2017

Structural study of membrane proteins by X-ray crystallography

Tsung-Han Chou
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd
Part of the Physics Commons

Recommended Citation
Chou, Tsung-Han, "Structural study of membrane proteins by X-ray crystallography" (2017). Graduate Theses and Dissertations. 17165. https://lib.dr.iastate.edu/etd/17165

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Structural study of membrane proteins by X-ray crystallography

by

Tsung-Han Chou

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Physics

Program of Study Committee:
Scott W. Nelson, Co-Major Professor
Kai-Ming Ho, Co-Major Professor
James W. Evans
Kristen M. Johansen
Jigang Wang
Xuefeng Wang

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

Copyright © Tsung-Han Chou, 2017 All rights reserved.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>CHAPTER 1 GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>General introduction</td>
<td>1</td>
</tr>
<tr>
<td>References</td>
<td>7</td>
</tr>
<tr>
<td>CHAPTER 2 CRYSTAL STRUCTURE OF MYCOBACTERIUM TUBERCULOSIS TRANSCRIPTIO</td>
<td>16</td>
</tr>
<tr>
<td>NAL REGULATOR RV0302</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>17</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>20</td>
</tr>
<tr>
<td>Conclusion</td>
<td>33</td>
</tr>
<tr>
<td>Materials and methods</td>
<td>36</td>
</tr>
<tr>
<td>References</td>
<td>43</td>
</tr>
<tr>
<td>CHAPTER 3 CRYSTAL STRUCTURE OF THE CHLAMYDOMONAS REINHARDTII LCi1 CHA</td>
<td>51</td>
</tr>
<tr>
<td>NEL</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>52</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>54</td>
</tr>
<tr>
<td>Methods</td>
<td>67</td>
</tr>
<tr>
<td>References</td>
<td>73</td>
</tr>
<tr>
<td>CHAPTER 4 STRUCTURAL STUDY OF RECEPTOR LIKE KINASE FERONIA OF ARABIDO</td>
<td>85</td>
</tr>
<tr>
<td>PSIS THALIANA</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>85</td>
</tr>
<tr>
<td>Results</td>
<td>87</td>
</tr>
<tr>
<td>Conclusion and discussion</td>
<td>98</td>
</tr>
<tr>
<td>Methods</td>
<td>101</td>
</tr>
<tr>
<td>References</td>
<td>107</td>
</tr>
<tr>
<td>CHAPTER 5 CONCLUSIONS AND FUTURE DIRECTIONS</td>
<td>111</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor Prof. Edward W. Yu for his support, motivation and patience throughout my Ph.D. study. His efforts towards shaping up my career gives me privilege to proudly accept that I could not have found a better advisor and mentor.

I would like to thank my committee chair, Prof. Scott W. Nelson and co-chair, Prof. Kai-Ming Ho for their patience and support in my last stage of my Ph.D. study as well as the rest of my committee members, Prof. James W. Evans, Prof. Kristen M. Johansen, Prof. Jigang Wang, and Prof. Xuefeng Wang, for their valuable advice from various perspectives. In addition, I would like to thank Prof. Drena Dobbs for her suggestions in my Preliminary exam.

I would like to thank all the lab members in Ed’s group, Dr. Chih Chia Su and Dr. Feng Long who taught me everything in our lab, Dr. Hsiang-Ting Lei and Dr. Jani Reddy Bolla for their supports and encouragements, Dr. David Lin and Dr. Sekhar Kambakam for their kind suggestions. Special thanks to my fellow lab mates Dr. Abhijith Radhakrishnan, Dr. Jared Delmar and Dr. Nitin Kumar for all the exciting discussions, countless encouragements, and enormous support in tough times.

I would like to thank all my collaborators during my graduate study. Dr. Kanagalaghatta R. Rajashankar, Dr. Surajit Banerjee and all the staff members of NECAT at Advanced Photon
Source, Argonne National Lab for X-ray data collection and structure determination. Prof. Georgiana E. Purdy for the studies of transcription regulators. Prof. Martin H. Spalding, Dr. Dan J. Stessman, and Alfredo Kono for LC11 functional study. Dr. Kannan Sankar for the computational simulation. Prof. Yanhai Yin and Dr. Hongqing Guo for the FERNOIA project. Special thanks to Dr. Bill Colonna and Dr. John Strohl for friendly support in yeast fermentation experiments.

I would like to thank my department chair Prof. Frank Krennrich and all the staff, especially Lori, Deb, Linda, and Gloria, in physics department for their kindness and assistance.

I would like to thank all my friends, my softball team in Ames for providing countless joy and lovely memories during these years.

I would like to thank my brother, my cousins, and parents for their support and encouragement, which provide me the strength and courage to continue my study. I would like to dedicate this dissertation to my mom and dad. Without their love and unconditional support, this accomplishment would be impossible.
ABSTRACT

Membrane proteins are essential for the cell to communicate with its environment. They function as a gateway across the lipid bilayer, allowing stimuli transmission and controlling molecular transport into or out of the cell. The structure of membrane proteins plays a pivot role in their function and mechanism. However, determination of membrane protein structure remains a great challenge due to difficulties associated with expression and purification. This dissertation focuses on utilizing X-ray crystallography to study the structure and the function of various membrane proteins from different biological systems. Chapter 2 explores the role of MmpL family transporters in the development of antibiotic resistance in *Mycobacterium tuberculosis*. The cell wall of *M. tuberculosis* is crucial to its virulence and antimicrobial resistance. The MmpL transporters are known to participate in cell wall formation by exporting fatty acid derivatives. We present the structural insights into the TetR transcriptional regulator Rv0302, which controls the expression of several MmpL proteins. Also, by combining functional studies and structural analysis, we demonstrate how the investigation of Rv0302 improves our understanding of substrate transport by the MmpL family proteins. Chapter 3 details our investigations of the the carbon concentrating mechanism (CCM) of the green algae *Chlamydomonas reinhardtii*. Due to their aquatic habitat, the photosynthetic efficiency of phytoplankton is hindered by the difficulty of maintaining sufficient supply of inorganic carbon (Ci). To overcome this hurdle, these microorganisms have developed CCM to enhance Ci uptake from its Ci limited environment. We reveal the structure of a CCM related membrane protein transporter LCI1, whose structure is the first membrane transporter solved in the CCM pathway of *C. reinhardtii* and
is unique in the protein database. Finally, in chapter 4 we apply our knowledge of protein
crystallography to study plant receptor-like kinase FERONIA from *Arabidopsis thaliana*.
FERONIA is known to play an important role in many plant signaling pathways such as
growth development, root growth, and drought response. We demonstrated different
expression, purification and crystallization attempts of acquiring structural information of
FERONIA and the plant hormones RALF1 and RALF23. While these efforts did not result in
a solved structure, based on our acquired experience with this system, we suggest several
directions for future structural study of membrane protein in higher evolved organisms.
CHAPTER I

GENERAL INTRODUCTION

Proteins are the most versatile macromolecules present in living organisms and essential to nearly all the biological activities. By interacting with other proteins, they perform various tasks to sustain biological processes, including catalysis, molecular transport, signaling, and structural element formation. In general, proteins are composed of a sequence of polymerized amino acids, which fold into a specific structure that defines the function of the protein. Thus, to understand the function of the proteins, it is crucial to obtain their structural information.

Currently, there are three major approaches for protein structure determination, which include nuclear magnetic resonance spectroscopy (NMR), X-ray crystallography, and cryo-electron microscopy. To date, X-ray crystallography has been the most reliable technique for solving protein structure. Up to now, more than 120,000 structures have been deposited to PDB and 90% of the protein structures are determined by X-ray crystallography (http://www.rcsb.org/pdb/statistics/holdings.do).

Since the first protein structure was determined half a century ago (myoglobin at 6 Å resolution),¹ X-ray crystallography has been the chief method for protein structure determination and the procedure of macromolecule crystallization has been standardized and described in great detail in many textbooks.²³⁴ In short, the proteins of interested are prepared by extracting them directly from the native organisms or purifying the recombinant protein by heterogeneous expression in various expression systems such as well-established bacterial expression host Escherichia coli, or in a variety of eukaryotic cells. To ensure the homogeneity of samples, the proteins are then further polished by size exclusion chromatography (SEC) or ion exchange chromatography (IEC) and then subjected to crystallization screens. Protein crystallization is
most often achieved by the vapor diffusion method\textsuperscript{5,6}. This method allows proteins to slowly reach the supersaturation and thus induces crystallization by equilibrating a droplet of the purified protein mixed with crystallization solution containing precipitants, salts, buffers, and additives, with a large reservoir with the similar crystallization solution at a higher concentration. By changing the combination and concentration of the components in the crystallization solution, the protein crystals can be optimized for suitable quality and size for data collection.

Thanks to the development of both software and hardware in data collection of protein X-ray crystallography,\textsuperscript{7,8,9,10} the time required for collecting diffraction data has been greatly reduced. However, a significant amount of crystal screening to acquire the best resolution data and experimental phase information is inevitable. Phase solving is essential to 3D electron density reconstitution in protein crystallography since the diffraction patterns provide only the intensity of the reflections. Phases are crucial to calculate the structural function in diffraction theory. There are several approaches to obtain the phase information of protein crystals, including acquiring anomalous signals to break Friedel’s law (Single/Multiple anomalous dispersion (SAD/MAD)),\textsuperscript{11} simplifying the diffraction patterns by introducing electron-rich (heavy) atoms into crystals (Single/Multiple isomorphous replacement (SIR/MIR)),\textsuperscript{11,12} matching the Patterson function with an existing model (Molecular replacement),\textsuperscript{13} direct method (applicable only for small peptides) and the combination of different methods (Single isomorphous replacement with anomalous signal (SIRAS)). Once the phase problem is solved, the protein model can be built based on the electron density map that reconstituted from the diffraction data and thus the reveals structure of the protein. Currently, there are many program packages available such as CCP4\textsuperscript{14} and Phenix,\textsuperscript{15} providing software for phase solving, model refining, and model building.
Despite the success in a variety of proteins, membrane protein structure determination is still a great challenge to X-ray crystallography. Among 70,000 unique protein structures in PDB, only less than 1% of them are membrane proteins (http://blanco.biomol.uci.edu/mpstruc/). Membrane proteins, which make up around 30% of all genes,\textsuperscript{16,17} are proteins that associate with or embed in the cell membrane. Some function as the bridges across the lipid bilayer, allowing stimuli transmission and molecular transport into or out of the cell. Also, more than 50% of modern drugs target membrane proteins,\textsuperscript{18} which makes the structural information of these proteins valuable to drug development. However, structural and functional studies of membrane proteins are often limited by difficulties in expression, purification, and crystallization.\textsuperscript{19}

Due to the low copies of membrane proteins in their native environment, recombinant overexpression is necessary to obtain a suitable quantity for crystallization. Membrane proteins are stable and well-folded only in the environment that is similar to their native environment. Also, membrane proteins, especially eukaryotic proteins are often modified by post-translational modification (PTM), including phosphorylation and glycosylation. Thus, choosing the appropriate expression host according to the properties of the target protein is important to acquire high quality recombinant protein.\textsuperscript{20} To date, several expression systems have established for membrane protein expression, including \textit{E. coli}, yeast (\textit{S. cerevisiae},\textsuperscript{21} \textit{Pichia}\textsuperscript{22}), insect cells (Sf9, Hi5),\textsuperscript{23} and mammalian cells (HEK293).\textsuperscript{24}

For membrane protein purification and crystallization, detergents are essential to extract proteins from lipid bilayer and stabilize proteins in the buffer by protecting the hydrophobic transmembrane domains from exposure to the solution. Generally, detergents used for membrane protein crystallization are an alkyl-chain between 7 and 14 carbons in length with varying headgroups.\textsuperscript{25} Among all the detergents, \textit{n}-dodecyl-D-maltoside (DDM) and \textit{n}-decyl-D-
maltoside (DM) are the most common for protein extraction. The use of detergents for protein extraction can be hinder protein crystallization. Generally, membrane proteins are more stable in detergents with longer alkyl-chain. These detergents occupy more space around the protein, which prevents it from interacting with another protein and thus prevent crystal packing. Therefore, detergent screen through gel filtration is important for optimizing the protein condition before being subjected to crystal screen. The choice of an appropriate detergent remains empirical, largely depending on the biochemical and structural properties of the target proteins.

Another strategy to crystallize membrane protein is lipid-protein co-crystallization. Introducing lipids increases the protein stability by mimicking the native membrane environment and thus improves crystal packing. Several related methods have been reported, including adding lipids throughout protein purification and crystallization, X-ray crystallography to study the function and the structure of membrane proteins in different biological systems, including bacterial drug resistance in Mycobacterium tuberculosis, Carbon Concentrating Mechanism (CCM) green algae in Chlamydomonas reinhardtii, and plant receptor-like kinase FERONIA in Arabidopsis thaliana. The brief introduction to these projects is described below.

Chapter II

For over a century, antimicrobial therapies have been the most effective ways for treating bacterial infections. Recently, however, these treatments have been complicated by the fast development of bacterial drug resistance. The bacterial resistance to antibiotics can be categorized into four major causes, which include alternating drug targeting site, enzymatic
degradation, decreased drug permeability by cell wall modification, and drug efflux by active transport.\textsuperscript{30,31} In \textit{Mycobacterium tuberculosis}, which infects a third of the population and causes more than 1 million deaths each year worldwide,\textsuperscript{32} the drug resistance is related to its unique cell wall.\textsuperscript{33} The cell wall of \textit{M. tuberculosis} comprises mycolic acids, which can only be found in Mycobacterium and play a crucial role in the virulence and antimicrobial resistance.\textsuperscript{34,35,36} Biosynthesis of this cell wall is facilitated by the Mycobacterium membrane protein large (MmpL) transporters,\textsuperscript{37} which belong to the resistance-nodulation-cell division (RND) superfamily.\textsuperscript{38} Previous studies have shown that these MmpL transporters are responsible for transporting fatty acid and lipid components of the cell wall.\textsuperscript{39,40,41,42} However, the regulation of MmpL protein expression and the role of MmpL proteins in cell wall remodeling has not been explored. In this chapter, we provide the structural insights that increase our understanding of the regulatory mechanism of TetR regulator Rv0302. We also study the regulation of MmpL proteins by Rv0302 with different biochemical and biophysical approaches. Moreover, we identify palmitic acid, a precursor of mycolic acids, is a possible natural ligand of Rv0302.

Chapter III

During photosynthesis, the essential element carbon is converted from the inorganic form, carbon dioxide (CO\textsubscript{2}), to carbohydrates by the carbon assimilation pathway. Since the carbon fixation enzyme Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) has low affinity for CO\textsubscript{2} and slow carboxylation rate,\textsuperscript{43,44,45,46} it is important for photosynthetic organisms to assure the concentration of CO\textsubscript{2} in the cell to maintain the photosynthetic efficiency. However, phytoplankton, which contribute more than half of total photosynthetic activity on Earth, find it difficult to retain the inorganic carbon (Ci) due to their aquatic habitats, where the CO\textsubscript{2} diffusion
rate is 10,000 times slower than the rate in air. To adapt to this hazardous environment, these microorganisms have developed the Carbon Concentrating Mechanism (CCM) to enhance Ci uptake from the Ci limited surrounding.\textsuperscript{47,48,49,50}

In unicellular green algae \textit{Chlamydomonas reinhardtii}, three membrane proteins involved in CCM have been characterized, including a bicarbonate (HCO$_3^-$) channel LCIA in Formate-Nitrate Transporter (FNT) family,\textsuperscript{51,52} a ATP binding cassette (ABC) transporter HLA3,\textsuperscript{53,54,55} and a putative transporter LCI1.\textsuperscript{56,57} HLA3 and LCIA, which are located at the plasma membrane and chloroplast membrane respectively, have been found to transport HCO$_3^-$ cooperatively into chloroplast envelope.\textsuperscript{58} However, LCI1, whose transformant has shown the increase of Ci uptake in \textit{C. reinhardtii}, appears to represent a completely novel type of transporter. In this chapter, we present the crystal structure of novel membrane protein LCI1, which is also the first structure of CCM in \textit{C. reinhardtii}, and provide the structural perspectives of elucidating its transport mechanism.

