A Study of Methyl-coenzyme M Reductase Maturation: Coenzyme F430 Biosynthesis and Post-translational Modifications

By

Kaiyuan Zheng

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirement for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 15, 2018

Approved by

Steven Mansoorabadi, Chair, Associate Professor of Chemistry
Holly Ellis, Professor of Chemistry
Eduardus Duin, Professor of Chemistry
Douglas Goodwin, Associate Professor of Chemistry

Abstract

This dissertation seeks to investigate the coenzyme F430 biosynthetic pathway and post-translational modification (PTM) enzymes of methyl-coenzyme M reductase (MCR), which will have applications in natural gas-to-liquid fuel conversion strategies and in the development of inhibitors to help reduce natural greenhouse gas emissions.

Methanogenesis, also named biomethanation, is the metabolic production of methane carried out by archaea called methanogens. MCR plays a significant role in methanogenesis and the anaerobic oxidation of methane (AOM). It catalyzes the terminal step of methanogenesis, resulting in the release of methane. Structural studies show MCR as a heterohexamer, with coenzyme F430, a highly reduced, nickel-chelated tetrapyrrole, at two active sites. However, the enzymology of F430 biosynthesis was unknown. Two compounds, precorrin 2 and sirohydrochlorin, had been proposed as biosynthetic precursors of F430. Molecular differences between these precursors and F430 include nickel chelation, two sites of amidation, the formation of two exocyclic rings, and a four- or six-electron ring reduction, respectively. Based on these differences, a comparative genomics approach was utilized to identify genes related to coenzyme F430 biosynthesis (MA3631, MA3626, MA3627, MA3628, and MA3630 in *Methanosarcina acetivorans* C2A). These genes were cloned, and the encoded enzymes were expressed. Activity assays were developed to determine the function of each enzyme. Besides *in vitro* experiments, all coenzyme F430 biosynthesis (*cfb*) genes were cloned in one vector.

Coexpression of the genes and following high-performance liquid chromatography (HPLC) analysis of cell extracts indicate that F430 is unable to be synthesized by the *cfb* genes alone in the heterologous host *Escherichia coli*. Additionally, it was found that ferredoxin and ferredoxin reductase (FNR) from spinach could support CfbCD catalysis *in vitro* and was therefore included in the *in vivo* F430 biosynthesis studies.

Moreover, there are several PTMs surrounding the MCR active sites. To elucidate the origin of the modifications, an *in vivo* coexpression approach with mass spectrometry (MS) protein sequencing was utilized. A cell line containing *mcrA* (or *mcrABG* or *ABGDC*) was used for McrA expression. The *mcrA* gene is equipped with either a C-terminal hexahistidine tag or a small ubiquitin-like modifier (SUMO) tag for ease of protein purification and increased protein solubility. MS data indicate that 1-*N*-methylhistidine and *S*-methylcysteine modifications are probably carried out by a protein methylation gene A (*prma*) homolog. Data also indicate that methanogenesis marker 1 (*mm1*) possibly catalyzes the biosynthesis of thioglycine modification with TfuA and ThiI.

Dedication

Five years ago, a young Chinese graduate student arrived in a small town located in the southeast United States. This place is ten thousand miles away from his hometown. Even with the fastest commercially available vehicle, it still takes over 16 hours to finish the trip. However, the long travel did not eliminate his passion and hope. Five years later, the young man became a Ph.D., and he feels grateful when he is writing these words. In these five years, there were so many memorable moments of joy and sorrow. In the hardest moments he thought about giving up studying for the degree, because it was such a lonely moment. Forturnately, his parents and his friends reignited him. He noticed that he is not alone, because there is God's love to fulfill him. He writes these words with appreciation and praise, even though there is a long way to go. He sincerely thanks the people that helped and loved him these past five years, and he sincerely thanks God for a completely different life and heart. Here, I, the young gradute student, gives thanks for such a precious gift in my life and offer all my praise!

Acknowledgements

I am full of appreciation when I am writing this acknowledgement for my dissertation and my five-year Ph.D. career. As an international student, it is a unique experience to study and live in a nation that is ten thousand miles away from my hometown. I feel very humble at this moment because I stand on a lot of selfless sharing and dedication. All these treasures set a firm foundation that enables me to see further. First, I would like to acknowledge my advisor, Dr. Steven Mansoorabadi. His sparks of thought always inspire me while involving me in such interesting projects. His erudition, preciseness, humility, and patience also motivate me to walk my way of self-development. He is the person who had the largest influence in my scientic career, for I am such an impatient and careless person. I would like to personally thank him. I would also like to acknowledge my collegues and my friends, Phong Ngo and Victoria Owens. Phong's passion of work and strict experimental spirit deeply impressed me. His guidance on experiments polished my skills as a scientist. Victoria is such a warm-hearted labmate who provided me a lot of encouragement in my difficult moments. I would like to acknowledge Patrick Donnan, a professional young scientist who provides me a lot of constructive advice. Selamawit Ghebreamlak, a responsible and reliable labmate who helped regulate all lab operation details. Trey Slaney, an honest reminder for experiments and daily life. I would also like to acknowledge Shuxin Li, Xingchen Huang, and Kenny Nguyen for your honest communication and selfless sharing. I would like to acknowledge Dr. Eduardus Duin for his

selfless sharing of experimental techniques and instruments. Also, for his advice on this dissertation and throughout my Ph.D. career, which was very constructive. I would like to acknowledge Dr. Holly Ellis who is such a gentle advisor that provided a lot of encouragement for my career. I would like to acknowledge Dr. Douglas Goodwin for his precise teaching and constructive structural advice for my proposal. I also would like to acknowledge the undergraduate students in the Mansoorabadi lab, including Benjamin Mueller and Stephen Chandler. Your passion for science inspires me because you are the future of science. I would like to acknowledge all of the chemistry department, graduate school, and international student office staff. Your warm-hearted help supported an outlander's road of study. And last, but certainly not least, I would like to acknowledge Eta Isiorho, Steve's wife. Thank you for your dedication to this lab and for such a warm family. Such a moment, I am full of appreciation. As a Christian, All Glory be to God. Thank you! God bless you!

Table of Contents

List of Tables

List of Schemes

List of Figures

List of Abbreviations

5-ALA	5-aminolevulinic acid
ADP	Adenosine diphosphate
AOM	Anaerobic oxidation of methane
ANME	Anaerobic methanotrophic archaea
AP	Alkaline phosphatase
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
BNF	Biological nitrogen fixation
cfb	Coenzyme F430 biosynthesis
CoA	Coenzyme A
CoB-S-S-CoM	Coenzyme B-Coenzyme M heterodisulfide
COR	Chlorophyllide *a* oxidoreductase
DAD	Diode Array Detector
DIET	Direct interspecies electron transfer
DPOR	Dark-operative protochlorophyllide *a*
DTT	Dithiothreitol
EPR	Electronic paramagnetic resonance
ESI	Electrospray ionization
ETC	Electron-transport component
FA	Formic acid

FAD	Flavin adenine dinucleotide
Fnr	Ferredoxin NADPH reductase
Fd	Ferredoxin
GWP	Global warming potential
HDR	Heterodisulfide reductase
HDS	Heterodisulfide
HPLC	High-performance liquid chromatography
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
isc	Iron-sulfur cluster
LB	Lysogeny broth
MCR	Methyl-coenzyme M reductase
MCS	Multiple cloning site
mm1	methanogenesis marker 1
mm10	methanogenesis marker 10
MS	Mass spectrometry
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
ORF	Open-reading frame
PBG	Porphobilinogen
PCR	Polymerase chain reaction
PEP	phosphoenolpyruvate
PK	Pyruvate kinase
RPL11	Ribosomal protein L11
PTM	Post translational modifications

RT	Room temperature
SAM	S-adenosyl-L-methionine
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SRB	Sulfate-reducing bacteria
SUMO	Small ubiquitin-like modifier
TB	Terrific broth
TBS	Tris buffered saline
TCA	Tricarboxylic acid
TEMED	Tetramethyl ethylenediamine
TOMMs	Thiazole/oxazole-modified microcins
TUV	Tunable Ultra-violet
tva	Thioviridamide
UMA	UDP-*N*-acetylmuramoyl-L-alanine
UV	Ultra-violet

Chapter 1

Introduction

1.1 Methanogens

Methanogenesis describes the process that produces methane in microorganisms called methanogens, all of which belong to the domain Archaea. There are three domains of life: Bacteria, Archaea, and Eukarya (*1*) (Figure 1). This system was introduced by Carl Woese and his group in 1977 based on differences in 16S rRNA genes (*2*). Species belonging to the archaeal domain are characterized by living in extreme environments, for example, hyperthermia, hypertonicity, or anaerobic conditions. Archaea carry out distinct biochemistry (e.g., RNA synthesis) when compared with Bacteria, even though both lack a nuclear membrane. Within Archaea, over fifty species are classified as methanogens. The species belong to many genera, including *Methanosarcina*, *Methanococcus* and *Methanobacterium.* All known methanogens are coccoid or bacilli. Most methanogens cannot function under aerobic conditions, with even trace amounts of oxygen being lethal. As an exception, recently a species named *Candidatus Methanothrix paradoxum* has been identified to function under aerobic conditions. Another exceptional example is *Methanosacina barkeri*. This species can survive longer than other methanogens in the presence of oxygen because it possesses a superoxide dismutase (*3*). Methanogens can be found living within several eukaryotic species. Microorganisms, including methanogens, help ruminants digest cellulose into nutrients that can then be absorbed by the animals. Other methanogens live in human digestive systems, producing methane-containing

flatus. Also, some methanogens have been found in soil and wetlands, contributing to the process of organic matter degradation (*4*) (*5*) (*6*) (*7*).

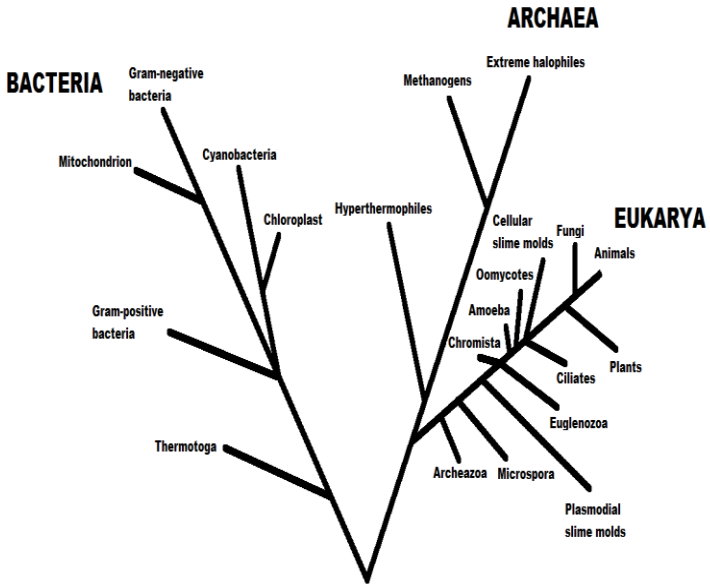
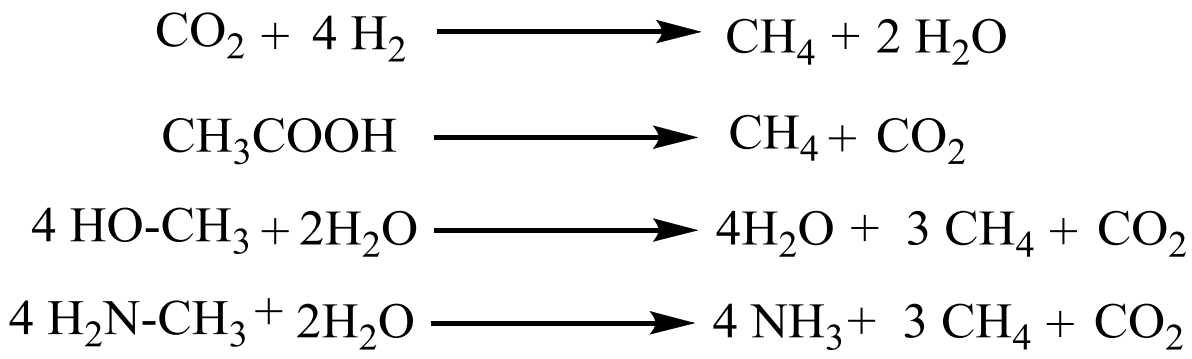


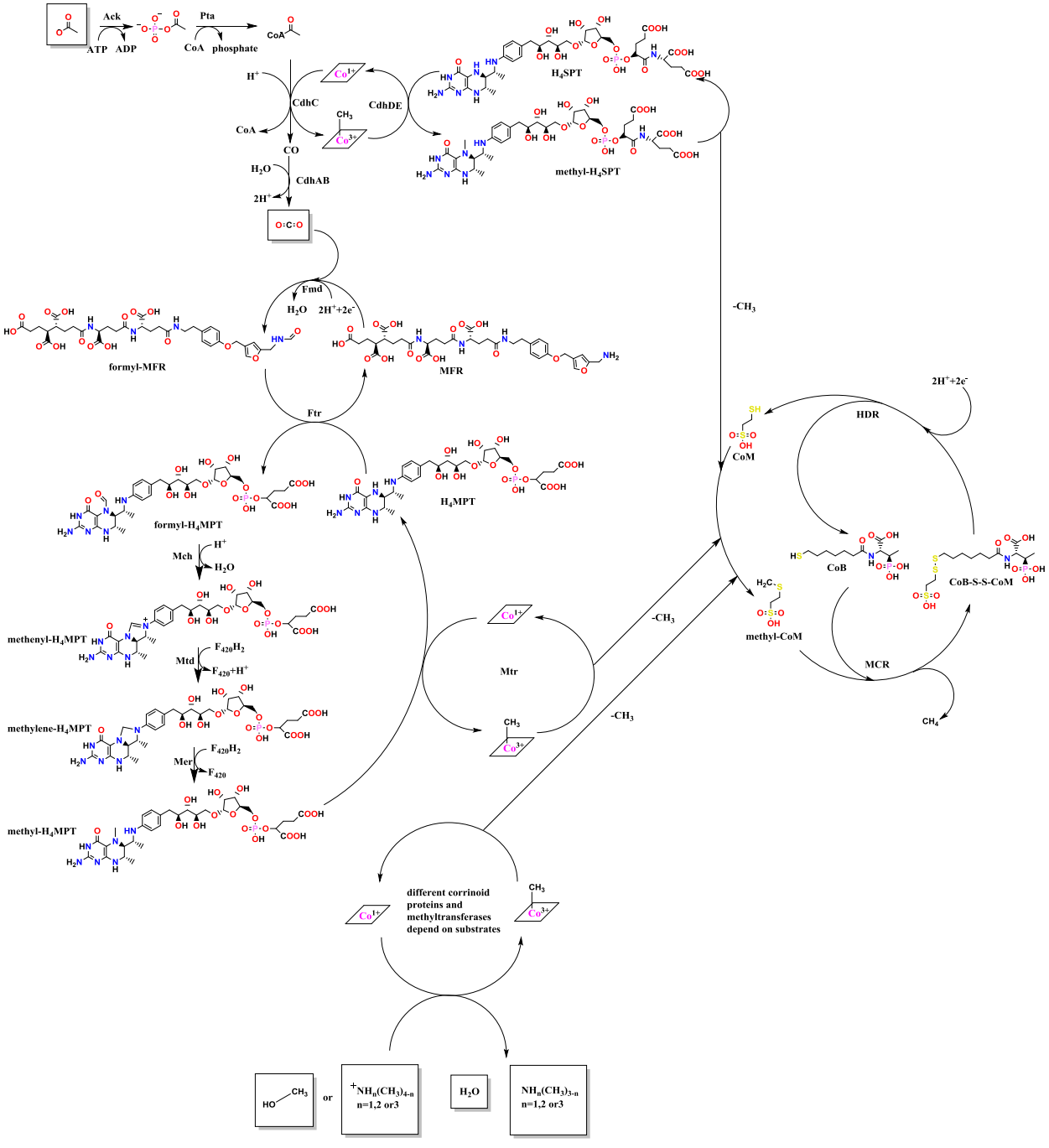
Figure 1: A phylogenetic tree based on rRNA data (*1*). This picture emphasizes the separation of three domain of life: Bacteria, Eukarya and Archaea.

Methanogenesis is a form of anaerobic respiration (*8*). Methanogens can use carbon dioxide, acetic acid, methanol, and/or methylamines as electron acceptors in energy metabolism (Scheme 1), which is an adaptation to life in anaerobic environments. The biochemistry of methanogenesis is relatively complex. Many coenzymes are involved in methanogenesis, including F420, coenzyme B, coenzyme M, methanofuran, and methanopterin. There are over twenty enzymes that participate in the superpathway of methanogenesis, with many containing multiple subunits and metallocofactors (Scheme 2). The key step of methanogenesis is catalyzed

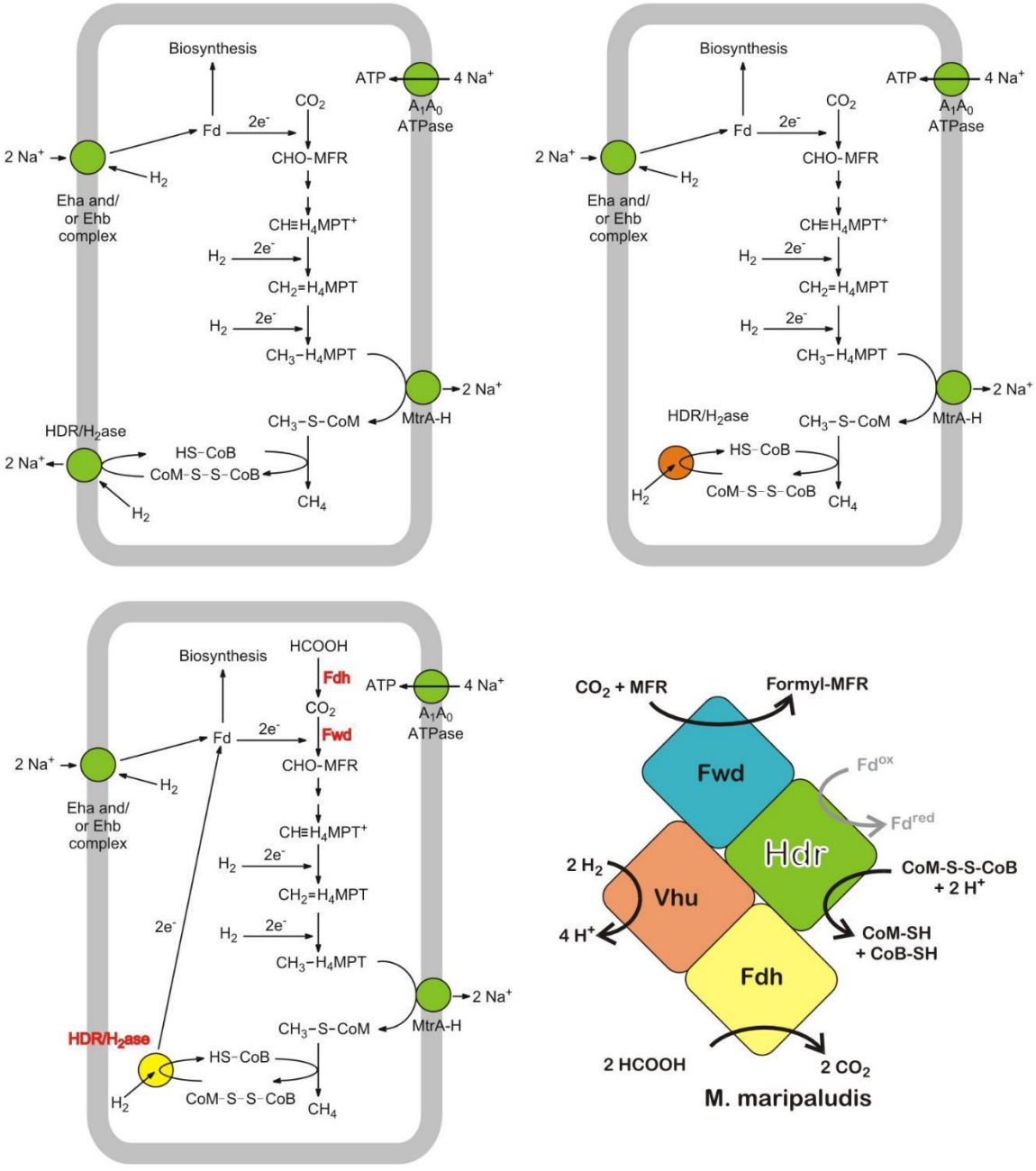
by methyl-coenzyme M reductase (MCR), which converts methyl-coenzyme M and coenzyme B to methane and the coenzyme M-coenzyme B heterodisulfide. Methanogens couple the heterodisulfide reductase (HDR) catalytic reaction with hydrogen oxidation and proton pumping by a hydrogenase-HDR complex (Scheme 3) (*8*). Also, methyltransfer reactions that produce methyl-coenzyme M are coupled by membrane-associated methyltransferases to sodium ion pumping. The sodium/proton concentration gradient across the memberane provides the driving force for adenosine triphosphate (ATP) generation via an ATP synthase (*8*).



Scheme 1: Four overall reactions of methanogensis. Carbon dioxide, acetic acid, methanol, and methylamine can serve as the electron acceptor for methane formation.



Scheme 2: Methanogenesis super pathway.



Scheme 3: The relationship between methanogenesis and ATP generation in methanogens.

1.2 Methyl-coenzyme M reductase and methanogenesis

Methyl-coenzyme M reductase (MCR), also named coenzyme B sulfoethylthiotransferase, is an enzyme that catalyzes the final step of methanogenesis (*9*). All electron acceptors and methyl group donors in methanogenesis, including carbon dioxide, acetic acid, methanol, and methylamines, transfer their methyl group (or carbon that will be reduced to a methyl group) by different pathways to coenzyme M. There are three main branches in this part of methane metabolism (Scheme 2). In the acetate branch, an acetyl group is ligated to coenzyme A enzymatically to form acetyl-coenzyme A firstly (*10*). Then, the methyl group from acetyl-coenzyme A is transferred to a corrinoid. At the same time, one molecule of carbon-monoxide is released from the reaction above (*11*). Carbon monoxide can then be enzymatically converted to carbon dioxide, another electron acceptor in methanogensis by carbon-monoxide dehydrogenase (*11*). In the carbon dioxide branch, the major branch of methanogenesis, the carbon dioxide is

transferred to different cofactors and reduced. These cofactors include methanofuran and tetrahydromethanopterin (*12*) (*13*). In the third branch, methanogens utilize methylamine or methanol as methyl group donors. A Co(I) methanol or methylamine specific corrinoid protein functions as the methyl transporter in this branch (*14*) (*15*) (*16*) (*17*).

All branches of methanogenesis result in methyl group transfer to coenzyme M to form methyl coenzyme M, which is one of the substrates of MCR. Another substrate is coenzyme B,

which is biosynthesized from 2-oxoglutarate (*18*). MCR catalyzes the reaction that ligates methyl coenzyme M and coenzyme B, resulting in the releasing of a molecule of methane (*19*).

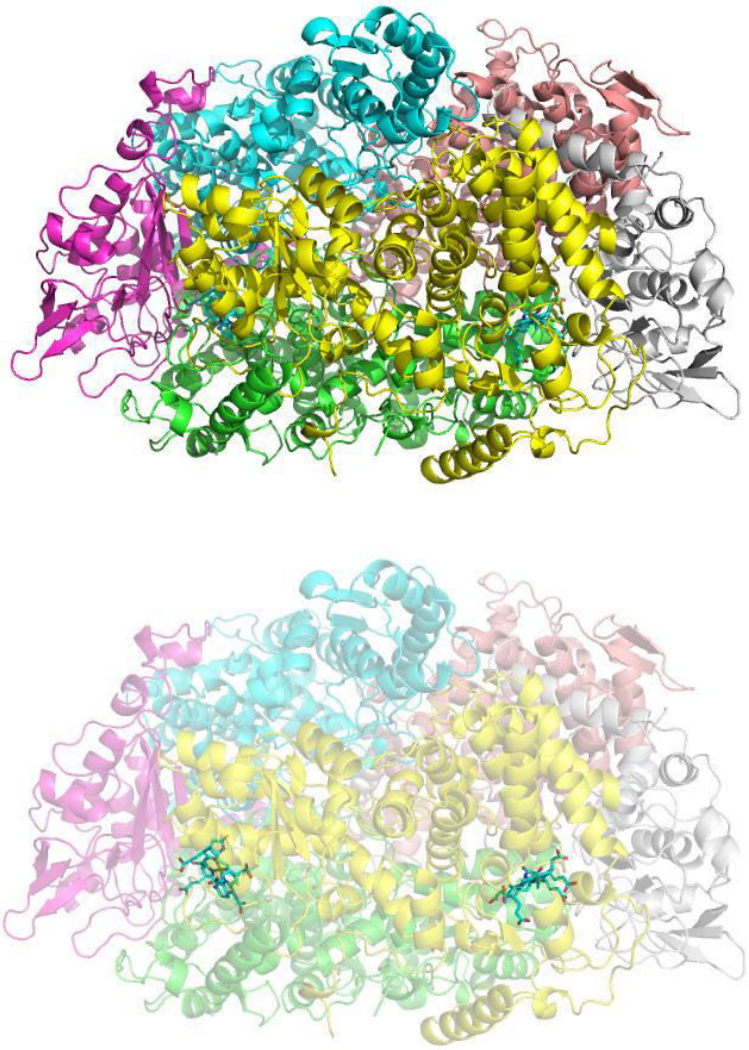


Figure 2: Structure of MCR (1MRO). Individual subunits are indicated by different colors. MCR is a 270 kDa complex that has two active sites. Coenzyme F430 is shown as a stick model.

Structural studies show MCR as a hexamer with two α, two β, and two γ subunits (Figure 2). There are four parts in the α subunit: An N-terminal region, an α + β domain, an α helical domain, and a C-terminal region. The structure of the α + β domain is similar to the α, β sandwich motif of the βαββαβ class (*20*), which has been found in formyltransferases (*21*) (*20*). The α helical domain is composed of eight α helices of differing length (*20*). The general architecture of the β subunit is similar in composition to that of the α subunit (*20*). The γ subunit is mainly built by the α, β sandwich motif of the βαββαβ class (*20*). MCR contains two nickel-containing tetrapyrroles, coenzyme F430 (*22*). There are five unprecedented post-translational modifications (PTMs) surrounding the active site: 2-(*S*)-methylglutamine, 5-(*S*)-methylarginine, 3-methylhistidine, *S*-methylcysteine, and thioglycine residues (Figure 3).

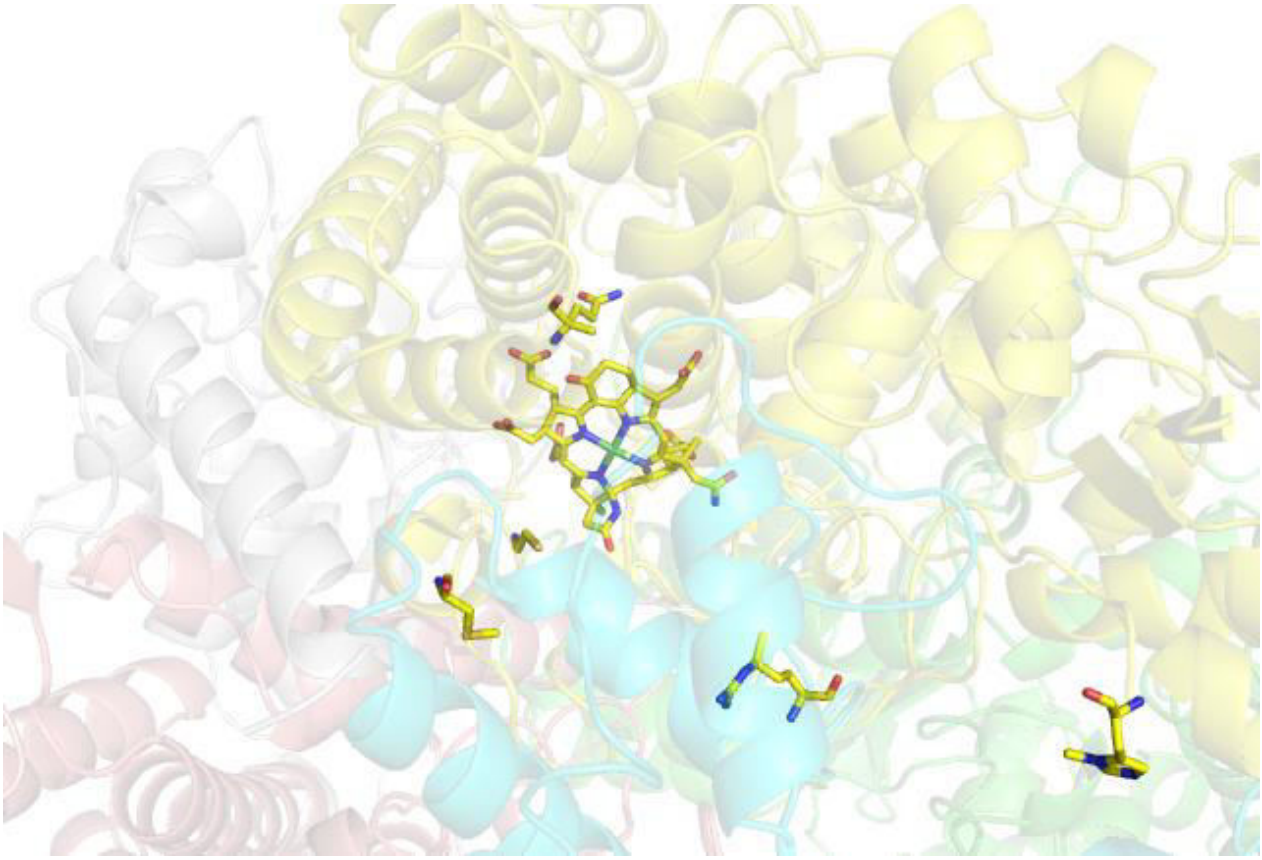


Figure 3: The coenzyme F430 binding site of MCR. Coenzyme F430, a nickel chelated tetrapyrrole, is located between the α and γ subunits.

1.3 Homolog of MCR in anaerobic methanotrophic archaea

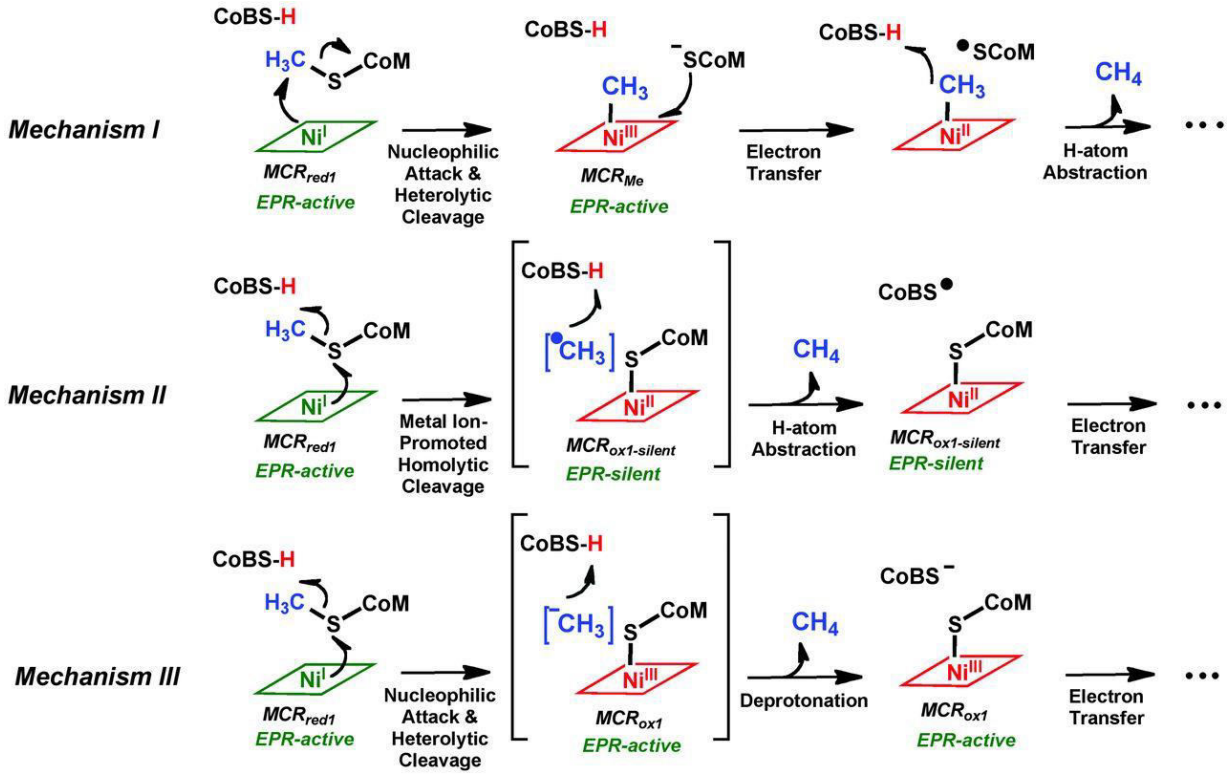
Homologs of MCR have been found in anaerobic methanotrophic archaea (ANME), which catalyze the anaerobic oxidation of methane (AOM) (*23*). Methane can be oxidized in this process using different electron acceptors, such as sulfate, nitrate, nitrite, and metals (*24*). Based on genomic data, it is believed that AOM is the reverse of methanogenesis, with MCR catalyzing the initial methane oxidizing step. There are some notable differences between authentic MCR

and ANME-1 MCR. ANME-1 MCR contains a modified F430 cofactor, 17^2-methylthio-F430. It also contains a different set of post-translationally modified amino acids, including 1-*N*-methylhistidine, 7-hydroxytryptophan, thioglycine and *S*-oxymethionine (*25*) (*26*).

1.4 Mechanism and activation of MCR

As discussed previously, MCR catalyzes the reaction between methyl-coenzyme M and coenzyme B. As the result, a molecule of CoB-S-S-CoM (the heterodisulfide, HDS) is produced with a molecule of methane. The biochemical standard Gibbs free energy change of this reaction is about -30 ± 10 kJ/mol (*27*), which indicates that it is a highly exergonic reaction. Studies showed that the Ni(II)-MCR, which is silent in electronic paramagnetic resonance (EPR) spectroscopy, is inactive (*28*). Ni(I)-MCR is the active form of MCR, and several different catalytic mechanisms have been proposed (Scheme 4). Currently, the catalytic mechanism of MCR is thought to proceed via a radical mechanism involving Ni(II)-thiolate and methyl radical intermediates (*29*). Ragsdale and coworkers used rapid kinetic approaches rule out the methyl-Ni(III) intermediate (*29*). Ultraviolet-visible (UV-vis) kinetic results demonstrated that the conversion rate of the methyl group of methyl-CoM to methane is the same rate as that of Ni(I) decay (*29*). Analysis of rapid freeze-quench (RFQ) EPR data reveals that no EPR active species accumulated with an amplitude similar to MCR_{red1} decay. This observation is consistent with mechanism II (Scheme 4), which involves a Ni(II) intermediate (*29*). Circular dichroism (CD) and magnetic circular dichroism (MCD) experiments also support this hypothesis. The data reveals an almost quantative conversion of MCR_{red1} to a species nearly identical to $MCR_{ox1\text{-}silent}$

(*29*). Density function theory (DFT) computations also suggest that mechanism II is thermodynamically favored (*29*).



Scheme 4: Proposed mechanisms of MCR-catalyzed methane formation. Current experimental and computational evidence support mechanism II.

Since nickel is normally found as Ni(II), a reductase is needed for MCR activation after F430 has been delivered into the active sites. In 2014, Duin and coworkers found that under the presence of an A2 complex, an A3a complex, an ATP carrier, and dithiothreitol (DTT), MCR can be reduced to its active state (*30*). The A3a complex is composed of several subunits, including an Fe-protein homolog, an iron-sulfur flavoprotein, protein components involved in

electron bifurcation, and McrC. The A2 complex has been suggested to be an ATP transporter that is delivering ATP to the Fe protein in the A3a complex (*30*). The A2 complex contains two ATP binding domains that have sequence similarities to the ATP-binding cassette family of transport systems (Figure 4) (*30*).

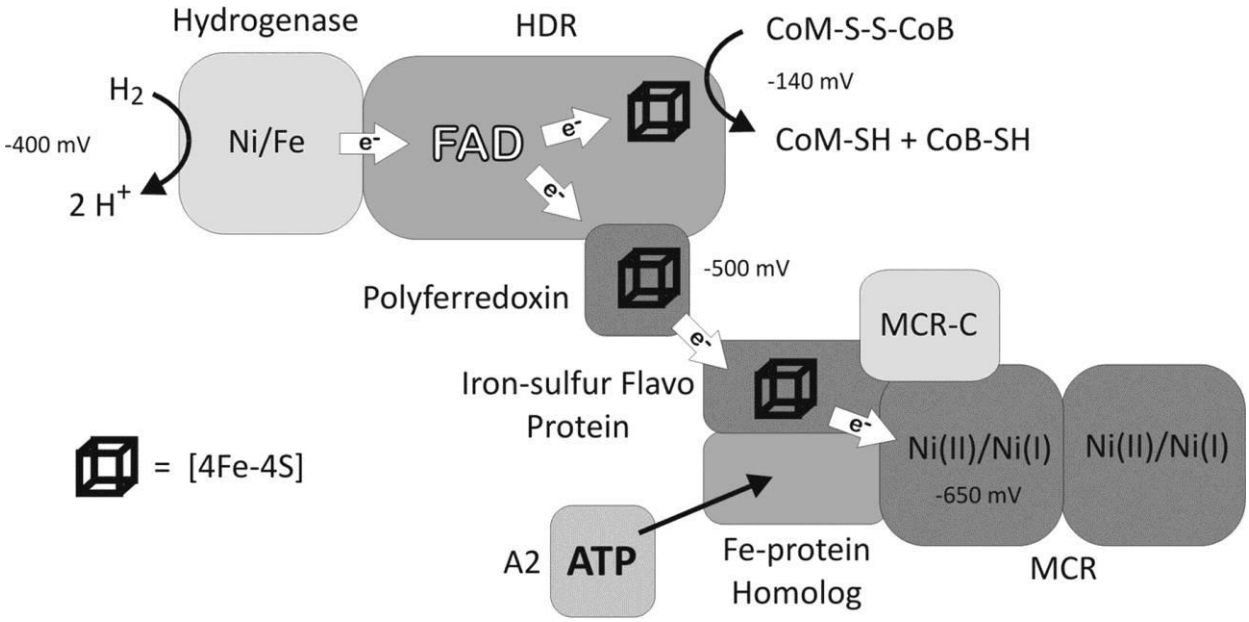


Figure 4: The process of MCR activation. Reducing power is provided by hydrogenase. The electons are then transferred through the heterodisulfide reductase (HDR), polyferredoxin, and the iron-sulfur flavoprotein before been used for reducing the Ni(II) of coenzyme F430 to Ni(I).

1.5 Reverse MCR reaction

ANME have been proposed to carry out the reverse (or partial reverse) of methanogenesis. The first step of these reactions is the oxidation of methane with CoB-S-S-CoM, which is an endergonic process (*23*) (*31*). Therefore, the reduction of a favorable electron

acceptor is needed in order to drive the oxidation of methane (*32*). This requirement explains why ANMEs often have a symbiotic syntrophic relationship with (*33*). The favorable reduction of sulfate by the SRB via direct interspecies electron transfer (DIET) provides the capability of reverse methanogenesis in the ANME (*34*).

1.6 Coenzyme F430 and tetrapyrrole biosynthesis

As discussed previously, MCR molecule contains two equivalents of coenzyme F430, a nickel-chelated, highly-reduced tetrapyrrole. Tetrapyrroles are a series of compounds that contain four pyrrole-type rings. These four rings are connected by single carbon bridges at the α position of each five-membered ring. Tetrapyrroles usual coordinate a metal ion and function as cofactors for various enzymes. There are two different, organism-specific precursors in tetrapyrrole biosynthesis: L-glutamate and glycine. Glycine has been found as the precursor in the tetrapyrrole biosynthetic pathways of animals, fungi, apicomplexan protozoa, and members of the α-proteobacteria. In other bacteria and archaea, tetrapyrrole biosynthesis generally begins with L-glutamate.

In the glycine branch of tetrapyrrole biosynthesis, the pathway begins with the decarboxylation of glycine and nucleophilic attack on succinyl-CoA from the tricarboxylic acid (TCA) cycle. This enzymatic reaction results in the formation of 5-aminolevulinic acid (5-ALA) (*35*). Enzymes capable of catalyzing this 5-aminolevulinic acid synthase reaction include HemA and HemT (*36*). 5-ALA provides an ideal compound for construction of pyrrole rings. Two molecule of 5-ALA undergo a condensation and cyclization reaction, catalyzed by

porphobilinogen synthase (HemB), to form a functionalized pyrrole containing propionate, acetate, and aminomethyl side chains (*37*). Then four porphobilinogen molecules link via a condensation reaction to produce preuroporphyrinogen (hydroxymethylbilane). This reaction is catalyzed by HemC, a porphobilinogen deaminase (*38*). Preuroporphyrinogen is the precursor of uroporphyrinogen III, the last common precursor across tetrapyrrole biosynthetic pathways. A dehydration and ring closing and inversion reaction catalyzed by HemD results in the formation of uroporphyrinogen III (*39*).

