift distributions were from generated using CcpNmr Analysis[28] with statistics from RefDB[41]. The methyl out-and-back experiments to correlate methyl groups to Cβ/Cα were shown in right with different COSY transfer steps (0 step, teal, HMQC; 1 step, purple, Cβ for Val and Cγ for Ile/Leu; 2 step, green, Cα for Val and Cβ for Ile/Leu; 3 step, navy, Cα for ILV with Cα for Val in negative sign and buried under 2 step plane).

But the chemical shift dispersions in Cβ/Cα/C' are still limited, since for a membrane protein system most of the ILV residues face lipids and experience very similar chemical environments. So we also collected the amide-detected assignment experiments which were much less sensitive, but worked unexpectedly well for this system and provided much better dispersion through the NH plane. A 3D Ile,Leu-(HM)CM(CGCBCA)NH and a 3D Val-(HM)CM(CGCBCA)NH experiments were used to link ILV methyl groups to amides.

The matching of Cβ/Cα/C' chemical shifts of this ILV sample to those of the uniformly deuterated sample was difficult. This was a result of the limited dispersion in the carbon dimension in membrane protein NMR and the multiple-bond deuterium isotope effect[42]. Therefore, we recollected backbone-walk data including a 3D TROSY-HNCA, a 3D TROSY-HN(CA)CB, and a 3D TROSY-HNCO using this ILV sample.

Besides these through-bond COSY-transfer type assignment experiments, several NOESY spectra including time-shared 3D C, N-HMQC-NOESY-TROSY and 3D C, N-HMQC-NOESY-HMQC experiments were collected to help the assignment process.

Using the above-mentioned comprehensive set of methyl assignment experiments, we were able to reach complete assignment of all 30 Ile-δ1 and 24 Val-γ methyl groups. For Leu, we have assigned 54 out of 64 methyl groups so far and the rest of Leu methyl assignment should be possible within a week.

Our assignment for Ile δ1 methyl groups is supported by a recently published work where a series of Ile mutants were generated to assign the δ1 methyl group by recording carbon 2D spectra for each Ile-to-Leu mutant[43]. In that study, the chain assignment for these Ile methyl groups was achieved using a mixed dimer strategy where WT EmrE was mixed with I54L (or I62L), and the mutated monomer was found to prefer to stay in monomer A conformation. It is worth pointing out that our assignments are more complete than what is available from the mutagenesis work. Our NMR experiments also assigned peaks in the crowded regions and are less susceptible to potential global peak perturbations from mutation. So far, no methyl assignment for Val and Leu have been published for EmrE.

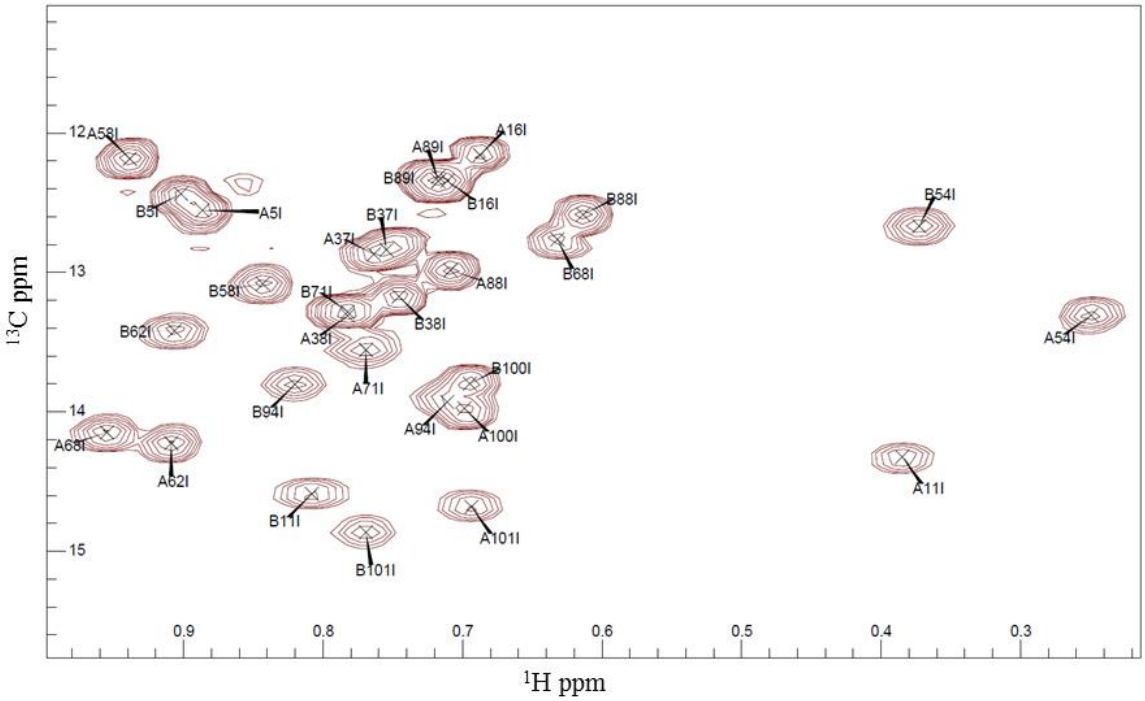


Fig. 3.15 Methyl assignment for Ile of EmrE-S64V in complex with TPP^{+}.

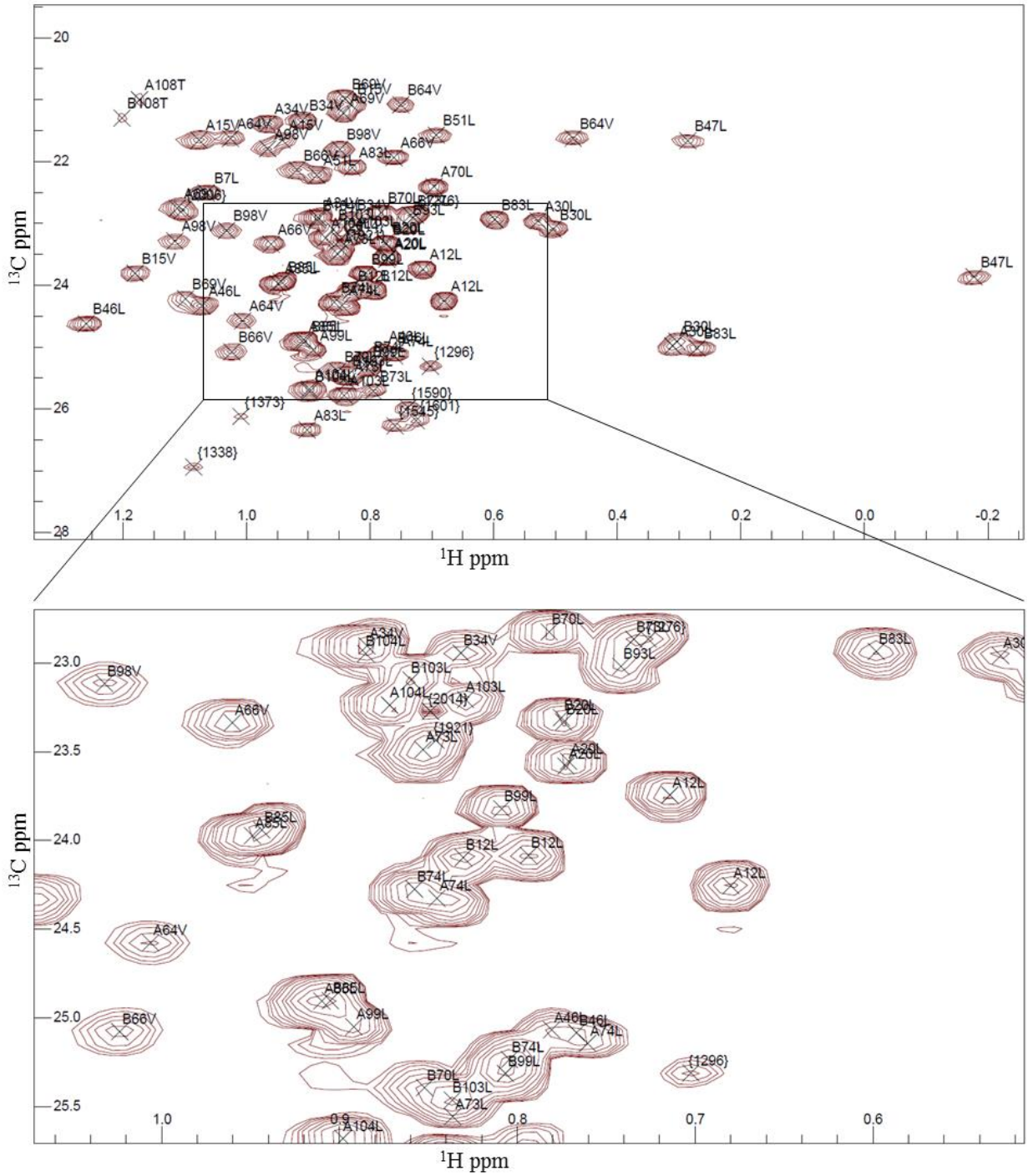


Fig. 3.16 Methyl assignment for Leu and Val of EmrE-S64V in complex with TPP^{+}.

Aromatic assignment For EmrE, aromatic groups are crucial to define the binding pocket of EmrE and protein-ligand interactions. This could be inferred even from the low-resolution

structure. The aromatic restraints were found to help pack TM helices in the NMR structure calculation of sensory rhodopsin II[10]. Thus we have invested major resources to facilitate aromatics assignment and restraints collection to improve the quality of final NMR structure. There are very limited studies for aromatic assignment, especially for large systems. Assignment experiment based on COSY transfer simply will not work as there are many transfer steps to reach already assigned resonances. This is further complicated by the large C-C coupling constants and large CSA of aromatic carbon. For aromatics, the assignment will be mainly structure-based[44]. We will rely on information from modeled structures (Fig. 3.17) and already assigned NH and methyl groups to facilitate aromatic assignment.

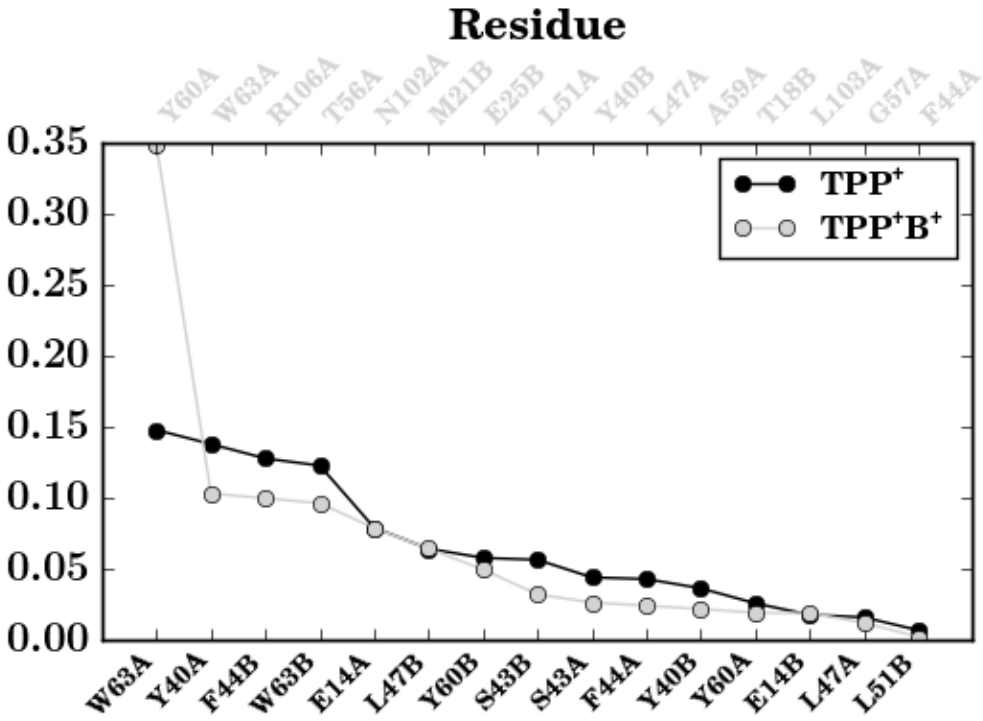


Fig. 3.17 Contact analysis between EmrE-S64V and TPP^+ from molecular dynamics simulations to facilitate structure-based aromatic assignment. The contribution of each amino acid to TPP^+ binding was plotted when the E14B was negatively charged or neutral. Work from Josh Vermaas.

We have collected a series of NOESY spectra using the protonated ^{13}C, ^{15}N labeled sample (Fig. 3.18). To improve both sensitivity and resolution for these aromatics, we have also tried alternate labeling where the large one-bond C-C coupling is eliminated and provides chances for aromatic TROSY. We have collected data using 1-^{13}C glucose labeling[45] (Fig. 3.19) and are collecting data using pyruvate labeling[25].

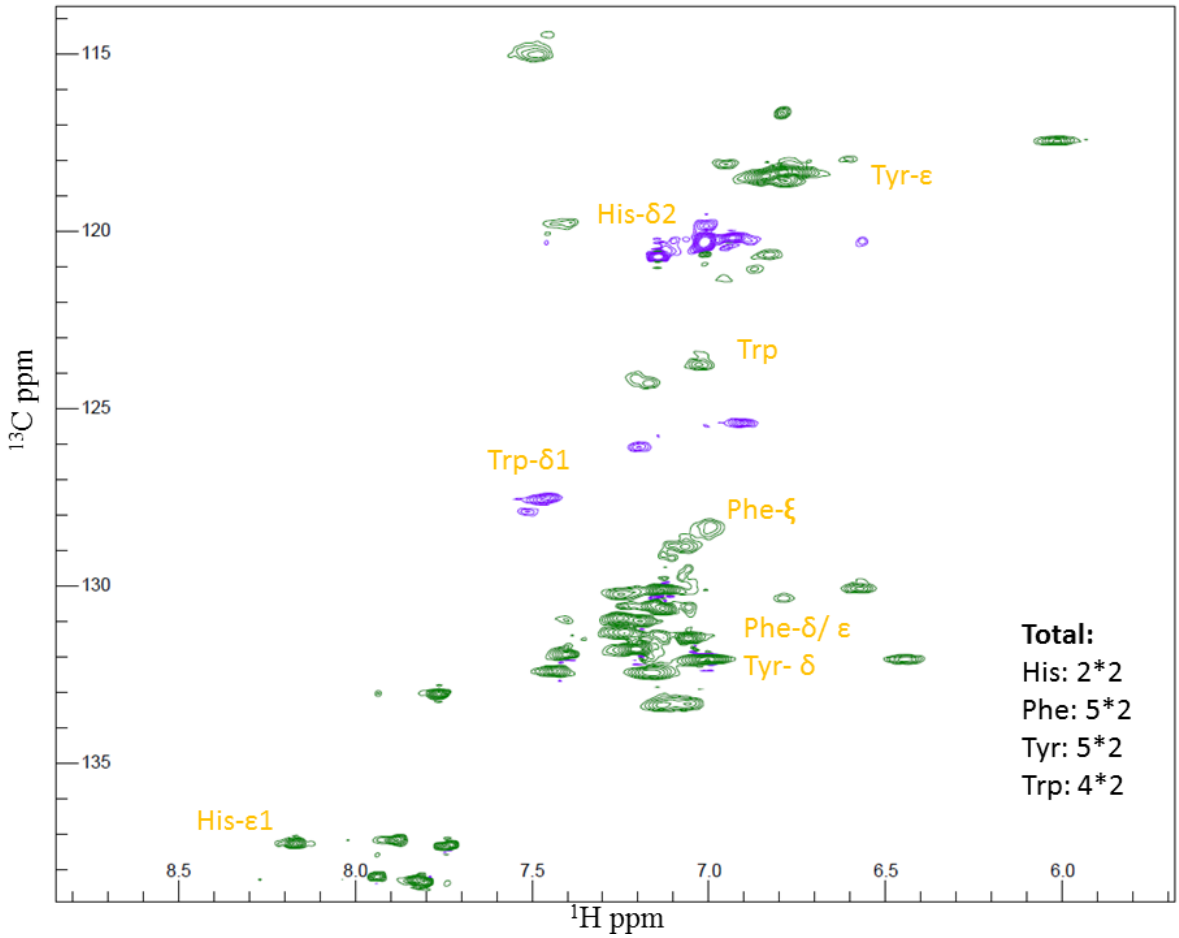


Fig. 3.18 Constant-time HSQC of the aromatic region for ^{1}H, ^{13}C, ^{15}N labeled EmrE-S64V in complex with TPP^{+} at 45˚C. The spectrum was acquired on a Varian 800 MHz spectrometer with a cryoprobe.

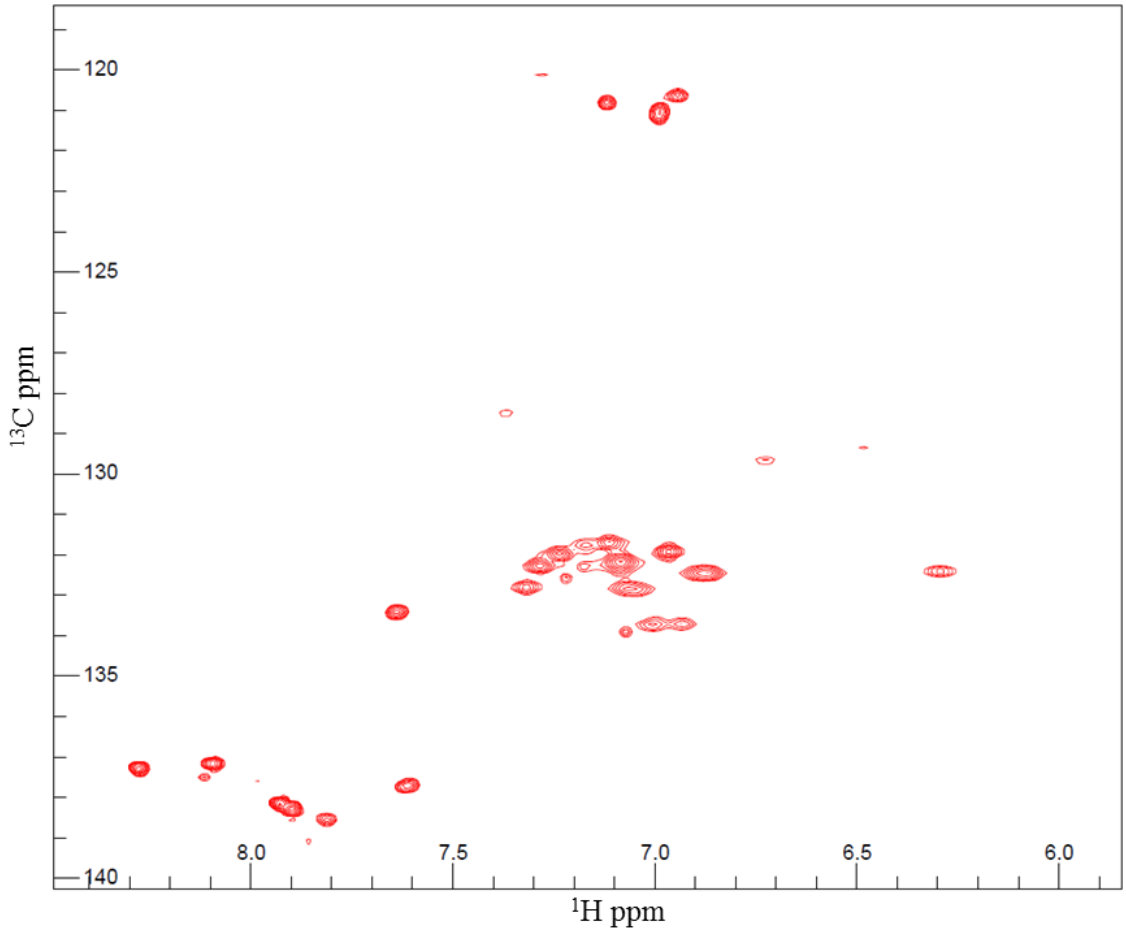


Fig. 3.19 Carbon TROSY of the aromatic region using the 1-^{13}C glucose labeling for EmrE-S64V in complex with TPP^{+} at 45˚C. The spectrum was acquired on a Varian 800 MHz spectrometer with a cryoprobe.

