

ELUCIDATING THE REACTION MECHANISM OF THE LARC NICKEL INSERTASE
FROM *MOORELLA THERMOACETICA* AND DEVISING A METHOD TO STUDY THE *LAR*
GENES IN *ESCHERICHIA COLI*

By

Aiko Turmo

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biochemistry and Molecular Biology – Doctor of Philosophy

2023

ABSTRACT

The nickel-pincer nucleotide (NPN) is a novel metallocofactor required for lactate racemase and similar activities. The name pincer derives from the metal ion being tri-coordinated in a planar orientation. These complexes are common in synthetic organic chemistry; however, the NPN cofactor is the first nickel pincer complex to be identified in Nature. Since its discovery in 2015, much work has been done to improve our understanding of the function and biosynthesis of this cofactor. Three enzymes, LarB, LarE, and LarC are involved in the biosynthesis of this cofactor. LarB adds a second carboxyl group to the pyridinium ring of the precursor nicotinic acid adenine dinucleotide (NaAD) and hydrolyzes the phosphoanhydride bond to form the product, pyridinium-3,5-biscarboxylic acid mononucleotide. LarE adds two molecules of sulfur which results in pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN). Finally, LarC completes the synthesis of the mature NPN cofactor by inserting the nickel ion.

Previously, LarC was shown to be a CTP dependent enzyme, but the function of this cofactor was not clear. Through mass spectrometry analysis and activity assay, I discovered a reaction intermediate, CMP-P2TMN, which provides insight on the role of CTP in the reaction. I speculate that the function of this adduct is to position the substrate correctly for the metal ion insertion. Working toward a better understanding of this process, I have obtained a preliminary cryo-electron microscopic protein structure of LarC from *Moorella thermoacetica*.

The NPN biosynthetic pathway and *larA* genes are found in almost a quarter of the analyzed prokaryotic genomes. The current process used to screen the functionality of these predicted homologs uses an *in vitro* method that is time consuming and error prone. I developed an efficient alternative method to confirm the roles of biosynthesis protein homologs and to

generate active NPN-containing proteins by implementing a co-expression system in genetically tractable *Escherichia coli*

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor and mentor, Professor Robert P. Hausinger, who guided me successfully to the completion of my Ph.D. His passion towards expanding the knowledge of the field has inspired me, and I hope to emulate his dedication and compassion in my future endeavors. I am also grateful to my lab mates in the Hausinger lab; I will always cherish the support everyone has given me and the laughter we shared in the lab.

Special thank you to my committee members, professors Tom Sharkey, Bjoern Hamburger, Jim Geiger and Jian Hu, for generously providing knowledge and expertise and guiding me throughout the years. I am deeply indebted to many of the wonderful MSU staff and researchers including professors, post-docs, facility scientists, program coordinators, lab technicians, and undergraduate researchers I had the privilege of working with and learning from. Thank you, Drs. Tina Dominguez Martin, Bryan Ferlez, and James Santiago for continuing to offer your guidance and friendship. I am grateful for the friendship that grew with other graduate students through my Ph.D. career, such as Drs. Ana-Maria Raicu, Alice Chu, and Basma Klump, who provided intellectual and moral support. Their camaraderie and encouragement helped me through the dissertation writing process.

I could not have undertaken this journey without my partner, Dalton Herbel. His kindness and patience know no bounds. Thank you to my family and friends, especially Michelle Beyers and Erica Nichols, for their unwavering belief in me, which has been a source of constant encouragement and motivation throughout this journey.

This endeavor was made possible by the generous support of the NSF Graduate Research Fellowship Program.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION: THE NICKEL-PINCER NUCLEOTIDE COFACTOR AND ITS BIOSYNTHETIC PATHWAY	1
REFERENCES	10
CHAPTER 2: CHARACTERIZATION OF THE NICKEL-INSERTING CYCLOMETALLASE LARC FROM <i>MOORELLA THERMOACETICA</i> AND IDENTIFICATION OF A CYTIDINYLYLATED REACTION INTERMEDIATE.....	13
REFERENCES	32
CHAPTER 3: PRELIMINARY STRUCTURAL STUDY OF <i>MOORELLA</i> <i>THERMOACETICA</i> LARC	34
REFERENCES	44
CHAPTER 4: EXPANDING THE METHOD TO STUDY THE <i>LAR</i> GENES IN <i>ESCHERICHIA COLI</i>	46
REFERENCES	60
CHAPTER 5: CONCLUSIONS AND FUTURE STUDIES	62
REFERENCES	67

CHAPTER 1

INTRODUCTION: THE NICKEL-PINCER NUCLEOTIDE COFACTOR AND ITS BIOSYNTHETIC PATHWAY

This chapter includes portions that were adapted from text originally published in:

Nevarez J. *, Turmo, A. *, Hu J., and Hausinger, R. P. (2020) Biological pincer complexes, *ChemCatChem*, 4;12(17): 4242-4254. *These authors contributed equally to the minireview.

Copyright © 2020 John Wiley & Sons, Inc. — Reproduced with permission.

Introduction

A novel organometallic cofactor was identified in the lactate racemase from *Lactiplantibacillus* (formerly *Lactobacillus*) *plantarum* and shown to be necessary for interconverting the L- and D- stereoisomers of lactic acid (Figure 1-1).¹ The lactate racemase

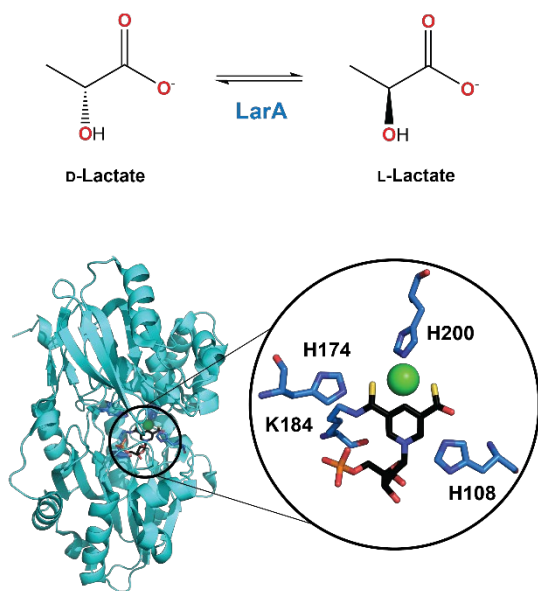


Figure 1-1 Lactate racemase, LarA.(Top) The interconversion of lactic acid stereoisomers by LarA. (Bottom) Crystal structure of LarA_{Lp} (PDB:5HUQ) with an insert highlighting the active site with the NPN cofactor. NPN in black stick; key residues in blue stick; nickel as a green sphere.

reaction was first observed in biology in 1936 from *Clostridium acetobutylicum* and *C. beijerincki*.² Racemization is a critical part of cell functionality since different isomers provide unique biological roles.^{3,4} For *L. plantarum*, D-lactate is essential for growth and is a component of its peptidoglycan component of the cell wall, where it provides resistance to vancomycin antibiotics.^{5,6}

Possibly due to this mandatory requirement, *L. plantarum* possesses both L- and D-lactate dehydrogenase and lactate racemase to ensure

this organism can obtain the stereospecific D-lactate through multiple means.

The nickel-pincer nucleotide (NPN) cofactor is one of two types of metal pincer complexes found in Nature, where the other type is the calcium- or lanthanide-pyrroloquinoline quinone complex.⁷ More commonly found in synthetic inorganic chemistry,^{8,9} pincers are characterized as planar ligands that tri-coordinate a metal ion; in the case of the NPN cofactor a nickel is coordinated by a carbon and two sulfur atoms (a so called SCS-type pincer) present in pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN). Recently it was shown that the

NPN cofactor is utilized by additional enzymes that catalyze racemase or epimerase reactions other than lactate racemization, e.g., enzymes were identified with specificity for malate, 2-hydroxyglutarate, phenyllactate, D-gluconate/D-mannonate, and other 2-hydroxyacid substrates.¹⁰ This finding expands the repertoire of the cofactor's functionality and opens the possibility of additional reactions that are still to be uncovered.

Characterization of lactate racemase and the role of NPN cofactor

Lactate racemase activity had been detected in several organisms since its discovery in 1936,² however the properties of the enzyme and its reaction mechanism had remained enigmatic until recently. Transcriptional and biochemical studies using *L. plantarum* revealed

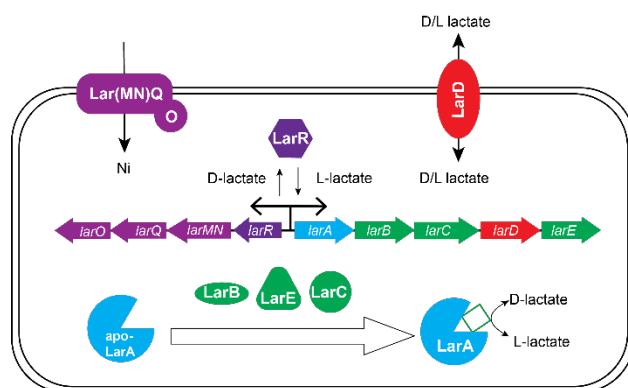


Figure 1-2 Overview of *L. plantarum* lar operons and gene functions.

two adjacent and oppositely transcribed operons (*larR(MN)QO* and *larABCDE*) that are associated with this activity (Figure 1-2).^{10,11} LarR is a transcriptional regulatory protein that increases expression of both operons when L-lactate is bound.¹² LarMN, LarQ, and LarO are three components of an

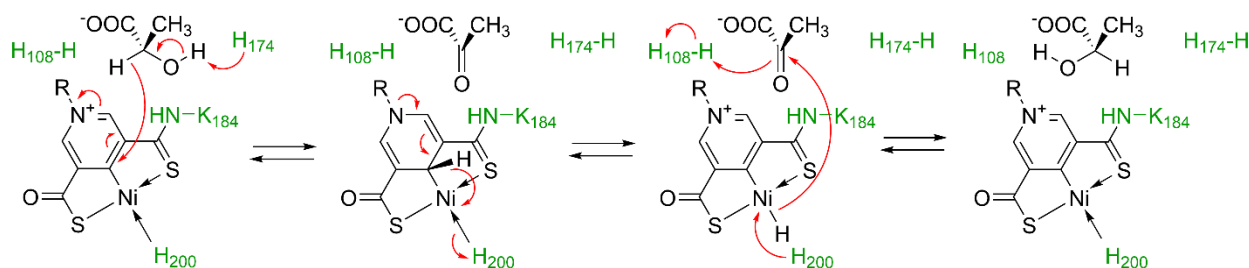
ATP-binding cassette type transporter that are most closely related to nickel uptake systems.

LarA is the protein responsible for lactate racemase activity. LarB is a homolog of PurE or carboxyaminoimidazole ribonucleotide mutase, but with an N-terminal extension. LarC is not related to any protein of known function. LarD is an aquaporin-type membrane protein that functions as a permease of both L- and D-lactate.¹³ Finally, LarE is a member of the diverse group of PP-loop ATPases. Significantly, lactate racemase activity requires the products of *larB*,

larC, and *larE* in addition to the presence of LarA,¹⁴ suggesting that the auxiliary components are needed to activate the enzyme.

Structural studies played a large role in characterizing the properties of lactate racemase. The report describing biochemical studies of LarA from *L. plantarum* (LarA_{Lp}) also had included the crystal structure of cofactor-free (apoprotein) LarA from *Thermoanaerobacterium thermosaccharolyticum* (LarA_{Tt}).¹⁵ More significantly, the structure of the LarA_{Lp} holoenzyme was solved and identified the composition of the cofactor, confirming mass spectrometric-based conclusions.¹ In LarA_{Lp}, a thioacid of the cofactor is covalently linked to Lys184 as a thioamide (Figure 1-1),¹ whereas the cofactor binds non-covalently to LarA_{Tt} even though a corresponding lysyl residue is present.¹⁶ The cofactor dissociates from LarA_{Tt} during purification or crystallization, thus explaining why the apoprotein structure was obtained. The apoprotein of LarA_{Tt} has been exploited for monitoring NPN cofactor biosynthesis; mixing of the protein with NPN cofactor rapidly confers lactate racemase activity.^{11,12} Likewise, the LarA_{Lp} holoenzyme has been useful for defining properties of the enzyme mechanism.

LarA_{Lp} catalyzes lactate racemization by using a proton-coupled hydride-transfer (PCHT) mechanism, possibly with two hydride binding sites on the cofactor (Scheme 1-1).¹³



Scheme 1-1 LarA activity using a proton-coupled hydride-transfer mechanism. Adopted from¹³.

Depending on whether L- or D-lactate binds, either His108 or His174 is thought to be appropriately positioned to serve as a general base to abstract the proton on the hydroxyl group. The C2 hydrogen atom transfers as a hydride to NPN as pyruvate is formed as a reaction

intermediate. It is plausible that the hydride binds to either of two distinct positions on the cofactor, C4 of the pyridinium ring or as a nickel-hydride (requiring dissociation of His200), which would account for the complexity of the NPN structure as opposed to the enzyme simply using NAD^+ (Figure 1-3). The hydride then returns to pyruvate by attacking either face, thus accounting for substrate racemization, as the corresponding His residue functions as a general acid. Evidence in support of this mechanism is derived from the direct identification of pyruvate during the quenched reaction, the demonstration of a substrate kinetic isotope effect when using substrate with ^2H at the C2 position, a change in visible absorption of the chromophore consistent with NPN reduction, and computational studies.¹⁴

Biosynthesis pathway of the NPN cofactor

The three auxiliary genes that co-localize and are co-regulated with *larA* in *L. plantarum* (*larB*, *larC*, and *larE*)¹⁶ were proposed to be involved in NPN cofactor biosynthesis.¹ Subsequent investigations revealed the pathway by which the corresponding gene products function in the biosynthesis of the novel SCS-type nickel-pincer complex.

In *L. plantarum* (and probably all other species containing the cofactor) the NPN structure is derived from nicotinic acid adenine dinucleotide (NaAD) with LarB_{Lp} catalyzing the first steps: carboxylation on C5 of the pyridinium ring followed by hydrolysis of the phosphoanhydride with release of AMP to form pyridinium-3,5-dicarboxylic acid mononucleotide (P2CMN).^{11,15} No external energy source is required for the carboxylation reaction, although it had been speculated that the energy released from the hydrolysis of NaAD was used for the carboxylation reaction.¹¹ A more recent investigation demonstrated that the two reactions are independent of each other and suggested that hydrolysis prevents the carboxylated NaAD product from binding to the enzyme and undergoing the reverse reaction, decarboxylation.¹⁵ That publication also solved

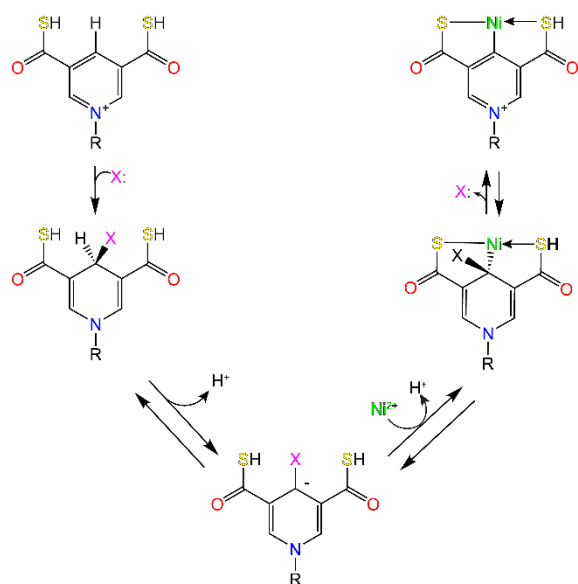
three structures of LarB_{Lp} in complex with a substrate analog, product AMP, and inhibitor zinc. Additional studies showed that LarB uses CO₂, not bicarbonate, for carboxylation, that a transient cysteinyl-pyridinium adduct is formed as a reaction intermediate (thus increasing the nucleophilicity of C5), and that hydrolysis occurs by water attack on the more distal phosphate (as shown by ¹⁸O-labeled water becoming incorporated into AMP, not P2CMN).¹⁵

In the second stage of the NPN biosynthetic pathway in *L. plantarum*, two molecules of LarE_{Lp} each sacrifice a cysteinyl side chain sulfur atom, forming dehydroalanine residues (Dha), while sequentially converting the P2CMN carboxyl groups into thiocarboxylates to form pyridinium-3-carboxy-5-thiocarboxylic acid mononucleotide (PCTMN) and P2TMN.^{11,17} On the basis of structural and mechanistic studies, each sulfur transfer reaction involves (i) ATP-dependent activation of a substrate carboxyl group by adenylation with the release of pyrophosphate, (ii) cysteinyl residue attack on the activated substrate to form a thioester with the release of AMP, (iii) deprotonation of the cysteinyl C α position, and (iv) sulfur transfer to form the product thioacid while generating Dha. This type of sacrificial sulfur transfer reaction is known to occur in only one other enzyme, thiamine thiazole synthase from *Saccharomyces cerevisiae*.¹⁸ The Dha-containing form of LarE_{Lp} is capable of being recycled *in vitro* by incubation with the persulfide of coenzyme A (CoA) followed by addition of a reductant.¹⁴ In this recovery reaction, the highly nucleophilic persulfide adds to the Dha residue yielding a CoA-LarE_{Lp} mixed disulfide that subsequently undergoes reduction. It is unclear whether Dha recycling is physiologically relevant; however, CoA binds to and stabilizes LarE_{Lp}.^{11,14} In addition to the LarE_{Lp} adducts with P2CMN and PCTMN formed during the sulfur transfer reactions, there is evidence for a LarE_{Lp} adduct of NPN, suggesting that nickel can insert into P2TMN while it is covalently bound to LarE_{Lp}.¹⁴ This

result explains why large amounts of isolated *L. plantarum* LarE, when purified from cells that co-produce LarB and LarC, can activate LarA apoenzyme.¹⁶

Homology models of some LarE homologs had indicated a tri-cysteine motif that could possibly bind an iron-sulfur cluster.¹⁷ Bioinformatics studies have shown that the majority of LarE homologs contain these three conserved cysteine residues, where the third is shifted in position by one residue from the single cysteine of LarE_{Lp}. This observation led to the hypothesis that an iron-sulfur cluster with a non-core sulfide could transfer the extra sulfur atom to substrate, thus potentially avoiding the energetically costly need to repair inactive protein containing a Dha residue. Recent biochemical studies have confirmed the presence of such a process using the homolog of LarE from *Thermotoga maritima* (LarE_{Tm}). The active form of this protein contains an oxygen-labile [4Fe-4S] cluster that appears capable of accepting a fifth sulfide from cysteine desulfurase (when it is provided with L-cysteine) and then catalytically transferring that sulfide to P2CMN. No mass change corresponding to Dha formation was detected for LarE_{Tm}. Because one iron atom of the [4Fe-4S] cluster is not coordinated by a cysteine residue, that open site is thought to be where the non-core sulfide is coordinated.¹⁹

In the terminal step of NPN cofactor biosynthesis, LarC catalyzes the remarkable reaction of installing nickel into the organic ligand by forming new nickel-carbon and nickel-sulfur sigma bonds.¹² This capability makes LarC the first enzyme identified to catalyze a cyclometalation reaction, a term used by inorganic chemists to reflect the formation of a metal-containing ring that includes carbon-metal and nucleophile-metal bonds.²⁰⁻²² The molecular mechanism of the LarC enzymatic reaction is unknown; however, a plausible nickel insertion mechanism is shown in



Scheme 1-2 Hypothetical nickel insertion mechanism of LarC. Adapted from²³.