Chapter IV

Plant receptor-like kinase FERONIA of \textit{Arabidopsis thaliana} is known to play an important role in many plant signaling pathways such as growth development,\textsuperscript{59,60} root growth,\textsuperscript{61,62} and drought response.\textsuperscript{60,63} FERONIA belongs to receptor-like kinase superfamily (RLK),\textsuperscript{64} a group of single transmembrane proteins which consist of an extracellular sensory domain and a cytoplasmic kinase domain that phosphorylates their downstream objects to activate the response pathways of incoming signals. The sensory of FERONIA is found to possesses two malectin-like domains, a homolog domain of animal sugar binding protein that
interacts with various sugars. RALFs, a group of plant growth peptide hormones which involve in root growth response, are believed to be the ligands of FERONIA. However, the ligand sensing mechanism of FERONIA remains elusive due to the lack of the structural information. In addition, a drought response related transcription factors RD26 is activated by FERONIA kinase domain. In this study, we investigated the efficiency of different expression and purification methods to acquire suitable amounts of protein for crystallization of FERONIA sensory domain and plant hormones RALF1 and RALF23. In addition, we purified and crystallized the protein kinase domain of FERONIA and DNA binding domain of RD26.

References


15. PHENIX: a comprehensive Python-based system for macromolecular structure solution.
   Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty,
   R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H.

   proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science*, **7**(4),
   1029-1038.

   transmembrane topology with a hidden markov model: application to complete genomes.


   DOI: 10.1016/bs.mie.2014.12.018

20. Yuan He, Kan Wang, Nieng Yan, The recombinant expression systems for structure
   determination of eukaryotic membrane proteins Protein Cell. 2014 Sep; **5**(9): 658–672

21. Strausberg RL, Strausberg SL. Overview of protein expression in saccharomyces


23. Unger T, Peleg Y. Recombinant protein expression in the baculovirus-infected insect cell


*PLoS Pathog.* **1**, 12–19


*PLoS Pathog.* **1**, 12–19


CHAPTER II

CRYSTAL STRUCTURE OF MYCOBACTERIUM TUBERCULOSIS TRANSCRIPTIONAL REGULATOR RV0302


Tsung-Han Chou¹ †, Jared A. Delmar¹ †, Catherine C. Wright², Nitin Kumar³, Abhijith Radhakrishnan³, Julia K. Doh², Meredith H. Licon², Jani Reddy Bolla³, Hsiang-Ting Lei³, Kanagalaghatta R. Rajashankar⁴, Chih-Chia Su¹, Georgiana E. Purdy², and Edward W. Yu¹,³

¹Department of Physics and Astronomy, Iowa State University, Ames, Iowa, 50011
²Department of Molecular Microbiology and Immunology, Oregon Health and Sciences University, Portland, Oregon, 97239
³Department of Chemistry, Iowa State University, Ames, Iowa, 50011
⁴NE-CAT and Department of Chemistry and Chemical Biology, Argonne National Laboratory, Cornell University, Argonne, Illinois, 60439

*Correspondence to: Edward W. Yu, Department of Chemistry, Iowa State University, Ames, IA

†J.A.D. and T.H.C. contributed equally to this work.
Introduction

Tuberculosis (TB) is a leading cause of death due to infectious disease despite the availability of antitubercular drugs. Its causative agent, *Mycobacterium tuberculosis* (Mtb), infects more than one third of the world's population. The unique architecture of the mycobacterial cell wall plays a key role in the host–pathogen interface since it is associated with the Mtb pathogenesis and provides a barrier against environmental stresses, antibiotics, and the host immune response. The outer membrane contains an inner leaflet of very long chain mycolic acids covalently bound to the arabinogalactan–peptidoglycan layer and an outer leaflet composed of noncovalently associated lipids, such as phthiocerol dimycocerosate, sulfolipids, and trehalose 6,6'-dimycolate. These surface-exposed lipids are immunomodulatory and play a role in host–pathogen interactions.

Recent work demonstrated that the mycobacterial membrane protein large (MmpL) proteins are cell wall lipid transporters. The MmpL transporters are crucial contributors to mycobacterial physiology and pathogenesis. MmpL3 is essential; MmpL4, MmpL5, MmpL7, MmpL8, MmpL10, and MmpL11 are required for full Mtb virulence. MmpL3 transports the trehalose dimycolate (TDM) precursor trehalose monomycolate to the mycobacterial surface. MmpL3 is therefore essential since TDM biosynthesis and incorporation into the mycobacterial cell wall is required for mycobacterial replication and viability.

Based on the genomic sequence of H37Rv, Mtb harbors 14 different MmpL proteins, belonging to the resistance-nodulation-cell division (RND) superfamily of transporters. Similar to the RND efflux pumps of Gram-negative bacteria, several of these MmpL transporters appear to work in conjunction with smaller accessory proteins called mycobacterial membrane protein...
small (MmpS)\textsuperscript{9,17,18}. However, unlike other RND family proteins, the MmpL proteins are not believed to export antibiotics.\textsuperscript{9} Instead, there is strong evidence that these MmpL transporters and their MmpS accessory proteins are responsible for shuttling fatty acid and lipid components of the cell wall, such as trehalose monomycolate, sulfolipids, phthiocerol dimycocerosate, diacyltrehalose, monomeromycolyl diacylglycerol, and mycolate wax ester\textsuperscript{3,9,11,12,19,20,21,22}.

The regulation of MmpL protein expression and the role of MmpLs in cell wall remodeling in different environmental conditions has not been explored. Thus, we capitalized on data made available by the TB Systems Biology Consortium to begin an in-depth analysis of how \textit{mmpL} and \textit{mmpS} genes are regulated. Currently, chromatin immunoprecipitation sequencing (ChIP-Seq) data for 82 of the 180+ Mtb transcription factors is available on the TBDatabase (TBDB)\textsuperscript{24,25,26,27}. We recently demonstrated that the MarR-family regulator Rv0678 regulates the \textit{mmpS2-mmpL2}, \textit{mmpS4-mmpL4}, and \textit{mmpS5-mmpL5} genes. We also identified that the crystal structure of Rv0678 is bound with a fatty acid glycerol ester 2-palmitoylglycerol (C\textsubscript{21}H\textsubscript{42}O\textsubscript{4}), suggesting that fatty acids may be the natural ligands of this regulator\textsuperscript{28}. This structure has allowed us to elucidate the induction mechanism, where the induced conformational change leading to substrate-mediated derepression is primarily caused by a rigid body rotational motion of the entire DNA-binding domain of the regulator toward the dimerization domain\textsuperscript{28}.

In this article, we report crystal structures of two conformational forms of the TetR-family transcriptional regulator Rv0302, which has predicted regulatory interactions within the \textit{mmpL3} and \textit{mmpL11} loci. Binding of this transcriptional regulator to the promoter and intragenic regions of \textit{mmpL} genes is summarized in Figure 1. Typically, the TetR-family
regulators are all helical dimeric proteins, consisting of a smaller N-terminal DNA-binding domain and a larger C-terminal regulatory domain. The N-terminal domains are quite conserved in protein sequences and form a helix-turn-helix (HTH) motif for DNA binding. However, the C-terminal sequences are poorly conserved, forming ligand-specific binding domains for inducing molecules. Our crystal structures of Rv0302 suggest that ligand binding at the C-terminal regulatory domain triggers a rotational motion of the regulator. This motion results in inducing the expression of the MmpL transporters by releasing the Rv0302 regulator from cognate DNAs. Using fluorescence polarization and electrophoretic mobility shift assay (EMSA), we demonstrate that Rv0302 can bind the promoter regions of these mmpL genes within a nanomolar range.

**Figure 1.** Schematic depiction of Rv0302 binding sites in the mmpL genes of interest. ChIPSeq data were obtained from TBDB (www.tbdb.org). In these experiments, FLAG-tagged (DYKDDDDK) transcription factors were episomally expressed in Mtb under the control of an anhydrotetracycline-inducible promoter (Galagan et al.). The red circles corresponding to the Rv0302 transcription factor are placed at the putative binding sites.
Results and discussion

Overall structure of Rv0302

*M. tuberculosis* Rv0302 is a 210 amino acid (aa) protein that belong to the TetR family of transcriptional regulators. Two distinct conformations of Rv0302 with space groups *P*6122 (form I) and *P*212121 (form II) were captured in two different forms of crystals. The form I structure was determined to a resolution of 2.04 Å using single isomorphous replacement. The form II conformation was resolved to a resolution of 2.65 Å using molecular replacement with anomalous scattering using the form I structure as a search model (Table 1 and Fig. 2). By applying the crystallographic symmetry operators, a dimeric arrangement of the structure was found. In the form II structure, two monomers were found in the asymmetric unit arranged as a dimer. Overall, the architecture of these two Rv0302 structures are in good agreement with those of the TetR-family regulators, including *TetR*,31,32 *QacR*,33,34 *CprB*,35 *EthR*,36,37 *CmeR*,38,39 *AcrR*,40 *SmeT*,41 *Rv3066*,42 and *Rv1219c*.43

Each subunit of Rv0302 is composed of nine helices (α1–α9 and α1’–α9’, respectively) that are organized to form two functional motifs: the N-terminal DNA-binding and C-terminal ligand-binding domains (Fig. 3). The helices of Rv0302 are designated numerically from the N-terminus as α1 (residues 14–29), α2 (residues 37–44), α3 (residues 48–55), α4 (residues 58–76), α5 (residues 88–104), α6 (residues 107–117), α7 (residues 123–147), α8 (residues 154–181), and α9 (residues 188–208). In this arrangement, the smaller N-terminal DNA-binding domain includes helices α1 through α3 and the N-terminal end of α4 (residues 58–65), with α2 and α3 forming a typical HTH motif. However, the larger C-terminal ligand-binding domain comprises
the C-terminal end of helices α4 (residues 66–76) through α9. Helices α6, α8, and α9 are involved in the dimerization of the regulator, and helix α9 contacts both α8 and α9′ to secure the dimerization interface.

In both form I and form II structures, the 21 aa helix α9 folds uniquely along the top of the dimer, forming the ceiling for the ligand-binding domain. To make space for this fold, helix α8 is oriented at an approximate 25° angle away from the dimerization interface. Comparing the dimeric structures of forms I and II suggests that these two structures depict two different transient states of the regulator. Superimposition of the forms I and II dimeric structures of Rv0302 results in an overall rms deviation of 3.0 Å. The difference between the two conformations is a 9° rotational motion of the right subunit with respect to the left protomer (Fig. 4). Based on this structural information, it is likely that ligand binding triggers a rotational motion within the dimer of the regulator. Presumably, this movement prohibits the binding of the dimeric regulator to its cognate DNA, which in turn releases the regulator from the promoter region and allows for the expression of the corresponding MmpL transporters. If this is the case, then the form I conformation should correspond to the ligand induced form of the Rv0302 regulator.

The C-terminal regulatory domain of each subunit of the Rv0302 structures forms a large cavity, presumably creating a ligand-binding pocket of the regulator. This cavity, which is predominately formed by helices α4–α9, orients more or less vertically and in parallel with the twofold symmetry axis of the dimer. At least 24 amino acids line the wall of this cavity. Among them, eight are aromatic residues (F73, F74, W80, F112, Y113, F140, Y176, and Y192), 11 are hydrophobic residues (I77, L95, L98, L109, V132, A136, L137, L147, V169, and L199), and
five are polar or charged residues (S91, Q94, Q102, S143, and D173). Based on these observations, ligand binding in Rv0302 is predominately governed by hydrophobic interactions.

An extra electron density was found within the ligand-binding pocket of each subunit of the form I structure of Rv0302 [Fig. 5(A)]. The shape of this extra density is compatible with an isopropanol molecule. This was not surprising because we used solutions containing isopropanol for crystallization. Each bound isopropanol molecule is completely buried in the Rv0302 binding pocket. Four aromatic and hydrophobic residues (L109, F112, Y113, and V132) make hydrophobic contacts with the bound isopropanol [Fig. 5(B)]. In addition, one of the side chain oxygens of D173 forms a hydrogen bond with the hydroxyl oxygen of the bound isopropanol to secure the binding.
Figure 2. Electron density maps of the *M. tuberculosis* Rv0302 regulator. (A) Stereo view of the experimental electron density map of the form I structure at a resolution of 2.04 Å. The electron density map is contoured at 1.0σ. The CA traces of the Rv0302 molecule in the asymmetric unit are colored green. Anomalous signals of the two Ta₆ cluster sites (contoured at 3σ) found in the asymmetric unit are colored red. (B) Representative section of electron density in the vicinity of helices α4 and α7. The solvent-flattened electron density (40–2.04 Å) is contoured at 1.0σ and superimposed with the final refined model (green, carbon; red, oxygen; blue nitrogen; yellow, sulfur). (C) Stereo view of the electron density map of the form II structure at a resolution of 2.65 Å. The electron density map is contoured at 1.0σ. The CA traces of the Rv0302 molecules in the asymmetric unit are in orange and yellow.
Table 1. Data collection, phasing and structural refinement statistics of Rv0302.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Rv0302 form I</th>
<th>Ta$<em>6$Br$</em>{12}^{2+}$ derivative</th>
<th>Rv0302 form II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9792</td>
<td>1.2550</td>
<td>0.9792</td>
</tr>
<tr>
<td>Space group</td>
<td>P6,22</td>
<td>P6,22</td>
<td>P2,2;2</td>
</tr>
<tr>
<td>Cell constants (Å)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>116.6</td>
<td>117.7</td>
<td>46.1</td>
</tr>
<tr>
<td>b</td>
<td>116.6</td>
<td>117.7</td>
<td>77.1</td>
</tr>
<tr>
<td>c</td>
<td>94.1</td>
<td>93.7</td>
<td>118.6</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90,90,120</td>
<td>90,90,120</td>
<td>90,90,90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.04 (2.11-2.04)</td>
<td>4.10 (4.25-4.10)</td>
<td>2.65 (2.74-2.65)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (100)</td>
<td>91.4 (92.1)</td>
<td>97.3 (97.6)</td>
</tr>
<tr>
<td>Total reflections</td>
<td>216,053</td>
<td>303,871</td>
<td>847,565</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>24,642</td>
<td>3,377</td>
<td>12,880</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.6 (7.6)</td>
<td>3.1 (3.0)</td>
<td>4.5 (4.5)</td>
</tr>
<tr>
<td>R$_{merge}$ (%)</td>
<td>4.3 (38.3)</td>
<td>13.0 (39.4)</td>
<td>10.5 (33.4)</td>
</tr>
<tr>
<td>⟨I/σ(I)⟩</td>
<td>52.6 (4.9)</td>
<td>8.3 (3.7)</td>
<td>12.2 (3.7)</td>
</tr>
<tr>
<td><strong>Phasing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of sites</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution used (Å)</td>
<td>4.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phasing power (acentric/centric)</td>
<td>0.81/0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R$_{Culitis}$ (acentric/centric)</td>
<td>0.88/0.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure of merit (acentric/centric)</td>
<td>0.34/0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R$_{work}$ (%)</td>
<td>19.2</td>
<td></td>
<td>23.3</td>
</tr>
<tr>
<td>R$_{free}$ (%)</td>
<td>21.5</td>
<td></td>
<td>26.8</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall(Å$^2$)</td>
<td>55.2</td>
<td></td>
<td>49.5</td>
</tr>
<tr>
<td>Ligands(Å$^2$)</td>
<td>51.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rms deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond (Å)</td>
<td>1.336</td>
<td></td>
<td>0.576</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>0.013</td>
<td></td>
<td>0.002</td>
</tr>
<tr>
<td><strong>Ramachandran analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favored (%)</td>
<td>97.8</td>
<td></td>
<td>98.0</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>2.2</td>
<td></td>
<td>2.0</td>
</tr>
<tr>
<td>Generously allowed (%)</td>
<td>0.0</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>Disallowed (%)</td>
<td>0.0</td>
<td></td>
<td>0.0</td>
</tr>
</tbody>
</table>
Figure 3. Structure of the *M. tuberculosis* Rv0302 regulator. (A) Ribbon diagram of a protomer of the form I structure of Rv0302. The molecule is colored using a rainbow gradient from the N-terminus (blue) to the C-terminus (red). (B) Ribbon diagram of the form I structure of the Rv0302 dimer. Each subunit of Rv0302 is labeled with a different color (red and yellow). The Figure was prepared using PyMOL (http://www.pymol.sourceforge.net).
**Figure 4.** Structural comparison of forms I and II of the Rv0302 regulator. (A) This is a superimposition of the dimeric structures of forms I and II (green, form I; orange, form II). For clarity, only the right subunit of helices α1–α9 the form II structure (orange) are labeled. The arrow indicates a change in orientation of the right subunit of form I when compared with the structure of form II. (B) Side view of the superimposition of the dimeric structures of forms I and II (green, form I; orange, form II). For clarity, only the right subunit of helices α1–α9 of form II structure (orange) are labeled. This view depicts a 9° rigid body rotation of the right subunit (α1–α9) of form I with respect to that of form II.
Figure 5. Electron density map and the isopropanol binding site. (A) Stereo view of the \( F_o - F_c \) electron density map of the bound isopropanol in Rv0302. The bound isopropanol is shown as a stick model (green, carbon; blue, nitrogen). The \( F_o - F_c \) map is contoured at 3.0\( \sigma \) (blue mesh). (B) The isopropanol-binding site of Rv0302. Residues involved in isopropanol binding are in orange sticks. The bound isopropanol is shown as green sticks. Dotted lines depict the hydrogen bonds.
Regulator–ligand interactions