Structural restraints I have already collected many structural restraints. The assigned backbone resonances were used to generate backbone dihedral angle restraints using TALOS-N, which will provide a major driving force for folding the primary sequence into helices. Alongside our assignment efforts, I have already collected a great deal of NOESY data. These include NOESY-TROSY and NOESY-HSQC (for both carbon aromatics and aliphatics) for samples with different level of deuteration. We also have C,N-HMQC-NOESY-TROSY, C,N-HMQC-NOESY-HMQC which contain hundreds of restraints related to ILV methyl groups.

Also we have used spectral filtering and editing to get protein-ligand NOE using the aromatic assignment sample to help define the binding pocket. The NOEs in these spectra are being assigned on the basis of the backbone and methyl assignments and the model structure in an iterative process by Claudia Cornilescu.

In addition, I have also collected RDCs orientational restraints. RDCs contain information about bond orientations with respect to the common molecular frame, therefore RDCs provide long-range restraints. Since the application to biomolecules, RDCs have been used to refine structures derived from either NOE or PRE distance restraints[5]. Here we have measured RDCs for NH and NC' bond vectors.

In order to measure RDCs, the protein should be slightly aligned to generate residual alignment of ~0.1%. Although most alignment media (including solvent-based systems, bicelles, and phage) do not work for membrane proteins due to their incompatibility with membrane-mimetics, we show that 3.8% stretched acrylamide gel can be used to align the EmrE-bicelle system. The soaking method does produce high concentration samples. Due to the osmotic pressure of the high lipid concentration needed to solubilize EmrE, the protein to long-chain lipid ratio has been decreased to 1:50 from 1:100. Copolymerization method did not work because of the presence of detergent. However, polyacrylamide doped with charged derives was found to promote the swelling process, although the measured RDCs were degenerate with those from neutral gels.

We tried both frequency-resolved methods and intensity-based methods to get precise measurements. Frequency-resolved experiments can directly measure the splitting as a result of residual alignment, but the number of peaks that can be measured is limited by the line shape of

the semi-TROSY peaks as some peaks will be broadened out due to the constructive relaxation of dipolar coupling and chemical shift anisotropy. Intensity-based methods use TROSY peaks primarily, and calculate RDCs from intensity ratios of a reference spectrum and an attenuated spectrum, therefore providing more precise measurements for large systems like EmrE in a bicelle. For both NH and NC' RDCs measurement, we have used intensity-based methods. The NH RDCs were measured using the ARTSY pulse sequence[33] and we have over 100 high quality orientational restraints for S64V-EmrE system (Fig. 3.19). The NC' RDCs measurements proved to be much more difficult as a result of the much smaller coupling constants. We got dozens of useful restraints using an intensity modulation method[34] primarily in loops where lack of regular secondary structure makes these restraints particular useful.

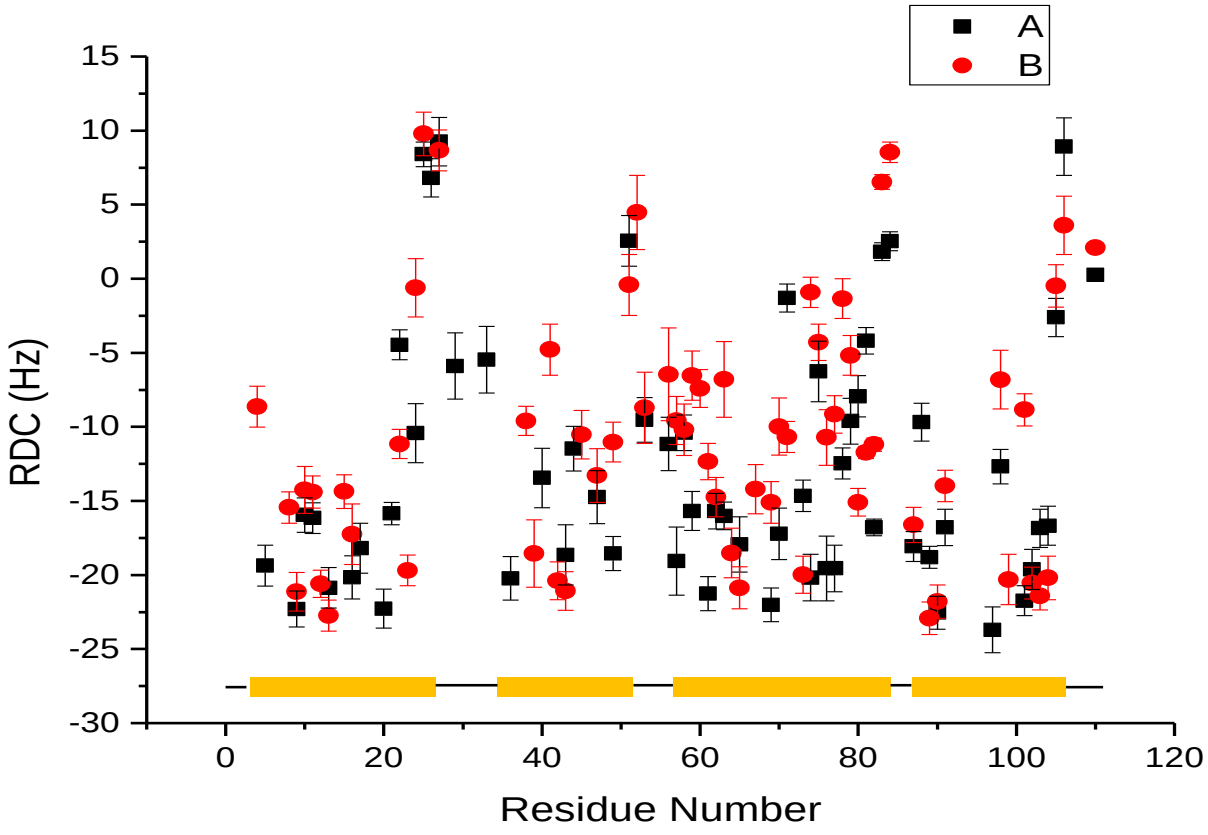


Fig. 3.20 RDCs for amide bond vectors for TPP^{+}-bound EmrE-S64V collected on a Varian 700 MHz spectrometer with a room temperature probe. The aligned condition was achieved using 3.8% neutral acrylamide gel.

Conclusion

A better structure of EmrE offers the opportunity for new insight into the multidrug transport mechanism. The conformational plasticity of EmrE, while essential for its function, makes NMR the best structural approach to obtain a better structure. However, NMR structure determination of a membrane transporter with eight unique transmembrane helices is highly challenging. Here we have made significant progress towards a better structure of EmrE using a mutant with reduced dynamics. Using the most recent cutting-edge NMR developments, we have achieved a near complete backbone assignment and side chain methyl assignment. We have also collected a great deal of structural restraints including NOEs and RDCs. Future integration of these NMR data and modeling studies for final structural calculation hold great promise for a better structure of EmrE, which will be crucial to the mechanistic understanding of this still-puzzling model system.

Author contributions

Dr. Katherine A. Henzler-Wildman and Chao Wu designed the experiments. Chao Wu collected and analyzed most of the data. Marco Tonelli was responsible for all pulse sequence testing and data collection at NMRFAM. Dr. Gregory T. DeKoster helped with local pulse sequence testing and labeling strategies for side chain assignment. Dr. Eva-Maria Uhlemann tested pyruvate labeling scheme and prepared samples for aromatic labeling. Dr. Katherine A. Henzler-Wildman and Dr. Claudia C. Cornilescu contributed to the assignment. Josh V. Vermaas from Dr. Emad

Tajkhorshid lab did the molecular modeling work. The deuterated TPP^+ was synthesized in Dr. Jim Davis lab. Chao Wu did the writing and Dr. Katherine A. Henzler-Wildman revised it.

References

1 Wüthrich, K. The way to NMR structures of proteins. *Nature structural biology* **8**, 923-925, doi:10.1038/nsb1101-923 (2001).

2 Simpson, P. J. in *Nuclear Magnetic Resonance: Volume 44* Vol. 44 348-384 (The Royal Society of Chemistry, 2015).

3 Kwan, A. H., Mobli, M., Gooley, P. R., King, G. F. & Mackay, J. P. Macromolecular NMR spectroscopy for the non-spectroscopist. *The FEBS journal* **278**, 687-703, doi:10.1111/j.1742-4658.2011.08004.x (2011).

4 Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of biomolecular NMR* **56**, 227-241, doi:10.1007/s10858-013-9741-y (2013).

5 Bax, A. & Grishaev, A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Current opinion in structural biology* **15**, 563-570, doi:10.1016/j.sbi.2005.08.006 (2005).

6 Rosenzweig, R. & Kay, L. E. Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annual review of biochemistry* **83**, 291-315, doi:10.1146/annurev-biochem-060713-035829 (2014).

7 Vitali, T. & Lewis, E. K. Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *Journal of the American Chemical Society* **125**, 1386813878, doi:10.1021/ja030345s (2003).

8 Cross, T. A., Ekanayake, V., Paulino, J. & Wright, A. Solid state NMR: The essential technology for helical membrane protein structural characterization. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **239**, 100-109, doi:10.1016/j.jmr.2013.12.006 (2014).

9 Murray, D. T., Das, N. & Cross, T. A. Solid state NMR strategy for characterizing native membrane protein structures. *Accounts of chemical research* **46**, 2172-2181, doi:10.1021/ar3003442 (2013).

10 Gautier, A., Mott, H. R., Bostock, M. J., Kirkpatrick, J. P. & Nietlispach, D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nature structural & molecular biology* **17**, 768-774, doi:10.1038/nsmb.1807 (2010).

11 Park, S. H. *et al.* Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* **491**, 779-783, doi:10.1038/nature11580 (2012).

12 Wang, S. *et al.* Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. *Nature methods* **10**, 1007-1012, doi:10.1038/nmeth.2635 (2013).

13 Reckel, S. *et al.* Solution NMR structure of proteorhodopsin. *Angewandte Chemie (International ed. in English)* **50**, 11942-11946, doi:10.1002/anie.201105648 (2011).

14 Berardi, M. J., Shih, W. M., Harrison, S. C. & Chou, J. J. Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching. *Nature* **476**, 109-113, doi:10.1038/nature10257 (2011).

15 Morrison, E. *et al.* Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature* **481**, 45-50, doi:10.1038/nature10703 (2012).

16 Morrison, E. & Henzler-Wildman, K. Reconstitution of integral membrane proteins into isotropic bicelles with improved sample stability and expanded lipid composition profile. *Biochimica et biophysica acta* **1818**, 814-820, doi:10.1016/j.bbamem.2011.12.020 (2012).

17 Morrison, E. A. & Henzler-Wildman, K. A. Transported substrate determines exchange rate in the multidrug resistance transporter EmrE. *The Journal of biological chemistry* **289**, 6825-6836, doi:10.1074/jbc.M113.535328 (2014).

18 Qing, G. *et al.* Cold-shock induced high-yield protein production in Escherichia coli. *Nature biotechnology* **22**, 877-882, doi:10.1038/nbt984 (2004).

19 Kelly, A. E., Ou, H. D., Withers, R. & Dötsch, V. Low-conductivity buffers for high-sensitivity NMR measurements. *Journal of the American Chemical Society* **124**, 12013-12019, doi:10.1021/ja026121b (2002).

20 Dutta, S., Morrison, E. A. & Henzler-Wildman, K. A. EmrE dimerization depends on membrane environment. *Biochimica et biophysica acta* **1838**, 1817-1822, doi:10.1016/j.bbamem.2014.03.013 (2014).

21 Morrison, E. A., Robinson, A. E., Liu, Y. & Henzler-Wildman, K. A. Asymmetric protonation of EmrE. *The Journal of general physiology* **146**, 445-461, doi:10.1085/jgp.201511404 (2015).

22 Takeuchi, K., Ng, E., Malia, T. & Wagner, G. 1-13C amino acid selective labeling in a 2H15N background for NMR studies of large proteins. *Journal of biomolecular NMR* **38**, 89-98, doi:10.1007/s10858-007-9152-z (2007).

23 Renee, O., Byron, C., Karla, D. K., Hans, J. V. & Frans, A. A. M. Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins. *Journal of the American Chemical Society* **132**, 2952-2960, doi:10.1021/ja907706a (2010).

24 Lundström, P. *et al.* Fractional 13C enrichment of isolated carbons using [1-13C]- or [2-13C]-glucose facilitates the accurate measurement of dynamics at backbone Cα and side-chain methyl positions in proteins. *Journal of Biomolecular NMR*, doi:10.1007/s10858-007-9158-6 (2007).

25 Milbradt, A. G. *et al.* Increased resolution of aromatic cross peaks using alternate 13C labeling and TROSY. *Journal of biomolecular NMR* **62**, 291-301, doi:10.1007/s10858-015-9944-5 (2015).

26 Chou, J. J., Gaemers, S., Howder, B., Louis, J. M. & Bax, A. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *Journal of biomolecular NMR* **21**, 377-382 (2001).

27 Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277-293, doi:10.1007/bf00197809 (1995).

28 Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696, doi:10.1002/prot.20449 (2005).

29 Hyberts, S. G., Arthanari, H. & Wagner, G. Applications of non-uniform sampling and processing. *Topics in current chemistry* **316**, 125-148, doi:10.1007/128_2011_187 (2012).

30 Hyberts, S. G., Takeuchi, K. & Wagner, G. Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data.

Journal of the American Chemical Society **132**, 2145-2147, doi:10.1021/ja908004w (2010).

31 Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H. & Wagner, G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *Journal of biomolecular NMR* **52**, 315-327, doi:10.1007/s10858-012-9611-z (2012).

32 Sun, S., Gill, M., Li, Y., Huang, M. & Byrd, R. A. Efficient and generalized processing of multidimensional NUS NMR data: the NESTA algorithm and comparison of regularization terms. *Journal of biomolecular NMR* **62**, 105-117, doi:10.1007/s10858-015-9923-x (2015).

33 Nicholas, C. F. & Ad, B. Facile measurement of ^{1}H-^{15}N residual dipolar couplings in larger perdeuterated proteins. *Journal of biomolecular NMR* **48**, 65-70, doi:10.1007/s10858-010-9441-9 (2010).

34 Liu, Y. & Prestegard, J. H. Measurement of one and two bond N-C couplings in large proteins by TROSY-based J-modulation experiments. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **200**, 109-118, doi:10.1016/j.jmr.2009.06.010 (2009).

35 Sousa da Silva, A. W. & Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* **5**, 1-8, doi:10.1186/1756-0500-5-367 (2012).

36 Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure (London, England : 1993)* **16**, 673-683, doi:10.1016/j.str.2008.03.005 (2008).

37 Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-

ray crystallography. *Methods (San Diego, Calif.)* **49**, 174-180, doi:10.1016/j.ymeth.2009.04.005 (2009).

38 Trabuco, L. G. *et al.* Applications of the molecular dynamics flexible fitting method. *Journal of structural biology* **173**, 420-427, doi:10.1016/j.jsb.2010.09.024 (2011).

39 Bezanilla, F. How membrane proteins sense voltage. *How membrane proteins sense voltage*, doi:10.1038/nrm2376 (2008).

40 Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature protocols* **1**, 749-754, doi:10.1038/nprot.2006.101 (2006).

41 Zhang, H., Neal, S. & Wishart, D. S. RefDB: A database of uniformly referenced protein chemical shifts. *Journal of Biomolecular NMR* **25**, 173-195, doi:10.1023/a:1022836027055.

42 Gardner, K. H. & Kay, L. E. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annual review of biophysics and biomolecular structure* **27**, 357-406, doi:10.1146/annurev.biophys.27.1.357 (1998).

43 Gayen, A., Leninger, M. & Traaseth, N. J. Protonation of a glutamate residue modulates the dynamics of the drug transporter EmrE. *Nature chemical biology*, doi:10.1038/nchembio.1999 (2016).

44 Hagn, F. & Wagner, G. Structure refinement and membrane positioning of selectively labeled OmpX in phospholipid nanodiscs. *Journal of biomolecular NMR* **61**, 249-260, doi:10.1007/s10858-014-9883-6 (2015).

45 Lundström, P. *et al.* Fractional 13C enrichment of isolated carbons using [1-13C]- or [2-13C]-glucose facilitates the accurate measurement of dynamics at backbone Calpha and

side-chain methyl positions in proteins. *Journal of biomolecular NMR* **38**, 199-212, doi:10.1007/s10858-007-9158-6 (2007).

Chapter 4: Transmembrane Helix Kinking in Conformational Interconversion of EmrE

Abstract

Transmembrane helix kinks are relatively common structural features of integral membrane proteins such as transporters, channels, and GPCRs. These kinks are often found in functionally important sites and act as hinges of the structural fold, which enable proteins to sample different conformational states required by their functions. In the small multidrug resistance transporter, EmrE, there is a kink in TM3 of one monomer but not the other in the asymmetric homodimer structure. Due to the way EmrE interconverts between open-in and open-out states, the kinking and unkinking of TM3 is likely central to the whole transport cycle. Understanding this mechanism will provide insight into the physical basis of the multidrug transport process.

Here I perturb the TM3 kink region using a series of mutants and ligands, and measure the interconversion rates for these systems using NMR. I find that mutations around the kink region at position 64 can dramatically alter the interconversion rates over several orders of magnitude. Dynamics data from both the mutants and the ligand series reveal the importance of hydrophobicity in determining the conformational interconversion rate, and support the idea that hydration around the kink region is very important for determining the rate of TM3 kinking and the conformational interconversion necessary for transport. This is reasonable, since backbone hydrogen bonds must break to form the kink and the energetics will depend on the effective dielectric constant in this area. This slow-down as a result of S64V mutation is even more dramatic in the absence of drug. The S64V mutation slows the conformational interconversion of drug-free EmrE from > 100 s^{-1} (WT) to ~0.5 s^{-1} (S64V mutant). This same mutation also raises

the pK_a of E14, the critical residue for coupling proton import to drug export. Thus, these results suggest the novel hypothesis that hydration provides the physiochemical basis linking proton import and drug export, as required for proton-coupled antiport.

Introduction

A common feature in helical membrane proteins is that transmembrane helices (TM) are frequently kinked[1-6]. These kinks usually locate at sites of functional importance as they provide exposed helix dipoles for specific ligand binding and increase structural complexities needed for diverse functions within a general fold space. In addition, these kinks act as hinge points to facilitate conformational sampling needed for functions.

In the small multidrug resistance transport, EmrE, there is a kink in TM3 that has been increasingly recognized as important for the conformational interconversion and multidrug recognition[7-10]. During the construction of the cryoEM model based on the electron map and sequence conservation within the Small Multidrug Resistance transporter family, a kink was located in TM3 of monomer A in the asymmetric dimer of EmrE[7]. This kink in TM3 was found to be important for the conformational plasticity of EmrE, facilitating its recognition of drugs of different shapes and charges[8]. This kink is also crucial to the conformational interconversion process central to alternating-access transport based on solid- and solution-state NMR studies[9-11].

Intrigued by the location of S64V mutation, the only slow dynamics mutation identified in our screening efforts, we set out to characterize a series of EmrE S64 mutants to probe the interconversion mechanism of EmrE as described in this chapter.

Materials and methods

Sample preparation WT and mutant EmrE were expressed, purified, and reconstituted into isotropic bicelles as previously described[9,11,12]. Single point mutants were constructed using QuickChange Mutagenesis (Stratagene).