Scheme 1-2²³. Activity assays demonstrated that LarC from *L. plantarum* (LarC_{Lp}) is a cytidine triphosphate (CTP)-dependent enzyme.¹² Efforts to obtain the structure of LarC_{Lp} yielded crystals of only the C-terminal domain, apparently derived by proteolysis from trace levels of protease in the sample. Fortuitously, x-ray crystallographic analysis of this domain in complex with CTP and manganese revealed a novel nucleotide-binding site. Substitution of the CTP-binding residues by

site-directed mutagenesis confirmed the importance of these residues in catalysis. Additional mutagenesis studies of residues in the N-terminal domain confirmed the importance of a His-rich region that is suspected to bind nickel and of several conserved carboxylate residues that may bind P2TMN or otherwise facilitate catalysis.¹² LarC_{Lp} hydrolyzes CTP to form CMP, but the role of this reaction was not defined. Of added interest, LarC_{Lp} appears to function stoichiometrically rather than catalytically; thus, LarC_{Lp} is a single-turnover enzyme.¹²

Prevalence of *lar* genes

Characterization of the NPN biosynthesis pathway has focused on enzymes from *L. plantarum*; however, an analysis of over 1,000 bacterial and archaeal genomes indicates about 9% contain genes that may encode homologs of LarA and the NPN biosynthetic proteins.¹⁶ Investigating selected homologs might lead to the discovery of alternative enzymes with enhanced stability or with distinct catalytic properties for generating the NPN cofactor. Furthermore, genome analyses have identified natural fusions of some of the biosynthetic enzymes, possibly allowing

for channeling of the pathway in Nature. An additional ~15% of the same list of genomes lack a homolog of the *larA* gene, but contain homologs to *larB*, *larC*, and *larE*. This finding suggests that the NPN cofactor is synthesized for purposes other than lactate racemization. Furthermore, some organisms have multiple paralogs of *larA*, again consistent with alternative roles for the cofactor.

16

Recent biochemical studies have determined the functions of seven out of 13 potential subgroups of *larA*.¹⁰ The reactions that were characterized so far all involve racemization or epimerization of 2-hydroxyacid substrates, however, there are possibilities of the NPN cofactor participating in other reactions. For example, a recent bioinformatics analysis reveals that genomes of cyanobacteria are especially noteworthy because they almost uniformly contain the NPN biosynthesis genes, but they lack any LarA homologs.²⁴ This consistent result suggests that these phototrophs might contain a non-LarA like enzyme that utilizes NPN cofactor.

This thesis seeks to fill some of the gaps in our knowledge of NPN biosynthesis and utilization. Chapter 2 focuses on LarC from *Moorella thermoacetica* (LarC_{Mt}), describing its purification and some of its properties, and testing a hypothesis for why LarC is a single-turnover enzyme. Most importantly, it establishes the function of CTP in the enzyme reaction mechanism. Chapter 3 extends our understanding of LarC_{Mt} by using cryo-electron microscopy to examine the structure of the intact protein and expanding our knowledge of the protein by using site-directed mutagenesis studies. Chapter 4 describes my efforts to create an expression system to allow the generation of NPN-containing enzymes in *Escherichia coli*. The goal of this undertaking is to be able to better examine the functions of LarA analogs and non-LarA NPN-binding proteins as well as to characterize the activities of LarB, LarC, and LarE homologs. Finally, Chapter 5 provides a summary of my results and offers a perspective of yet unanswered questions related to these topics.

REFERENCES

- (1) Desguin, B., Zhang, T., Soumillion, P., Hols, P., Hu, J., and Hausinger, R. P. (2015) A tethered niacin-derived pincer complex with a nickel-carbon bond in lactate racemase. *Science* **349**, 66-69
- (2) Tatum, E. L., Peterson, W. H., and Fred, E. B. (1936) Enzymic racemization of optically active lactic acid. *Biochem J* **30**, 1892-1897
- (3) Ribeiro, C., Santos, C., Gonçalves, V., Ramos, A., Afonso, C., and Tiritan, M. (2018) Chiral Drug Analysis in Forensic Chemistry: An Overview. *Molecules* **23**, 262
- (4) Cuesta, S. M., Rahman, S. A., and Thornton, J. M. (2016) Exploring the chemistry and evolution of the isomerases. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 1796-1801
- (5) Goffin, P., Deghorain, M., Mainardi, J. L., Tytgat, I., Champomier-Verges, M. C., Kleerebezem, M., and Hols, P. (2005) Lactate racemization as a rescue pathway for supplying D-lactate to the cell wall biosynthesis machinery in *Lactobacillus plantarum*. *J Bacteriol* **187**, 6750-6761
- (6) Ferain, T., Hobbs, J. N., Jr., Richardson, J., Bernard, N., Garmyn, D., Hols, P., Allen, N. E., and Delcour, J. (1996) Knockout of the two *ldh* genes has a major impact on peptidoglycan precursor synthesis in *Lactobacillus plantarum*. *J Bacteriol* **178**, 5431-5437
- (7) Nevarez, J. L., Turmo, A., Hu, J., and Hausinger, R. P. (2020) Biological pincer complexes. *Chemcatchem* **12**, 4242-4254
- (8) Singleton, J. T. (2003) The uses of pincer complexes in organic synthesis. *Tetrahedron* **59**, 1837-1857
- (9) van Koten G., M. D. (2013) *Organometallic Pincer Chemistry*, Springer, Germany
- (10) Desguin, B., Urdiain-Arraiza, J., Da Costa, M., Fellner, M., Hu, J., Hausinger, R. P., Desmet, T., Hols, P., and Soumillion, P. (2020) Uncovering a superfamily of nickel-dependent hydroxyacid racemases and epimerases. *Sci Rep* **10**, 18123
- (11) Desguin, B., Soumillion, P., Hols, P., and Hausinger, R. P. (2016) Nickel-pincer cofactor biosynthesis involves LarB-catalyzed pyridinium carboxylation and LarE-dependent sacrificial sulfur insertion. *Proc Natl Acad Sci USA* **113**, 5598-5603
- (12) Desguin, B., Fellner, M., Riant, O., Hu, J., Hausinger, R. P., Hols, P., and Soumillion, P. (2018) Biosynthesis of the nickel-pincer nucleotide cofactor of lactate racemase requires a CTP-dependent cyclometallase. *J Biol Chem* **293**, 12303-12317

- (13) Rankin, J. A., Mauban, R. C., Fellner, M., Desguin, B., McCracken, J., Hu, J., Varganov, S. A., and Hausinger, R. P. (2018) Lactate racemase nickel-pincer cofactor operates by a proton-coupled hydride transfer mechanism. *Biochemistry* **57**, 3244-3251
- (14) Fellner, M., Rankin, J. A., Desguin, B., Hu, J., and Hausinger, R. P. (2018) Analysis of the active site cysteine residue of the sacrificial sulfur insertase LarE from *Lactobacillus plantarum*. *Biochemistry* **57**, 5513-5523
- (15) Rankin, J. A., Chatterjee, S., Tariq, Z., Lagishetty, S., Desguin, B., Hu, J., and Hausinger, R. P. (2021) The LarB carboxylase/hydrolase forms a transient cysteinyl-pyridine intermediate during nickel-pincer nucleotide cofactor biosynthesis. *Proc Natl Acad Sci U S A* **118**, e2106202118
- (16) Desguin, B., Goffin, P., Viaene, E., Kleerebezem, M., Martin-Diaconescu, V., Maroney, M. J., Declercq, J. P., Soumilion, P., and Hols, P. (2014) Lactate racemase is a nickel-dependent enzyme activated by a widespread maturation system. *Nat Commun* **5**, 3615
- (17) Fellner, M., Desguin, B., Hausinger, R. P., and Hu, J. (2017) Structural insights into the catalytic mechanism of a sacrificial sulfur insertase of the N-type ATP pyrophosphatase family, LarE. *Proc Natl Acad Sci USA* **114**, 9074-9079
- (18) Chatterjee, A., Abeydeera, N. D., Bale, S., Pai, P.-J., Dorrestein, P. C., Russell, D. H., Ealick, S. E., and Begley, T. P. (2011) *Saccharomyces cerevisiae* THI4p is a suicide thiamine thiazole synthase. *Nature* **478**, 542-546
- (19) Chatterjee, S., Parson, K. F., Ruotolo, B. T., Mccracken, J., Hu, J., and Hausinger, R. P. (2022) Characterization of a [4Fe-4S]-dependent LarE sulfur insertase that facilitates nickel-pincer nucleotide cofactor biosynthesis in *Thermotoga maritima*. *J Biol Chem* **298**, 102131
- (20) Dehand, J., and Pfeffer, M. (1976) Cyclometallated compounds. *Coord Chem Rev* **18**, 327-352
- (21) Albrecht, M. (2010) Cyclometalation using d-block transition metals: fundamental aspects and recent trends. *Chem Rev* **110**, 576-623
- (22) Klein, A., Sandleben, A., and Vogt, N. (2016) Synthesis, structure and reactivity of cyclometalated nickel(II) complexes: a review and perspective. *Proc Natl Acad Sci India Sect A Phys Sci* **86**, 533-549
- (23) Desguin, B., Soumilion, P., Hols, P., Hu, J., and Hausinger, R. P. (2017) Lactate Racemase and Its Niacin-Derived, Covalently-Tethered, Nickel Cofactor. in *The Biological Chemistry of Nickel* (Zamble, D., Rowińska-Żyrek, M., and Kozłowski, H. eds.), The Royal Society of Chemistry. pp 220-236

- (24) Chatterjee, S., Gatreddi, S., Gupta, S., Nevarez, J. L., Rankin, J. A., Turmo, A., Hu, J., and Hausinger, R. P. (2022) Unveiling the mechanisms and biosynthesis of a novel nickel-pincer enzyme. *Biochem Soc Trans* **50**, 1187-1196

CHAPTER 2

CHARACTERIZATION OF THE NICKEL-INSERTING CYCLOMETALLASE LARC FROM *MOORELLA THERMOACETICA* AND IDENTIFICATION OF A CYTIDINYLYLATED REACTION INTERMEDIATE

This chapter was adapted from text originally published in:

Turmo, A., Hu, J., and Hausinger, R.P. (2022) Characterization of the nickel-inserting cyclometallase LarC from *Moorella thermoacetica* and identification of a cytidinylylated reaction intermediate, *Metallomics* 14, 2022, mfac014

Copyright, © 2022 Oxford University Press — Reproduced with permission.

Introduction

The nickel-pincer nucleotide (NPN) is a recently discovered cofactor of lactate racemase that also functions in other racemase and epimerase reactions.^{1,2} This complex, with nickel coordinated to pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN), is covalently tethered by one thiocarboxylic acid to a lysyl residue in some, but not all, NPN-containing proteins. The biosynthesis of NPN is best characterized in *Lactobacillus plantarum* (Figure 2-1).

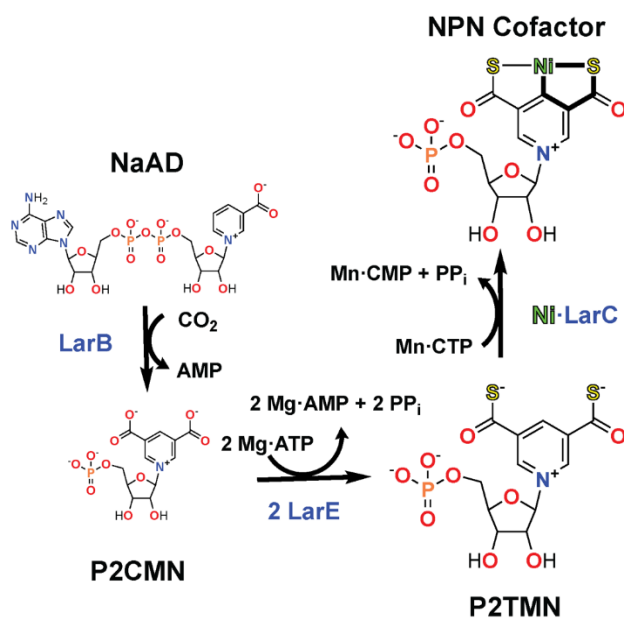


Figure 2-1 Biosynthesis and structure of the nickel-pincer nucleotide (NPN) cofactor. LarB catalyzes both pyridinium ring C5 carboxylation of nicotinic acid adenine dinucleotide (NaAD) and the hydrolysis of the phosphoanhydride, releasing AMP, to produce pyridinium-3,5-biscarboxylic acid (P2CMN). LarE uses ATP to activate the pyridinium ring carboxyl groups of P2CMN by adenylation, and then transfers a cysteinyl sulfur atom to this substrate to release AMP and produce dehydroalanine. Two molecules of LarE are needed to produce each molecule of pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN). LarC transfers a protein-bound nickel ion into P2TMN in a CTP-driven reaction producing the NPN cofactor. The metallacycle generated by this reaction is highlighted by bold lines.

The pathway initiates from nicotinic acid adenine dinucleotide (NaAD),³ with LarB catalyzing the addition of a second carboxyl group to the pyridinium ring and hydrolyzing the phosphoanhydride bond to form pyridinium-3,5-biscarboxylic acid mononucleotide (P2CMN).⁴ Two molecules of LarE sequentially catalyze ATP-dependent sacrificial sulfur insertion reactions, resulting in P2TMN.^{5,6} Finally, LarC completes the synthesis of the mature NPN cofactor by generating nickel-carbon and nickel-sulfur σ bonds in a CTP-dependent reaction.⁷

Inorganic chemists use the term cyclometallation to describe metal insertion reactions that form a metallacycle in which the

metal becomes coordinated to carbon and an electrophilic atom.⁸ LarC creates such a

metallacycle (indicated by the thicker lines in Figure 2-1) and represents the first cyclometallase identified in nature.⁷

The sequences of LarC proteins are not homologous to other proteins of known function.⁹ LarC of *L. plantarum* (LarC_{Lp}) is among the ~8% of LarC homologs that are encoded by two open reading frames, *larC1* and *larC2*, and separated by a programmed ribosomal frameshift (PRF).⁹ The PRF can be eliminated by gene fusion without compromising the activity of the enzyme. The N-terminal sequence (LarC1) contains a His-rich region that is presumed to bind nickel, and ~90% of LarC_{Lp} as purified from nickel-enriched growth medium is loaded with this metal ion.⁷ The full-length protein undergoes apparent proteolysis during crystallization, with only the C-terminal portion (LarC2) being crystallizable. The crystal structure of this protein fragment, a hexamer containing two domains, was solved at a resolution of 2.0 Å [protein database (PDB) ID: 6BWO].⁷ A full-length LarC protein structure has not yet been reported.

The reaction mechanism of this enzyme remains unclear, but several intriguing aspects of catalysis have been uncovered using LarC_{Lp}.⁷ Nickel incorporation into P2TMN requires the hydrolysis of CTP, forming CMP and presumably pyrophosphate (PP_i). Also required is the presence of Mg²⁺ or Mn²⁺, with a preference for the latter metal. Of great interest, the structure of LarC2•Mn•CTP was solved by soaking the protein fragment with this metal ion and nucleotide (PDB ID: 6BWQ).⁷ Surprisingly, LarC_{Lp} appears to be a single-turnover enzyme, with a single molecule of CTP undergoing hydrolysis for each molecule of NPN synthesized. Site-directed mutagenesis of the fused version of *larC* was used to replace several C-terminal domain residues involved in CTP binding, generally resulting in severe diminishment of LarC activity. Mutagenesis of *larC* was also used to delete the His-rich region or to substitute several acidic residues in the LarC1 region, demonstrating the importance of these components for the enzyme

activity. The combined results led to a proposal that LarC utilizes a carboxylate-associated mechanism for transferring nickel into P2TMN in a CTP-dependent manner.⁷

In this study, I characterized a LarC homolog from *Moorella thermoacetica* (LarC_{Mt}) encoded by a gene lacking an internal stop signal and not subject to a PRF. I demonstrated that LarC_{Mt} is more resistant to proteolysis when compared to LarC_{Lp}, and I characterized selected properties of the protein. For example, I ruled out the hypothesis that enzyme inhibition by its product PP_i accounts for its apparent single-turnover activity. Of greatest interest, I identified a cytidinylated (CMPylated)-substrate intermediate that is formed during the reaction of LarC_{Mt}. Selected variants with substitutions at the predicted CTP-binding site retained substantial activity, but they exhibited greatly reduced levels of the intermediate. In contrast, use of LarC_{Mt} from cells grown on medium without supplemental nickel led to enhanced amounts of the intermediate. On the basis of these results, I propose a functional role for CTP in the unprecedented nickel-insertase reaction during NPN biosynthesis.

Methods

Materials

Carbenicillin, kanamycin, chloramphenicol, and β-D-1-thiogalactopyranoside were purchased from Gold Bio (St. Louis, MO, USA). Desthiobiotin and NaAD were acquired from Sigma (St. Louis, MO, USA). All other chemicals used were reagent grade or better.

Genes, plasmids, and cloning

The gene encoding LarC_{Mt}, flanked by NdeI and XhoI restriction sites, was chemically synthesized (Integrated DNA Technologies, Coralville, IA, USA). The DNA fragment was inserted into the vector pLW01¹⁰ resulting in the production of LarC_{Mt} with an N-terminal His₆-tag followed by a tobacco etch virus protease cleavage site. Site-directed mutagenesis of *larC_{Mt}*

was carried out using the gap-repair method,¹¹ and the constructs were verified by Sanger sequencing (Azenta, South Plainfield, NJ, USA). The constructs were transformed into competent *Escherichia coli* BL21 (DE3) cells for gene expression and protein purification studies.¹² The strains, plasmids, and primers used in this study are provided in Table 2-1.

Table 2-1 Strains, plasmids and primers.