The glutamate branch of tetrapyrrole biosynthesis pathway is different from the glycine branch at several of the beginning steps. L-Glutamate must be ligated to $tRNA^{Glu}$ first before any other modification. This process is catalyzed by glutamate-tRNA ligase (GltX) (*40*). Then, a glutamyl-tRNA reductase (HemA) catalyzes the reduction of the ester linkage to an aldehyde, producing (*S*)-4-amino-5-oxopentanoate (*41*). 5-ALA is then biosynthesized from (*S*)-4-amino-5-oxopentanoate by glutamate-1-semialdehyde aminotransferase (HemL) (*42*).

The two tetrapyrrole biosynthesis pathways converge at 5-ALA and continue on through the synthesis of uroporphyrinogen III, at which point the various pathways diverge once more. One branch produces precorrin 2, which is biosynthesized from uroporphyrinogen III by a methyltransfer reaction catalyzed by uroporphyrinogen III *C*-methyltransferase (SirA) (*43*). Precorrin 2 is the biosynthetic precursor of heme d_1, siroheme, cobalamin, and coenzyme F430. A second branch produces coproporphyrinogen, resulting from the decarboxylation of uroporphyrinogen III. This is an enzymatic reaction carried out by uroporphyrinogen III

decarboxylase (HemE) (*44*). Coproporphyrinogen is the biosynthetic precursor of heme, chlorophyll, bacteriochlorophyll, and dinoflagellate luciferin (Figure 5).

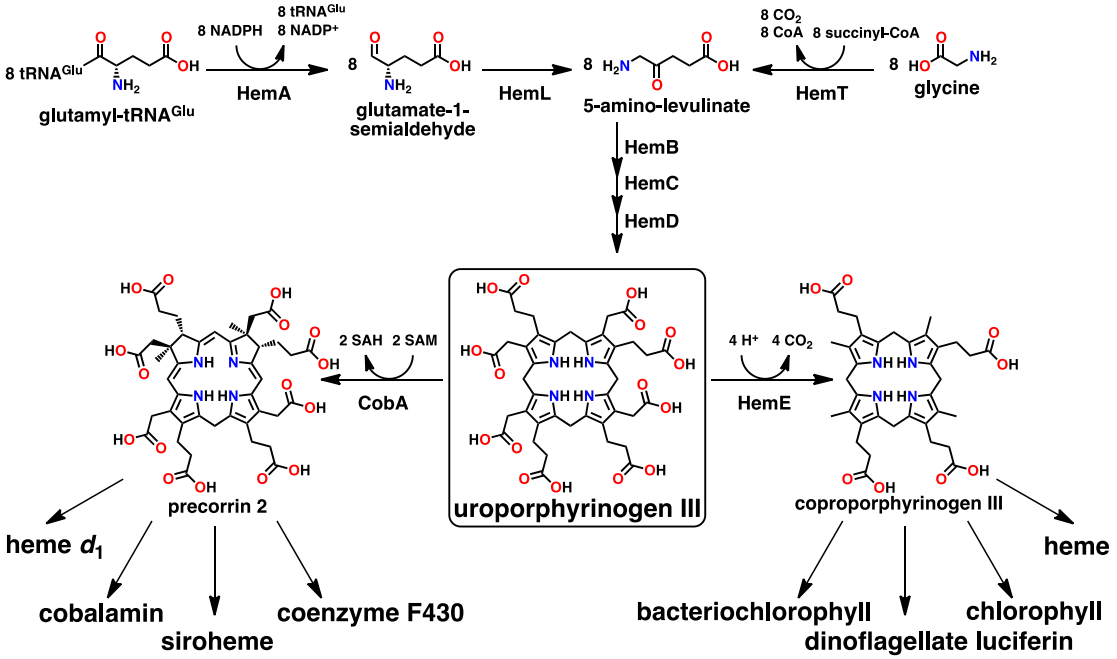


Figure 5: Tetrapyrrole biosynthesis pathways. 5-Amino-levulinate, which is the precursor of uroporphyrinogen III, can be biosynthesized from glycine or glutamate. Precorrin 2 and coproporphyrinogen III are synthesized from uroporphyrinogen III. All known tetrepyrroles can be biosynthesized from these three compounds.

1.7 Coenzyme F430 biosynthesis

In order to locate the genes encoding the enzymes related to coenzyme F430 biosynthesis (*cfb*), we utilized a comparative genomics strategy. Currently, the genomes of over thirty species of methanogenic archaea have been sequenced. We screened the genomic contexts of chelatase homologs (which are responsible for metal ion insertion into tetrapyrroles) for encoded enzymes catalyzing the requisite chemistry for F430 biosynthesis from known pathway intermediates,

such as precorrin 2 or sirohydrochlorin. Differences between the proposed precursors precorrin 2 or sirohydrochlorin and F430 are: 1) nickel chelation, 2) two amidations, 3) two cyclizations, and 4) a multi-electron ring reduction. The analysis revealed a homolog of sirohydrochlorin cobaltochelatase (*cbiX*) and four additional genes conserved in all methanogens that are excellent candidates for coenzyme F430 biosynthesis. They include homologs of cobyrinic acid *a*,*c*-diamide synthase (*cbiA*), nitrogenase (*nifD*/*nifH*), and *N*-acetylmuramoyl-L-alanine:D-glutamate ligase (*murD*). One-third of the species we scanned contained these four genes clustered with the *cbiX* homolog. The genes are partially clustered in another one-third of species and completely scattered in the remaining genomes.

1.8 Sirohydrochlorin Cobaltochelatase

Many chelatases have been sequenced and characterized, including some chelatases involved in tetrapyrrole metal chelation. The identified homolog of sirohydrochlorin cobaltochelatase (MA3631 in *Methanosarcina acetivorans*) potentially is the first nickel chelatase and belongs to the class II chelatases. There are three main classes of chelatases. The class I chelatases are comprised of three subunits and couple the hydrolysis of adenosine triphosphate (ATP) to metal ion insertion. These enzymes, which include the magnesium chelatase (ChlHID) in chlorophyll biosynthesis and the cobaltochelatase (CobNST) in the aerobic adenosylcobalamin biosynthetic pathway, belong to the chaperone-like (AAA) family of ATPases (*45*). The class II chelatases are either "long" monomers containing two chelatase domains or "short" homodimers. They are ATP-independent chelatases and include the

ferrochelatases in heme and siroheme biosynthesis and the cobaltochelatase from the anaerobic adenosylcobalamin biosynthetic pathway. The third type of chelatases are ATP-independent multifunctional enzymes, such as siroheme synthase (CysG), which catalyzes methylation and dehydrogenation in addition to ferrochelation (Figure 6)(*46*).

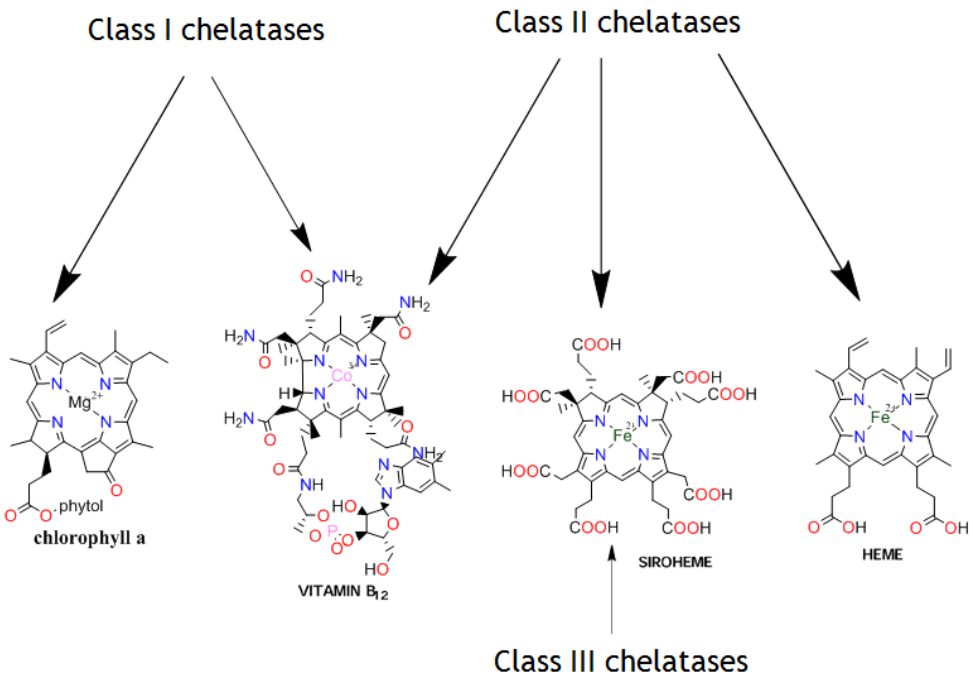
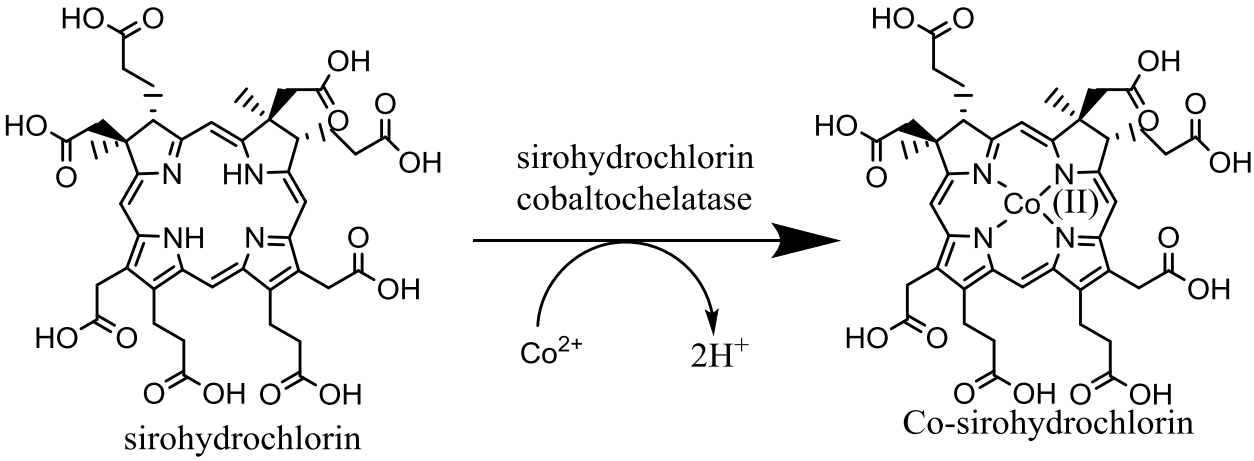


Figure 6: The 3 different types of chelatases. The class I chelatases are comprised of three subunits and couple the hydrolysis of adenosine triphosphate (ATP) to metal ion insertion. The class II chelatases are either "long" monomers containing two chelatase domains or "short" homodimers. The third type of chelatases are ATP-independent multifunctional enzymes.

Sirohydrochlorin cobaltochelatase (CbiX) is a dimeric class II chelatase (Figure 7) (*45*). It chelates Co^{2+} into sirohydrochlorin, forming Co-sirohydrochlorin and releasing two protons

(Scheme 5). This enzyme is a lyase. The systematic name of this enzyme is cobalt-sirohydrochlorin cobalt-lyase (sirohydrochlorin-forming) (*46*).



Scheme 5: Enzymatic reaction catalyzed by sirohydrochlorin cobaltochelatase.

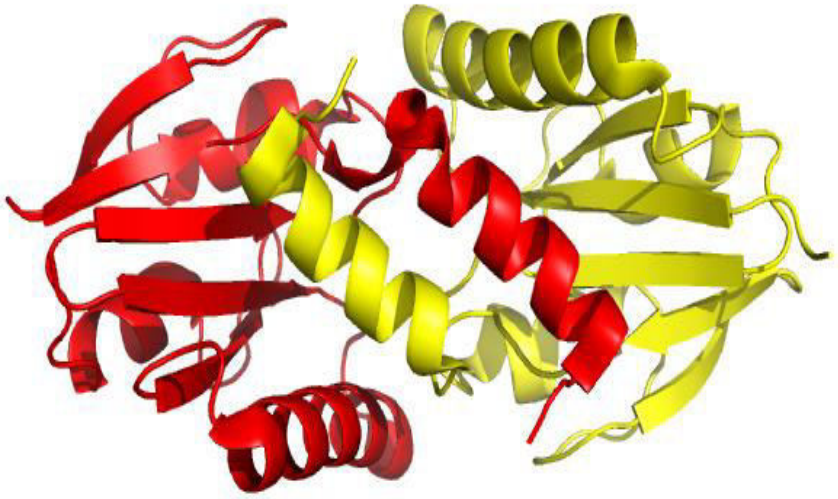
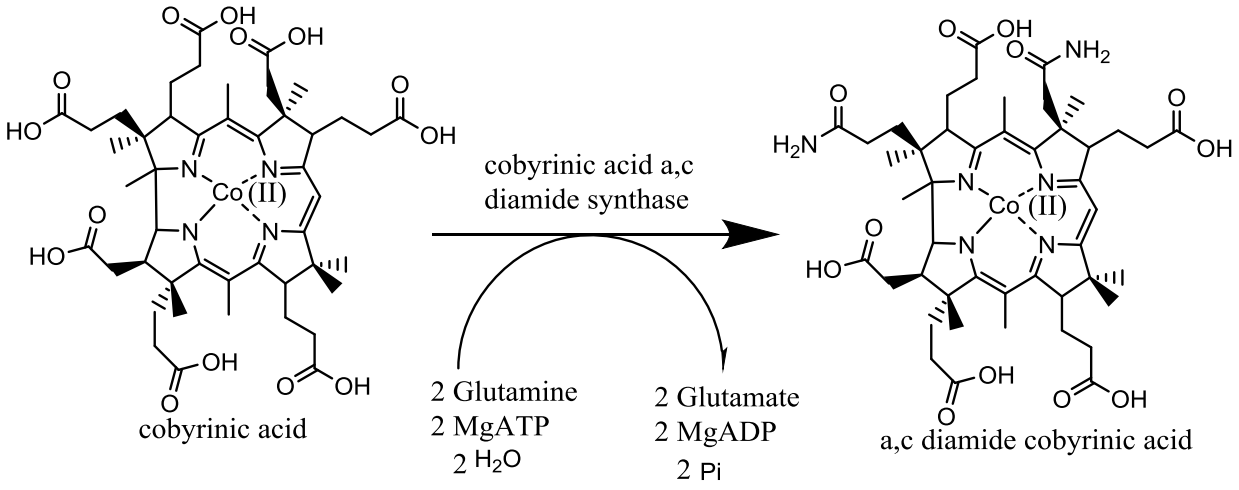


Figure 7: A structure of the homodimeric sirohydrochlorin cobaltochelatase (2DJ5).

1.9 Cobyrinic acid *a*,*c*-diamide synthase

MA3626 encodes a homolog of cobyrinic acid *a*,*c*-diamide synthase (CbiA) that is located at the end of the *cfb* cluster. Homologs of CbiA have been found in several biosynthetic pathways, including the vitamin B_{12} biosynthesis pathway. CbiA catalyzes the amidation of the *a* and *c* acetate side chains of cobyrinic acid (*47*) (*48*) (Scheme 6).



Scheme 6: Enzymatic reaction catalyzed by cobyrinic acid *a*,*c*-diamide synthase (CbiA).

CbiA is the first glutamine amidotransferase in the anaerobic adenosylcobalamin biosynthetic pathway. This enzyme has multiple active sites. One active site is in the C-terminal domain and catalyzes the hydrolysis of glutamine. This active site contains a histidine that activates a cysteine residue for nucleophilic attack on glutamine to release ammonia (*47*). The other active site, which is located in the N-terminal domain, utilizes ATP to phosphorylate the *a*-

and *c*-carboxylate oxygens of cobyrinic acid to activate them as leaving groups for subsequent amidation using the ammonia liberated by the glutaminase domain (*47*).

1.10 Nitrogenase

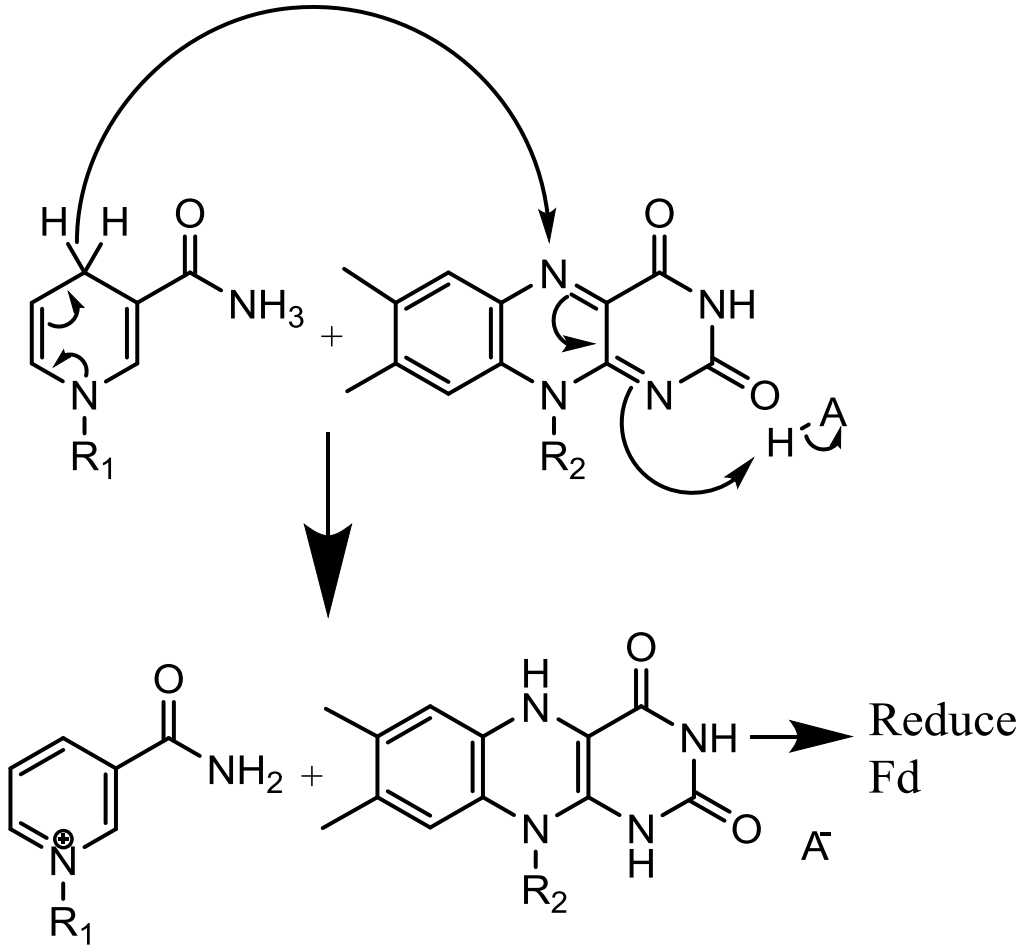
MA3627 and MA3628 encode a complex homologous to nitrogenase. Nitrogenase is involved in the N_2 fixation pathway (Scheme 7) (*49*). Both a reductant and chemical energy in the form of ATP are required to reduce dinitrogen to ammonia. Nitrogenase is a two-component metalloenzyme consisting of the catalytic MoFe protein ($NifK_2$-$NifD_2$) and its reductase, the Fe protein ($NifH_2$). Homologs of nitrogenase are also found in the biosynthetic pathways for the photosynthetic pigments chlorophyll and bacteriochlorophyll. These include the dark-operative protochlorophyllide *a* (DPOR) and chlorophyllide *a* oxidoreductases (COR). DPOR and COR both have similar structural topologies to nitrogenase and are comprised of BchLNB and BchXYZ subunits, which are homologous to the NifHDK subunits of nitrogenase, respectively.

$$N_2 + 8H^+ + 8e^- + 16ATP \xrightarrow{\text{nitrogenase}} 2NH_3 + H_2 + 16ADP + 16Pi$$

Scheme 7: The overall reaction of nitrogen fixation.

Nitrogenase-catalyzed N_2 reduction involves repeated cycles of Fe and MoFe protein association, intercomponent electron transfer, ATP hydrolysis, complex disassociation, and Fe protein reduction by ferredoxin/flavodoxin. Ferredoxins and flavodoxins can be reduced by numerous physiological electron donors, including Nicotinamide adenine dinucleotide phosphate

(NADPH) via ferredoxin $NADP^+$ reductase (Fnr). Flavin adenine dinucleotide (FAD) is the intermediate of this electron transfer reaction (Scheme 8) (*50*). There is a glutamate residue and a serine residue in the active site of Fnr. When Fnr binds to ferredoxin, a hydrogen bond forms between these two residues. This highly conserved glutamate residue both stabilizes the semiquinone form of FAD and serves as a proton donor/acceptor in the reaction (*51*) (*52*).



Scheme 8: Fnr-catalyzed transfer of electrons from NADPH to ferredoxin. Flavin adenine dinucleotide (FAD) is an intermediary in this electron transfer reaction.

Yi-ping Wang and his group devised an electron-transport component (ETC) model to describe the three main modules in systems of biological nitrogen fixation (BNF). They utilized *Escherichia coli* as a chassis to study the compatibility between molybdenum and iron-only nitrogenases with ETC modules from different organelles. In this study, they found that different modules in ETC can be replaced by similar modules from other species and still retain functionality (*53*) (Figure 8). This study has implications for the engineering of functional nitrogenase homologs (such as MA3627/MA3628) in a heterologous host.

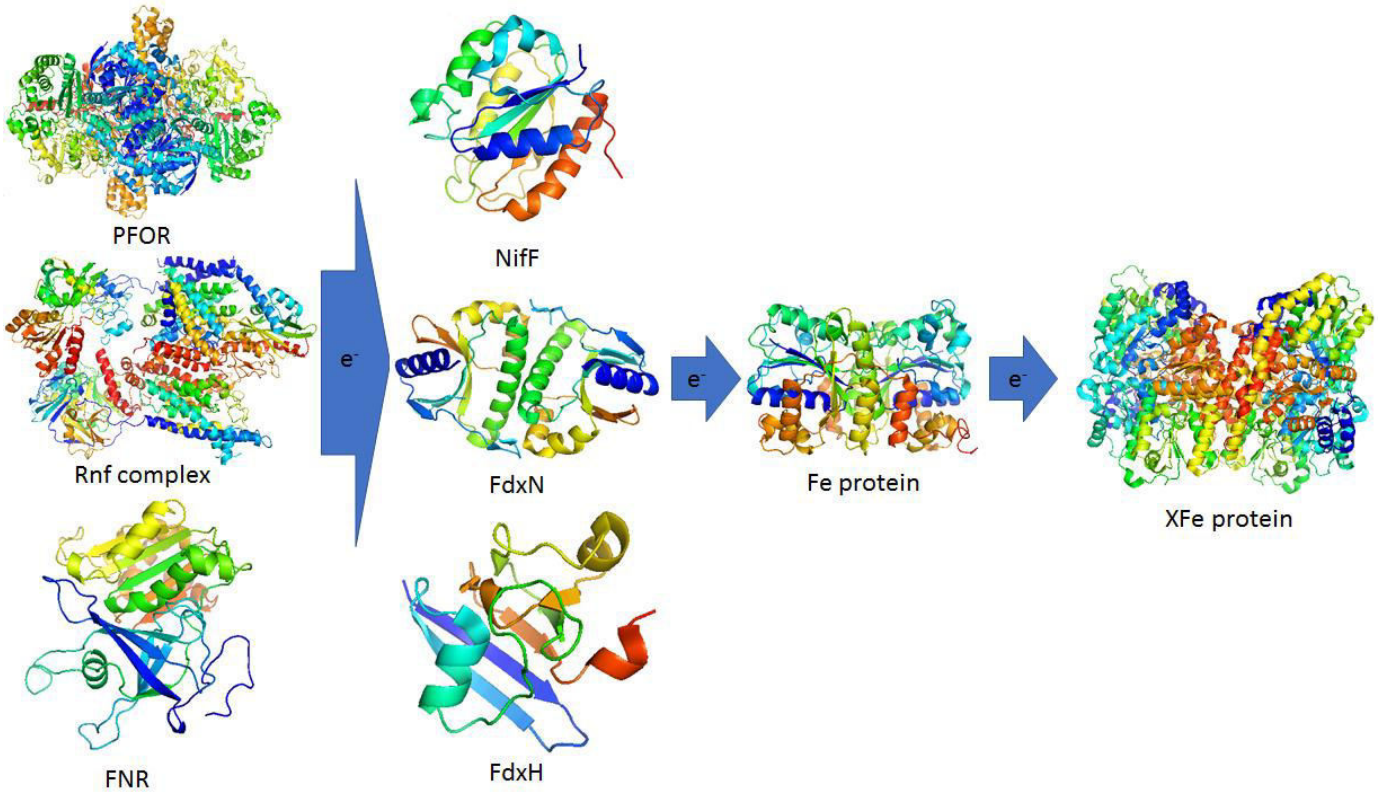
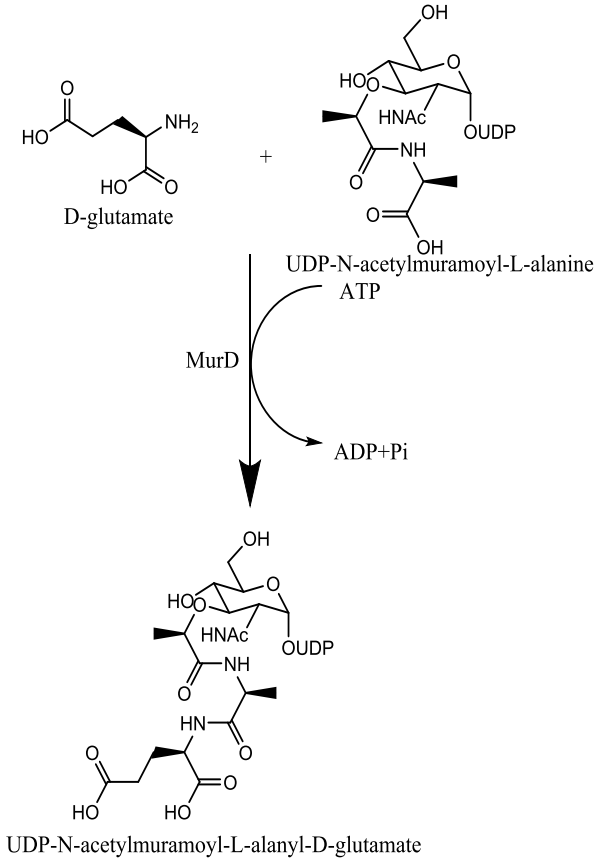


Figure 8: Electron-transport component (ETC) model. Different ETC modules for nitrogenase can be replaced by similar modules from other species and still retain functionality.

1.11 Mur Ligase

UDP-*N*-acetylmuramoyl-L-alanine:D-glutamate ligase (MurD) catalyzes the ATP-dependent formation of a peptide bond between UDP-*N*-acetylmuramoyl-L-alanine (UMA) and D-glutamate during peptidoglycan biosynthesis (Scheme 9). Mur ligases, such as MurD, share a similar three-domain structure. Domain 1 binds the nucleotide substrate, domain 2 binds ATP, and domain 3 binds the amino acid substrate (*54*).



Scheme 9: UDP-*N*-acetylmuramoyl-L-alanine:D-glutamate ligase (MurD) catalyzes the ATP-dependent formation of a peptide bond between UDP-*N*-acetylmuramoyl-L-alanine (UMA) and D-glutamate during peptidoglycan biosynthesis.

1.12 Post-translational modifications of MCR

Ermler and coworkers determined the first MCR high-resolution crystal structure, which revealed five PTMs in MCR α subunit: 1-*N*-methylhistidine, 5-(*S*)-methylarginine, 2-(*S*)-methylglutamine, thioglycine, and *S*-methylcysteine residues (*22*). These PTMs are not present in all methanogens. MCR in *Methanopyrus kandleri* and *Methanocaldococcus jannaschii* lack the *S*-methylcysteine residue. *Methanosarcina barkeri* does not conserve the 2-(*S*)-methylglutamine residue, but a compensating alanine-to-threonine substitution in a nearby position orders a water molecule near the site occupied by the 2-methyl group (*55*). Because of the lower resolution of this structure (2.7 Å), compensatory changes caused by the absence of the *S*-methylcysteine in the active site region were less discernable than in *M. kandleri*.

We propose that several genes (*mm1*, *mm3*, *mm10*, *tfuA*, and *prmA*) are involved in the PTMs of MCR. The 1-*N*-methylhistdine and *S*-methylcysteine modifications are likely accomplished via methyl cation transfer from S-adenosyl-L-methionine (SAM) (*56*). Adjacent to the *mcr* cluster in *Methanosarcina spp.* is a gene, *prmA*, encoding a ribosomal protein L11 methyltransferase homolog that may be responsible for one or both of these modifications. The 2-(*S*)-methylglutamine and 5-(*S*)-methylarginine modifications are different in that their formation involves *C*-methylation of an unactivated C-H bond, likely involving a radical mechanism. A member of the radical SAM superfamily, methanogenesis maker 10 (*mm10*), is conserved in all methanogens and is typically found adjacent to the *mcr* gene cluster. We propose that it encodes an enzyme that catalyzes the formation of the 5-(*S*)-methylarginine

residue. Another methanogenesis marker (*mm1*), which encodes a YcaO superfamily protein, is clustered with a *tfuA* homolog. A TfuA-associated YcaO homolog is involved in the production of the thioamide-containing natural product thioviridamide. Therefore, these two genes may be involved in the formation of the thioglycine residue.

We also noticed that P162 in the MCR β subunit (*M. acetivorans* numbering) is a *cis*-proline. Interestingly, the methanogenesis marker 3 (*mm3*) gene encodes a hypothetical protein that has a conserved domain belonging to the cyclophilin superfamily. This family contain proteins that exhibit peptidylprolyl *cis-trans* isomerase (PPIase) activity in all 3 domains of life. PPIases can catalyze the *cis-trans* isomerization reaction for proline imidic peptide bonds to help protein folding (*106*) (*107*). Thus, we hypothesize that Mm3 is a *cis*-prolyl isomerase that catalyzes the proline *cis-trans* isomerization reaction in the MCR β subunit.

1.13 1-*N*-Methylhistidine and *S*-methylcysteine

As noted, the 1-*N*-methylhistidine and *S*-methylcysteine PTMs are likely to proceed via methyl cation transfer from SAM to the activated N and S of the corresponding deprotonated amino acids (*56*). An enzyme involved in ribosome PTMs, ribosomal protein L11 methyltransferase (PrmA), proceeds via an analogous enzymatic mechanism. There are 34 proteins belonging to the 50S large subunit of the prokaryotic ribosome, and there are 49 proteins belonging to the 60S large subunit of the eukaryotic ribosome, which are critical for proteins synthesis. Ribosomal protein L11 (RPL11) is a component conserved in the large subunits of all ribosomes, and there is evidence showing that it binds to 23S rRNA (*57*). Before RPL11

participates in the assembly of the ribosomal large subunit, a couple of amino acid residues must be methylated. PrmA is responsible for multiple methylations in RPL11 (*58*). As noted, there is a homolog of *prmA* conserved in all methanogens close to the *mcr* gene cluster, providing a good candidate to encode a methylase for 1-*N*-methylhistidine and/or *S*-methylcysteine.

1.14 2-(*S*)-Methylglutamine and 5-(*S*)-methylarginine

The 2-(*S*)-methylglutamine and 5-(*S*)-methylarginine PTMs involve *C*-methylation. These methylated amino acid are uniquely found in MCR (*56*). The hydrogens in the methylation sites of these amino acid are hard to deprotonate. The pK_a of the α-proton of an amino acid has been estimated to be ~32 in a peptide, which is larger than that of the free amino acid due to the lack of the positively charged ammonium group and the weaker electron-withdrawing effects of a carbonyl from an amide (rather than a carboxylic acid) (*59*)(*60*). If the 2-(*S*)-methylglutamine modification is to be biosynthesized through the methyl cation mechanism described above, the α-proton must be removed, generating an enolate anion. The enolate can then accept the methyl cation from SAM. This pK_a is close to the maximum value known for enzymatic proton transfer reactions. Likewise, the pK_a for the δ-proton of arginine is expected to be similarly unactivated. These *C*-methyl modifications are therefore unlikely to proceed via polar mechanisms since the energetic requirements are prohibitive.

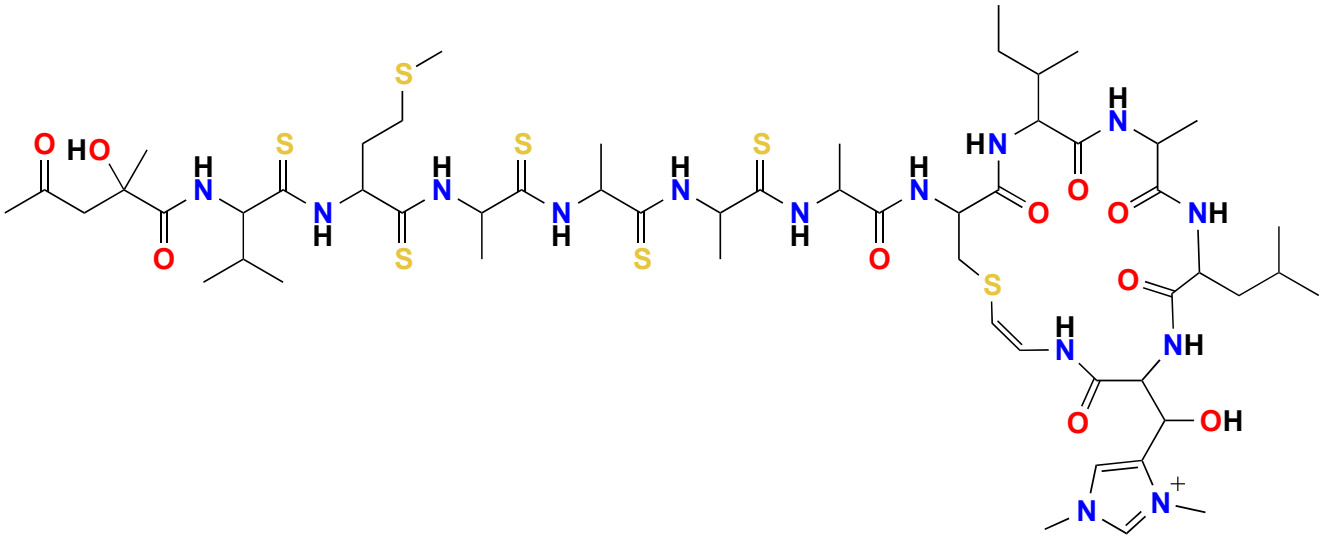
Instead of the methyl cation mechanism, 2-(*S*)-methylglutamine and 5-(*S*)-methylarginine are likely to be generated via a radical mechanism, which is only known to be catalyzed by members of the radical SAM superfamily (*61*). The radical SAM superfamily is a

large superfamily of enzymes that contains over 113,000 members. More than 10,000 are thought to be methyltransferases. To date, four classes of radical SAM methyltransferases have been discovered and characterized. They are classified based on protein architecture, cofactor requirements, and proposed mechanisms (*61*)(*62*). Class A proteins have a single canonical radical core domain that includes a $(\beta\alpha)_6$ partial barrel. They are mainly rRNA methyltransferase. Class B proteins contain a N-terminal cobalamin binding domain and a C-terminal radical SAM domain, which conserves a canonical CxxxCxxC motif. This motif can bind a [4Fe-4S] cluster that binds and reduces SAM. Class C proteins share significant sequence similarity with coproporphyrinogen III oxygenase (HemN). Class D proteins use methylenetetrahydrofolate as the source of methyl carbon. They contain a second auxiliary iron-sulfur cluster and a N^5,N^{10}-methylenetetrahydrofolate-binding site. Class B and class D radical SAM methyltransferases are capable of methylating sp^3-hybridized centers like the α-carbon of glutamine or the δ-carbon of arginine. It is possible that one of these classes of methyltransferases is responsible for the 2-(*S*)-methylglutamine and/or 5-(*S*)-methylarginine PTMs. Another possibility would be a member of the radical SAM superfamily utilizing novel chemistry to catalyze these PTMs.

A gene encoding a radical SAM enzyme conserved in all methanogens, methanogenesis marker 10 (*mm10*), is often adjacent to, and divergently transcribed from, the *mcr* gene cluster. Sequence analysis of *mm10* reveals that the encoded enzyme has a conserved CxxCxPxxxGCxxC (or CxxCxxxxGCxYC) motif, which may bind an auxiliary iron-sulfur cluster in an N-terminal radical SAM domain. It also contains a C-terminal domain of unknown function.

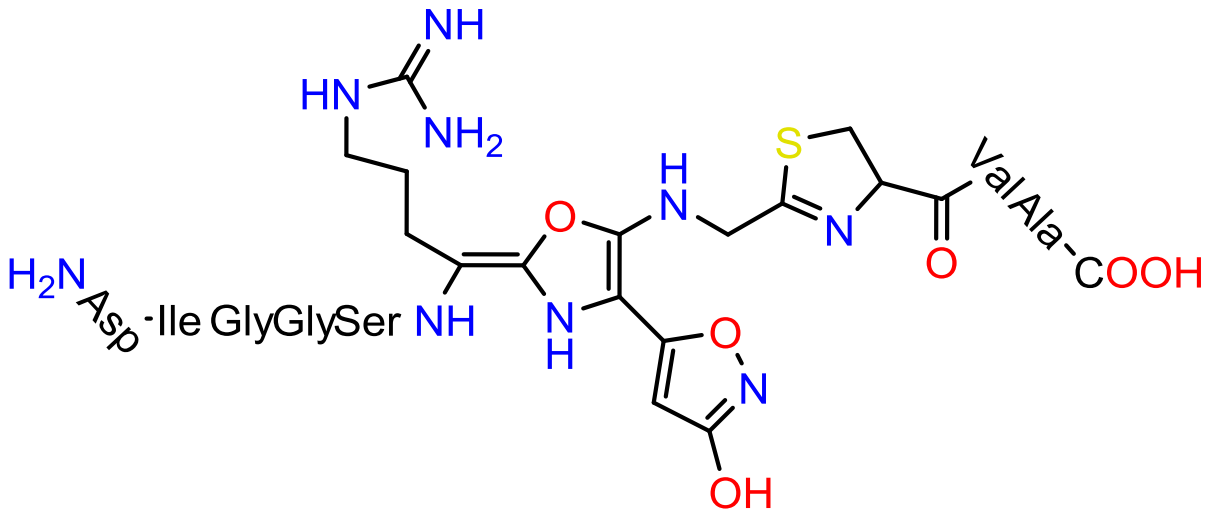
1.15 Thioglycine

The thioglycine residue has only been observed in MCR (*56*). However, similar modifications are reported by Hayakawa and coworkers in thioviridamide, a thiopeptide-containing natural product (*63*). Thioviridamide contains thiovaline, thiomethionine, and three thioalanine residues (Scheme 10) (*64*). The thioviridamide (*tva*) biosynthetic gene cluster from *Streptomyces olivoviridis* contains 15 genes. A TfuA-associated YcaO homolog in this cluster has been implicated in the installation of the thioamide moieties (*65*).



Scheme 10: The stucture of thioviridamide.

Another genetic marker of methanogenesis, methanogenesis marker 1 (*mm1*), encodes a homolog of the YcaO superfamily and is found clustered with a *tfuA* homolog in the genomes of methanogens. Mitchell and coworkers reported that ATP binds to members of the YcaO superfamily that are involved in the biosynthesis of thiazole/oxazole-modified microcins (TOMMs) (*66*). They also showed these enzymes contain a unique ATP binding motif that is required for the cyclodehydration reactions they catalyze (*67*). It is thought that ATP is used to activate the carbonyl oxygen of the peptide backbone for nucleophilic attack by a cysteine or serine residue to construct the thiazole or oxazole rings. Trifolitoxin is a genetically encoded, post-translationally modified thiazole-containing peptide antibiotic (Scheme 11). TfuA is a protein of unknown function, but has been found to be important for trifolitoxin biosynthesis (*68*). It is possible that the thioglycine modification of MCR is catalyzed by the enzymes encoded by these genes.



Scheme 11: The structure of trifolitoxin.