In-cell transport assay These assays were carried out using the BL21(DE3) strain of *E. coli* transformed with empty pET15b vector or pET15b containing WT or mutant EmrE. The cells were grown in M9 minimal media with 100 μg/ml of ampicillin at 37 ˚C until the OD_{600} reached 0.4. Then cells were induced with 0.33 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) for 30 minutes at 37˚C. 2.5 μM ethidium bromide and 40 μM carbonyl cyanide p-chlorophenylhydrazone (CCCP) were added and the cells were incubated at 37˚C for an additional hour. Assays were started immediately with excess cell culture stored on ice until it was in an assay. For each experiment, 2 ml of cell culture was spun down and immediately resuspended in 1 ml fresh M9 media with 2.5 μM ethidium bromide. Ethidium bromide fluorescence was monitored with excitation at 545 nm and emission at 610 nm. The time course of fluorescence was plotted after normalization to the initial value of each run.

Isothermal titration calorimetry All isothermal titration calorimetry (ITC) experiments were performed on a TA instruments LV Nano calorimeter. Data were fit simultaneously to a model of ligand binding to n independent and identical sites plus a constant baseline due to mixing[11]. All ITC experiments were performed at least in triplicate. EmrE was reconstituted into DMPC/DHPC (q = 0.33) isotropic bicelles in 20 mM potassium phosphate, 20 mM NaCl, pH 7, 45 ºC and loaded into the cell at 40-60 μM for tetraphenylphophonium (TPP^+) titrations and 500 μM for titration with methyltriphenylphophonium ($MeTPP^+$) or ethyltriphenylphosphonium ($EtTPP^+$). TPP^+ concentration in the syringe was 150-200 μM. $MeTPP^+$ and $EtTPP^+$

concentrations were 2-4 mM. For these two weaker ligands, only duplicates were carried out, therefore, the reported errors are standard error of mean (SEM). All solutions contained buffer and bicelle concentrations matched to the protein samples.

NMR spectroscopy and data analysis NMR data were collected using samples with 0.8-1.5 mM $^{2}H,^{15}N$ EmrE in DMPC/DHPC bicelles (q = 0.33, with a protein to DMPC molar ratio of 1:50) and 100 mM MOPS, 10-30 mM NaCl, 2 mM TCEP, 8-10 % D_2O, pH 7 at 45 ˚C on a Varian 700 MHz spectrometer with a room temperature probe unless otherwise noted. All NMR spectra were processed with NMRPipe[13] and analyzed in CcpNmr Analysis[14]. For the TPP^+-bound EmrE, 2 mM TPP^+ was added to ensure EmrE saturation. For EmrE-S64V in complex with weaker ligands, 10 mM $MeTPP^+$ or 5 mM $EtTPP^+$of ligands was added. 2D $^{1}H,^{15}N$ TROSY-HSQC and TROSY-selected ZZ-exchange experiments[15] with a lipid flip-back pulse[9] were carried out with a recycle delay of 2 s and 128-144 increments. The conformational interconversion rate, k_{conf}, were analyzed from the ZZ-exchange data as previously described[16] using the composite peak ratio method with an 11.1 ms offset time, t_0, to account for the back-transfer time in the pulse sequence[11]. The composite peak ratios of intensities of the auto-peaks (I_{AA}, I_{BB}) and cross-peaks (I_{AB}, I_{BA}) were fit to the following equation as a function of the delay time, t,:

$$\Xi(\mathrm{t}) = \frac{I_{AB}(t)I_{BA}(t)}{I_{AA}(t)I_{BB}(t) - I_{AB}(t)I_{BA}(t)} \cong k_{conf}^2(t - t_0)^2. \qquad (1)$$

For TPP^+ bound EmrE, two planes were collected with different mixing times and the mixing times were adjusted according to the conformational interconversion rate of each mutant. The mixing times were 40, 60 ms for S64A, 100, 200 ms for S64T, and 200, 225 ms for S64V. For S64V bound to weaker ligands, the mixing times were 40, 60, 80, 100 ms for $MeTPP^+$ and 80,

120, 160, 180 ms for $EtTPP^+$. The standard deviation of individual fits was used to estimate the error of the rates.

pH titrations of S64V drug-free EmrE were performed on a Varian 800 MHz spectrometer with a cryoprobe using a sample prepared as described above except that ether-linked DHPC (eDHPC) was substituted for the standard ester-linked DHPC in the bicelles. The ether-linked lipid is more pH stable, so this substitution reduces pH drift in higher pH-samples that normally occur as a result of lipid hydrolysis. The DMPC was not substituted to minimize the perturbation to the protein[17]. To span a wide range of pH, a mixed buffer system was used with 50 mM boric acid, 50 mM bicine, 50 mM MOPS, and 20 mM acetate. The buffer system and eDHPC were observed to have minimal effect on the NMR spectra and interconversion rates of EmrE. For the S64V drug-free system, a single ZZ plane with a 200 ms mixing time was collected at each pH condition to estimate the interconversion rate. As a result of the increased solvent exchange, only limited cross peaks could be observed at the long mixing times required to monitor exchange in this mutant. For the pH titration data, amide proton and nitrogen chemical shifts were fit separately[18] to either one or two pK_a values using the following equations[19,20]. For the single pK_a model, the generalized Henderson-Hasselbalch equation was used to describe the chemical shift of each residue, δ, as a function of pH:

$$\delta = \frac{\delta_P 10^{-pH} + \delta_D 10^{-pK_A}}{10^{-pH} + 10^{-pK_A}}, \qquad (2)$$

where δ_P and δ_D are the chemical shift of the protonated and deprotonated state of the residue. For the two pK_a model, a generalized macroscopic model was used:

$$\delta = \frac{\delta_P 10^{-2pH} + \delta_I 10^{-(pH+pK_{A1})} + \delta_D 10^{-(pK_{A1}+pK_{A2})}}{10^{-2pH} + 10^{-(pH+pK_{A1})} + 10^{-(pK_{A1}+pK_{A2})}}, \qquad (3)$$

where δ_P and δ_D are the chemical shifts of the fully protonated and deprotonated state of the residue, and δ_I is chemical shift of the intermediate singly protonated species.

Results

EmrE-S64V is a kink region slow dynamics mutant that maintains multidrug binding. From our previous screening efforts to find dynamics mutants of EmrE suitable for structural studies, we identified EmrE-S64V as a slow dynamics mutant. As can be seen from ethidium efflux assay data in *E. coli* (Fig. 4.1), this mutation significantly impairs EmrE transport activity in its native environment, although it still export ethidium faster than the negative control. There are several possible origins of this behavior: impaired substrate binding, leaking or uncoupling of the transporter due to loss of proper structure, or reduced conformational interconversion rates between open-in and open-out states.

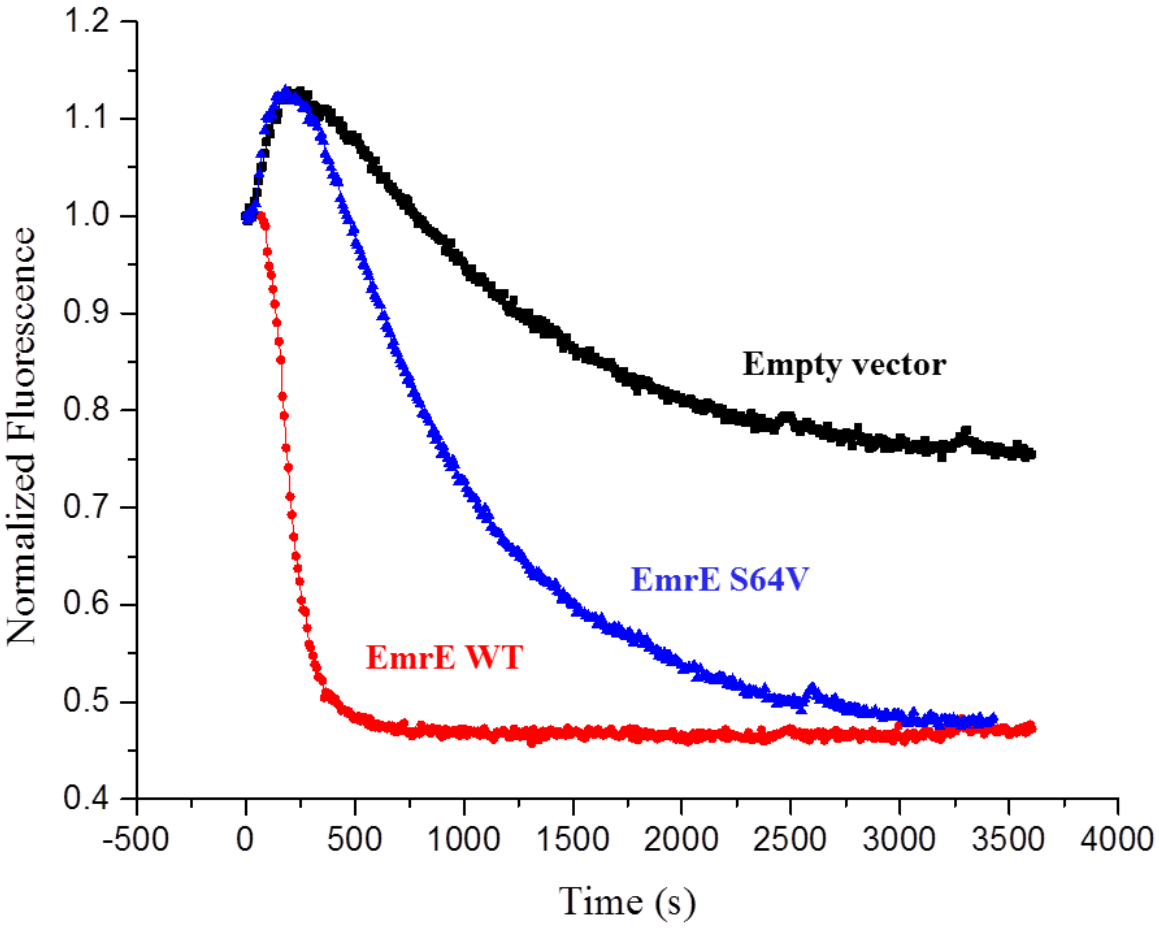
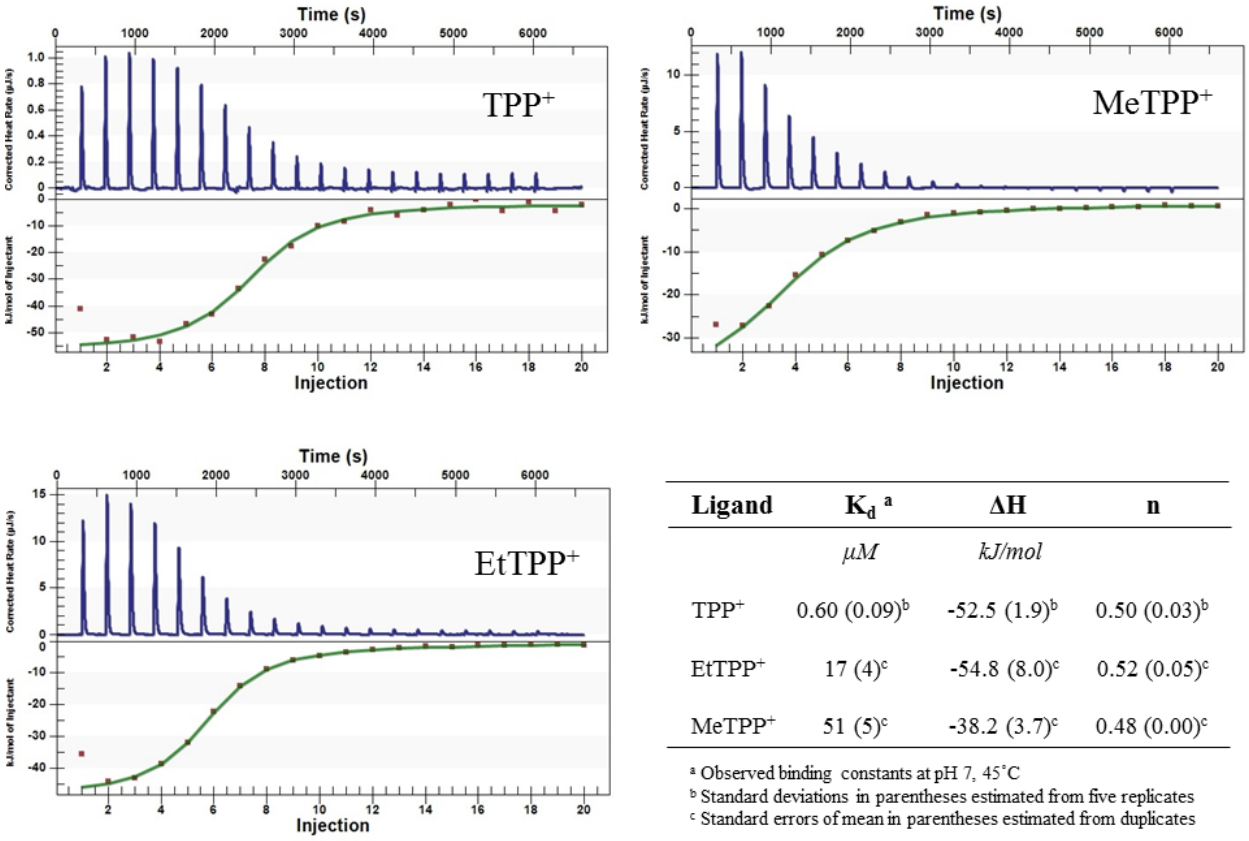


Fig. 4.1 Normalized fluorescence as a function of time from in-cell assay for EmrE-S64V. Data from the black curve were collected for *E. coli* cells transformed with empty pET15b vectors whereas data from the red and blue curve were from cells with EmrE overexpression.

This impaired transport by S64V-EmrE is mostly likely not due to impaired ligand binding since the screening process that identified this mutation selected against mutants that did not bind ligand. I performed more quantitative binding assays to confirm this. Ethidium is a weak binding ligand and the K_d is difficult to precisely determine. However, as determined by isothermal titration calorimetry (ITC) for binding to a series of TPP^+ derivatives, TPP^+, $MeTPP^+$, and $EtTPP^+$, EmrE-S64V has very similar ligand binding properties compared to WT EmrE in general. These TPP^+ derivatives were shown to have very interesting properties: both their

affinities to EmrE and the interconversion rates of EmrE bound to these ligands vary over two orders of magnitude[11]. Based on these results, it seems unlikely that the reduced ethidium export of EmrE-S64V would arise from significantly altered ligand binding. In addition, these data demonstrate that EmrE-S64V maintains the multidrug binding capability.



Ligand	K_d [a]	ΔH	n
	μM	*kJ/mol*	
TPP^+	0.60 (0.09)[b]	-52.5 (1.9)[b]	0.50 (0.03)[b]
$EtTPP^+$	17 (4)[c]	-54.8 (8.0)[c]	0.52 (0.05)[c]
$MeTPP^+$	51 (5)[c]	-38.2 (3.7)[c]	0.48 (0.00)[c]

[a] Observed binding constants at pH 7, 45˚C
[b] Standard deviations in parentheses estimated from five replicates
[c] Standard errors of mean in parentheses estimated from duplicates

Fig. 4.2 Isothermal titration calorimetry data for EmrE-S64V binding to TPP^+, $EtTPP^+$, and $MeTPP^+$. These ligands differ in only one of the four phenyl rings. These experiments were performed at pH 7, 45 ˚C in DMPC/DHPC (q = 0.33) bicelles. The first data points of these titrations were omitted for fitting since it was an artifact from titrant dilution during incubation. As a comparison, for WT EmrE in DLPC/DHPC bicelles[11], the K_d values for TPP^+, $EtTPP^+$, and $MeTPP^+$ are 0.45 ± 0.01, 21.8 ± 0.7, and 130 ± 20 μM.

The TROSY-HSQC spectrum (Fig. 4.3) of EmrE-S64V in complex with TPP^+ is strikingly similar to TPP^+-bound WT EmrE, with two sets of peaks that is characteristic of the asymmetric dimer structure. Thus, EmrE-S64V in complex with TPP^+ remains an asymmetric dimer structure

with an overall structure that must be quite similar to WT. There are global peak shifts in the vicinity of the mutation, as well as some more remote chemical shift changes as a result of the altered kink region. As observed for the WT EmrE, the TROSY-HSQC spectra of EmrE-S64V in complex with TPP^+ are also very similar in DLPC/DHPC and DMPC/DHPC bicelles confirming that EmrE is relatively structurally insensitive to the thickness of the lipid bilayer. However, the conformational interconversion rate between open-in and open-out drug-bound states is significantly different. Whereas WT EmrE in complex with TPP^+ interconverts at a rate of 4.7 ± 0.6 (4.4 ± 0.4) s^{-1} in DMPC/DHPC (DLPC/DHPC) bicelles, EmrE-S64V in complex with TPP^+ has a conformational rate of 0.6 ± 0.1 (1.2 ± 0.2) s^{-1} in DMPC/DHPC (DLPC/DHPC) bicelles. This data clearly shows that EmrE-S64V is a slow-dynamics mutant and that altered conformational interconversion dynamics is the most likely cause of the reduced ethidium efflux rate in *E. coli*.