Isothermal titration calorimetry

Recently, we have found that the *M. tuberculosis* Rv3249c regulator is able to recognize palmitic acid, a saturated fatty acid containing 16 carbons with the molecular formula C₁₆H₃₂O₂.⁴⁴ Since both Rv0302 and Rv3249c were predicted to regulate the expression of the *mmpS1/L1, mmpL3, mmpL7,* and *mmpL11* genes, it is possible that these two regulators share a similar set of ligands. Therefore, we decided to test if Rv0302 is capable of binding palmitic acid. Isothermal titration calorimetry (ITC) was then used to study the interaction between Rv0302 and this ligand. This titration depicts a typical hyperbolic binding curve, with thermodynamic parameters of −543.6 ± 43.5 cal mol⁻¹ (ΔH) and 21.0 cal mol⁻¹ deg⁻¹ (ΔS). The equilibrium dissociation constant (\(K_D\)) for the binding of Rv0302 to palmitic acid was measured to be 10.5 ± 2.3 µM (Fig.6). Indeed, our data indicate that Rv0302 is capable of recognizing this fatty acid.

**Figure 6.** Representative ITC for the binding of palmitic acid to Rv0302. (A) Each peak corresponds to the injection of 10 µL of 500 µM palmitic acid in buffer containing 10 mM Na-phosphate (pH 7.2), 100 mM NaCl, and 0.001% DDM into the reaction containing 14.5 µM Rv0302 dimer in the same buffer. (B) Cumulative heat of reaction is displayed as a function of the injection number. The solid line is the least-square fit to the experimental data, giving a \(K_D\) of 10.5 ± 2.3 µM.
Regulator–DNA interactions

Fluorescence polarization assay

Fluorescence polarization was used to quantify the strength of regulator–DNA interactions. To identify regulatory targets of these proteins, we used ChIP-Seq data from Galagan et al. and the TBDB.\textsuperscript{11,22,36} Regions of the \textit{M. tuberculosis} H37Rv genome that were identified by these experiments to interact with Rv0302 were first examined to find potential binding sequences for each individual protein. Typically, TetR-family proteins interact with DNAs \textit{via} symmetric palindromic stretches called inverted repeats (IRs), approximately 15–30 nucleotides long. Thus, the search was narrowed to include sequences that contain these patterns. For Rv0302, we were able to identify a putative IR sequence located in one or more of the \textit{M. tuberculosis} H37Rv genes encoding MmpL transporter proteins (Table 2). These DNA sequences are in good agreement with both the consensus binding sequences and protein–DNA interactions determined by others.\textsuperscript{37} In short, we have compiled additional evidence that the Rv0302 protein may act as a regulator for \textit{mmpL6} and \textit{mmpL11}.

Fluorescence polarization assays were then performed using the purified Rv0302 regulator protein and duplex DNAs. We quantified the interaction of this regulator with the DNA sequences listed in Table 2. These DNA sequences are located within the operons of \textit{mmpL11} and \textit{mmpL6}. In addition, we were able to locate an IR sequence within the promoter region of \textit{rv0302}. The experiments suggest that Rv0302 binds these DNA sequences with \(K_D\) values in the nanomolar range (Table 2 and Fig. 7). Interestingly, the fluorescence polarization data indicate that Rv0302 binds these DNA with a stoichiometry of one Rv0302 dimer per DNA duplex.
**Figure 7.** Representative fluorescence polarization of Rv0302. (A) The binding isotherm of Rv0302 with the 28-bp DNA located within the promoter region of *mmpL11*, showing a $K_D$ of $40.4 \pm 4.9$ nM. (B) The binding isotherm of Rv0302 with the 29-bp DNA located within the promoter region of *mmpL6*, showing a $K_D$ of $58.5 \pm 2.9$ nM. (C) The binding isotherm of Rv0302 with the 24-bp DNA located within the promoter region of *rv0302*, showing a $K_D$ of $80.9 \pm 9.2$ nM. Fluorescence polarization is defined by the equation, $FP = (V - H)/(V + H)$, where $FP$ equals polarization, $V$ equals the vertical component of the emitted light, and $H$ equals the horizontal component of the emitted light of a fluorophore when excited by vertical plane polarized light. $FP$ is a dimensionless entity and is not dependent on the intensity of the emitted light or on the concentration of the fluorophore. $mP$ is related to $FP$, where $1 \text{ mP}$ equals one thousandth of a $FP$. 
Table 2. Affinity for DNA Binding by Rv0302

<table>
<thead>
<tr>
<th>DNA sequence</th>
<th>Location</th>
<th>$K_D$ (nM)</th>
<th>Hill coefficient(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5′-GCCTGCGGCCGCTCGTCGCGGTGCCTGT-3$^a$</td>
<td>mmpL11</td>
<td>12.6 ± 1.7</td>
<td>1.2. ± 0.2</td>
</tr>
<tr>
<td>5′-F-ACAGGCACCCGCACGACGCACGACGCACGGCAGGC-3$^b$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5′-TGCCCGGGCGCGCACCACGCCCGGTACCT-3′</td>
<td>mmpL6</td>
<td>19.1 ± 5.2</td>
<td>1.1 ± 0.1</td>
</tr>
<tr>
<td>5′-F-AGGTACGGGCGTGGTCGCGCCCGGGCA-3′</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5′-TTTCTTGGGCGGAACGCCACTGG-3′</td>
<td>rv0302</td>
<td>13.7 ± 2.8</td>
<td>1.1 ± 0.1</td>
</tr>
<tr>
<td>5′-F-CCAGTGCCGGTCCCGGCAAGAAA-3′</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ The IR sequence was underlined.

$^b$ F denotes the fluorescein which was covalently attached to the 5′ end of the oligodeoxynucleotide (reversed) by a hexamethylene linker.

**Gel filtration**

To confirm the dimeric oligomerization of Rv0302 depicted by the crystal structures, we performed a gel filtration experiment using the purified Rv0302 protein. The result suggests an average molecular weight of 46.3 ± 2.5 kDa. This value is in good agreement with the theoretical value of 48.4 kDa for two Rv0302 molecules, indicating that the Rv0302 regulator is dimeric in solution.
Our fluorescence polarization experiments suggest that the Rv0302 protein uses a simple binding stoichiometry with a 1:1 dimeric Rv0302-to-duplex DNA molar ratio to interact with the DNA sequences located within the operons of *mmpL11*, *mmpL6*, and *rv0302*, respectively. To confirm this protein–DNA binding stoichiometry, gel filtration experiment was carried out using the purified Rv0302 protein pre-incubated with the purified, complementary, annealed oligonucleotides that contain the sequences within the operons of *mmpL11*, *mmpL6*, and *rv0302*, individually. The results suggest average molecular weights of 65.5 ± 3.1, 70.7 ± 4.6 and 58.9 ± 4.5 kDa for these Rv0302–DNA complexes (Fig. 8). These values are in good agreement with the corresponding theoretical values of 65.8, 66.4 and 63.3 kDa for two Rv3066 molecules bound to the respective DNAs, confirming the stoichiometry of these Rv0302-DNA bindings is 1:1 dimeric Rv0302-to-DNA molar ratio.

**Figure 8.** Representative gel filtration experiment. The experiment demonstrated that Rv0302 is dimeric in solution. In addition, one Rv0302 dimer is found to bind one duplex DNA. The y-axis values were defined as: 

\[ K_{av} = (V_e - V_0)/(V_T - V_0) \]

where \( V_T \), \( V_e \), and \( V_0 \) are the total column volume, elution volume, and void volume of the column, respectively.

Standards used were: A, cytochrome C (\( M_r 12,400 \)); B, carbonic anhydrase (\( M_r 29,000 \)); C, albumin bovine serum (\( M_r 66,000 \)); D, alcohol dehydrogenase (\( M_r 150,000 \)); and E, β-amylase (\( M_r 200,000 \)). The void volume was measured using blue dextran (\( M_r 2,000,000 \)). Samples for the measurements were: red triangle, Rv0302; green diamond, Rv0302-*mmpL11*; orange inverted triangle, Rv0302-*mmpL6*; blue square, Rv0302-*rv0302*. 
**Electrophoretic mobility shift assay**

ChIP-Seq data suggests that Rv0302 regulates expression of rv0302, mmpS2/L2, and mmpL11 [Fig. 9(A)]. We performed EMSAs using purified Rv0302 to demonstrate direct transcriptional regulation by Rv0302. We observed a concentration-dependent shift of the rv0302, mmpL2, and mmpL11 probes [Fig. 9(B–D)]. As a negative control, we used a DNA probe that has no predicted binding sites for Rv0302. EMSAs were also performed in the presence of nonlabeled “cold” probe. Release of Dig-labeled probe was observed consistent with specific binding of Rv0302 to the mmpL11 probe [Fig. 9(C)]. As fluorescence polarization study suggested that Rv0302 binds palmitic acid, we performed an EMSA in the presence and absence of palmitic acid to demonstrate this experimentally. Indeed, addition of palmitate reduced binding of Rv0302 to the rv0302 probe [Fig. 8(D)].

**Conclusion**

In this article, we describe the crystal structures of the Rv0302 transcriptional regulator, which contribute to the regulatory network that controls the expression levels of the MmpL transporters. Specifically, the Rv0302 protein should regulate the genes mmpL1, mmpL2, mmpL3, mmpL6, mmpL7, mmpL9, and mmpL11. MmpL transporters significantly contribute to the export of important lipid components of the mycobacterial cell wall and are necessary for the virulence of this pathogen. Our experimental data demonstrate a direct binding of this transcriptional regulator to intragenic and promoter DNAs, providing evidence for the transcriptional control of mmpL gene expression. Multiple transcriptional factor binding sites exist within the promoter
Figure 9. Rv0302 binds to promoter regions of mmpL11 and rv0302 and intragenic region of mmpL2. (A) A schematic depicting the DNA probes used in EMSAs. (B) EMSAs were performed with 6 nM Dig-labeled probe and the indicated micromolar concentrations of protein. (C) To demonstrate specificity, the MmpL11 EMSA was performed in the presence of nonlabeled (“cold”) probe. Reactions were performed with 6 nM Dig-labeled probe, the indicated micromolar concentrations of protein, and 360 nM cold probe. (D) Ligand-bound Rv0302 does not bind target probes. EMSA was performed using 12 nM Dig-labeled probe and 0.1 µM Rv0302 in the absence or presence of the indicated concentration of palmitic acid. An arrow denotes the shifted probes and the asterisk notes the accumulation of free Dig-labeled probe.
and intragenic region of the *mmpL* genes, and each transcriptional regulator recognizes several *mmpL* regulatory regions. For example, both Rv0302 and Rv3249c are able to bind to different regulatory sequences within the *mmpL11* gene. Indeed, our experimental data indicate that these two regulators may also share the same palmitate ligand. These findings suggest that *mmpL* gene expression may rely on a complex interplay of multiple transcription regulators. Further experiment is needed to confirm this observation.

The TetR family of regulators uses a few distinct mechanisms for modulating transcriptional regulation. However, the net consequence of binding of inducing ligands to these regulators is essentially the same. Ligand binding at the C-terminal regulatory domain triggers a long distance conformational change at the N-terminal DNA binding domain, resulting in the release of the regulator from its operator DNA. The TetR-family regulators use the N-terminal recognition helix α3 to bind the major groove of B-DNA. The two crystal structures of Rv0302 have allowed us to understand how this regulator controls gene expression. It appears that ligand binding may trigger a rotational motion of one subunit of Rv0302 in relation to the next subunit within the dimer. This rotational motion presumably makes the relative orientation of the two N-terminal DNA-binding domains of the regulator incompatible with the two consecutive major grooves of the operator B-DNA. Similar rigid body rotational movement has been found in the SimR and Rv3066 regulators, where rigid body rotation within subunits of the dimer in relation to each other contributes to the induction process.\(^{34,38}\) The net result is that this dimeric Rv0302 regulator is released from the promoter, which in turn initiates the expression of the *mmpL* genes.
Materials and methods

Cloning of \textit{rv0302}

The \textit{rv0302} ORF from genomic DNA of \textit{M. tuberculosis} strain H37Rv was amplified by polymerase chain reaction (PCR) using the primers 5′-CTTTAAGAA GGAGATATACCATGGTGCGGCTCCGCAAGAAAAAC-3′ and 5′-GATCCTCAGTGATGATGGTGATGATGTCTCCTCAGGAGGACGGGAATC-3′. The corresponding PCR product was digested with NcoI and BamHI, extracted from the agarose gel and inserted into pET15b as described by the manufacture (Merck, Kenilworth, NJ). The recombinant plasmid, pET15b\textsubscript{Ω}rv1219c, was transformed into DH10b cells, and the transformants were selected on Luria Broth (LB) agar plates containing 100 µg/mL ampicillin. The presence of the correct \textit{rv1219c} sequence in the plasmid construct was verified by DNA sequencing.