In addition to the two proteins discussed above, the thioglycine PTM will also need a source of sulfur. Thiamine is a sulfur-containing heterocyclic vitamin, and its biosynthetic pathway has been studied. In general, the sulfur inside the heterocyclic ring is derived from cysteine. A cysteine desulfurase transfers the sulfur from cysteine to a carrier protein, ThiI, which then produces a sulfane sulfur. In most cases, ThiF and ThiS will also participate in thio-cofactor biosynthesis together with ThiI. ThiF catalyzes a reaction that facilitates the transfer of the sulfur atom on ThiI to ThiS. The sulfur in ThiS is a C-terminal thiocarboxylate. ThiS can serve as the immediate sulfur donor in several different heterocyclic ring constructions (*69*). Mueller and coworkers reported that ThiI can serve as the immediate sulfur donor for the 4-thiouridine modification in tRNAs of bacteria and archaea (*70*). In this case, instead of cysteine, the sulfur that binds to ThiI is derived from hydrogen sulfide. According to genomic analysis of methanogens, most genomes contain conserved homologs of ThiI, but some do not have a homolog of ThiS (*71*). Thus, it is possible that a homolog of ThiI can serve as the sulfur donor for the biosynthesis of the thioglycine modification.

Chapter 2

The Heterologous Expression and purification of Coenzyme F430 Biosynthesis Enzymes in *Escherichia coli* and *in vitro* activity assays

A version of this chapter was published in *Science*.

K. Zheng, P. D. Ngo, V. L. Owens, X. Yang, S. O. Mansoorabadi, The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science* 354, 339–342 (2016)

2.1 Background

Methanogens are archaea that are responsible for the production of methane. Over one billion tons of methane are produced by methanogens per year, which is over 70% of all methane emissions (*72*). Much focus has recently been placed on reducing the production of carbon dioxide, one of the major greenhouse gases. However, the global warming potential (GWP) of methane is about 30 times greater than that of carbon dioxide, and this hazard is a growing concern (*73*).

Methanogens produce methane via methanogenesis. This metabolic pathway can utilize several C1 compounds, including carbon dioxide, methanol, and methylamines. These compounds act as carbon resources for methane production, and are all converted to the

intermediate methyl-coenzyme M. The last step of methanogenesis is catalyzed by methyl-coenzyme M reductase (MCR), which releases these C1 compounds as methane (*74*).

Currently, there is great interest in the heterologous expression of *holo* MCR. Success in this area would allow engineering of an industrial host microorganism for methane production or conversion. As discussed above, methane is a severe greenhouse gas, and it is also highly flammable. These properties result in difficulties and dangers in methane capture and transport. The discovery of a MCR homolog in anaerobic methanotrophic archaea (ANME) provides a potential way to liquify methane for safe transporting and extensive usage (*23*) (*26*) (*31*) (*32*) (*75*) (*76*).

The MCR complex is a heterohexamer that contains three pairs of subunits. Two α, two β and two γ subunits together form a large 270 kDa metalloenzyme. The enzyme has two active sites, and each site tightly binds an essential prosthetic group known as coenzyme F430. In prior literature, there was little known about the biosynthesis of F430. However, elucidation of thecoenzyme F430 biosynthesis pathway is a large obstacle that must be overcome before heterologous expression of *holo* MCR or further engineering of an industrial host organism for methane conversion can be achieved.

As discussed in the Introduction, over 60 genomes of methanogenic species have been sequenced. The genomes from representative methanogens were analyzed for potential coenzyme F430 biosynthesis (*cfb*) genes. Five candidate genes that are not only conserved in all methanogen genomes, but also present in the available ANME genome, were identified (Figure

9). These five genes are clustered or partially clustered in most of representative species, which suggests functional correlation between the genes.

We chose *Methanosarcina acetivorans* as our model organism to study coenzyme F430 biosynthesis. It is a versatile methanogen that survives in diverse environments, including trash dumps, oil wells, deep-sea hydrothermal vents, and oxygen-depleted sediments beneath kelp beds. Also, 37 °C is suitable for the growth of *M. acetivorans*, and its enzymes are active at this temperature (*77*). Therefore, this gave us confidence to try to express its enzymes heterologously in *Escherichia coli*, which multiplies efficiently at 37 °C. Furthermore, the encoded enzymes from *M. acetivorans* can be co-assayed with enzymes from *E. coli* in *in vitro* experiments if needed.

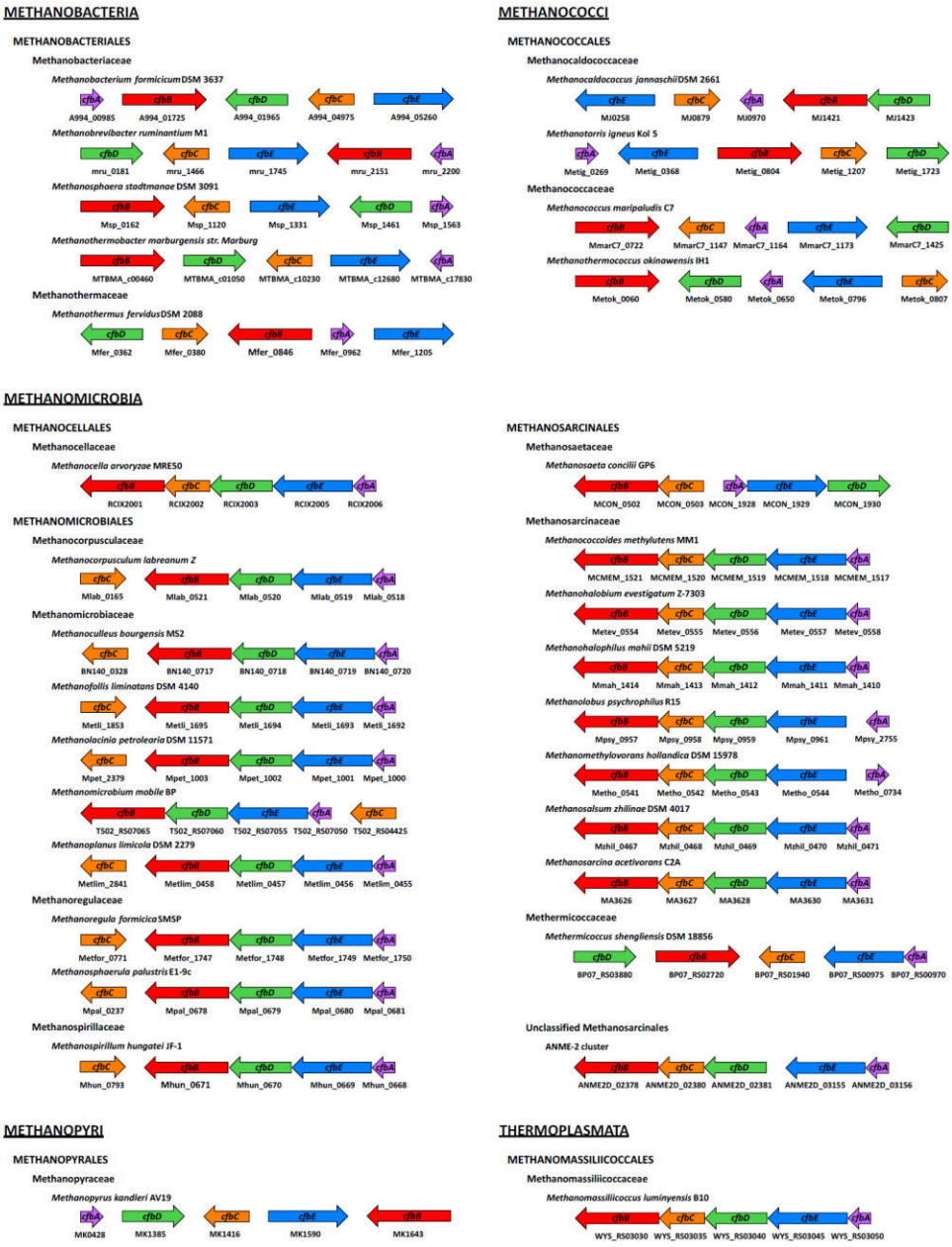


Figure 9: The *cfb* genes from representative methanogen genomes.

The genome of *M. acetivorans* has been sequenced completely and is one of the largest known archaeal genomes. *M. acetivorans* uses all three metabolic pathways for methanogenesis

(*78*), which provides us other benefits as a research object, as it conserves all genes related to methanogenesis. In addtion, all five of the identified coenzyme F430 biosynthesis candidate genes are clustered together in its genome, which are designated MA3626, MA3627, MA3628, MA3630 and MA3631. *M. acetivorans* are also capable of forming multicellular colonies, a trait rarely found in archaea (Figure 10).

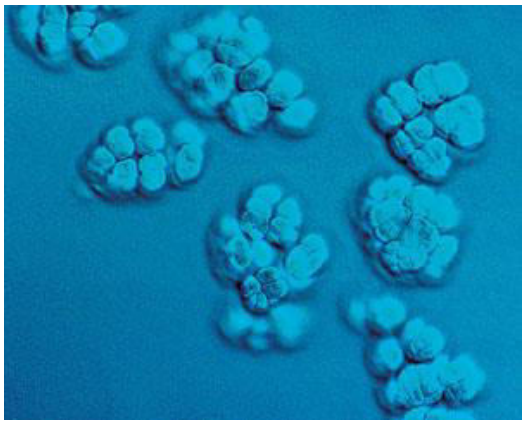


Figure 10: Micrograph of *Methanosarcina acetivorans*, which is capable of forming multicellular colonies.

E. coli is commonly used as a host for heterologous expression. We decided to use the pET-28b(+) vector to introduce exogenous genes into this host strain. The pET expression vectors are a powerful system for recombinant protein expression in *E. coli.* Target genes ligated into pET vectors are strictly controlled by the bacteriophage T7 promoter. Induction of the target gene relies on the expression of the T7 RNA polymerase in the host cell. The T7 RNA polymerase gene is controlled by the lacUV5 promoter in λ DE3 lysogen cells. In this case, there

is still some expression of the target protein even without induction with isopropyl-β-D-thiogalactopyranoside (IPTG). Therefore, these strains are suitable for the expression of nontoxic proteins. However, when the host cell contains pLysS or pLysE, the expression level of the target protein is lowered or even close to eliminated without IPTG, due to the expression of T7 lysozyme, a natural inhibitor of T7 RNA polymerase. This strategy provides a possibility for the expression of toxic proteins (*79*). We chose BL21(DE3) as our expression host since it was available and compatible with the pET-28b(+) vector. BL21(DE3) is deficient in *Ion* protease and lacks *ompT* protease (*80*), which gives a higher chance of preventing degradation of the target protein during purification.

2.2 Methods

2.2.1 *Genes cloned from Escherichia coli*

The *hemC* (BL21_03628) and *hemD* (ECD_0367) genes were amplified by polymerase chain reaction (PCR) from the genomic DNA of *E. coli* BL21 (DE3) (New England Biolabs). Primers were synthesized by Sigma-Aldrich and their sequences are provided in Table 1.

2.2.2 *Genes cloned from Methanosarcina acetivorans*

The *sirA* (MA3033), *sirC* (MA0567), *cfbA* (MA3631), *cfbB* (MA3626), *cfbC* (MA3627), *cfbD* (MA3628), *cfbE* (MA3630), and *mcrD* (MA3649) genes were amplified from the genomic DNA of *Methanosarcina acetivorans* C2A (DSM-2834). Primers were synthesized by Sigma-Aldrich and their sequences are provided in Table 1. Phusion High-Fidelity DNA Polymerase

(New England Biolabs) was utilized for all PCR reactions in accordance with the manufacturer's protocol.

Gene	Primer
hemC	Forward: 5′-GCGGCCATATGTTAGACAATGTTTTAAGAATTGCC-3′ Reverse: 5′-TATAACTCGAGTCATGCCGGAGCGTC-3′
hemD	Forward: 5′-TGGGCCATATGAGTATCCTGGTC-3′ Reverse: 5′-TAGGACTCGAGTTATTGTAATGCCCG-3′
sirA	Forward: 5′-CGGCGCATATGTCAGAAAATTACGG-3′ Reverse: 5′-ATGAGCTCGAGTCAGAAATCCTTTCCTGC-3′
sirC	Forward: 5′-GAGGACATATGATGGCTGAAACAAATAATTTTC-3′ Reverse: 5′-TAGGACTCGAGTTATTCGAGCTTATCCGAG-3′
cfbA	Forward: 5′-GGCACCATATGACTGAGAAACTCGG-3′ Reverse: 5′-ATTACGGATCCTTACAGGGCTTCCTG-3′
cfbB	Forward: 5′-CCACACATATGTCCCACAGCAAACAATC-3′ Reverse: 5′-ATTAAGGTACCCTACCGGGGAGCCC-3′
cfbC	Forward: 5′-CGCTGCATATGAAAAAACAAAAGATCGTTGC-3′ Reverse 1: 5′-CCGCGAAGCTTTTATTTTGTCATTTCCC-3′ Reverse 2: 5′-ATTATGGCCGGCCTTATTTTGTCATTTCCC-3′
cfbD	Forward: 5′-CGCCGTCATGACTCAAAAAGAGATCTC-3′ Reverse: 5′-ATCACAAGCTTTCAGGCTTCTTTTGCAAC-3′
cfbE	Forward: 5′-GACACCATATGGACCTGTTCCGG-3′ Reverse: 5′-CGCACCTCGAGTTAACGGAAACATTTC-3′
mcrD	Forward: 5′-AATCTCATATGTCAGACTCTGCTTCAAACACG-3′ Reverse: 5′-GCTCTCTCGAGTCACTCATCTTTATCAGTGTC-3′

Table 1: The sequences of primers used for PCR of the *mcrD*, *cfb*, and sirohydrochlorin biosynthesis genes.

2.2.3 *Plasmid Construction*

PCR products were purified using the OMEGA Bio-tek E.Z.N.A.® Cycle-Pure Kit and inserted into the pCR-Blunt vector (Life Technologies) by blunt end ligation, giving rise to the

pCR-Blunt:*hemC*, pCR-Blunt:*hemD*, pCR-Blunt:*sirA*, pCR-Blunt:*sirC*, pCR-Blunt:*cfbA*, pCR-Blunt:*cfbB*, pCR-Blunt:*cfbC*, pCR-Blunt:*cfbD*, and pCR-Blunt:*cfbE* plasmids. After sequence verification, *hemC*, *hemD*, *sirA*, *sirC*, and *cfbE* were digested with the NdeI and XhoI restriction enzymes and subcloned into pET-28b(+) (Novagen), producing pET-28:*hemC*, pET-28:*hemD*, pET-28:*sirA*, pET-28:*sirC*, and pET-28:*cfbE*. The *cfbA* and *cfbB* genes in pCR-Blunt were digested using NdeI and BamHI and subcloned into pET-28b(+), producing pET-28:*cfbA* and pET-28:*cfbB*. The pCR-Blunt:*cfbD* vector was digested with either NdeI and HindIII or BspHI and HindIII. The former *cfbD* insert was subcloned into pET-28b(+), producing pET-28:*cfbD*, while the latter was subcloned into the first multiple cloning site (MCS) of pRSFDuet-1 (Novagen), producing pRSFDuet-1:*cfbD*. The pCR-Blunt:*cfbC* vector was digested with either NdeI and HindIII or NdeI and FseI. The former *cfbC* insert was subcloned into pET-28b(+), producing pET-28:*cfbC*, while the latter was subcloned into the second MCS of pRSFDuet-1, producing pRSFDuet-1:*cfbC*. Also, *cfbC* digested by NdeI and FseI was subcloned into the second MCS of pRSFDuet-1:*cfbD* to construct pRSFDuet-1:*cfbD-cfbC*. Additionally, the pET-28:*cfbD* and pET-28:*cfbC* vectors were digested with NcoI and HindIII, and then subcloned into the first MCS of pCDFDuet-1 (Novagen), producing pCDFDuet-1:*cfbD* and pCDFDuet-1:*cfbC*. Vector maps for these constructed plasmids are given in Figures 11-13.

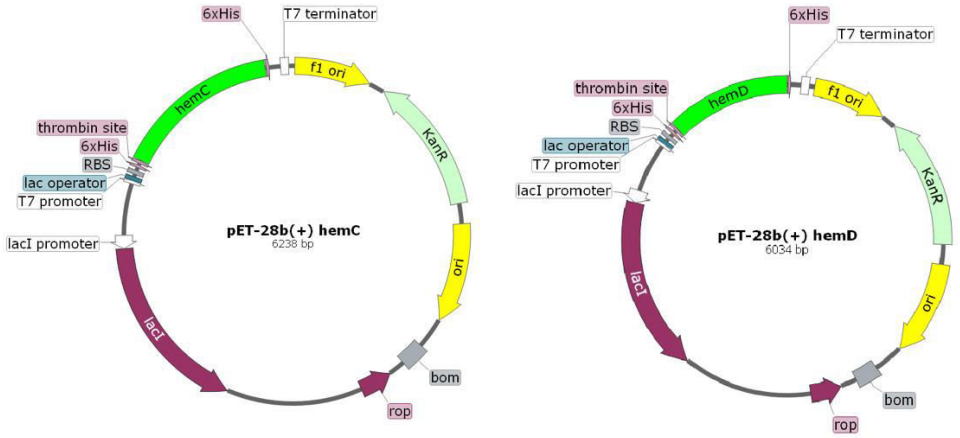


Figure 11: Plasmids containing the uroporphyrinogen III biosynthesis genes (*hemC* and *hemD*).

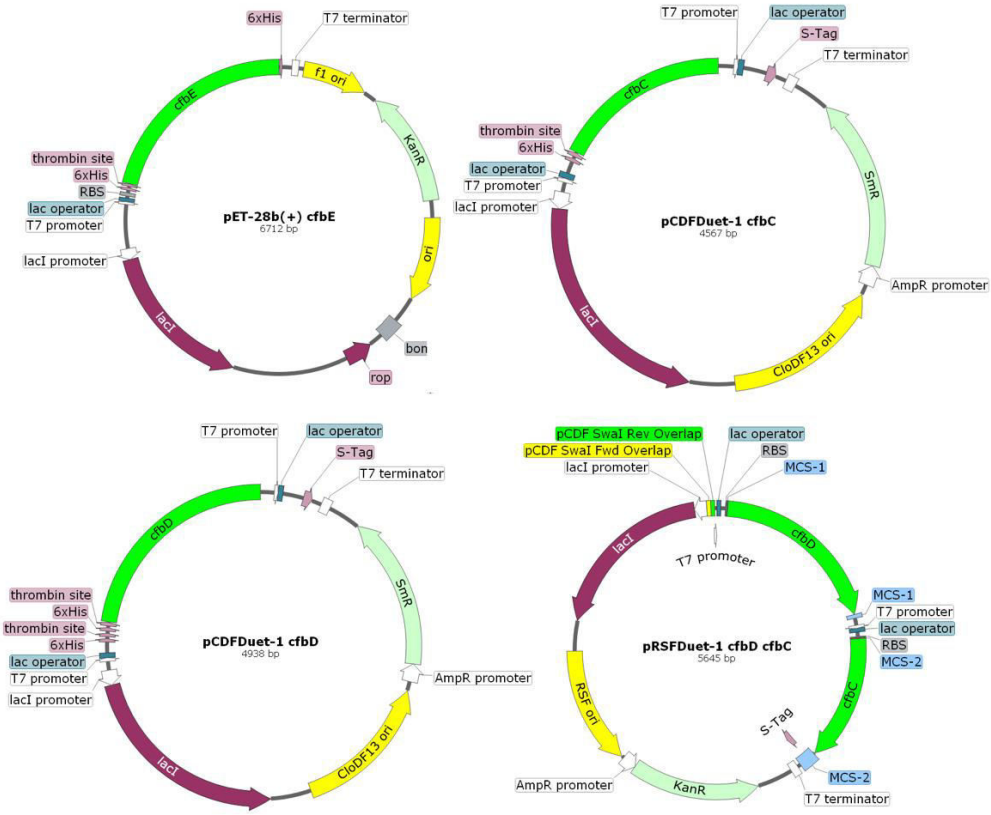


Figure 12: Additional plasmids used for the expression of the *cfbCDE* genes.

Figure 13: Plasmids used for the expression of sirohydrochlorin biosynthesis genes (*sirA* and *sirC*) and the *cfbABCD* genes.

2.2.4 *Expression and purification of the Cfb enzymes*

The pET-28:*hemC* plasmid was transformed into *E. coli* BL21 (DE3) cells (New England Biolabs) and grown in Lysogeny broth (LB), supplemented with kanamycin (50 μg/mL final concentration) at 37 °C. After three hours, the temperature was lowered to 15 °C and allowed to shake for an additional one hour. The cells were then induced with IPTG at a final concentration of 40 μM and 5-aminolevulinate (5-ALA) at a final concentration of 10 μM and incubated with shaking at 15 °C for 8 hours.

The pET-28:*hemD* plasmid was transformed into *E. coli* BL21 (DE3) cells and grown in LB medium, supplemented with kanamycin (50 μg/mL final concentration). When an OD_{600} of 0.6 was reached, the cells were induced with IPTG at a final concentration of 40 μM and incubated with shaking at 18 °C for eight hours.

Both pET-28:*sirA* and pET-28:*sirC* were transformed into *E. coli* BL21 (DE3) cells and grown in LB medium, supplemented with kanamycin (50 μg/mL final concentration). When an OD_{600} of 0.5 was reached, the cells were induced with IPTG at a final concentration of 0.1 mM and incubated with shaking overnight at 18 °C.

The pET-28:*cfbA* plasmid was transformed into *E. coli* BL21 (DE3) cells and grown in LB medium, supplemented with kanamycin (50 μg/mL final concentration). When an OD_{600} of 0.5 was reached, the cells were induced with IPTG at a final concentration of 0.4 mM and incubated with shaking overnight at 25 °C.

Both pET-28:*cfbB* and pET-28:*cfbE* were transformed into *E. coli* BL21 (DE3) cells and grown in LB medium, supplemented with kanamycin (50 μg/mL final concentration). When an OD_{600} of 0.5 was reached, the cells were induced with IPTG at a final concentration of 0.1 mM, and incubated with shaking overnight at 25 °C.

Both pET-28:*cfbC* and pET-28:*cfbD* were transformed into *E. coli* BL21 (DE3) cells and grown in LB medium, supplemented with kanamycin (50 μg/mL final concentration). When an OD_{600} of 0.5 was reached, the cells were induced with IPTG at a final concentration of 0.1 mM, and incubated with shaking overnight at 25 °C.

The pDB1282 vector containing the iron-sulfur cluster (*isc*) biosynthetic gene cluster from *Azotobacter vinelandii*, pRSFDuet-1: *cfbD-cfbC* and pCDFDuet-1:*cfbC* were transformed into *E. coli* BL21 (DE3) cells and grown in LB medium, supplemented with kanamycin (50 μg/mL final concentration), ampicillin (100 μg/mL final concentration), and spectinomycin (25 μg/mL final concentration). When an OD_{600} of 0.5 was reached, the cells were induced with L-(+)-arabinose at a final concentration of 3 g/L. At the same time, 4 mM of $FeSO_4$ and cysteine were added. After a three hour incubation, the cells were induced with IPTG at a final concentration of 0.3 mM and incubated with shaking overnight at 18 °C.

The cells expressing His_6-tagged HemC or His_6-tagged CfbC with untagged CfbC and CfbD were pelleted and frozen at -80 °C. The cell pellets were then transferred to a controlled anaerobic nitrogen atmosphere, where they were defrosted and re-suspended in anaerobic lysis

buffer (50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, lysozyme (1 mg/ml), Ameresco's Protease Inhibitor Cocktail, pH 8.0). The cell suspension was kept on ice for several hours and was degassed. All subsequent steps were performed anaerobically unless otherwise noted. The cell suspension was then sonicated and centrifuged at 30,000 × G for one hour. The supernatant was then applied to a Bio-Rad Econo-Pac column packed with Ni^{2+}-charged Profinity IMAC Resin. The column was then washed with 50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, pH 8.0 buffer. The protein was then eluted with 50 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole, pH 8.0 buffer, and buffer exchanged with 100 mM Tris-HCl, pH 8.0 buffer. Then 20% glycerol (from an 80% stock solution) was added for storage at -80 °C.

The cells expressing His_6-tagged HemD, SirA, SirC, or CfbE were pelleted and re-suspended aerobically in 50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, pH 8.0 lysis buffer. After the cell suspension was sonicated and centrifuged at 15,900 × G for thirty minutes, the supernatant was applied to a Bio-Rad Econo-Pac column packed with Ni^{2+}-charged Profinity IMAC Resin. The column was washed with 50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, pH 8.0 buffer. The protein was then eluted with 50 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole, pH 8.0 buffer, and buffer exchanged with 100 mM Tris-HCl, pH 8.0 buffer. Then 20% glycerol (from an 80% stock solution) was added for storage at -80 °C.

The cells expressing His_6-tagged CfbA or CfbB were pelleted and re-suspended aerobically in 50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, pH 8.0 lysis buffer.

After the cell suspension was sonicated and centrifuged at 15,900 × G for thirty minutes, the supernatant was applied to a Bio-Rad Econo-Pac column packed with Ni^{2+}-charged Profinity IMAC Resin. The column was then washed with 50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, pH 8.0 buffer and 100 mM Tris-HCl, pH 8.0 buffer. After the washing step, 80 units of thrombin were added to 1 ml of resin, then 1 ml 100 mM Tris-HCl, pH 8.0 buffer was added. The thrombin resin buffer mixture was incubated at 25 °C for sixteen hours. The protein was then eluted with 100 mM Tris-HCl, pH 8.0 buffer and filtered through a benzamidine sepharose column to remove the thrombin. The protein sample was then concentrated and 20% glycerol (from an 80% stock solution) was added for storage at -80 °C.

2.2.5 *SirA Activity Assays*

The SirA reaction was performed anaerobically for 2 hours at 37 °C and contained: HemC (0.06 mg/mL), HemD (0.06 mg/mL), CobA (0.12 mg/mL), porphobilinogen (PBG) (0.2 mg/mL), and *S*-adenosyl-L-methionine (SAM) (0.78 mg/mL) in 1 mL of 100 mM Tris-HCl, pH 8.0 buffer. Ultraviolet-visible (UV-vis) spectrophotometry was used to follow the reaction progress.

2.2.6 *SirC Activity Assays*

The SirC reaction was performed anaerobically for 12 hours at 37 °C and contained: HemC (0.06 mg/mL), HemD (0.06 mg/mL), SirA (0.12 mg/mL), SirC (0.36 mg/mL), porphobilinogen (PBG) (0.2 mg/mL), SAM (0.78 mg/mL), $NAD(P)^+$ (1 mM), and $MgCl_2$ (40

μM) in 1 ml of 100 mM Tris-HCl, pH 8.0 buffer. UV-vis spectrophotometry was used to follow the reaction progress.

2.2.7 *CfbA Activity Assays*

The CfbA reaction was performed anaerobically for 12 hours at 37 °C and contained: HemC (0.06 mg/mL), HemD (0.06 mg/mL), SirA (0.12 mg/mL), SirC (0.36 mg/mL), thrombin cleaved CfbA (0.09 mg/mL), porphobilinogen (PBG) (0.2 mg/mL), SAM (0.78 mg/mL), $NAD(P)^+$ (1 mM), $MgCl_2$ (4 mM), and $NiCl_2$ (200 μM, half added at the beginning of the assay, and another half added after three hours) in 1 mL of 100 mM Tris-HCl, pH 8.0 buffer. UV-vis spectrophotometry was used to follow the reaction progress.

2.2.8 *CfbB Activity Assays*

The CfbB reaction was performed anaerobically for 12 hours at 37 °C and contained: 200 μL of the completed CfbA reaction (HPLC verified), CfbB (0.63 mg/mL), glutamine (4.08 mM), ATP (1.77 mM), pyruvate kinase (400 units), and phosphoenolpyruvate (PEP) (7.10 mM) in 1mL of 100 mM Tris-HCl, pH 8.0 buffer. UV-vis spectrophotometry was used to follow the reaction progress.

2.2.9 *CfbCD Activity Assays*

The CfbCD reaction was performed anaerobically for 12 hours at 37 °C and contained: 180 μL of the completed CfbB reaction 180 μL (HPLC verified), the purified CfbCD complex (0.63 mg/mL), sodium dithionite (26 mM, added at 2 hour intervals over the first 6 hours of the

reaction), ATP (1.75 mM), pyruvate kinase (400) units, and PEP (17.5 mM) in 1 mL of 100 mM Tris-HCl, pH 8.0 buffer.

2.2.10 *CfbE Activity Assays*

The CfbE reaction was performed anaerobically for 12 hours at 37 °C and contained: 80 μL of the completed CfbCD reaction (HPLC verified), CfbE (1.4 mg/mL), ATP (1.61 mM), pyruvate kinase (200) units, and PEP (10.0 mM) in 1 mL of 100 mM Tris-HCl, pH 8.0 buffer. Another CfbE reaction was also prepared that included 2.6 mg/mL of McrD.

2.2.11 *HPLC Analysis*

High-performance liquid chromatography (HPLC) analysis was performed using an Agilent Infinity 1260 HPLC with an Agilent Poroshell 4.6 × 150 mm C18 column. Data analysis utilized the Agilent ChemStation Software. The method consisted of solvent A (water + 0.5% formic acid (FA)) and solvent B (acetonitrile + 0.5% FA) with a flow rate established at 1.0 mL/min. The column was equilibrated with 100% A and remained isocratic for 2 minutes after sample injection. Solvent B was increased to 20% over 3 minutes and remained isocratic for 5 minutes. Solvent B was then increased to 25% over 5 minutes and remained isocratic for 5 minutes. Solvent B was then increased to 30% over 5 minutes and subsequently to 100% over 5 minutes. Finally, solvent B remained isocratic for an additional 15 minutes. The total run time of the solvent gradient program was 45 minutes. A detection wavelength of 400 nm was selected on the diode array detector.

2.2.12 *LC-MS Analysis*

Concurrently, samples were also subjected to mass spectrometry (MS) analysis on a Waters Acquity UHPLC TUV Q-TOF LC-MS equipped with an Agilent Poroshell 4.6 × 150 mm C18 column. The data was analyzed using Waters MassLynx software. The solvent method consisted of an identical solvent A and B utilized in the HPLC analysis with a 1.0 mL/min flow rate. The gradient initialized with 100% solvent A and increased to 100% solvent B over the course of 15 minutes. The electrospray ionization mass detector was configured to positive ion mode and to scan between 0 and 1100 *m/z*. The inline Waters TUV UV-vis detector was also engaged at 400 nm to verify UV-visible peaks corresponded with MS peaks on their respective chromatograms.

2.3 Results

2.3.1 *Expression and Purification of cfb Enzymes*

The *sirA*, *sirC*, *cfbA*, *cfbB*, *cfbC*, *cfbD*, *cfbE*, and *mcrD* genes were cloned from the genomic DNA of *M. acetivorans* C2A using polymerase chain reaction (PCR). To biosynthesize the precursor of F430 pathway, the genes *hemC, hemD* were cloned from the genomic DNA of *E. coli* using PCR. These genes were ligated into pET vectors, which allowed protein expression with an N-terminal His_6-tag. The vectors were transformed into *E. coli* BL21 (DE3). The cells were induced with IPTG, harvested, and the soluble protein was extracted from the cell. The proteins were purified by immobilized metal affinity chromatography (IMAC). Purification of the CfbCD complex was carried out by expressing His_6-tagged CfbC in the presence of untagged

CfbC and CfbD and the *isc* gene cluster. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed to check the size and purity and if was found that the optimized expression and purification conditions provided high purity protein (Figure 14).

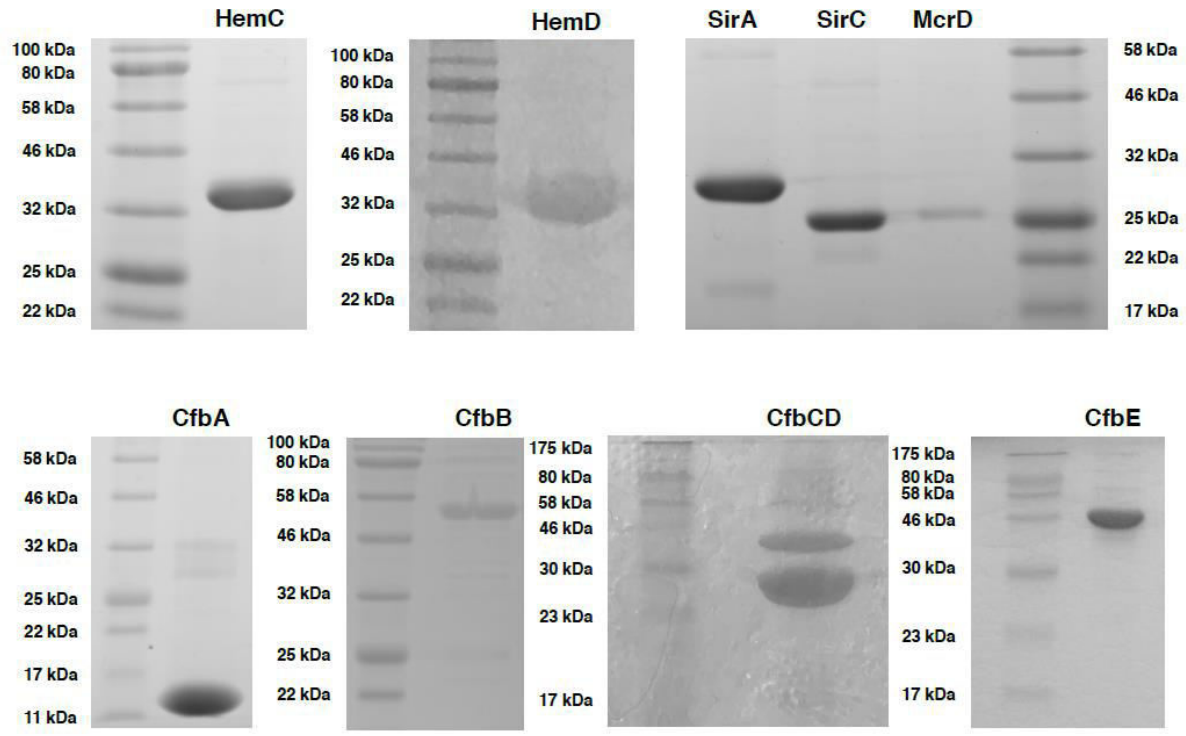


Figure 14: SDS-PAGE gel for McrD and the uroporphyrinogen III, sirohydrochlorin, and coenzyme F430 biosynthesis enzymes. These enzymes were used for the *in vitro* assays reported above. The approximate size of these proteins are: CfbA, 14 kDa; CfbB, 53 kDa; CfbC, 28 kDa; CfbD, 41 kDa; CfbE, 50 kDa; SirA, 28 kDa; SirC, 24 kDa; HemC, 32 kDa; HemD, 30 kDa; McrD, 24 kDa.

2.3.2 *SirA and Precorrin 2*

Sirohydrochlorin and/or precorrin 2 are proposed intermediates in the biosynthesis of all

C2 and C7 methylated tetrapyrroles (*81*). Uroporphyrinogen III, the last common precursor of all tetrapyrrole biosynthesis pathways (*82*), is not commercially available. To biosynthesize precorrin 2, uroporphyrinogen III is needed. HemC (PBG deaminase) and HemD (uroporphyrinogen III synthase) can be used to synthesize uroporphyrinogen III from porphobilinogen (PBG), which is commercially available. Subsequently, SirA can convert uroporphyrinogen III to precorrin 2 (*83*) (*84*). Sirohydrochlorin can then be prepared using precorrin 2 dehydrogenase (SirC), which converts precorrin 2 to sirohydrochlorin using $NAD(P)^+$ (*81*) (*85*).

PBG has a broad absorption band around 500 nm, while the characteristic absorption maximum (λ_{max}) of precorrin 2 is around 400 nm. When HemC, HemD, and SirA were incubated in pH 8.0 Tris buffer with PBG and *S*-adenosyl-L-methionine (SAM), a slight decrease at 500 nm and an increase at 400 nm revealed precorrin 2 formation, indicating that all the enzymes are active (Figure 15).

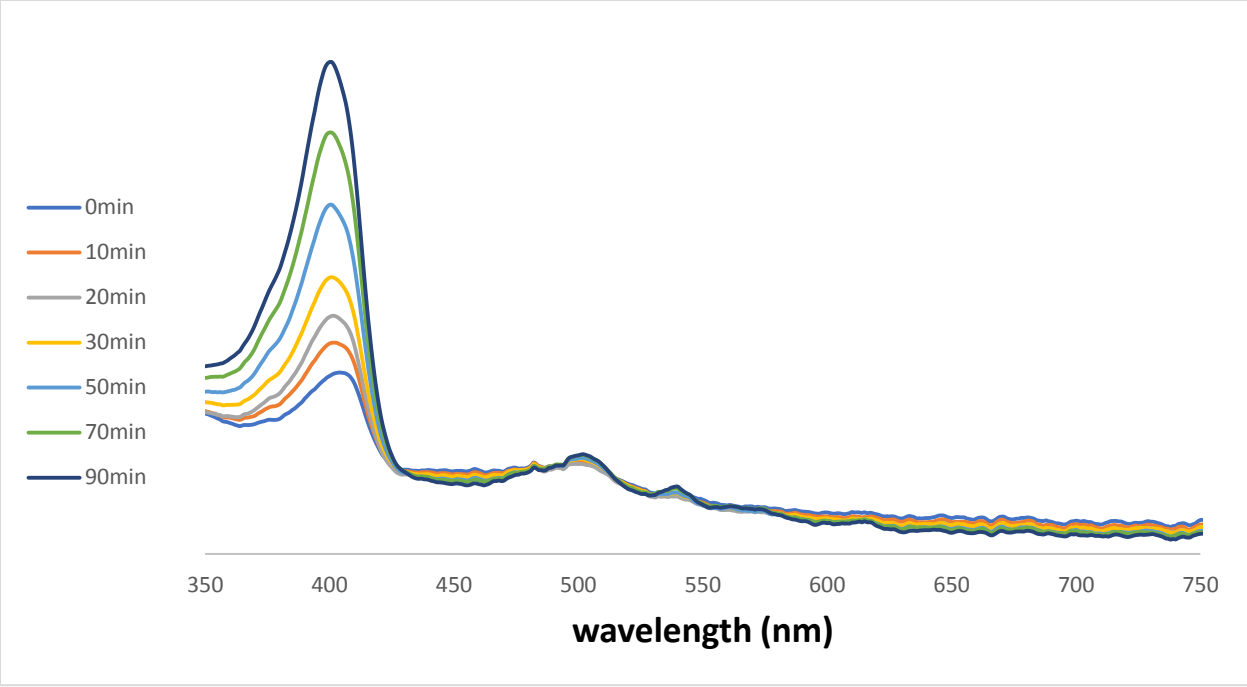


Figure 15: UV-vis kinetic traces of the SirA reaction, showing the synthesis of precorrin 2 from PBG. The bottom blue line was monitored when reaction started. First line to Third line above the grey line were monitor every 10 min. Fourth line to sixth line above grey line were monitored every 20 min after the third line.

2.3.3 *SirC and Sirohydrochlorin*

We then tested the activity of SirC. SirC and NAD^+ were included in the SirA reaction. Broad peaks near 350 and 600 nm formed, indicating the biosynthesis of sirohydrochlorin and Nicotinamide adenine dinucleotide (NADH) as a coproduct, which has a λ_{max} of 340 nm (Figure 16).

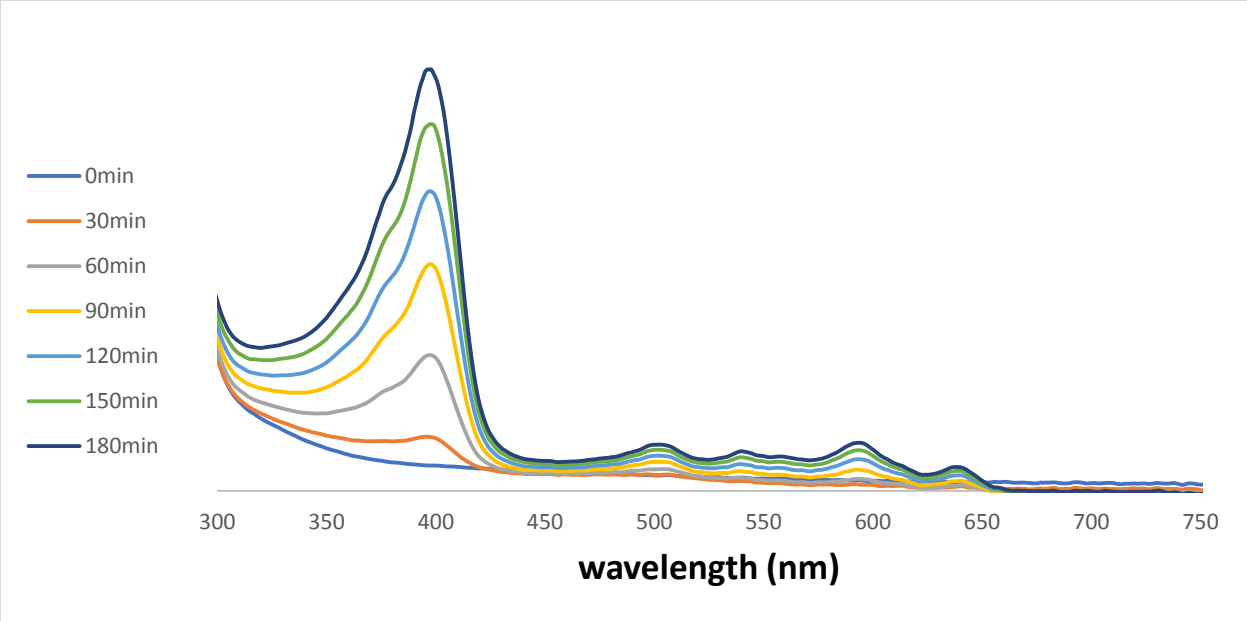


Figure 16: UV-vis kinetic traces of the SirC reaction, showing the synthesis of sirohydrochlorin and NADH from precorrin 2 and NAD^+. Data were monitored per 30 min.