Strain, plasmid or primer	Characteristic(s) or sequence		Source or reference
Strains			
<i>Lc. Lactis</i>			
NZ3900	MG1363 derivative		¹²
<i>E. coli</i>			
DH5α	F- φ80lacZΔ M15 Δ (<i>lacZYA-argF</i>) U169 <i>recA1 endA1 hsdR17</i> (rK- mK+) <i>phoA supE44 λ- thi-1 gyrA96 relA1</i>		ThermoFisher
BL21 (DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS</i> <i>λ DE3 = λ sBamHI ΔEcoRI-B int:: (lacI::PlacUV5::T7 gene1)</i> <i>i21 Δnin5</i>		NEB
ArcticExpress	Contains <i>Cpn60</i> and <i>Cpn10</i> from <i>Oleispira antarctica</i>		Agilent
Plasmids			
pLW01	<i>Amp^r</i>		¹⁰
pLW01:LarCMT	<i>Amp^r</i> ;	LarC _{Mt} purification	This study
pET:LarBLp	<i>Km^r</i> ; pET28a with a 0.77-kb insert containing <i>larB</i> translationally fused to DNA encoding the StrepII-tag	LarB _{Lp} purification (Strain057-BL21)	⁴
pBAD:LarELp	<i>Amp^r</i> ; construct- <i>L. plantarum</i> LarE overexpression with strep tag (C-term)	LarE _{Lp} purification (Strain036-Arctic)	⁵
pGIR082	<i>Cm^r</i> ; pNZ8048 with a 1.31-kb insert after PnisA containing larATt translationally fused to DNA encoding the StrepII-tag	LarA _{Tt} purification	⁹
pET:LarCMT D256A	<i>Amp^r</i> ;	LarC _{Mt} D256A variant purification	This study
pET:LarCMT E261A	<i>Amp^r</i> ;	LarC _{Mt} E261A variant purification	This study
pET:LarCMT E364A	<i>Amp^r</i> ;	LarC _{Mt} E364A variant purification	This study
pET:LarCMT D256A E261A	<i>Amp^r</i> ;	LarC _{Mt} D256A E261A variant purification	This study
Primers			
LarC-MT NdeI fw	GATCCATATGAAGATCGCCTATTTTGAT		Subcloning
LarC-MT XhoI rv	GTCACCTCGAGATTAAAATGCTTTTCAGTGCACGTGC		Subcloning
LarC-MT E261A_fw	GATGATATGAACCCGGCGTTTTTCCGGCACTGCTGGA AGAAACC		Gap Repair

Table 2-1, (cont'd)

LarC-MT E261A_rv	CGCCGGGTTCATATCATCAATGGTGGTTTCAATAACCA GG	Gap Repair
LarC-MT D256A_fw	GGTTATTGAAACCACCATTGCGGATATGAACCCGGAAT TTTTCCGGC	Gap Repair
LarC-MT D256A_rv	CGCAATGGTGGTTTCAATAACCAGGCTGCTTTCTTCACC	Gap Repair
LarC-MT E364A_fw	GGTTATTACCAATATTGCACCGGCGTATGAAAGCTGTC G	Gap Repair
LarC-MT E364A_rv	CGCCGGTGCAATATTGGTAATAACCTGACCGGTCGGAT CACGATACAGACC	Gap Repair
LarC-MT D256A E261A fw	GCGGATATGAACCCGGcgTTTTTCCGGCACTGC	Gap Repair
LarC-MT D256A E261A rv	cgCCGGGTTCATATCCGCAATGGTGGTTTC	Gap Repair

Gene overexpression and protein purification

I grew *E. coli* BL21 (DE3) strains containing plasmids with wild-type and mutant *larC_{Mt}* in an autoinduction medium¹⁴ amended with 100 mg/L carbenicillin. The cultures were grown at 20°C while agitating at 220 RPM and, except where indicated, 1 mM NiCl₂ was added after 4 h of growth. Cells were harvested after ~20 h by centrifugation at 8000 rpm, resuspended in an equal volume of 100 mM Tris, pH 8.0, buffer containing 300 mM NaCl, and stored at –80°C until needed.

Thawed cells were lysed by use of a French pressure apparatus operating at 16,000 psi and 4 °C. The debris was removed by centrifugation (45 min at 115,955 ×g) at 4 °C. His-tagged LarC_{Mt} and its variants were purified using a His60 Ni Superflow resin by following the manufacturer's protocol (Takara Bio, San Jose, CA, USA). For native molecular weight determination, the sample was subjected to size exclusion chromatography (SEC) in 100 mM Tris-HCl buffer, pH 8.0, containing 300 mM NaCl on a Superdex 200 Increase 10/300 GL column (GE Healthcare, Chicago, IL, USA) while monitoring with miniDAWN TREOS multi-

angle light scattering (MALS) and TRex refractive index detectors (Wyatt, Santa Barbara, CA, USA). The data were analyzed with the ASTRA software (Wyatt).

Overexpression and purification of two other NPN biosynthesis proteins, LarB_{Lp} [from *E. coli* BL21 (DE2) cells] and LarE_{Lp} (expressed in *E. coli* ArcticExpress cells), were carried out as previously described.^{4,5} The lactate racemase apoprotein from *Thermoanaerobacterium thermosaccharolyticum*, LarA_{Ti}, was obtained from *Lactococcus lactis* NZ3900 cells containing pGIR082, following the previously reported purification protocol.⁹

The protein concentrations were determined either by the absorbance at 280 nm ($\epsilon_{280} = 23,740 \text{ M}^{-1} \text{ cm}^{-1}$ using the ExPASy protein parameter tool) or the Bradford protein assay reagent (Bio-Rad, Hercules, CA, USA) using bovine serum albumin as the standard.

LarC activity assay

The LarC_{Mt} substrate, P2TMN, was synthesized by incubating NaAD (0.2 mM) with LarB_{Lp} (10 μM) and NaHCO₃ (50 mM), to generate P2CMN, along with LarE_{Lp} (200 μM), ATP (2 mM), and MgCl₂ (20 mM) at room temperature for 1 h in 100 mM Tris-HCl buffer, pH 7.0.⁷

Conversion of P2TMN to NPN was achieved by incubation of an aliquot of the mixture mentioned earlier with an equal volume of CTP (0.2 mM), MgCl₂ (10 mM), β -mercaptoethanol (β -ME, 10 mM), and LarC_{Mt} or LarC_{Lp} (2.5 μM) in 100 mM 2-(*N*-morpholino)ethanesulfonic acid, pH 6.0, at room temperature for 30 min. Synthesis was terminated by heat treatment at 95 °C for 10 min. A 5- μl aliquot of the resulting NPN was mixed with LarA_{Ti} apoprotein (0.8 μM) and L-lactate (45 mM) in 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (100 mM), pH 7.0, for 5 min at 50°C, and the reaction was terminated by incubation in 95°C for 10 min. The lactate racemase activity was measured using a commercial kit (Neogen, Lansing, MI, USA) as previously described.⁹ Variations to these parameters were performed in specific experiments

that are described in the “Results and discussion” section. In some cases, P2TMN was purified by chromatography on a Q-Sepharose column (5 mL column volume) in 30 mM Tris-HCl buffer, pH 8, with elution using a NaCl gradient (to 1 M) and detection at 254 nm, analogous to a previously described procedure.⁷

To assess whether the apparent single-turnover reactivity of LarC_{Mt} is due to inhibition by PP_i, I tested the effect of adding 10 mM PP_i (Avantor, Radnor, PA, USA) and of providing 2 units of pyrophosphatase (PPase; Sigma) to the LarC assay reaction.

Mass spectrometric (MS) analysis

The intact LarC_{Mt} protein mass was determined using a Waters G2-XS Q-TOF (time of flight) mass spectrometer by injecting 10 µl of sample onto a Thermo Hypersil Gold CN guard column (1.0 × 10 mm) for desalting. A gradient of water + 0.1% formic acid (solvent A) and acetonitrile (solvent B) was run as follows at a flow rate of 0.1 ml/min: initial conditions were 98% A/2% B, hold at 2% B until 5 min with the flow diverted to waste for the first 3 min, ramp to 75% B at 10 min and hold at 75% B until 12 min, return to 2% B at 12.01 min and hold until 15 min. Mass spectra were obtained using electrospray ionization in positive ion mode with a source temperature of 100 °C, cone voltage of 35 V, desolvation temperature of 350 °C, desolvation gas flow of 600 L/h, cone gas flow of 50 L/h, and capillary voltage of 3.0 kV. Data were acquired using a 0.5 s TOF MS scan across an *m/z* range of 200–2000 and the spectra were deconvoluted in Masslynx using the Max Ent I algorithm.

LarC_{Mt} reaction samples were analysed using a Waters G2-XS Q-TOF mass spectrometer interfaced with a Waters Acquity UPLC. The 10-µl samples were injected onto a Waters Acquity UPLC BEH-C18 column (2.1 × 100 mm) that was held at 40 °C. Compounds were separated by ion-pairing chromatography using a binary gradient as follows: initial conditions were 100%

mobile phase A (10 mM tributylamine and 15 mM acetic acid in a 97:3 water/methanol (v/v) mixture) and 0% mobile phase B (methanol), hold at 100% A for 1 min, linear ramp to 99% B at 7 min, hold at 99% B to 8 min, return to 100% A at 8.01 min and hold until 10 min. The flow rate was 0.3 mL/min. Mass spectra were obtained by electrospray ionization operating in negative ion mode with a capillary voltage of 2.0 kV, source temperature of 100 °C, cone voltage at 35 V, desolvation temperature of 350 °C, desolvation gas flow of 600 L/h and cone gas flow of 50 L/h. Data were acquired using a data-independent MS^e method (scans with fast switching between no collision energy and using a collision energy ramp of 20–80 V) across an m/z range of 50–1500. Daughter ion spectra were acquired for $m/z = 715.02$ using an MS/MS method with selection in the quadrupole and fragmentation using a collision energy ramp of 10–60 V. Lockmass correction was performed in MassLynx software using leucine enkephalin as the reference compound.

Results and discussion

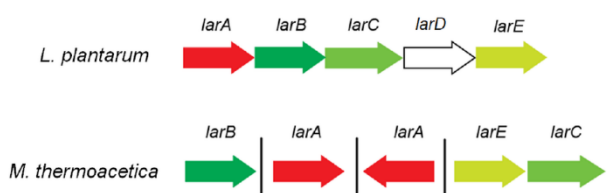


Figure 2-2 Comparison of the organization of genes related to NPN biosynthesis and use in *Lactobacillus plantarum* versus *Morella thermoacetica*. In *L. plantarum*, the single lactate racemase gene *larA* is clustered with *larB*, *larC*, and *larE* that encode enzymes for NPN biosynthesis. A non-essential lactate permease is encoded by *larD*. In *M. thermoacetica*, the *lar* genes are distributed at four sites in the genome, with only *larE* and *larC* grouped together and with two copies of *larA*-like genes of unknown function. The GeneBank accession numbers for the *L. plantarum* genes are WP_011100883.1 (*larA*), WP_011100884.1 (*larB*), WP_003641713.1 (*larC1*), WP_003641714.1 (*larC2*), WP_003643656.1 (*larD*), and WP_003641716.1 (*larE*), whereas those for *M. thermoacetica* are WP_011392116.1 (*larB*), WP_071541324.1 (*larA1*), WP_231114104.1 (*larA2*), WP_011393988.1 (*larE*), and WP_011393989.1 (*larC*).

Characterization of *M. thermoacetica* *LarC*

The genome of *M. thermoacetica* exhibits four widely dispersed sites of *lar* genes, with an isolated *larB*, two widely separated *larA* homologs of undefined roles, and *larE* grouped with *larC*; this organization contrasts with the situation in *L. plantarum*, where a single copy of *larA* encoding lactate racemase is located immediately adjacent to the three NPN

biosynthetic genes (Figure 2-2). LarC_{Mt} and LarC_{Lp} exhibit 38% sequence identity with only the latter protein containing a PRF (Figure 2-3).

Species	Protein	Length
Mt-LarC	MKIAYFDFCSGISGDMCLGALIACGLSQDELTSGLKELGLEGWELRVREVKQHSIAATDV	60
Lp-LarC	MQTLYLDAFSGISGDMFLGALLDLGLDFEQLKTELAKLHVHGYELTQQREAQSSIIYGTSTF	60
*: *:***** *****: **:::*: *: *:*** :. * ** .*		
Mt-LarC	AVQVTGSQ-----PFRHLADILGLINNSSLPAVPVKEKSAAVFKNL	100
Lp-LarC	DVQVAGGKDHGFVEHHHHQHEAGHHHDHEARHLADIEALIDGSDLSDTVKHHAKAIFMEI	120
:*. : ** .***:.* **::: ** :		
Mt-LarC	ARAEGQVHHGIDASQVHHFEVGAVDAAIIDIVGSILGLHLLGIEKVISSPLPAGSGWVDCRH	160
Lp-LarC	AQAEAAVHHMPLAEVHHFEVGAALDSIVDIVGCCIGLELMQIDTIMASPLSDGSGGFINVAAH	180
:. ** : :*****:***:***. :*. *: *:***:*** **:: *		
Mt-LarC	GKLPVPAPATLYLLQGYPVY--GTEDKAELVTPTGAALITTLADSFSGFPFAMNL-TRVG	216
Lp-LarC	GQMPVPVPVAVMQMRVGSAPIQQRLDVHTELITPTGMGLVKTLVREFGPLPENAVPTRVG	240
*:***.***: : * : :*:***. *.***. ***:* : ****		
Mt-LarC	FGAGKTELPHPNLLRLALGEINS--GQ--LEGEESSLVIETTDMDNPFHFFPALLEETMA	272
Lp-LarC	YGFGKRDGTGGFNALRAVLFEKKKLSQQIVNRTADAVLMIEANLDDQTGEGLGYVMNQLLT	300
: * ** : * ** . * * :. * . : *:***:*** . * : :*** :		
Mt-LarC	AGAVDAFFTPVQMKKGRPGILFALCPENKLPAAAAIFTSSSTLGLRFRDQRLVCQRR	332
Lp-LarC	AGAYDVFFTPIQMKKDRPATKLTVLGNVNDKDLLTKLILQETTIGVRYQTWQRTIMQRH	360
*** *.***:***.***. :*. * . : : *: :***:*** ** : **:		
Mt-LarC	MAEVVTPYGTVPVKLGLYRDPTGQVITNIAPFYESCRQIAKSAGVPVKEVYAAALAAARA	392
Lp-LarC	FLTVAATPYGDVQVKVATYQDI----EKKMPEYADCAQLAQQFHIPFRTVYQAALVAVDQ	415
: *.*** * **: . **: : *** . *:***. :*: ** **.*.		
Mt-LarC	LKAF-	396
Lp-LarC	LDEEA	420
*		

Figure 2-3 Comparison of the LarC sequences from *M. thermoacetica* (Mt-LarC) and *L. plantarum* (Lp-LarC). The sequences exhibit 38% identity with 163 identical positions. This alignment was created using Clustal-Omega.¹⁵ His residues are highlighted in cyan and residues that were substituted by mutation of the corresponding codons are highlighted in pink. A more extensive alignment of LarC sequences that highlighted CTP-binding residues was published previously.⁷

The endogenous His-rich region is shorter in LarC_{Mt} compared to LarC_{Lp} and it has a smaller overall abundance of His (10 versus 23 residues), especially in the N-terminus. Only five His residues are conserved between the two proteins. Homogeneity of the His₆-tagged protein was established by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Figure 2-4A) and confirmed by ESI-MS, which yielded $m/z = 44,202.5$ (Figure 2-4B), consistent with the

calculated mass of the nickel-free, full-length protein subunit (average $M_r = 44,204.83$).¹⁶ SEC

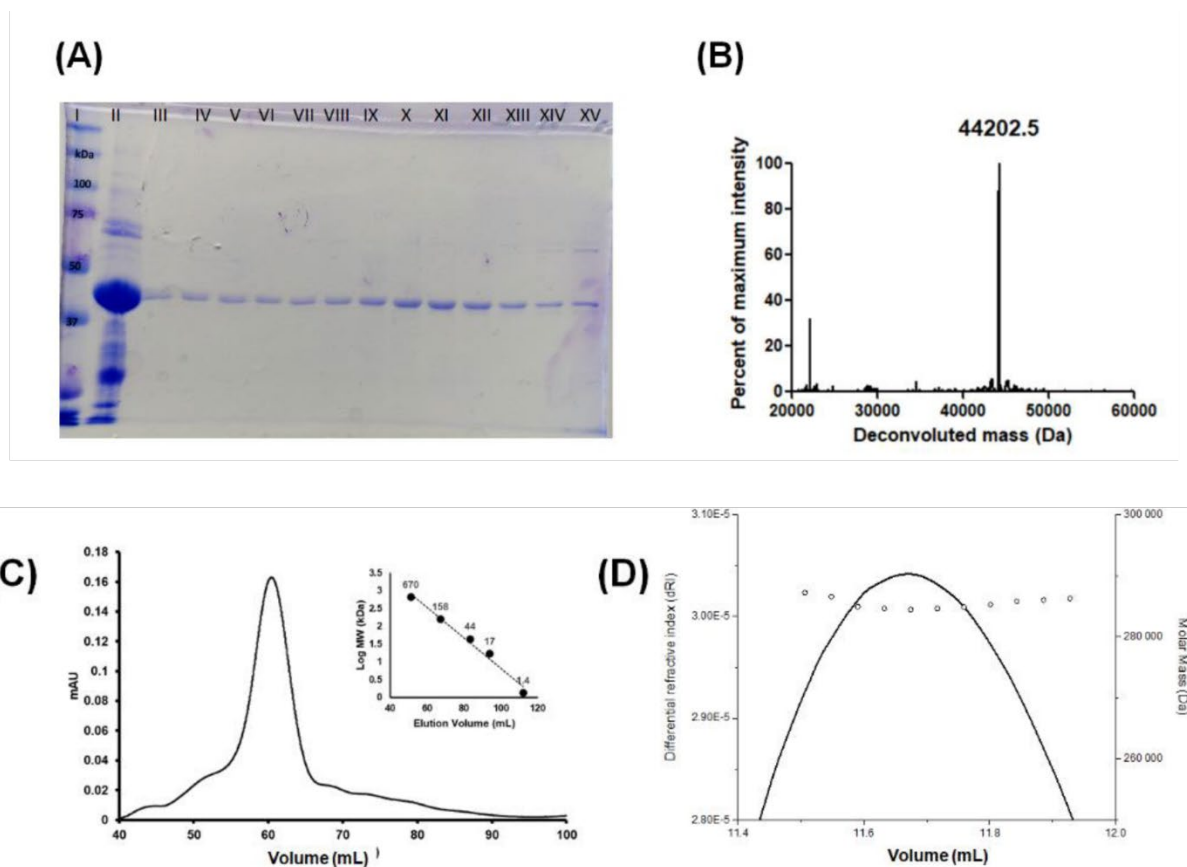


Figure 2-4 Homogeneity and size analyses of His₆-tagged LarC_{Mt}. (A) Sodium dodecyl sulfate polyacrylamide gel electrophoretic analysis of LarC_{Mt}. Lanes: (I) Precision Plus Protein Standard – Dual Color (Bio-Rad, Hercules, CA), (II) pooled elution fraction from the Ni-NTA column, (III-XV) fractions from the size exclusion chromatogram shown in panel (C). (B) Subunit mass determined by electrospray ionization-mass spectrometry. The expected monoisotopic weight is 44176.84 Da, and the expected average mass is 44204.83 Da. (C) Size exclusion chromatography while monitoring the absorbance at 280 nm. The inset depicts the elution positions for standards of the indicated sizes. The predominant peak of the sample corresponds approximately to $M_r = 321$ kDa. (D) SEC-MALS determination of molecular mass. The differential refractive and MALS-determined molecular mass determination are indicated.

comparison to standards provided an approximate $M_r = 321$ kDa (Figure 2-4C), whereas SEC-MALS analysis indicated a 6- or 7-mer oligomeric state of LarC_{Mt} in solution ($M_r = 287,500$) (Figure 2-4D).