Expression and purification of Rv0302

Briefly, the full-length protein Rv0302 containing a 6×His tag at the C-terminus was overproduced in \textit{E. coli }BL21(DE3) cells possessing pET15b\textsubscript{Ω}rv0302. Cells were grown in 6 L of LB medium with 100 µg/mL ampicillin at 37°C. When the OD\textsubscript{600} reached 0.5, the culture was treated with 0.2 mM IPTG to induce Rv0302 expression, and cells were harvested within 3 h. The collected bacterial cells were suspended in 100 mL ice-cold buffer containing 20 mM Na-HEPES (pH 7.2) and 250 mM NaCl. The cells were then lysed with a French pressure cell. Cell debris was removed by centrifugation for 45 min at 4°C and 20,000 rev/min. The crude lysate
was filtered through a 0.2 µm membrane and loaded onto a 5 mL Hi-Trap Ni$^{2+}$-chelating column (GE Healthcare Biosciences, Pittsburgh, PA) pre-equilibrated with 20 mM Na-HEPES (pH 7.2) and 250 mM NaCl. To remove unbound proteins and impurities, the column was first washed with eight column volumes of buffer containing 20 mM imidazole, 250 mM NaCl, and 20 mM Na-HEPES (pH 7.2), and then five column volumes of buffer containing 50 mM imidazole, 250 mM NaCl, and 20 mM Na-HEPES (pH 7.2). The Rv0302 protein was then eluted with three column volumes of buffer containing 300 mM imidazole, 250 mM NaCl, and 20 mM Na-HEPES (pH 7.2). The purity of the protein was judged using 12.5% SDS-PAGE stained with Coomassie Brilliant Blue. The purified protein was extensively dialyzed against buffer containing 100 mM imidazole, 150 mM NaCl, and 20 mM Na-HEPES (pH 7.5) and concentrated to 20 mg/mL.

**Crystallization of Rv0302**

All crystals of the 6×His Rv0302 regulator were obtained using hanging drop vapor diffusion. The form I Rv0302 crystals were grown at room temperature in 24-well plates with the following procedures. A 1 µL protein solution containing 20 mg/mL Rv0302 protein in 20 mM Na-HEPES (pH 7.5), 150 mM NaCl, and 100 mM imidazole was mixed with 1 µL of reservoir solution containing 30% polyethylene glycol (PEG) 400, 0.1M Na-HEPES (pH 7.5), 0.2M NaCl, and 10% isopropanol, with an addition of 2% benzamidine·HCl. The resultant mixture was equilibrated against 500 µL of the reservoir solution. Crystals appeared overnight and grew to a full size in the drops within 2 weeks. Typically, the dimensions of the crystals were 0.1 mm × 0.3 mm × 0.3 mm. Further cryoprotection was not necessary.
Crystals of the tantalum derivative were prepared by incubating the form I crystals overnight in a solution containing 30% PEG 400, 0.1M Na-HEPES (pH 7.5), 0.2M NaCl, 10% isopropanol, 2% benzamidine·HCl, and 1 mM (Ta₆Br₁₂)²⁺·2Br⁻ (Jena Bioscience, Jena, Germany).

The form II Rv0302 crystals were grown at room temperature in 24-well plates by mixing 1 µL of protein solution with 1 µL of reservoir solution containing 18% PEG 2000, 0.1M K-MES (pH 6.5), 0.2M NaCl, and 10% isopropanol. The resultant mixture was equilibrated against 500 µL of the reservoir solution. Crystals appeared overnight and grew to a full size in the drops within 1 week. Typically, the crystals were plate-like with dimensions 0.2 mm × 0.2 mm × 0.05 mm. Cryoprotection was achieved by raising the PEG 2000 concentration stepwise to 25%.

**Data collection, structural determination, and refinement**

All diffraction data were collected at 100K at beamline 24ID-C located at the Advanced Photon Source, using a Pilatus 6M detector (Dectris, Switzerland). Diffraction data were processed using DENZO and scaled using SCALEPACK. The form I crystals of Rv0302 belong to the space group P₆₁₂₂ (Table 1). Based on the molecular weight of Rv0302 (23.8 kDa), the asymmetric unit is expected to contain one regulator molecule with a solvent content of 68.7%. Two tantalum cluster sites were identified using SHELXC and SHELXD as implemented in the HKL2MAP package. Single isomorphous replacement with anomalous scattering was used to obtain experimental phases using the program MLPHARE. The resulting phases were then subjected to density modification using the program PARROT.
phases were of excellent quality and allowed for tracing of most of the molecule in PHENIX AutoBuild,\textsuperscript{51} which led to an initial model with over 90% amino acid residues containing side chains. The remaining part of the model was manually constructed using the program Coot.\textsuperscript{52} Then, the model was refined using PHENIX\textsuperscript{51} leaving 5% of reflections in Free-R set. Iterations of refinement using PHENIX\textsuperscript{51} and CNS\textsuperscript{53} and model building in Coot\textsuperscript{52} lead to the current model with excellent geometrical characteristics (Table 1).

The form II crystal took the space group P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1}. This structure was determined by molecular replacement, using the form I structure as the search model. The program PHASER\textsuperscript{54} was used to carry out the MR calculations. Structural refinements were then performed using PHENIX\textsuperscript{51} and CNS\textsuperscript{53} (Table 1).

**Isothermal titration calorimetry**

We used ITC to determine the binding affinity of palmitic acid to the purified Rv0302 regulator. Measurements were performed on a VP-Microcalorimeter (MicroCal, Northampton, MA) at 25°C. Before titration, the protein was thoroughly dialyzed against buffer containing 10 mM Na-phosphate pH 7.2, 100 mM NaCl, and 0.001% n-dodecyl-β-D-maltoside (DDM). The protein concentration was determined using the Bradford assay and then adjusted to a dimeric concentration of 14.5 µM. The ligand solution containing 500 µM palmitic acid, 10 mM Na-phosphate pH 7.2, 100 mM NaCl, and 0.001% DDM was used as the titrant. Binding experiments were carried out with the protein solution (1.4 mL) in the cell and the ligand solution as the injectant. Thirty injections of 10 µL each of the ligand solution were used for data collection.
Injections occurred at intervals of 240 s, and the duration time of each injection was 20 s. Heat transfer (µcal/s) was measured as a function of elapsed time (s). The mean enthalpies measured from injection of the ligand in the buffer were subtracted from raw titration data before data analysis with ORIGIN software (MicroCal, Westborough, MA). Titration curves were fitted by a nonlinear least squares method to a function for the binding of a ligand to a macromolecule. Nonlinear regression fitting to the binding isotherm provided us with the equilibrium binding constant \((K_A = 1/K_D)\) and enthalpy of binding \((\Delta H)\). Based on the values of \(K_A\), the change in free energy \((\Delta G)\) and entropy \((\Delta S)\) were calculated with the equation: \(\Delta G = -RT \ln K_A = \Delta H - T\Delta S\), where \(T\) is 273 K and \(R\) is 1.9872 cal/K per mol.

Calorimetry trials were also carried out in the absence of Rv0302 in the same experimental conditions. No change in heat was observed in the injections throughout the experiment.

**Fluorescence polarization assay**

Fluorescence polarization assays were used to determine the affinity for DNA binding by Rv0302. All oligodeoxynucleotides and fluorescein-labeled oligodeoxynucleotides were purchased from Integrated DNA Technologies (Coralville, IA). The sequences of these oligodeoxynucleotides are summarized in Table 2. The fluoresceinated ds-DNAs were prepared by annealing the oligodeoxynucleotide and its corresponding fluorescein-labeled oligodeoxynucleotide together. Fluorescence polarization experiment was done using a DNA binding solution containing 10 mM Na-phosphate (pH 7.2), 100 mM NaCl, 2.5 nM fluoresceinated DNA, and 1 µg of poly(dI-dC) as nonspecific DNA. The protein solution
containing 500 nM dimeric Rv0302 and 2.5 nM fluoresceinated DNA was titrated into the DNA binding solution until the millipolarization (mP) become unchanged. All measurements were performed at 25°C using a PerkinElmer LS55 spectrofluorometer equipped with a Hamamatsu R928 photomultiplier. The excitation wavelength was 490 nm, and the fluorescence polarization signal (in $\Delta P$) was measured at 525 nm. Each titration point recorded was an average of 15 measurements.

Data were analyzed using the equation, $P = \{(P_{\text{bound}} - P_{\text{free}})[\text{protein}]/(K_D + [\text{protein}])\} + P_{\text{free}},$

where $P$ is the polarization measured at a given total protein concentration, $P_{\text{free}}$ is the initial polarization of free fluorescein-labeled DNA, $P_{\text{bound}}$ is the maximum polarization of specifically bound DNA, and [protein] is the protein concentration. The titration experiments were repeated three times to obtain the average $K_D$ value. Curve fitting was accomplished using the program ORIGIN (OriginLab, Northampton, MA).

**Gel filtration**

A protein liquid chromatography Superdex 200 16/60 column (GE Healthcare Biosciences, Pittsburgh, PA) with a mobile phase containing 20 mM Na-phosphate (pH 7.2) and 100 mM NaCl was used in the gel filtration experiments. Blue dextran (Sigma-Aldrich, St. Louis, MO) was used to determine the column void volume, and proteins for use as gel filtration molecular weight standards were cytochrome C ($M_r$ 12,400), carbonic anhydrase ($M_r$ 29,000), albumin bovine serum ($M_r$ 66,000), alcohol dehydrogenase ($M_r$ 150,000), and β-Amylase ($M_r$ 200,000). All these standards were purchased from Sigma-Aldrich (St. Louis, MO). The molecular weights of the experimental samples were determined following the protocols supplied by the manufacturers.
Electrophoretic mobility shift assay

Probes were amplified from the H37Rv genome using the primers listed in Table 3. All probes were labeled with Digoxigenin using the Roche DIG Gel Shift kit. For EMSA analysis, 12 nM Dig-labeled probe and the indicated micromolar concentrations of protein were incubated for 45 min at room temperature in the Roche binding buffer modified by the addition of 0.25 mg/mL herring sperm DNA, and 0.75 mg/mL poly(d[I-C]). For ligand competition assays, the stock solution of palmitic acid was made in dimethyl sulfoxide (DMSO) and a solvent control reaction included at the highest concentration of DMSO. All reactions were resolved on a 6% native polyacrylamide gel in TBE buffer, transferred to nylon membrane and Dig-labeled DNA–protein complexes detected following the manufacturer's recommendations. Chemiluminescent signals were acquired using an ImageQuant LAS 4000 (GE).

Table 3. Primers used to Amplify EMSA Probes

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0302F.0302</td>
<td>5′-CGGTACTGCACGTCGACAA-3′</td>
</tr>
<tr>
<td>Rv0302R.0302</td>
<td>5′-GTTCGGTCGCGTCGAGAATC-3′</td>
</tr>
<tr>
<td>mmpL11F.0302</td>
<td>5′-CCGAGATGGCAGGATGACGG-3′</td>
</tr>
<tr>
<td>mmpL11R.0302</td>
<td>5′-TCGCTGATGGTTCGGCAAC-3′</td>
</tr>
<tr>
<td>mmpL2F.0302</td>
<td>5′-TTATCTGGCATGCGACGTT-3′</td>
</tr>
<tr>
<td>mmpL2R.0302</td>
<td>5′-TTGCCGTCGGAGACAAA-3′</td>
</tr>
</tbody>
</table>

Protein data bank accession code

Coordinates and structural factors for the structures of Rv0302 have been deposited in the RCSB Protein Data Bank with accession codes 5D18 and 5D19 for the form I and form II conformations, respectively.
Acknowledgement

This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Pilatus 6M detector on 24-ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

References


   (2012) Inhibition of mycolic acid transport across the *Mycobacterium* 

   major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science* **276**:1420– 
   1422.

   KasA, a key enzyme of mycolic acid biosynthesis, leads to mycobacterial cell lysis. *J 

15. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, 

   (1999) The RND permease superfamily: an ancient, ubiquitous and diverse family that 
   125.


   Riccardi G (2009) Azole resistance in *Mycobacterium tuberculosis* is mediated by the 
   MmpS5–MmpL5 efflux system. *Tuberculosis* **89**:84–90.


CHAPTER III

CRYSTAL STRUCTURE OF THE CHLAMYDOMONAS REINHARDTII LCI1 CHANNEL

Manuscript submitted

Tsung-Han Chou\textsuperscript{1,†}, Abhijith Radhakrishnan\textsuperscript{2,†}, Jani Reddy Bolla\textsuperscript{3}, Kannan Sankar\textsuperscript{4}, Chih-Chia Su\textsuperscript{1}, Nitin Kumar\textsuperscript{2}, Robert L. Jernigan\textsuperscript{4,5}, Carol V. Robinson\textsuperscript{3}, Martin H. Spalding\textsuperscript{6}, and Edward W. Yu\textsuperscript{1,2,*}

\textsuperscript{1}Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, USA
\textsuperscript{2}Department of Chemistry, Iowa State University, Ames, IA 50011, USA
\textsuperscript{3}Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford, South Park Road, OX1 3QZ, UK.
\textsuperscript{4}Bioinformatics and Computational Biology Interdepartmental Graduate Program, Iowa State University, Ames, IA 50011, USA
\textsuperscript{5}Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011, USA
\textsuperscript{6}Department of Genetics, Developmental and Cell Biology, Iowa State University, Ames, IA 50011, USA

*Correspondence to: Edward W. Yu, Department of Chemistry, Iowa State University, Ames, IA

† T.H.C. and A.R contributed equally to this work.
Introduction

Carbon dioxide (CO$_2$) is an essential substrate for photosynthesis, which provides the basis for the synthesis of organic compounds and nutrients. Without CO$_2$, the life of photosynthetic organisms and even animals would not be possible. However, the central enzyme of photosynthetic CO$_2$ assimilation, rubilose-1,5-bisphosphate carboxylase oxygenase (Rubisco$^{1,2,3}$), which is probably the most abundant enzyme on earth, has a low affinity for CO$_2$ and slow catalytic turnover rate for carboxylation$^{4,5}$. As a result, the efficiency of this photosynthetic process is strongly dependent upon the concentration of CO$_2$ in the vicinity of Rubisco. Given the fact that the diffusion rate of Ci in water is 10$^4$ times slower than that in air, aquatic photosynthetic species appear to endure even more obstacles in terms of scavenging inorganic carbon (Ci) resources. To overcome this hurdle, microalgae and cyanobacteria have developed a CO$_2$-concentrating mechanism (CCM)$^{6,7,8,9}$, which is one of the most effective strategies for CO$_2$ enrichment. These CCMs utilize diverse and active Ci uptake systems capable of accumulating intracellular Ci concentration up to 1,000-fold from low CO$_2$ environments. As microalgae and cyanobacteria contribute significantly to global CO$_2$ sequestration and have shown enormous potential as an alternative future energy source, enhanced knowledge of CCM will offer us guidance to boost biomass production and improve the efficiency of photosynthetic carbon fixation in crop species that lack a CCM.

In comparison with cyanobacterial CCMs, the CCMs of eukaryotic microalgae are more complex mainly due to the involvement of a larger number of intracellular compartments. The green alga Chlamydomonas reinhardtii has served as a model organism
for decades, allowing us to identify key components that are involved in CO₂ and bicarbonate (HCO₃⁻) uptake of the CCM. However, a complete picture of the CCM remains unclear. In most natural environments, CO₂ and HCO₃⁻ are the two major sources of Ci for *C. reinhardtii* assimilation. As the Ci concentration, pH and temperature vary, the availability of both the Ci species also changes. This change may activate and induce the expression of various Ci transport proteins to accommodate intracellular Ci uptake. At least three acclimation states have been identified in *C. reinhardtii* based on physiological responses to various CO₂ concentrations. These concentrations are high CO₂ (0.5-5%), low CO₂ (0.03-0.4%) and very low CO₂ (<0.02%)⁷.