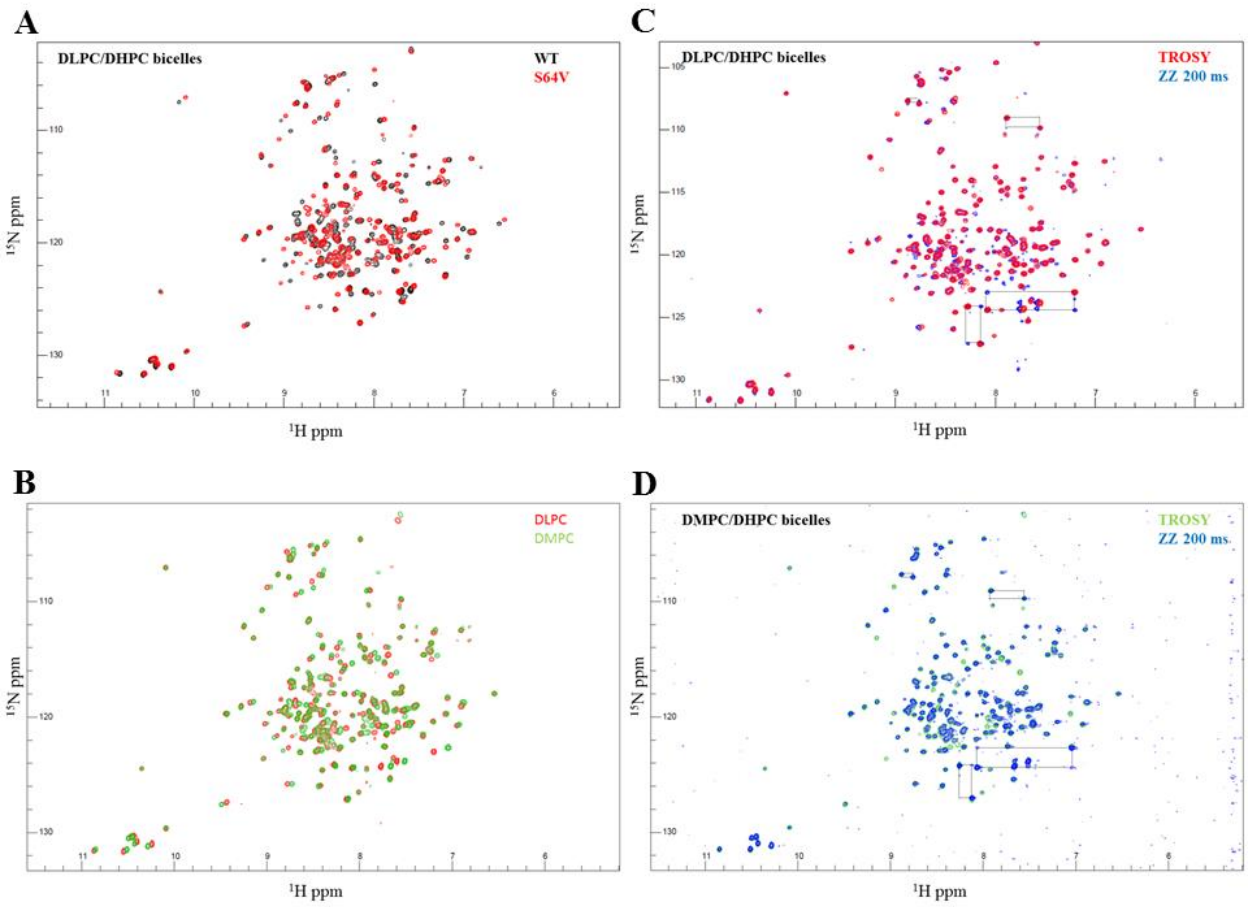
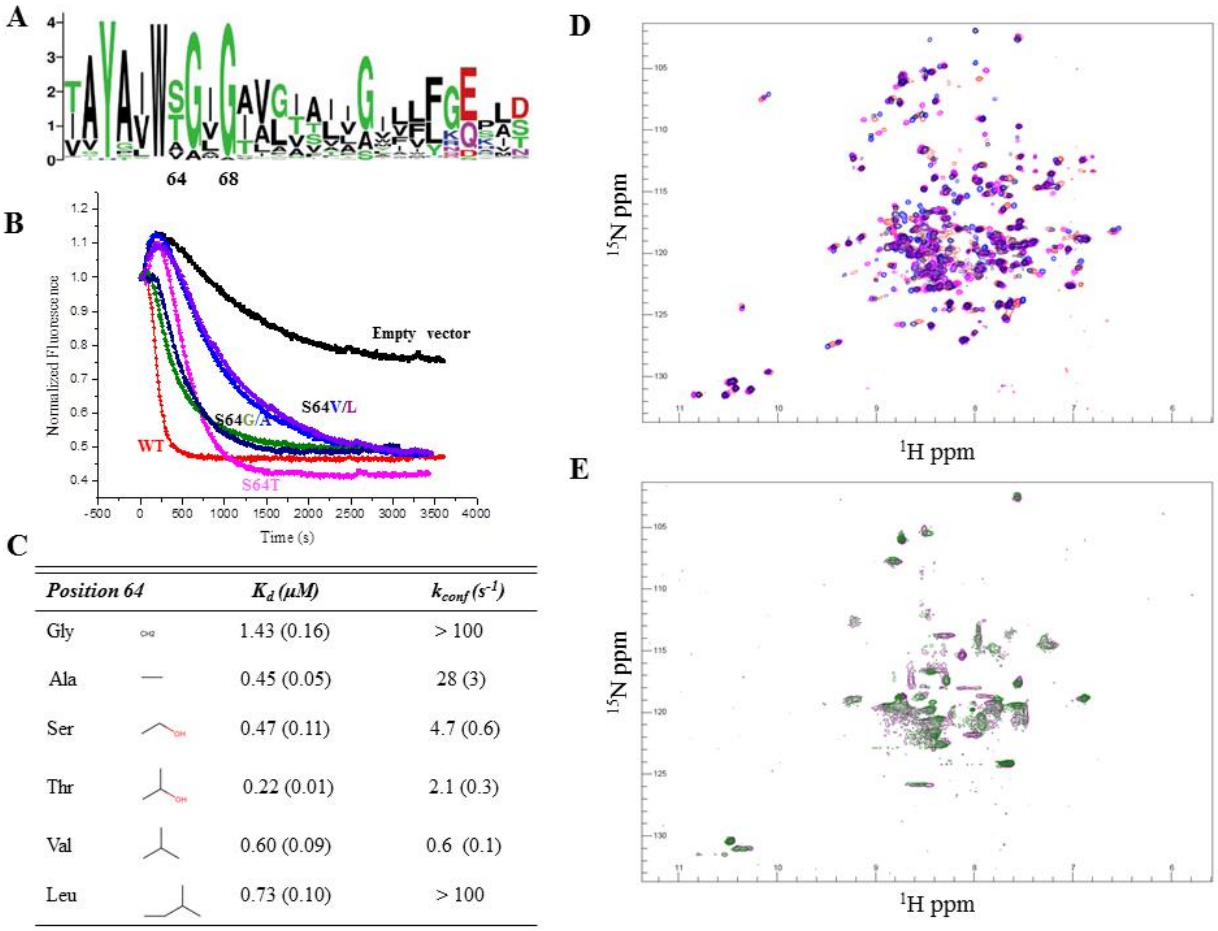


Fig. 4.3 TROSY-HSQC and ZZ-exchange spectra for EmrE in complex with TPP^+ at pH 7, 45 ˚C in bicelles (q = 0.33). A. TROSY-HSQC overlay for WT (black) and S64V (red) in DLPC/DHPC bicelles. B. TROSY-HSQC overlay for S64V in DLPC (red) and DMPC (green). C. Overlay of TROSY-HSQC (red) and a ZZ plane with 200 ms mixing time (blue) for S64V in DLPC/DHPC bicelles. D. Overlay of TROSY-HSQC (green) and a ZZ plane with 200 ms mixing time (blue) for S64V in DMPC/DHPC bicelles. Dotted boxes in C. and D. connect auto- and cross-peaks of pairs of amides that correspond to the same residue in each monomer and the exchange peaks due to the two monomers swapping conformations as EmrE interconverts between open-in and open-out. From the ZZ data, the conformational interconversion rates of EmrE-S64V in complex with TPP^+ are 0.6 ± 0.1 s^{-1} in DMPC/DHPC bicelles and 1.2 ± 0.2 s^{-1} in DLPC/DHPC bicelles. As a comparison, the rates for WT EmrE in complex with TPP^+ are 4.7 ± 0.6 s^{-1} in DMPC/DHPC bicelles and 4.4 ± 0.4 s^{-1} in DLPC/DHPC bicelles[11].

Characterization of S64 mutant series Our previous screening efforts for putative dynamics mutants (Chapter 2) revealed that most of the mutations that led to impaired transport activity, actually increased the conformational interconversion rate. S64V mutation was the only mutation consistent with our initial hypothesis. The position of Ser64 in TM3 is quite interesting due to its

location (Fig. 4.4 A). First, this residue is located right before the ^{65}GVG^{67} motif[21] which is known to introduce a kink in TM3 of EmrE. From the cryoEM structure, this kink is only found in monomer A[7] and is shown to contribute to the plasticity of the drug binding site of EmrE, which is important for multidrug recognition[8]. The difference in helix kinking between the two TM3 in the dimer was also observed using oriented solid state NMR[10]. Also chemical shift difference mapping from solution NMR data highlights the TM3 kink region as the most distinct in each monomer of the asymmetric dimer[9] and also one of the most sensitive regions to identity of the bound ligand[11]. The medium resolution X-ray and cryoEM structures[7,22] both suggest that the polar Ser64 residues from both monomers are facing inside and forming interhelix contacts, consistent with a functional role in conformational interconversion process. In addition, Ser64 is right after Trp63 which is one of the most conserved residues in its polyaromatic cation binding and critical for determining the substrate specificity of the transporter[23]. Thus, the residue at position 64 may also be critical for linking drug binding to transport. Sequence conservation data[21] suggests that there are three other frequently occurring residues at this position: Thr, Ala, and Val. I further characterized these EmrE S64 mutants to probe the conformational interconversion mechanism. I also included S64A and S64L mutations to get more comprehensive information about the role of TM3 kink in conformational interconversion.



Position 64		K_d *(μM)*	k_{conf} *(s^{-1})*
Gly		1.43 (0.16)	> 100
Ala		0.45 (0.05)	28 (3)
Ser		0.47 (0.11)	4.7 (0.6)
Thr		0.22 (0.01)	2.1 (0.3)
Val		0.60 (0.09)	0.6 (0.1)
Leu		0.73 (0.10)	> 100

Fig. 4.4 Characterization of EmrE S64 mutants. A. Sequence logo for TM3 of EmrE. Figure was modified from Elbaz et. al., 2008[21]. Colors: green, polar; blue, basic; red, acidic; black, hydrophobic. B. Normalized fluorescence as a function of time for EmrE WT and S64 mutants from in-cell assay data. Data from the black curve were collected for *E. coli* cells transformed with empty pET15b vectors whereas data from the others were from cells with EmrE overexpression. WT, red; S64G, olive; S64A, navy; S64T, magenta; S64V, blue; S64L, purple. C. Measured K_d for EmrE WT and S64 mutants with the ligand TPP^+ from ITC and conformational interconversion rates, k_{conf}, from ZZ-exchange experiments. Numbers in parentheses are standard deviations. D. TROSY-HSQC spectra for EmrE in complex with TPP^+, in DMPC/DHPC bicelles at pH 7, 45˚C, for WT (red), S64A (navy), S64T (magenta), and S64V (blue) show two sets of peaks. E. TROSY-HSQC spectra for EmrE in complex with TPP^+ for S64G (olive) and S64L (purple) show only one set of peaks suggesting significantly faster conformational interconversion.

The in-cell transport assay data provides a quick overview of the ethidium bromide export properties of cells expressing the different EmrE mutants (Fig. 4.4 B). Not surprisingly, cells expressing WT EmrE export ethidium bromide most efficiently followed by the S64G and S64A mutants. Cells expressing S64T-EmrE behave differently and are more transport-impaired in the

initial stage. Cells with the two largest and most hydrophobic mutations at this site, S64V-EmrE and S64L-EmrE, behave similarly and are the most transport-impaired. These *in vivo* drug export data are interesting but only provide a rather qualitative picture about the behavior of cells expressing EmrE. Factors like different expression, folding, and insertion levels of the various EmrE mutants would require many additional controls to seriously compare the drug export properties of these EmrE mutants.

I went on to characterize TPP^+ binding affinities of this series of EmrE mutants using ITC to more quantitatively evaluate differences in ligand binding properties (Fig. 4.4 C). Interestingly, the measured K_d values for these EmrE mutants binding to TPP^+ vary only slightly; all the K_d values fall between 200-1450 nM under identical conditions. These similar drug affinities among the various S64 mutants, along with the similarity between the binding affinities of S64V- and WT-EmrE for different TPP^+ derivatives, collectively suggest that the binding site of EmrE is fairly plastic and perturbations to either ligands or proteins can be accommodated without loss of affinity. Thus defects in ligand binding unlikely contribute to the observed different phenotypes.

Intrigued by the distinct phenotypes of different S64 mutations on the drug export function of EmrE, we characterized these TM3 kink mutants using solution NMR to identify possible dynamics-function correlations. As shown in the TROSY-HSQC spectra (Fig. 4.4 D,E) of EmrE in complex with TPP^+, WT EmrE and S64A-, S64T-, and S64V-EmrE all show two sets of peaks corresponding to the asymmetric dimer. These mutations do not change the structure dramatically. In the case of EmrE S64A, cross-peaks due to conformational interconversion appear in the TROSY-HSQC spectrum indicating faster interconversion between open-in and open-out for this mutant, on the order of 30-100 s^{-1}. In the case of S64G-EmrE bound to TPP^+, the conformational interconversion is so fast that only one set of peaks appear in the spectrum.

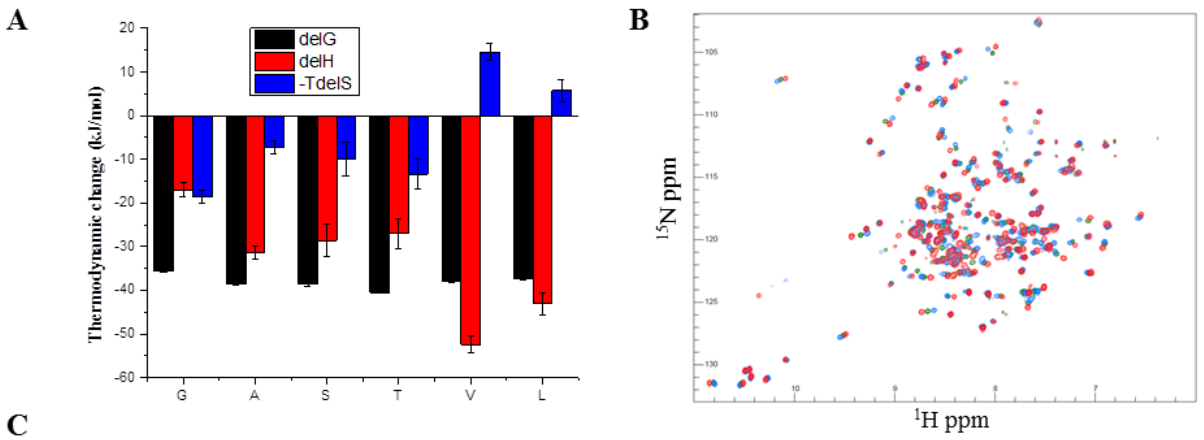
This is similar to WT EmrE bound to $MeTPP^+$, and a rough estimate of the rate suggests that the conformational interconversion should be faster than 100 s^{-1} for S64G-EmrE in complex with TPP^+. Glycine residues in transmembrane helices are known to render the helices more flexible as a result of the larger conformational space that can be sampled by glycine backbone dihedral angles. Therefore glycine is one of the two residues that are known to cause kinks in helices. In light of this, the increased dynamics by the S64G mutation of EmrE (which results in a GGVG sequence in the TM3 kink region) makes sense. Using ZZ-exchange experiments, we calculated that the rates for mutants that still had two sets of peaks in the spectrum. The interconversion rates for S64A and S64T are 28 ± 3 and 2.1 ± 0.3 s^{-1}. Thus, within the mutation series, as the size of the residue 64 in EmrE gets larger, the interconversion rates for TPP^+-bound state decrease. However, this trend does not hold for the S64L mutation. Like S64G-EmrE, the spectrum for S64L-EmrE in complex with TPP^+ has only one set of peaks and significant line broadening suggesting significantly faster dynamics or loss of structural asymmetry or stability. Thus, although S64L behaves similarly to S64V in the in-cell ethidium efflux assay, its behavior is quite distinct based on the NMR dynamics data. One possible explanation is that the bulky side chain of S64L (Leu is rare at this position in the SMR family) might not fit within the tightly packed helix-helix interface where S64 is located and may disrupt the structure of this critical region more generally. As pointed out previously, the conformational interconversion step is only one of the potential rate-limiting steps in the transport cycle[11] and S64L may disrupt other aspects of the transport cycle, such as proton coupling.

Hydration around the kink region in conformational interconversion A closer look into the thermodynamic components of different EmrE mutants binding to TPP^+ reveals the importance of hydration within the transport pore. As shown above, the free energies of binding to TPP^+ do

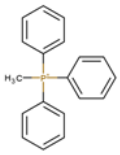
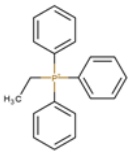
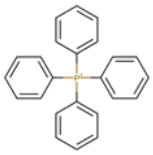
not change much for different S64 mutants of EmrE. However, the enthalpic and entropic components differ significantly (Fig. 4.5 A). For S64G/A/T substitutions, both ΔH and $-T\Delta S$ of TPP^+ binding vary but remain negative as for WT EmrE. However, for S64V/L substitutions, ΔS is positive. This observed enthalpy-entropy compensation[24-31], is almost universally observed in weak interactions in water. This suggests that the different S64 mutants may result in distinct hydration patterns around the TM3 kink region. Altered hydration in this region would affect the conformational interconversion rate because TM3 is only bent in one monomer of the asymmetric dimer. As EmrE interconverts between open-in and open-out states, kinking and unkinking must occur as the two monomers swap conformations within the asymmetric dimer. This requires breaking and rebuilding of backbone hydrogen bonds near the middle of TM3, which is likely a relatively hydrophobic environment. Altered hydration patterns will change the local effective dielectric constant and perturb both the thermodynamics and kinetics of hydrogen bonding[32,33].

The idea that hydration around TM3 kink region may be important for conformational interconversion is supported when comparing interconversion rates of the closely related TPP^+ derivatives (Fig. 4.5 B,C). Previously, dynamics studies[11] using this series of TPP^+ derivatives revealed that the interconversion rate decreases as the ligand becomes larger and more hydrophobic. This is true all the way from the smallest ligand, $MeTPP^+$ ($130 \pm 20\ s^{-1}$, in DLPC/DHPC bicelles), to $EtTPP^+$ ($21.8 \pm 0.7\ s^{-1}$, in DLPC/DHPC bicelles), to TPP^+ ($4.4 \pm 0.4\ s^{-1}$, in DLPC/DHPC bicelles; $4.7 \pm 0.6\ s^{-1}$, in DMPC/DHPC bicelles), and to the largest, $DPhTPP^+$ ($0.4 \pm 0.1\ s^{-1}$, in DLPC/DHPC bicelles). These could be understood using the hydration concept: as the ligand gets bigger and more hydrophobic, it blocks water access into the transport pore and the TM3 kink region becomes dehydrated increasing the energetic barrier of helix kinking

and leading to slower interconversion. This trend is also observed in S64V-EmrE with the conformational interconversion rate decreasing from 7.9 ± 2.3 s^{-1} for $MeTPP^+$, to 1.1 ± 0.1 s^{-1} for $EtTPP^+$, to 0.6 ± 2.3 s^{-1} for TPP^+ (I did not measure the interconversion rate for EmrE-S64V in complex with $DPhTPP^+$ as that would require prohibitively long mixing times for ZZ-exchange experiments). Interestingly, the calculated ratios of interconversion rates from ligand and protein perturbations (Fig. 4.5 D) further suggest there is a shared dehydration process. For example, in the ligand perturbation for WT EmrE, the ratio is ~6 from $MeTPP^+$ to $EtTPP^+$ and ~5 from $EtTPP^+$ to TPP^+. For EmrE-S64V, the ratio is ~7 from $MeTPP^+$ to $EtTPP^+$. However, the ratio is only ~2 from $EtTPP^+$ to TPP^+. This drop in ratio is also true as seen from the protein perturbation side. For $MeTPP^+$ and $EtTPP^+$, the ratios are ~16 and ~20 when the protein is perturbed from Ser to Val. The ratio drops significantly for TPP^+ and is now ~8. These results lead to a shared dehydration hydration hypothesis: increasing either ligand or protein hydrophobicity in the transport pore near the TM3 kink will dehydrate the binding site, reduce the rate of conformational interconversion, and reduce the rate of drug efflux. However, a limit is reached where no more water molecules can be displaced by further increase of either ligand or protein hydrophobicity.



C

	k_{conf} *in drug-bound state* (s^{-1})[a]			k_{conf} *in drug-free state* (s^{-1})		
	 *MeTPP$^+$*	 *EtTPP$^+$*	 *TPP$^+$*	*pH 5.2*	*pH 6.0*	*pH 7.4*
WT	130 (20)[b]	21.8 (0.7)[b]	4.7 (0.6)	> 200[c]	> 150[c]	> 100[c]
S64V	7.9 (2.3)	1.1 (0.1)	0.6 (0.1)	~ 0.5	~ 0.5	~ 0.5

[a] At pH 7, 45°C
[b] Rates in DLPC/DHPC bicelles from Morrison, *et al.*, 2014.
[c] Estimated from Gayen, *et al.*, 2016. EmrE is in aligned large bicelles for solid state NMR at 37°C.