To confirm the biosynthesis of sirohydrochlorin, we utilized HPLC to separate the products from the reaction mixture. We applied water + 0.5% formic acid and acetonitrile + 0.5% formic acid as the mobile phases on the C18 column. Chromatogram traces were collected with a UV-vis diode array detector set to monitor at 400 nm. The data reveal formation of sirohydrochlorin (Figure 17a) at a retention time of 16.1 minutes with its characteristic UV-vis absorption spectrum (Figure 17b). This peak also had a 863.30 *m*/*z* in the UPLC-MS spectrum, which is representative of sirohydrochlorin (Figure 17c).

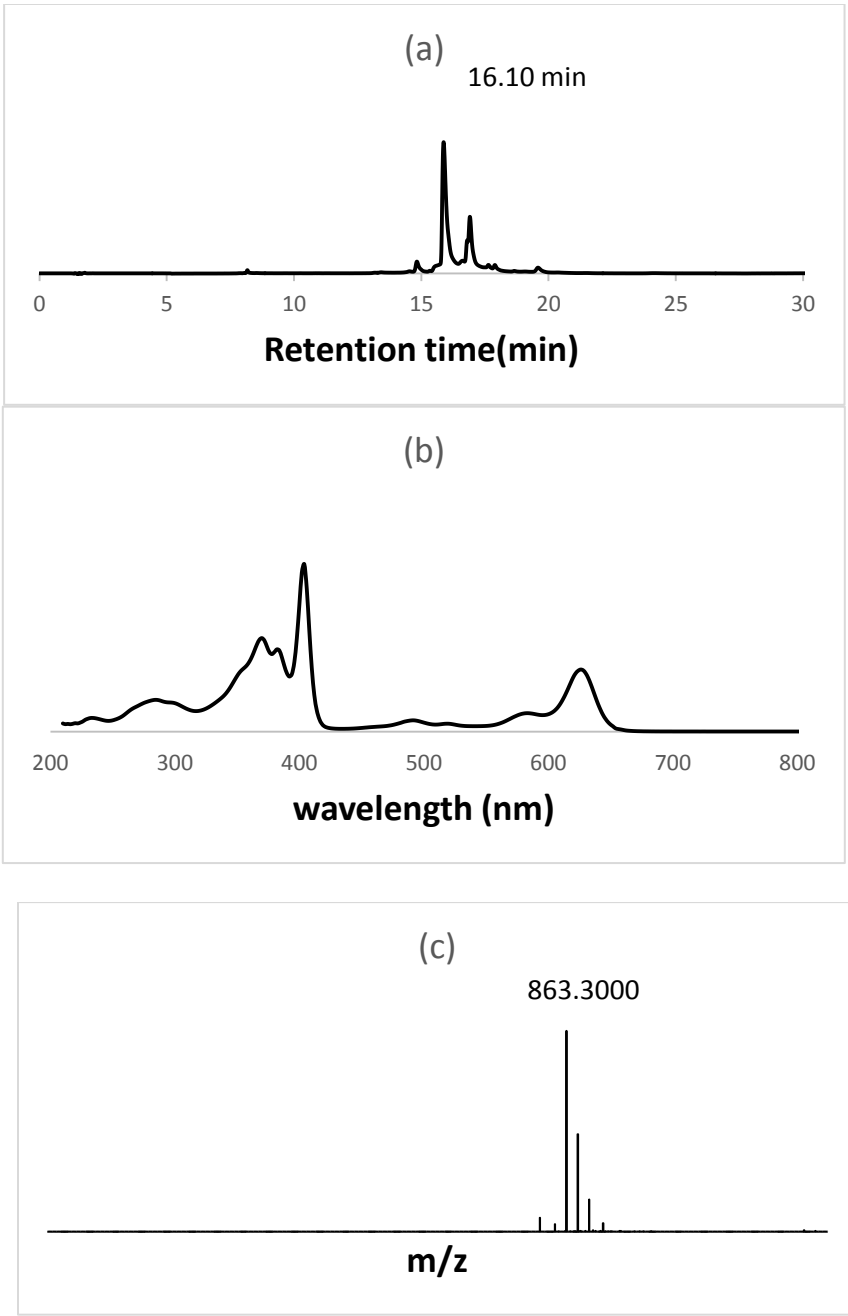


Figure 17: Analysis of the SirC reaction and confirmation of the production of sirohydrochlorin. (a) HPLC chromatogram of the SirC reaction with detection at 400 nm, (b) UV-vis absorption and (c) UPLC-MS spectra of the peak with a retention time of 16.1 minutes (sirohydrochlorin).

2.3.4 *CfbA and Ni-sirohydrochlorin*

We included His_6-tagged CfbA and $NiCl_2$ into the SirA reaction, and no obvious changes were observed using UV-vis absorption spectroscopy. Then we injected the sample onto HPLC and UPLC-MS, and no meaningful peaks were found other than the SirA reaction products. The same phenomena were observed when the assay was repeated with thrombin-cleaved CfbA.

These assays indicate that precorrin 2 is not the substrate of CfbA. Another candidate for the immediate precursor of coenzyme F430 biosynthesis is sirohydrochlorin. Although His_6-tagged CfbA still did not yield the expected product, when CfbA treated with thrombin (to remove the His_6-tag) was used in the reaction, the formation of Ni-sirohydrochlorin was observed. When CfbB was also included to help alleviate product inhibition of CfbA, full conversion to Ni-sirohydrochlorin was observed within 6 hours. When sirohydrochlorin becomes metallated, the UV-vis absorption peaks at 400 and 600 nm becomes sharper. At the same time, the 600 nm peak is split into several small peaks. The shape and absorption wavelength of these peaks depend on the type of metal ion (Figure 18). We then analyzed the reaction mixture using HPLC and UPLC-MS and found peaks representative of Ni-sirohydrochlorin (20.2 minute retention time, 919.22 *m*/*z*) (Figure 19).

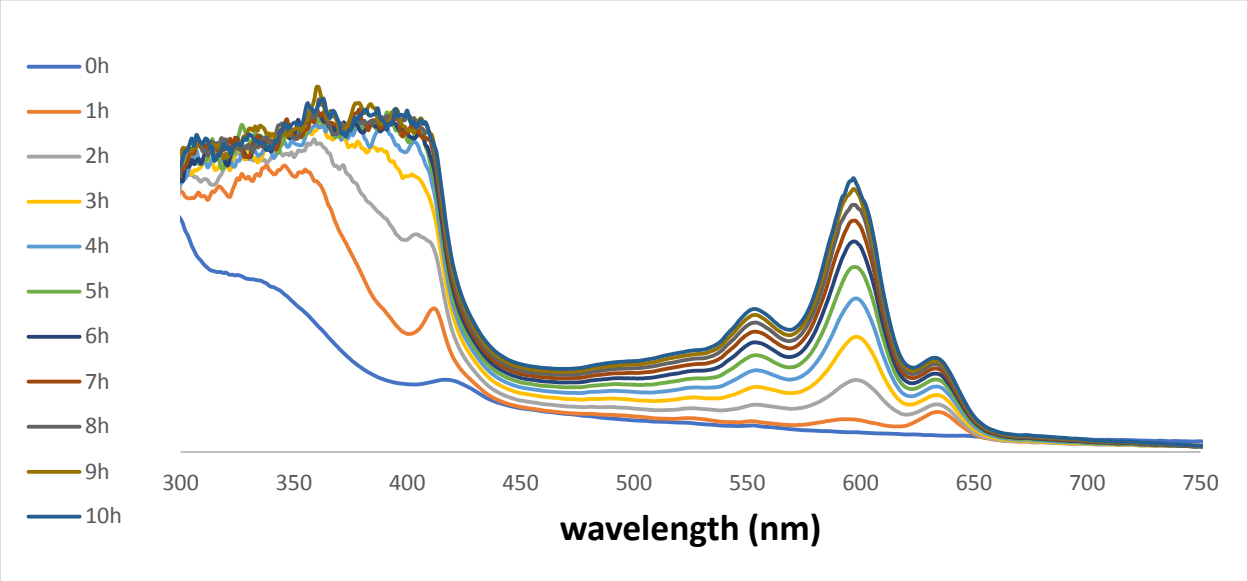


Figure 18: UV-vis kinetic traces of the CfbA reaction, revealing the synthesis of Ni-sirohydrochlorin from sirohydrochlorin. Data were monitored every hour.

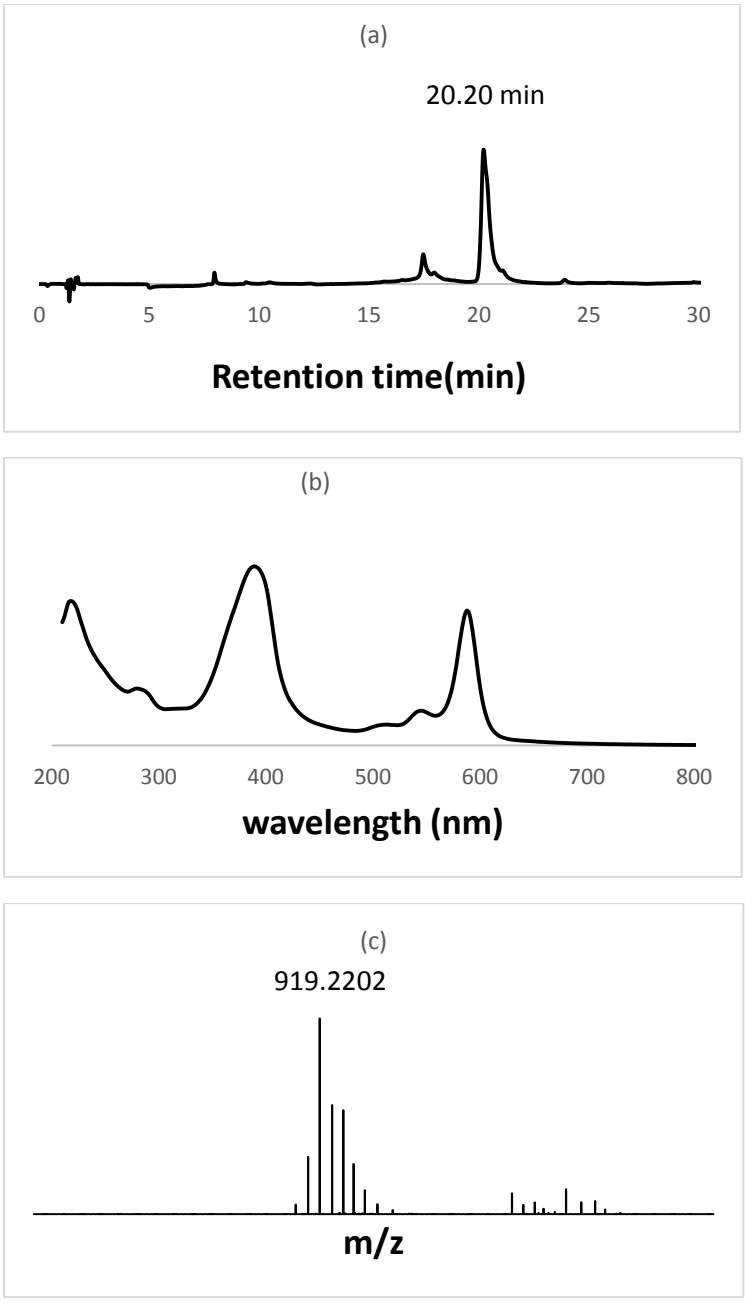


Figure 19: Analysis of the CfbA reaction and confirmation of the production of Ni-sirohydrochlorin. (a) HPLC chromatogram of the CfbA reaction with detection at 400 nm, (b) UV-vis absoption and (c) UPLC-MS spectra of the peak with a retention time of 20.2 minutes (Ni-sirohydrochlorin).

2.3.5 *CfbB and Ni-sirohydrochlorin a,c-diamide*

We next tested the activity of CfbB by adding it, ATP, glutamine, and an ATP regeneration system to the CfbA reaction (CfbCD (without the reductant required for its activity) also was included to alleviate product inhibition). An analogous change occurred in the UV-vis spectrum as what was observed with the CfbA reaction (Figure 20). We used HPLC to separate the products from the CfbB reaction mixture. We found that the peak of Ni-sirohydrochlorin on HPLC chromatogram disappeared and a new peak with a retention time of 17.2 min was formed. According to the UPLC-MS spectrum, this new peak has a *m/z* of 917.25, which is consistent with the structure of Ni-sirohydrochlorin *a*,*c*-diamide (Figure 21).

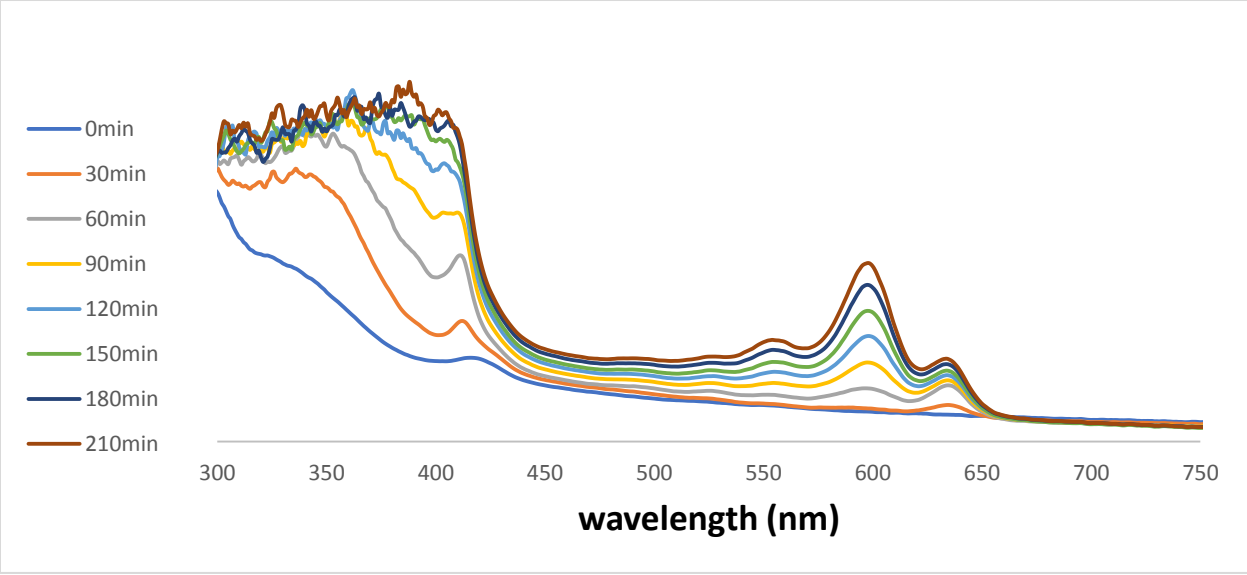


Figure 20: UV-vis kinetic traces of the CfbB reaction, revealing the synthesis of Ni-sirohydrochlorin *a*,*c*-diamide. Data were monitored per 30 min.

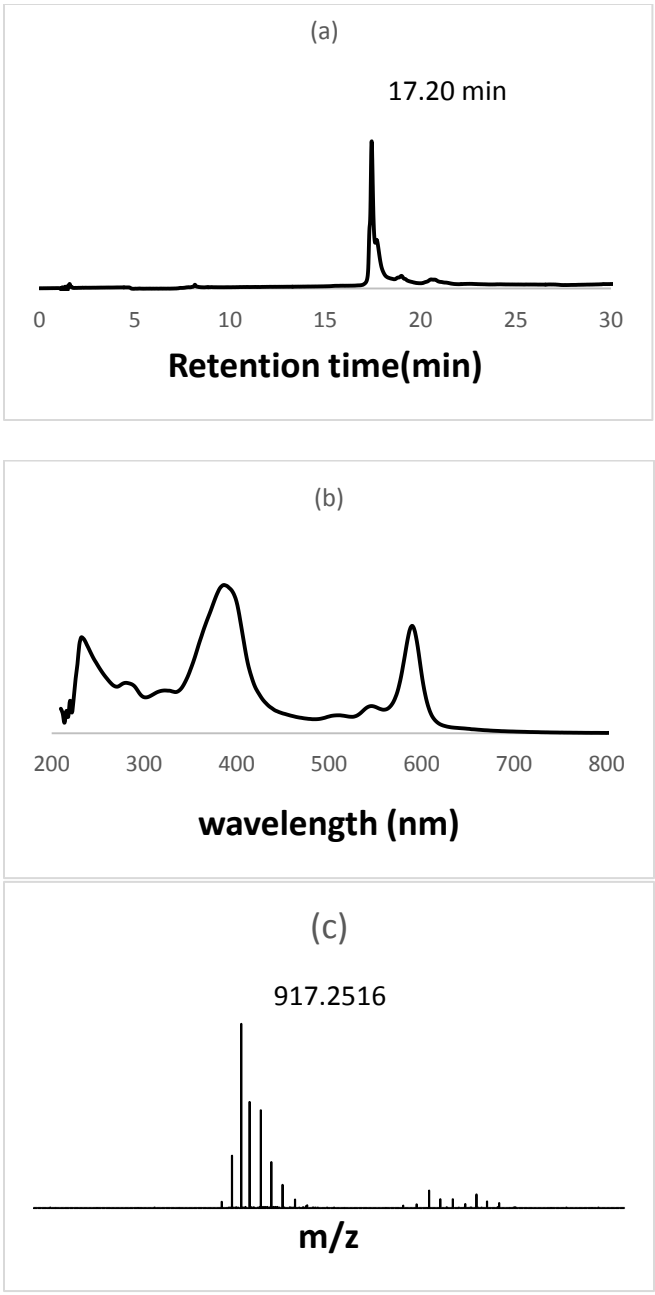


Figure 21: Analysis of the CfbB reaction and confirmation of the production of Ni-sirohydrochlorin *a*,*c*-diamide. (a) HPLC chromatogram of the CfbB reaction with detection at 400 nm, (b) UV-vis absorption and (c) UPLC-MS spectra of the peak with a retention time of 17.2 minutes (Ni-sirohydrochlorin *a*,*c*-diamide).

2.3.6 *CfbCD and 15,17^3-seco-F430-17^3-acid*

After yielding Ni-sirohydrochlorin *a*,*c*-diamide, we set up the CfbCD reaction. We added the CfbCD complex and dithionite into the CfbB assay without an ATP regeneration system. No new compounds were detected by the assay. In the nitrogenase and DPOR reactions, ATP plays two critical roles: as the energy source, and the regulator of association and disassociation of the complex. In both cases, the reactions are strongly inhibited by Adenosine diphosphate (ADP). We hypothesized that ADP is also an inhibitor of the CfbCD reaction. To remove synthesized ADP, we introduced pyruvate kinase (PK) and phosphoenol pyruvate (PEP). Since PK utilizes PEP and ADP to produce pyruvate and ATP, we were able to remove ADP and regenerate the ATP substrate at the same time. Moreover, sodium dithionite participates in this reaction as an electron donor. After a 12-hour reaction, we collected the reaction mixture and analyzed it using HPLC and UPLC-MS. A new peak with a retention time of 8.1 minutes was observed, with a UV-visible absorption spectrum (λ_{max} = 425 nm) identical to that of the only previously known intermediate in coenzyme F430 biosynthetic pathway, 15,17^3-seco-F430-17^3-acid. Formation of this intermediate was also verified by UPLC-MS, which showed a peak with a *m*/*z* of 923.30, consistent with the calculated *m*/*z* for 15,17^3-seco-F430-17^3-acid (Figure 22).

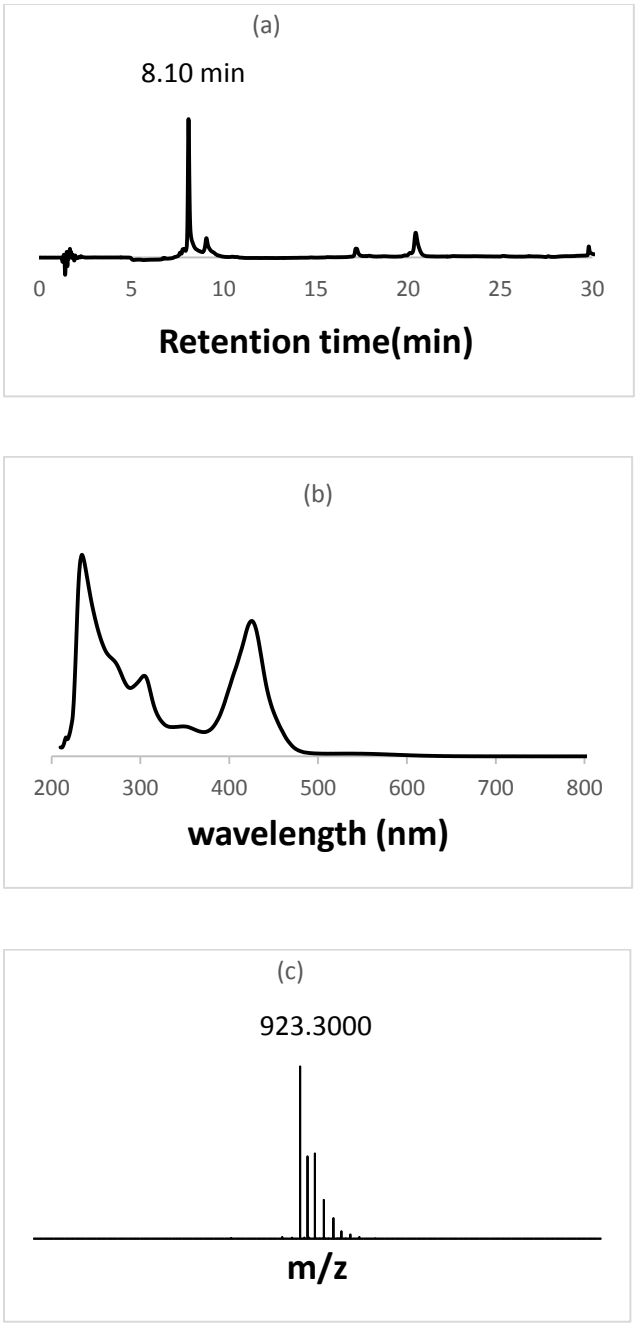


Figure 22: Analysis of the CfbCD reaction and confirmation of the production of 15,17^{3}-seco-F430-17^{3}-acid. (a) HPLC chromatogram of the CfbCD reaction with detection at 400 nm, (b) UV-vis absorption and (c) UPLC-MS spectra of the peak with a retention time of 8.1 minutes (15,17^{3}-seco-F430-17^{3}-acid).

2.3.7 *CfbE and coenzyme F430*

The final step of the *cfb* pathway is cyclization to form the carbocyclic F ring. The carboxyl group on C17 ligates to C15, and a molecule of water is released. We hypothesized that the final enzyme in the *cfb* pathway is CfbE. In order to test the hypothesis, we added CfbE into the CfbCD reaction. After a 12 hour incubation, we analyzed the mixture using HPLC, and found a new peak with the same exact retention time (7.8 minutes) as a coenzyme F430 standard. The conversion rate of this reaction was such that the ratio of substrate to product was 7:3 (Figure 23).

2.3.8 *McrD*

We hypothesized that McrD is the chaperone protein that delivers coenzyme F430 to MCR. A McrD assay was established from the CfbE assay. After addition of McrD to the CfbE assay, complete conversion of 15,17^3-seco-F430-17^3-acid to coenzyme F430 is observed. UPLC-MS analysis shows a peak at 905.29 *m*/*z*, which demonstrates that F430 has been synthesized as expected (Figure 23).

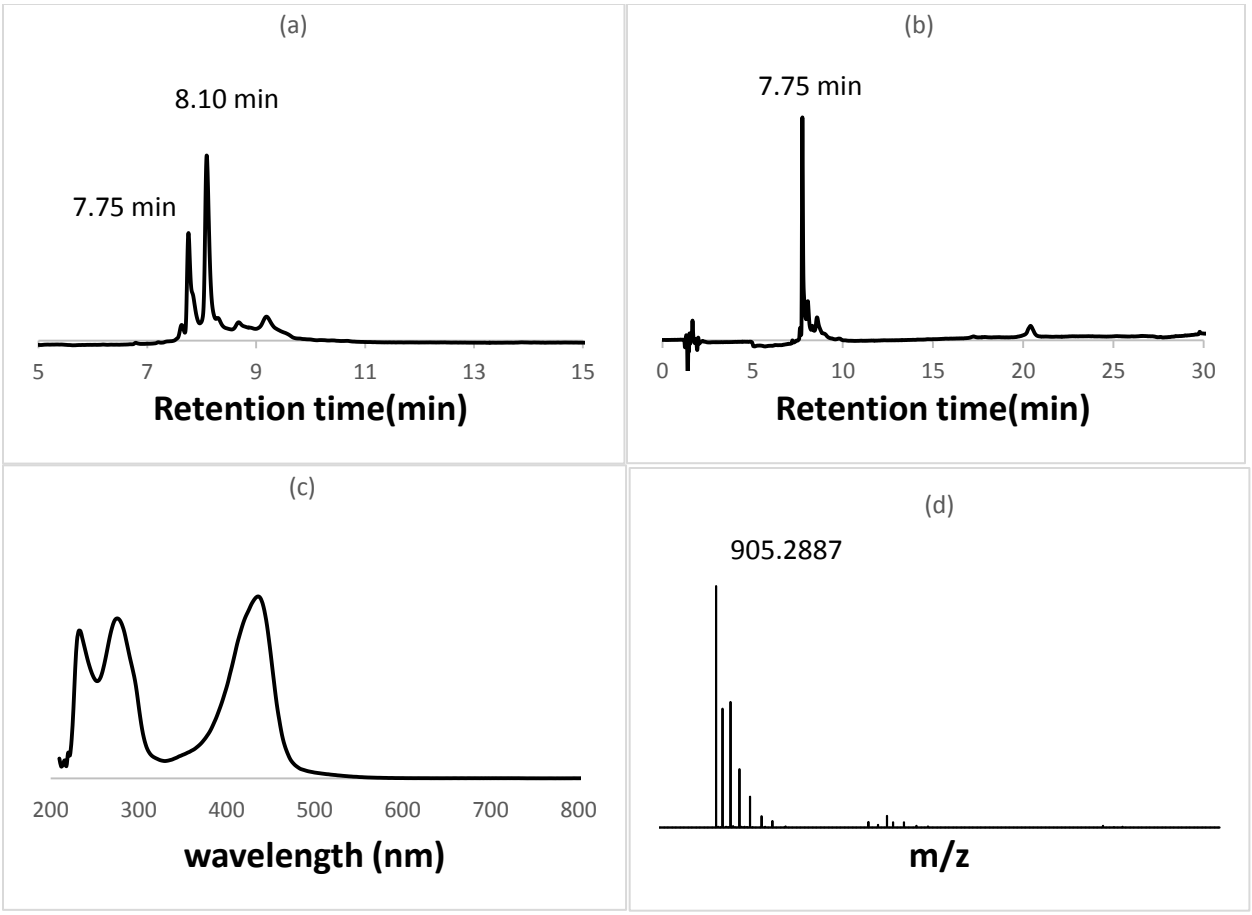


Figure 23: Analysis of the CfbE reaction and confirmation of the *in vitro* biosynthesis of coenzyme F430. HPLC chromatograms of the CfbE reaction (a) before and (b) after addition of McrD, showing the effect of the yield of F430. (c) UV-vis absorption and (d) UPLC-MS spectra of the peak with a retention time of 7.8 minutes (coenzyme F430).

2.4 Discussion

We heterologous expressed and purified SirA, SirC, the coenzyme F430 biosynthesis enzymes (CfbABCDE), and McrD using pET-28b(+), pRSFDuet and/or pCDFDuet vectors and the *E. coli* BL21(DE3) expression host. All enzymes have been proven to be active during assays according to characterization using UV-vis spectrophotometry, HPLC, and LC-MS. The study

described in this chapter clearly elucidates the enzymes that are necessary and sufficient for *in vitro* coenzyme F430 biosynthesis (Figure 24), and all enzymes encoded by the *cfb* genes were characterized for the first time in this study. Moreover, we identified several intermediates in this pathway, two of which were not previously reported. Last but not least, this study showed that all *cfb* enzymes could be heterologous expressed in an active form by non-methanogenic hosts, specifically by *E. coli*. This fact indicates that there is potential to biosynthesize F430 (and thus *holo* MCR) in an industrially relevant host, since all the enzymes directly involved in this pathway can be expressed together using genetic engineering.

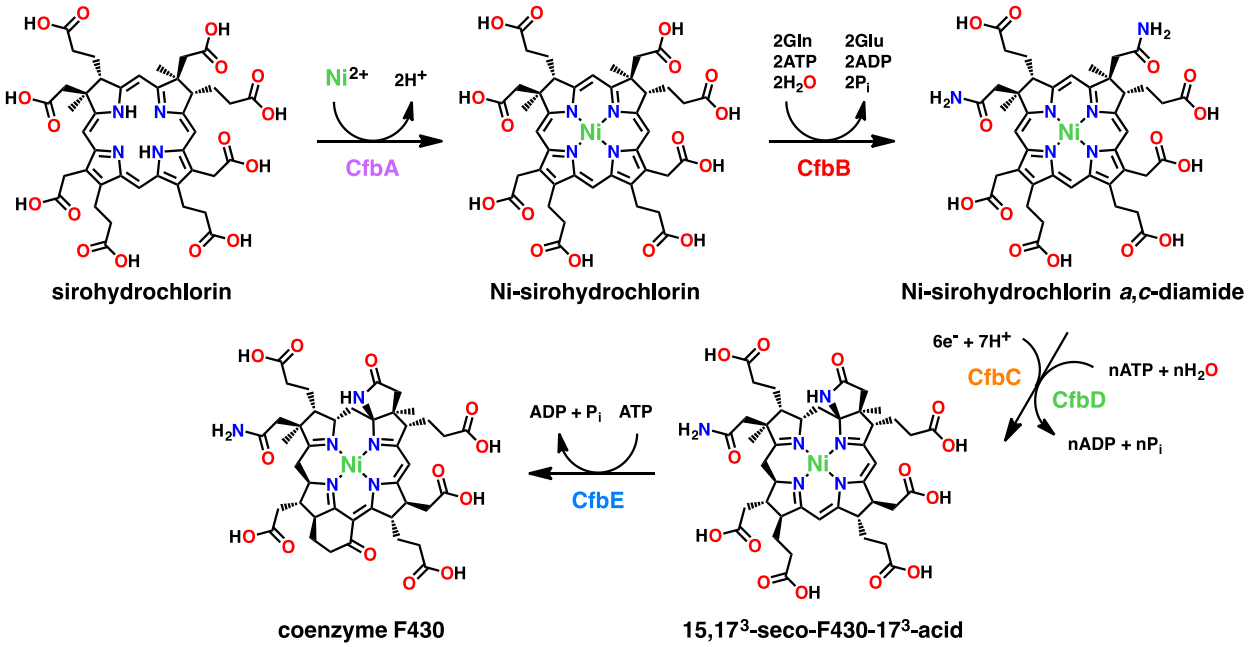


Figure 24: The identified coenzyme F430 biosynthesis pathway.

The experimentally elucidated pathway matched our hypothesis, except for the pathway precursor. Initially, we proposed that precorrin 2 was the immediate precursor of the *cfb* pathway because its molecular framework is closer to F430 than that of sirohydrochlorin (i.e., it is more highly reduced). However, the sirohydrochlorin cobaltochelatase homolog, CfbA, only showed activity with sirohydrochlorin. We noticed that removal of the N-terminal His_6-tag of CfbA revives its activity after expression and purification using Ni-IMAC. Moreover, wild-type CfbA containing a C-terminal His_6-tag is also inactivity. Interestingly, CfbA is the first reported tetrapyrrole nickelochelase that is a class II, ATP-independent chelatase. Certain members of this class of chelatase are homodimeric. We proposed that the active form of CfbA is a homodimer based on its size and comparison with other chelatase homologs. It is possible that the dimer interface is close to either or both of N- and C-terminal regions of the protein and is perturbed by the His_6-tag. Alternatively, residues important for the activity of CfbA may be influenced by the infinity tag. Either way, an interesting chelatase was identified that can now be mechanistically and structurally studied.

We expressed CfbCD as a complex in *E. coli* instead of reconstituting it *in vitro*. This expression was carried out in a cell including pDB1282:*isc*, pRSFDuet-1:*cfbD-cfbC*, and pCDFDuet-1:*cfbC*. The pDB1282:*isc* vector contains eight genes related to iron-sulfur cluster assembly from *Azotobacter vinelandii* (*86*) (*87*). They are *hscA*, *iscS*, *fdx*, *hscB*, *iscU*, *orf3*, *ndk* and *iscA*. HscA is a bacterial Hsp70-class molecular chaperone that interacts with IscU (*88*). HscB is a cochaperone of HscA that regulates ATPase activity and binding specificity (*89*). IscS is a cysteine desulfurase that converts cysteine to alanine, and serves as a sulfur donor for iron

sulfur cluster construction (*90*) (*91*) (*92*). Fdx is a ferredoxin containing a [2Fe-2S] cluster (*93*). IscU functions with IscA as a scaffold during the construction process of iron sulfur clusters (*94*) (*95*). Orf3 functions as an iron donor (*96*). The *ndk* gene is located adjacent to *orf3* in *E. coli* (*91*). The CfbCD complex is homologous to nitrogenase and DPOR and has similar functionality. Compared with nitrogenase and DPOR, the CfbCD complex carries out more a complicated reaction, but has a simpler structure. Also, whether the γ-lactamization reaction is enzymatic or not is still a topic to be elucidated.

The *mcrD* gene is part of the *mcr* gene cluster and encodes a protein of unknown function. Some co-precipitation experiments have shown that McrD interacts with MCR physically, even though it is not required for MCR activity. Our McrD assay suggests that McrD could be a chaperone for coenzyme F430 delivery. If this postulation can be proven by further interaction experiments, it is likely required for *holo* MCR assembly. The phenomena that the inclusion of McrD enhances the conversion of 15,17^3-seco-F430-17^3-acid to coenzyme F430 also highlights a noteworthy observation, present in almost every step of coenzyme F430 biosynthesis experiments: enzymes catalyzing a subsequent reaction on the pathway helps alleviate product inhibition of the enzyme catalyzing the preceding step.

In summary, this project attempted to express and purify all of the hypothesized *cfb* encoded enzymes in a heterologous host in order to biosynthesize coenzyme F430 *in vitro*. According to the data and analysis discussed above, the goal has been accomplished and has provided several new questions to be theoretically and experimentally explored. The coenzyme

F430 biosynthesis pathway has been elucidated and shown to include five different enzymes and four unique steps. Each of the enzymes that have been discovered represents a new target for inhibitors of methanogenesis. Furthermore, this study not only sets a stable foundation for understanding the maturation of MCR, but also provides a possibility to express *holo* MCR in non-methanogenic hosts.

Chapter 3

Heterologous Expression of Multiple *cfb* Genes in *Escherichia coli* and *in vivo* Biosynthesis of Coenzyme F430

3.1 Background

Methane is a greenhouse gas that has approximately thirty times the global warming potential (GWP) of carbon dioxide (*73*). One of the main sources of methane is methanogenic archaea (*97*), which metabolize carbon dioxide and other C1 compounds via the methanogenesis pathway. Methyl groups on methyl coenzyme M are obtained from reduced carbon dioxide or methyl groups from other C1 compounds. Methane is released by the reaction between methyl coenzyme M and coenzyme B, which is catalyzed by methyl coenzyme M reductase (MCR) (*98*). MCR is a large 270 kDa heterohexamer ($\alpha_2\beta_2\gamma_2$) complex. MCR contains two active sites (*22*), each of which house a tightly bound coenzyme F430 molecule (*99*).

Methane is the main component of natural gas, which is often found accompanying petroleum in oil fields. However, natural gas is difficult and expensive to capture and transport during the extraction of petroleum. As a result, natural gas is often vented or flared, which wastes this important resource and contributes to global warming through the release of vast quantities of greenhouse gas (Figure 25) (*100*). Therefore, new strategies for methane conversion are needed to deal with this global crisis.

There has been great interest in engineering a non-methanogenic host for biological methane conversion to liquid biofuels and other high value products. MCR is an ideal research

target for this purpose, as a homolog has been discovered in anaerobic methanogenic archaea (ANME), where there is some evidence that it catalyzes the anaerobic oxidation of methane (AOM) (*23*). The study of heterologous *in vivo holo* MCR assembly could lead to the ability to engineer an organism for industrial methane conversion.

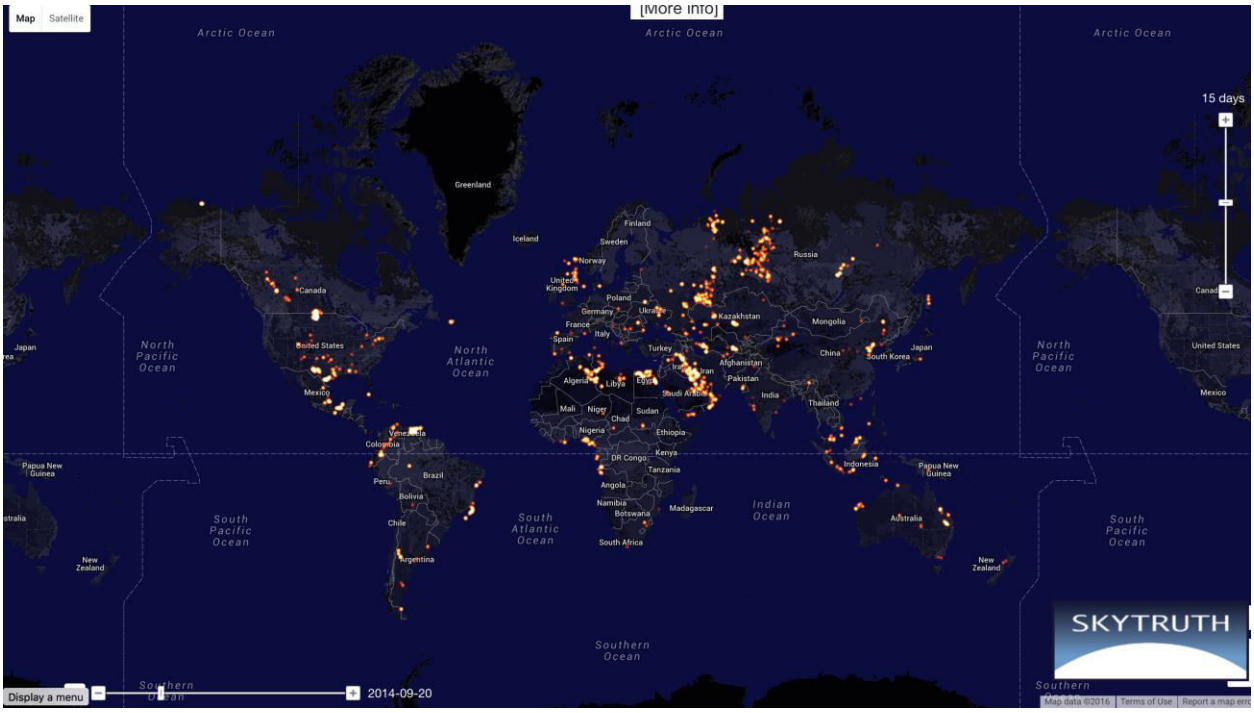


Figure 25: Infrared satellite heatmap showing the flaring of natural gas across the planet. It is difficult and expensive to transport natural gas. The venting and flaring of natural gas cause the release of vast amounts of greenhouse gas. Thus, novel approaches for natural gas conversion are significant for global warming and the energy crisis.

As discussed in Chapter 2, our group discovered that coenzyme F430 is biosynthesized from sirohydrochlorin using five different enzymes (*101*). All assays that proved the activity of these enzymes were performed *in vitro*. The obvious next step would be to try and biosynthesize coenzyme F430 *in vivo* using a non-methanogenic host through a coexpression strategy. However, there are still several obstacles preventing us from achieving this goal. For example, we used sodium dithionite as the reductant for the CfbCD catalyzed reaction, but this is not a viable *in vivo* reductant. Also, sufficient iron sulfur clusters must be provided to ensure activity of the CfbCD complex. Moreover, concentration of cofactors and ATP must be enhanced to a range that supports the energy intensive production of coenzyme F430. Thus, there are several significant challenges to overcome to enable the *in vivo* biosynthesis of coenzyme F430.