The activity of LarC_{Mt} was assessed by an indirect assay that measured the lactate racemase activity of NPN-activated LarA_{Tt} apoprotein.⁷ The LarC_{Mt} substrate was generated from NaAD by the combined actions of LarB and LarE in the presence of CO₂/bicarbonate and Mg•ATP. Transformation of P2TMN into the NPN cofactor was achieved by LarC_{Mt} in the

presence of Mn•CTP. Incubation of the resulting NPN with LarA_{Tl} was followed by measurement of lactate racemase activity. The general enzymatic properties of LarC_{Mt} were similar to those of LarC_{Lp}, but the activity of the *M. thermoacetica* enzyme (see following text) exhibited only approximately 3% of the *L. plantarum* enzyme activity⁷ when assayed using the reported standard conditions for the latter enzyme.

Relationship of PP_i to the reactivity of LarC_{Mt}

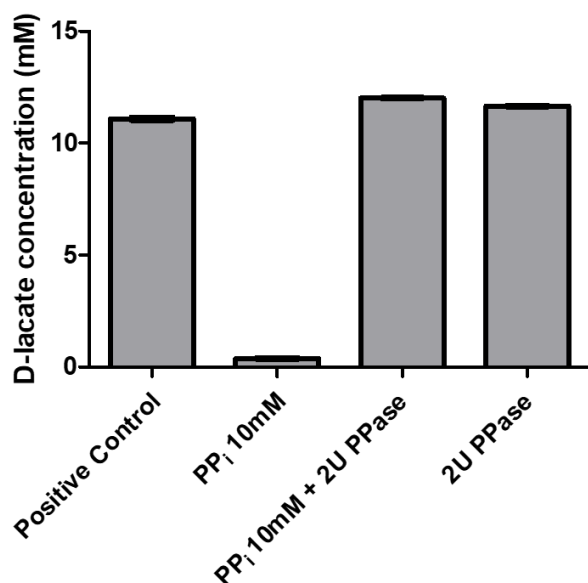


Figure 2-5 Pyrophosphate (PP_i) affects the activity of LarC_{Mt}. The nickel-inserting activity of LarC_{Mt} was assayed by an indirect assay that involved the activation of LarA_{Tl} apoprotein by enzymatically produced NPN and subsequent measurement of the conversion of L-lactate into D-lactate. The LarC_{Mt} substrate, pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN), was generated by the combined action of LarB_{Lp} and LarE_{Lp} acting on NaAD. Shown are the control reaction along with the effects of including 10 mM PP_i, both PP_i and pyrophosphatase (PPase), and only PPase (n = 2 biological replicates using separate enzyme preparations).

To assess whether the enzyme reaction product PP_i leads to the low activity of LarC_{Mt} and to investigate whether such inhibition accounts for the previously described apparent single-turnover reactivity of the enzyme,⁷ I examined the effects of adding PP_i or introducing pyrophosphatase (PPase) to the assay (Figure 2-5). PP_i was demonstrated to be a potent inhibitor of LarC_{Mt} activity, but full activity was restored by the introduction of 2 units of PPase. Inclusion of PPase in a sample lacking PP_i, however, did not lead to greater levels of activity. Thus, while PP_i inhibits LarC_{Mt}, it does not account for the apparent

single-turnover reactivity of this enzyme. As previously proposed,⁷ it is more likely that the stoichiometric nickel-insertion reaction is attributed to transfer of the inaccessible metal ion within the protein that cannot be replaced by adding nickel ions to the assay solution.

Identification of a P2TMN-CMP reaction intermediate

Time-dependent MS analysis of metabolites associated with the LarC_{Mt} reaction revealed the expected decrease in P2TMN levels as NPN was synthesized (demonstrated by incorporating the cofactor into LarA_{Tl} and measuring the lactate racemase activity, Figure 2-6A).

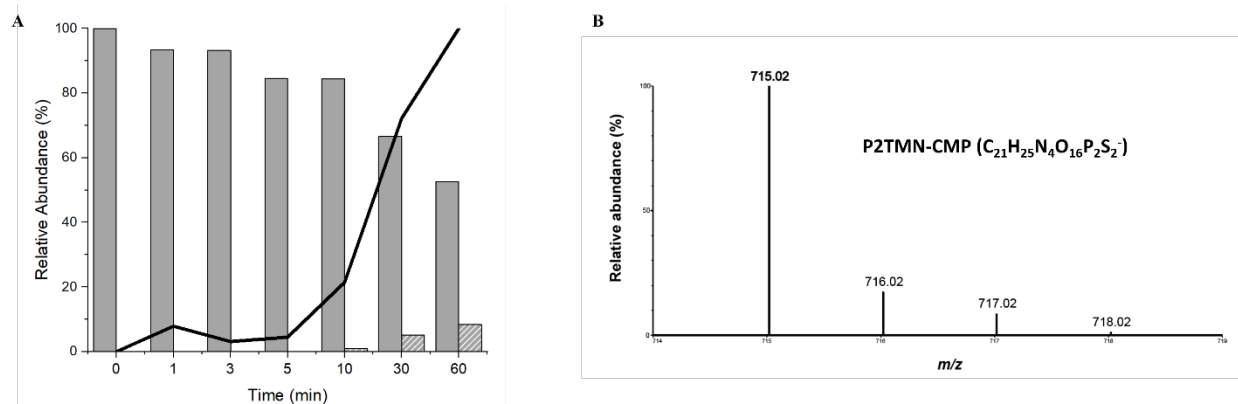


Figure 2-6. LarC_{Mt} forms a reaction intermediate. (A) The LarC_{Mt} reaction time course reveals a decrease in the concentration of pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN; solid gray bars), an increase in NPN synthesis (monitored by the ability to activate LarA_{Tl} to generate lactate racemase activity, line), and an increase in another metabolite (striped gray bars). The relative abundance of P2TMN and the novel metabolite are based on the intensities of the mass spectrometric (MS) peaks relative to that of P2TMN in the zero-time sample, which did not contain the other metabolite. (B) MS analysis of the intermediate species provides $m/z = 715.02$ and the relative intensities of the isotopic species indicate the presence of two sulfur atoms, consistent with a P2TMN-CMP linkage. These representative data are from a single experiment, but a replicate with another enzyme preparation showed the same trends.

The concentrations of synthesized P2TMN are not known because a standard is not available, so only the relative abundances are indicated for the representative data shown. Although NPN was able to be detected, the very weak intensity of the feature associated with the cofactor prevented its quantification using these conditions. Notably, an unidentified species ($m/z = 715.02$) was shown to be generated by LarC_{Mt} as P2TMN was consumed (Figure 2-6A and B). The relative abundance of this new metabolite was based on comparison of its peak intensities to that of P2TMN at zero time; no metabolite was detected at the initial time point. Significantly, the novel species was not formed in the absence of CTP. The mass of this species is consistent with that of P2TMN linked to CMP. MS–MS fragmentation analysis of this species supported a structure in which the substrate forms a phosphoanhydride bond with the nucleotide (Figure 2-7), i.e., a

CMPylated P2TMN. In particular, a species with an m/z of 402.01 is consistent with a phosphoanhydride-containing molecule. The proposed formation and decay of this species

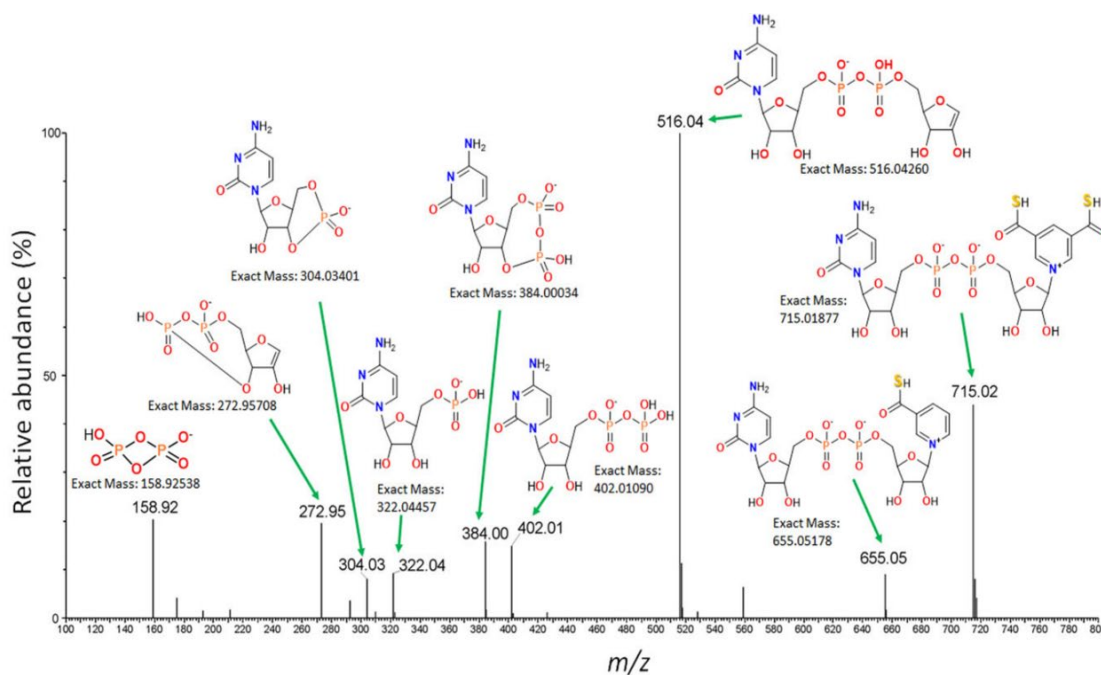


Figure 2-7 The mass spectrometric (MS)/MS fragmentation spectrum of the LarC_{Mt} intermediate. It is consistent with cytidinylated (CMPylated) pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN).

(Figure 2-8) is reminiscent of an intermediate formed by molybdenum insertase (Cnx1) during the synthesis of the molybdenum cofactor (Moco, Figure 2-9).

In that pathway, molybdopterin is thought to be adenylylated to properly position its dithiolene

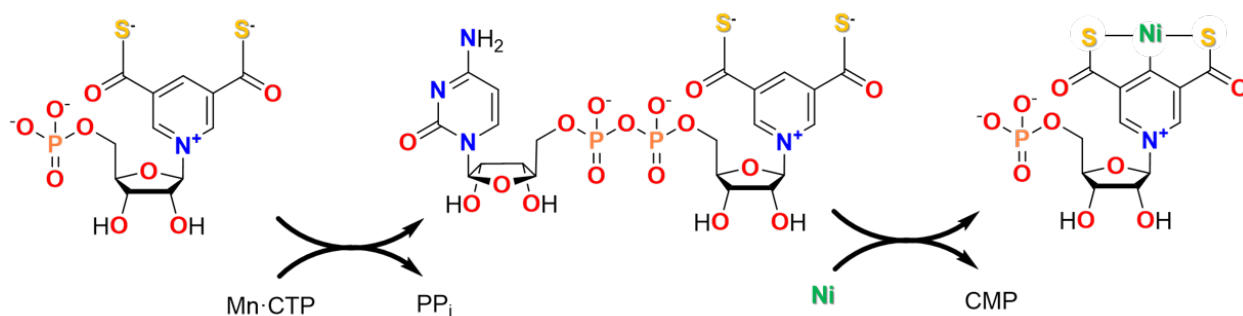


Figure 2-8 Two-step reaction of LarC. CTP-dependent CMPylation of pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN), with release of PP_i, is proposed to position the pyridinium ring in a proper orientation to allow nickel ion transfer accompanied by phosphoanhydride hydrolysis by LarC_{Mt}.

moiety near the molybdenum-binding site, followed by molybdate insertion and

phosphoanhydride hydrolysis to release AMP and the Moco.¹⁷⁻²⁰ I speculate that a similar

process occurs during NPN cofactor biosynthesis, i.e. P2TMN undergoes CMPylation to assist in

orienting the pyridinium ring near the buried nickel-binding site followed by metal insertion and

phosphoanhydride cleavage; however, additional studies (such as structure determination of the

LarC_{Mt}•CMP-P2TMN complex) are required to verify this hypothesis. Of additional interest, the

Moco-forming enzyme is inhibited by PP_i,¹⁷ as shown above for LarC_{Mt}.

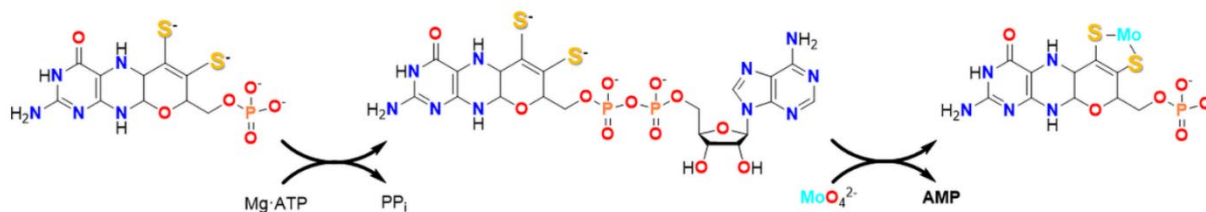


Figure 2-9 ATP-dependent AMPylation of MPT, with release of PP_i. The adenylylated intermediate is proposed to position the dithiolene moiety in a proper orientation to allow molybdate insertion accompanied by phosphoanhydride hydrolysis by *Arabidopsis thaliana* Cnx1.

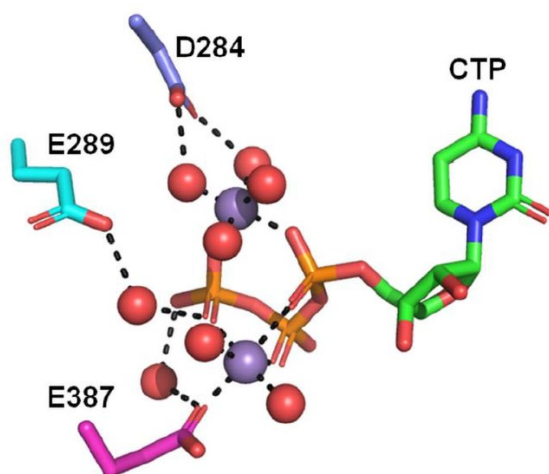


Figure 2-10 Residues predicted to be at the CTP-binding site of LarC_{Mt} that may participate in ATP hydrolysis. Acidic residues associated with the Mn·CTP in LarC_{Mt} are based on the LarC₂·Mn·CTP structure of LarC_{Lp} (PDB ID: 6BWQ). The LarC_{Lp} residues corresponding to D256 and E364 (D284 and E387, respectively) were previously substituted by Ala, resulting in proteins with approximately 10% and 0% of wild-type activity.

The D256A, E261A, E364A, and D256A/E261A variants of LarC_{Mt} were created and characterized for two reasons. First, the three side chains correspond to residues (Asp284, Glu289, and Glu387, Figure 2-3) that coordinate the manganese within the LarC_{Lp} CTP-binding domain (PDB ID: 6BWQ).⁷ D256A and E364A variants of LarC_{Mt} were created to confirm their importance for activity, based on the significant effects of D284A and D387A variants of LarC_{Lp} (~10% active and inactive,

respectively).⁷ No variant of Glu289 was examined for LarC_{Lp} because that position is not conserved (e.g. Gln was noted in some LarC sequences). Second, these residues were targets for mutagenesis because their predicted positions in LarC_{Mt} (Figure 2-10) were appropriate for facilitating hydrolysis of the intermediate. Thus, I wondered whether their substitution by alanine might increase the production of CMPylated-P2TMN.

The activities of the LarC_{Mt} variants were compared to that of the wild-type enzyme, and to two of the corresponding LarC_{Lp} variants, by using the indirect lactate racemase-based assay (Figure 2-11A). I found the D256A variant of LarC_{Mt} exhibited about 50% of the wild-type enzyme activity, compared to the 90% activity loss for the D284A variant of LarC_{Lp}. The E261A variant of LarC_{Mt} retained even greater levels of activity. Surprisingly, the E364A variant of LarC_{Mt} exhibited near wild-type activity levels, whereas the corresponding E387A variant of LarC_{Lp} was inactive. The basis of this difference is unclear, but may relate to protein folding issues, functional redundancy by another residue in close proximity, or other effects.