To better understand how the CCM works, it is important to elucidate the structure and function of Ci transport proteins participating in Ci enrichment. In *C. reinhardtii*, Ci uptake occurs at the plasma membrane, chloroplast envelope and inside the chloroplasts. Many of these CO₂-responsive genes encode putative Ci transporters and CO₂ uptake systems. These membrane proteins include LCI1 (low CO₂ inducible gene 1)¹¹ and HLA3 (high light activated 3, also known as MRP1)¹² found in the plasma membrane, as well as LCIA (low CO₂ inducible gene A, also known as NAR1.2)¹³ located at the chloroplast envelope.

Several Ci transporters functionally confirmed in cyanobacteria and algae, including BCT1,¹⁴,¹⁵ BicA¹⁶,¹⁷ and SbtA¹⁷,¹⁸ in cyanobacteria and HLA3¹⁹,²¹ and LCIA²¹,²² in *C. reinhardtii*, belong to well-characterized transporter families (BCT1 and HLA3 to the ATP binding cassette (ABC) transporter family, BicA to the sulphate transporter (SuP) family, and LCIA to the Formate-Nitrite Transporter (FNT) family). Although diversified by their molecular identities, energization modes, substrate affinity and Ci-dependent expression
patterns, these Ci transporters exclusively belong to typical energy-dependent, integral membrane proteins like other nutrient transporters. Therefore, the biochemical basis for their HCO$_3^-$ uptake is similar to the process catalyzed by carrier-mediated transport, which is well characterized in many prokaryotic and eukaryotic organisms. However, LCI1 appears to represent a completely novel type of transporter.

LCI1 is encoded by nuclear DNA, translated in the cytoplasm, and localized in the plasma membrane, where it is involved in Ci transport and accumulation.° LCI1 was predicted to possesses four hydrophobic domains, but the first domain at the N-terminal end was presumed to serve as a signaling peptide.° To elucidate the structural basis of the mechanisms for Ci enrichment, we here report a crystal structure of the full-length LCI1 membrane protein, which is consistent with LCI1 being able to catalyze the transport of CO$_2$ across the plasma membrane.

**Results and discussion**

The crystal structure of LCI1 was determined to a resolution of 3.2 Å using single isomorphous replacement (Fig. 1, Fig. S1 and Table S1). Three LCI1 molecules, which form a homotrimer, were found in the asymmetric unit. Superimposition of these three protomers gives RMSDs between 0.39 and 0.41 Å (over 165 Cα atoms), suggesting that their conformations are nearly identical to each other. The composition of LCI1 is predominantly hydrophobic, with most of the amino acids embedded in the transmembrane. Apparently, *LCI1* does not contain any signaling peptide. Thus, the N- and C-terminal ends of this protein
are located at the cytoplasm. Each LCI1 protomer comprises four helical transmembrane segments (TMs) and two extra-membrane $\alpha$-helices ($\alpha$s), which are located right above the plasma membrane surface. The TMs and $\alpha$s are designated numerically from the N- to C-termini: TM1 (6-38), $\alpha$1 (47-62), TM2 (73-95), TM3 (108-133), $\alpha$2 (136-141) and TM4 (144-169). Of the four TMs, TM1 is relatively long and part of this helix (residues 34-38) is exposed to solvent. In addition to this top portion of TM1, helices $\alpha$1 and $\alpha$2 make up a small periplasmic domain of the membrane protein. The LCI1 trimer orients in a way that its pseudo threefold axis is perpendicular to the membrane surface. This threefold axis is surrounded by the three closely packed TM3 helices, which seal the central trimer interface against the passage of substrates.

In a view parallel to the membrane plane, the LCI1 trimer is about 60 Å tall, 40 Å wide and 40 Å thick. A bundle of four antiparallel helices constitute an LCI1 subunit and it folds into a cylindrical structural feature within the transmembrane region. Based on the crystal structure, it is likely that each protomer forms a pathway for transport or conductance. Indeed, the bundle of four antiparallel helices of each LCI1 molecule contributes to create a tunnel that spans approximately two-thirds of the transmembrane (Fig. 2). Interestingly, the two short $\alpha$1 and $\alpha$2 helices, which are exposed to solvent in the periplasm, form the entrance of this channel. In particular, residues L42, Y52, V55, T137 and V141 are found to encircle this entrance. The tunnel is hydrophobic in nature. Nineteen aromatic and hydrophobic residues, including V26, A30, L33, L42, Y52, Y56, F60, I66, F77, I83, I84, A128, V141, F142, I147, A150, L154 and V157, participate to construct this tunnel. The locations of these hydrophobic residues suggest that they may be important to forming a path
for substrate transport. In addition, the three polar residues Q80, N125 and T137 are also involved in lining the wall of the tunnel.

Apparently, the structure captures a closed conformation of the LCI1 channel. A negatively charged residue E87 is found within the transmembrane region of LCI1. The side chain of this residue likely makes a hydrogen bond with the side chain of N161 to form a gate and blocks this transmembrane tunnel (Fig. 2). Interestingly, the interior surface of this LCI1 channel is strikingly electronegative (Fig. 3), suggesting that this channel may tend to transport neutral or positively charged substrates. Thus, LCI1 may be in favor of CO$_2$ rather than HCO$_3^-$ and facilitate the passage of this neutral molecule through the plasma membrane.

Surprisingly, an extra electron density was found right below this gate and towards the cytoplasmic side of the elongated tunnel of each subunit of the LCI1 protein, indicating the existence of a fortuitous bound ligand co-purified and co-crystallized with this membrane protein. The shape of this extra density is compatible with a CO$_2$ molecule (Fig. 4). Within 4.5 Å of the bound ligand, it was found that there are at least five amino acids, including T19, T94, F91, I115 and T164, involved in the binding. Ab initio calculations have suggested that these amino acids have a strong affinity for CO$_2$\textsuperscript{15,16}, posing an idea that the fortuitous ligand may be CO$_2$.

We then used AutoDock Vina\textsuperscript{17} to search for potential CO$_2$-binding sites within the LCI1 channel protein. AutoDock Vina could only find one potential CO$_2$-binding site located within the bundle of four antiparallel helices of the membrane protein. Interestingly, this site coincides with the extra electron density (Fig. 4), suggesting that LCI1 may be capable of binding CO$_2$.
To confirm the trimeric architecture of LCI1, we turned to use non-denaturing mass spectrometry (native MS) to elucidate the oligomerization state of this membrane protein in aqueous phase. This approach should allow us to obtain valuable information regarding the sample mass, oligomerization state and subunit stoichiometry. In addition, it should lead us to identify bound lipid molecules if they are in complex with the membrane protein. The spectra showed a well-resolved charge state series that correspond to the trimeric oligomerization of LCI1. Additionally, native MS revealed the presence of bound lipids co-purified with this protein (Fig. 5). Using lipidomic approach, we identified that these bound lipids are phosphatidic acid (16:0/18:1) (PA(16:0/18:1)) and phosphatidic acid(18:1/18:1) (PA(18:1/18:1)). These lipids are bound within the purified LCI1 protein in different combinations. The observed masses of these lipid-LCI1 complexes are listed in Table S2.

To elucidate if LCI1 is capable of transporting CO₂ across the membrane, we performed steered molecular dynamics (SMD) simulations on this protein embedded in palmitoyloleoylphosphatidylethanolamine (POPE) membrane bilayer. We measured the magnitude of applied forces necessary to maintain a constant velocity of the CO₂ molecule moving through the channel created by the LCI1 protomer. We initially pushed CO₂ into the LCI1 channel via the opening created by the α1 and α2 helices. SMD suggests that CO₂ can migrate smoothly into the channel. The force applied to maintain a constant velocity of CO₂ migration is shown in Fig. 6. Interestingly, as CO₂ reaches the vicinity of the CO₂-binding site identified by the crystal structure, a noticeable reduction in the applied force is observed. The calculation suggests that residues lining the wall of this region may have the capacity to push CO₂ further into the channel, allowing the CO₂ molecule to continue to propagate farther down the channel and then exit to the cytoplasm.
We also performed SMD simulations by applying a reversed force to push CO₂ into the channel from the cytoplasmic side. The simulations indicate that the CO₂ molecule can only migrate a short distance into the channel. As it arrives the CO₂-binding site, we observed a significant conformational change between the transmembrane helices TMs 1 and 2 of LCI1. These two helices seemingly perform a scissor-like motion and push CO₂ back into the cytosol, suggesting that this channel is unidirectional.

Similar SMD simulations were performed to investigate the possibility of HCO₃⁻ permeation through the channel. SMD simulations reveal that HCO₃⁻ can migrate from the periplasmic to cytoplasmic sides of LCI1 using a similar pathway. However, the forces involved to push the HCO₃⁻ ion through this channel are at least five times higher, suggesting that LCI1 is more favorable to CO₂ passage (Fig. 6). These data are in good agreement with the electronegative nature of the LCI1 channel, which may not tolerate negatively charged ions. Like the CO₂ simulations, SMD calculations suggest that HCO₃⁻ can only migrate a short distance when a reverse force was applied to push this ion from the cytoplasmic side into the LCI1 channel.

The crystal structure of LCI1 suggests a substrate translocation mechanism, in which each LCI1 molecule within the trimer forms a channel for transport. Based on our structural model, we hypothesize that the reversible conversion of CO₂ and HCO₃⁻ is a necessary prerequisite for importing Ci into the cytoplasm through this channel. This conversion process is driven by the electronegativity of local protein environments. The exterior surface of the small periplasmic domain of LCI1 does not seem to form extensive positively or negatively charged patches, suggesting that both CO₂ and HCO₃⁻ are popular Ci species at this region. However, the entrance of each channel is surrounded with residues Y52, V55,
T137 and V141, which are more favorable for the passage of CO₂ molecules. Thus, we expect that CO₂ is a more preferred Ci species than HCO₃⁻ to enter the LCI1 channel from the periplasm (Fig. 7). Within the LCI1 channel, the interior surface is electronegative. This suggests that CO₂ should be the predominant Ci species inside the channel and LCI1 should primarily translocate CO₂ across the plasma membrane. At the cytoplasmic face of LCI1, it is found that its outermost surface, facing the cytoplasm, is highly electropositive. This local environment makes it more favorable for the negatively charged HCO₃⁻ ions. Thus, the last step for Ci translocation through LCI1 is that CO₂ must convert to HCO₃⁻ before releasing to the cytoplasm. Since LCI1 itself has no recognizable carbonic anhydrase active site in this region, or elsewhere, this apparent need for hydration of transported CO₂ to HCO₃⁻ suggests a close interaction of LCI1 with a carbonic anhydrase at the cytoplasmic surface. LCI1, therefore, apparently operates via a transport mechanism unlike those of other Ci transporters functionally confirmed as HCO₃⁻ transporters in cyanobacteria and algae, and unlike any previously described Ci transporter. It is a unique protein that transports Ci via a completely novel mechanism.
Figure 1. Structure of the *Chlamydomonas reinhardtii* LCI1. (a) Ribbon diagram of a monomer of LCI1 viewed in the membrane plane. The molecule is colored using a rainbow gradient from the N-terminus (blue) to the C-terminus (red). (b) Ribbon diagram of a trimer of LCI1 viewed in the membrane plane. The three protomers are colored green, red and yellow, respectively. The transmembrane segments (TMs) and α-helices (αs) of the front protomer (green) of LCI1 are labeled. The Figure was prepared using PyMOL (http://www.pymol.sourceforge.net).
Figure 2. Channel of LCI1. (a) Each subunit of LCI1 forms a channel (colored green) spanning approximately two-thirds of the transmembrane. This channel was calculated using the program HOLE\textsuperscript{27}. The channel is closed by an interaction between the side chains of residues E87 and N161. (b) The channel (colored pink) of LCI1 is surrounded with 22 amino acids, including V26, A30, L33, L42, Y52, Y56, F60, I66, F77, Q80, I83, I84, N125, A128, T137, V141, F142, I147, A150, L154 and V157. The potential CO$_2^-$ binding site is located right below residues E87 and N161. This channel was calculated using the program CAVER (http://loschmidt.chemi.muni.cz/caver). The secondary structural elements of the LCI1 protomer are in gray. Residues that are involved in forming this channel are in green sticks. E87 and N161 are also included in the figure.
Figure 3. Electrostatic surface potentials of LCII1. Surface representations of the (a) inside, (b) top view and (c) bottom view of the LCII1 channel colored by charge (red; negative -15 kT/e, blue; positive +15 kT/e).
Figure 4. Carbon dioxide binding site. (a) Stereo view of the F_c – F_c electron density map of bound CO_2 in LCI1. The bound CO_2 molecule is shown as a stick model (cyan, carbon; blue, nitrogen). The F_o – F_c map is contoured at 3.0 σ (blue mesh). Residues involved in CO_2 binding are in green sticks. The bound CO_2 ligand is shown as green sticks. (b) This is a composite figure showing the locations of the predicted bound CO_2 ligand (yellow) and bound CO_2 (cyan) found in the crystal structure of LCI1.
Figure 5. Mass spectrum of LCI1 obtained under native conditions. The spectra indicate the presence of the LCI1 trimer (purple circles) and its charge state series. The measured mass of this trimer is 66113.01 ± 0.27 Da, which is in good agreement with the theoretical mass of the trimeric LCI1 protein (66115.8 Da) without the first methionine residue on each monomer. The other species (cyan, yellow and red triangles) correspond to the lipid bound trimers and these are endogenously purified. Lipid analysis on this sample revealed the presence of PA(16:0/18:1) and PA(18:1/18:1) as shown in Fig. S2. These lipids are bound in different numbers and combinations. All of the molecular masses observed are listed in Table 2.
Figure 6. Steered molecular dynamics (SMD) simulations of the migration of CO$_2$ and HCO$_3^-$ through LCI1. (a) The trajectory of CO$_2$ through LCI1 in SMD is shown as blue mesh. LCI1 is oriented with the channel along the z-axis and the position of ligand along the channel is measured as distance from center of mass the protein along the z-axis. (b) Plots of applied forces as a function of ligand positions along the channel (blue, CO$_2$; red, HCO$_3^-$). The plots indicate that pushing of HCO$_3^-$ through the channel requires much larger force, suggesting that CO$_2$ may be the preferred ligand. The region corresponds to the observed CO$_2$-binding site is highlighted by a black rectangle. The two peaks showing in the CO$_2$ plot (blue) indicate local electrostatic interactions between (d) bound and residues E87 and N161; and (e) bound CO$_2$ and residue Q16. (f) Root mean square fluctuations (RMSF) of the LCI1 residues during the SMD simulations with CO2 (blue) and HCO$_3^-$ (red).
Figure 7. Proposed CO₂ transport across the plasma membrane via LC11. The LC11 trimer is represented by a green envelope showing in the cartoon. The channel formed within a monomer of LC11 is colored red.
Methods

Cloning, expression and purification of *C. reinhardtii* LCI1

Briefly, the full-length LCI1 membrane protein containing a 6xHis tag and a GFP tag at the C-terminus was overproduced in *Pichia pastoris* SMD1168H cells possessing the pPICZaΩLCI1 expression vector. Cells were grown at 29°C in 2 l of fermentation basal salts medium containing 26.7 ml/l 85%-phosphoric acid, 18.2 g/l K₂SO₄, 0.93 g/l CaSO₄, 18.2 g/l K₃SO₄, 14.9 g/l MgSO₄·7H₂O, 4.13 g/l KCl, 20 g/l glycerol and 4.4 ml/l PTM₁ trace salts (6.0 g/l CuSO₄·5 H₂O, 0.08 g/l NaI, 3.0 g/l MgSO₄·H₂O, 0.2 g/l Na₂MoO₄·2H₂O, 0.02 g/l H₃BO₃, 0.5 g/l CoCl₂, 20.0 g/l ZnCl₂, zinc chloride, 65.0 g/l FeSO₄·7H₂O, 0.2 g/l biotin, 5.0 ml/l H₂SO₄) (Invitrogen). The fermentation batch was bubbled with pure oxygen with a rate of 2 l/min. Temperature and pH were automatically maintained at 29 °C and 5.0, respectively, by water-cooling and the addition of 30% NH₄OH solution. Glycerol fed-batch process was then initiated when the glycerol concentration was dropped to zero. 50% glycerol and 12 ml/l PTM₁ were fed in a rate of 0.33 ml/min for 8 h. The expression was then induced using 100% methanol and 12 ml/l PTM₁ with a rate of 0.06 ml/min. Cells were harvested within 32 h of induction by centrifugation.