In many cases, ferredoxin serves as the physiological reductant for oxidoreductase enzymes. Initially we attempted to identify ferredoxin homologs in the *Methanosarcina acetivorans* C2A genome. We found that there are many homologs, and each is possibly involved in a variety of pathways, making identification of the correct homolog prohibitively difficult. Thus, we decided to test spinach ferredoxin (Fd) and its associated ferredoxin:$NADP^+$ reductase (Fnr), which is commercially available. These two proteins were tested for ability to support CfbCD catalysis in *in vitro* assays in place of sodium dithionite. Moreover, we constructed a compatible expression vector encoding this reducing system for use in the *in vivo* assays.

As noted, another difficulty for *in vivo* F430 biosynthesis is the heterologous construction of the iron sulfur clusters. Oftentimes *E. coli* lacks the required genes for the biogenesis of specialized cofactors for oxidoreductases, such as the iron sulfur clusters of nitrogenase. This phenomenon was observed when we tried to heterologously express CfbC and CfbD separately. The purified enzymes had low cluster content and lacked activity even when the cells were supplemented with iron and cysteine. The expression was improved by introducing pDB1282 vector, which includes all of the iron sulfur cluster (*isc*) biosynthetic genes from *Azotobacter vinelandii*. Thus, the pDB1282 vector was also included in the study of *in vivo* coenzyme F430 biosynthesis to ensure activity of CfbCD complex.

We selected the Duet plasmids system for our coexpression study. Duet vectors are expression vectors that utilize the T7 promoter and carry compatible replicons and antibiotic resistance markers. Each Duet vector is designed for carrying two exogenous genes. These features enable the Duet vector system to introduce eight exogenous genes in one cell for coexpression. We optimized this system and enabled each Duet vector to carry eight exogenous genes. Thus, a single *E. coli* cell is capable of containing thirty-two genes for coexpression. Based on this strategy, we sequentially introduced all of the *cfb* genes, *sirA* and *sirC*, the ferredoxin/Fnr genes, the *isc* genes, and a nickel transporter (*nixA*) from *Helicobacter pylori* together in *Escherichia coli* BL21(DE3).

3.2 Methods

3.2.1 *Construction of plasmids containing the cfb genes*

All plasmids were from Novagen EMD Millipore, Darmstadt, Germany. The *sirC* gene was amplified by PCR with primers that contained the NcoI and Hind III restriction enzyme sites. It was then ligated into the first multiple cloning site (MCS-I) of pETDuet after NcoI and Hind III digestion (Figure 26). The *cfbA* and *cfbE* genes were ligated into the first and second multiple cloning sites of pCDFDuet, respectively (Figure 26). The *cfbA* gene was ligated into the NcoI and BamHI restriction sites of MCS-I, while the *cfbE* gene was ligated into the NdeI and XhoI restriction sites of MCS-II. The *cfbB* and *sirA* genes were ligated into the MCS-I and MCS-II of pACYCDuet (Figure 26). The *cfbB* gene was ligated into the NcoI and BamHI restriction sites of MCS-I and the *sirA* gene was ligated into the NdeI and XhoI restriction sites of MCS-II. The *cfbD* and *cfbC* genes were ligated into the MCS-I and MCS-II of pRSFDuet (Figure 26). NcoI and Hind III restriction sites were utilized to ligate *cfbD* into MCS-I and the NdeI and XhoI restriction sites were utilized for the *cfbC* ligation. All primers are shown in Table 2.

The genes encoding Fd (LOC110805667) and Fnr (LOC110801208) were from *Spinacia oleracea*. The *Fd* gene was cloned directly from genomic DNA, while the *Fnr* gene was codon optimized and synthesized by Invitrogen (Thermo Fisher Scientific) in order to remove introns. The genomic DNA was directly extracted from *Spinacia oleracea* using the Plant DNA Extraction Protocol (*102*). We then constructed the pRSFDuet:*Fd-Fnr* vector using Gibson Assembly (see the Appendix for experimental details). The gene encoding Fd was amplified by

PCR using the Fd FW and Fd_ REVprimers with Phusion High-Fidelity DNA Polymerase (New England Biolabs). An empty pRSFDuet vector was utilized as a template vector for PCR during Gibson Assembly, which was amplified using the pRSFDuet_FW_MCSI and pRSFDuet_REV_MCSI primers. All PCR products were purified using double-tiered agarose gel electrophoresis (FlashGel System and FlashGel Recovery Cassettes from Lonza). The purified PCR products were then assembled using the Gibson Assembly Master Mix (New England BioLabs) following the manufacturer's protocol. The reacted mixture was transformed into NEB 5α competent cells (Invitrogen Thermo Fisher Scientific). The constructed pRSFDuet:*fd* vector was sequence verified by Eurofins Genomics. Then we PCR amplified the synthesized gene encoding Fnr from using the Fnr FW and fnr REV primers (Table 2). The constructed pRSFDuet:*fd* vector served as the template for a subsequent PCR using the pRSFDuet_FW_MCSIIand pRSFDuet_REV_MCSIIprimers (Table 2). The PCR products were again purified using agarose gel electrophoresis and assembled by Gibson Assembly. The final pRSFDuet:*Fd-Fnr*) vector was sequence verified by Eurofins Genomics.

In addition to the pRSFDuet:*Fd-Fnr* vector, we constructed a pACYC:*sirC-sirA* vector using Gibson Assembly. The *sirC* gene was cloned from pCR-Blunt:*sirC* using the ga sirc FW and ga sirc REV primers (Table 2). An empty pACYCDuet vector served as the template for PCR using the pACYCDuet_FW_MCSI and pACYCDuet_REV_MCSI primers (Table 2). The two PCR products were purified by agarose gel electrophoresis and assembled following the standard protocol. The constructed pACYCDuet:*sirC* vector was sequence verified by Eurofins Genomics. Then we used the restriction enzymes NdeI and XhoI to digest the pET-28: *sirA* and

gel purified the insert. The *sirA* gene was then ligated into the NdeI and XhoI restriction sites of pACYCDuet:*sirC* using T4 ligase (New England Bio Labs). Finally, we obtained a sequence verified (Eurofins Genomics) pACYCDuet:*sirC-sirA* vector. Using a similar strategy, we also constructed a pACYCDuet:*cfbB-nixA* vector. The *cfbB* gene was cloned from pCR-Blunt:*cfbB* using the *cfbB* FW and *cfbB* REV primers (Table 2). Then we used the restriction enzymes PciI and BamHI to digest *cfbB* the PCR products. We also used the restriction enzymes NcoI and BamHI to digest the empty pACYCDuet vector. The *cfbB* gene was then ligated into the NcoI and BamHI restriction sites of pACYCDuet vector using T4 ligase (New England Bio Labs). The *nixA* gene was cloned from synthesized gene encoding *nixA* using the ga nixA FW and ga nixA REV primers (Table 2). The pACYCDuet:*cfbB* vector served as the template for PCR using the pACYCDuet_FW_MCSII and pACYCDuet_REV_MCSII primers (Table 2). The two PCR products were purified by agarose gel electrophoresis and assembled following the standard protocol. Then we obtained a sequence verified (Eurofins Genomics) pACYCDuet:*cfbB-nixA* vector.

Gene	Primer
sirA	Forward: 5'-CGGCGCATATGTCAGAAAATTACGG-3' Reverse: 5'-ATGAGCTCGAGTCAGAAATCCTTTCCTGC-3'
sirC	Forward: 5'-ATATACCATGGCTGAAACAAATAATTTTC-3' Reverse: 5'-ATATAAAGCTTTTATTCGAGCTTATCCG-3' ga sirc FW: 5′-TTTAACTTTAATAAGGAGATATACCATGGCTGAAACAAAT AATTTT-3′ ga sirc REV: 5′- GTACAATACGATTACTTTCTGTTCGATTATTCGAGCTTATC CGAG-3′ pACYCDuet_FW_MCSI: 5′- TCGAACAGAAAGTAATCG-3′ pACYCDuet_REV_MCSI: 5′- GGTATATCTCCTTATTAAAGTTAAAC-3′
cfbA	Forward: 5'-GCGACTCATGACTGAGAAACTCGG-3' Reverse: 5'-ATTACGGATCCTTACAGGGCTTCCTG-3'
cfbB	Forward: 5'-ACACAACATGTCCCACAGCAAACAATC-3' Reverse: 5'-ATTAAGGATCCCTACCGGGGAGCCC-3'
cfbC	Forward: 5'-CGCTGCATATGAAAAAACAAAAGATCGTTGC-3' Reverse: 5'-ATTATGGCCGGCCTTATTTTGTCATTTCCC-3'
cfbD	Forward: 5′-CGCCGTCATGACTCAAAAAGAGATCTC-3′ Reverse: 5′-ATCACAAGCTTTCAGGCTTCTTTTGCAAC-3′
cfbE	Forward: 5'-GACACCATATGGACCTGTTCCGG-3' Reverse: 5'-CGCACCTCGAGTTAACGGAAACATTTC-3'
Fd	Fd FW: 5′-GATATACCATGGATGGCTGCCTACAAGG-3′ Fd_ REV: 5′- TCGACTTAAGCATCAGGCAGTAAGCTCCTC-3′ pRSFDuet_FW_MCS: 5′- CTGCCTGATGCTTAAGTCGAACAGAAA-3′ pRSFDuet_REV_MCSI: 5′- GCAGCCATCCATGGTATATCTCCTTATTAAA-3′
Fnr	Fnr FW: 5′- AAGGAGATATACATATGCAGATTGCCAGTGATG-3′ fnr REV: 5′- ACCAGACTCGAGCTCGAGTTAATACACTTC-3′ pRSFDuet_FW_MCSII: 5′- TTAACTCGAGCTCGAGTCTGGTAAAGAA-3′ pRSFDuet_REV_MCSII: 5′- GGCAATCTGCATATGTATATCTCCTTCTTATACTTAAC-3′
nixA	ga nixA FW: 5′- TATTAGTTAAGTATAAGAAGGAGATATACATATGAAGCTG TGGTTCCC-3′ ga nixA REV: 5′- ATTAAGCTGCGCTAGTAGATTAGCTTTCCAGTTTGCTG-3′ pACYCDuet_FW_MCSII: 5′- TCTACTAGCGCAGCTTAAT-3′ pACYCDuet_REV_MCSII: 5′-ATGTATATCTCCTTCTTATACTTAACTAATA-3′
	SwaI Overlap Fwd: 5′- GGGATCTCGACGCTCTCCCTTATTT-3′ SwaI Overlap Rev: 5′- CCTAATGCAGGATTTGCTAGTTATTGCTCAGCGG -3′

Table 2: Primers used for plasmid construction for the *in vivo* coenzyme F430 biosynthesis experiments.

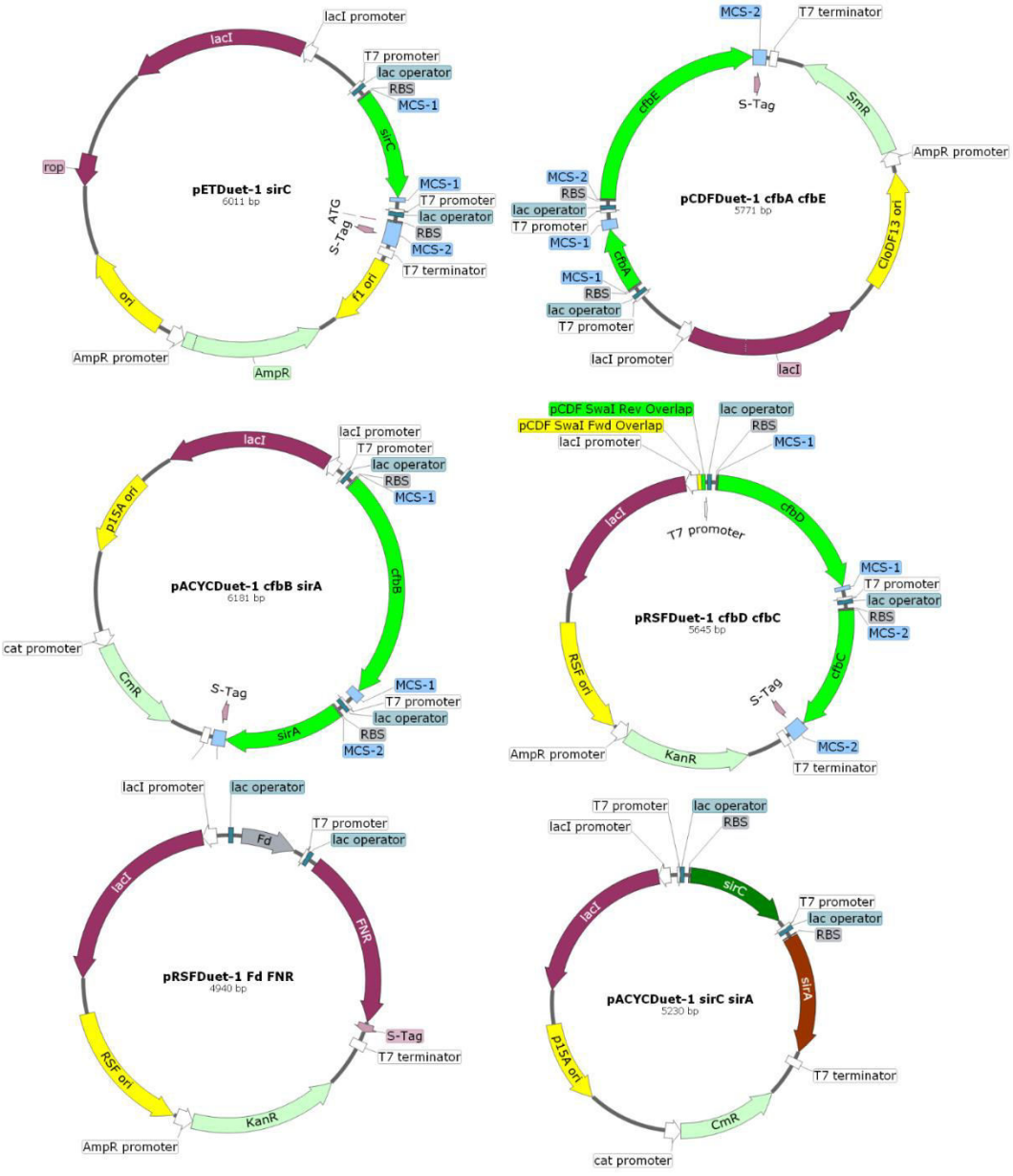


Figure 26: Plasmids constructed for the coexpression of the *cfb* and accessory genes.

3.2.2 *Superplasmid construction*

The pCDFDuet:*cfbA-cfbE*, pRSFDuet:*cfbC-cfbD*, and pACYCDuet:*cfbB-nixA* vectors were the starting material for the construction of a *cfb* 'superplasmid'. The pCDFDuet:*cfbA-cfbE* vector was used as the base plasmid. We introduced a SwaI restriction enzyme site approximately 100 bp upstream of the MCS-I site of pCDFDuet:*cfbA-cfbE*, pRSFDuet:*cfbC-cfbD*, and pACYCDuet:*cfbB-nixA*. We treated SwaI-pCDFDuet:*cfbA-cfbE* with the SwaI restriction enzyme (New England BioLabs) to linearize the plasmid. The SwaI Overlap Fwd and SwaI Overlap Rev were used to clone the *cfbB-nixA* fragment from pACYCDuet:*cfbB-nixA* using PCR. The two fragments were purified by agarose gel electrophoresis (Lonza) and were combined using Gibson Assembly. The constructed four gene superplamid SwaI-pCDFDuet:*cfbB-nixA-cfbA-cfbE* was sent to Eurofins Genomics for sequence verification. The restriction enzyme SwaI was then used to digest SwaI-pCDFDuet:*cfbB-nixA-cfbA-cfbE* for Gibson Assembly with the *cfbC-cfbD* fragment obtained via PCR using the primers SwaI Overlap Fwd and SwaI Overlap Rev. After transformation, plasmid preparation, and sequence verification, we obtained the SwaI-pCDFDuet:*cfbC-cfbD-cfbB-nixA-cfbA-cfbE* construct (Figure 27).

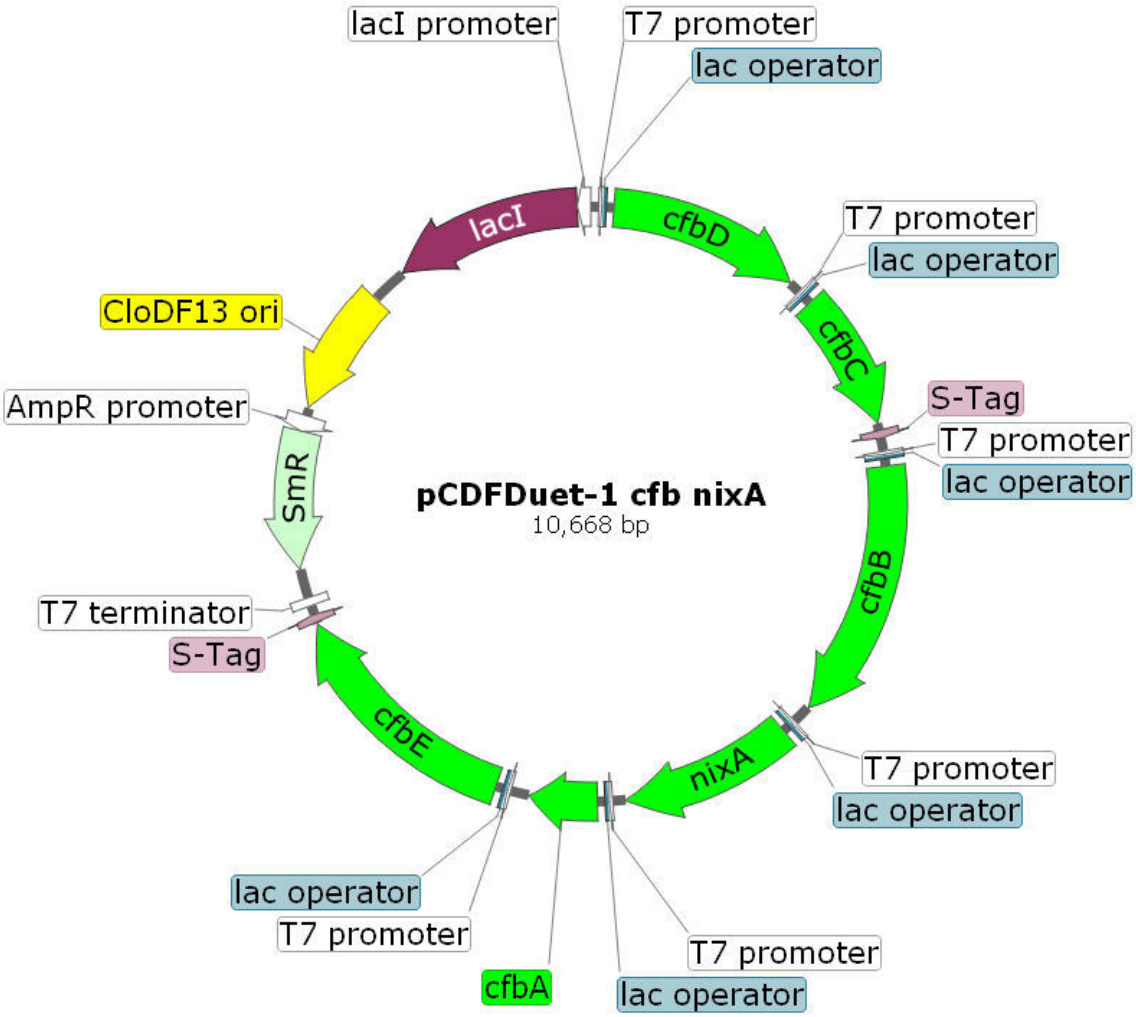


Figure 27: Diagram of a superplasmid containing *nixA* and all of the *cfb* genes.

3.2.3 *Supercell construction*

We utilized *E. coli* BL21(DE3) cells (New England Labs) as the host organism for our *in vivo* coenzyme F430 biosynthesis experiments. A chemical competency protocol was used to

sequentially transform this strain with several different multi-gene plasmids. *E. coli* BL21(DE3) cells were cultured in 10 mL of lysogeny broth with the proper antibiotic for overnight growth after a plasmid was transformed into it and confirmed to be present using colony PCR. Then the cultures were transferred to 20 mL of LB with a 1:100 ratio. The cultures were chilled on ice for 30 min when the OD_{600} was between 0.2 to 0.5. The cultures were then centrifuged for 10 min in a microcentrifuge at 5500 rcf to collect the cell pellets. Next, 5 mL of a sterilized 0.1 M $CaCl_2$ solution was prepared and used to resuspend the cells. The resuspension mixture was chilled on ice for 30 min, then centrifuged again for 10 min at 5500 rcf. The pellets were resuspended again and chilled on ice for another 30 min. Then the cells were pelleted and resuspended in 0.8 mL of a sterilized 0.1 M $CaCl_2$ solution containing 15% glycerol. Finally, the resulting chemically competent *E. coli* BL21(DE3) cells harboring the desired plasmid were separated into 100 uL aliquots.

We sequentially transformed the pETDuet:*sirC*, pRSFDuet:*cfbD-cfbC*, pCDFDuet:*cfbA-cfbE*, and pACYCDuet:*cfbB-sirA* vectors into one *E. coli* BL21(DE3) cell. This cell has four selectable markers for antibiotic resistance, which allows it to grow on plates and media containing ampicillin, kanamycin, spectinomycin, and chloramphenicol.

3.2.4 *Coexpression of F430 genes and tetrapyrrole extraction*

Both Terrific Broth (TB) and LB were utilized as the expression media. We used four different combinations of supplementary chemicals for the experiments. Supply 1 consisted of nickel chloride (200 μM final concentration), glutamate (100 μM final concentration), 5-

aminolevulinic acid (5-ALA) (10 μM final concentration), and SAM (1 mM final concentration). Supply 2 consisted of nickel chloride (200 μM final concentration), ferrous sulfate (200 μM final concentration), methionine (50 μM final concentration), cysteine (200 μM final concentration), glutamate (100 μM final concentration), 5- ALA (10 μM final concentration), and SAM (1 mM final concentration). Supply 3 consisted nickel chloride (200 μM final concentration), ferrous sulfate (200 μM final concentration), cobalamin (2 μM final concentration), riboflavin (50 μM final concentration), methionine (50 μM final concentration), cysteine (200 μM final concentration), histidine (100 μM final concentration), glutamate (100 μM final concentration), 5-ALA (10 μM final concentration), and SAM (1 mM final concentration). The supplemental chemicals were prepared in powder form before expression. The antibiotics ampicillin (100 μg/mL final concentration), kanamycin (50 μg/mL final concentration), spectinomycin (25 μg/mL final concentration), and chloramphenicol (35 μg/mL final concentration) were premixed with the media. Expression of the Cfb enzymes, SirA, and SirC was carried out by incubation at 37 °C until the OD_{600} was approximately 0.6. Then all of the supplemental chemicals and isopropyl-β-D-1-thiogalactopyranoside (IPTG) were added to the media. The final concentration of IPTG was 0.4 μM, and the temperature was decreased to 18 °C. After approximately 16 h of incubation, the temperature was raised back to 37 °C for 2-4 more hours.

The cell pellets were collected using centrifugation at 15900 ×g for 10 min. The weight of cell pellets was then recorded. In order to extract the tetrapyrrole from the cell pellet, we mixed 1 g of cells with 1 mL of methanol under anaerobic conditions. The mixture was then centrifuged at 7000 ×g for 30 min after a 1 h incubation. Finally, the supernatant was filtered

using a 0.22 μm filter. The filtrated supernatant was diluted two times using methanol. It was then analyzed using high-performance liquid chromatography (HPLC) after another 30 min centrifugation at 7000 ×g.

3.2.5 *HPLC and LC-MS analysis*

HPLC analysis was performed on an Agilent 1260 Infinity Quaternary LC System using a Diode Array Detector (DAD) VL+. The column was an Agilent Poroshell 120 EC-C18 (4.6 ×150 mm, 2.7 μm) reverse phase column. We utilized the Agilent OpenLAB ChemStation Edition software for data analysis. The chromatographic method for characterization of the coenzyme F430 biosynthetic reactions consisted of the following gradient of water (solvent A) and acetonitrile (solvent B), with each containing 0.5% formic acid: 0% B for 2 min, 0-20% B over 3 min, 20% B for 5 min, 20-25% B over 5 min, 25% B for 5 min, 25-30% B over 5 min, and 30-100% B over 5 min. The flow rate was 1.0 mL/min. The chromatogram was monitored at 400 nm. LC-MS analysis was carried out using a Waters Acquity UPLC/Q-TOF Premier Mass Spectrometer equipped with the identical Agilent Poroshell 120 EC-C18 column. The software used for data analysis was Waters MassLynx MS. The LC method consisted of the same solvent system and gradient as described above. The electrospray ionization (ESI) mass detector was applied with positive ion mode scanning between 0-1100 *m*/*z*. The chromatography detector was an inline Tunable UV (TUV) detector, set to 400 nm to match the peaks observed in the mass chromatograms to those observed by HPLC analysis.

3.3 Results

3.3.1 *Sirohydrochlorin in vivo experiments*

We utilized the *E. coli* BL21(DE3) strain as the host for enzymatic coexpression. The genome of *E. coli* BL21(DE3) contains early stage tetrapyrrole biosynthetic genes, including *hemA, hemL*, *hemB*, *hemC* and *hemD*, giving *E. coli* BL21(DE3) the ability to biosynthesize uroporphyrinogen III from glutamic acid. Since the precursor of the *cfb* pathway is sirohydrochlorin, theoretically a precorrin 2 synthase (SirA) and a precorrin 2 dehydrogenase (SirC) are required. We constructed the pACYCDuet:*sirC-sirA* vector to be transformed into *E. coli* BL21(DE3), because this strain lacks homologs of *sirC* and *sirA*, instead containing the trifunctional *cysG* gene that converts uroporphyrinogen III directly to siroheme.

The *E. coli* BL21(DE3) cells containing the pACYCDuet:*sirC-sirA* vector were supplied with SAM, 5-ALA, and glutamic acid. The cell pellets were collected after incubation and induction with IPTG. In contrast to the normal tan color of *E. coli*, the pellet color was green. Then the cell extracts were obtained by the protocol discussed above. We analyzed the cell extracts using HPLC and found a main peak with a retention time of 16.1 minutes. The UV-vis spectrum of the peak was consistent with the spectroscopic properties of sirohydrochlorin (Figure 28). Thus, *E. coli* BL21(DE3) can biosynthesize sirohydrochlorin, the precursor to the coenzyme F430 biosynthesis pathway, after introduction of exogenous *sirA* and *sirC* genes from methanogens.

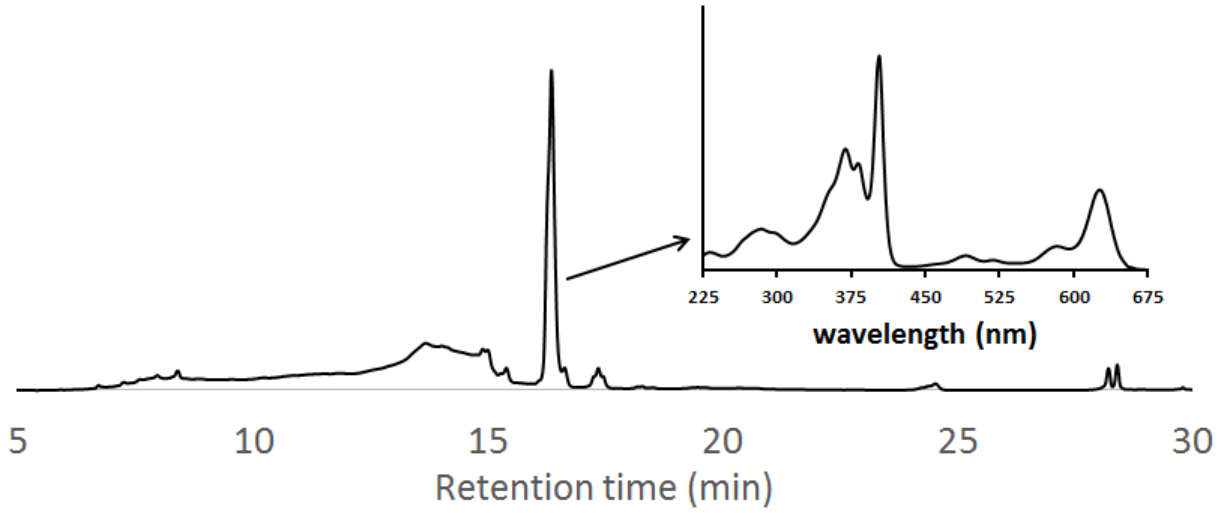


Figure 28: HPLC and UV-vis spectroscopy of cell free extracts from *E. coli* BL21(DE3) cells expressing the genes contained within the pACYCDuet:*sirC-sirA* vector.

We also sequentially transformed pETDuet:*sirC*, pRSFDuet:*cfbC-cfbD,* and pACYCDuet:*cfbB-sirA* plasmids in *E. coli* BL21(DE3). The cells were incubated with nickel chloride, ferrous sulfate, cobalamin, riboflavin, methionine, cysteine, histidine, glutamine, 5-aminolevulinic acid (5-ALA), and porphobilinogen (PBG). Cell extracts were collected using the above protocol and analyzed using HPLC. The results showed that a considerable amount of sirohydrochlorin was synthesized. no other compounds in the *cfb* pathway were observed because of no CfbA expressed.

3.3.2 *Coexpression of all cfb genes*

The feasibility of sirohydrochlorin biosynthesis in *E. coli* has been demonstrated by the experimental results discussed above. In this section, the experimental feasibility of coenzyme

F430 biosynthesis *in vivo* in a non-methanogen host by including all identified *cfb* genes will be examined. An *E. coli* BL21(DE3) cell line was constructed by sequentially transforming it with the pETDuet:*sirC*, pRSFDuet:*cfbC-cfbD*, pACYCDuet:*cfbB-sirA,* and pCDFDuet:*cfbA-cfbE* plasmids. This "supercell" containing 4 plasmids was cultured in TB using the protocol of F430 coexpression, which means expressing multiple enzymes in one cell for *in vivo* study. All supplementary chemicals were the same as described in the Methods section above. The harvested cell pellets were dark purple in color. Cell extracts characterized by HPLC and LC-MS contained Ni-sirohydrochlorin *a*,*c*-diamide in high abundance (Figure 29). Unfortunately, we observed no detectable 15,17^3-seco-F430-17^3-acid or coenzyme F430 synthesis.

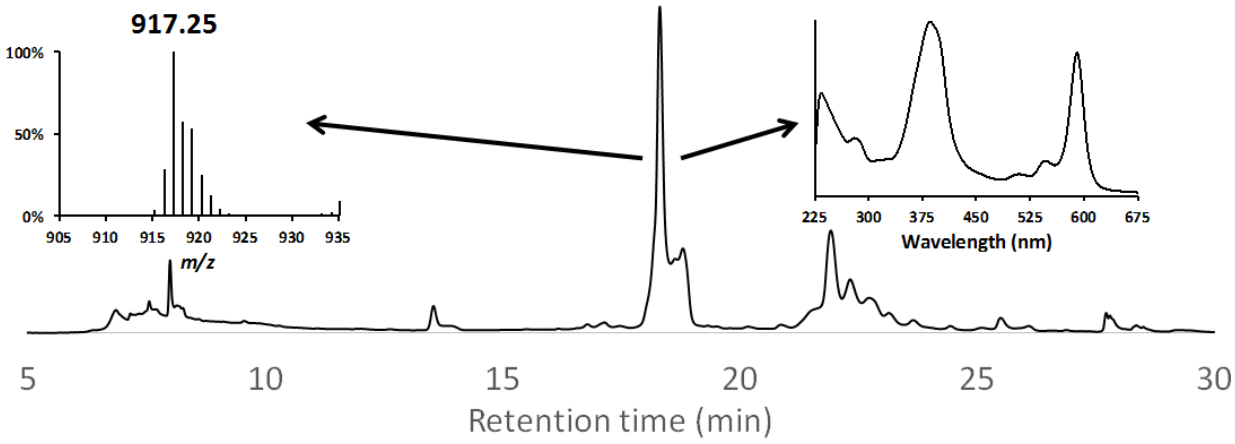


Figure 29: HPLC, LC-MS, and UV-vis spectroscopy of cell free extracts from *E. coli* BL21(DE3) cells expressing *sirA*, *sirC*, and all of the *cfb* genes. The 17.1 min peak observed by HPLC has a UV-vis spectrum and a 917.25 *m/z* that are consistent with the formation of Ni-sirohydrochlorin *a*,*c*-diamide in high abundance.

3.3.3 *Biological reductant for the CfbCD complex*

Since the *in vivo* biosynthesis of coenzyme F430 stopped after the second step, we reasoned that a biological reductant was needed to activate the CfbCD complex. We acquired ferredoxin (Fd) and ferredoxin:$NADP^+$ reductase (Fnr) from spinach and tested whether these enzymes could function as the reducing system to activate the complex. HPLC analysis revealed that Fd from spinach can indeed reduce the CfbCD complex to the active state using NADPH as the reductant, as $15,17^3$-seco-F430-17^3-acid was synthesized (

Figure 30). Additional evidence of the ability of spinach Fd to reduce the CfbCD complex was provided by EPR spectroscopy. The $[4Fe\text{-}4S]^{2+}$ cluster in CfbC, which is EPR silent, was successfully reduced to $[4Fe\text{-}4S]^{1+}$ state in the presence of Fd, Fnr, and NADPH (Figure 31).

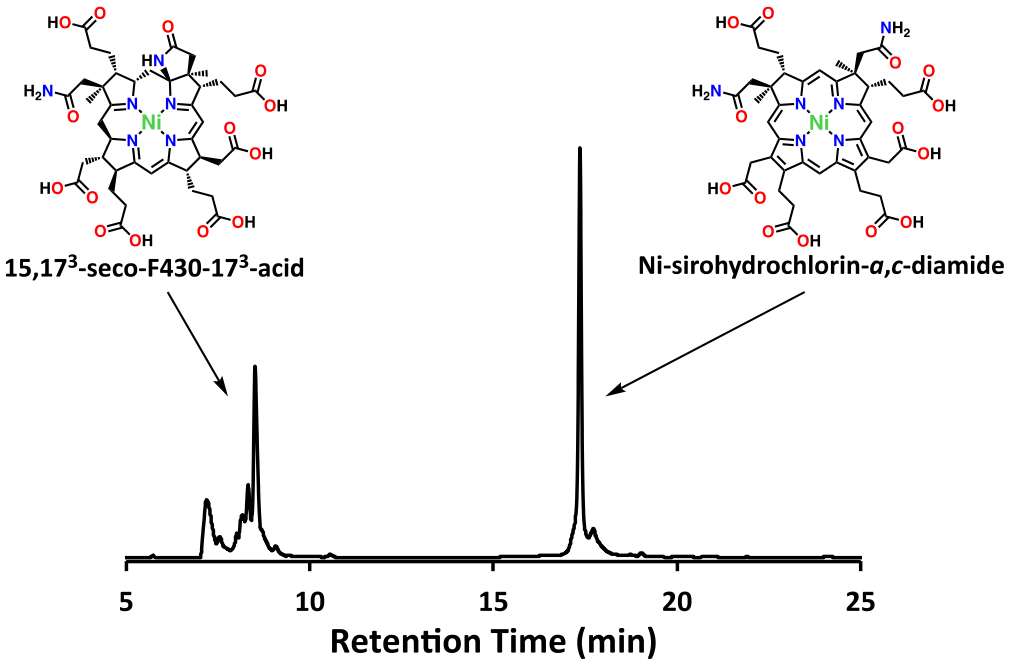


Figure 30: HPLC analysis of *in vitro* CfbCD assays using Fd-Fnr and NADPH as the reducing system.

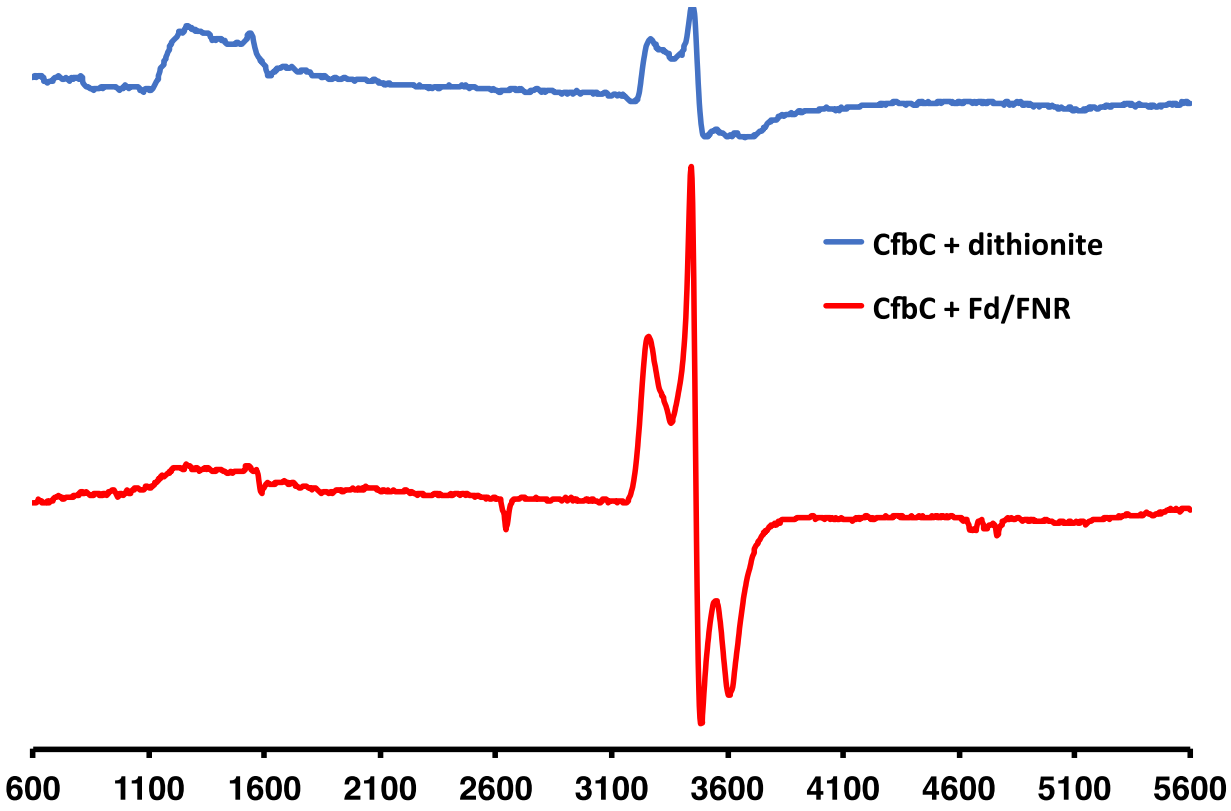


Figure 31: EPR spectroscopy of reduced CfbC. The blue trace was obtained from samples treated with sodium dithionite and the red trace was obtained from samples treated with NADPH and the Fd-Fnr system.

3.3.4 *Coenzyme F430 in vivo biosynthesis*

We constructed four different cell lines in order to examine the genetic requirements of the heterologous biosynthesis of coenzyme F430 *in vivo*. They are the *cfb* cell line (*E. coli* BL21(DE3) containing the pACYCDuet:*sirC-sirA* and pCDFDuet:*cfb-nixA* plasmids), the *Fd-cfb* cell line (*E. coli* BL21(DE3) containing the pACYCDuet:*sirC-sirA*, pCDFDuet:*cfb-nixA*, and

pRSFDuet:*Fd-Fnr* plasmids), the *isc-cfb* cell line (*E. coli* BL21(DE3) containing pACYCDuet:*sirC-sirA*, pCDFDuet:*cfb-nixA*, and pDB1282:*isc* plasmids), and the *Fd-isc-cfb* cell line (*E. coli* BL21(DE3) containing pACYCDuet:*sirC-sirA*, pCDFDuet:*cfb-nixA*, pRSFDuet:*Fd-Fnr,* and pDB1282:*isc* plasmids).

First, the *cfb* cell line was incubated with supplementary chemicals (supply 1, 2, or 3 described above) in different coexpression experiments. Cell pellets from these experiments were harvested separately and the cell free extracts were analyzed using HPLC (Figure 32). The purpose of these experiments was to verify that pCDFDuet:*cfb-nixA* is able to function as well as the separated Duet vectors comprising all *cfb* genes together in one cell.