The E261A variant of LarC_{Mt} was slightly more active than the D256A variant. The reductions in activity by the D256A and E261A variants were approximately additive for the D256A/E261A double variant. MS analysis of the metabolites associated with these reactions revealed substantial reduction in the amount of P2TMN for the wild-type enzyme, clear decreases of P2TMN for the more active variant enzymes, and less utilization of the substrate by the

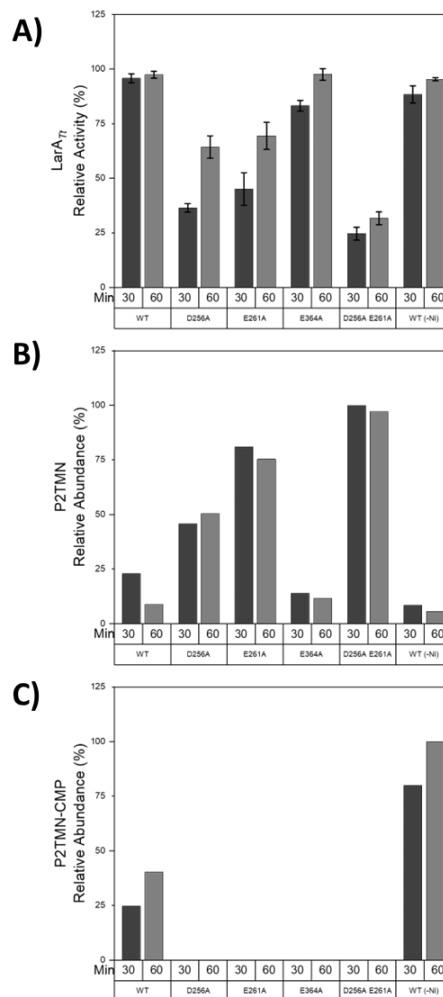


Figure 2-11 LarC_{Mt} variant analysis and effect of nickel limitation. Relative levels of (A) lactate racemase activity, (B) remaining pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN) substrate, and (C) CMP-P2TMN intermediate after 30 or 60 min of incubation of LarC_{Mt} samples with P2TMN and Mn•CTP. Samples included wild-type LarC_{Mt} and its D256A, E261A, E364A, and D256A/E261A variants that were purified from cultures supplemented with 1 mM NiCl₂ as well as wild-type LarC_{Mt} that was isolated from cells grown in medium without supplemental nickel ions. The samples were incubated with enzymatically produced P2TMN for the times indicated, the relative abundance levels of P2TMN and CMP-P2TMN were quantified by comparing the intensities of these mass spectrometric (MS) peaks to that of P2TMN at zero time (n = 2, technical replicates), and the products were mixed with LarA₇ apoprotein and the resulting lactate racemase activities were determined (n = 1).

double variant (Figure 2-11B), as expected. Notably, all of the variant proteins showed insignificant levels of the intermediate $m/z = 715.02$ species (Figure 2-11C). These results suggest that the decreases in variant enzyme activities are primarily associated with reduced rates of synthesis of the CMP-P2TMN intermediate while not affecting the hydrolysis of this species. Also shown in Figure 2-11 are the activity and metabolite level results obtained using LarC_{Mt} that was purified from cells grown in medium without supplemental nickel addition. This form of the enzyme was active, demonstrating the ability of the enzyme to sequester trace levels of nickel ions from the medium during growth. Accordingly, the relative levels of P2TMN exhibited substantial decreases over time. Significantly, the relative level of the CMP-P2TMN intermediate was greater than that associated with enzyme purified from cells grown with excess nickel ions. These results suggest, but still require further verification, that the intermediate is generated prior to nickel insertion, and that limited nickel levels increase the amount of the intermediate.

Conclusions

The gene encoding LarC_{Mt} was expressed, the protein was purified, and several of its properties were determined. The enzyme converts P2TMN to NPN, but is inhibited by the product of the reaction, PP_i. This PP_i inhibition does not account for the apparent single-turnover reaction kinetics of the enzyme. I identified a novel intermediate in which the precursor, P2TMN is CMPylated. Substitution of residues that are predicted to be positioned at the Mn•CTP-binding site resulted in only partial reduction of LarC_{Mt} activity, but a significant reduction in the formation of CMP-P2TMN, indicating the rate-determining step of NPN synthesis is associated with formation of the intermediate. By contrast, enhanced levels of CMP-P2TMN are produced when nickel ions are limiting, consistent with the metal ion binding to the CMPylated

intermediate. My discovery of the CMP-P2TMN reaction intermediate provides insight into the mechanism of the nickel insertion reaction and clarifies the role of CTP in this reaction.

REFERENCES

- (1) Desguin, B., Zhang, T., Soumillion, P., Hols, P., Hu, J., and Hausinger, R. P. (2015) A tethered niacin-derived pincer complex with a nickel-carbon bond in lactate racemase. *Science* **349**, 66-69
- (2) Desguin, B., Urdiain-Arraiza, J., Da Costa, M., Fellner, M., Hu, J., Hausinger, R. P., Desmet, T., Hols, P., and Soumillion, P. (2020) Uncovering a superfamily of nickel-dependent hydroxyacid racemases and epimerases. *Sci Rep* **10**, 18123
- (3) Desguin, B., Soumillion, P., Hols, P., and Hausinger, R. P. (2016) Nickel-pincer cofactor biosynthesis involves LarB-catalyzed pyridinium carboxylation and LarE-dependent sacrificial sulfur insertion. *Proc Natl Acad Sci USA* **113**, 5598-5603
- (4) Rankin, J. A., Chatterjee, S., Tariq, Z., Lagishetty, S., Desguin, B., Hu, J., and Hausinger, R. P. (2021) The LarB carboxylase/hydrolase forms a transient cysteinyl-pyridine intermediate during nickel-pincer nucleotide cofactor biosynthesis. *Proc Natl Acad Sci U S A* **118**, e2106202118
- (5) Fellner, M., Desguin, B., Hausinger, R. P., and Hu, J. (2017) Structural insights into the catalytic mechanism of a sacrificial sulfur insertase of the N-type ATP pyrophosphatase family, LarE. *Proc Natl Acad Sci USA* **114**, 9074-9079
- (6) Fellner, M., Rankin, J. A., Desguin, B., Hu, J., and Hausinger, R. P. (2018) Analysis of the active site cysteine residue of the sacrificial sulfur insertase LarE from *Lactobacillus plantarum*. *Biochemistry* **57**, 5513-5523
- (7) Desguin, B., Fellner, M., Riant, O., Hu, J., Hausinger, R. P., Hols, P., and Soumillion, P. (2018) Biosynthesis of the nickel-pincer nucleotide cofactor of lactate racemase requires a CTP-dependent cyclometallase. *J Biol Chem* **293**, 12303-12317
- (8) Albrecht, M. (2010) Cyclometallation using d-block transition metals: fundamental aspects and recent trends. *Chem Rev* **110**, 576-623
- (9) Desguin, B., Goffin, P., Viaene, E., Kleerebezem, M., Martin-Diaconescu, V., Maroney, M. J., Declercq, J. P., Soumillion, P., and Hols, P. (2014) Lactate racemase is a nickel-dependent enzyme activated by a widespread maturation system. *Nat Commun* **5**, 3615
- (10) Bridges, A., Gruenke, L., Chang, Y. T., Vakser, I. A., Loew, G., and Waskell, L. (1998) Identification of the binding site on cytochrome P450 2B₄ for cytochrome b₅ and cytochrome P450 reductase. *J Biol Chem* **273**, 17036-17049
- (11) Ruyter, P. G. d., Kuipers, O. P., and Vos, W. M. d. (1996) Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. *Appl Environ Microbiol* **62**, 3662-3667

- (12) García-Nafria, J., Watson, J. F., and Greger, I. H. (2016) IVA cloning: A single-tube universal cloning system exploiting bacterial *in vivo* assembly. *Sci Rep* **6**, 27459
- (13) Hanahan, D., Jessee, J., and Bloom, F. R. (1991) Plasmid transformation of *Escherichia coli* and other bacteria *Meth Enzymol* **204**, 63-113
- (14) Studier, F. W. (2014) Stable expression clones and auto-induction for protein production in *E. coli*. *Methods Mol Biol* **1091**, 17-32
- (15) Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-W641
- (16) Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. in *The Proteomics Protocols Handbook*, Humana Press. pp 571-607
- (17) Llamas, A., Mendel, R. R., and Schwarz, G. (2004) Synthesis of adenylated molybdopterin. *J Biol Chem* **279**, 55241-55246
- (18) Krausze, J., Hercher, T. W., Zwerschke, D., Kirk, M. L., Blankenfeldt, W., Mendel, R. R., and Kruse, T. (2018) The functional principle of eukaryotic molybdenum insertases. *Biochem J* **475**, 1739-1753
- (19) Probst, C., Yang, J., Krausze, J., Hercher, T. W., Richers, C. P., Spatzal, T., Kc, K., Giles, L. J., Rees, D. C., Mendel, R. R., Kirk, M. L., and Kruse, T. (2021) Mechanism of molybdate insertion into pterin-based molybdenum cofactors. *Nat Chem* **13**, 758-765
- (20) Hercher, T. W., Krausze, J., Hoffmeister, S., Zwerschke, D., Lindel, T., Blankenfeldt, W., Mendel, R. R., and Kruse, T. (2020) Insights into the Cnx1E catalyzed MPT-AMP hydrolysis. *Biosci Rep* **40**

CHAPTER 3

PRELIMINARY STRUCTURAL STUDY OF *MOORELLA THERMOACETICA* LARC

Introduction

The structure of the C-terminal domain of LarC (LarC2) from *L. plantarum* was solved previously, revealing a novel CTP binding pocket.¹ Mutagenesis analysis suggested that the N-terminal portion of the protein is responsible for binding of nickel and the substrate, P2TMN.¹ Support for a possible carboxylate-assisted mechanism of nickel transfer/cyclometallation was shown through residue substitution experiments. Specifically, the replacement of highly conserved acidic residues resulted in the loss of LarC activity, leading to the speculation that these carboxylate residues are involved in enzyme function.¹

Due to its susceptibility to proteolysis during extended incubation of the protein while attempting crystallization, the full-length structure of *L. plantarum* LarC (LarC_{Lp}) protein has yet to be solved. The protein cleavage occurs in the region where a programmed ribosomal frameshift (PRF) allows the ribosome to bypass a stop codon to yield a larger fusion protein that is needed for activity. A PRF is present at this position in 8% of the LarC homologs and results in two isoforms of this protein, but most genomes encoding LarC produce only the longer form of the protein and lack the “slip site” to introduce the internal stop codon during transcription.² An engineered transcriptional fusion version of *L. plantarum* *larC* was created to remove the PRF sequence and avoid synthesis of the non-functional truncated species; however, the engineered fusion protein did not eliminate the proteolysis issue. A separate attempt to obtain the structure of the truncated N-terminal domain of LarC (LarC1) was unsuccessful due to aggregation of the purified protein.

I used two approaches to attempt to solve the full-length structure of LarC. One tactic was to introduce residue modification at the proteolysis sites of LarC_{Lp} in hopes of preventing the cleavage reaction. The other effort was to study homologs of LarC that don't have a PRF and

are less susceptible to proteolysis. For both approaches, I could attempt crystallography of the protein, but I also investigated the use of cryo-electron microscopy (cryo-EM) for structure determination. In addition to these experimental efforts, I used computational tools to create and investigate a full-length structural model of LarC with a special interest in identifying plausible substrate-binding sites.

Methods

Gene, plasmids, and cloning

Bacterial strains, plasmids and primers used for this study are listed in Table 3-1. The primers used were purchased from IDT (Newark, NJ, USA). Site-directed mutagenesis was done following the *in vivo* assembly method.^{3,4} Subcloning was performed using the NdeI and XhoI cut sites as mentioned in the Chapter 2 methods section.²² The bacterial cell transformations were performed using chemically competent *E. coli* DH5 α or BL21 (DE3) cells following a standard protocol.⁶

Table 3-1 Strains, plasmids and primers

Strain, plasmid or primer	Characteristic(s) or sequence		Source or reference
Strains			
<i>E. coli</i>			
DH5 α	F- ϕ 80lacZ Δ M15 Δ (<i>lacZYA-argF</i>) U169 <i>recA1 endA1 hsdR17</i> (rK- mK+) <i>phoA supE44 λ- thi-1 gyrA96 relA1</i>		ThermoFisher
BL21 (DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS</i> <i>λ DE3 = λ sBamHIo ΔEcoRI-B int::(<i>lacI</i>::<i>PlacUV5</i>::T7 <i>gene1</i>) i21 Δnin5</i>		NEB
Plasmids			
pET22b:fused_LarCLp			⁷
pET22b:fused_LarCLp alanine	<i>Amp^r</i> ;	S277A, Q278A, Q279A, R283A substitution for <i>larC fused Lp</i>	This study
pLW01	<i>Amp^r</i>		⁸
pLW01:LarCMT	<i>Amp^r</i> ;	LarC _{MT} purification	This study
Primers			
LP_LarC2_A_fw	CGTAAATGCAACGGCTGATGCTGTCTTA		Gap Repair
LP_LarC2_A_rv	ATGGCGGCAGCTAGTTTCTTTTTCGAATAATACGG		Gap Repair
LarC-MT NdeI_fw	GATCCATATGAAGATCGCCTATTTTGAT		Subcloning
LarC-MT XhoI_rv	GTCACCTCGAGATTAATGCTTTTCAGTGACGTGC		Subcloning

Gene overexpression and protein purification

E. coli BL21 (DE3) strains were used to overexpress the recombinant proteins in autoinduction medium.⁹ Cultures were grown at room temperature with shaking at 220 rpm. NiCl₂ (1 mM final concentration) was added to the medium 4 h after growth and the cultures were grown an additional 20 – 24 h. Harvested cells were pelleted and resuspended in 35 mL of 100 mM Tris, pH 8.0, buffer containing 300 mM NaCl (TBS 100/300 pH 8.0) and stored in -80 °C until further use.

To lyse the cells, 1 mM DTT, 1 mM lysozyme, cOmplete™ EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland) and 1 unit of benzonase were added and the mixture was incubated on ice for 30 min. The cells were passed through a French press apparatus twice at 16,000 psi. The lysate was centrifuged at 115,955 x g for 45 min at 4 °C. The His-tagged protein in the supernatant solution was purified using gravity His60 Ni Superflow resin following the manufacture's protocol (Takara Bio, San Jose, CA, USA).

Further preparation of the sample prior to analysis by cryo-EM included chromatography of the Ni-nitrilotriacetic acid (NTA) fraction on a Superdex 200 Increase 10/300 GL column (GE Healthcare, Chicago, IL, USA) in TBS 100/300 buffer at pH 8.0. The eluted fractions were concentrated using an Amicon concentrator (Sigma-Aldrich, St. Louis, MO, USA) with a 10-kDa molecular weight cutoff filter. Final protein concentrations were determined by using the Bradford protein assay reagent (Bio-Rad, Hercules, CA) with bovine serum albumin as the standard.

Methods for cryo-EM specimen preparation and data collection

The peak fractions containing LarC_{Mt} from the gel filtration column were collected and concentrated to 1.1 mg/mL. Cryo-EM grids were frozen using a Vitrobot Mark IV (Thermo

Fisher Scientific, Waltham, MA, USA) as follows: 3.5 μ l of protein samples was applied to a glow-discharged Quantifoil Cu 1.2/1.3 holey carbon 200 mesh grid (Quantifoil, Großlobichau, Thüringen, Germany), and the grid was blotted for 3.5 s prior to plunge freezing in liquid ethane. The cryo-EM images were collected on the Talos Arctica microscope (Thermo Fisher Scientific, Waltham, MA, USA) operated at 200 kV and equipped with a Falcon 3EC direct electron detector camera. 1,586 movies were collected in counting mode using EPU software at a nominal magnification of 92,000 x (corresponding to a calibrated pixel size of 1.12 Å/pixel), with a defocus range of -0.8 – 2.5 μ m. The total exposure dose of 35 $e^-/\text{Å}^2$ was fractionated into 42 frames.

Image processing

The cryo-EM movies were corrected for beam-induced motion by performing patch motion-correction in cryoSPARC (Structura Biotechnology Inc., Toronto, Canada). Contrast transfer function (CTF) parameters were determined by patch CTF estimation also in cryoSPARC. After removing micrographs of poor quality (bad CTF estimation, ice contamination, etc.), 1,253,520 particles were initially picked using blob picking followed by 2D classification in cryoSPARC. Particles belonging to bad 2D classes were discarded, and the remaining 1,148,996 particles were used to calculate an initial 3D map using *ab initio* reconstruction in cryoSPARC. Using the initial 3D map as a reference map, 3D classification was performed in Relion 3.0¹⁰ to isolate a particle population (421,288 particles) showing the highest resolution features. Final 3D refinement was performed in cryoSPARC, yielding a cryo-EM map of ~6 Å.

Alpha-Fold and HADDOCK

The AlphaFold predictions were obtained from the AlphaFold Protein Structure Database.¹¹ The ligand used for HADDOCK¹² analysis was downloaded from the RCSB Protein Data Bank (rcsb.org).¹³ The specific ligand used was the dithiodinicotinic acid mononucleotide from PDB: 5HUQ, ligand 4EY.

Results and Discussion

LarC_{Lp} alanine variants

In an effort to reduce the amount of proteolysis during extended incubations of wild-type LarC_{Lp} protein, I substituted residues that were previously identified by mass spectrometry¹ to form the cleavage site by replacing the corresponding codons to encode alanine. More specifically, I switched residues LSQQIVNRT (positions 266 to 272) to LAIVNAT. I purified the fused LarC_{Lp} variant and its alanine derivatives by using the standard NTA-resin approach. I stored the proteins at 4 °C for 2 weeks and assessed the amount of proteolysis by subjecting the samples to sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Unfortunately, I detected no diminishment in protein degradation for the variant forms of LarC_{Lp} (data not shown). Because this approach was not successful, I refocused my efforts to follow the lead of studying LarC homologs.

Analysis of LarC homologs

The genes encoding ten LarC homologs from various microorganisms were purchased, and Dexin Sui, a former lab member of Prof. Hu's laboratory constructed vectors for their expression in *E. coli*. I replicated several of his initial characterization efforts to identify the most tractable LarC system. The gene from *Synechocystis* was poorly expressed under all conditions tested, whereas that from *Clostridiodes difficile* produced high levels of the desired protein, but it

was cleaved near the carboxyl terminus. The genes from most other sources also were discarded for various reasons. The most suitable homolog was that from *Moorella thermoacetica* which showed a promisingly high expression level. Unfortunately, I noticed that after the SEC purification there was a concentration dependent aggregation of the protein when present in concentrations greater than 1 mg/mL. I attempted the crystallization of the protein, but these low concentration samples failed to provide promising results.

Preliminary Cryo-EM of LarC_{Mt}

Combining my observations that (i) the protein formed a possible octamer according to gel filtration analysis (Chapter 2, Figure 2-4) and (ii) that solubility required the maintenance of low concentrations of samples, I chose to pursue the path of using cryo-EM to solve the structure of this enzyme. In collaboration with the laboratory of Professor Kelly Kim at MSU, we were able to obtain preliminary data that provided interesting insights. Although the resolution of the structure is only $\sim 9 \text{ \AA}$, we could conclude from the density map compared to the predicted AlphaFold structure that the protein forms a dimer using these cryoEM conditions (Figure 3-1). This finding supports the notion of a concentration dependence on the oligomeric state of this enzyme.