The collected yeast cells were resuspended in low salt buffer containing 100 mM sodium phosphate (pH 7.2), 10 % glycerol, 1 mM ethylenediaminetetraacetic acid (EDTA) and 1 mM phenylmethanesulfonyl fluoride (PMSF), and then disrupted with a French pressure cell. Debris was removed by centrifugation at 1500 x g. The membrane fraction was collected and washed twice with high salt buffer containing 20 mM sodium phosphate (pH
7.2), 2 M KCl, 10% glycerol, 1 mM EDTA and 1 mM PMSF, and once with 20 mM HEPES-NaOH buffer (pH 7.5) containing 1 mM PMSF as described previously\(^{28}\). The membrane protein was then solubilized in 2% (w/v) n-dodecyl-β-D- maltoside (DDM). Insoluble material was removed by ultracentrifugation at 100,000 x g. The extracted protein was purified with a Co\(^{2+}\)-affinity column. The purified protein was loaded into a PD-10 desalting column (GE Healthcare Bio-Sciences) to remove imidazole and then concentrated to 10 mg/ml in a buffer containing 20 mM Na-HEPES (pH 7.5) and 0.03% DDM. The 6xHis tag and GFP tag at the C-terminus were then cleaved by adding 1 unit of thrombin (Fisher Scientific) per 10 mg of purified LCI1 at room temperature for 20 h. A final purification step was performed using a G200 size exclusion column loaded with buffer solution containing 20 mM Na-HEPES (pH 7.5) and 1.1% n-octyl-β-D-glucopyranoside (β-OG). The purity of the LCI1 protein (>95%) was judged using 15% SDS-PAGE stained with Coomassie Brilliant Blue. The purified protein was then concentrated to 10 mg/ml in a buffer containing 20 mM Na-HEPES (pH 7.5) and 1.1% β-OG.

**Crystallization of C. reinhardtii LCI1**

Crystals of the LCI1 protein were obtained using hanging-drop vapor diffusion. The LCI1 crystals were grown at room temperature in 24-well plates with the following procedures. A 0.5 µl protein solution containing 10 mg/ml LCI1 in 20 mM Na-HEPES (pH 7.5) and 1.1% (w/v) β-OG was mixed with a 0.5 µl of reservoir solution containing 18% polyethylene glycol (PEG) 2000 MME, 0.1 M sodium citrate (pH 6.5) and 0.01 M NiCl\(_2\). The resultant mixture was equilibrated against 500 µl of the reservoir solution at 25°C. Crystals of
LCI1 grew to a full size in the drops within two months. Typically, the dimensions of the crystals were 0.1 mm x 0.1 mm x 0.05 mm. These LCI1 crystals diffracted x-rays to >4 Å resolution. To improve the resolution limit, crystals of the mercury derivative were prepared by incubating the crystals of LCI1 in solution containing 18% PEG 2000 MME, 0.1 M sodium citrate (pH 6.5), 0.01 M NiCl₂ and 1 mM HgCl₂ for 15 min at 25°C. Cryoprotection of these crystals was achieved by raising the glycerol concentration stepwise to 25% with a 5% increment in each step.

**Data collection, structural determination and refinement**

All diffraction data were collected at 100K at beamline 24ID-C located at the Advanced Photon Source, using a Platus 6M detector (Dectris Ltd., Switzerland). Diffraction data were processed using DENZO and scaled using SCALEPACK²⁹. Crystals of LCI1 belong to space group P3₁21 (Table S1). Based on the molecular weight of LCI1 (21.4 kDa), the asymmetric unit is expected to contain three membrane protein molecules with a solvent content of 64.1%. Three mercury sites were identified using SHELXD³⁰ as implemented in the HKL2MAP package³¹. These heavy-atom sites were refined by single anomalous dispersion (SAD) at a resolution of 4 Å using the program AutoSol implemented in PHENIX³². Phases were then subjected to density modification, NCS averaging, and phase extension to 3.19 Å-resolution using the program RESOLVE³³. The resulting phases were of excellent quality, which allowed us to trace most of the secondary structural features of the three LCI1 molecules within the asymmetric unit. After tracing the initial model manually using the program Coot³⁴, the model was refined using PHENIX³² leaving 5% of reflections
in Free-R set. Feature-enhanced maps calculated using PHENIX and B-factor sharpening maps created using CCP4 were employed to ascertain loop-regions and side chains in the structure. Iterations of refinement using PHENIX$^{32}$ and CNS$^{35}$ and model building in Coot$^{34}$ lead to the 3.19 Å-resolution structural model of the LCI1 trimer with excellent geometrical characteristics (Table S1).

**Docking of CO$_2$**

The program AutoDock Vina$^{26}$ was used to predict the CO$_2$-binding mode. A monomer of the structure of LCI1 with the bound CO$_2$ molecule removed was used for docking. A grid of 35 Å x 35 Å x 35 Å with 0.375 Å spacing was calculated around the docking area using AutoGrid. The iterated local search global optimizer algorithm was used to predict the binding free energies for these compounds.

**Native mass spectrometry**

Purified LCI1 protein was buffer exchanged in 200 mM ammonium acetate (pH 8.0) and 0.05% lauryldimethylamine oxide (LDAO) using a Biospin-6 column (BioRad) prior to mass spectrometry analysis. The protein was directly introduced into the mass spectrometer using gold-coated capillary needles prepared in house$^{36}$. Data was collected on a modified QExactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific Inc., Berman, Germany) optimized for analyzing high mass complexes$^{37}$. Optimized instrument parameters were as follows: capillary voltage 1.2kV, S-lens RF potential 100V, quadrupole
selection range between 2,000 and 20,000 m/z, collisional activation in the HCD cell 100V, argon pressure in the HCD cell 1.12x10-9 mbar and resolution of the instrument was acquired at 17,500 with m/z = 200 (transient time = 64 ms).

**Lipid analysis**

Lipid analysis was carried out by following the protocol as described previously\textsuperscript{38}. Briefly, intact LCI1 protein was digested with trypsin overnight at 37\textdegree C, lyophilized and re-dissolved in 35% acetonitrile. The peptide/lipid mixture was loaded onto a C18 column (Acclaim PepMap 100, C18, 75 µm × 15 cm; Thermo Scientific) and separated with a linear gradient of 35-100% acetonitrile. The column eluent was then delivered to a hybrid LTQ-Orbitrap XL mass spectrometer (Thermo Scientific) that was coupled to the column. The LTQ-Orbitrap XL was operated in negative ion mode and in data-dependent acquisition set-up to perform five MS/MS scans per MS scan. Survey full-scan MS spectra were acquired in the orbitrap (m/z 350-2000) with a resolution of 60,000.

**Steered Molecular Dynamics**

A 120 Å x 120 Å palmitoyl-oleyl-phosphatidylethanolamine (POPE) membrane bilayer was constructed using the membrane builder plugin in NAMD\textsuperscript{39} with the membrane normal parallel to the z-axis. The protein was embedded in the membrane such that its threefold symmetrical axis was perpendicular to the membrane plane. The system was then solvated by the addition of 5 Å of water layers on both sides of the lipid bilayer. Na\textsuperscript{+} and Cl\textsuperscript{−}
ions were also included to ensure the balance of charges, resulting in a total number of 112,470 atoms (112,472 atoms in the case of bicarbonate) in the system. After, the system was equilibrated at 10K and the temperature was slowly increased to 310 K in steps of 1 K per 1 ps each step for 200,000 steps. A harmonic restraint of 1 kcal/mol Å² was applied to the protein during the process of raising the temperature. A 100 ps equilibration run was then performed at 310K, in which all atoms were allowed to move freely. Langevin dynamics (1 ps⁻¹ damping coefficient) was used to maintain the constant temperature. A cutoff distance of 12 Å was used to mimic the effect of van der Waals interactions. A periodic boundary condition was imposed using the particle mesh Ewald method with a 1 Å grid spacing to evaluate long-range full electrostatic interactions. In all simulations, NAMD³⁹ with CHARMM27 parameter set⁴⁰,⁴¹ was used. The topology and parameter files for CO₂ were generated using the SwissParam server⁴¹,⁴² and those for bicarbonate ion were obtained using the Paratool plugin in VMD⁴³.

SMD simulations were carried out to investigate the transport pathway of CO₂ and HCO₃⁻. These ligands were placed accordingly at one end of the channel. Constant velocity steered MD (cv-SMD) was applied. Specifically the carbon atom of CO₂ or HCO₃⁻ was pulled in a direction perpendicular to the membrane plane (Z-direction) using a harmonic constraint at a velocity of 30 Å/ns with a force constant (k) of 5 kcal mol⁻¹Å⁻². To prevent overall translational motion of the system, specific residues at the periphery of the periplasmic side (T39, V59, P73, N135 and D143) and cytoplasmic side (V10, V99, M107 and S170) of the protein were harmonically restrained with a force constant of 5 kcal/mol Å² in the z-direction.
Protein data bank accession code

Atomic coordinates and structure factors for the structure of LCI1 have been deposited at the RCSB Protein Data Bank with an accession code 5TSM.

Acknowledgements

This work was supported by an NIH Grant R01GM086431 (E.W.Y.). This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines of the Advanced Photon Source, supported by an award GM103403 from the National Institutes of General Medical Sciences. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357

References


**Figure S1a.** Stereo view of the experimental electron density map of the LCI1 trimer at a resolution of 3.19 Å. The electron density map is contoured at 1.2σ. The Cα traces of the LCI1 molecule in the asymmetric unit are showed as colored lines. Anomalous signals of the three Hg binding sites (contoured at 3σ) found in the asymmetric unit are colored red.
Figure S1b. Representative section of electron density near helices TM3 and TM4 of LCI1 protomer. The solvent-flattened electron density (50–3.19 Å) is contoured at 1.2σ and superimposed with the final refined model (green, carbon; red, oxygen; blue nitrogen; yellow, sulfur)
Figure S2. Identification of the bond lipids of lipid-LC11 complexes by native MS. Lipid analysis showed the two most abundant precursor ion peaks at m/z 673.48 and 699.49 are PA(16:0/18:1) and PA (18:1/18:1) respectively.
Table S1. Data collection, phasing and structural refinement statistics of LCI1.

<table>
<thead>
<tr>
<th></th>
<th>LCI1-Hg derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data set</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.008</td>
</tr>
<tr>
<td>Space group</td>
<td>P3₁21</td>
</tr>
<tr>
<td>Cell constants (Å)</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>85.24</td>
</tr>
<tr>
<td>b</td>
<td>85.24</td>
</tr>
<tr>
<td>c</td>
<td>209.78</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90,90,120</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50-3.19 (3.32-3.19)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100.0)</td>
</tr>
<tr>
<td>Total reflections</td>
<td>1,214,125</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>15,317</td>
</tr>
<tr>
<td>CC₁/₂ (outer shell)</td>
<td>0.51</td>
</tr>
<tr>
<td>Redundancy</td>
<td>9.3 (9.6)</td>
</tr>
<tr>
<td>R_pim (%)</td>
<td>4.2 (52.1)</td>
</tr>
<tr>
<td>⟨I/σ(I)⟩</td>
<td>19.83 (1.55)</td>
</tr>
<tr>
<td><strong>Phasing</strong></td>
<td></td>
</tr>
<tr>
<td>Number of sites</td>
<td>3</td>
</tr>
<tr>
<td>Figure of merit (acentric/centric)</td>
<td>0.91/0.85</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>R_work (%)</td>
<td>21.08</td>
</tr>
<tr>
<td>R_free (%)</td>
<td>28.41</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
</tr>
<tr>
<td>Overall (Å²)</td>
<td>24.52</td>
</tr>
<tr>
<td>Chain A/B/C</td>
<td>22.74/23.14/27.44</td>
</tr>
<tr>
<td>Rms deviations</td>
<td></td>
</tr>
<tr>
<td>Bond (Å)</td>
<td>0.012</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>1.224</td>
</tr>
<tr>
<td><strong>Ramachandran analysis</strong></td>
<td></td>
</tr>
<tr>
<td>Most favored (%)</td>
<td>98.0</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>2.0</td>
</tr>
<tr>
<td>Disallowed (%)</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Table S2. Observed masses of the lipid-LC11 complexes.

<table>
<thead>
<tr>
<th>Observed mass (Da) ± standard deviation (Da)</th>
<th>Adduct mass (Da)</th>
<th>Number of lipids bound</th>
<th>Type of lipid bound (confirmed by lipidomics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>66113.01 ± 0.27</td>
<td>670.26</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>66783.27 ± 0.08</td>
<td>705.77</td>
<td>1</td>
<td>PA (16:0/18:1)</td>
</tr>
<tr>
<td>66818.78 ± 0.05</td>
<td>1343.19</td>
<td>2</td>
<td>PA (16:0/18:1)</td>
</tr>
<tr>
<td>67456.20 ± 0.30</td>
<td>1377.67</td>
<td>2</td>
<td>PA (16:0/18:1) + PA (18:1/18:1)</td>
</tr>
<tr>
<td>68150.16 ± 1.47</td>
<td>2047.15</td>
<td>3</td>
<td>2×PA (16:0/18:1) + 1×PA (18:1/18:1)</td>
</tr>
</tbody>
</table>
CHAPTER IV

STRUCTURAL STUDY OF RECEPTOR LIKE KINASE FERONIA OF ARABIDOPSIS THALIANA

Introduction

Unlike animals, plants are unable to mobilize themselves to adapt the environment changes. Thus, plant cells have developed the strategies that involve signal perception and transduction across the cell membrane. On the surface of the plasma membrane, which is the first line of the defense and signal reception of cells, a large group of transmembrane proteins are responsible for receiving incoming signals and stimulate the cell interior reaction pathways. These receptor-like kinases (RLK) are single transmembrane helix proteins with an extracellular receptor domain for signal reception and a cytoplasmic kinase domain that induces cellular responses by phosphorylating downstream targets. In Arabidopsis thaliana, there are more than 610 RLK homologs encoded in its genome but only few of them have known function as well as their ligands and signaling targets.\textsuperscript{1,2}

Among the RLK superfamily, a group of 17 proteins led by FERONIA have been identified by their extracellular sensory domain and are found to be involved in growth and reproduction.\textsuperscript{3,4} FERONIA was first found to be regulator in fertilization and pollen perception.\textsuperscript{5,6,7} The follow-up functional studies on FERONIA indicated that feronia mutants inhibited the root growth and root hair elongation in seedling phenotypes.\textsuperscript{8,9} Also, feronia seedlings are hypersensitive to abscisic acid (ABA), which plays a crucial role in plant stress response, which significantly reduced the root growth in the presence of ABA.\textsuperscript{9,10} In addition, studies showed that plants are more resistant to certain bacterial and fungal pathogens in the absence of functional FERONIA.\textsuperscript{11,12,13} These studies suggested that FERONIA precipitates in
various signaling pathways in plant growth, reproduction, pathogen response, and environmental stress response. As for the signal reception of FERONIA, a group of plant growth peptide hormones, which involve in root growth response, are believed to be the ligands of FERONIA. Rapid alkalization factors (RALFs) were found in many plant species and showed inhibition of root growth. In particular, RALF1 has been reported the capability of interacting with FERONIA binding domain thus induced phosphorylation at its C-terminal region and feronia mutated seedlings are insensitive to RALF1 treatment, which suggested RALFs are the ligands of FERONIA.