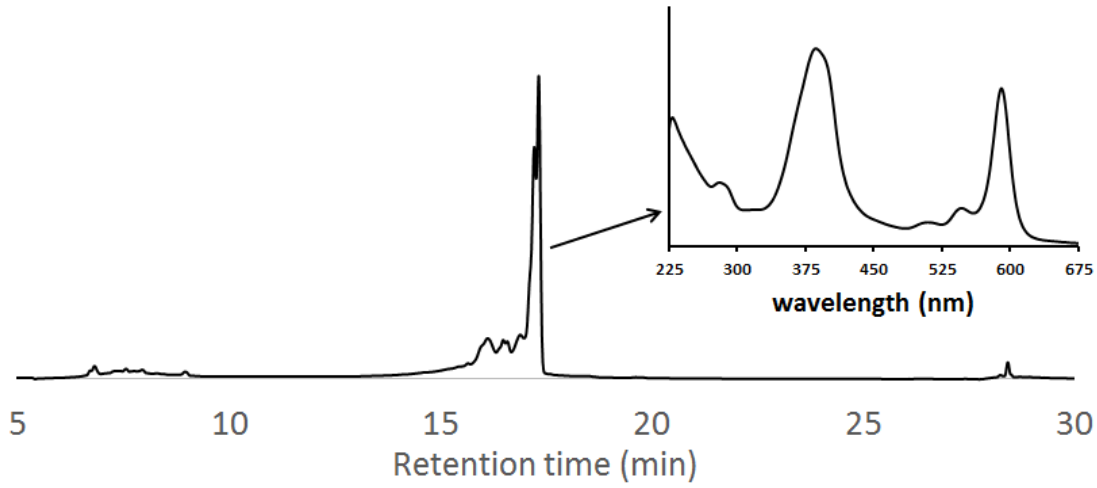


Figure 32: HPLC and UV-vis spectroscopy of cell free extracts from the *cfb* cell line obtained with chemical supply 1. These cells contain *sirA*, *sirC, nixA*, and all of the *cfb* genes.

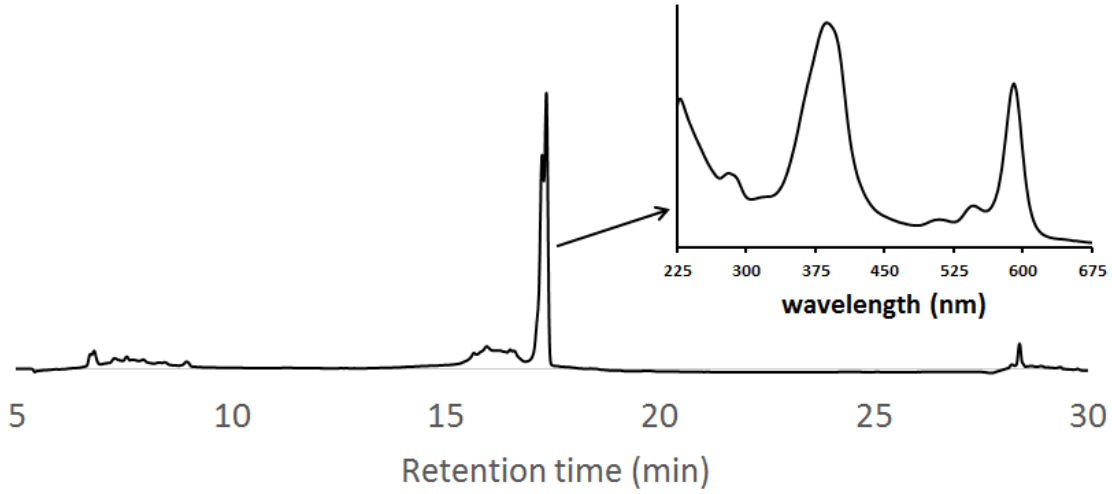


Figure 33: HPLC and UV-vis spectroscopy of cell free extracts from the *cfb* cell line obtained with chemical supply 2. These cells contain *sirA*, *sirC, nixA*, and all of the *cfb* genes.

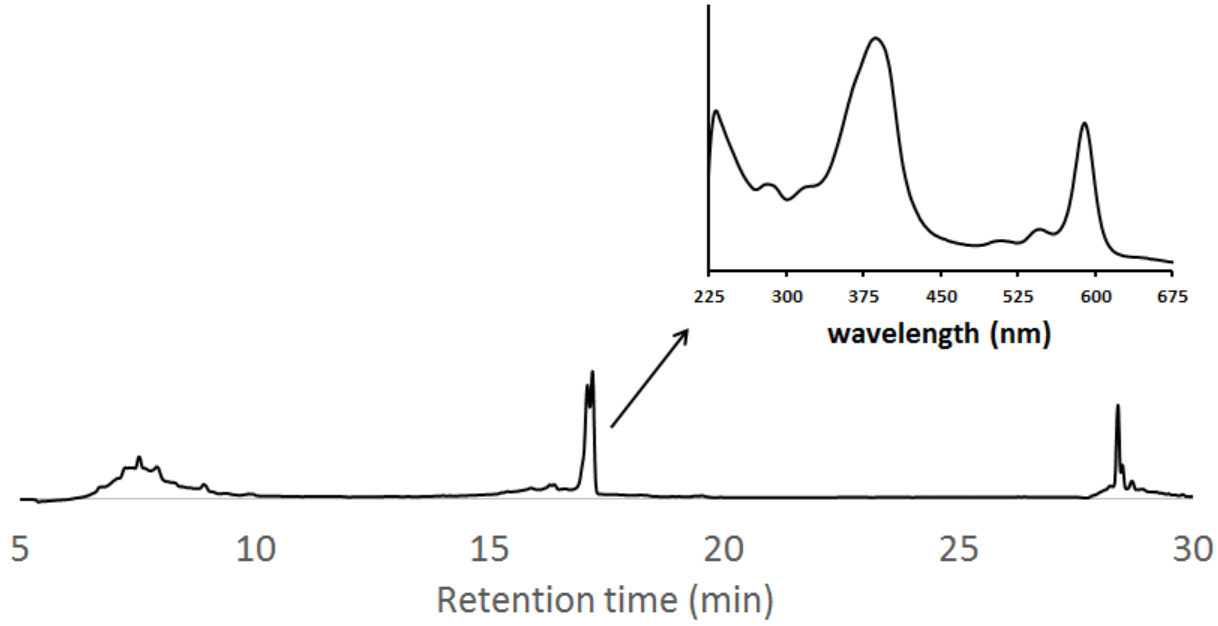


Figure 34: HPLC and UV-vis spectroscopy of cell free extracts from the *cfb* cell line obtained with chemical supply 3. These cells contain *sirA*, *sirC, nixA*, and all of the *cfb* genes.

Cell free extracts from the *cfb* cell line that were supplemented with chemical supply 1, 2, or 3 all produced a product in high abundance whose HPLC retention time and UV-vis spectrum are consistent with Ni-sirohydrochlorin *a*,*c*-diamide. However, preliminary experiments suggest that supply 1 and 2 provided a better yield of Ni-sirohydrochlorin *a*,*c*-diamide even with less variety of supplementary chemicals.

Then, we coexpressed ferredoxin and ferredoxin:$NADP^+$ reductase with all of the *cfb* genes encoded enzymes by the *Fd-cfb* cell line. Similarly, supplementary chemical supply 1, 2, or 3 were added to the TB medium, and cell free extracts were collected. HPLC was used to analyze any tetrapyrrole compounds that were biosynthesized ([35-37]). These experiments aimed to verify whether the reducing system was the determining factor for CfbCD complex activity in heterologous coexpression. A high yield of Ni-sirohydrochlorin *a*,*c*-diamide was detected in cell free extracts of the *Fd-cfb* cell line supplemented with chemical supply 1, 2, or 3.

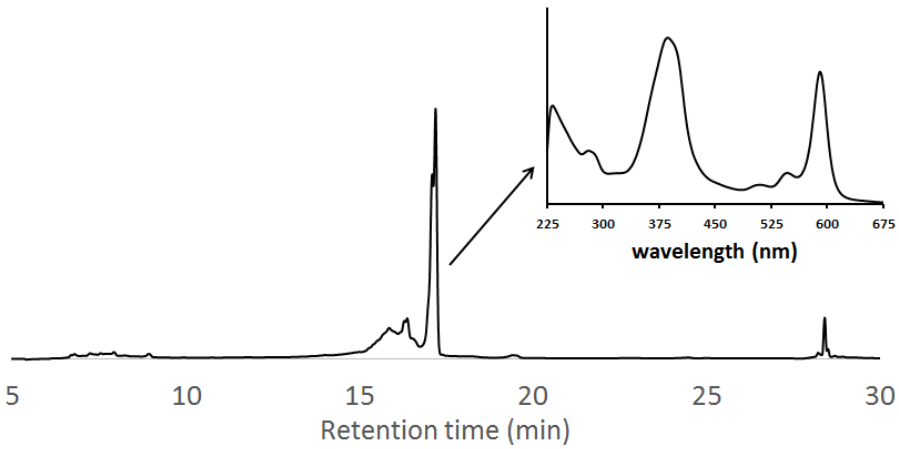


Figure 35: HPLC and UV-vis spectroscopy of cell free extracts from the *Fd-cfb* cell line obtained with chemical supply 1. These cells contain *sirA*, *sirC, nixA, Fd, Fnr*, and all of the *cfb* genes.

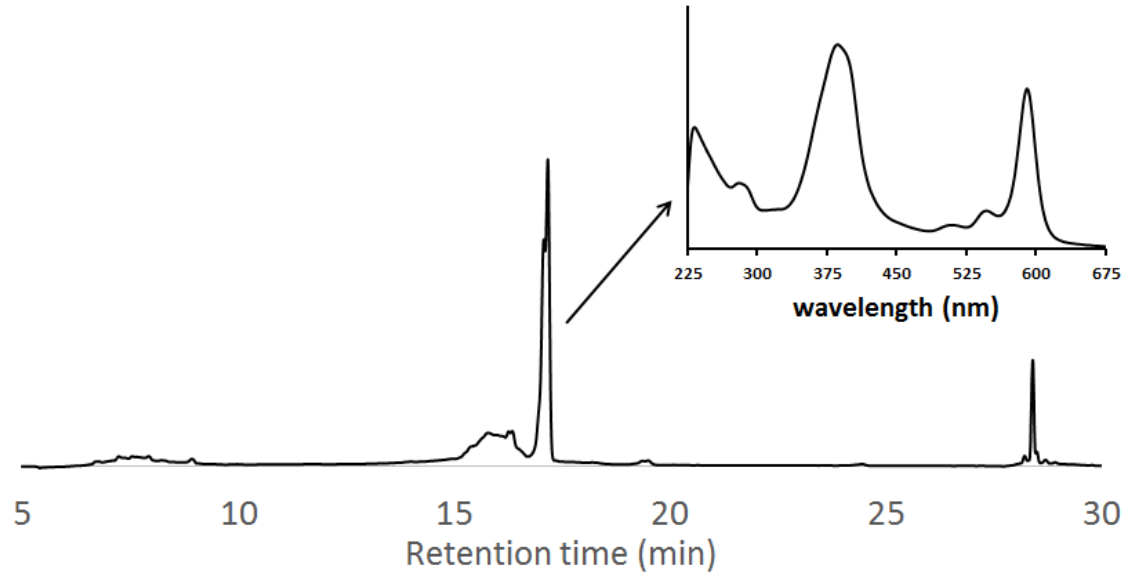


Figure 36: HPLC and UV-vis spectroscopy of cell free extracts from the *Fd-cfb* cell line obtained with chemical supply 2. These cells contain *sirA*, *sirC, nixA, Fd, Fnr*, and all of the *cfb* genes.

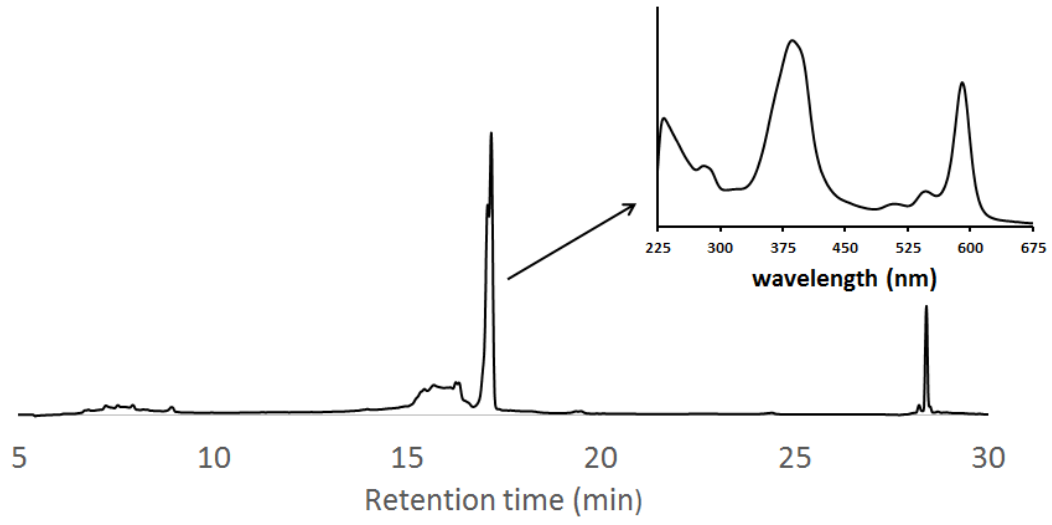


Figure 37: HPLC and UV-vis spectroscopy of cell free extracts from the *Fd-cfb* cell line obtained with chemical supply 3. These cells contain *sirA*, *sirC, nixA, Fd, Fnr*, and all of the *cfb* genes.

In addition, the *isc* genes were expressed from the pDB1282:*isc* vector to construct the iron-sulfur cluster for the CfbCD complex in the *isc-cfb* cell line. Cells were supplemented with chemical supply 1, 2, or 3 and grown in TB medium. Cell free extracts were obtained and analyzed by HPLC (Figure 38Figure 40). These experiments were intended to verify whether coexpression with exogenous *isc* enzymes is indispensible for CfbCD complex activity in heterologous coexpression.

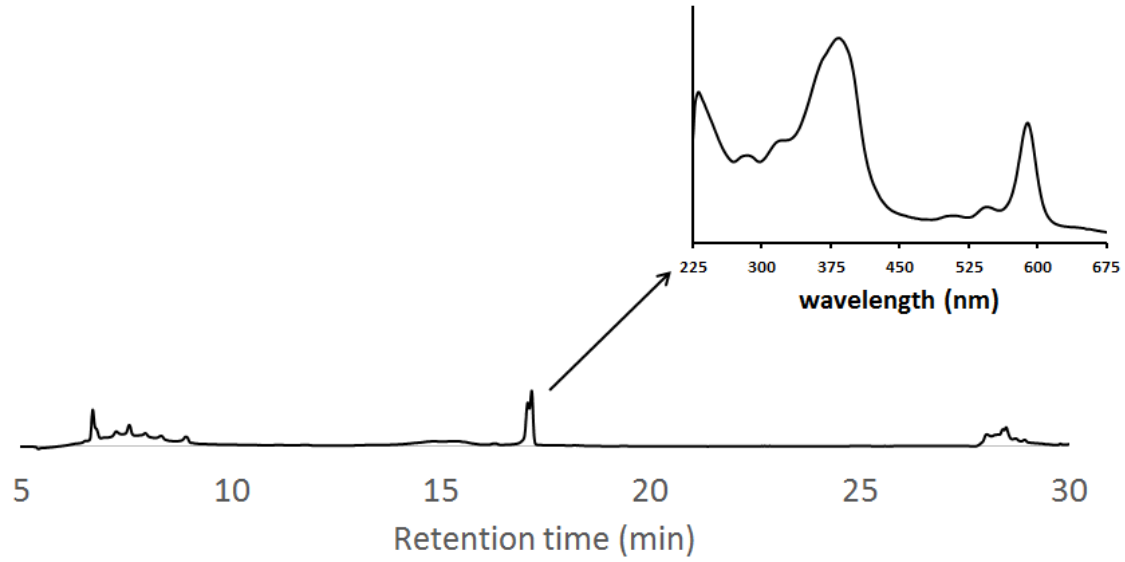


Figure 38: HPLC and UV-vis spectroscopy of cell free extracts from the *isc-cfb* cell line obtained with chemical supply 1. These cells contain *sirA*, *sirC, nixA*, the *isc* cluster, and all of the *cfb* genes.

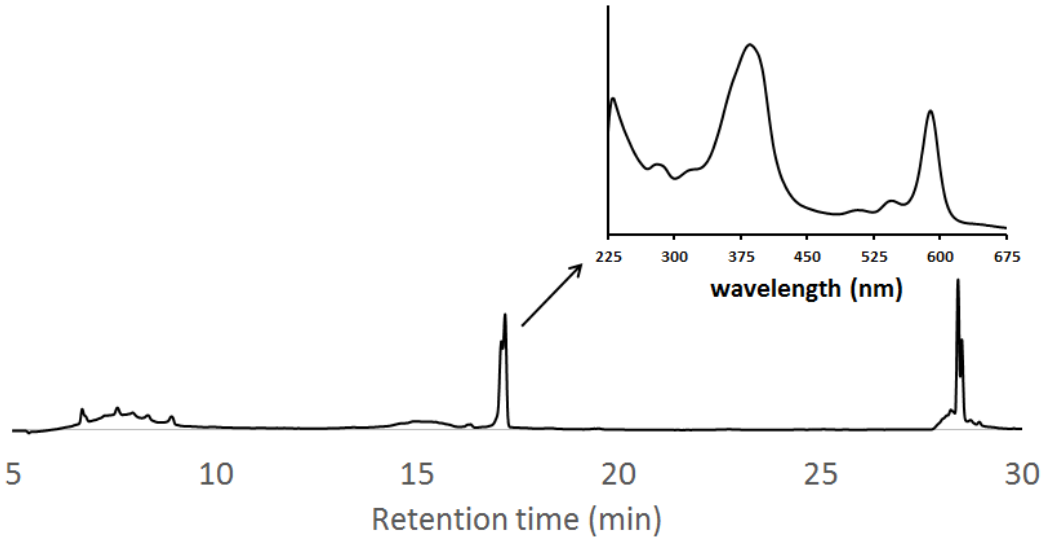


Figure 39: HPLC and UV-vis spectroscopy of cell free extracts from the *isc-cfb* cell line obtained with chemical supply 2. These cells contain *sirA*, *sirC, nixA*, the *isc* cluster, and all of the *cfb* genes.

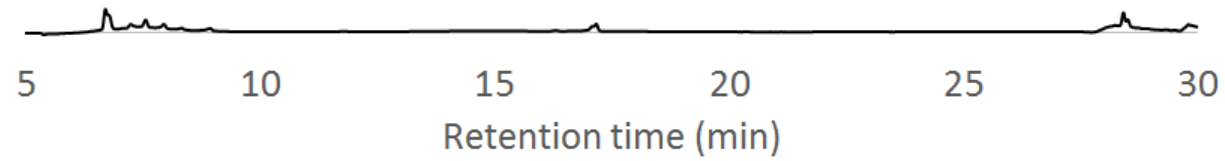


Figure 40: HPLC analysis of cell free extracts from the *isc-cfb* cell line obtained with chemical supply 3. These cells contain *sirA*, *sirC, nixA*, the *isc* cluster, and all of the *cfb* genes.

Cell free extracts of the *isc-cfb* cell line have been analyzed by HPLC. The results indicate that a low amount of Ni-sirohydrochlorin *a*,*c*-diamide was biosynthesized. Trace amounts of this product were observed in samples from cells supplemented with chemical supply 3.

Finally, all of the genes (*sirA*, *sirC, nixA*, *Fd*, *Fnr*, the *isc* cluster, and all of the *cfb* genes) were included in the *Fd-isc-cfb* cell line. The enzymes encoded by these genes were expressed with supplementary chemical supply 1, 2, or 3 in TB medium. We obtained cell free extracts and analyzed them using HPLC (Figure 41Figure 43).

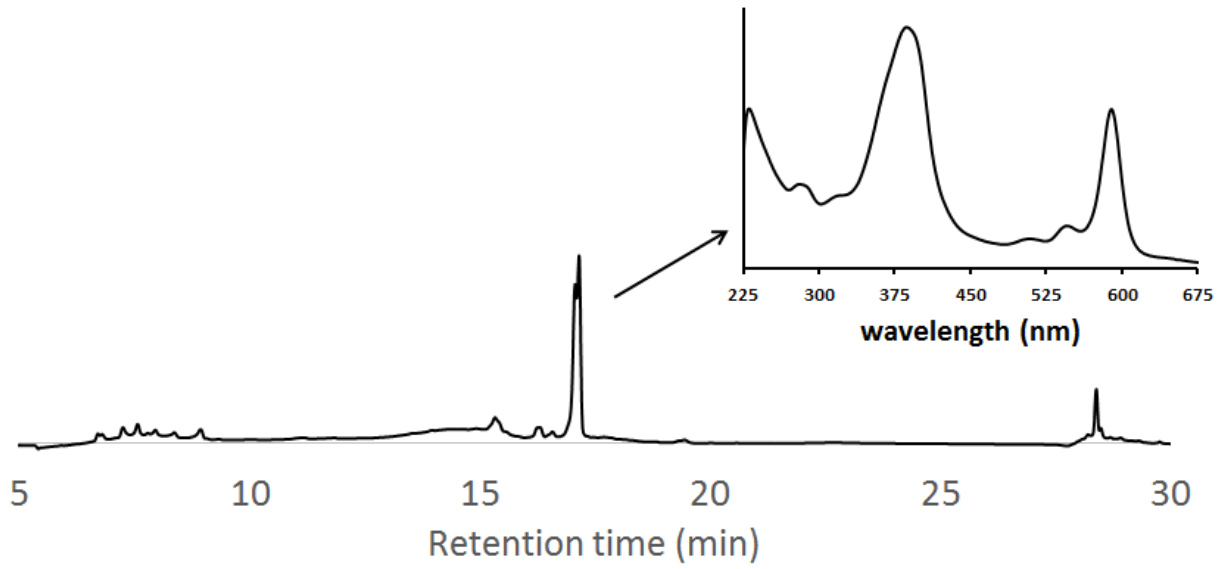


Figure 41: HPLC and UV-vis spectroscopy of cell free extracts from the *Fd-isc-cfb* cell line obtained with chemical supply 1. These cells contain *sirA*, *sirC, nixA*, *Fd*, *Fnr*, the *isc* cluster, and all of the *cfb* genes.

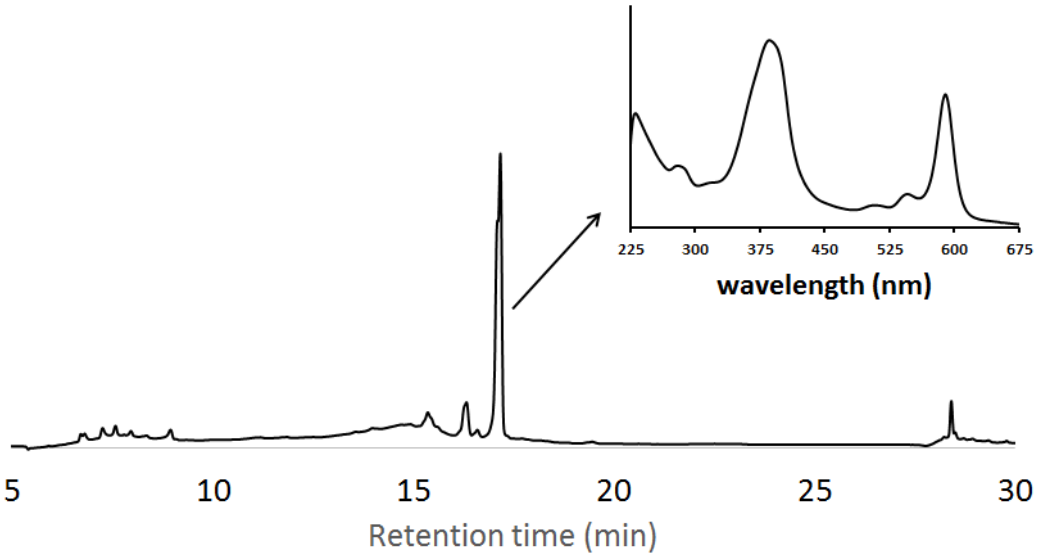


Figure 42: HPLC and UV-vis spectroscopy of cell free extracts from the *Fd-isc-cfb* cell line obtained with chemical supply 2. These cells contain *sirA*, *sirC, nixA*, *Fd*, *Fnr*, the *isc* cluster, and all of the *cfb* genes.

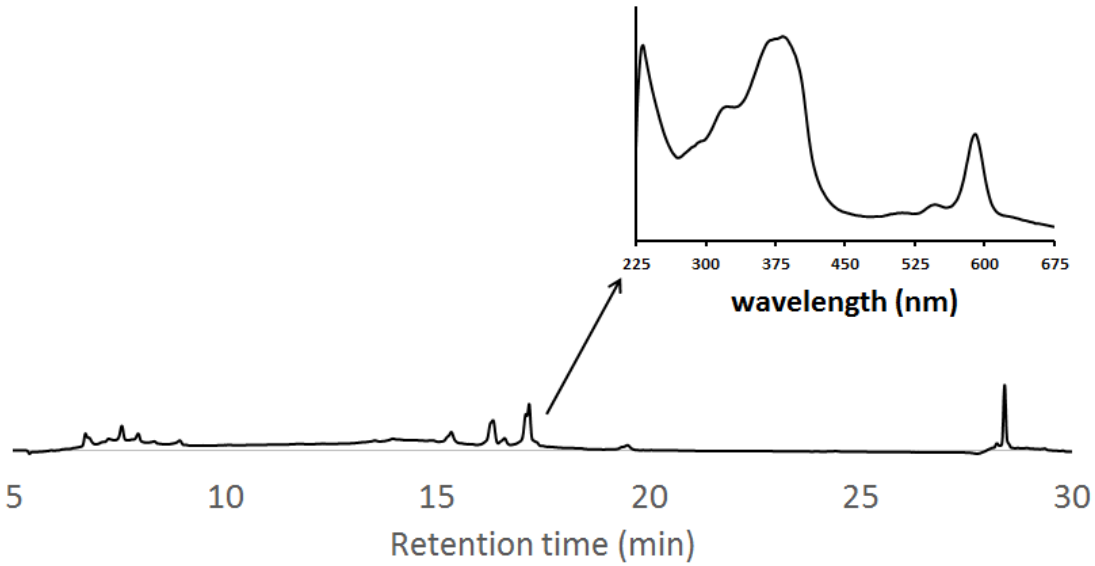


Figure 43: HPLC and UV-vis spectroscopy of cell free extracts from the *Fd-isc-cfb* cell line obtained with chemical supply 3. These cells contain *sirA*, *sirC, nixA*, *Fd*, *Fnr*, the *isc* cluster, and all of the *cfb* genes.

These preliminary experiments suggest that more Ni-sirohydrochlorin *a*,*c*-diamide is present in the cell extracts of the *Fd-isc-cfb* cell line than in the *isc-cfb* cell line and supplementary chemical supply 2 provide the greatest yield. However, additional experiments are still needed to verify this observation. Unfortunately, again no 15,17^3-seco-F430-17^3-acid or coenzyme F430 were detected.

3.4 Discussion

Coenzyme F430 is a significant cofactor in methanogenesis that is critical for the activity of MCR. Our group discovered that the coenzyme F430 biosynthetic pathway involves 5 enzymes: CfbA, CfbB, CfbC, CfbD, and CfbE (*101*). The activities of these enzymes were confirmed *in vitro*. We have discussed the potential industrial applications related to MCR research above. Methanotrophic archaea are not suitable industrial hosts because of their long doubling time and growth requirements. There has been great interest in engineering a non-methanogenic host to express *holo* MCR for biological methane conversion. The *in vivo* biosynthesis of coenzyme F430 in non-methanogenic hosts is a significant aim of this project.

However, a bacterial cell is a self-organizing system. While we can artificially adjust every enzyme, reagent, cofactor, and supplementary chemical in *in vitro* studies, it is often difficult to accurately regulate these features *in vivo*. There are three main obstacles to the *in vivo* biosynthesis of coenzyme F430. First, substrates and enzymes must accumulate to sufficient extents for the biosynthetic pathway to move forward. *In vitro,* we are able to adjust the concentrations of each reactant so as to drive the reaction to completion. In contrast, cells

regulate chemical concentrations metabolically, making control of chemical concentrations difficult. Enzymes and substrates are synthesized and degraded, until metabolic balance is reached. Moreover, other co-related pathways may interact with the main pathway being pursued. Second, CfbCD, the enzyme complex that catalyzes the third step of the *cfb* pathway, requires reducing equivalents supplied by an external reductase. Sodium dithionite was used as the reductant for the CfbCD complex during *in vitro* assays, but it is incompatible with *in vivo* experiments. Third, the CfbCD complex must be assembled correctly to ensure its activity. Some model hosts like *E. coli* may lack the enzymatic system for nitrogenase-like iron-sulfur cluster construction, namely the *nif* or the *isc* systems. The consequence is an inability to express active CfbCD in these hosts without the inclusion of these systems.

Before conducting the *in vivo* coenzyme F430 biosynthesis experiments, we ensured that the *E. coli* BL21(DE3) strain can biosynthesize sirohydrochlorin, the precursor of F430. Only SirA and SirC, uroporphyrinogen III methylase and precorrin 2 dehydrogenase, are needed since *E. coli* can biosynthesize uroporphyrinogen III. A pACYCDuet vector containing the *sirC* and *sirA* genes was transformed into the cell. Cell culture experiments and cell free extract assays demonstrate that the *E. coli* BL21(DE3) strain containing the exogenous genes can biosynthesize sirohydrochlorin. Moreover, sirohydrochlorin accumulates abundantly in the cell, showing that it is not cytotoxic to the strain.

We utilized the Duet vector system to coexpress all *cfb* genes together with *sirA* and *sirC* after the capability of sirohydrochlorin biosynthesis was verified. The Duet vectors

weresequentially transformed into *E. coli* BL21(DE3) cells. Cell free extract analysis indicated that the coenzyme F430 biosynthesis pathway stopped after the second step, as the experimental data showed the production of Ni-sirohydrochlorin *a*,*c*-diamide without observable amounts of 15,17^3-seco-F430-17^3-acid or coenzyme F430. Thus, we constructed a six gene superplasmid containing all *cfb* genes and the *nixA* gene. Similar results were observed from the cell free extract of *E. coli* BL21(DE3) cells containing the *cfb* superplasmid and pACYCDuet:*sirC-sirA*.

The experimental results above are consistent with the obstacles to the *in vivo* biosynthesis of coenzyme F430 discussed above. The cells containing all *cfb* genes are unable to biosynthesize coenzyme F430 because the pathway is stopped after the second step. This phenomenon implies that the CfbCD complex is inactive.

We examined the ability of ferredoxin (Fd) and ferredoxin:NADP$^+$ reductase (Fnr) from spinach to support CfbCD catalysis. Preliminary exprerimental results indicate that spinach Fd/Fnr can reduce CfbC and activate the CfbCD complex with NADPH for 15,17^3-seco-F430-17^3-acid biosynthesis. The HPLC and EPR evidence inspired us to introduce the Fd and Fnr genes exogenously for *in vivo* coenzyme F430 biosynthesis. If *E. coli* is able to express the CfbCD complex in an active form without the *isc* genes, Fd and Fnr can use NADPH to reduce the complex. As a result, 15,17^3-seco-F430-17^3-acid (and subsequently coenzyme F430) would then be synthesized. Unfortunately, we were not able to detect any signals that indicated the biosynthesis of 15,17^3-seco-F430-17^3-acid or coenzyme F430. However, a better yield of Ni-sirohydrochlorin *a*,*c*-diamide was observed compared to the *cfb* cell line according to the

obtained HPLC traces and UV-vis spectra. However, these experiments need to be repeated to verify this claim.

Another possible explanation for the inactivity of the CfbCD reaction is that *E. coli* BL21(DE3) may fail to incorporate the necessary iron-sulfur clusters. We were informed by our experience with the heterologous expression of the CfbCD complex. A pDB1282 vector comprising the whole *isc* cluster from *A. vinelandii* greatly increased the amount of correctly assembled CfbCD complex. If iron-sulfur cluster assembly is the restrictive condition for CfbCD complex activity, a possible solution is to transform the pDB1282:*isc* vector into the *cfb* cell line. We constructed an *isc-cfb* cell line to verify this hypothesis. However, it seems that a lower amount of Ni-sirohydrochlorin *a*,*c*-diamide was obtained compare to the *Fd-cfb* cell line.

Finally, we constructed the *Fd-isc-cfb* cell line that contains all *cfb* genes and accessory genes (*sirA*, *sirC*, *Fd*, *Fnr*, and the *isc* cluster). Compare with the *isc-cfb* cell line, a better yield of Ni-sirohydrochlorin *a*,*c*-diamide was observed. Of all the different supplementary chemical conditions, supply 2 provided the best result. Nevertheless, no 15,17^3-seco-F430-17^3-acid or coenzyme F430 was detected. If true, the better yield of Ni-sirohydrochlorin *a*,*c*-diamide is likely indicative of the presence of the CfbCD complex (due to its ability to bind to the product and alleviate inhibition of CfbB). Although the *isc* gene cluster is necessary and sufficient for the formation of *holo* CfbCD, the production and activity of the encoded enzymes competes for resources and energy with tetrapyrrole biosynthesis. This interference from the *isc* gene cluster might be the main reason for the failure to biosynthesize coenzyme F430.

Chapter 4

Heterologous Coexpression of Putative MCR Post-translational Modification Genes and *in vivo* MCR Maturation

4.1 Background

The last step of methanogenesis is catalyzed by methyl coenzyme M reductase (MCR), which results in the synthesis of the coenzyme B-coenzyme M heterodisulfide (CoB-S-S-CoM) and methane. Methanogenesis is the energy-yielding metabolic pathway for some archaea that produces over 1 billion tons of methane per year. This natural process, which has been exacerbated through the anthropogenic raising of ruminant livestock, is a major contributor to global warming and there is a need for the design of novel methanogenesis inhibitors. This motivation pushes us to study MCR, a 270 kDa heterohexamer (Figure 44).

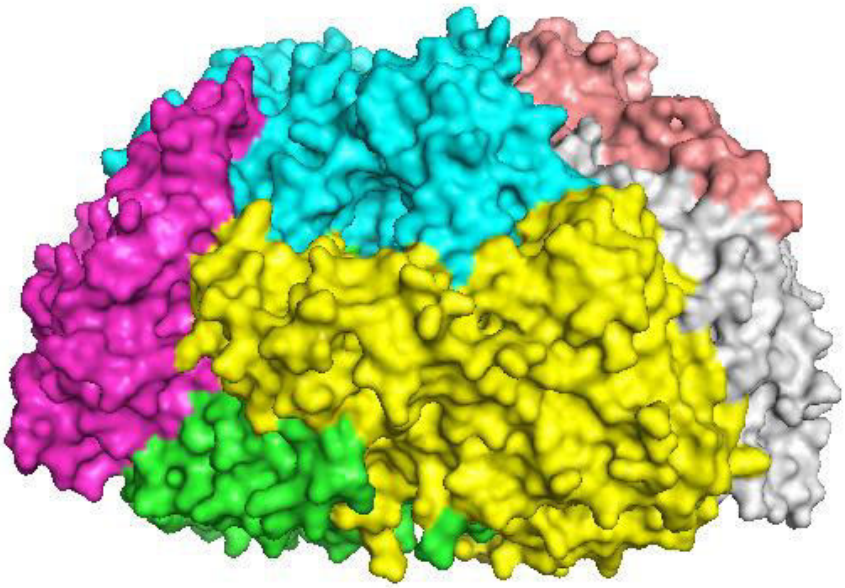


Figure 44: Structure of MCR (1MRO), a large 270 kDa complex with two active sites. Each of the subunits of the $\alpha_2\beta_2\gamma_2$ heterohexamer are shown using a different color.

There are two main aspects in our investigation of MCR. One of them is focused on coenzyme F430, a highly reduced nickel chelated tetrapyrrole. Coenzyme F430 is biosynthesized from sirohydrochlorin by 5 enzymes through 4 unique steps, which has been described in Chapters 2 and 3. Another is centered on the five post-translational modifications (PTMs) located in the α subunit of MCR. They are 2-(*S*)-methylglutamine, 5-(*S*)-methylarginine, 3-methylhistidine, *S*-methylcysteine, and thioglycine residues (Figure 45). Not all methanogens contain all five PTMs in the α subunit; some (like the *Methanosarcina* genus) only have four (*103*). The function of these PTMs are unknown and the detailed genetics and enzymology of their incorporation are unclear.

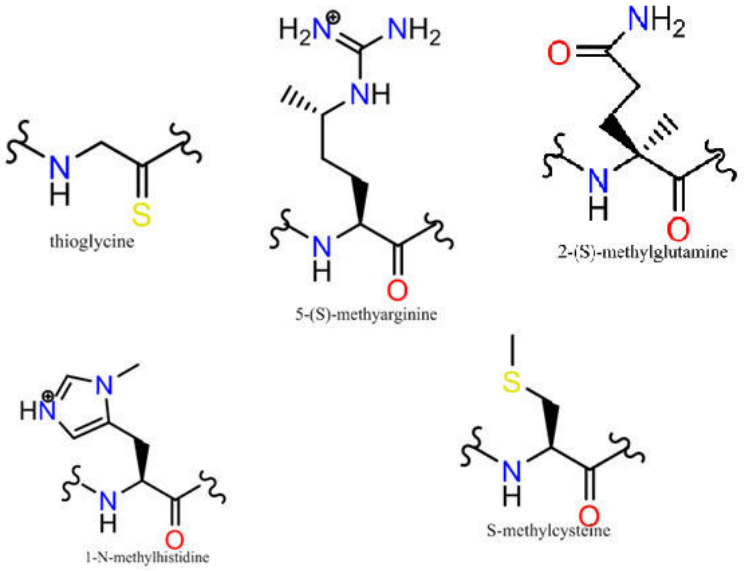


Figure 45: The five post-translational modifications (PTMs) located in the α subunit of MCR: 2-(*S*)-methylglutamine, 5-(*S*)-methylarginine, 3-methylhistidine, *S*-methylcysteine, and thioglycine residues.

Our research uses *Methanosarcina acetivorans* C2A as the model organism, which contains MCR with only four PTMs (the 2-(*S*)-methylglutamine modification is absent). The lack of the 2-(*S*)-methylglutamine modification is compensated by an alanine-to-threonine substitution, which orders a water molecule in the position that would have been occupied by the 2-methyl group (*55*).

The chemical requirements for the incorporation of the PTMs can be deduced from the structure of the modifications. We classified these five PTMs in three different group. Group one contains 3-methylhistidine and *S*-methylcysteine; group two includes 2-(*S*)-methylglutamine and 5-(*S*)-methylarginine; group three is thioglycine. The pK_a values of the methylation sites for

group one residues are less than ten, which indicates that these protons can be easily removed with a catalytic base. In other word, these methylations could be carried out by the alkylation of an activated nucleophile. Thus, a canonical *S*-adenosyl-L-methionine (SAM)-dependent methylase is sufficient for this (methyl cation transfer) reaction. In contrast, the C-H bonds in the group two residues are extremely stable (the estimated pK_a values are very high, ~30 according to the MarvinSketch software), thus making a polar mechanism unlikely. A radical mechanism catalyzed by a member of the radical SAM superfamily is required for such difficult alkylation reactions. We therefore propose that radical SAM methylases catalyze group two PTMs (*56*).

We noticed that there is a hypothetical SAM-dependent protein methyltransferase located downstream of the *mcr* gene cluster in the genome of *M. acetivorans* (Figure 46). It is homologous to ribosomal protein L11 methyltransferase (protein methylation gene A, *prmA*), which catalyzes the methylation of the N-terminal α-amino group and ε-amino groups of two lysine residues. Thus, we proposed that this homolog of *prmA* encodes an enzyme catalyzing the biosynthesis of one or more of the group one PTMs. Moreover, we also found a radical SAM superfamily member upstream of the *mcr* gene cluster (Figure 46). This gene is annotated as methanogenesis marker 10 (*mm10*). Sequence analysis suggests that Mm10 contains an auxiliary iron-sulfur cluster in addition to the canonical radical SAM cluster. In radical SAM enzymes, SAM binds to the latter iron-sulfur cluster and a single electron is transferred to it to create a 5'-deoxyadenosyl radical that is used for C-H bond activation. Mm10 also contains a domain of unknown function, DUF512, which might bind an additional cofactor needed for methyl radical

donation. Thus, we hypothesize that *mm10* may encode a radical SAM methyltransferase, which can catalyze the biosynthesis of the group two 5-(*S*)-methylarginine PTM (*56*).

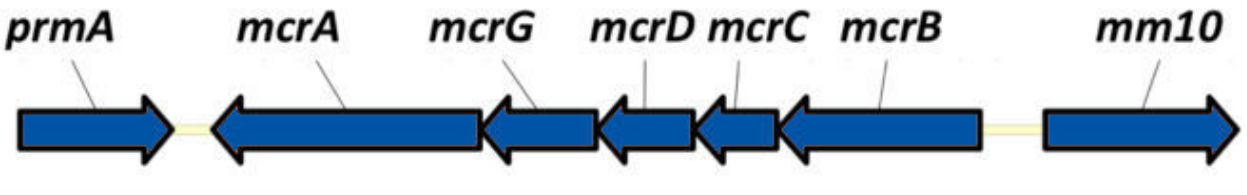


Figure 46: The *prmA* and *mm10* genes are located near the *mcr* cluster.