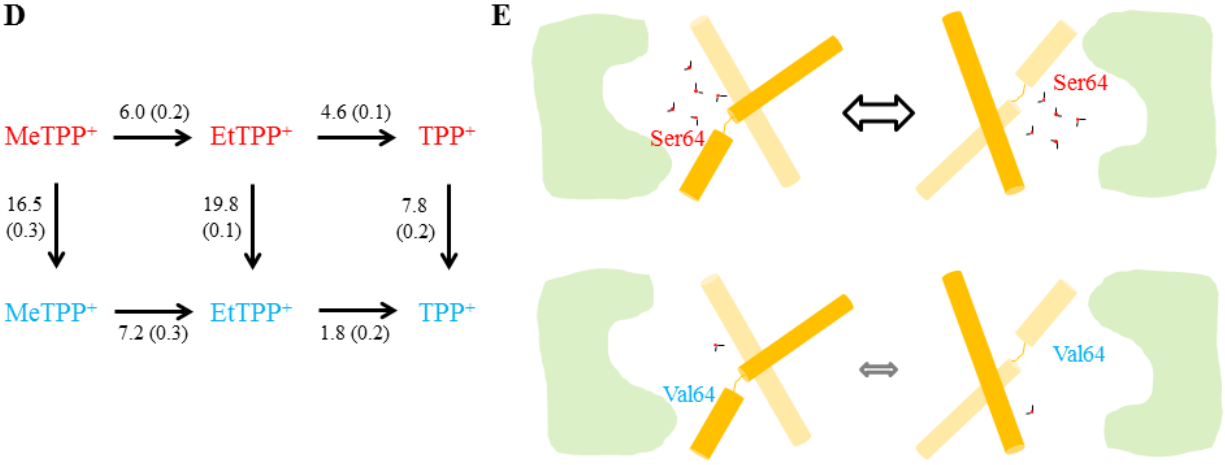


Fig. 4.5 The role of hydration in proton-driven drug efflux by EmrE. A. Enthalpy-entropy compensation in TPP^+ binding to EmrE S64 point mutants (ΔG, black; ΔH, red; -TΔS, blue). B. TROSY-HSQC spectra of EmrE-S64V in complex with TPP^+ (red), $MeTPP^+$ (blue), and $EtTPP^+$ (green) show no major structural changes. C. Conformational interconversion rates, k_{conf}, for WT- and S64V-EmrE in the presence and absence of drug. The rates for WT EmrE in drug-free state were estimated from solid-state NMR PUREX experiments[34] and our own ZZ-exchange measurements of drug-free EmrE[20]. The rates for EmrE S64V in drug-free state were estimated from ZZ-exchange data although the slow rate limits the number of cross peaks that can be quantified. D. Ratios of conformational

interconversion rates as a result of perturbation to the protein and the bound ligand. WT EmrE systems are colored in red while EmrE-S64V systems are colored in blue. Numbers in parentheses are propagated errors in ratios. E. A cartoon model to explain the role of hydration around kink region affects helix kinking and unkinking in conformational interconversion. The two TM3 helices in the antiparallel asymmetric dimer are drawn in yellow while the rest of EmrE is shown as green blob. During conformational interconversion, one TM3 kinks and the other straightens. When the position 64 is a serine residue, there are more water molecules around the kink effectively lowering the energetic barrier for helix kinking and unkinking.

As the saturation trend in our ligand and protein perturbation series suggests a shared dehydration process, we predicted that the differences in interconversion rates between WT and S64V EmrE will be even larger in the drug-free state. This is because the substitution of the serine with valine should displace more waters than that when the binding site is already occupied by a large hydrophobic ligand. This prediction is borne out by comparison of the interconversion rates between drug-free WT- and S64V-EmrE (Fig. 4.5 C). The interconversion rate determined from solid-state NMR[34] for WT EmrE in aligned DMPC/DHPC bicelles is over 100 s^{-1} below neutral pH at 37˚C. This is consistent with rates determined from solution NMR in bicelles at 45 ˚C of ~50 s^{-1} at pH 8.8, ~75 s^{-1} at pH 8.4, and much faster (>100 s^{-1}) at lower pH values[20]. In contrast, the interconversion rates for EmrE-S64V in the drug-free state are much slower. The rate is estimated to be ~0.5 s^{-1} across the pH range from 5.2 to 7.4.

Collectively, these interconversion rates comparisons support the hypothesis that hydration around the TM3 kink regions may affect the rate of conformational interconversion. This is illustrated in the cartoon model in Fig 4.5 E. During the conformational interconversion, the two monomers in the asymmetric dimer of EmrE swap conformations. In particular, the TM3 from monomer A is straightened from the kinked state while the other TM3 from monomer B becomes kinked. When the 64 position of TM3 is mutated from the polar serine residue to nonpolar valine, the region around TM3 kink region is more dehydrated which effectively

increases the energetic barrier of helix kinking and unkinking by changing the local dielectric constant. This could be reasoned from both a thermodynamic and kinetic point of view. A nonpolar mutation would increase the strength of hydrogen bonding[33], therefore lowering the free energy of the ground state. Meanwhile, this nonpolar mutation would increase the effective dielectric constant making exposing of polar groups during kinking and unkinking of helices kinetically more difficult.

E14 pK_a fine-tuning by hydration The pH titration of the drug-free state of S64V-EmrE reveal another interesting insight into the transport process. As previously discussed, WT EmrE is very sensitive to pH in the absence of drug. The two active site E14 residues in the antiparallel asymmetric homodimer have pK_a values of 7.0 ± 0.1 and 8.2 ± 0.3[20], and the rate of open-in to open-out exchange varies from ~220 s^{-1} to ~40 s^{-1} between pH 5 and 9.5[34]. Unlike WT EmrE, the interconversion for drug-free S64V-EmrE is very slow (~0.5 s^{-1}), with two sets of peaks visible in TROSY-HSQC spectra over the entire pH range studied (4.8 - 10.5), and is independent of pH from 5.2 to 7.4. Solvent exchange prevents the acquisition of meaningful ZZ-exchange data at higher pH for S64V due to the long mixing times required to monitor the slow exchange process (the faster exchange rate of WT EmrE allows shorter mixing times to be used and ZZ-exchange data can be acquired up to pH 8.5). This dramatically different behavior was intriguing, particularly in light of the questions regarding how EmrE couples proton import to drug efflux given our recent findings that contradict the long-standing single-site alternating access for coupled antiport. Our recent discoveries of asymmetric protonation of EmrE[20], its ability to interconvert between open-in and open-out in multiple different protonation states[20] and the ability to simultaneously bind drug and proton (unpublished) are all inconsistent with traditional single-site alternating access antiport models. We have developed a kinetic model for proton-

coupled TPP^+ transport using rates measured for microscopic steps within the transport cycle, but this naturally leads to another question: if proton coupling is achieved by careful tuning of the proton and drug transport halves of the transport cycle, how does EmrE harness the proton-motive-force to actively pump such diverse substrates, particularly since we know that the kinetics of the open-in to open-out exchange process varies over several orders of magnitude for the substrate transport side of the cycle?

We hypothesize that hydration provides the key, linking both substrate and proton transport by tuning both the conformational interconversion rate and the pK_a of the critical E14 residue. Changing hydration patterns can affect pK_a values and protein functions as demonstrated in several other cases[35-43]. We therefore predicted that reduced hydration of the transport pore in S64V-EmrE should further upshift the pK_a values of the two Glu14 residues.

This is indeed the case and is immediately obvious from the TROSY-HSQC spectra at several pH values (Fig. 4.6). The peaks of drug-free S64V-EmrE shift only slightly below pH 8, in stark contrast to WT EmrE[20]. With the use of ether-linked DHPC, we were able to collect spectra up to pH 10.8. At pH values higher than 8.6, significant number of peaks, especially those that are exposed to the solvent, disappear as a result of increased amide exchange. But it is very obvious from the residues that remain visible that two sets of peaks remain. From the curved pattern as the peaks shift across the spectrum with pH, there are at least two protonation events. I fit the chemical shifts as a function of pH using both a single pK_a and double pK_a model[20]. The double pK_a model clearly gives better fits to the data, as was also observed in WT[20]. The fitted two pK_a values, 8.6 ± 0.1 and 9.3 ± 0.2, are both at least one pH unit higher compared to WT EmrE. This not only supports the idea that hydration in the binding site is reduced as a result of the S64 mutation, but also links the proton coupling E14 region to the TM3 kink region.

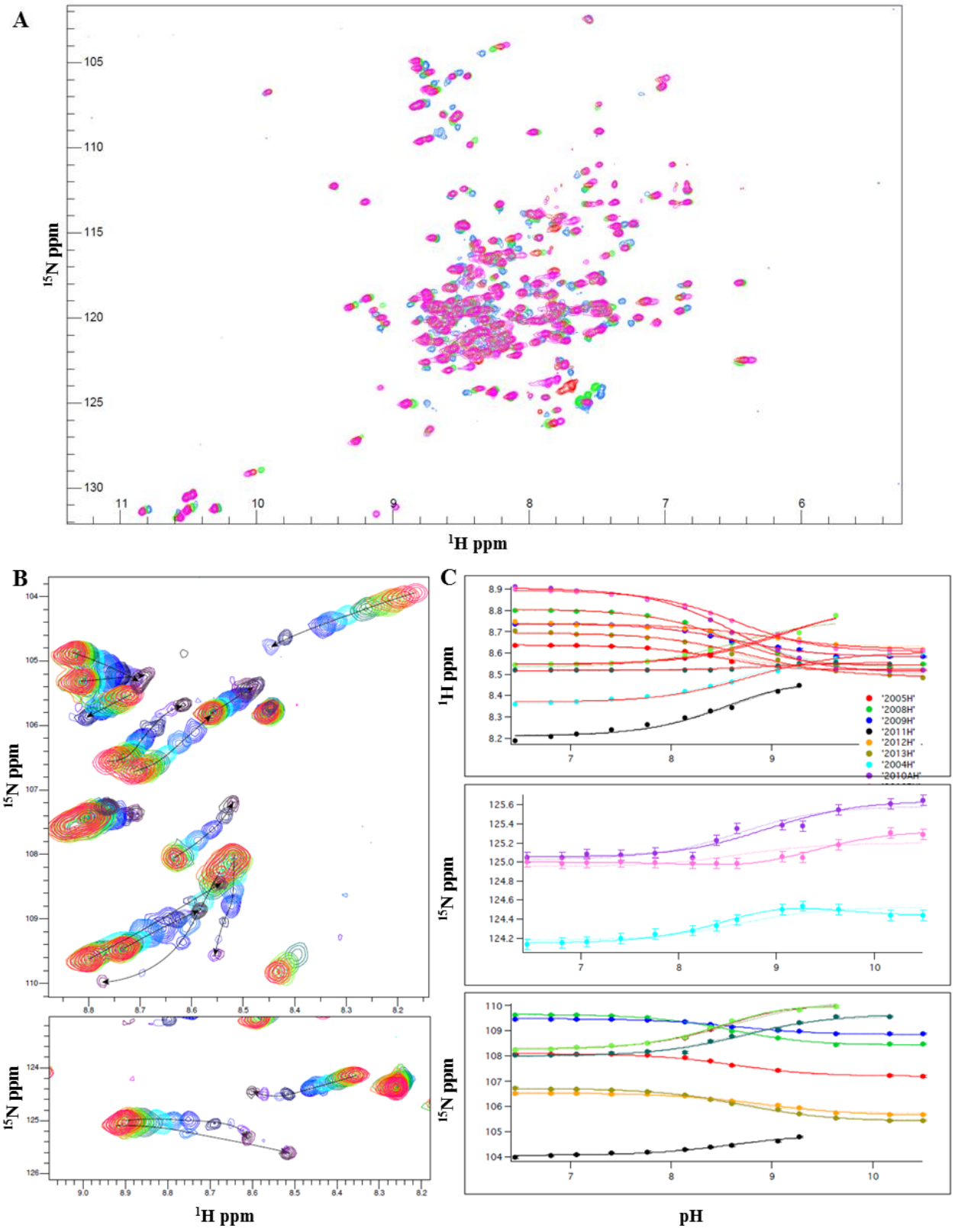

A
^{15}N ppm
^{1}H ppm
B
^{15}N ppm
^{1}H ppm
C
^{1}H ppm
'2005H'
'2008H'
'2009H'
'2011H'
'2012H'
'2013H'
'2004H'
'2010AH'
^{15}N ppm
^{15}N ppm
pH

Fig. 4.6 pH titration of EmrE-S64V in drug-free state reveals a link between hydration and fine-tuning of E14 pK_a value which is critical for energetic coupling in the transport cycle. A. TROSY-HSQC spectra at pH 5.9, 6.8, and 7.8. Colors: pH 5.9, magenta; pH 6.8, red; pH 7.8, lime; pH 8.6, bright blue. B. Regions of spectra from pH titration show movements of peaks as the pH is increased. Upper figure, an enlarged glycine region; lower figure, a region including Ala10. Colors: pH 5.9, magenta; pH 6.4, pink; pH 6.8, red; pH 7.1, orange; pH 7.4, yellow; pH 7.8, lime; pH 8.1, green; pH 8.4, cyan; pH 8.6, bright blue; pH 9.0, navy; pH 9.3, lavender; pH 9.6, purple; pH 10.2, brown; pH 10.5, grey. C. Fitting of chemical shifts as a function of pH to a single pK_a (dotted lines) and two pK_a values (solid lines). Peaks are not assigned. The two pK_a model fits better to the experimental data and gives 8.6 ± 0.1 and 9.3± 0.2 while the single pK_a model gives 8.6 ± 0.1. As a comparison, the two pK_a model fitting for WT^{20} give 7.0 ± 0.1 and 8.2 ± 0.3 while the single pK_a model gives 7.1 ± 0.1.

Discussions

TM kinking central to interconversion As the number of membrane protein structures grows, it is clear that kinked helices are more common than initially expected, and TM helix kinks are usually found at sites of functional importance. Kinks commonly occur at Gly or Pro residues[2,3,5,6] and may provide the correct functional groups and geometry for ligand coordination, or create hinge points for functionally important conformational changes required by many membrane protein functions. Such functions include moving molecules or transducing signals across the hydrophobic membrane barrier: transporters must alternately open to either side of the membrane, GPCRs must transfer information from one side of membrane to the other[44], and ion channels must open and close in response to signals[45].

In the case of EmrE, substantive evidence points to the importance of the TM3 kink in the transport process[7,9-11,46]. Here, I show directly the importance of the TM3 kink for conformational exchange between open-in and open-out by demonstrating that mutation of S64 leads to almost three orders of magnitude differences in the interconversion rate. For example, the glycine at position 64 increases the interconversion rate to > 100 s^{-1}, consistent with

flexibility allowed by the small side chain and consequent propensity of glycine to facilitate kink formation in helices[47].

The role of hydration in helix kinking thus interconversion By comparing interconversion rates of different EmrE mutants and WT EmrE bound to different drugs, we identified a common theme that size and hydrophobicity of the amino acid side chain or substrate correlated with the conformational exchange rate. Thermodynamic characterization of substrate binding revealed enthalpy-entropy compensation, leading to the hypothesis that hydration around the TM3 kink may be important for substrate interaction, and perhaps also helix kinking and conformational interconversion of EmrE. As the residue at position 64 (or the substrate) gets bigger and more hydrophobic, the interconversion between open-in and open-out slows down, although this effect eventually plateaus. This suggests that there may be a shared dehydration process around the TM3 kink resulting from a combination of protein and ligand factors.

Helix kinking within the hydrophobic environment of the membrane is generally unfavorable unless water or other proteins interactions can stabilize the broken backbone hydrogen bonding partners and lower the energetic barrier of exposing helix dipoles within bilayer. Residue 64 in WT EmrE is serine. Serine residues in helices can stabilize a TM helix kink since the side-chain hydroxyl group can hydrogen bond with the carbonyl group of the i - 4 residue[48-50]. However, mutation of S64 to alanine increases the interconversion rate, suggesting that hydrogen bonding potential is not a critical factor. Based on the observed pattern of effects across both mutant and ligand, neither hydrogen bonding potential nor the presence of polar groups are critical for rapid conformational exchange, supporting the more general hydration model.

Furthermore, the effect of S64 mutation on conformational exchange rate is even more dramatic in the drug-free state where the mutation is the sole perturbation, as expected from the proposed dehydration model. While WT EmrE in the absence of drug interconverts between open-in and open-out at rates between ~30 s^{-1} to > 200 s^{-1} in a pH-dependent manner. EmrE-S64V has a conformational exchange rate of ~0.5 s^{-1} near neutral pH in the absence of drug substrates.

The hydration concept explains most of the experimental data, but it is clearly not the only factor affecting interconversion. S64L does not fit the trend of decreased conformational exchange with increasing side-chain size and hydrophobicity. Instead, the interconversion rate of TPP^+-bound EmrE-S64L is significantly faster even than TPP^+-bound EmrE-S64A. We hypothesize that leucine is large enough that it does not fit well in the tightly packed region of the TM3 hinge where it interfaces with the substrate binding pocket. The interconversion rate of drug-free EmrE-S64L is 1.2 ± 0.2 s^{-1}, similar to drug-free EmrE-S64V and much slower than TPP^+-bound EmrE-S64L. Thus S64L is more consistent with the overall pattern in the absence of substrate, suggesting that the binding site of EmrE, although plastic, still has limited space. Perhaps similar factors explain the behavior of EmrE-S64V in complex with $MeTPP^+$/$EtTPP^+$. This is also suggested by the global peak shifts observed between the drug-free and substrate-bound states of EmrE-S64V.

Hydration as a unifying factor in transport In addition to the importance for TM3 kinking and open-in to open-out conformational exchange of EmrE, hydration could also influence the energetics of first solvent shell within the transport pore, affecting the pK_a of titratable residues important for proton-coupled transport. The various kinetic steps, e.g. drug-off rate and interconversion rate, in multidrug transport by EmrE vary several orders of magnitude[11] and pK_a shifts upon drug binding will affect the efficiency of proton-coupled transport within our new

kinetic model for EmrE transport. The simultaneous effect of altered hydration within the transport pore on the rate of TM3 kinking and the pK_a of E14 suggests a new model for how EmrE achieves proton-coupled transport of various drugs with different sizes, shapes, and charges.

To date, studies of EmrE structure and dynamics have all been performed in symmetric environments. Yet, *in vivo*, the membrane asymmetry brought by the negative-inside membrane potential could be a defining feature affecting the kinetics and thermodynamics of multiple steps in the transport cycle. Hydration within the transport pore will also influence movement of charged residues within the pore across the membrane potential as the barrier switches from one side of the transport pore to the other. As a result, hydration may be an important unifying factor to simultaneously fine-tune multiple steps in the transport cycle and achieve efficient proton-coupled transport.

Conclusion

In this chapter, I presented experimental data supporting the hypothesis that the TM3 kink in EmrE is crucial for conformation interconversion and transport activity. The perturbations of position 64 in TM3 dramatically changed the interconversion rate by over three orders of magnitude. Combined with the ligand perturbation series, this suggest that hydration within the transport pore may fine-tune the interconversion rate. In addition, these changes are also linked to more than 1 unit change of pK_a values of E14, suggesting hydration may provide a key link between substrate binding site and the TM3 kink important for open-in to open-out exchange, as required for coupled antiport.

Author contributions

Katherine A. Henzler-Wildman and Chao Wu designed the experiments. Chao Wu collected and analyzed most of the data. Gregory T. DeKoster collected and analyzed NMR data for EmrE-S64V in complex with $MeTPP^+$/$EtTPP^+$. Eva-Maria Uhlemann replicated ITC data for EmrE S64 mutants with TPP^+ and collected ITC data for EmrE S64V with $MeTPP^+$ and $EtTPP^+$. Katherine A. Henzler-Wildman did the mutagenesis and analyzed the pH titration data for drug-free EmrE-S64V. Chao Wu did the writing and Katherine A. Henzler-Wildman revised it.