FERONIA consists of an extracellular sensory domain and a cytoplasmic kinase domain connected by single transmembrane helix. The extracellular sensory domain possesses two malectin-like domains, which distinguishes it from other members of RLK superfamily. Malectin is an animal protein known to bind dimeric and oligomeric glucose, which suggests that carbohydrates are possible ligand candidates of FERONIA and its homologs. Several crystal structures of the sensory domain in RLK superfamily have been reported, which provided insights into signal reception mechanism in RLK superfamily. However, these structures provided less information to understand the sensory mechanism of FERONIA due to its unique extracellular domain. Thus, the mechanism of ligand-protein interaction and downstream regulation of FERONIA remain elusive. In this study, we used structural and biochemical approaches to explore the function of the FERONIA receptor kinase and its peptide ligands and its downstream transcriptional factor. (nomenclature is usually “signal peptide”)

Figure 1. Conserved domains of FERONIA.
Results

Expressing FERONIA sensory domain in *E. coli.*

To prepare the FERONIA sensory domain for crystallization, we began by modifying the N-terminus MBP-fused FERONIA sensory domain expression construct provided by Dr. Yin’s group (GDBC, Iowa State university) by introducing a thrombin cleavage site between the MBP and FERONIA sensory domain. The resulting fusion protein was over-expressed in *E. coli* and purified using an amylose column. However, during the thrombin cleavage test, the fusion protein remained intact in 4 °C after treating thrombin for 16 hours. At room temperature, degradation of the sample was observed, which suggested the FERONIA sensory domain was not stable at room temperature. While MBP tags are known to highly promote the expression level of the target protein, it is not guaranteed that the expressed protein is functional and well folded. Based on the proteolysis results, the expressed FERONIA might not be folded properly.

![Figure 2. The proteolysis assay of purified MBP-FERONIA sensory domain by thrombin.](image-url)
**Refolding of FERONIA sensory domain**

To avoid over-expressing poorly-folded protein, we removed the MPB from the construct and cloned the full-length FERONIA sensory domain into expression vector pET15b. The expression test showed the target protein was highly expressed but aggregated in an inclusion body. We were able to obtain a soluble form of the FERONIA sensory domain by the rapid dilution method. However, size exclusion chromatography indicated that the protein was aggregated. After several attempts with different refolding methods, the aggregation issue was still unable to be resolved.

![Diagram](image)

**Figure 3.** a) Schematic depiction of FERONIA sensory domain construct used for expression. b) The SDS-page analysis of the refold FERONIA sensory domain and c) the elusion diagram of size exclusion chromatography.
Expressing two FERONIA melactin domains in *E. coli*.

Based on the results from the MPB fused construct and refolding attempts, it is unlikely that stable full length FERONIA sensory domain protein which is suitable for crystallization can be produced. However, there are two melactin sub-domains that have been identified in the FERONIA sensory domain. It would be a good alternative way to understand the architecture of FERONIA sensory domain by determining the structure of the two melactin domains. To test the expression of two melactin domains of FERONIA in *E. coli*, we generated the N-terminus 6xHis-GFP fused constructs for the each melactin domains. The expression results from Western blots and fluorescent intensity indicated low yield and degradation of the two proteins. It is suggested that *E. coli* might not be a suitable expression system for expressing the FERONIA sensory domain.

![Western blot and GFP fluorescent intensity](image)

**Figure 4.** a) The western blot and b) GFP fluorescent intensity of the GFP fused FERONIA melactin domains expression test with different *E. coli* strains and temperature.
**Expressing FERONIA sensory domain in *P. pastoris*.

To investigate the expression of FERONIA sensory domain in *P. pastoris* we started by checking the solubility of the expressed protein. The western blot result of yeast cell lysate indicated that the FERONIA sensory domain can be expressed and is soluble in *P. pastoris*. However, the size corresponding to the signal was two times larger than the calculated size of the FERONIA sensory domain. This size difference might cause by post-translational modification in yeast. There are ten predicted glycosylation sites found in the FERONIA sensory domain. It is possible that the protein was glycosylated while being expressed in yeast.

For the scaled-up expression, we used two different approaches to select the cell line that yielded the highest expression of recombinant protein. By fusing alpha-factor, the secretion peptide that can export the protein of interest out of the yeast cell, it should be possible to harvest the FERONIA sensory domain from the growth medium without disrupting the yeast cell. However, the western blot failed to detect the protein in the medium due to low yield. To corroborate this, we generated a C-terminus GFP fused construct which is widely used for checking membrane protein expression in yeast. The fluorescent intensities of selected colonies were significantly lower than the positive control after 48 hours of induction. The results suggest a low yield of FERONIA sensory domain in *P. pastoris*. 
**Figure 5.** a) The western blot analysis of expression FERONIA sensory domain in *P. pastoris*. Schematic depiction of the construct used and GFP fluorescence intensity results for expression test of a) C-terminus GFP fused FERONIA sensory domain and b) N-terminus MBP, C-terminus GFP fused FERONIA sensory domain.
Crystallization of FERONIA protein kinase domain

Based on the secondary structure analysis and crystal structure of the homolog (PDBID: 2QKW), we generated a construct of the FERONIA protein kinase domain (496-818) excluding the flexible loop containing phosphorylated sites at the C-terminus into the expression vector pET15b with a C-terminal His tag for protein crystallization. However, no colony appeared on the ampicillin selection plate after transformation which was consistent with the observation in Dr. Yin’s group when they attempted to express MBP fused full length C-terminus FERONIA in E. coli. The results may indicate that the FERONIA PK domain is toxic to E. coli.

To remove the toxicity, we introduced single-site mutagenesis at nucleotide binding domain (K565R) to deactivate the kinase activity. The resulting construct exhibited a good yield of the recombinant protein. The homogeneity of the protein was verified by the size exclusion chromatography. The sample was pre-incubated with ATP, ADP, or AMP before being subjected to crystallization screening. Crystals of FERONIA PK domain were observed in the AMP co-crystallization drop. After optimization, the best diffraction of the crystals extended up to 4Å. However, the overall diffraction patterns of the crystals were smeared and exhibited the behavior of multi-crystals. The following crystallization optimization and screening attempts of FERONIA PK domain were unable to resolved the problem. We then rebuilt the construct by removing a predicted flexible loop and N-terminal His tag followed by a thrombin cleavage site to increase the protein integrity. Still, no crystal was observed in the crystal screening.
Figure 6. a) The construct used for FERONIA kinase domain crystallization. b) Gel filtration and c) SDS-page analysis of the purified FERONIA kinase domain.
Figure 7. a) The crystal and b) the X-ray diffraction pattern of FERONIA kinase domain.
Crystallization of transcriptional factors RD26

Based on the study of Dr. Yin’s group (data unpublished), FERONIA inhibits the expression of the drought response related transcription factor RD26 which plays an important role in Abscisic acid (ABA) pathway. We expressed and purified the DNA binding domains of RD26. The purified protein was then subjected to crystal screening. The crystals of RD26 DNA binding domain were apparent after two weeks and grew to full size after a month. The crystals were unable to diffract beyond 7Å. Additionally, we tried to study the DNA protein complex by crystallography; however, no crystal was carried out during the crystal screening.

**Figure 8.** The crystal and the X-ray diffraction pattern of RD26 DNA binding domain.
Figure 9. a) Schematic depiction of RD26 DNA binding domain construct used for crystallization. b) The western blot results of RD26 expression in *E. coli*. c) The SDS-page analysis of His tag removal and gel filtration of the purified RD26 DNA binding domain.
**Purification of RALF and RALF23**

The plant growth peptide hormone RALF1 is believed to be one of the ligands for the FERONIA receptor kinase. Studies have shown that RALF1 inhibited root growth in a seeding growth assay. Also, RALF1 stimulated the phosphorylation of the C-terminus of FERONIA, which might lead to a down-regulation of AHA2, a H⁺-ATPase that plays an important role in yeast growth and root growth in plant. In collaboration with Dr. Yanhai Yin’s group, we focused on studying a RALF-like peptide RALF23 from both a functional and structural perspective. In our lab, we expressed C-terminus His tagged RALF1 and RALF23 in E. coli with MBP fused construct and purified in combined with immobilized metal affinity chromatography and reversed phase chromatography. The purified RALF and RALF23 were then subjected to crystal screens and bio-functional studies in Dr. Yin’s group. However, we were unable to crystallize RALF or RALK23 after several attempts.

![Image](image-url)
Conclusion and discussion

In this study, we first utilized different methods of expression and purification protocols to obtain recombinant FERONIA sensory domain for crystallization. Using the *E. coli* system, which is the most established prokaryotic expression system, we were able to over-express the FERONIA sensory domain as inclusion bodies or MBP-fused protein in soluble form. However, the MBP-fused protein appeared to be misfolded, resulting in complete degradation during thrombin treatment. For the attempts of refolding recombinant protein from solubilized inclusion bodies, only aggregated protein was observed in size exclusion chromatography after the refolding process. The results suggested that the expression of FERONIA sensory domain in *E. coli* failed to retain its structural integrity, which is a common issue for expressing recombinant eukaryotic proteins using prokaryotic expression systems. Protein maturation in eukaryotic organisms, especially membrane proteins, undergoes a more sophisticated process compared with expression in prokaryotic organisms. In eukaryotes, membrane proteins are synthesized at the surface of endoplasmic reticulum (ER) and folded into their matured form with the help of chaperones in the ER. The folded proteins are translocated to the Golgi apparatus for further processing before being transported to their final destinations. During the process proteins are modified to their final functional forms by post-translation modifications (PTM) such as phosphorylation and glycosylation. However, prokaryotic expression systems such as *E. coli* cannot perform this PTM and lacks the necessary enzymes and chaperones for proteins to mature. Although the previous studies indicated that the MBP-fused FERNOIA sensory domain exhibited the capability of binding ligand RALF1\textsuperscript{16}, it does not guarantee that the sample was suitable for
crystallization. The overexpressed protein might be heterogeneous or partially folded, however, the essential condition for protein crystallization is its homogeneity.

To acquire the proper folded and homogenous protein, we tested the expression of FERONIA sensory domain in eukaryotic expression system *Pichia pastoris*. By the western blot assay, a potentially post-translationally modified protein was observed, which agreed with the prediction on the open protein database (http://www.uniprot.org). However, we were unable to produce enough sample for analysis due to the low expression level in *P. pastoris*. This suggested the expression system might not be compatible for expressing the target protein. To overcome expression problems of the FERONIA sensory domain, a different expression system in higher evolved species might be required. The previous structural studies of the sensory domain in the RLK superfamily suggested that insect expression systems such as Sf-9 and high-five cells may be preferable for producing recombinant protein for crystallization.\textsuperscript{18,19,20,21} Recently, a new expression system in plant has been proposed,\textsuperscript{22} which could be the better approach for expressing plant proteins in future studies.

On the other hand, we crystallized the FERONIA kinase domain in the presence of AMP. The crystals diffracted to 5 Å overall but the diffraction patterns were not of an analyzable quality. Also, we were able to crystallize the DNA binding domain of RD26, which is a possible downstream target of FERONIA signaling pathway. However, the diffraction could not be extended higher than 7 Å. After extensive crystallization screening and condition optimization, the resolution of these crystals was still far from 4 Å, which is the cut off resolution of determining side chain location. One of the possible reasons for the poor diffraction is the disturbance in the crystal lattice. This may be due to the fact that proteins are not tightly packed while forming crystals, allowing other substances like water
molecules and small chemicals that are not involved in crystal formation into the crystals. The loose packing of the protein crystal lattice also permits spaces for the flexible domains of the protein to disorient the crystal, resulting in loss of high-resolution information in the diffraction pattern. We removed the flexible regions on the target proteins by secondary structure prediction and protein engineering. Nevertheless, the new constructs failed to crystallize. We then attempted to co-crystallize RD26 DNA binding domain with its target DNA sequence, hoping that the complex could stabilize the packing. Unfortunately, little progress has been made. One alternative method for acquiring structural information is to screen suitable candidates for crystallization among the closest homologs. By crystallizing the homologs which function similarly in the different species, we can utilize the homolog structure as the template to model the structure of the target protein for structural analysis.

Finally, we have shown the purification of two hormone peptides RALF1 and RALF23 which are the ligands of FERONIA. Currently, the protein samples are being subjected to growth phenotype assay in Dr. Yin’s group to study the signal response of FERONIA. Our lab is currently working on the co-crystallization of RALFs and a functional FERONIA sensory domain which may provide structural insight into understanding the sensory mechanism of FERONIA.
Methods

Purification of MBP-FERONIA sensory domain

A short linker and a thrombin cleavage site was placed between the MBP and FERONIA sensory domain in the vector provided by Prof. Yin’s group. The construct was over-expressed in *E. coli* BL21 cells. Cells were grown in 6L of LB medium with 50 µg/ml kanamycin at 37 °C. The cultures were then induced with 0.2mM IPTG overnight at room temperature. The cells were pelleted and suspended in 100 ml of ice-cold buffer containing 20 mM Na-HEPES (pH 7.5), 150 mM NaCl and 10% glycerol. The cells were then lysed and the cell debris was removed by centrifugation for 45 min at 4 °C and 36,000 rpm. The crude lysate was loaded onto a 3-ml amylose resin (NEB) pre-equilibrated with the lysis buffer. The column was washed with buffer 20 mM Na-HEPES (pH 7.5), 150 mM NaCl and eluded with the same buffer containing extra 10mM maltose. The maltose was remove by dialysis against the buffer 20 mM Na-HEPES (pH 7.5), 150 mM NaCl for 4L. The purity of the protein was judged using 12.5% SDS-PAGE stained with Coomassie Brilliant Blue.

Refolding of FERONIA sensory domain

FERONIA sensory domain protein was cloned into vector pET15b containing a 6xHis tag at the N-terminus was over-expressed in *E. coli* BL21 ΔAcrB cells. Cells were grown in 3L of Luria broth (LB) medium with 100 µg/ml ampicillin at 37 °C. The cultures were then induced with 0.2mM IPTG for 3 hours at 37 °C. The cells were pelleted and suspended in 100 ml of ice-cold buffer containing 20 mM Na-HEPES (pH 7.5), 150 mM NaCl and 10% glycerol and then lysed with a French pressure cell. Cell debris
containing inclusion body was collected by centrifugation for 45 min at 4 °C and 36,000 rpm. The pellet was then homogenized and washed with 1% Triton X-100 two times to remove detergent soluble impurities and left the pure inclusion body. The resided Triton X-100 was removed by homogenized and washed with the final buffer 20 mM Na-HEPES (pH 7.5), 150 mM NaCl. The pure inclusion bodies were then dissolved in the final buffer containing desired denaturant (6M Guanidine·HCl or 8M Urea) for 30 minutes at room temperature by stirring. The insoluble portion was removed by centrifugation and the denatured protein was loaded onto a Hi-Trap Ni²⁺-chelating column (GE Healthcare) pre-equilibrated with the denaturant buffer. The column was washed with the imidazole gradient and eluded with denaturant buffer with extra 300 mM imidazole. The purified denatured sample was ready for different refolding methods.