Group three of the PTMs only contains one member, thioglycine. Unlike the group one and two residues, it is not a methylated amino acid. A further comparative genomics study identified three viable candidate genes that may be involved in the formation of the thioglycine PTM. They are methanogenesis marker 1 (*mm1*), a homolog of *tfuA*, and a homolog of *thiI*. The gene *mm1* is hypothesized to encode a YcaO superfamily member that associates with TfuA. It has been found that a homolog of *mm1* from the *tva* gene cluster is responsible for thioamide formation during thioviridamide biosynthesis (*104*)(*65*). Moreover, a *tfuA* homolog has also been shown to be essential for thioviridamide biosynthesis (*65*). There is a *tfuA* homolog in methanogens that clusters with *mm1*. Other than *mm1* and *tfuA*, we also identified a *thiI* homolog in methanogens (*105*). ThiI serves as a sulfur donor for thiamine biosynthesis and may play a role in delivering sulfur for thioglycine formation.

In addition to the PTMs in the α subunit, we also noticed that there is a *cis*-proline residue in the MCR β subunit. Methanogenesis marker 3 (*mm3*) encodes a homolog of a *cis*-prolyl isomerase, and thus may introduce the *cis*-proline residue that is present in the MCR β subunit.

In order to investigate the PTM of MCR, we chose to use Duet vectors for a coexpression study in which the *mcrA* gene was coexpressed with various combinations of putative PTM genes and the purified McrA was examined for the presence of the relevant modifications. Duet vectors are T7 promoter expression vectors, which carry compatible replicons and antibiotic resistance markers. We also utilized a 'superplasmid' strategy for the PTM study because of the large number of genes involve in these biosynthetic processes. All *mcr* and putative PTM genes were co-transformed into *E. coli* BL21(DE3) as exogenous genes for the *in vivo* study.

We also used the pETSUMO vector to see if this system could improve the expression/solubility of McrA. pETSUMO is a T7 promoter expression vector that carries a kanamycin resistance marker and a replicon that is compatible with certain Duet vectors. Proteins can be expressed from this vector with a small ubiquitin-like modifier (SUMO)-tag and a N-terminal His_6-tag for protein purification. SUMO proteins are a family of small proteins, which can be reversibly and covalently linked to other proteins in the cell to modify their functions. This SUMOylation is involved in the regulation of different cellular process (*108*). Fusing a SUMO-tag with a recombinant protein was reported to improve its solubility and

expression level (*109*). All this information motivated us to use the pETSUMO vector for McrA expression.

4.2 Methods

4.2.1 *Construction of plasmids containing the mcr genes*

The genomic DNA of *Methanosarcina acetivorans* C2A was purchased from DSMZ in Germany (DSM2834). The *mcrABCDG* genes were amplified from the purchased genomic DNA by polymerase chain reaction (PCR). Primers were synthesized by Sigma-Aldrich (Table 3). Phusion High-Fidelity DNA Polymerase (New England Biolabs) was utilized for all PCR reactions. We also synthesized a codon optimized version of *mcrA* (*optmcrA*) from Invitrogen by Thermo Fisher Scientific.

Gene	Primer
mcrA	Forward: 5'-GATACTCATATGATGGCAGCAGACATTTTC-3' Reverse: 5'-ATGTGCTCGAGCTAGTGATGGTGATGATGGTGTTTTGCCGGGATGACGAG-3'
mcrB	Forward: 5'-CGCGGCATATGTCTGACACAGTAGACATCTACGACG-3' Reverse: 5'-TTATAGGCCGGCCTTAGAGCGCTCCTG-3'
mcrG	Forward: 5'-TTACTCCATGGCATACGAAGCACAGTATTATCC-3' Reverse: 5'-TTATTGGATCCTCATTTCGGCTGGAATCC-3'
mcrC	Forward: 5'-AATCTCATATGTCAGACTCTGCTTCAAACACG-3' Reverse: 5'-GCTCTCTCGAGTCACTCATCTTTATCAGTGTC-3'
mcrD	Forward: 5'-AGAATTCATGATGATCGACCGGGAAACAC-3' Reverse: 5'-CACTTGGATCCTCATGCAACTCCTTTATGATC-3'
mm1	Forward: 5′-ATATACCATGGCCGAGATAAAAATTG-3′ Reverse: 5′-AGATAAAGCTTCATCTTCTTTTCCA-3′
mm10	Forward: 5'-GCTCGCATATGGAAGTAGTTGTCGACG-3' Reverse: 5'-ATATAACTCGAGTTACTCTAGAGGCAATCCCAG-3'

tfuA	Forward: 5′-ACAGACCATGGAAAAGAAGATGAAAG-3′ Reverse: 5′- ATATAGGATCCTCAGGCCTCTTCAATAAG-3′
thiI	Forward: 5′- ATATAATTAATGACAGGCAATTCC-3′ Reverse: 5′- ATATAGGTACCTCAGAGCTTGAGGATTTTTAC-3′
prmA	Forward: 5′- GGCGCCATATGGAAATAAGATGTAGGTG-3′ Reverse: 5′- AGCTAGCTCGAGTCAAATCACAACAAC-3′
sumo mcrA	Primer 1: 5′- AGATTGGTGGTATGGCAGCAGACATTTTC-3′ Primer 2: 5′- AAGCTTGTCTTTATTTTGCCGGGATG-3′ Primer 3: 5′- GCTGCCATACCACCAATCTGTTCTCTG-3′ Primer 4: 5′- CCGGCAAAATAAAGACAAGCTTAGGTATTTATTCG-3′
sumo opt mcrA	Primer 1: 5′- GATTGGTGGTATGGCAGCAGATATCTTC-3′ Primer 2: 5′- CTAAGCTTGTCTTTATTTTGCAGGAATAACCA-3′ Primer 3: 5′- TTATTCCTGCAAAATAAAGACAAGCTTAGGTATTTATTCG-3′ Primer 4: 5′- GATATCTGCTGCCATACCACCAATCTGTTCTCTG-3′
opt mcrA petD GA	Primer 1: 5′- ATGGCAGCAGACATTTTCG-3′ Primer 2: 5′- GCATGAACTCTCTGATGGCAC -3′ Primer 3: 5′- GTGCCATCAGAGAGTTCATGC-3′ Primer 4: 5′- CGAAAATGTCTGCTGCCATC -3′

Table 3: Primers used for the construction of plasmids for the MCR PTM study.

We choose the Duet vector system from Novagen EMD Millipore of Germany for our coexpression. The *mcrA* gene from *M. acetivorans* was ligated into the second multiple cloning site (MCS-II) of pETDuet-1. This *mcrA* has been fused with a C-terminal His-tag for purification. In addition to the Duet vectors, we also used the pETSUMO vector from Thermo Fisher Scientific for our expression. The pETSUMO vector contains a SUMO-tag and a N-terminal His_6-tag for protein purification. The SUMO-tag is directly followed by the gene that encodes the target protein, which allows wild-type protein to be obtained after cleavage with SUMO protease. The *mcrA* gene was ligated into pETSUMO by Gibson Assembly using specially

designed primers (Table 3). The Gibson Assembly kit was purchased from New England Biolabs. In addition to the *mcrA* gene cloned from *M. acetivorans*, we also ligated the synthesized *optmcrA* gene into pETSUMO by Gibson Assembly. Thus, we obtained pETDuet:*mcrA*, pETSUMO-*mcrA* and pETSUMO-*optmcrA* (Figure S2-S9).

The *mcrC* and *mcrD* genes were sequentially ligated into MCS-I and MCS-II of the pRSFDuet-1 vector, respectively. The *mcrB* and *mcrG* genes were sequentially ligated into MCS-I and MCS-II of the pETDuet-1 vector, respectively. Thus, we obtained pRSFDuet:*mcrD-mcrC* and pETDuet:*mcrG-mcrB* (Figure 47).

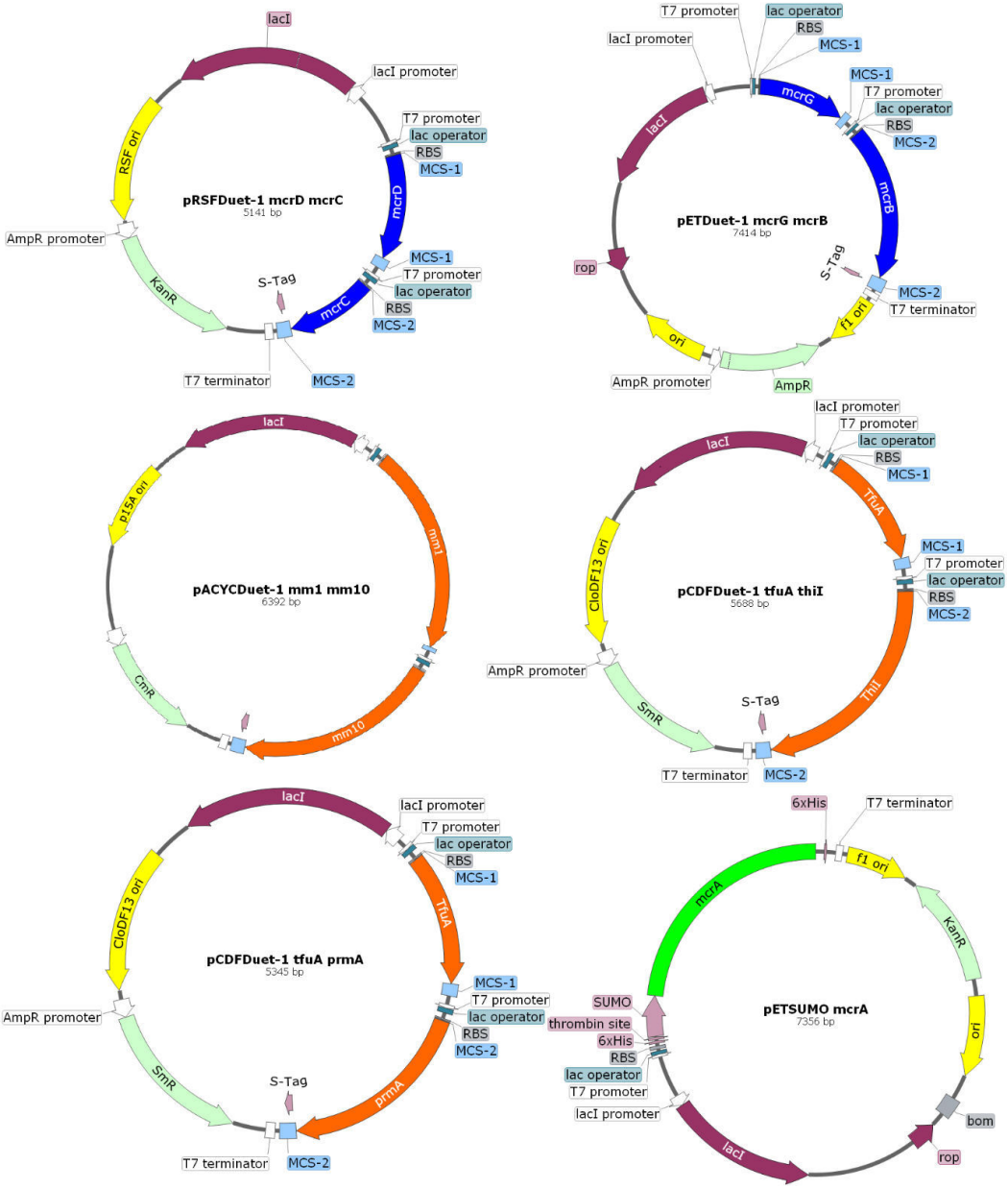


Figure 47: Plasmids constructed for MCR PTM studies (part I).

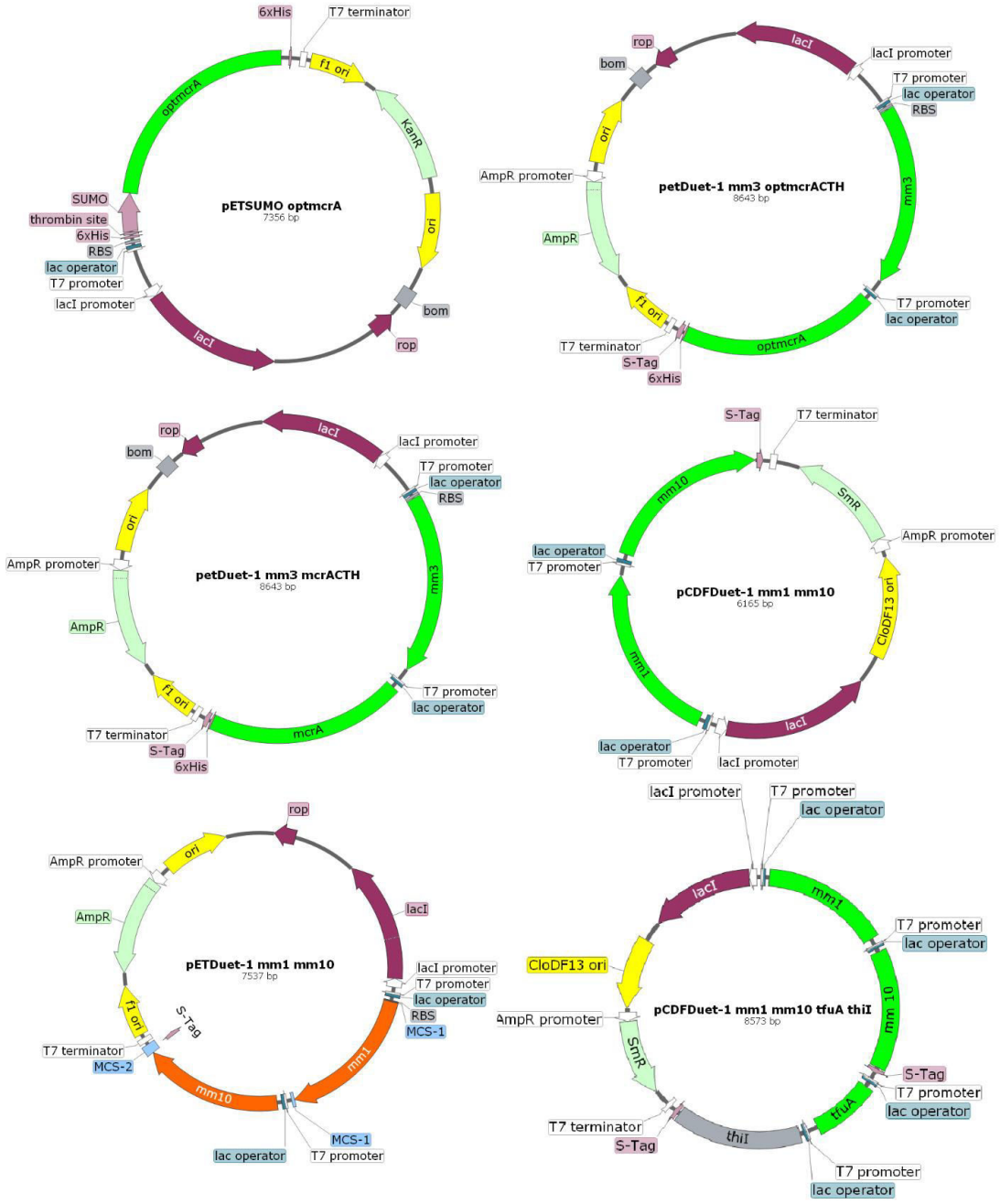


Figure 48: Plasmids constructed for MCR PTM studies (part II).

4.2.2 *Construction of plasmids containing the PTM genes*

The *mm1, mm3, mm10, prmA, tfuA,* and *thiI* genes were cloned from *M. acetivorans* C2A by PCR. The primers were synthesized by Sigma-Aldrich (Table 3). The *mm1* and *mm10* genes were ligated into MCS-I and MCS-II of pACYCDuet-1. There are two different versions of pCDFDuet-1 for the PTM analysis. We ligated the *tfuA* gene into the first MCS of pCDFDuet-1. Then, either the *prmA* or *thiI* gene was ligated into the second MCS (Figure 47).

After constructing the pETDuet:*mcrA* vector, we ligated the *mm3* gene into the first MCS of pETDuet:*mcrA*. Thus, we obtained pETDuet:*mm3-mcrA* (Figure 48). Then we used Gibson Assembly to replace the *mcrA* gene in pETDuet:*mm3-mcrA* with the codon optimized version to construct the pETDuet:*mm3-optmcrA* vector (Figure 48).

In addition to the pACYCDuet:*mm1-mm10* vector mentioned above, we also constructed pETDuet:*mm1-mm10* and pCDFDuet:*mm1-mm10* to allow for the coexpression of different combinations of genes *in vivo* (Figure 47 and Figure 48). We sequentially ligated the *mm1* and *mm10* genes from *M. acetivorans* into MCS-I and MCS-II of pETDuet-1. The pCDFDuet:*mm1-mm10* vector was obtained by double digesting pACYCDuet:*mm1-mm10* with the EcoNI and XhoI restriction enzymes (New England BioLabs) to obtain the *mm1-mm10* fragment. Then this fragment was ligated into double digested pCDFDuet-1.

4.2.3 *Superplasmid construction*

The Duet vector system allows eight different exogenous genes to be expressed in one cell. However, there are over eight genes involved in the assembly and PTM of MCR. We therefore developed a "superplasmid" strategy to overcome this obstacle. Starting from the previously constructed pACYCDuet:*mm1-mm10* and pCDFDuet:*tfuA-thiI* vectors, a SwaI restriction enzyme site was introduced approximately 100 bp upstream of the MCS-I of each vector. Then, pCDFDuet:*tfuA-thiI* was digested with SwaI to obtain a linearized fragment. The *mm1-mm10* fragment was cloned from pACYCDuet:*mm1-mm10* using a specially designed pair of primers. Finally, the *mm1-mm10* fragment and SwaI-digested pCDFDuet:*tfuA-thiI* were fused together by Gibson Assembly. Thus, we obtained the pCDFDuet:*mm1-mm10-tfuA-thiI* superplasmid (Figure 49). Moreover, this pCDFDuet:*mm1-mm10-tfuA-thiI* superplasmid conserves the upstream SwaI site, which allowed us to insert the *prmA* fragment similarly. This insertion will result in a pCDFDuet:*prmA-mm1-mm10-tfuA-thiI* superplasmid, which will be constructed in the future (Figure 49).

Since we have the pRSFDuet:*mcrD-mcrC* and pETDuet:*mcrG-mcrB* vectors, a similar mutagenesis and Gibson Assembly protocol was applied on them to obtain pRSFDuet:*mcrG-mcrB-mcrD-mcrC*. Moreover, this vector will allow for the construction of the pRSFDuet:*mm3-optmcrA-mcrG-mcrB-mcrD-mcrC* superplasmid in the future (Figure 50).

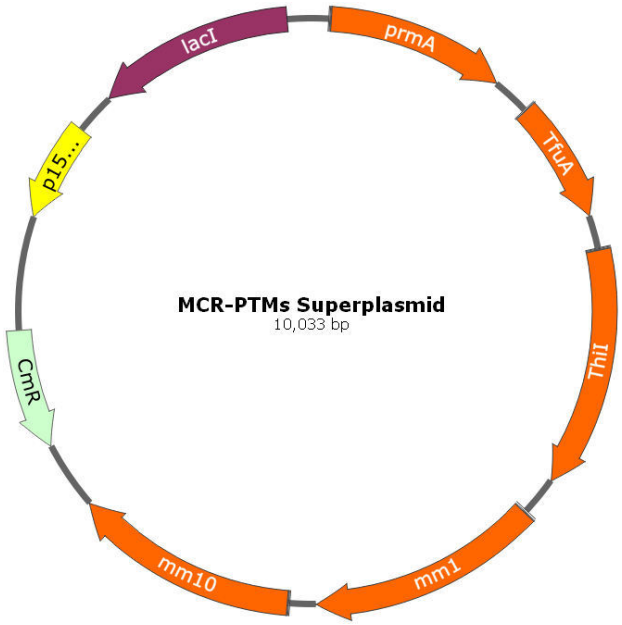


Figure 49: Superplasmid under construction containing all putative MCR PTM genes.

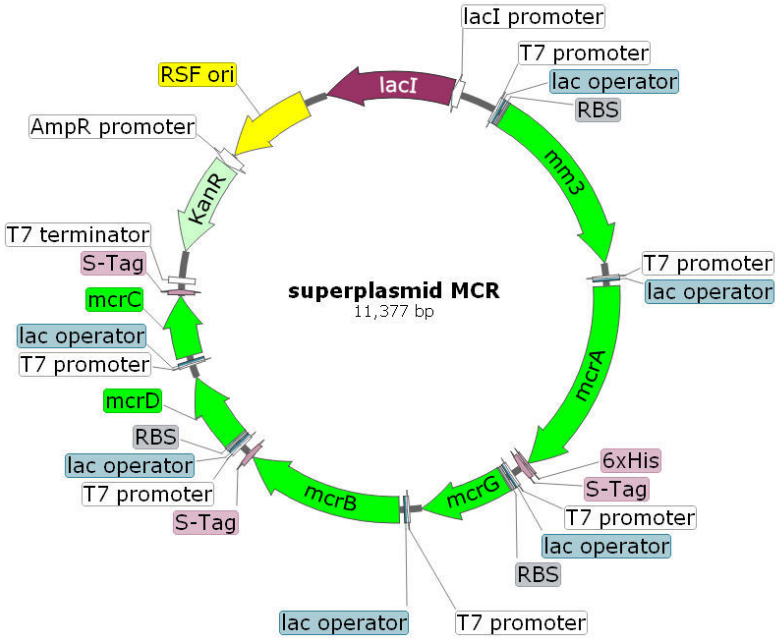


Figure 50: Superplasmid under construction containing all *mcr* genes.

4.2.4 *Coexpression and post-translational modifications*

All plasmids were sequence verified and *E. coli* BL21(DE3) from New England BioLabs served as the host. We sequentially transformed different vectors into a single cell. The pRSFDuet:*mcrG-mcrB-mcrD-mcrC*, pETDuet:*mcrA*, pACYCDuet:*mm1-mm10*, and pCDFDuet:*tfuA-prmA* vectors were included in the McrA-PrmA cell line. The McrA-ThiI cell line contains pRSFDuet:*mcrG-mcrB-mcrD-mcrC*, pETDuet:*mcrA*, pACYCDuet:*mm1-mm10*, and pCDFDuet:*tfuA-thiI* vectors. Both cell lines contain four different antibiotic resistance markers.

Terrific Broth (TB) was prepared for expression with several supplementary chemicals: ferrous sulfate (200 μM final concentration), cobalamin (2 μM final concentration), riboflavin (50 μM final concentration), methionine (50 μM final concentration), cysteine (200 μM final concentration), ampicillin (100 μg/mL final concentration), kanamycin (50 μg/mL final concentration), spectinomycin (25 μg/mL final concentration), and chloramphenicol (35 μg/mL final concentration). The cultures were incubated at 37 °C until the OD_{600} increased to 0.6. Then the cells were induced with β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.4 μM and incubated at 18 °C for 16 h.

The pellets were harvested by centrifugation at 15900 × g for 10 min. Three different types of buffer were prepared: lysis buffer, wash buffer, and elution buffer. Wash buffer contains 50 mM sodium phosphate, 300 mM NaCl, and 5 mM imidazole at pH 8.0. Lysis buffer was prepared by adding lysozyme (1 mg/mL final concentration) and Ameresco's Protease Inhibitor

Cocktail into wash buffer with 20% glycerol. Elution buffer contains 500 mM imidazole instead of 5 mM imidazole in wash buffer. The harvested cell pellets were resuspended in lysis buffer, then sonicated and centrifuge at 30000 × g for 1 h. The supernatant was applied on a Bio-Rad Econo-Pac column packed with Ni^{2+}-charged Profinity IMAC Resin to purify McrA. The column was then washed with five to ten column volumes of wash buffer. Finally, the column was eluted with five column volume of elution buffer. The buffer of the eluted target protein (the α subunit of MCR) was exchanged with 0.1 M Tris-HCl, pH 8.0 buffer containing 20% glycerol. The protein was then concentrated using a Microsep™ Advance Centrifugal Device (30 kDa) from Pall Corporation for further analysis.

We also constructed pETSUMO-*mcrA* and several cell lines have been constructed using this vector: SUMO-McrA (pETSUMO-*mcrA*), SUMO-McrABG (pETSUMO-*mcrA*, pETDuet:*mcrB-mcrG*), SUMO-McrABG-Mm (pETSUMO-*mcrA*, pETDuet:*mcrB-mcrG*, pACYCDuet:*mm1-mm10*), SUMO-McrABG-Mm-PrmA (pETSUMO-*mcrA*, pETDuet:*mcrB-mcrG*, pACYCDuet:*mm1-mm10*, pCDFDuet:*tfuA-prmA*), SUMO-McrA-PrmA (pETSUMO-*mcrA*, pCDFDuet:*tfuA-prmA*), SUMO-McrA-Mm (pETSUMO-*mcrA*, pACYCDuet:*mm1-mm10*), and SUMO-McrA-Mm-PrmA (pETSUMO-*mcrA*, pACYCDuet:*mm1-mm10*, pCDFDuet:*tfuA-prmA*). All these cell lines used *E. coli* BL21(DE3) as the host strain.

Several *E. coli* Rosetta cell lines have also been constructed: RSUMO-McrA (pETSUMO-*mcrA*), RSUMO-McrABG (pETSUMO-*mcrA*, pETDuet:*mcrB-mcrG*), RSUMO-McrABG-PrmA (pETSUMO-*mcrA*, pETDuet:*mcrB-mcrG*, pCDFDuet:*tfuA-prmA*), RSUMO-

McrA-Mm-ThiI (pETSUMO-*mcrA*, pCDFDuet:*mm1-mm10-tfuA-thiI*), RSUMO-McrA-Mm-PrmA (pETSUMO-*mcrA*, pCDFDuet:*mm1-mm10*, pETDuet:*tfuA-prmA*), and RSUMO-McrABG-Mm-ThiI (pETSUMO-*mcrA*, pETDuet:*mcrB-mcrG*, pCDFDuet:*mm1-mm10-tfuA-thiI*).

We also constructed several cell lines based on the codon optimized *mcrA* gene in pETSUMO using *E. coli* BL21(DE3): optSUMO-McrA (pETSUMO-*optmcrA*) and optSUMO-McrABG (pETSUMO-*optmcrA*, pETDuet:*mcrB-mcrG*). Also, there are several cell lines based on the codon optimized C-terminal His_6-tag *mcrA* gene under construction: optMcrA (pETDuet:*mm3-optmcrA*), optMcrABG (pETDuet:*mm3-optmcrA*, pRSFDuet:*mcrG-mcrB-mcrD-mcrC*), optMcrABG-PrmA (pETDuet:*mm3-optmcrA*, pRSFDuet:*mcrG-mcrB-mcrD-mcrC*, pCDFDuet:*tfuA-prmA*), and optMcrABG-Mm-ThiI (pETDuet:*mm3-optmcrA*, pRSFDuet:*mcrG-mcrB-mcrD-mcrC*, pCDFDuet:*mm1-mm10-tfuA-thiI*).

All of the cell lines constructed above were expressed in LB and TB. Cells were cultured in 10 mL LB overnight with appropriate antibiotics, then transferred to 1L LB/TB with several supplementary chemicals: ferrous sulfate (200 μM final concentration), cobalamin (2 μM final concentration), riboflavin (50 μM final concentration), methionine (50 μM final concentration), and cysteine (200 μM final concentration). Antibiotics including ampicillin (100 μg/mL final concentration), kanamycin (50 μg/mL final concentration), spectinomycin (100 μg/mL final concentration), and/or chloramphenicol (35 μg/mL final concentration) were added according to the cell line. All cells were incubated in an incubator/shaker at 37 °C and 200 rpm until the

OD_{600} increased to 0.6. Then the cells were induced with IPTG at a final concentration of 0.4 μM and incubated at 15 °C and 90 rpm for 24 h.

The cell pellets were harvested by centrifugation at 15900 × g for 10 min. The collected pellets were resuspended in lysis buffer, then sonicated and centrifuged at 7197 × g for 1 h. The supernatant was loaded on a Bio-Rad Econo-Pac column packed with Ni^{2+}-charged Profinity IMAC Resin to purify McrA. The column was washed with five column volumes of wash buffer, then eluted with five column volumes of elution buffer to obtain McrA.

The insoluble fraction from the centrifugation step after sonication was treated with urea buffer: 50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, and 8 M urea at pH 8.0 (1 mL/0.5 g). The mixture was diluted 10 times with wash buffer after a 3 h treatment in order to refold McrA from inclusion bodies. The diluted mixture was centrifuged at 46500 × g for 20 min, then was filtered by a 0.22 μm polystyrene filter (VWR Bottle Top Filtration Funnel Only). The filtered solution was concentrated with a 10 kDa semipermeable membrane until the volume of the solution was decreased 10-fold. The concentrated solution was treated according to the above purification protocol in order to extract McrA from inclusion bodies. The purified protein was then concentrated for further analysis.

The size and purity of the target protein, McrA, was examined by SDS-PAGE. The protein samples were analyzed using a 12% resolving gel because the size of the MCR α subunit is approximately 62 kDa. Details of the SDS-PAGE protocol is given in the Appendix. Western blot analysis was performed on the MCR α subunit to confirm its presence and exact location on

the gel. Details of the western blot protocol are given in the Appendix. Gel slices containing the target protein were cut from the gel under the guidance of the western blot. These gel slices were then sent to the Proteomics and Mass Spectrometry Facility at the University of Gerogia (UGA) for peptide MS analysis.

4.2.5 *MALDI-TOF and LC-MS*

A trypsin or pepsin digestion was performed on the proteins from the gel slices, and the resultant peptide fragments were analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) for peptide and post-translational modification identification. MALDI-TOF MS analysis was performed on a Bruker Autoflex MALDI spectrometer. The resultant peptide fragments were also analyzed by liquid chromatography-mass spectrometry (LC-MS) for further peptide and post-translational modification verification. LC-MS analysis was performed on an Orbitrap Elite system with an electrospray ion source and an interchangeable nanospray source. The data was submitted to the publicly available Matrix Science software package. All the MASCOT Peptide Mass Fingerprint server, MASCOT Sequence Query server and the University of Georgia Proteomics and Mass Spectrometry server were used for data analysis to identify the PTMs.

4.3 Results

4.3.1 *Expression and purification*

Minute amounts of soluble McrA is expressed by the McrA-PrmA cell line. Also, the purity of the eluted sample is low. There are several *E. coli* protein bands on the gel, including a band that of McrA. Western blot analysis examined the position of the MCR α subunit band, which enabled it to be excised from the gel. Similar results were also observed using the McrA-ThiI cell line. We also purified the MCR α subunit from the insoluble fraction of the cell free extract using a refolding strategy. Expression of MCR α subunit in both cell lines were hard to be repeated. SDS-PAGE Gel slices were cutted approximately according to the size of MCR α subunit and sended for MS analysis.

The SUMO-tag was fused with McrA to try to improve the expression level and solubility (*109*). The MCR α subunit containing N-terminal His- and SUMO-tags was expressed using either *E. coli* BL21(DE3) or Rosetta cells. After several rounds of optimization of the expression conditions, we were able to produce SUMO-McrA in the *E. coli* Rosetta cell line, although not in *E. coli* BL21(DE3). However, the expressed SUMO-McrA appears to be unstable and is truncated into smaller pieces (as confirmed by western blot analysis using antibodies against the N-terminal His-tag). Moreover, we noticed that the expression results have less truncated bands on Western-Blot when the MCR α subunit was coexpressed with the β and γ subunits (Figure S2Figure S9).

Analyzing the sequence of the *mcrA* gene, there are several rare codons conserved in the sequence (Figure 51). This may be an explanation for why the expression level of the MCR α subunit in *E. coli* BL21(DE3) is so low compared to *E. coli* Rosetta. We therefore codon

optimized the sequence of *mcrA* and had it directly synthesized. The new so called *optmcrA* gene was ligated into both the pETSUMO vector and the pETDuet vector with a C-terminal His-tag.

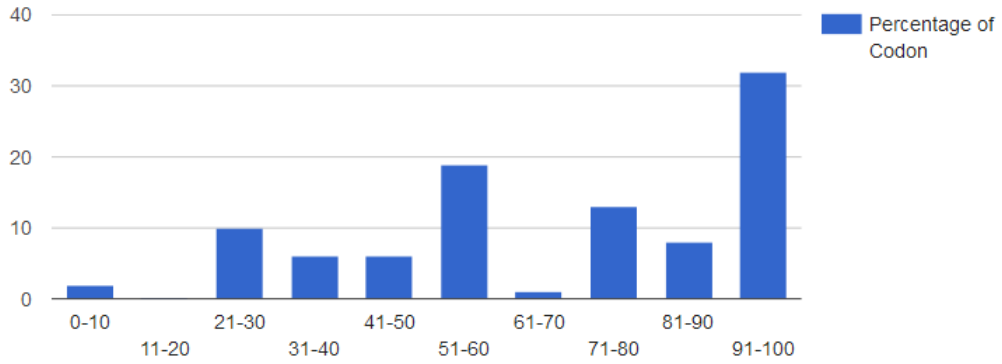


Figure 51: Codon usage frequency distribution of the *mcrA* gene with *E. coli* as the expression host.

4.3.2 *MALDI-TOF and LC-MS analysis*

We obtained the MCR α subunit from coexpression experiments with the putative post-translational modification genes in *E. coli* using SDS-PAGE. The appropriate protein bands were excised and digested with pepsin. Then the samples were analyzed using MALDI-TOF MS to obtain the fingerprint spectrum for each peptide.

We obtained gel slices that contained the MCR α subunit expressed from the McrA-ThiI cell line. A search of the MALDI-TOF MS fingerprint spectrum indicates that the gel slices from the insoluble fraction of the cell free extract contains the MCR α subunit (Figure 52 and

Figure S11). The pepsin-treated peptides were then analyzed using LC-MS to identify if any PTMs of interest were present (Figure 53). The McrA-ThiI cell line contains the *mm1*, *mm10*,

tfuA, and *thiI* genes. We expected to observe that G465 has been modified to thioglycine. The peptide sequence containing G465 after digestion with pepsin is FGFDLQDQCGATNVLS, and the predicted monoisotopic *m/z* is 857.9. The signal of this peptide with a modified glycine residue has been identified in the MASCOT search results (Figure 54 and Figure S12), which is consist with the incorporation of the expected thioglycine PTM.

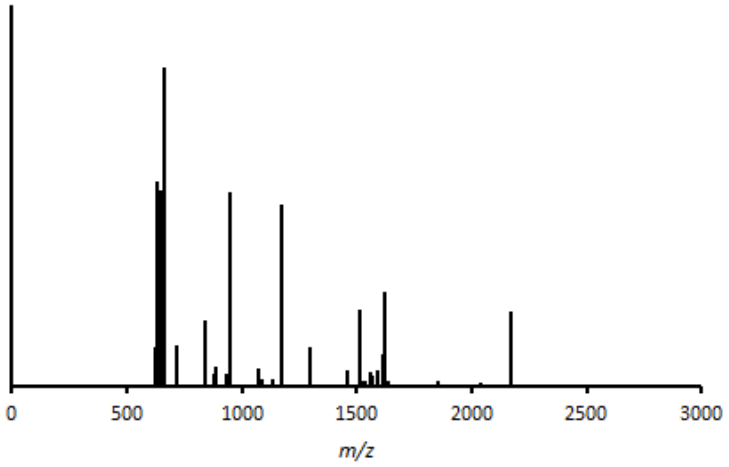


Figure 52: MALDI-TOF MS spectrum of trypsin-digested McrA produced by the McrA-ThiI cell line. This spectrum has been inputted into MASCOT for fingerprint analysis, which confirmed the presence of the MCR α subunit.

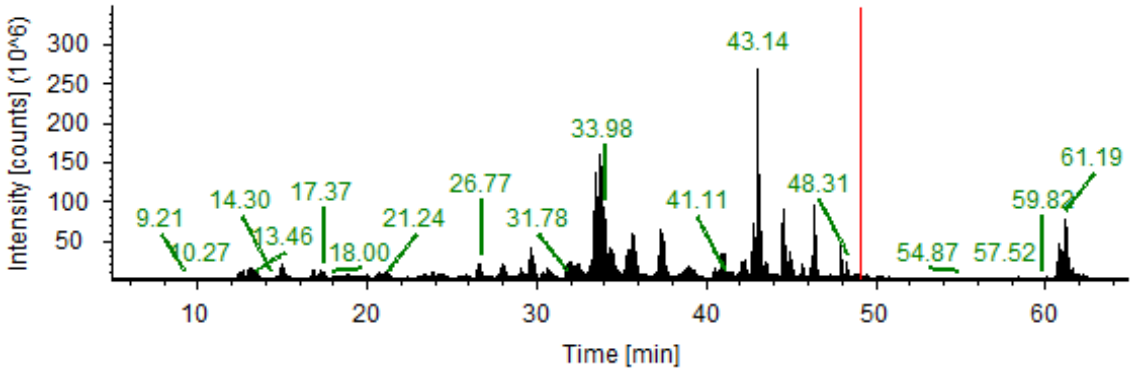


Figure 53: LC-MS results of pepsin-digested McrA produced by the McrA-ThiI cell line. The peak highlighted in red with a retention time of 49.1 min is consistent with the FGFDLQDQCGATNVLS peptide containing a thioglycine modification.

Query	Start - End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	Peptide
3138	378 - 399	773.4159	2317.2260	2317.2377	-5.03	4	14	0.049	1		L.GTDNKVKATLDVVKDIATESTL.Y
3139	378 - 399	773.4159	2317.2260	2317.2377	-5.03	4	44	0.002	1		L.GTDNKVKATLDVVKDIATESTL.Y
1613	388 - 399	645.8425	1289.6705	1289.6715	-0.74	2	12	0.61	1		L.DVVKDIATESTL.Y
1615	388 - 399	645.8425	1289.6705	1289.6715	-0.74	2	31	1.7	2		L.DVVKDIATESTL.Y
1617	388 - 399	645.8432	1289.6718	1289.6715	0.30	2	12	3.1	4		L.DVVKDIATESTL.Y
1618	388 - 399	645.8432	1289.6718	1289.6715	0.30	2	1	1	3		L.DVVKDIATESTL.Y
1619	388 - 399	645.8434	1289.6723	1289.6715	0.67	2	21	0.49	1		L.DVVKDIATESTL.Y
1622	388 - 399	645.8494	1289.6843	1289.6715	9.96	2	54	0.046	1		L.DVVKDIATESTL.Y
2234	412 - 425	765.3902	1528.7658	1528.7634	1.59	3	39	0.33	1		A.LEDHFGGSQRATVL.A
2235	412 - 425	765.3902	1528.7658	1528.7634	1.59	3	23	0.17	1		A.LEDHFGGSQRATVL.A
2236	412 - 425	510.5960	1528.7660	1528.7634	1.71	3	13	0.73	1		A.LEDHFGGSQRATVL.A
2237	412 - 425	510.5960	1528.7660	1528.7634	1.71	3	3	0.69	1		A.LEDHFGGSQRATVL.A
1113	413 - 422	552.2471	1102.4796	1102.4792	0.36	1	19	0.19	1		L.EDHFGGSQRA.T
1114	413 - 422	552.2471	1102.4796	1102.4792	0.36	1	20	0.34	1		L.EDHFGGSQRA.T
1986	413 - 425	708.8474	1415.6801	1415.6794	0.56	2	15	0.19	1		L.EDHFGGSQRATVL.A
1987	413 - 425	708.8474	1415.6801	1415.6794	0.56	2	19	0.94	1		L.EDHFGGSQRATVL.A
1989	413 - 425	472.9010	1415.6811	1415.6794	1.22	2	6	1.2	8		L.EDHFGGSQRATVL.A
2463	464 - 478	849.8903	1697.7661	1697.7719	-3.45	4	7	0.23	1		F.FGFDLQDQCGATNVL.S + Propionamide (C)
2492	464 - 478	857.8864	1713.7583	1713.7491	5.35	4	5	1.6	3		F.FGFDLQDQCGATNVL.S + Thioglycine Carboxy->Thiocarboxy (G); Propionamide (C)
2493	464 - 478	857.8864	1713.7583	1713.7491	5.35	4	24	0.1	1		F.FGFDLQDQCGATNVL.S + Thioglycine Carboxy->Thiocarboxy (G); Propionamide (C)
3103	479 - 498	762.0174	2283.0303	2283.0080	9.79	4	10	2.4	7	U	L.SYQGDEGLPDELRGPNYPNY.A
3105	479 - 498	1142.5226	2283.0306	2283.0080	9.91	4	41	2	5	U	L.SYQGDEGLPDELRGPNYPNY.A
762	491 - 498	490.7333	979.4521	979.4512	0.99	1	18	0.14	1		L.RGPNYPNY.A
1787	499 - 512	681.3166	1360.6187	1360.6194	-0.47	4	20	0.027	1		Y.AMNVGHQGGYAGIA.Q + Oxidation (M)
1788	499 - 512	681.3166	1360.6187	1360.6194	-0.47	4	59	0.0014	1		Y.AMNVGHQGGYAGIA.Q + Oxidation (M)
1585	509 - 522	427.8814	1280.6223	1280.6221	0.13	5	31	0.024	1	U	Y.AGIAQAAHSGRGDA.F
1586	509 - 522	427.8814	1280.6223	1280.6221	0.13	5	46	0.0017	1	U	Y.AGIAQAAHSGRGDA.F
946	512 - 522	520.7470	1039.4795	1039.4795	-0.038	3	32	0.56	1	U	I.AQAAHSGRGDA.F
947	512 - 522	520.7470	1039.4795	1039.4795	-0.038	3	33	0.27	1	U	I.AQAAHSGRGDA.F
738	513 - 522	485.2287	968.4428	968.4424	0.46	2	13	0.56	1	U	A.QAAHSGRGDA.F
740	513 - 522	485.2288	968.4430	968.4424	0.65	2	24	0.091	1	U	A.QAAHSGRGDA.F

Figure 54: Results from MASCOT analysis of the LC-MS data of McrA produced by the McrA-ThiI cell line. The yellow highlighted part shows the thioglycine-containing FGFDLQDQCGATNVLS peptide.