References

1 Bowie, J. U. Solving the membrane protein folding problem. *Nature* **438**, 581-589, doi:10.1038/nature04395 (2005).

2 Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P. & Bowie, J. U. The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 959-963, doi:10.1073/pnas.0306077101 (2004).

3 Hall, S. E., Roberts, K. & Vaidehi, N. Position of helical kinks in membrane protein crystal structures and the accuracy of computational prediction. *Journal of molecular graphics & modelling* **27**, 944-950, doi:10.1016/j.jmgm.2009.02.004 (2009).

4 Langelaan, D. N., Wieczorek, M., Blouin, C. & Rainey, J. K. Improved helix and kink characterization in membrane proteins allows evaluation of kink sequence predictors.

Journal of chemical information and modeling **50**, 2213-2220, doi:10.1021/ci100324n (2010).

5 Meruelo, A. D., Samish, I. & Bowie, J. U. TMKink: a method to predict transmembrane helix kinks. *Protein science : a publication of the Protein Society* **20**, 1256-1264, doi:10.1002/pro.653 (2011).

6 Wilman, H. R., Shi, J. & Deane, C. M. Helix kinks are equally prevalent in soluble and membrane proteins. *Proteins* **82**, 1960-1970, doi:10.1002/prot.24550 (2014).

7 Fleishman, S. *et al.* Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *Journal of molecular biology* **364**, 54-67, doi:10.1016/j.jmb.2006.08.072 (2006).

8 Korkhov, V. M. & Tate, C. G. Electron crystallography reveals plasticity within the drug binding site of the small multidrug transporter EmrE. *Journal of molecular biology* **377**, 1094-1103, doi:10.1016/j.jmb.2008.01.056 (2008).

9 Morrison, E. *et al.* Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature* **481**, 45-50, doi:10.1038/nature10703 (2012).

10 Gayen, A., Banigan, J. R. & Traaseth, N. J. Ligand-induced conformational changes of the multidrug resistance transporter EmrE probed by oriented solid-state NMR spectroscopy. *Angewandte Chemie (International ed. in English)* **52**, 10321-10324, doi:10.1002/anie.201303091 (2013).

11 Morrison, E. A. & Henzler-Wildman, K. A. Transported substrate determines exchange rate in the multidrug resistance transporter EmrE. *The Journal of biological chemistry* **289**, 6825-6836, doi:10.1074/jbc.M113.535328 (2014).

12 Morrison, E. & Henzler-Wildman, K. Reconstitution of integral membrane proteins into isotropic bicelles with improved sample stability and expanded lipid composition profile. *Biochimica et biophysica acta* **1818**, 814-820, doi:10.1016/j.bbamem.2011.12.020 (2012).

13 Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277-293, doi:10.1007/bf00197809 (1995).

14 Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696, doi:10.1002/prot.20449 (2005).

15 Li, Y. & Palmer, A. G. TROSY-selected ZZ-exchange experiment for characterizing slow chemical exchange in large proteins. *Journal of biomolecular NMR* **45**, 357-360, doi:10.1007/s10858-009-9385-0 (2009).

16 Miloushev, V. Z. *et al.* Dynamic properties of a type II cadherin adhesive domain: implications for the mechanism of strand-swapping of classical cadherins. *Structure (London, England : 1993)* **16**, 1195-1205, doi:10.1016/j.str.2008.05.009 (2008).

17 Wu, H., Su, K., Guan, X., Sublette, M. E. & Stark, R. E. Assessing the size, stability, and utility of isotropically tumbling bicelle systems for structural biology. *Biochimica et biophysica acta* **1798**, 482-488, doi:10.1016/j.bbamem.2009.11.004 (2010).

18 Tomlinson, J. H., Green, V. L., Baker, P. J. & Williamson, M. P. Structural origins of pH-dependent chemical shifts in the B1 domain of protein G. *Proteins* **78**, 3000-3016, doi:10.1002/prot.22825 (2010).

19 McIntosh, L. P. *et al.* Dissecting electrostatic interactions in Bacillus circulans xylanase through NMR-monitored pH titrations. *Journal of biomolecular NMR* **51**, 5-19, doi:10.1007/s10858-011-9537-x (2011).

20 Morrison, E. A., Robinson, A. E., Liu, Y. & Henzler-Wildman, K. A. Asymmetric protonation of EmrE. *The Journal of general physiology* **146**, 445-461, doi:10.1085/jgp.201511404 (2015).

21 Elbaz, Y., Salomon, T. & Schuldiner, S. Identification of a glycine motif required for packing in EmrE, a multidrug transporter from Escherichia coli. *The Journal of biological chemistry* **283**, 12276-12283, doi:10.1074/jbc.M710338200 (2008).

22 Chen, Y.-J. *et al.* X-ray structure of EmrE supports dual topology model. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18999-19004, doi:10.1073/pnas.0709387104 (2007).

23 Brill, S., Falk, O. S. & Schuldiner, S. Transforming a drug/H+ antiporter into a polyamine importer by a single mutation. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 16894-16899, doi:10.1073/pnas.1211831109 (2012).

24 Gilli, P., Ferretti, V., Gilli, G. & Borea, P. A. Enthalpy-entropy compensation in drug-receptor binding. *The Journal of Physical Chemistry* **98**, 1515-1518, doi:10.1021/j100056a024 (1994).

25 Movileanu, L. & Schiff, E. A. Entropy-enthalpy Compensation of Biomolecular Systems in Aqueous Phase: a Dry Perspective. *Monatshefte fur chemie* **144**, 59-65, doi:10.1007/s00706-012-0839-9 (2013).

26 Olsson, T. S., Ladbury, J. E., Pitt, W. R. & Williams, M. A. Extent of enthalpy-entropy compensation in protein-ligand interactions. *Protein science : a publication of the Protein Society* **20**, 1607-1618, doi:10.1002/pro.692 (2011).

27 Makhatadze, G. I., Clore, G. M. & Gronenborn, A. M. Solvent isotope effect and protein stability. *Nature structural biology* **2**, 852-855 (1995).

28 Chodera, J. D. & Mobley, D. L. Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annual review of biophysics* **42**, 121-142, doi:10.1146/annurev-biophys-083012-130318 (2013).

29 Fenley, A. T. *et al.* Entropy–enthalpy transduction caused by conformational shifts can obscure the forces driving protein–ligand binding. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1213180109 (2012).

30 Chervenak, M. C. & Toone, E. J. A Direct Measure of the Contribution of Solvent Reorganization to the Enthalpy of Binding. *Journal of the American Chemical Society*, doi:10.1021/ja00102a021 (1994).

31 Duff, M. R. & Howell, E. E. Thermodynamics and solvent linkage of macromolecule–ligand interactions. *Methods* **76**, 51-60, doi:10.1016/j.ymeth.2014.11.009 (2015).

32 Joh, N. *et al.* Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* **453**, 1266-1270, doi:10.1038/nature06977 (2008).

33 Gao, J., Bosco, D. A., Powers, E. T. & Kelly, J. W. Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. *Nature structural & molecular biology* **16**, 684-690, doi:10.1038/nsmb.1610 (2009).

34 Gayen, A., Leninger, M. & Traaseth, N. J. Protonation of a glutamate residue modulates the dynamics of the drug transporter EmrE. *Nature chemical biology*, doi:10.1038/nchembio.1999 (2016).

35 Blasic, J. R., Worcester, D. L., Gawrisch, K., Gurnev, P. & Mihailescu, M. Pore Hydration States of KcsA Potassium Channels in Membranes. *The Journal of biological chemistry*, doi:10.1074/jbc.M115.661819 (2015).

36 Feng, J. & Mertz, B. Proteorhodopsin Activation Is Modulated by Dynamic Changes in Internal Hydration. *Biochemistry* **54**, 7132-7141, doi:10.1021/acs.biochem.5b00932 (2015).

37 Krepkiy, D. *et al.* Structure and hydration of membranes embedded with voltage-sensing domains. *Nature* **462**, 473-479, doi:10.1038/nature08542 (2009).

38 Kaila, V. R., Wikström, M. & Hummer, G. Electrostatics, hydration, and proton transfer dynamics in the membrane domain of respiratory complex I. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 6988-6993, doi:10.1073/pnas.1319156111 (2014).

39 Han, W., Cheng, R. C., Maduke, M. C. & Tajkhorshid, E. Water access points and hydration pathways in CLC H+/Cl- transporters. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1819-1824, doi:10.1073/pnas.1317890111 (2014).

40 Baday, S., Wang, S., Lamoureux, G. & Bernèche, S. Different hydration patterns in the pores of AmtB and RhCG could determine their transport mechanisms. *Biochemistry* **52**, 7091-7098, doi:10.1021/bi400015f (2013).

41 Goyal, P., Lu, J., Yang, S., Gunner, M. R. & Cui, Q. Changing hydration level in an internal cavity modulates the proton affinity of a key glutamate in cytochrome c oxidase. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 18886-18891, doi:10.1073/pnas.1313908110 (2013).

42 Zhao, C. & Noskov, S. The Role of Local Hydration and Hydrogen-Bonding Dynamics in Ion and Solute Release from Ion-Coupled Secondary Transporters. *Biochemistry* **50**, 1848-1856, doi:10.1021/bi101454f (2011).

43 Mahmmoud, Y. A., Kopec, W. & Khandelia, H. K+ congeners that do not compromise Na+ activation of the Na+,K+-ATPase: hydration of the ion binding cavity likely controls ion selectivity. *The Journal of biological chemistry* **290**, 3720-3731, doi:10.1074/jbc.M114.577486 (2015).

44 van der Kant, R. & Vriend, G. Alpha-bulges in G protein-coupled receptors. *International journal of molecular sciences* **15**, 7841-7864, doi:10.3390/ijms15057841 (2014).

45 Akitake, B., Anishkin, A., Liu, N. & Sukharev, S. Straightening and sequential buckling of the pore-lining helices define the gating cycle of MscS. *Nature structural & molecular biology* **14**, 1141-1149, doi:10.1038/nsmb1341 (2007).

46 Tate, C. G., Ubarretxena-Belandia, I. & Baldwin, J. M. Conformational changes in the multidrug transporter EmrE associated with substrate binding. *Journal of molecular biology* **332**, 229-242, doi:10.1016/S0022-2836(03)00895-7 (2003).

47 Dong, H., Sharma, M., Zhou, H.-X. X. & Cross, T. A. Glycines: role in α-helical membrane protein structures and a potential indicator of native conformation. *Biochemistry* **51**, 4779-4789, doi:10.1021/bi300090x (2012).

48 Gray, T. M. & Matthews, B. W. Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. *Journal of molecular biology* **175**, 75-81 (1984).

49 Ballesteros, J. A., Deupi, X., Olivella, M., Haaksma, E. & Pardo, L. Serine and Threonine Residues Bend α-Helices in the χ1=g− Conformation. *Biophysical Journal* **79**, 2754-2760, doi:10.1016/S0006-3495(00)76514-3 (2000).

50 Del Val, C., White, S. H. & Bondar, A.-N. N. Ser/Thr motifs in transmembrane proteins: conservation patterns and effects on local protein structure and dynamics. *The Journal of membrane biology* **245**, 717-730, doi:10.1007/s00232-012-9452-4 (2012).

Chapter 5: Asymmetric C-termini of EmrE

Abstract

EmrE, a small multidrug resistance transporter, is a model system for studying the transport process. Sequence conservation and prior NMR data suggest that the C-termini of EmrE may be functionally important. Unfortunately, currently available structures of EmrE lack information about the C-termini. Here we use the highly conserved C-terminal His110 residue as a reporter to probe the chemical environment of the C-termini in EmrE. Our results show that His110 from the two monomers within the asymmetric dimer have different upshifted pK_a values and dynamic behaviors. These results reveal that the C-termini of the two monomers are not just flexible, water exposed tails but interact with the rest of the asymmetric dimer, providing the first insight into the structure and environment of the highly conserved C-terminal histidine.

Introduction

The C-terminal residue of EmrE, His110, is highly conserved,[1-3] suggesting that it may be functionally important. Other residues conserved to this degree have been shown to be critical to function: Tyr60 and Trp63 (substrate recognition), Gly97 (dimerization), and Gly65 and Gly67 (transmembrane helix 3 hinge). However, the role of His110 remains ambiguous. One potential role of His110 is in determining the membrane topology of EmrE due to its partial positive charge. EmrE has very low charge bias, consistent with its dual topology in the membrane, and addition of a single positively charged residue or multiple histidines to the C-terminus has been shown to alter the topology of the entire protein.[4] However, the contribution of His110 to net charge bias does not sufficiently explain its conservation as the C-terminal amino acid in many

SMR family members since most of the other charged residues in EmrE are not highly as conserved. This suggests that His110 may have an additional function.

A high resolution structure of EmrE could provide hypotheses for the role of His110. Unfortunately, the conformational plasticity that enables EmrE to transport diverse substrates also makes it a very challenging system for high resolution structural studies.[5] As a result, there are currently only low-resolution cryoEM[6] and X-ray[7] structures available with no structural information on the C-terminal tail. Terminal regions of proteins are frequently missing from crystal structures due to their dynamic nature, and NMR is an ideal technique for studying flexible regions elusive to traditional structural approaches. Interestingly, the NMR spectra of EmrE suggest that the C-terminal region may have a unique structure or environment.[8] The residues in this region, including 104-108, have very unique chemical shifts that are not typical of an α-helix or a random coil, and the peaks are not as intense as they would be in an unstructured tail.[8,9]

Furthermore, the NMR spectra of EmrE in complex with different substrates show that the chemical shifts of residues 105, 106, and 108 from the C-terminal tail of one monomer in the homodimer are very sensitive to the identity of bound substrates while the residues in the tail of the other monomer are nearly insensitive to the bound substrate.[8] Without direct structural data it is difficult to understand this stark asymmetry. However, considering the orientation and position of the fourth transmembrane helix in the best available structures, it is possible that the C-terminus on the open side of the antiparallel dimer might extend into the substrate binding site while the C-terminal tail located on the closed side would be relatively isolated from the rest of the protein. Taken together, the experimental data available to date suggest that the C-terminus may have a functional role based on its high conservation and sensitivity to substrate identity.

To better understand the potential function of the C-terminal tail and conserved terminal histidine residue, we performed an NMR pH titration of the His110 side chains. These experiments probe the pK_a and dynamics of the highly conserved His110 and provide additional insight into the environment of the C-terminus of EmrE. Histidine side chains can exist in two neutral tautomeric states (ε and δ) and one charged state (+). Studies of model compounds have established a set of characteristic ^{15}N histidine chemical shifts for each of these states (Figure S1).[10-14] The protonated and deprotonated nitrogen atoms in either of the two neutral forms have chemical shifts of 167.5 ppm and 249.5 ppm and the two nitrogen atoms in the charged state resonate at 174 and 178 ppm. The chemical shifts of histidine side chains are readily recorded, even in large membrane protein systems such as EmrE, using HMBC experiments where the magnetization is transferred from the non-labile carbon ^{1}H through $^{2}J/^{3}J$ to ^{15}N. Rapid exchange between the different states of the histidine side chain, as commonly observed in aqueous solution[14,15], leads to ^{15}N chemical shifts that are a population-weighted average of all species present. ^{15}N chemical shifts of the histidine side chain as a function of pH can be analyzed to determine the pK_a value as well as the tautomerization equilibrium constant, $K_T = [\delta]/[\varepsilon]$.[13] By comparing the pK_a and K_T values with those of model compounds in water (pK_a of 6.2 and K_T of 0.25),[13] we can get insight into the chemical environment surrounding the histidine side chains.

Materials and Methods

Protein purification and sample preparation

The pET15b vector expresses EmrE with a N-terminal hexhistidine tag that can be removed by thrombin cleavage to leave three extra N-terminal residues (GSH).[7] EmrE was overexpressed, purified, and reconstituted into DMPC/DHPC (q=0.33) isotropic bicelles as previously

described.[9,16] Briefly, BL21-DE3 cells containing EmrE expression plasmid were grown in M9 minimal media. ^{15}N labeling media was supplemented with 2 g/L glucose and 1 g/L $^{15}NH_4Cl$. ^{2}H, ^{15}N labeling was grown in D_2O M9 media with 4 g/L glucose, 1g/L $^{15}NH_4Cl$, 0.5g/L ^{2}H,^{15}N ISOGRO (Sigma Isotec), and one generic multivitamin per liter. EmrE overexpression was induced with 0.33 mM IPTG at an OD_{600} of ~1.0 at 17 ˚C for ~16 hours. His-tagged EmrE was purified in n-Decyl-β-D-maltopyranoside (Anatrace) via Ni-affinity chromatography and the tag was removed by thrombin cleavage. The tag-less EmrE was then subject to size exclusion chromatography (Sephadex 200) and pooled fractions were reconstituted into DMPC/DHPC isotropic bicelles.

^{1}H-^{15}N HMBC NMR spectroscopy

All NMR experiments were performed on a 700 MHz Varian NMR spectrometer equipped with an HCN room temperature probe. ^{15}N labeled NMR samples in DMPC/DHPC bicelles (100 DMPC lipids per EmrE dimer) containing 1.2-1.5 mM EmrE (monomer concentration) and 2 mM TPP^+ were buffered by 20 mM acetate, 50 mM MOPS, and 50 mM bicine. NMR samples also contained 10% D_2O, 0.05% NaN_3. DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was used for ^{1}H chemical shift referencing and ^{15}N chemical shifts were indirectly referenced. Sample pH was adjusted using a pH electrode calibrated at the same temperature and was measured again after each NMR experiment to account for any pH drift. pH drift during 2D NMR experiments was only observed for titration points above pH 8 and the average pH value was used.

Two-dimensional ^{1}H-^{15}N HMBC spectra of TPP^+-bound EmrE were recorded from pH 5.1 to pH 8.8 at 45 ˚C using a HMQC pulse sequence. The delay during which ^{1}H and ^{15}N magnetizations

become antiphase was set to 22.8 ms so that magnetizations from $^{1}J_{NH}$ couplings are refocused.[10] In the HMBC experiment, magnetizations start from non-labile carbon protons (Hε1 and Hδ2) and are transferred through $^{2}J/^{3}J$ to ^{15}N atoms (Nε2 and Nδ1) in histidine side chains. So, for each histidine residue in a unique environment, there are two ^{1}H and two ^{15}N chemical shifts corresponding to at most four peaks (the Hδ2-Nδ1 peaks are sometimes not seen as a result of the smaller ^{3}J coupling).[10-12,14]

Paramagnetic Relaxation Enhancement ^{1}H-^{15}N TROSY HSQC experiments

Four separate uniformly deuterated, ^{15}N labeled NMR samples were prepared as described above. Samples contained 0.6-0.8 mM EmrE monomer as well as 20 mM NaCl, 0.05% NaN_3, 10% D_2O, 2 mM TCEP, 2 mM TPP^{+}, and 2 mM DSS and were buffered by 100 mM MOPS, 100 mM bicine, and 20 mM acetate. Each sample was adjusted to pH at 35°C as described for the HMBC experiments. ^{1}H-^{15}N TROSY HSQC spectra were collected on each sample with a 700 MHz Varian Inova spectrometer at 35°C. Following sample acquisition, the pH was checked again to confirm pH stability and the pH reported are averages of the pre-run and post-run pH measurements. To assess paramagnetic relaxation effects, 0.5 mM $MnCl_2$ was added to each sample[17-19] and another, identical, spectrum was collected in the same manner.

Spectra processing

All spectra were processed with NMRPipe[20] and analyzed using CcpNmr Analysis[21].