The $\sim 100 \text{ kDa}$ size of the protein dimer is quite small for analysis by cryo-EM. Furthermore, there is an issue with the protein taking a preferred orientation on the grid, hindering the ability to obtain a set of images of protein in different views to get a better-quality image. Moving forward, different additives were used successfully to be used to address this

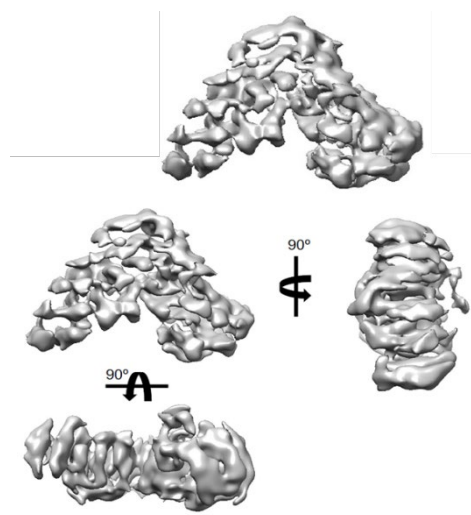


Figure 3-1 Low resolution cryo-EM density map of LarC_{Mt}.

issue. In particular, Dr. Kim and her colleague Dr. Robert Wolfe found that the addition of 0.05% n-dodecyl-B-D-maltoside (DDM) overcame the preferred orientation problem. He has further optimized the purification methods, resulting in the ability to obtain samples at 3.3 mg/mL. New structural data were obtained using cryo-EM resulting in a greatly improved structure prediction at a resolution of ~ 9 Å of what now appears to be a hexamer in agreement with what was observed for the LarC2 crystal structure. Dr. Wolfe will be continuing the structural analysis of this protein.

LarC_{Mt} structural prediction

The vast improvement in structural prediction using artificial intelligence has resulted in the development of the AlphaFold protein structure database.^{11,14} Using this resource, I retrieved the high confidence structural model of LarC from *Moorella thermoacetica*. From this model we can see the two distinct domains of this

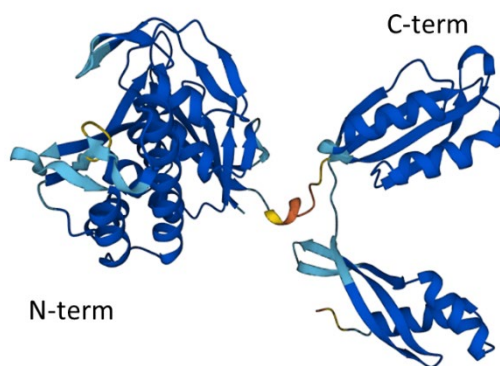


Figure 3-2 AlphaFold structure prediction of LarC_{Mt}. The color indicates the model confidence level; from navy being very high confidence to orange being very low confidence. Figure taken from EMBL-EBI AlphaFold protein structure database.

enzyme covering the N-terminal and C-terminal regions along with a low confidence linking connector (Figure 3-2). Notably the N-terminal domain with the histidine-rich region that is predicted to hold the nickel has a lower confidence level in the AlphaFold model, perhaps the flexibility is needed to accommodate the movement required to insert the nickel to the substrate to create the mature NPN cofactor.

Using the model obtained with AlphaFold, I docked the ligand the dithiodinicotinic acid mononucleotide (P2TMN missing an oxygen on one of the thiocarboxylate)¹⁵ using the HADDOCK server in their easy mode.¹² I restricted the docking site to fit the N-terminal domain containing the histidine-rich region that is predicted to coordinate the nickel. Figure 3-3 shows the predicted ligand binding pocket of the docking result generated using LigPlot.¹⁶ Residues Val62, Asp 128, Asp124, Thr190, and Thr192 are shown to have hydrogen bonds with the ligand. Notably, previous mutagenesis work showed that converting residue Asp124 to alanine resulted in diminishment of the enzyme activity.¹ This finding might indicate

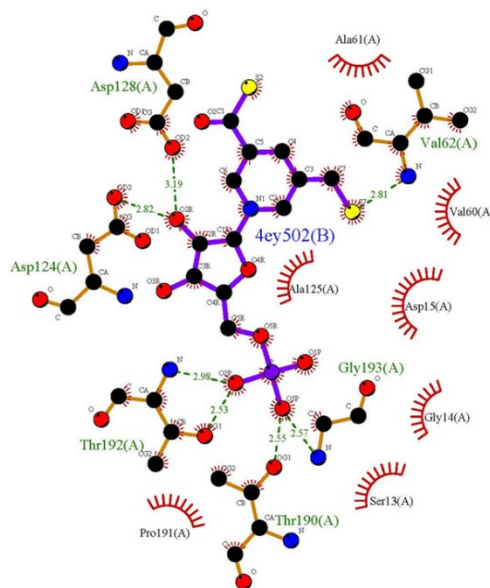


Figure 3-3 LarC residues interaction with bound ligand, 4YE. LigPlot of residues proximate to the docking result. Green dash indicated hydrogen bonds.

there is some significance to the docking result, however there are several issues that argue against this docking result having much weight. First, I do not know the specific active site of LarC, resulting in a broad area of the protein surface being included for the potential binding site during docking analysis. Secondly, the site of the predicted cofactor binding is quite far from the binding site of the substrate, CTP, requiring major conformational changes to allow for creating the CMP-P2TMN adduct reaction intermediate. Furthermore, the planar portion of the ligand into which the nickel is inserted faces out toward the solvent and not towards a buried nickel-

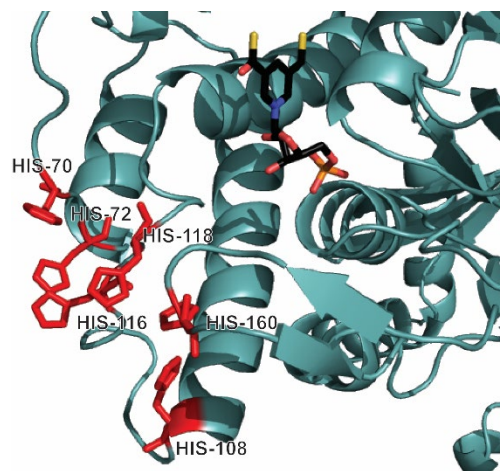


Figure 3-4 Histidine-rich region on the AlphaFold model of LarC_{Mt}. The histidines of interest are shown in red sticks with the residue numbers. The docked ligand, 4YE, is shown in black stick.

binding site that is presumed to involve the histidine-rich region. Finally, a cluster of histidine residues, a possible nickel binding site, is not located near the predicted position for the thiocarboxylate of the ligand (Figure 3-4). Multiple alternative docking programs are available that could be utilized to examine whether a consensus P2TMN docking site occurs, but all such programs are limited by working with only a model of the LarC1 portion of the enzyme instead of an experimentally

determined structure.

Overall, the most promising way to determine the substrate binding site will be to co-crystallize the protein with P2TMN or to determine the structure of the complex via cryo-EM.

REFERENCES

- (1) Desguin, B., Fellner, M., Riant, O., Hu, J., Hausinger, R. P., Hols, P., and Soumillion, P. (2018) Biosynthesis of the nickel-pincer nucleotide cofactor of lactate racemase requires a CTP-dependent cyclometallase. *J Biol Chem* **293**, 12303-12317
- (2) Caliskan, N., Peske, F., and Rodnina, M. V. (2015) Changed in translation: mRNA recoding by -1 programmed ribosomal frameshifting. *Trends Biochem Sci* **40**, 265-274
- (3) García-Nafria, J., Watson, J. F., and Greger, I. H. (2016) IVA cloning: A single-tube universal cloning system exploiting bacterial *in vivo* assembly. *Sci Rep* **6**, 27459
- (4) Huang, F., Spangler, J. R., and Huang, A. Y. (2017) *In vivo* cloning of up to 16 kb plasmids in *E. coli* is as simple as PCR. *PLOS ONE* **12**, e0183974
- (5) Turmo, A., Hu, J., and Hausinger, R. P. (2022) Characterization of the nickel-inserting cyclometallase LarC from *Moorella thermoacetica* and identification of a cytidinylated reaction intermediate. *Metallomics* **14**, 8
- (6) Hanahan, D., Jessee, J., and Bloom, F. R. (1991) Plasmid transformation of *Escherichia coli* and other bacteria *Meth Enzymol* **204**, 63-113
- (7) Desguin, B., Goffin, P., Viaene, E., Kleerebezem, M., Martin-Diaconescu, V., Maroney, M. J., Declercq, J. P., Soumillion, P., and Hols, P. (2014) Lactate racemase is a nickel-dependent enzyme activated by a widespread maturation system. *Nat Commun* **5**, 3615
- (8) Bridges, A., Gruenke, L., Chang, Y. T., Vakser, I. A., Loew, G., and Waskell, L. (1998) Identification of the binding site on cytochrome P450 2B4 for cytochrome b(5) and cytochrome P450 reductase. *J Biol Chem* **273**, 17036-17049
- (9) Studier, F. W. (2014) Stable expression clones and auto-induction for protein production in *E. coli*. *Methods Mol Biol* **1091**, 17-32
- (10) Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., and Scheres, S. H. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**
- (11) Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zidek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., and Velankar, S. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439-D444
- (12) van Zundert, G. C. P., Rodrigues, J., Trellet, M., Schmitz, C., Kastiris, P. L., Karaca, E., Melquiond, A. S. J., van Dijk, M., de Vries, S. J., and Bonvin, A. (2016) The

HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol* **428**, 720-725

- (13) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The protein data bank. *Nucleic Acids Res* **28**, 235-242
- (14) Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589
- (15) Desguin, B., Zhang, T., Soumillon, P., Hols, P., Hu, J., and Hausinger, R. P. (2015) A tethered niacin-derived pincer complex with a nickel-carbon bond in lactate racemase. *Science* **349**, 66-69
- (16) Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778-2786

CHAPTER 4

EXPANDING THE METHOD TO STUDY THE *LAR* GENES IN *ESCHERICHIA COLI*

Introduction

Bioinformatic analysis of over 1,000 eubacterial and archaeal genomes indicated that homologs of LarA and the nickel-pincer nucleotide (NPN) biosynthesis pathway enzymes are present in approximately 9% of this population.¹ Recent biochemical studies of seven out of 13 potential *larA* homologs showed that they often carry out a reaction distinct from lactate racemase (Lar); namely racemization or epimerization of other 2-hydroxyacid substrates such as malate, 2-hydroxyglutarate, and the sugar D-gluconate.² Moreover, ~15% of the genomes analyzed contain homologs of *larB*, *larE* and *larC*, but lack a *larA* gene homolog, suggesting that some microorganisms synthesize the NPN cofactor using the usual pathway, but then incorporate the molecule into a non-LarA NPN-binding protein.¹ Given the widespread appearance and diverse functionality of the LarA homologs and the likely presence of non-LarA NPN cofactor-binding proteins, it is important to develop a process allowing for the routine generation of cofactor-containing (active) forms of these enzymes to further characterize their properties. In addition, such a system could be used to confirm the reactivities and characterize the attributes of LarB-, LarE-, and LarC-like proteins. Such a capability will allow scientists to better analyze the NPN superfamily and to obtain a better understanding of the full potential of the NPN cofactor.

In previous studies, two general approaches were used to assess the function of *lar*-like genes. In one case, the set of *Lactiplantibacillus plantarum* *lar* genes under the control of a nisin-inducible promoter were transformed into *Lactococcus lactis* lacking the *lar* genes. Homologs of *larA* from other microorganisms were substituted for the corresponding *L. plantarum* gene within this Gram-positive host resulting in a few active enzymes, but this method was not generally successful.¹ Furthermore, efforts to substitute a gene encoding an NPN biosynthesis enzyme with a corresponding homolog have been hindered by plasmid instability and other confounding issues

(unpublished observations). In parallel with the *L. lactis* expression studies, individual homologs of *lar* genes were separately expressed in and purified from *Escherichia coli* for biochemical analysis.²⁻⁵ Significantly, only the substrate of LarB (NaAD) is commercially available, whereas

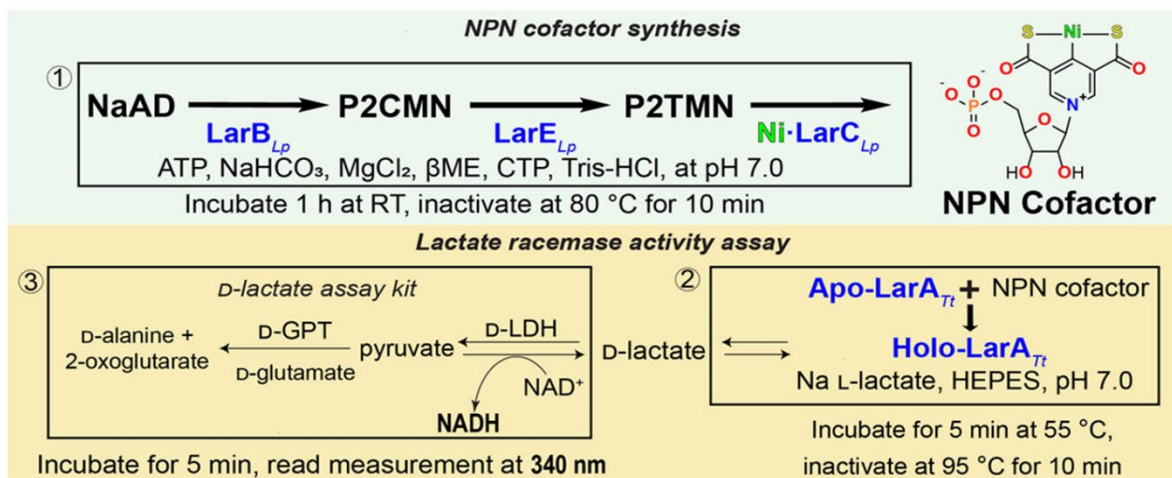


Figure 4-1 Workflow of *in vitro* Lar assay. LarB, LarE, LarC, and LarA apoprotein were separately expressed and purified for use in an *in vitro* assay. NaAD; nicotinic acid adenine dinucleotide, P2CMN; pyridinium-3,5-biscarboxylic acid mononucleotide, P2TMN; pyridinium-3-5-bisthiocarboxylic acid mononucleotide, RT; room temperature, NPN; nickel-pincer cofactor, D-LDH; D-lactate dehydrogenase, D-GPT; glutamate-pyruvate transaminase, NAD⁺ & NADH; nicotinamide adenine dinucleotide.

the substrates of LarE (P2CMN) or LarC (P2TMN) and the cofactor itself (NPN) must currently be generated via biosynthesis. The *E. coli*-produced proteins have allowed for testing of enzyme activities by a time-consuming and error-prone process (Figure 4-1) involving the sequential chemical transformations of NaAD with *L. plantarum* LarB, LarE, and LarC (or their homologs), as monitored by incorporating the cofactor into the LarA apoprotein from *T. thermosaccharolyticum* and subsequent assaying of Lar activity.¹

Here, I describe my successful co-expression of the four *lar* genes of *L. plantarum* within *E. coli* using the Duet expression vector system⁶ and the production of Lar activity in this genetically tractable and easily manipulated host microorganism. I show that genes encoding homologs of the NPN biosynthetic enzymes can be swapped for the corresponding genes in this system to test their biosynthetic functionality. In addition, I show the ability to exchange genes

encoding LarA homologs with the initial *L. plantarum* *larA* gene, allowing for synthesis of the holoprotein forms of these LarA-like proteins that can be further characterized. This new approach avoids the problems of the *L. lactis* expression system and the need to purify from *E. coli* each of the individual enzymes separately for the NPN biosynthesis requirement. I also describe a fluorescent staining procedure to identify proteins with covalently bound P2TMN via the reactivity of its thiocarboxylic acid.

Methods

Gene, plasmids, and cloning

Bacterial strains, plasmids, and primers used for this study are listed in Table 4-1. The plasmids were created through the *in vivo* subcloning assembly method.^{7,8} PCR amplifications were performed using Q5 high-fidelity DNA polymerase following the manufacture's protocol (NEB, Ipswich, MA). The primers used were purchased from IDT (Newark, NJ, USA). The transformations were performed using a standard chemical method in *E. coli* DH5 α and BL21 (DE3) cells for plasmid amplification and protein expression purposes, respectively.⁹

Table 4-1 Strains, plasmids and primers.