The refolding attempts of FERONIA sensory domain were carried out by rapid dilution and in-column refolding methods. For the rapid dilution method, the purified denatured protein solution was diluted into the refolding buffer (20 mM Na-HEPES (pH 7.5), 150 mM NaCl, 10mM β-Mercaptoethanol and 0.5M L-arginine) to the final concentration of 1mg/ml by vigorous mixing. The diluted sample was then dialysis against 4L of final buffer (20 mM Na-HEPES (pH 7.5), 150 mM NaCl) to remove denaturant and additives which may inhibit protein crystallization. As for the in-column refolding method, the purified protein was retained in the Ni²⁺ column without elusion. The refolding process was performed by washing the column with a step-wise decreased of denaturant as following, 10ml of final buffer (20 mM Na-HEPES (pH 7.5), 150 mM NaCl, 10mM β-Mercaptoethanol) containing 6M, 4M, 2M, 1M Urea respectively. The column was washed with 20ml final buffer before
elusion. The qualities of the refold FERONIA sensory domain in different methods were verified by size exclusion chromatography.

**Expression screen of FERONIA sensory domain in *P. pastoris***

The transformation protocols of both C terminus GFP fused FERONIA sensory domain and N terminus MBP, C terminus GFP fused FERONIA sensory domain into *P. pastoris* generally followed the *Pichia* Expression kit (Invitrogen). In brief, the genes containing the fusion protein were cloned into expression vector pPICZ-A (Invitrogen) and purified using High-Speed Plasmid Mini kit (IBI Scientific). 10 µg of DNA was linearized by treating with enzyme PmeI (NEB) in 37°C for 5 hours. The reaction was disrupted by heat inactivation for 20 min at 65°C. Electrocompetent cells SMD1168H were prepared following the protocol outlined in the *Pichia* Expression kit (Invitrogen). Linearized plasmid DNA was incubated with 100 µL of electrocompetent SMD1168H on ice and electroporated using a BioRad Gene Pulser. The cells were plated on YPDS containing 200 µg/ml and 500 µg/ml zeocin. The plates were then placed in the incubator at 29°C until the colonies grown to desirable size (approximately 3 days).

The colonies for expression screening were suspended in 60 µL BMGY(2.0% Peptone, 1.0% Yeast extract, 100mM Potassium phosphate pH 6.0, 1.34% Yeast Nitrogen Base (w/o AA), 0.4µg/mL Biotin, 1.0% Glycerol) separately. 1 µL of each suspended colonies were set on the induction plate containing BMMY(2.0% Peptone, 1.0% Yeast extract, 100mM Potassium phosphate pH 6.0, 1.34% Yeast Nitrogen Base (w/o AA), 0.4µg/mL Biotin, 0.5% Methanol) at the designated positions. The expression levels of each
colony were verified by comparing the GFP fluorescent signals using Typhoon FLA 9500 (GE Healthcare) after 48 hours of induction at 29 °C.

**Purification and crystallization of FERONIA kinase domain**

FERONIA kinase domain (PK) protein containing a 6xHis tag at the C terminus and a thrombin cleavage site was over-expressed in *E. coli* BL21 cells. Cells were grown in 3L of Luria broth (LB) medium with 50 µg/ml kanamycin at 37 °C. The cultures were then induced with 0.2mM IPTG overnight at room temperature. The cells were pelleted and suspended in 100 ml of ice-cold buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl and 10% glycerol. The cells were then lysed with a French pressure cell. Cell debris was removed by centrifugation for 45 min at 4 °C and 36,000 rpm. The crude lysate was loaded onto a 3-ml Hi-Trap Ni²⁺-chelating column (GE Healthcare) pre-equilibrated with the lysis buffer. The column was washed with the following buffers, 50mL of 20 mM imidazole (pH 7.2), 150 mM NaCl, and 20 mM Tris-HCl (pH 7.5), 20mL of 60 mM imidazole (pH 7.2), 150 mM NaCl, and 20 mM Na-HEPES (pH 7.5). FERONIA PK domain was then eluted using 20mL of 300 mM imidazole (pH 7.2), 150 mM NaCl, and 20 mM Tris-HCl (pH 7.5). The purity of the protein was judged using 12.5% SDS-PAGE stained with Coomassie Brilliant Blue. The His tag was removed by Ni-NTA column after treating with thrombin for 16h at 4 °C. The sample protein was further purified with size exclusion chromatography with superdex 200 column, resulting a single homogenous peak which is suitable for crystallization.
Crystals of the FERONIA PK domain were obtained using sitting-drop vapor diffusion. The crystals were grown at room temperature in 96-well plates with the following procedures. The final concentration of 2mM AMP was mixed with 10mg/ml FERONIA PK domain and incubated on ice for 30 minutes before crystallization. A 0.5 µl protein solution containing was mixed with a 0.5 µl of reservoir solution containing 12-16% PEG 2000MME, 0.1M Na-Citrate 6.0. The resultant mixture was equilibrated against 500 µl of the reservoir solution at 25°C. Crystals of FERONIA PK domain grew to a full size in the drops in two weeks. The dimensions of the crystals were 0.05 mm x 0.2 mm x 0.1 mm. Cryoprotection of these crystals was achieved by raising the glycerol concentration stepwise to 25% with a 5% increment in each step. The diffraction data were collected at 100K at beamline 24ID-C located at the Advanced Photon Source, using a Platus 6M detector (Dectris Ltd., Switzerland).

**Purification and crystallization of RD26 DNA binding domain**

RD26 DNA binding domain protein was cloned into vector pET15b containing a 6xHis tag at the N terminus and a thrombin cleavage site. The construct was over-expressed in *E. coli* C43 codon plus ΔAcrB cells. Cells were grown in 3L of LB medium with 100 µg/ml ampicillin at 37 °C. The cultures were then induced with 0.2mM IPTG for 3 hours at 37 °C. The cells were pelleted and suspended in 100 ml of ice-cold buffer containing 20 mM Na-HEPES (pH 7.5), 150 mM NaCl and 10% glycerol. The cells were then lysed with a French pressure cell. Cell debris was removed by centrifugation for 45 min at 4 °C and 36,000 rpm. The crude lysate was loaded onto a 3-ml Hi-Trap Ni2+ chelating column (GE Healthcare) pre-equilibrated with the lysis buffer. The column was washed by imidazole
gradient in buffer 20 mM Na-HEPES (pH 7.5), 150 mM NaCl and eluded with the same buffer containing extra 300 mM imidazole. The purity of the protein was judged using 12.5% SDS-PAGE stained with Coomassie Brilliant Blue. The His tag was removed by Ni-NTA column after treating with thrombin for 16h at 4 °C. The sample protein was further purified with size exclusion chromatography with superdex 200 column.

Crystals of the RD26 DNA binding domain were obtained using hanging-drop vapor diffusion. The crystals were grown at room temperature in 24-well plates with the following procedures. A 0.5 µl protein solution containing 6 mg/ml RD26 DNA binding domain in 20 mM Na-HEPES (pH 7.5) and 150mM NaCl was mixed with a 0.5 µl of reservoir solution containing 3M NaCl, 0.1M Bis-Tris 6.0. The resultant mixture was equilibrated against 500 µl of the reservoir solution at 25°C. Crystals of RD26 DNA binding domain grew to a full size in the drops in one month. The dimensions of the crystals were 0.1 mm x 0.1 mm x 0.1 mm. Cryoprotection of these crystals was achieved by incubating the crystals in reservoir solution containing additional 25% glycerol for 1 minute. The diffraction data were collected at 100K at beamline 24ID-C located at the Advanced Photon Source, using a Platus 6M detector (Dectris Ltd., Switzerland).

**Purification of RALF1 and RALF23**

RALF1 and RALF23 core domain proteins containing a His6 tag at the C terminus were cloned into vector pMAL-c5x and were overproduced in *Escherichia coli* BL21 cells. Cells were grown in 3 liters of Luria broth (LB) medium with 50 µg/ml kanamycin at 37 °C. The cultures were then induced with 0.2mM IPTG overnight at room temperature. The
pelleted cells were suspended in 50 ml of ice-cold buffer containing 20 mM Na-HEPES (pH 7.5) and 150 mM NaCl. The cells were then disrupted with a French pressure cell. Cell debris was removed by centrifugation for 45 min at 4 °C and 36,000 rpm. The crude lysate was loaded onto a Hi-Trap Ni²⁺-chelating column (GE Healthcare) pre-equilibrated with the lysis buffer. To remove the impurities the column was washed with the lysis buffer containing additional 20mM Imidazole. MBP fused RALF1 and RALF23 core domain proteins were then eluted using the same buffer with 30mM Imidazole. The purity of the proteins was judged using 12.5% SDS-PAGE stained with Coomassie Brilliant Blue. The MPB was then cleaved by 1U thrombin (GE Healthcare Bio-Sciences, Pittsburgh, PA) per 10mg of protein for 16h at 4 °C. The Ralf and Ralf23 core domain were then further purified by HPLC with the C8 reverse phase column in 0.1% TFA and eluted with acetonitrile gradient. The corresponding fractions were then freeze-dried and stored at -20°C.

References


Mycobacterium tuberculosis possesses a unique cell wall, which is known to play a crucial role in virulence and antimicrobial resistance. In M. tuberculosis, the MmpL family, a group RND transporters that can be only found in mycobacteria family, contributes to cell wall biosynthesis by exporting fatty acid derivatives. The expression of the M. tuberculosis MmpL proteins is controlled by a complicated regulatory network system. In Chapter II, we demonstrated two high-resolution crystal structures of two forms of the TetR-family transcriptional regulator Rv0302, which was predicted to regulate the expression of multiple MmpL proteins. Comparison of the two Rv0302 crystal structures suggested that the conformational changes leading to disengagement from DNA consist of a rigid body rotational motion within the dimer interface of the regulator. Using fluorescence polarization and electrophoretic mobility shift assays, we demonstrated the recognition of promoter and intragenic regions of multiple MmpL genes by Rv0302. In addition, the isothermal titration calorimetry and electrophoretic mobility shift experiments indicate that fatty acids may be the natural ligand of this regulator.

Although our study has shown the possibility that Rv0302 regulates multiple MmpL genes in vitro, the real interactions of these genes and Rv0302 cells are yet to be determined. Besides, palmitic acid, which is the product of the fatty acid synthesis pathway, is the probable nature ligand of Rv0302; it is interesting to explore if Rv0302 is involved in regulating fatty acid synthesis. To study the regulation of Rv0302 in vivo, one
of the approaches could be to measure the change of mRNA levels of different genes in *M. tuberculosis* while Rv0302 is mutated.

The large cavity in the ligand binding domain of form I Rv0302 revealed the structural insight of ligand induced conformational change, providing the information for the ligand-protein interaction. This information can be subjected to future ligand discovery by the combination of high throughput docking simulation and biochemical analysis. Furthermore, the same strategy can lead to inhibitor development of this regulators, which occupy the ligand binding site without inducing disengage the protein from DNA and thus disrupting the expression of MmpL transporters. On the other hand, it is known that the ligands of transcriptional regulators are likely to be the substrates of their regulating targets. Therefore, exploring the possible ligands of Rv0302 could contribute to understanding the function of MmpL transporters as well as their roles in pathogenicity and antimicrobial resistance. In summary, the study of this regulator might provide a different perspective for future study of MmpL family and the cell wall formation of *M. tuberculosis*.

Carbon acquisition via photosynthetic CO₂ assimilation is responsible for the production of essentially all biological carbon and thus is essential for almost all life on earth. Phytoplankton such as microalgae, contribute more than half of all global photosynthetic CO₂ assimilation and require a carbon concentrating mechanism (CCM) to survive and flourish. Despite the importance of photosynthetic CO₂ assimilation and of microalgae specifically, it still is not fully understood how microalgae scavenge atmospheric CO₂. In Chapter III, we present a crystal structure of the *Chlamydomonas reinhardtii* LCI1 channel, which has shown a direct involvement in inorganic carbon
transport by the CCM. Combined with X-ray crystallography, mass spectrometry, and computational simulation, our data indicate that the LCI1 membrane protein forms a trimeric assembly, in which each protomer conducts uncharged CO\(_2\) and shuttles this inorganic carbon species across the cell membrane. LCI1 does not belong to a well-characterized transporter family or transport inorganic carbon by known carrier-mediated transporters. LCI1 not only is the first of these CCM-related transporters to have its structure solved, but it also represents a completely novel type of transporter.

While we have determined the crystal structure of a novel membrane transporter LCI1 and elucidated its structural details, the function and transportation mechanism of this protein remain elusive. Hence, future work should focus on investigating the biofunctional importance of LCI1 in CCM combined with structural information. The structural analysis and computational simulation of the crystal structure suggested that LCI1 might transport CO\(_2\) but \(\text{\textsuperscript{\text{\textminus}}HCO}_3\), however, there is no experimental evidence \textit{in vitro} or \textit{in vivo} to support analysis. Due to pH dependence of the interconversion between CO\(_2\) and \(\text{\textsuperscript{\text{\textminus}}HCO}_3\), which the conversion is in favor of CO\(_2\) in acidic environment and \(\text{\textsuperscript{\text{\textminus}}HCO}_3\) in alkaline environment, one of the plausible methods is to conduct the growth phenotype assay in \textit{C. reinhardtii} and compare the difference between the wild type and LCI1 mutant at different pH. Moreover, the structure of LCI1 has revealed a small and rigid soluble domain that allows no conformational change for scavenging substrate from the environment. It is unlikely that the LCI1 can concentrate Ci under very low Ci concentration without the support from other proteins. Therefore, the pull-down assay of LCI1 against cell lysate could be performed to identify the proteins that interact with LCI1 and investigate the involvement of these proteins in CCM.
Alternatively, the proteoliposome related assays may offer a new route to inspect transport mechanism LCI1 \textit{in vitro}. By incorporating the protein into the liposome, the proteoliposome mimics the natural environment of the membrane transporter. In addition, various substrate specific indicators can be encapsulated in the proteoliposome which serve as suitable markers for recognizing substrate transport. In our lab, we currently focus on stop-flow proteoliposome assay incorporated with pH sensitive indicator to study the transport mechanism of LCI1 as well as residues involved in substrate transportation by site-directed mutagenesis.

Structure determination of membrane protein in higher evolved species is always a great challenge for X-ray crystallography. The major obstacle is not the complexity of protein structure but the acquisition of the well-folded recombinant protein for crystallization. In Chapter IV, we demonstrate our efforts of determining the structure of a plant RLK protein FERONIA in \textit{Arabidopsis thaliana} and its possible downstream target RD26. In our attempts to expressing and purify the FERONIA sensory domain in prokaryotic host \textit{Escherichia coli}, the recombinant proteins failed to maintain its structural integrity and were unable to be recovered by protein refolding methods. The expression tests in yeast expression system \textit{Pichia pastoris} have shown that the sensory domain was modified by PTM but low yield, suggesting the need for more sophisticated expression host, such as insect cell or tobacco, for producing well-folded and homogenous protein for crystallization. On the other hand, we crystallized the FERONIA kinase domain and RD26 DNA binding domain but with poor diffraction. Considerable efforts are still being made in our lab for optimizing the crystal conditions.
To understand how FERONIA triggers the kinase activities at its C-terminus after receiving signal at sensory domain, it is crucial to acquire the structural information of the full-length protein. Unfortunately, crystallization of proteins with high flexibility as well as large protein complexes is almost an impossible task. Single particle cryo-electron microscopy (cryo-EM), on the other hand, is able to image protein structure directly without crystallizing it. This technique was limited by poor resolution data acquisition. However, due to the recent developments of the detector and data processing, single particle cryo-EM is now able to elucidate the protein structural information to atomic level and thus it could be the ultimate approach for revealing the overall structure of FERONIA.