Interpretation of ^{1}H-^{15}N HMBC spectra

Histidine side chains exist in equilibrium between the charged state (+) and two neutral states (δ and ε) where the two neutral states are in a tautomeric equilibrium. The protonation equilibrium constant K_a is given by $[\delta+\varepsilon][H^{+}]/[+]$, while the tautomerization constant K_T is given by $[\delta]/[\varepsilon]$. The recorded ^{15}N chemical shifts of histidine side chains contain a wealth of information about

their protonation and tautomerization equilibria since the different states have very characteristic ^{15}N chemical shifts[10,11,13] (Fig. 5.1). In the charged state (+), both nitrogen atoms of histidine side chains have chemical shifts at around 176 ppm; this is an α+ type nitrogen. In the neutral states (δ and ε), protonated nitrogen atoms (Nε2 for ε state and Nδ1 for δ state) have α type chemical shifts of 167.5 ppm while unprotonated nitrogen atoms (Nδ1 for ε state and Nε2 for δ state) have β type chemical shifts at around 249.5 ppm. So at the low pH limit where the histidine is in the charged state (+), the spectra from HMBC experiments will have four peaks with an α+ type nitrogen chemical shift for both ^{15}N atoms of a model compound.[13] At the high pH limit, if there is a preference for ε or δ, there will be both α and β types of chemical shifts and the downfield peaks will correspond to β type nitrogens (Nδ1 or Nε2 for ε or δ preferences, respectively). If there are both ε and δ states at the high pH limit, the measured chemical shifts will be population-weighted averages of both states, since solvent exchange is fast in most occasions.[14,15] At intermediate pH, the chemical shifts will be population-weighted averages of all exchanging states including the charged state (+) and the neutral states (ε and δ).

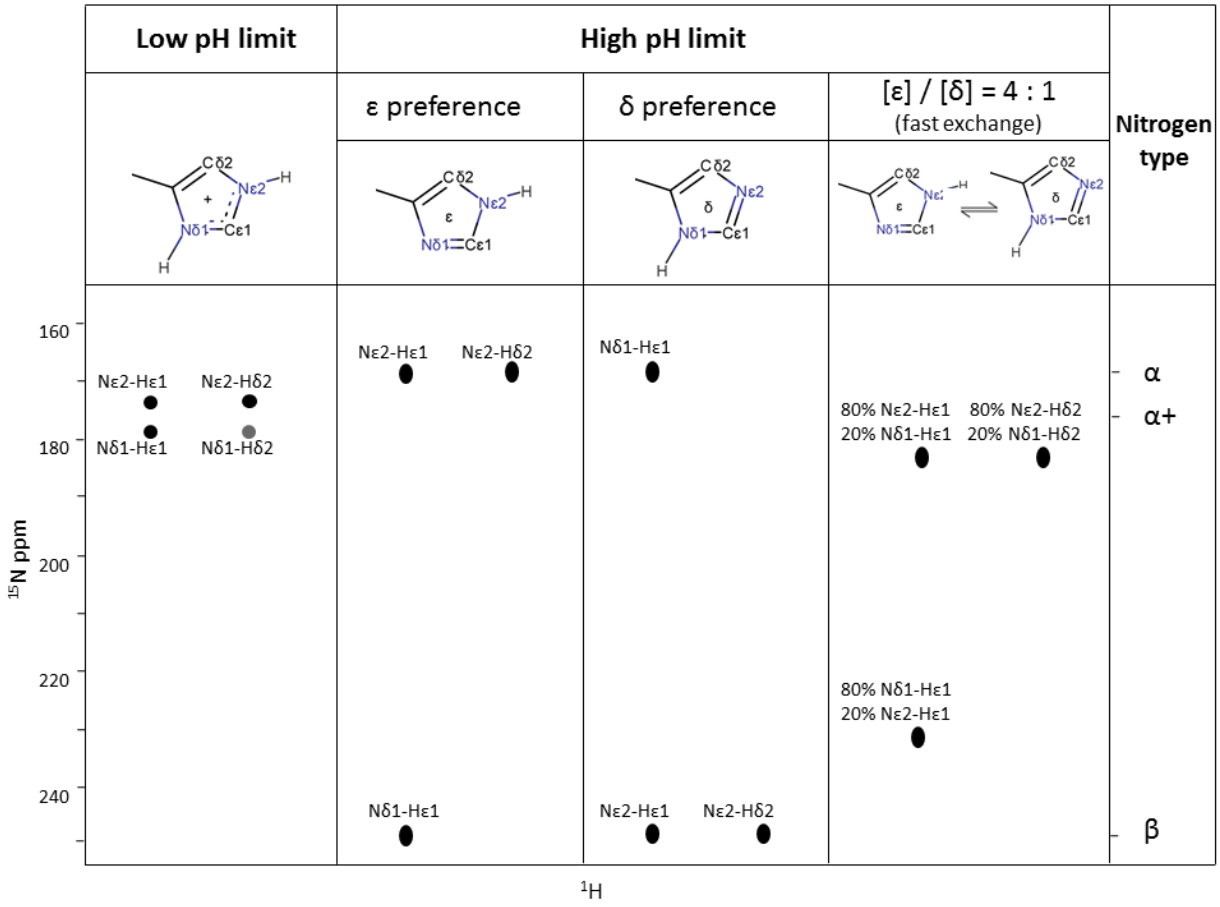


Fig. 5.1 ^{1}H-^{15}N HMBC peak patterns expected for different states of a model histidine side chain.

Determination of pK_a and K_T values from pH titration spectra

The pK_a and K_T were determined using a previously published method.[13] Briefly, ^{15}N chemical shifts at various pH values are fit to the modified Henderson-Hasselbalch equation,

$$CS_{obs} = CS_{high} \times \frac{1}{1+10^{(pKa-pH)}} + CS_{low} \times (1 - \frac{1}{1+10^{(pKa-pH)}}) \qquad (1)$$

where CS_{obs} is the observed chemical shift. The chemical shifts at extreme low and high pH are represented by CS_{low} and CS_{high} respectively. CS_{high} is the population-weighted average of the ^{15}N chemical shifts of both neutral states.

Using the value of CS_{high} determined from equation (1), the proportion, P, of ε or δ states is calculated from,

$$P = \frac{249.5 - CS_{high}}{249.5 - 167.5} \quad (2)$$

using typical chemical shifts of model compounds as reference values (167.5 ppm for α type and 249.5 ppm for β type nitrogen atoms). The tautomeric equilibrium constant, K_T, is given by the following equation.

$$K_T = \frac{P(\delta)}{P(\epsilon)} \quad (3)$$

The measured chemical shifts for both Nδ1 and Nε2 at different pH values for each residue were fit using equation (1) with a single pK_a parameter. The residue-specific K_T constants were then estimated from equations (2) and (3).

Results and discussion

We have collected twelve ^{1}H-^{15}N HMBC spectra from pH 5.1 to 8.8 for TPP^{+}-bound EmrE in DMPC/DHPC isotropic bicelles (q = 0.33). For the construct we use in our experiments, there are three additional residues remaining at the N-terminus after thrombin cleavage: GSH.[7] The N-terminal His-1 residue serves as a convenient internal control in this study. Since there are two His residues per monomer and two monomers per asymmetric EmrE dimer, we expect to see a maximum of four sets of peaks, assuming each histidine has only one major conformational state. Each set will in turn be composed of at most four peaks since the magnetization can start from either Hε1 or Hδ2 and is transferred through $^{2}J/^{3}J$ to Nδ1 and Nε2. Fewer peaks may be observed due to weaker ^{3}J coupling and/or dynamic exchange on an intermediate timescale. The resonance assignments for His110 and His-1 were made by performing a similar set of pH

titrations for a mutant with His-1 replaced by alanine (His-1Ala EmrE). Assignment to individual monomers could not be made since His110 is preceded by Pro109, which precludes through-bond assignment methods, so the A, B designations for the two monomers within the dimer are arbitrary throughout this chapter.

Across the pH range studied, both N-terminal His-1 residues behave like model compounds and there is little difference between the two His-1 residues in either chemical shift or dynamic behavior (Fig. 5.2). At pH 5.10, there are four peaks for each His-1 residue (connected with grey lines), and the two upfield peaks from both His-1 residues have degenerate chemical shifts suggesting they are experiencing very similar chemical environments. As the pH increases, the proportion of histidine side chains in the neutral states will increase, causing the ^{15}N chemical shifts of all four peaks to move downfield. The Nδ1 peaks will move faster as long as the ε state is the preferred neutral state, as observed here. At around neutral pH, the two upfield peaks for the two His-1 residues remain very close. Meanwhile, both lower-right peaks (Nδ1-Hδ2) disappear possibly as a result of increased protonation equilibria in the context of an already small ^{3}J coupling. The downfield movement of peaks continues as the pH increases and plateaus above pH 8. Note that as the chemical shift differences between the Nδ1 and Nε2 peaks become really large at pH 8.8, the peak widths increase significantly, implying that the exchange between the two neutral states is no longer in the fast regime.

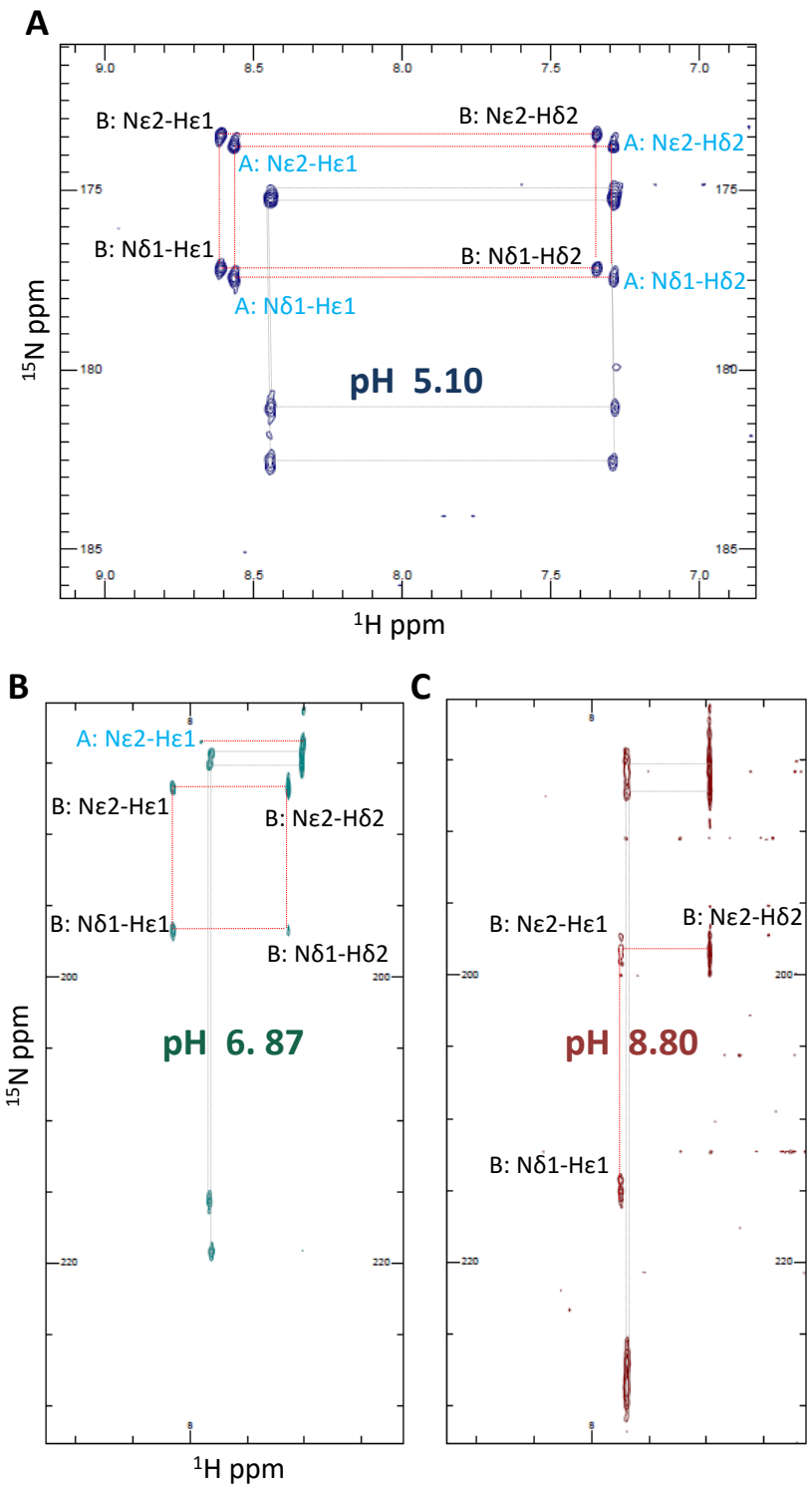


Fig. 5.2 ^{1}H-^{15}N HMBC spectra of TPP^{+}-bound WT EmrE collected on a Varian 700 MHz spectrometer with a room temperature probe at varying pH: (A) pH 5.10 ± 0.02, (B) pH 6.87 ± 0.00, and (C) pH 8.80 ± 0.15. Peaks belonging to His110/His-1 are labeled and connected with red/grey dotted lines. The assignments of Nε2 and Nδ1 peaks were made by comparing relative peak intensities. A, B designations are arbitrary here. Assignments were made by comparison with spectra of His-1Ala EmrE.

To quantitatively estimate the pK_a and K_T values, we fitted the ^{15}N chemical shifts over this pH range using equations 1-3 as described in more detail in methods section.[13] The modified Henderson-Hasselbalch equation describes the chemical shift as a function of pH:

$$CS_{obs} = \frac{CS_{high}}{1+10^{(pK_a - pH)}} + CS_{low}\left(1 - \frac{1}{1+10^{(pK_a - pH)}}\right) \quad (1)$$

where CS_{obs} is the observed chemical shift and CS_{low} and CS_{high} represent the extreme chemical shifts at high and low pH, respectively. Since CS_{high} is the population weighted average of the ^{15}N chemical shifts of both neutral states, the proportion of ε or δ states can be approximated by,

$$P = \frac{249.5 - CS_{high}}{249.5 - 167.5} \quad (2)$$

with reference to the ε or δ state chemical shifts of model compounds. The tautomeric equilibrium constant, K_T, is then simply

$$K_T = \frac{P(\delta)}{P(\epsilon)} \quad (3).$$

As can be seen in Fig. 5.3, the titration series for the two His-1 residues can be well fit. The same results are obtained whether the two nitrogens are analyzed individually or simultaneously for each residue. Here we only report the globally fit parameters. The two His-1 residues have pK_a values of 6.26 ± 0.03 and 6.13 ± 0.03, similar to model compounds (pK_a of 6.2). The K_T values for the two His-1 residues are 0.37 and 0.33, which are also close to model compounds (K_T of 0.25). Both these quantitative results and the qualitative behavior of the His-1 peaks in the NMR spectra suggest that the twoHis-1 residues exist in a symmetric, water-exposed environment, consistent with a flexible tail that does not interact with the asymmetric EmrE dimer.

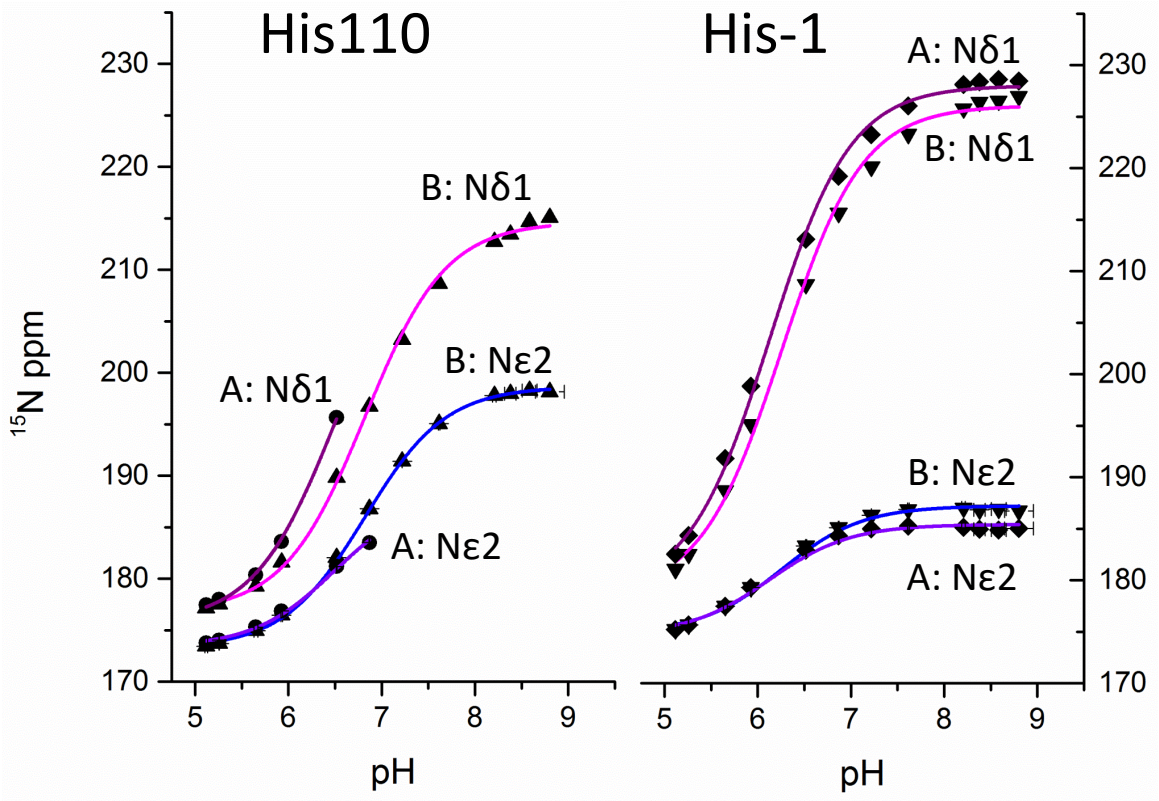


Fig. 5.3 Histidine ^{15}N chemical shifts as a function of pH (symbols) together with the global fitting (lines). Error bars represent standard errors and are shown only when they are larger than the size of the symbol.

Based on recorded chemical shifts and titration behaviors, the two C-terminal His110 residues are clearly different from His-1 and are also different from each other (Fig. 5.2). At pH 5.10, the two His110 residues have more downfield chemical shifts compared to His-1, implying the His110 residues are more protonated and have higher pK_a values. As the pH increases, differences between the two His110 residues become apparent. At pH 6.87, His110B still has four visible peaks that are distinct from the other three histidines, while His110A only has two upfield peaks visible with chemical shifts more similar to His-1. His110A completely disappears from the spectra above neutral pH. Given that the chemical shift difference between protonated and unprotonated ^{15}N corresponds to 5.7 kHz on a 700 MHz spectrometer, this signal loss suggests there are significant dynamics on a microsecond-millisecond timescale. At pH 8.80,

well above the pK_a, His110B differs dramatically from His-1 and the spectral pattern indicates that His110B not only has a significantly upshifted pK_a but also a different K_T value. The fitted pK_a value for His110B is 6.82 ± 0.01, and the K_T value is ~0.69 (Fig. 5.3), which indicates more of the δ species is present in the neutral state than is typical for a model compound. The stabilization of the δ species could come from H-bonding with any electronegative group of the system, including the C-terminal carboxylate group.[13] For His110A, as a result of a lack of data above pH 6.87, the pK_a cannot be confidently determined but is estimated to be in the range of 6.3-6.6, somewhat upshifted from model compounds. Nevertheless, both the titration behaviors and the fitted pK_a values demonstrate that the two C-terminal His110 residues are experiencing different environments.

To get more insight into the environment surrounding His110, we measured solvent accessibility of the backbone of TPP^+-bound EmrE by determining the paramagnetic relaxation enhancement broadening effect of Mn^{2+} [17-19] on the ^1H-^{15}N TROSY-HSQC across the same pH range (Fig. 5.4). At pH 5.2 and 6.0, the His110 signals for both monomers are not fully relaxed by the paramagnetic ion, indicating they are partially protected from the solvent. However, at pH 7.1 and 8.4, the signals become fully relaxed. While it is not possible to deduce what interaction could be protecting the backbone from solvent at low pH, this data does begin to provide an explanation for the shifted pK_a values: the protonated state appears to be interacting with either the membrane or the protein itself. Since this technique is monitoring the backbone and is not sensitive enough to detect minor differences, it is not surprising that we cannot see a distinction between the environments of the two His110.