Strain, plasmid or primer	Characteristic(s) or sequence		Source or reference
Strains			
<i>Lc. Lactis</i>			
NZ3900	MG1363 derivative		10
<i>E. coli</i>			
DH5 α	F– ϕ 80lacZ Δ M15 Δ (<i>lacZYA-argF</i>) <i>U169 recA1 endA1 hsdR17</i> (rK– mK+) <i>phoA supE44 λ- thi–1 gyrA96 relA1</i>		ThermoFisher
BL21 (DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS</i> λ DE3 = λ <i>sBamH1o ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1)</i> <i>i21 Δnin5</i>		NEB
Plasmids			
pETDuet	<i>Amp^r</i>		Novagen
pRSFDuet	<i>Kan^r</i>		Novagen
pAT035	<i>Amp^r</i>	LarA _{Lp} and LarB _{Lp} expression; LarA _{Lp} purification	This study
pAT038	<i>Kan^r</i>	LarE _{Lp} and LarC _{Lp} expression	This study
pAT039	<i>Kan^r</i>	LarE _{Lp} and LarC _{Mt} expression	This study
pAT040	<i>Kan^r</i>	LarE _{Lp} and LarC _{Sc} expression	This study

Table 4-1, (cont'd)

pGIR112		Overexpression of LarA fused with Strep-tag at the C-terminus in the whole <i>larABC1C2DE</i> operon	¹
Primers	The orientation is 5' to 3' in all cases.		
LarALp-Strep_fw	CTTTAAGAAGGAGATATACCATGTCCGTTGCAATTGATTTACCATATGACAA		Subcloning
LarALp-Strep_rv	CCGCAAGCTTGTCGACCTACTTCTCAAATTGTGGATGACTCCAGC		Subcloning
pETDuet_MCS1_fw	CATCCACAATTTGAGAAGTAGCTTAAGTCGAACAGAAAGTAATCGTATTGTAC		Subcloning
pETDuet_MCS1_rv	CATATGGTAAATCAATTGCAACGGACATGGTATATCTCTTCTTAAAG		Subcloning
LarBLp_fw	GTAAAGTATAAGAAGGAGATATACATATGGCAACCACAGCAGAAATATTACAACAAGTG		Subcloning
LarBLp_rv	CCAGACTCGAGGGTACCTTACATTTGATTGACCATACTAGCTGAGTAGG		Subcloning
pETDuet_MCS2_fw	GTATGGTCAATCAAATGTAAGGTACCCTCGAGTCTGGTAAAG		Subcloning
pETDuet_MCS2_rv	CTGCTGTGGTTGCCATATGTATATCTCCTTCTTATACTTAACTAATATAC		Subcloning
LarELp_fw	CTTTAATAAGGAGATATACCATGGCAACATTAGCAACAATAAGCAACGTTAGTA		Subcloning
LarELp_rv	CTGTTGCGACTTAAGCTAGGCGAAAGTGGCCAATTG		Subcloning
pRSFDuet_MCS1_fw	CCACTTTCGCCTAGCTTAAGTCGAACAGAAAGTAATCGTATTGTACA		Subcloning
pRSFDuet_MCS1_rv	GCTAATGTTGCCATGGTATATCTCCTTATTAAAGTTAAACAAAATTATTTC		Subcloning
LarCLp_fw	GTATAAGAAGGAGATATACATATGGGTGCTCAAACACTTTATTTAGACGCTTTTTTC		Subcloning
LarCLp_rv	CCAGACTCGAGGGTACCTTACGCCTCCTCATCTAATTGATCTACCG		Subcloning
pRSFDuet_MCS2_fw	GATGAGGAGGCGTAAGGTACCCTCGAGTCTGGTAAAG		Subcloning
pRSFDuet_MCS2_rv	GCGTCTAAATAAAGTGTTTGAGCACCCATATGTATATCTCCTTCTTATACTTAAC		Subcloning
LarCSc_fw	GAAGGAGATATACATATGGGTCTGATCGCC		Subcloning
LarCSc_rv	GACTCGAGGGTACCTTAGCTTTCCGG		Subcloning
pRSFDuet_MCS2_Sc_fw	CTGAGTCCGGAAGCTAAGGTACCCTCGAG		Subcloning
pRSFDuet_MCS2_Sc_rv	CAAAATAGGCGATCAGACCCATATGTATATCTCC		Subcloning
LarCMt_fw	GAAGGAGATATACATATGAAGATCGCCTATTTTGATTGCTTTAGC		Subcloning
LarCMt_rv	GACTCGAGGGTACCTTAAAATGCTTTCAGTGCACGTGCGC		Subcloning
pRSFDuet_MCS2_Mt_fw	GCACTGAAAGCATTTTAAGGTACCCTCGAGTCTGG		Subcloning
pRSFDuet_MCS2_Mt_rv	CAAAATAGGCGATCTTCATATGTATATCTCCTTCTTATACCTTAAC		Subcloning

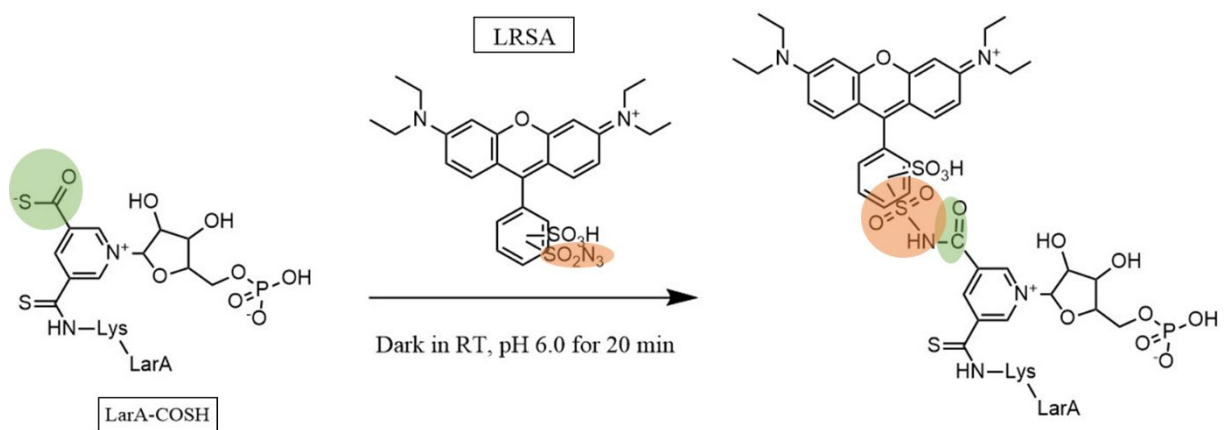
Gene overexpression and protein purification of LarA or its homologs

E. coli BL21 (DE3) strains containing the modified pETDuet and/or pRSFDuet plasmids were grown with the appropriate antibiotics in autoinduction medium.¹¹ The cultures were grown

at room temperature while shaking at 220 RPM. When indicated, 1 mM of NiCl₂ and/or 1 mM of nicotinic acid (final concentrations) were added after 4 h of growth. Cells were harvested after ~24 h. Cell pellets were resuspended in 100 mM Tris, pH 7.5, buffer containing 150 mM NaCl and stored at -80 °C until use. Once the cells were thawed, final concentrations of 0.5 mM Na₂SO₃, 1 mM phenylmethylsulphonyl fluoride (PMSF), one tablet of cOmplete™ EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland), 1 mM lysozyme, 1 mM dithiothreitol (DTT), and 1 unit of benzonase were added. The cells were lysed by two passes through a French pressure cell at 16,000 psi. Strep-tagged LarA were purified using StrepTactin XT resin (IBA, Göttingen, Germany) with buffers that included 0.05 mM Na₂SO₃ for NPN cofactor stabilization and the proteins were eluted with 50 mM biotin.¹² Protein concentrations were determined by the Bradford protein assay reagent (Bio-Rad, Hercules, CA) using bovine serum albumin as the standard.

Lissamine rhodamine B sulfonyl azide (LRSA) labeling for detection of P2TMN bound LarA

To assess whether LarA homologs covalently incorporated the NPN cofactor, I desired a reagent that would react with the thiocarboxylate of protein bound P2TMN and selected for study LRSA. This reagent reacts with thiocarboxylic acids according to the reaction shown in



Scheme 4-1 LRSA reaction. The “click reaction” between LRSA and the LarA thiocarboxylate. RT; room temperature, LRSA; lissamine rhodamine B sulfonyl azide. The highlighted region is where the click chemistry occurs.

Scheme 4-1. Synthesis of the LRSA reagent was carried out as previously described.¹³ The protocol for LRSA labeling of P2TMN-bound proteins was based on and modified from the procedure for labeling proteins that terminate in a thiocarboxylic acid at their carboxyl end.^{13,14} For this analysis, 0.5 g of cells were resuspended in 0.7 mL of 100 mM Tris-buffered saline containing 150 mM NaCl at pH 7.5 and transferred to a 2 mL tube to be lysed with a bead beater.

The lysates were centrifuged, and the supernatant solutions were collected. The samples were roughly normalized based on the overall protein content using the absorbance at 280 nm, and buffer exchanged into 50 mM potassium phosphate, 300 mM NaCl, and 6 M urea, at pH 6.1. To each of the samples 10 μ L of 15 mM LRSA in dimethyl sulfoxide was added and the vials were left to react in the dark at room temperature for 20 min. The protein portions of the samples were precipitated using the chloroform-methanol method³² and resuspended in the phosphate urea buffer stated above. Each sample (20 μ l) was mixed with 5 μ l of 5-fold concentrated sodium dodecyl sulfate (SDS)-loading buffer and 20 μ l was loaded onto a 12% acrylamide gel, subjected to SDS-polyacrylamide gel electrophoresis (PAGE), and used for imaging the proteins with bound LRSA followed by staining with Coomassie brilliant blue. The rhodamine-bound gel bands were excited at 530 nm while monitoring the emission at 580 nm and documented using the ChemiDoc MP imaging system (BioRad, Hercules, CA, USA).

LarA UV-visible spectra

The absorbance (200 – 800 nm) of the purified LarA proteins was measured using a quartz cuvette with a Shimadzu UV-2600 spectrophotometer at room temperature. Sample volumes were 1 mL.

Nickel content analysis

Quantification of the LarA_{Lp} nickel content was carried out by using an Agilent 710 Series (Santa Clara, CA, USA) inductively coupled plasma optical emission spectrometer (ICP-OES). The samples were prepared by adjusting to 35% w/v HNO₃ and heating at 95 °C for one h to mineralize the components. A final concentration of 0.1 ppm Yttrium (Sigma-Aldrich, St. Louis, MO, USA) was added to all samples as an internal standard. A nickel standard curve and a buffer control were used to account for background nickel contamination. Data were collected and analyzed using the ICP Expert II software.

Mass spectrometric analysis of NPN-binding to LarA

The LarA_{Lp} was analyzed by using a Waters G2-XS Q-TOF (time of flight) mass spectrometer by injecting 10 µl of sample onto a Thermo Hypersil Gold CN guard column (1.0 × 10 mm) for desalting. A gradient was run using 0.1% formic acid in water (solvent A) and acetonitrile (solvent B) as follows at a flow rate of 0.1 ml/min: initial conditions were 98% A/2% B, hold at 2% B to 5 min with the flow diverted to waste for the first 3 min, ramp to 75% B at 10 min and hold at 75% B to 12 min, return to 2% B at 12.01 min and hold to 15 min. Mass spectra were obtained using electrospray ionization in positive ion mode with a source temperature of 100 °C, cone voltage of 35 V, desolvation temperature of 350 °C, desolvation gas flow of 600 L/h, cone gas flow of 50 L/h, and capillary voltage of 3.0 kV. Data were acquired using a 0.5 s TOF MS scan across an *m/z* range of 200–2000. The spectra were deconvoluted in Masslynx using the maximum entropy (MaxEnt) I algorithm.

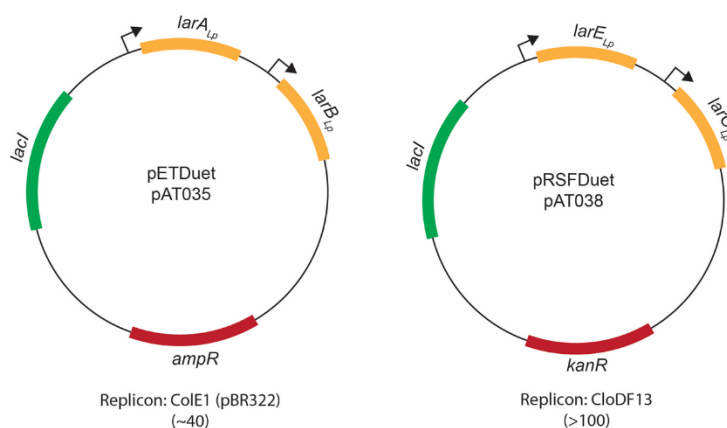
LarA activity assay

The purified Strep-tagged LarA protein from *L. plantarum* (LarA_{Lp}) was buffer exchanged to remove the Na₂SO₃ and biotin from the buffer using a PD-10 desalting column

(Marlborough, MA, USA). To assess the Lar activity, a LarA_{Lp} sample (1 pmol) was mixed with sodium L-lactate (5-400 mM) in 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, 100 mM) buffer, pH 7.0, for 1-12 min at 35 °C, then boiled for 10 min at 95 °C to inactivate the enzyme. The precipitated protein was removed by centrifugation at 17,000 x g for 10 min and the supernatant was collected. The amount of D-lactate in the sample, produced by the lactate racemase activity of LarA, was measured using a commercial kit (Neogen, Lansing, MI) as previously described.¹

Results and Discussion

Construction of the Duet system plasmids for synthesis of active LarA in *E. coli*



To co-express the genes encoding the NPN biosynthesis pathway proteins (LarB, LarE, and LarC) along with that encoding lactate racemase (LarA), I chose to use the Duet Expression system (Novagen, Merck KGaA, Darmstadt, Germany).⁶

Figure 4-2 Plasmid design for *lar* genes expression in *E. coli*. The number in the parentheses indicates the copy number.

The system includes four plasmids that have compatible replication origins and different antibiotic resistance cassettes. Each of the plasmids contains two multiple cloning sites, allowing for up to eight genes to be co-expressed in one culture of *E. coli*. Among the four plasmids, I chose pETDuet and pRSFDuet to express the Lar-related genes from *L. plantarum* (Figure 4-2). The genes encoding LarE_{Lp} and LarC_{Lp} were selected for expression using the pRSFDuet plasmid (pAT038), a high copy number vector, since these proteins are thought to catalyze single turnover reactions, whereas *larA* and *larB* from *L. plantarum* were expressed from the pETDuet

plasmid (pAT035). Importantly, the gene encoding LarA_{Lp} was cloned with a sequence for a Strep-tag on its C-terminus for easy purification prior to analyzing activity and testing for the presence of covalently bound NPN. These plasmids were designed to allow homologous genes to be substituted for the *L. plantarum* genes to test for the biosynthetic abilities and functionality of the corresponding proteins.

Testing for the presence of protein bound NPN by LRSA labeling, chromophore absorbance, nickel content, and mass spectrometry

Cell-free lysates derived from the *E. coli* Duet system expressing the *L. plantarum* genes were tested for the presence of NPN-bound LarA_{Lp} by four approaches.

First, the presence of protein bound cofactor was qualitatively investigated by reacting the P2TMN thiocarboxylic acid adduct with LSRA, resolving the proteins by SDS-PAGE, and visualizing the labeled protein bands by fluorescence imaging (Figure 4-3). I compared the intensity of labeling for cell lysates derived from the *E. coli* Duet system with both pAT035 and

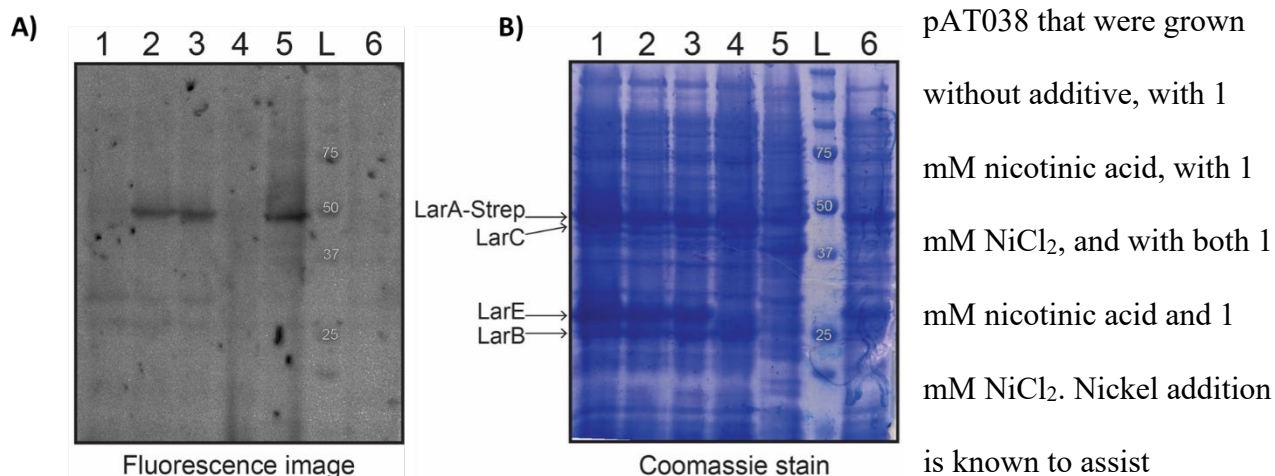


Figure 4-3 LRSA labeling of crude *E. coli* lysates. LRSA-labeled protein samples were subjected to denaturing gel electrophoresis, imaged for (A) fluorescence (with excitation and emission wavelengths of 530 nm and 580 nm), and (B) stained with Coomassie brilliant blue. Lane 1: lysate of culture expressing pAT035 and pAT038 with 1 mM nicotinic acid supplemented; lane 2: lysate of culture with 1 mM NiCl₂; lane 3: lysate of culture with 1 mM nicotinic acid and NiCl₂; lane 4: lysate of culture expressing pAT035 alone with 1 mM nicotinic acid and NiCl₂; lane 5: *L. lactis* lysate with 1 mM NiCl₂ supplemented during growth; lane L: protein ladder; lane 6: culture with no additive.

Nicotinic acid is an upstream precursor of the substrate required for the NPN biosynthesis pathway, NaAD,¹⁶ and the supplementation was necessary for LarA activation in the *L. lactis* cells expressing the *lar* operon genes.¹⁷ In addition, I examined a sample from *E. coli* containing only pAT035 as a negative control. Furthermore, I investigated cell extracts of *L. lactis* (pGIR112) expressing the *lar* genes as a positive control.¹

As shown in the fluorescence image of Figure 4-3A, a band corresponding in size to Strep-tagged LarA_{Lp} (47.5 kDa) was fluorescently labeled when using samples derived from *E. coli* (pAT035/pAT038) or *L. lactis* (pGIR112) grown in the presence of nickel ions. In the absence of added nickel ions, the LarA band was not labeled, presumably due to a requirement for complete NPN cofactor biosynthesis prior to covalent attachment to LarA in the *E. coli* cells. No band was labeled for *E. coli* (pAT035) that produced LarA and LarB but was incapable of NPN cofactor biosynthesis because it lacked LarE and LarC. Analysis of the SDS-PAGE gel by Coomassie staining (Figure 4-3B) revealed approximately equal loading of proteins in each lane and revealed the presence of LarC_{Lp} (46.5 kDa), LarE_{Lp} (30.5 kDa), and LarB_{Lp} (25.3 kDa). Of additional interest, the relative intensities of several bands were altered in the *L. lactis* lysate

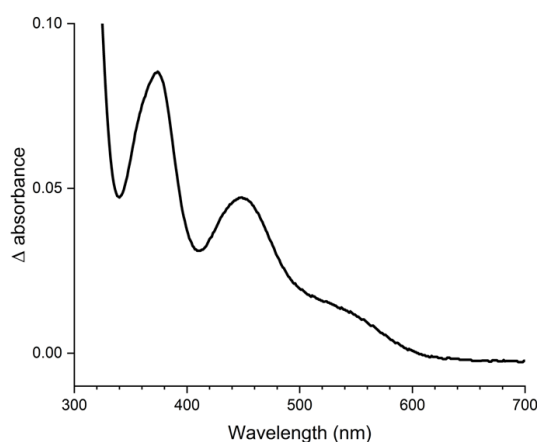


Figure 4-4 UV-Vis spectroscopic difference of LarA_{Lp} holoprotein and apoprotein from *E. coli* expression system. Both samples contained 50μM sulfite. The concentrations of the samples were adjusted to have an absorbance of 1.0 at 280 nm.

which has a different expression system.

Although these results indicate that NPN-bound LarA_{Lp} is generated in the *E. coli* cells, it does not allow for precise quantification of the labeling.

A second approach to monitor NPN cofactor binding to LarA_{Lp} in *E. coli* was to test for the presence of the chromophore in

the purified sample. The UV-vis difference spectrum of the enzyme versus the apoprotein (Figure 4-4) revealed electronic transitions at 375 nm and 440 nm along with a shoulder at 550 nm that are not present in the apoenzyme sample, in agreement with prior findings for the enzyme purified from *L. lactis*.^{12,17} The intensities of these absorptions divided by the protein concentration was less than that reported previously for LarA_{Lp} isolated from *L. lactis* suggesting somewhat less incorporation of the cofactor.¹

As an additional method to quantify the NPN cofactor content I quantified the metal content in the purified protein sample. ICP-OES result indicated ~20% nickel loading (Figure 4-5) which agrees with what was previously seen with LarA_{Lp} purified from *L. lactis*.⁹

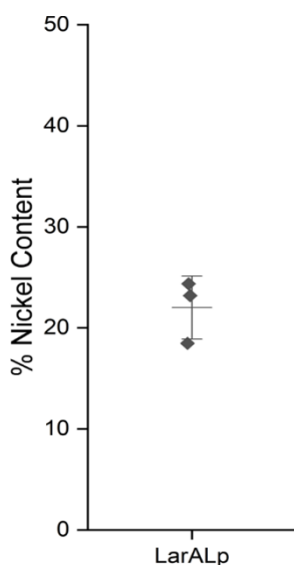


Figure 4-5 Nickel content analysis of LarA_{Lp}. The samples were measured using ICP-OES. The protein concentration was determined using Bradford. Each point represents an average of triplicates of one biological replicate (n=3).

Finally, as a fourth method to assess the content of NPN cofactor in LarA_{Lp}, I subjected the apoprotein and holoprotein samples to mass spectrometry. Figure 4-6 shows that each sample is nearly homogeneous and the *m/z* values are consistent with the masses expected for the proteins missing their *N*-terminal methionine residues, with an additional mass difference of 450.9 Da in the holoprotein. This result is consistent with the NPN cofactor being covalently bound to lysine 184, as previously reported.¹⁷

Demonstration of Lar activity in E. coli

The above experiments demonstrate that the NPN-cofactor had become covalently bound to LarA_{Lp} at high levels in the *E. coli* expression system. To further examine the utility of this system for analysis of function for LarA homologs, I tested the activity of the LarA_{Lp} holoprotein. I detected lactate racemase activity in all lysate samples from Figure 4-3 except for the negative control

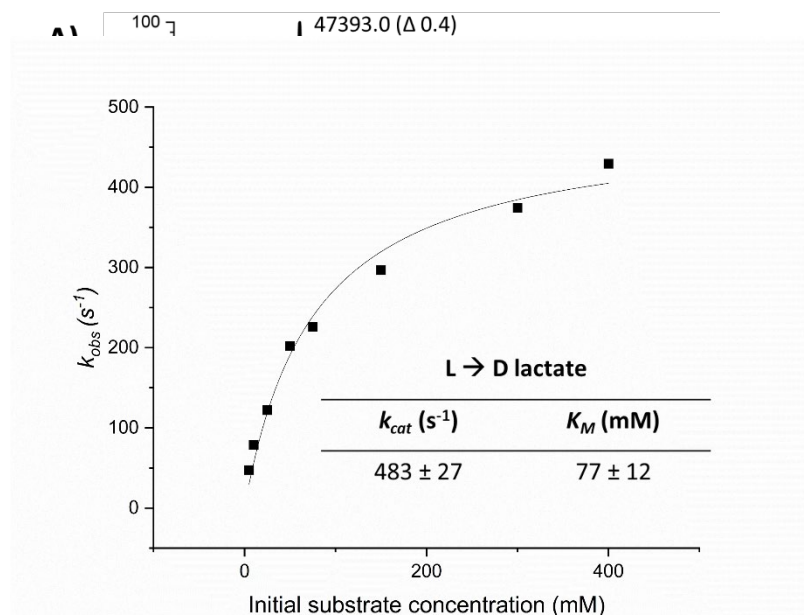


Figure 4-7 LarA_{Lp} purified from *E. coli* kinetic. Michaelis-Menten curve of the lactate racemase specific activity (k_{obs}) of LarA_{Lp} in the L- to D-lactate direction. The curve was fitted using the OriginLab app Enzyme Kinetics ver. 1.10. Each point represents an average of quadruplicates from one representative experiment.

(data not shown). A possible reason to explain the presence of activity in samples that were not labeled by LRSA labeling (from cells provided with only 1 mM nicotinic acid or no additive (lanes 1 and 7) could be due to insufficient levels of covalently bound NPN cofactor due to insufficient amounts of nickel to be detected by the labeling.

The kinetic information of the purified LarA_{Lp} is provided in Figure 4-7. Compared to the kinetics of LarA_{Lp} purified from *L. lactis*, the K_m is increased by ~1.5-fold and k_{cat} is decreased by 10-fold. The reduction in the kinetics parameters shouldn't be a hindrance to assess functionality of selected *lar* gene homologs since qualitative activity assays are done at much higher protein concentrations (e.g., 1 pmol of enzyme was used for the kinetics assay vs. the 0.8 μM of protein used for qualitative activity assays).⁷

Testing homologous biosynthesis pathway genes

To test the versatility of this system, I examined whether homologs of genes in the NPN biosynthesis pathway could substitute for genes encoding the *L. plantarum* enzymes. For example, I swapped *larC_{Lp}* with *larC* from *M. thermoacetica* which I had shown to encode an enzyme with nickel insertase activity in previous work.¹⁸ In parallel, I constructed a plasmid with the *larC_{Lp}* swapped with the homolog from *Synechococcus* sp. PCC 6803, which has not yet

been tested for its ability to insert nickel into P2TMN and create a mature NPN cofactor. Mass spectrometric analysis of the LarA_{Lp} resulting from both constructs demonstrated the presence of covalently bound NPN (data not shown). Furthermore, the Lar activity was detected for both enzymes (data not shown). These findings demonstrate that this *E. coli* expression system can be used to test the function of potential biosynthesis pathway enzymes identified only as being sequence homologs.

This methodology already has been implemented by other researchers in the laboratory to demonstrate the covalent attachment of NPN to LarA of *Megasphaera elsdenii* and the generation of Lar activity when using the gene encoding LarE of *Latilactobacillus sakei*.

REFERENCES

- (1) Desguin, B., Goffin, P., Viaene, E., Kleerebezem, M., Martin-Diaconescu, V., Maroney, M. J., Declercq, J. P., Soumillion, P., and Hols, P. (2014) Lactate racemase is a nickel-dependent enzyme activated by a widespread maturation system. *Nat Commun* **5**, 3615
- (2) Desguin, B., Urdiain-Arraiza, J., Da Costa, M., Fellner, M., Hu, J., Hausinger, R. P., Desmet, T., Hols, P., and Soumillion, P. (2020) Uncovering a superfamily of nickel-dependent hydroxyacid racemases and epimerases. *Sci Rep* **10**, 18123
- (3) Desguin, B., Soumillion, P., Hols, P., and Hausinger, R. P. (2016) Nickel-pincer cofactor biosynthesis involves LarB-catalyzed pyridinium carboxylation and LarE-dependent sacrificial sulfur insertion. *Proc Natl Acad Sci USA* **113**, 5598-5603
- (4) Fellner, M., Desguin, B., Hausinger, R. P., and Hu, J. (2017) Structural insights into the catalytic mechanism of a sacrificial sulfur insertase of the N-type ATP pyrophosphatase family, LarE. *Proc Natl Acad Sci USA* **114**, 9074-9079
- (5) Desguin, B., Fellner, M., Riant, O., Hu, J., Hausinger, R. P., Hols, P., and Soumillion, P. (2018) Biosynthesis of the nickel-pincer nucleotide cofactor of lactate racemase requires a CTP-dependent cyclometallase. *J Biol Chem* **293**, 12303-12317
- (6) Tolia, N. H., and Joshua-Tor, L. (2006) Strategies for protein coexpression in *Escherichia coli*. *Nat Methods* **3**, 55-64
- (7) García-Nafria, J., Watson, J. F., and Greger, I. H. (2016) IVA cloning: A single-tube universal cloning system exploiting bacterial *in vivo* assembly. *Sci Rep* **6**, 27459
- (8) Huang, F., Spangler, J. R., and Huang, A. Y. (2017) *In vivo* cloning of up to 16 kb plasmids in *E. coli* is as simple as PCR. *PLOS ONE* **12**, e0183974
- (9) Hanahan, D., Jessee, J., and Bloom, F. R. (1991) Plasmid transformation of *Escherichia coli* and other bacteria *Meth Enzymol* **204**, 63-113
- (10) Ruyter, P. G. d., Kuipers, O. P., and Vos, W. M. d. (1996) Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. *Appl Environ Microbiol* **62**, 3662-3667
- (11) Studier, F. W. (2014) Stable expression clones and auto-induction for protein production in *E. coli*. *Methods Mol Biol* **1091**, 17-32
- (12) Rankin, J. A., Mauban, R. C., Fellner, M., Desguin, B., McCracken, J., Hu, J., Varganov, S. A., and Hausinger, R. P. (2018) Lactate racemase nickel-pincer cofactor operates by a proton-coupled hydride transfer mechanism. *Biochemistry* **57**, 3244-3251

- (13) Krishnamoorthy, K., and Begley, T. P. (2010) Reagent for the detection of protein thiocarboxylates in the bacterial proteome: lissamine rhodamine B sulfonyl azide. *J Am Chem Soc* **132**, 11608-11612
- (14) Xie, L. (2018) *Enzyme-catalyzed acylium ion formation and reagents for the detection of thiocarboxylates involved in the biosynthesis of the NAD derived pincer cofactor*. Doctor of Philosophy Doctoral dissertation, Texas A & M University
- (15) Wessel, D., and Flugge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **138**, 141-143
- (16) Foster, J. W., and Moat, A. G. (1980) Nicotinamide adenine dinucleotide biosynthesis and pyridine nucleotide cycle metabolism in microbial systems. *Microbiol Rev* **44**, 83-105
- (17) Desguin, B., Zhang, T., Soumillion, P., Hols, P., Hu, J., and Hausinger, R. P. (2015) A tethered niacin-derived pincer complex with a nickel-carbon bond in lactate racemase. *Science* **349**, 66-69
- (18) Turmo, A., Hu, J., and Hausinger, R. P. (2022) Characterization of the nickel-inserting cyclometallase LarC from *Moorella thermoacetica* and identification of a cytidinylylated reaction intermediate. *Metallomics* **14**

CHAPTER 5

CONCLUSIONS AND FUTURE STUDIES

Conclusions from my studies

In this thesis, I have reported new discoveries pertaining to the nickel insertase or cyclometallase protein, LarC, that functions in the biosynthesis pathway of the nickel-pincer nucleotide (NPN) cofactor and I have expanded the set of methods that can be used to study the activity of homologs of LarA and the NPN biosynthetic enzymes. Of particular interest, I discovered that the LarC reaction possesses an intermediate, cytidinylylated P2TMN, that explains the previously unknown purpose of CTP in the reaction (Chapter 2). I also obtained preliminary structural information for the full-length LarC protein from *Moorella thermoacetica* by cryo-electron microscopy and I generated *in silico* structural models of this LarC structure allowing for identification of possible binding sites of the substrates (Chapter 3). Finally, I developed a *lar* expression system for use in *Escherichia coli* utilizing the well-established Duet expression vectors (Chapter 4). I coupled this system with several other experimental approaches, allowing me to expand our knowledge of the NPN cofactor biosynthesis pathway by testing the functionality of Lar-related protein homologs. Below is a summary of each of the chapters along with suggestions for potential studies that could extend the work to address yet answered questions related to LarC and the other *lar* operon expressed enzymes.

Additional studies related to the LarC reaction intermediate

In chapter 2, I addressed the question of CTP's role in the nickel inserting activity of LarC. My investigation resulted in the mass spectrometric identification of an intermediate of the reaction, P2TMN-CMP, that becomes metalated and is hydrolyzed to release NPN. Moving forward, studies should be conducted to further elucidate the function of this intermediate in the overall nickel inserting mechanism. I speculated that the function is probably analogous to how molybdenum insertase installs its metal to produce the mature Moco cofactor. In that system, the

enzyme was shown to form the AMP adduct of molybdopterin (MPT), the AMP-MPT intermediate was shown to correctly position the modified substrate to facilitate molybdenum insertion, and the AMP was subsequently removed by hydrolysis.¹ Notably, the AMP-MPT hydrolysis reaction was not utilized as an energy source thus providing an excellent parallel to what I suggest happens during the LarC reaction.²

To test whether this hypothesis applies to LarC, additional variants could be made to identify residues that are critical for CMP adduct formation and release during the reaction. In addition, it should be established whether elimination of nickel from the growth medium generates higher levels of the intermediate as my preliminary experiments suggest. Finally, emphasis should be placed on solving several structures of the enzyme, ideally with bound P2TMN, bound nickel, and, most importantly, the bound CMP-P2TMN intermediate.

Follow-up efforts to obtain structural insights on LarC_{Mt}

Chapter 3 described my preliminary low-resolution structural data for *M. thermoacetica* LarC using cryo-electron microscopy that was carried out in collaboration with Prof. Kelly Kim. Prior studies had solved the structure of the C-terminal portion of *L. plantarum* LarC, but crystallization of the N-terminal or full-length structure was not possible due to protein aggregation and proteolysis.³ Moving forward, Dr. Kim's lab has already solved the preferred orientation problem by including DDM in the buffer and obtained evidence for a hexameric structure of the full-length enzyme. It seems like that cryo-electron microscopy studies will be able to achieve the full-length structure of this protein. Additional avenues to consider would be to create a truncated version of LarC_{Mt} for attempts to crystallographically resolve the N-terminal structure if a full-length structure is not feasible. The truncation would likely be located at the flexible linker region predicted by using the AlphaFold modeling. The *L. plantarum* and *M.*

thermoacetica homologs have differences in the length and abundance of His residues in the His-rich region within the *N*-terminal domain, but either structure would provide keen insights to better understand the enzyme. Additional future studies could include efforts to structurally characterize other homologs of LarC. For example, I obtained preliminary mass spectrometric data indicating the truncated *C. difficile* LarC homolog was soluble. An attempt at purification using a Ni-NTA column could be made to purify and solve the structure of the *N*-terminal domain.

Future studies using the *E. coli* system to study *lar* gene homologs

My efforts have demonstrated that active lactate racemase can be generated in significant quantities by expressing the appropriate set of *lar* gene constructs in *E. coli*, rather than using the less tractable *L. lactis* system (Chapter 4). I showed that expression of *L. plantarum* genes leads to the production of LarA that stains with a fluorescent reagent specific for thiocarboxylic acids, exhibits an NPN-based chromophore, has the expected nickel content, and possesses the appropriate mass for the protein linked to an NPN adduct. Using this experimental toolbox, I have confirmed that homologs of LarC from *M. thermoacetica* and *Synechococcus* are able to replace the function of the *L. plantarum* LarC to create NPN. Ongoing studies are examining the NPN cofactor biosynthetic ability of a LarE homolog from *Latilactobacillus sakei* and the phenyllactate racemase activity of a LarA homolog of *Megasphaera elsdenii*. This system can also be used to examine the function of NPN-binding non-LarA proteins that are likely to be present in organisms that contain *larB*-, *larE*-, *larC*-like genes for the NPN biosynthesis pathway, but lack a homolog of *larA*. A gene suspected to encode such an NPN-binding protein can easily be swapped with the gene encoding LarA_{Lp} in the Duet system and the NPN cofactor-bound enzyme could be purified and characterized.

It is also worth mentioning that use of the Duet Expression system can be further expanded by using additional compatible plasmids with so-far unused overexpression sites. Indeed, four additional genes could be co-expressed with the present four-gene system. For example, another plasmid could encode *iscS* or other *isc* genes to facilitate the synthesis of LarE protein homologs that contain a [4Fe-4S] cluster or that need to synthesize and recycle a [4Fe-5S] cluster.⁴ The Duet system potentially could be improved by utilizing an *E. coli* strain lacking a nickel exporter (e.g., *rcnA*) or a strain with altered nickel homeostasis (such as one with a defective NikR).⁵ Additional engineering of the strain could boost the production of NaAD in the cell by increasing expression of genes in the Preiss-Handler pathway (i.e., genes encoding nicotinic acid phosphoribosyl transferase and nicotinate-nucleotide adenylyltransferase).⁶

Concluding remarks

Since the discovery of the NPN cofactor in 2015⁷ there have been many breakthroughs to understand this novel biological pincer complex. My contributions have already expanded our understanding of an important step for synthesis of this biological pincer complex. It is very likely that further use of my *lar* gene Duet plasmid system will greatly expand our knowledge of the function and importance of these genes in other organisms.

REFERENCES

- (1) Krausze, J., Hercher, T. W., Zwerschke, D., Kirk, M. L., Blankenfeldt, W., Mendel, R. R., and Kruse, T. (2018) The functional principle of eukaryotic molybdenum insertases. *Biochem J* **475**, 1739-1753
- (2) Probst, C., Yang, J., Krausze, J., Hercher, T. W., Richers, C. P., Spatzal, T., Kc, K., Giles, L. J., Rees, D. C., Mendel, R. R., Kirk, M. L., and Kruse, T. (2021) Mechanism of molybdate insertion into pterin-based molybdenum cofactors. *Nat Chem* **13**, 758-765
- (3) Desguin, B., Fellner, M., Riant, O., Hu, J., Hausinger, R. P., Hols, P., and Soumillon, P. (2018) Biosynthesis of the nickel-pincer nucleotide cofactor of lactate racemase requires a CTP-dependent cyclometallase. *J Biol Chem* **293**, 12303-12317
- (4) Chatterjee, S., Parson, K. F., Ruotolo, B. T., Mccracken, J., Hu, J., and Hausinger, R. P. (2022) Characterization of a [4Fe-4S]-dependent LarE sulfur insertase that facilitates nickel-pincer nucleotide cofactor biosynthesis in *Thermotoga maritima*. *J Biol Chem* **298**, 102131
- (5) Iwig, J. S., Rowe, J. L., and Chivers, P. T. (2006) Nickel homeostasis in *Escherichia coli* the *rcnR-rcnA* efflux pathway and its linkage to NikR function. *Mol Microbiol* **62**, 252-262
- (6) Yang, L., Mu, X., Nie, Y., and Xu, Y. (2021) Improving the production of NAD⁺ via multi-strategy metabolic engineering in *Escherichia coli*. *Metab Eng* **64**, 122-133
- (7) Desguin, B., Zhang, T., Soumillon, P., Hols, P., Hu, J., and Hausinger, R. P. (2015) A tethered niacin-derived pincer complex with a nickel-carbon bond in lactate racemase. *Science* **349**, 66-69