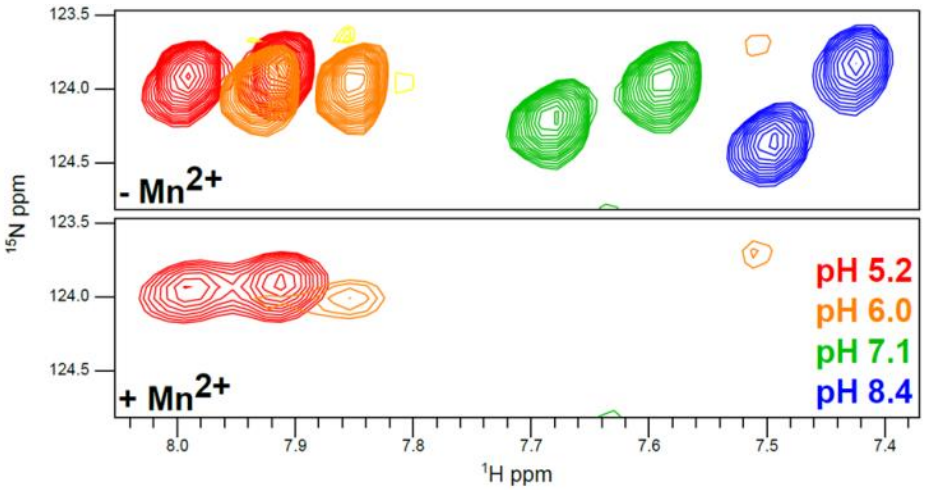


Fig. 5.4 His110 peaks in the ^{1}H-^{15}N TROSY-HSQC of TPP^{+}-bound EmrE. At each pH, spectra were collected first in the absence and then in the presence of the paramagnetic ion, Mn^{2+}.

Conclusion

In summary, we have made use of the ^{15}N chemical shifts of histidine side chains to probe the chemical environments of the two C-termini of TPP^{+}-bound EmrE. Our data reveal that the two His110 residues each display a distinct dynamic behavior and both have upshifted pKa values. This is in sharp contrast with the two N-terminal His-1 residues which behave very similarly to each other as well as to model compounds. These findings suggest that the two C-termini of the asymmetric EmrE dimer are not simply disordered, solvent-exposed tails, a claim which is further supported by the observed protection from a soluble paramagnetic agent at low pH. However, if the upshifted pK_a values were simply due to stabilization of the protonated state by a relatively non-specific interaction with the water-lipid interface, a phenomenon not uncommon to histidines in membrane environments,[22] one would expect the two histidines to display very similar dynamic profiles and pK_a values. Likewise, a simple interaction with the carboxyl groups of each terminus should also have similar effects on both His110. The fact that the dynamics of

the two C-termini are distinct suggests that interaction with the asymmetric EmrE dimer is likely contributing to the unique environments of the two His110 residues. Although the data presented here cannot speak directly to the nature of this interaction, it is interesting to note that the timescale of dynamics which is the likely cause of signal loss from one of the two His110 is not in the open-in to open-out conformational exchange regime[9] but is of the same order of magnitude as loop dynamics and ring flipping. Interaction with either the loops or one of the several aromatic residues in the substrate binding site or even the substrate itself could lead to dynamics on this timescale. As it has previously been observed that residues in one of the C-termini are sensitive to the drug identity,[8] it is tempting to speculate that one of the C-termini may interact with the protein near the drug binding site. However, further information, such as a high-resolution structure, will be needed to test this possibility. Nonetheless, this research begins to reveal the environment of the highly conserved C-terminal histidine and highlights the need to consider the role of the C-termini in future mechanistic studies.

Author contributions

Katherine A. Henzler-Wildman, Chao Wu, and Anne E. Robinson designed the experiments Chao Wu collected and analyzed the histidine pH titration data. Anne E. Robinson collected and analyzed the PRE data. All authors contributed to the writing.

References

1 Brill, S., Sade-Falk, O., Elbaz-Alon, Y. & Schuldiner, S. Specificity determinants in small multidrug transporters. *Journal of molecular biology* **427**, 468-477, doi:10.1016/j.jmb.2014.11.015 (2015).

2 Elbaz, Y., Salomon, T. & Schuldiner, S. Identification of a glycine motif required for packing in EmrE, a multidrug transporter from Escherichia coli. *The Journal of biological chemistry* **283**, 12276-12283, doi:10.1074/jbc.M710338200 (2008).

3 Ninio, S., Rotem, D. & Schuldiner, S. Functional analysis of novel multidrug transporters from human pathogens. *The Journal of biological chemistry* **276**, 48250-48256, doi:10.1074/jbc.M108231200 (2001).

4 Seppälä, S., Slusky, J. S., Lloris-Garcerá, P., Rapp, M. & von Heijne, G. Control of membrane protein topology by a single C-terminal residue. *Science (New York, N.Y.)* **328**, 1698-1700, doi:10.1126/science.1188950 (2010).

5 Korkhov, V. M. & Tate, C. G. Electron crystallography reveals plasticity within the drug binding site of the small multidrug transporter EmrE. *Journal of molecular biology* **377**, 1094-1103, doi:10.1016/j.jmb.2008.01.056 (2008).

6 Fleishman, S. *et al.* Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *Journal of molecular biology* **364**, 54-67, doi:10.1016/j.jmb.2006.08.072 (2006).

7 Chen, Y.-J. *et al.* X-ray structure of EmrE supports dual topology model. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18999-19004, doi:10.1073/pnas.0709387104 (2007).

8 Morrison, E. A. & Henzler-Wildman, K. A. Transported substrate determines exchange rate in the multidrug resistance transporter EmrE. *The Journal of biological chemistry* **289**, 6825-6836, doi:10.1074/jbc.M113.535328 (2014).

9 Morrison, E. *et al.* Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature* **481**, 45-50, doi:10.1038/nature10703 (2012).

10 Van Dijk, A. A., Scheek, R. M., Dijkstra, K., Wolters, G. K. & Robillard, G. T. Characterization of the protonation and hydrogen bonding state of the histidine residues in IIAmtl, a domain of the phosphoenolpyruvate-dependent mannitol-specific transport protein. *Biochemistry* **31**, 9063-9072, doi:10.1021/bi00152a050 (1992).

11 Pelton, J. G., Torchia, D. A., Meadow, N. D. & Roseman, S. Tautomeric states of the active site histidines of phosphorylated and unphosphorylated IIIGlc, a signal transducing protein from Escherichia coli, using two dimensional heteronuclear NMR techniques. *Protein Science* **2**, 543-558, doi:10.1002/pro.5560020406 (1993).

12 Singer, A. U. & Forman-Kay, J. D. pH titration studies of an SH2 domain-phosphopeptide complex: unusual histidine and phosphate pKa values. *Protein science : a publication of the Protein Society* **6**, 1910-1919, doi:10.1002/pro.5560060912 (1997).

13 Shimahara, H. *et al.* Tautomerism of histidine 64 associated with proton transfer in catalysis of carbonic anhydrase. *Journal of Biological Chemistry* **282**, 9646-9656, doi:10.1074/jbc.M609679200 (2007).

14 Hass, M. A., Hansen, D. F., Christensen, H. E., Led, J. J. & Kay, L. E. Characterization of conformational exchange of a histidine side chain: protonation, rotamerization, and tautomerization of His61 in plastocyanin from Anabaena variabilis. *Journal of the American Chemical Society* **130**, 8460-8470, doi:10.1021/ja801330h (2008).

15 Sehgal, A., Duma, L., Bodenhausen, G. & Pelupessy, P. Fast Proton Exchange in Histidine: Measurement of Rate Constants through Indirect Detection by NMR Spectroscopy. *Chemistry-A European Journal* **20**, 6332-6338, doi:10.1002/chem.201304992 (2014).

16 Morrison, E. & Henzler-Wildman, K. Reconstitution of integral membrane proteins into isotropic bicelles with improved sample stability and expanded lipid composition profile. *Biochimica et biophysica acta* **1818**, 814-820, doi:10.1016/j.bbamem.2011.12.020 (2012).

17 Gong, X.-M. M., Ding, Y., Yu, J., Yao, Y. & Marassi, F. M. Structure of the Na,K-ATPase regulatory protein FXYD2b in micelles: implications for membrane-water interfacial arginines. *Biochimica et biophysica acta* **1848**, 299-306, doi:10.1016/j.bbamem.2014.04.021 (2015).

18 Porcelli, F. *et al.* Structure and orientation of pardaxin determined by NMR experiments in model membranes. *The Journal of biological chemistry* **279**, 45815-45823, doi:10.1074/jbc.M405454200 (2004).

19 Schievano, E. *et al.* pH-Dependent conformational changes and topology of a herpesvirus translocating peptide in a membrane-mimetic environment. *Biochemistry* **43**, 9343-9351, doi:10.1021/bi0496438 (2004).

20 Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* **6**, 277-293 (1995).

21 Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696, doi:10.1002/prot.20449 (2005).

22 Martin, B. U., Mark, S. P. S. & Alfredo Di, N. Properties of integral membrane protein structures: Derivation of an implicit membrane potential. *Proteins: Structure, Function, and Bioinformatics* **59**, 252-265, doi:10.1002/prot.20334 (2005).

Chapter 6: Conclusion and Future Direction

My doctoral thesis focuses on a small multidrug resistance transporter, EmrE, which is a model system for secondary active transporters. Despite decades of intensive studies, mechanistic understanding about how EmrE couples proton import and drug export remains rather immature[1]. This results from, to a large extent, a lack of atomic descriptions for different structural states in the transport cycle and the dynamic processes linking these structures states, since transport is inherently a dynamic process. Therefore, my thesis studies both structure and dynamics of EmrE in order to provide mechanistic insights about proton-coupled multidrug transport. This is enabled by the use of NMR, a powerful technique to study structure and dynamics simultaneously with atomic resolution. Complemented by other techniques, our structure and dynamics studies have made good progresses towards building a higher-resolution structure of EmrE and understanding the physicochemical basis underlying conformational interconversion central to transport processes in general.

We have overcome several challenges involved in NMR structure determination of EmrE. The conformational interconversion, where EmrE exchanges between open-in and open-out, is constantly ongoing. This makes the long-sought high-resolution structure of EmrE extremely challenging to obtain by any structural approach. The relatively small size and constant dynamics of EmrE make NMR the best approach for structural studies. Yet, the conformational interconversion still poses significant problems for NMR structure determination which can only tolerate certain levels of dynamics. Therefore, through collaborative screening, we set out to find potential dynamics mutants and identified EmrE-S64V as a slow dynamics mutant suitable for structural studies. Though this mutant makes NMR structure determination of EmrE possible, this is still a highly challenging project. Current NMR structures are mostly for molecules with

molecular-weight less than 25 kDa, and NMR structure determination of helical membrane proteins are further complicated by signal overlap and limited tertiary restraints. Therefore, considering the asymmetric dimer of EmrE has eight unique transmembrane helices, we have made very good progress by assigning most of backbone and side chain methyl resonances. These resonance assignments were only possible using the most cutting-edge technologies including sparse sampling and specific labeling designed for large and challenging systems. This is aided by our novel approach making use of the often challenging factors inherent to EmrE, e.g. the use of ZZ-exchange data from conformational interconversion to assign pairs of amide resonances in combination with backbone-walk data. Our assignment processes and mechanistic insights are also coupled. This is reflected by the convergent chain assignment within the asymmetric EmrE dimer aided by the pH titration data and concurrent realization that protonation of E14 is asymmetric, causing systematic peak shifts in monomer B, which is mechanistically relevant to proton coupling. Another beneficial factor in our assignment process is the combination of experiment and theory through integration of our NMR data with collaborative modeling studies. These modeling studies not only provided good starting points for structure-based assignment, but also provided feedback on inconsistencies for refinement during the iterative assignment process.

There are very limited NMR studies for systems as complicated as this one. My research helps push the boundary on several frontiers of biomolecular NMR structure and dynamics of asymmetric, dynamic assemblies and integral membrane proteins, which will be of general interest to the broader biomolecular NMR community. I have tried various labeling strategies for side chain resonance assignment designed for large systems with mixed success, and these lessons will help future studies of membrane proteins, which have some distinct challenging

features from large soluble protein complexes. For example, our strategy of initiating more extensive side chain resonance assignment from methyl groups is broadly applicable for membrane protein systems. However, due to the hydrophobic environment of membrane proteins, a significant portion of methyl groups face lipids and therefore are not as valuable as those in soluble proteins, necessitating an approach that goes beyond backbone and methyl groups. Nevertheless, the technological advances in ILV methyl labeling provide starting point to bootstrap assignment of other side chain resonances. Methyl groups also provide tertiary restraints for initial definition of helix packing.

Our methyl assignments are only possible because we have collected high quality methyl-detected out-and-back and amide-detected spectra. It is worth to point out that the high quality of these methyl assignment experiments based on COSY was unexpected considering the general pessimistic impression from the literature. This case encourages future research into this COSY-transfer route for large systems.

My work also highlights the importance of aromatic residues in NMR structure determination of membrane proteins, which should be seriously considered for future studies. This is in part because of the limited amount of restraints that can be collected for helical membrane proteins due to requirements for deuteration and the fact that many methyl groups face lipids and do not contribute long-range restraints for 3D structure determination. This is especially important considering the controversies involved in several membrane protein NMR structures. Despite the challenges involved, side chain restraints from methyl and aromatic groups should both be considered for future NMR structure determination of helical membrane proteins.

Our large sets of NOESY spectra on samples with various labeling schemes will provide a reasonable number of distance restraints. In addition, these distance restraints are further complemented with orientational restraints from high precision RDCs measured using strained-gel induced alignment.

To finish the structure, iterative assignment of the large sets of NOESY spectra must be completed along with structural calculation. Structure calculations will be aided by the starting structures from our collaborative modeling studies and the newly implemented implicit membrane model within Xplor-NIH[2]. Ideally, more extensive testing of the resulting final structures in explicit membrane environment should be performed by continued collaboration. These modeling studies certainly will continue to offer insights and generate hypotheses for future experimentation. For example, further NMR and MD research will likely contribute key insights to the proton-coupling mechanism in atomic detail. How is the proton transferred within the pore and how is this affected by the presence of drug? How does the C-terminus from monomer A behave with different protonation states of both E14 and H110 residues? What is the hydration pattern within the pore under different conditions? How will the presence of membrane potential change the system?

The EmrE-S64V mutant also makes structural studies of drug-free states possible. The next structure(s) that could be pursued should be the proton-bound state(s). After the first structure is solved, it will be much easier to solve related structures. The help from modeling will reduce the barrier even more as well. These structures will provide further information about the structural basis of the proton-coupling mechanism, e.g. how EmrE avoids leaking proton even though it can interconvert between open-in and open-out in multiple protonation states.

Our dynamic studies of the series of S64 mutants, following the identification of EmrE-S64V, demonstrate for the first time that mutations near the TM3 kink in EmrE alter rates of global conformational exchanges dramatically. This provides experimental confirmation of the hypothesis that the TM3 kink of EmrE serves a as hinge enabling global conformational transition needed for transport activity. In addition, our characterization of the mutant series together with the ligand series suggest a very interesting trend that the interconversion rate decreases with increasing hydrophobicity within the pore. These data pointed me to the importance of hydration for enabling the helix kinking and unkinking required for conformational interconversion. What is more, the hydration concept could be more broadly involved in the overall transport process. The S64V mutation not only changes the rate of interconversion, but also perturbs the pK_a values of E14 by more than one pH unit, suggesting that hydration could fine tune important steps involving proton transport as well. This observation makes hydration a unifying factor to understand the proton coupling mechanism of EmrE. Furthermore, the multidrug transport activity of EmrE has several rate-limiting steps, e.g. drug off-rate and interconversion rate, which are orders of magnitude different for different ligands. Hydration could provide the underlying explanation for how EmrE achieves proton-coupled multidrug transport with such different kinetics of individual steps by simultaneously tuning the pK_a shift upon substrate binding.

Further structural characterization of the TM3 kink under a variety of conditions, combining different substrates and kink region mutants will help develop future mechanistic insights into the interconversion process. To get more precise measurements of helix orientation information for EmrE, oriented solid-state NMR experiments such as PISEMA should be performed for specific amino acid- labeled EmrE samples[3].

In order to get direct information about the hydration pattern around kink region, NMR H/D experiments and molecular dynamics simulation for several mutants should be performed. Also ITC experiment in H_2O vs D_2O[4-6] would directly probe the solvent reorganization contribution involved when comparing different mutants. Neutron scattering[7] might also be helpful if changes in hydration are significant enough. Interpretation of these experimental data will benefit greatly from accompanying simulations that provide atomic structural and kinetic models of hydration within the transport pore.

More studies are needed to further explore the effect of ligands and mutants on the pK_a of E14. This will develop a deeper understanding of coupling between E14 protonation and TM3 helix kinking.

Most of our dynamic studies so-far focus on the interconversion, and therefore fall into the ms-s timescale regime. Studies that extend beyond these timescales are needed to fully link the dynamics to the function. For example, ps-ns dynamics of different mutants could provide further information about how slight changes in local flexibility in the TM3 hinge are linked to the rates of global conformational change. Also further CPMG dynamic studies on the sub-millisecond dynamics of the C-terminus (Chapter 5) may provide further insight into its role in transport.

I have aimed to understand the workings of this biological machinery by reconstituting it *in vitro*. While necessary for initial characterization, this provides a very simplified environment. In the case of EmrE, the symmetric bilayer environment in particular is an oversimplification. Adding another layer of complexity by introducing membrane asymmetry and membrane potential will be necessary to truly understand how EmrE works.

In conclusion, our structural and dynamic studies on EmrE not only push boundaries on many frontiers of NMR, but also demonstrate the usefulness as well as the need to employ the new paradigm of structure-dynamics-function to understand the mechanisms of inherently dynamic biological molecules.

References

1 Sanders, C. R. Perplexing new insight into the dynamics of the EmrE transporter. *The Journal of general physiology* **146**, 441-444, doi:10.1085/jgp.201511523 (2015).

2 Tian, Y., Schwieters, C. D., Opella, S. J. & Marassi, F. M. A Practical Implicit Membrane Potential for NMR Structure Calculations of Membrane Proteins. *Biophysical journal* **109**, 574-585, doi:10.1016/j.bpj.2015.06.047 (2015).

3 Gayen, A., Banigan, J. R. & Traaseth, N. J. Ligand-induced conformational changes of the multidrug resistance transporter EmrE probed by oriented solid-state NMR spectroscopy. *Angewandte Chemie (International ed. in English)* **52**, 10321-10324, doi:10.1002/anie.201303091 (2013).

4 Chervenak, M. C. & Toone, E. J. A Direct Measure of the Contribution of Solvent Reorganization to the Enthalpy of Binding. *Journal of the American Chemical Society*, doi:10.1021/ja00102a021 (1994).

5 Ozen, C., Norris, A. L., Land, M. L., Tjioe, E. & Serpersu, E. H. Detection of specific solvent rearrangement regions of an enzyme: NMR and ITC studies with aminoglycoside phosphotransferase(3')-IIIa. *Biochemistry* **47**, 40-49, doi:10.1021/bi701711j (2008).

6 Tzeng, S.-R. R. & Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **488**, 236-240, doi:10.1038/nature11271 (2012).

7 Krepkiy, D. *et al.* Structure and hydration of membranes embedded with voltage-sensing domains. *Nature* **462**, 473-479, doi:10.1038/nature08542 (2009).