The SUMO-tagged MCR α subunit was expressed in the RSUMO-McrA, RSUMO-McrABG, RSUMO-McrABG-PrmA, and RSUMO-McrABG-Mm-ThiI cell lines. We chose the largest fragment on the gel for MALDI-TOF MS fingerprint analysis. Unfortunately, we did not detect MCR α subunit expression in the RSUMO-McrABG cell line. Although the MS analysis with the RSUMO-McrABG cell line failed, we did confirm that the MCR α subunit was expressed by the RSUMO-McrABG-PrmA and RSUMO-McrABG-Mm-ThiI cell lines (Figure S13 and Figure S14). However, these experimental results were hard to be repeat consistently.

4.4 Discussion

Heterologous expression of the MCR α subunit in *E. coli* is the main difficulty in studying the PTMs. Either low expression levels or high insolubility has been observed during the experiments. Several approaches have been used to try and overcome this obstacle, including codon optimization, fusing with a SUMO-tag, and changing expression conditions. Unfortunately, none of these have been able to prominently improve expression or solubility.

MALDI-TOF MS analysis still allowed for the characterization of PTMs of interests, even though the expression level was low. The production of the MCR α subunit by the McrA-PrmA cell line was confirmed. This sample was then treated with trypsin and the proteolyzed peptide sequences were examined by MALDI-TOF MS to confirm the presence of the PTMs of interest, which are 1-*N*-methylhistidine (H271) and *S*-methylcysteine (C472). According to this analysis, the relevant peptides are HAALVSMGEMLPARR and

LGFFGFDLQDQCGATNVLSYQGDEGLPDELR. The unmodified peptides have a monoisotopic *m*/*z* of 1638.8 and 3404.6, respectively. The predicted (Me-H)AALVSMGEMLPARR peptide has a *m*/*z* of 1652.8, and the LGFFGFDLQDQ(Me-C)GATNVLSYQGDEGLPDELR *m*/*z* is 3418.6, which were consistent with observed peaks in the MS spectrum (Figure 55-Figure 56). Thus, peptides consistent with both the 1-*N*-methylhistidine and *S*-methylcysteine PTMs have been identified when *prmA* is coexpressed with *mcrA*. These experimental results are consistent with the hypothesis that PrmA is responsible for the 1-*N*-methylhistidine and *S*-methylcysteine PTMs. Thus, these preliminary experiments suggest that MA4545 encodes a methylase that catalyzes 1-*N*-methylation of histidine and *S*-methylation of cysteine. However, additional experiments are still needed to definitively confirm this hypothesis.

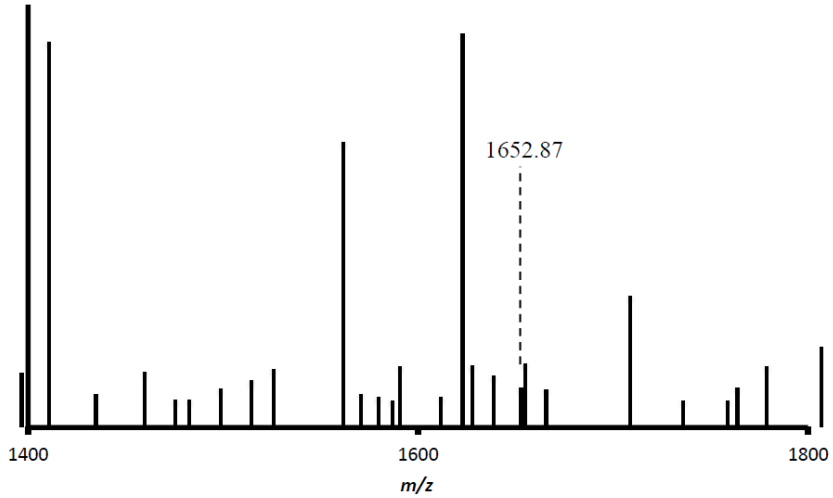


Figure 55: MALDI-TOF MS spectrum of trypsin-digested MCR α-subunit containing 1-*N*-methylhistidine. The 1652.87 *m/z* peak is consistent with a HAALVSMGEMLPARR peptide containing the 1-*N*-methylhistidine modification.

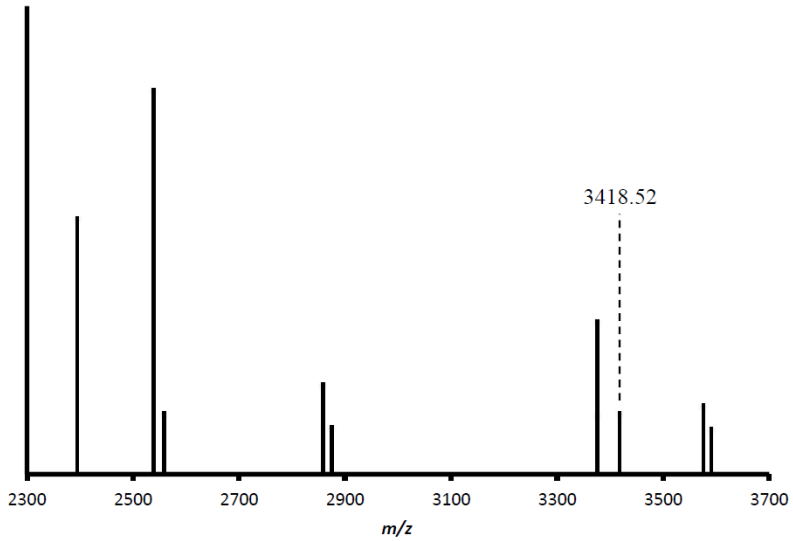


Figure 56: MALDI-TOF MS spectrum of trypsin-digested MCR α-subunit containing *S*-methylcysteine. The 3418.5260 *m/z* peak is consistent with a LGFFGFDLQDQCGATNVLSYQGDEGLPDELR peptide containing the *S*-methylcysteine modification.

The thioglycine PTM has also been detected when the MCR α-subunit is coexpressed with the *mm1*, *tfuA*, and *thiI* genes. Even though the MCR α-subunit was purified from inclusion bodies, the MS results indicate the presence of thioglycine. It is not expected that the protein aggregates in inclusion bodies would be substrates for Mm1 and its associated enzymes to install this PTM. Moreover, the thioglycine PTM can be hydrolyzed in water-based buffer systems. Considering all of these aspects, it is possible that inclusion bodies may help protect the thioglycine PTM from hydrolysis and the enzymatic reactions to install this PTM may be performed faster than inclusion body formation.

Recently, Gunhild Layer and coworkers reported that Mm10, which is encoded by the gene MA4551, is a radical SAM methyltransferase catalyzing the formation of the 5-(*S*)-methylarginine PTM (*110*). This result was confirmed through gene deletion experiments in *M. acetivorans* and is consistent with the results our collaborator William (Barny) Whitman (University of Georgia) and coworkers. A mutant of *Methanococcus maripaludis* that lacked the *mm10* gene was constructed by gene knockout techniques. The MCR expressed by this mutant was compared with the wild-type MCR. Both studies above implicate that Mm10 is required for 5-(*S*)-methylarginine biosynthesis. Initially, we were expecting to detect the 5-(*S*)-methylarginine modification by coexpressing *mm10* and *mcrA* as exogenous genes in *E. coli*. However, Mm10 is a radical SAM methyltransferase, and iron-sulfur cluster(s) are significant for both its structure and activity. *E. coli* may not be an ideal host to express this iron-sulfur protein. Therefore, a supplementary set of genes, such as the *isc* system used for the expression of the

CfbCD complex, might be required for iron-sulfur cluster manufacturing. Moreover, the absence of the appropriate methyl donor for this enzymatic reaction may be another cause for concern. Methylcobalamin is a potential candidate for the methyl donor for this reaction in methanogens. Unfortunately, *E. coli* is incapable of the *de novo* biosynthesis of cobalamin (although it does have the genes for the B_{12} salvage pathway). Therefore, one of the keys to ensure that Mm10 is functioning properly in *E. coli* is to provide the correct methyl donor for the reaction. We have constructed a 'superplasmid' containing the cobalamin transporter genes that can be coexpressed with *mcrA*/*mm10* and the *isc* operon to enhance B_{12} uptake in *E. coli* for use in future experiments.

In summary, based on our discoveries and recently reported results from our colleagues, the genes required for the biosynthesis of all of the PTMs that are present in the MCR from *M. acetivorans* have been identified. Combining this knowledge with that of coenzyme F430 biosynthesis, *holo* MCR expression and assembly becomes the next long-term goal, with the solubility and expression level problems being one of the major remaining obstacles.

Chapter 5

Discussion and future work

Methane, as the simplest organic compound, is the major component of natural gas. It is also one of the main culprits for the greenhouse gas effect. Because of its global warming and fuel potential, the study of the bioproduction and bioconversion of methane is of current interest. Methyl-coenzyme M reductase (MCR) is a significant research focus in this field because it catalyzes the last step of methanogenesis, the energy-yielding metabolism found in methanogenic archaea. The C-H bond strength of methane is very high, which leads to its chemical stability and difficulty to be activated for oxidation (*111*)(*25*). MCR and its homolog in anaerobic methanotrophic archaea (ANME) provided a biological solution for methane activation.

The activation of the C-H bond of methane using reverse methanogenesis provides an opportunity for the conversion of methane to several important C1 compounds like methanol and methylamine. As alluded to above, ANME couple the anaerobic oxidation of methane (AOM) using a MCR homolog to the reduction of a thermodynamically favorable electron acceptor, such as sulfate or nitrate (*112*). The resulting activated methyl can then be involve in the production of acetyl-coenzyme A (acetyl-CoA). Acetyl-CoA is a significant precursor for several biosynthetic pathways that enable the production of ethanol, fatty alcohols, alkanes, and acyl esters (*111*). Thus, MCR and related enzymes in the metabolism of C1 compounds are key links for multiple bioconversion chemistries.

The study of *holo* MCR production can be classified into three different aspects: coenzyme F430 biosynthesis, MCR post-translational modification (PTM), and heterologous expression/assembly. We elucidated the biosynthesis of coenzyme F430 through *in vitro* approaches in Chapter 2. This unique biosynthesis pathway is comprised of four steps and involves five different enzymes. Sirohydrochlorin is the direct precursor of coenzyme F430 biosynthesis. The first step of this pathway is catalyzed by CfbA, a sirohydrochlorin nickelochelatase. Ni-sirohydrochlorin serves as the substrate in the second step, which is an diamidation reaction. This step is catalyzed by CfbB, a Ni-sirohydrochlorin *a*,*c*-diamide synthetase. Then the isobacteriochlorin ring system of sirohydrochlorin is reduced by six electrons and the γ-lactam E ring is formed by the CfbCD complex, a homolog of nitrogenase, to generate 15,17^3-seco-F430-17^3-acid. The final step of the pathway is accomplished through a six-member ring cyclization reaction, which is catalyzed by the Mur ligase homolog, CfbE, the coenzyme F430 synthetase. McrD improves the conversion rate of the last step reaction. We hypothesized it is a chaperone that helps deliver coenzyme F430 to the MCR complex.

We also discussed our progress in the heterologous *in vivo* biosynthesis of coenzyme F430 in Chapter 3. A significant conclusion is that coenzyme F430 cannot be biosynthesized *in vivo* with only the *cfb* genes when *E. coli* is used as the expression host. Large quantities of Ni-sirohydrochlorin *a*,*c*-diamide was detected, which indicates that the pathway stops after the second step. In other words, the CfbCD reaction is not proceeding *in vivo*. In our initial *in vitro* assays, the CfbCD complex is supplied with sodium dithionite as the reductant, which is unavailable *in vivo*. A ferredoxin (Fd) and ferredoxin:NADP$^+$ reductase (Fnr) pair from spinach

was purchased for preliminary tests as a physiological reductant, and it was shown that it is capable of supporting CfbCD catalysis. Moreover, an *isc* gene cluster was also introduced as exogenous genes to support iron-sulfur cluster manufacturing. With all these supplemental genes, we built five different cell lines to test coenzyme F430 biosynthesis capability of *E. coli*.

Unfortunately, we were not able to biosynthesize coenzyme F430 *in vivo* even in the presence of all supplementary genes. The introduction of Fd and Fnr improves the yields of Ni-sirohydrochlorin *a*,*c*-diamide. In contrast, the presence of the *isc* genes inhibits its biosynthesis. We hypothesize that the expression of the *isc* enzyme system uses large amount of energy and resources, which may diminish the production of the Cfb enzymes and the energetically demanding formation of Ni-sirohydrochlorin *a*,*c*-diamide. We will express these enzymes sequentially in future experiments. IPTG will be used for inducing the expression of SirA, SirC, the Cfb enzymes, and Fd/Fnr. We will then harvest cells containing large quantities of Ni-sirohydrochlorin *a*,*c*-diamide. Then these cells will be resuspended in fresh media. The expression of the Isc enzymes and the CfbCD complex will then be induced to try to overcome the "second step barrier" of the *cfb* pathway in *E. coli*.

We discussed our study of the PTMs of MCR in Chapter 4. MCR has four or five different PTMs in its α subunit depending on the species of methanogen. They are 2-(*S*)-methylglutamine, 5-(*S*)-methylarginine, 3-methylhistidine, *S*-methylcysteine, and thioglycine. We identified several genes, including *prmA*, *mm1*, *mm10*, *tfuA*, and *thiI*, that appear to be involved in introducing these PTMs. Various approaches were used for expressing abundant

modified MCR subunits in *E. coli*. All experiments were performed using Duet vectors and a "superplasmid" strategy. In this way, *prmA* was implicated as a SAM-dependent methylase that catalyzes the biosynthesis of 3-methylhistidine and *S*-methylcysteine. Moreover, a thioglycine PTM was detected via coexpression of *mcrABG*, *mm1*, *tfuA*, and *thiI*. This is consistent with *mm1* encoding a TfuA-associated YcaO superfamily enzyme capable of catalyzing thioamide biosynthesis. We have not yet successfully expressed Mm10 or detected the proposed 5-(*S*)-methylarginine PTM. Heterologous expression and activation of Mm10, a hypothesized radical SAM methyltransferase, might be blocked by difficulties in iron-sulfur cluster biosynthesis and the lack of an appropriate methyl donor (e.g., methylcobalamin).

Considering the ultimate goal of constructing active *holo* MCR in non-methanogenic hosts, the studies conducted thus far set a stable foundation. Genomic and enzymological research elucidated the coenzyme F430 biosynthesis pathway, details concerning the incorporation of most of the MCR PTMs, and some hints for achieving heterologous expression in *E. coli*. In the future, we will continue to work on optimizing the *in vivo* biosynthesis of coenzyme F430. This research will provide abundant and controllable coenzyme F430 production in a non-methanogen host as a module for MCR assembly. Moreover, heterologous expression and activation of Mm10 is still under investigation. Furthermore, assembly of the MCR complex will be studied with the combination of information from the *in vivo* coenzyme F430 biosynthesis and MCR PTM studies. Finally, the MCR activation system will be transferred into non-methanogens hosts for *holo* MCR activation. The strategy of *holo* MCR

heterologous expression and activation can be extended to the ANME MCR field or other C1 compound bioconversion systems.

In summary, the work discussed in this dissertation is mainly focus on coenzyme F430 biosynthesis and MCR PTM gene identification. The coenzyme F430 biosynthesis pathway was elucidated by *in vitro* approaches. Then the focus shifted to *in vivo* coenzyme F430 biosynthetic studies in a non-methanogen host. Most of the MCR PTM genes have been identified. However, additional experiments must be performed before the ultimate goals of the research are achieved. All these studies set a stable foundation for the development of methane bioproduction and bioconversion strategies, which have potential applications in improving global warming and the energy crisis.

Appendix

A1. Gibson Assembly

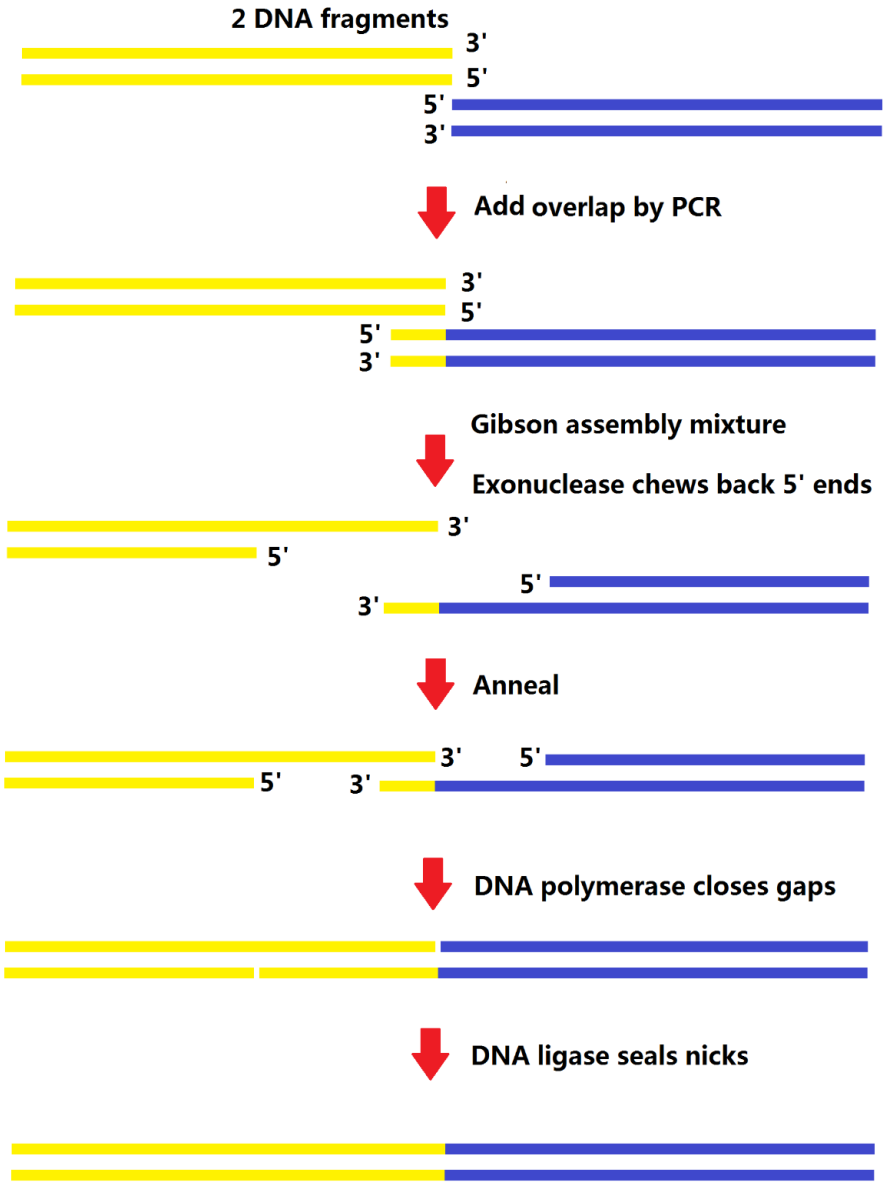


Figure S1: Flowchart of the Gibson Assembly method.

Gibson Assembly is a molecular cloning method that was created by Daniel G. Gibson, which allows the joining of multiple DNA fragments together in a single isothermal reaction (*113*). The ligation points created by Gibson Assembly are traceless. This feature enhances the opportunity for the heterologous expression of active tagged fusion proteins.

Gibson Assembly fragments were obtained by polymerase chain reaction (PCR). Primers were designed to introduce a 15-40 base pair overlap with an adjacent DNA fragment. Three different enzymes participate in Gibson Assembly with the proper buffer system. They are exonuclease, DNA polymerase, and DNA ligase. The exonuclease chews DNA fragments back from 5′ end without inhibiting DNA polymerase and DNA ligase. This process creates a single stranded region that enables the "overlapping region" to anneal. Then the DNA polymerase will fill the gap after annealing. Finally, the DNA ligase covalently joins the DNA of adjacent segments (Figure S1).

In our case, we utilized the Gibson Assembly 2× Master mix. The protocol is briefly summarized below:

1. Design 4 proper PCR primers that fit the overlap principle for Gibson Assembly.
2. Clone 2 fragments for Gibson Assembly by PCR using these primers.
3. Purify the 2 resulting fragments using agarose gel electrophoresis.
4. Mix these 2 DNA fragments with a final concentration 0.02-0.5 pmol. Then add the appropriate amount of water to reach 10 μL.

5. Add 10 μL of Gibson Assembly 2× Master mix to the solution containing the mixed fragments.
6. Incubate the mixture at 50 °C for 20 min to 1 h.
7. Transform the ligated plasmids into NEB 5α Competent *E. coli* cells.
8. Subculture the obtained colonies, then make a plasmid preparation from the subculture and sequence verify the obtained plasmid sample.

A2. SDS-PAGE Gel Preparation

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) is an electrophoretic method that separates charged biomolecules, like proteins, in mixtures according to their molecular mass.

The protocol for SDS-PAGE gel preparation is shown below:

1. Clean glass spacers and plates with 95% ethanol and distilled water. Allow to air-dry sufficiently on paper towel until ready to cast gel.
2. Make resolving gel (Table S1). SDS is sodium dodecyl sulfate. APS is ammonium persulfate. TEMED is tetramethyl ethylenediamine. The recipe is for 4 gels.

	7.5%	10.0%	12%	15%
1.5M Tris-HCl pH 8.8	5ml	5ml	5ml	5ml
30% acrylamide	3.75ml	4ml	6ml	7.5ml
water	6.15ml	5.9ml	3.9ml	2.4ml

20% SDS	75µl	75µl	75µl	75µl
10% APS	75µl	75µl	75µl	75µl
TEMED	25µl	25µl	25µl	25µl

Table S1: Recipe for the SDS-PAGE resolving gel.

3. Assemble gel spacers and holders in a mini-gel casting apparatus. Add resolving gel mixture immediately inbetween spacer and plate. Fill until there is a 0.75 inch space from top of the plate.
4. Immediately add ethanol to overlay the resolving gel in order to remove bubbles. Allow to polymerize for at least 10 min.
5. Rinse off ethanol with distilled water. Ensure there is no water left on the resolving gel.
6. Make stacking gel (Table S2). Immediately add the stacking gel mixture onto the top of resolving gel until it is full. Add appropriate gel combs to stacking gel immediately. Allow to polymerize for at least 10 min.
7. After stacking gel is polymerized, the SDS-PAGE gel is ready for electrophoretic analysis of protein samples. Running the gel at 200 V for 35 min is recommended.

1.5M Tris-HCl pH 8.8	0.62ml
30% acrylamide	0.83ml

water	3.82ml
20% SDS	25µl
10% APS	50µl
TEMED	5µl

Table S2: Recipe for the SDS-PAGE stacking gel.

In our case, we choose a 15% resolving gel for SDS-PAGE analysis of the coenzyme F430 biosynthesis enzymes. A 12% resolving gel is appropriate for SDS-PAGE analysis of the MCR α subunit.

A3. Western-Blot

The western blot, also called protein immunoblot, has been widely used in biology and biochemistry studies for specific protein identification. Synthetic or animal-derived antibodies that binding with specific proteins or tags are utilized. The protein samples are denatured and separated by gel electrophoresis (i.e., SDS-PAGE). Then these samples are transfer to a nitrocellulose membrane. The membrane containing sample proteins are blocked and incubated with the specific antibody. Excess antibody is washed off, then a secondary antibody is usually added for visualization of the protein.

In our case, we do not need to use a secondary antibody due to the specific primary antibody that we used. A brief western blot protocol is shown below:

1. Separate the protein samples using SDS-PAGE. Follow the usual SDS-PAGE protocol up to washing gels with distilled deionized water (ddH_2O). Rinse gels once with western blot transfer buffer and proceed to the protein transfer step.
2. Soak foam pad, filter paper, and nitrocellulose membrane in western blot transfer buffer and chill at 4 °C for 20 min. Transfer buffer: 25 mM Tris, 192 mM glycine, 20% v/v methanol, pH 8.3.
3. In the gel holder cassette, assemble the gel sandwich as follows: Grey side of cassette - foam pad - filter paper - gel (SDS-PAGE) - membrane - filter paper - foam pad. Then gently remove air bubbles from between the layers.
4. Close the cassette firmly; avoid moving the gel and filter paper sandwich.
5. Place the cassette in the electrode module, with the grey side of the cassette aligned with the black and the transparent side of the cassette aligned with the red side of the module.
6. Place the module, a stir bar, and the blue cooling unit (kept at -20 °C) in the buffer tank. Fill the tank with transfer buffer to the "blotting" mark. Transfer for 1 hour at 100 V and 0.35 A at 4 °C.
7. Remove the membrane from the sandwich and place in a separate container. Add blocking buffer and shake for 2 h at room temperature (RT). Blocking buffer (3% gelatin-TBS): 20 mM Tris, 500 mM NaCl, 3% gelatin, pH 7.5.
8. Wash with 1× TTBS and shake for 5 min at RT. Repeat once. Incubate with antibody buffer and shake for 2 – 4 h. Tris buffered saline (1× TBS): 20 mM Tris, 500 mM NaCl, pH 7.5. TTBS (wash solution, 1×): 20 mM Tris, 500 mM NaCl, 0.05% Tween-20, pH 7.5.

Antibody buffer (1% gelatin-TTBS): 20 mM Tris, 500 mM NaCl, 1% gelatin, 0.05% Tween-20, pH 7.5. Primary antibody: Monoclonal Anti-poly Histidine-Alkaline Phosphatase, Clone HIS-1 (reusable for 5 times), supplied by Sigma-Aldrich, containing 12.5 µl of antibody in 25 mL antibody buffer and store at -4 °C.

9. Wash with 1× TTBS and shake for 5 min at RT. Then wash with 1× TBS, shake for 5 min at RT, and discard solution.
10. Add 25 mL of 1× AP (alkaline phosphatase) color development buffer to membrane and shake at RT until bands become visible. Wash membrane with distilled deionized water (ddH_2O) for 5 min with shaking and repeat twice. Air dry the membrane. Alkaline phosphatase (AP) color development buffer: 250 µL of AP color reagent A (kept at -20 °C), 250 µL AP color reagent B (kept at -20 °C), 1 mL of 25× AP Color development buffer (kept at -4 °C), diluted to 25 mL with ddH_2O.

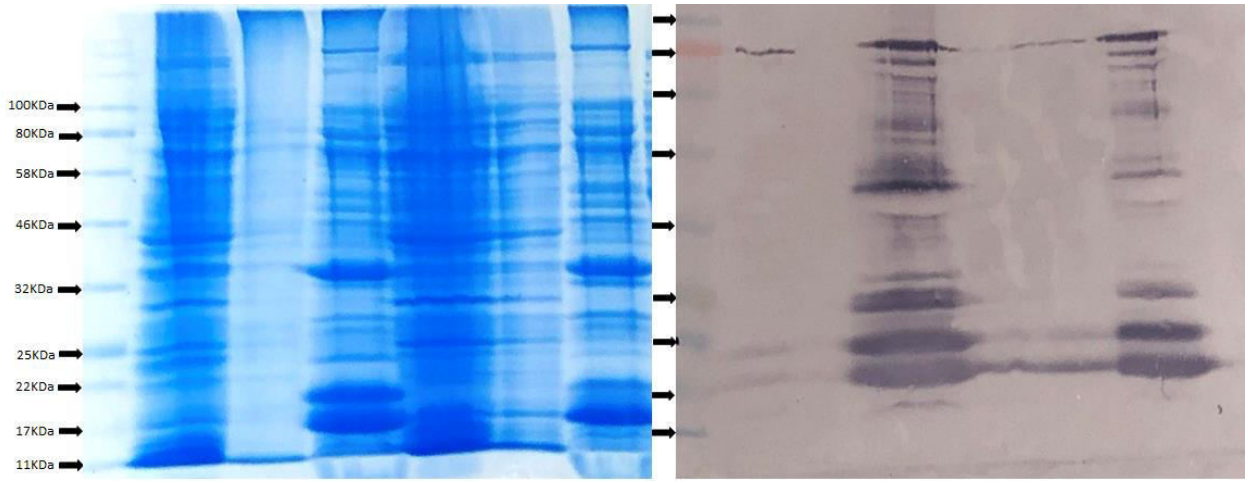


Figure S2: The SDS-PAGE gel (left) and western blot (right) from proteins expressed by the RSUMO-McrA cell line, which contains pETSUMO-*mcrA* vector. Lanes 1-7 are as follows:

molecular weight standard; flow-through, wash, and elution from LB expressed cells; flow through, wash, and elution from TB expressed cells.

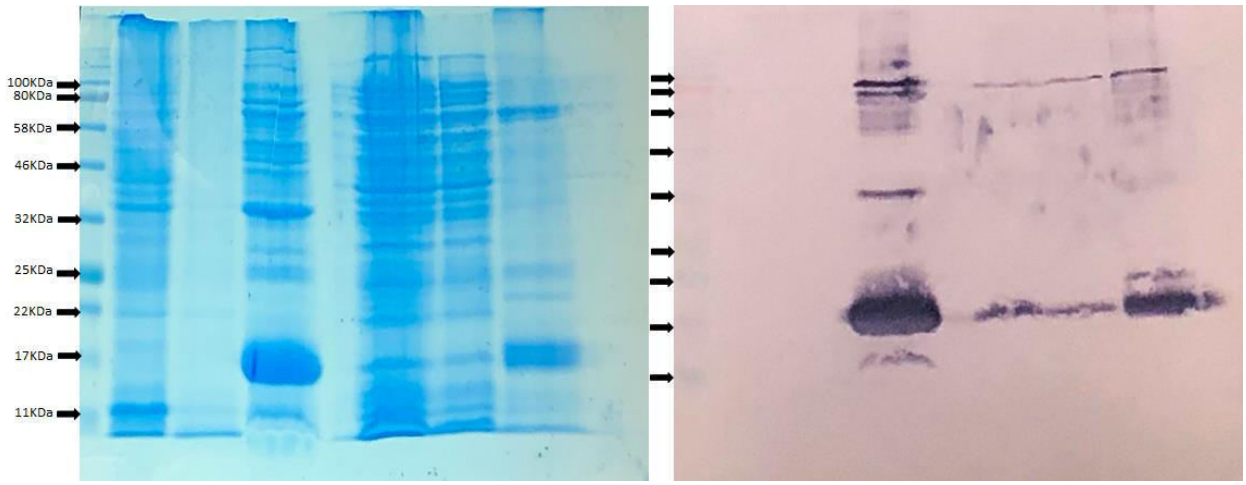


Figure S3 The SDS-PAGE gel (left) and western blot (right) from proteins expressed by the optSUMO-McrA cell line, which contains the pETSUMO-*optmcrA* vector. Lanes 1-7 are as follows: molecular weight standard; flow-through, wash, and elution from LB expressed cells; flow through, wash, and elution from TB expressed cells.

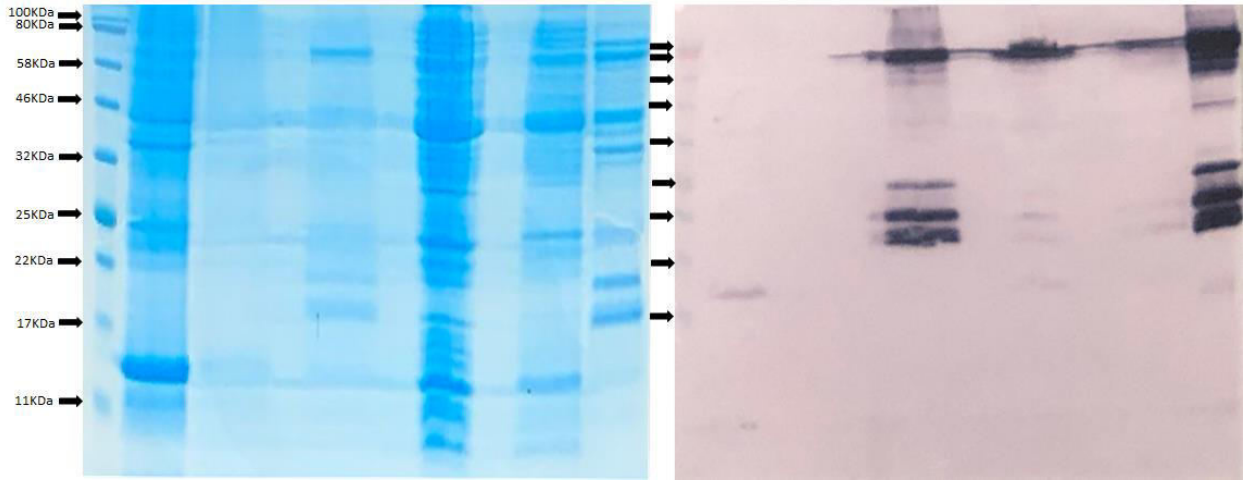


Figure S4: The SDS-PAGE gel (left) and western blot (right) from proteins expressed by the RSUMO-McrA-Mm-PrmA cell line, which contains the pETSUMO-*mcrA*, pCDFDuet:*mm1-mm10*, and pETDuet:*tfuA-prmA* vectors. Lanes 1-7 are as follows: molecular weight standard;

flow-through, wash, and elution from LB expressed cells; flow-through, wash, and elution from TB expressed cells.

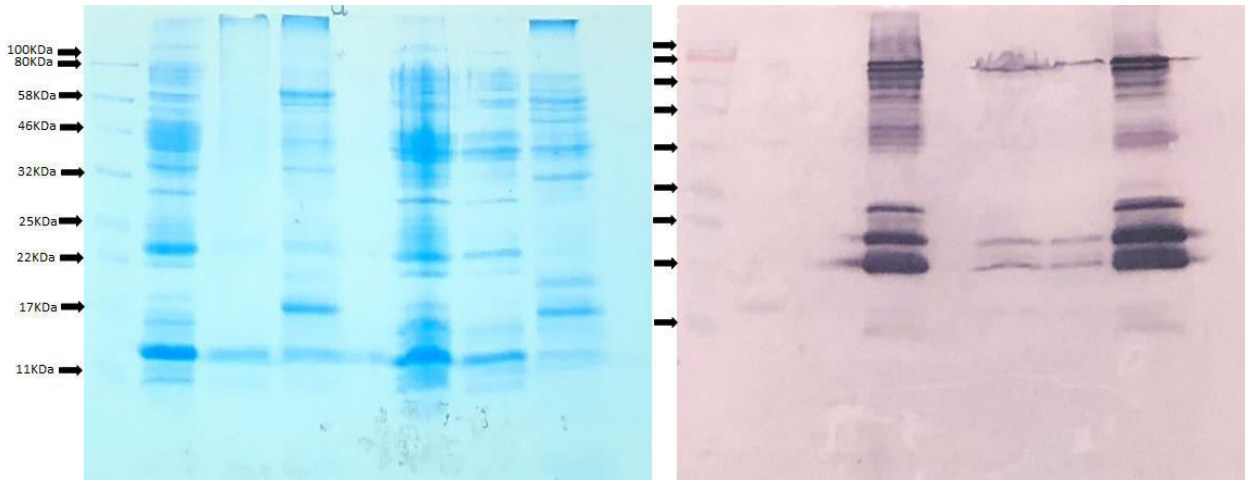


Figure S5: The SDS-PAGE gel (left) and western blot (right) from proteins expressed by the RSUMO-McrA-Mm-ThiI cell line, which contains the pETSUMO-*mcrA* and pCDFDuet:*mm1-mm10-tfuA-thiI* vectors. Lanes 1-7 are as follows: molecular weight standard; flow-through, wash, and elution from LB expressed cells; flow-through, wash, and elution from TB expressed cells.

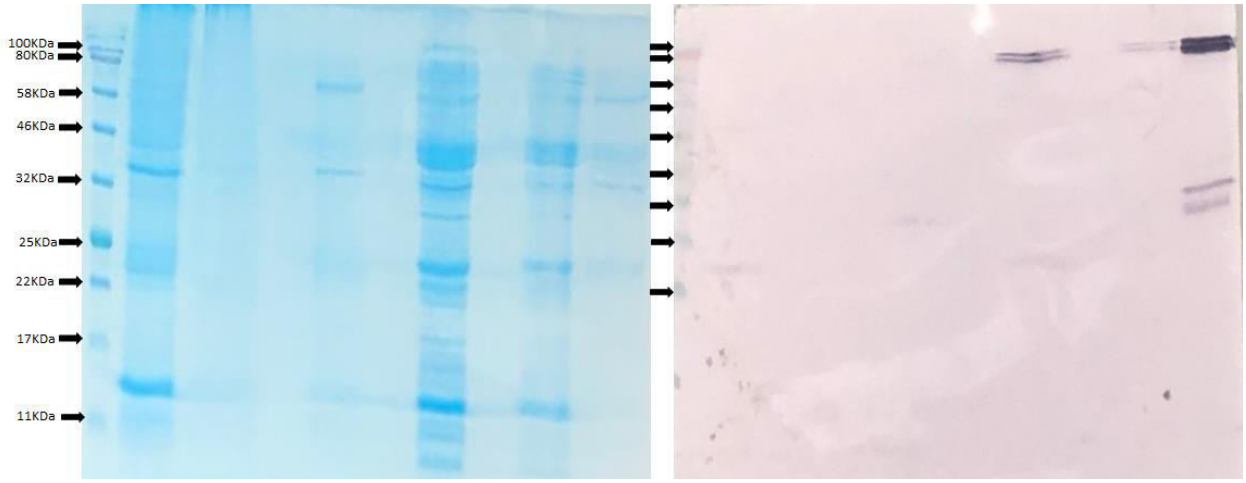


Figure S6: The SDS-PAGE gel (left) and western blot (right) from proteins expressed by the RSUMO-McrABG-Mm-ThiI cell line, which contains the pETSUMO-*mcrA*, pETDuet:*mcrB-mcrG*, and pCDFDuet:*mm1-mm10-tfuA-thiI* vectors. Lanes 1-7 are as follows: molecular weight

standard; flow-through, wash, and elution from LB expressed cells; flow-through, wash, and elution from TB expressed cells.

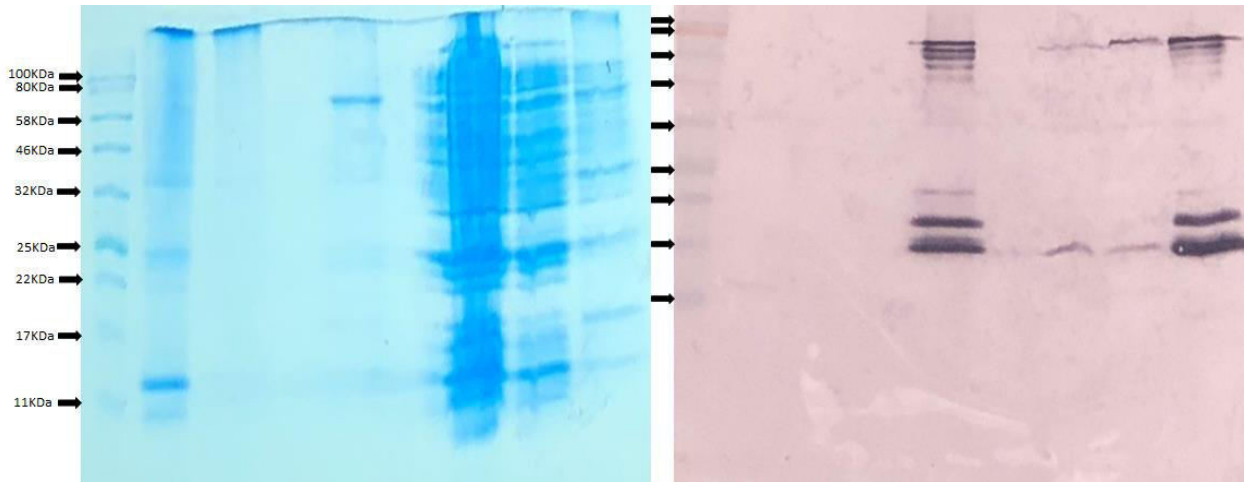


Figure S7: The SDS-PAGE gel (left) and western blot (right) from proteins expressed by the RSUMO-McrABG-PrmA cell line, which contains the pETSUMO-*mcrA*, pETDuet:*mcrB-mcrG*, and pCDFDuet:*tfuA-prmA* vectors. Lanes 1-7 are as follows: molecular weight standard; flow-through, wash, and elution from LB expressed cells; flow-through, wash, and elution from TB expressed cells.

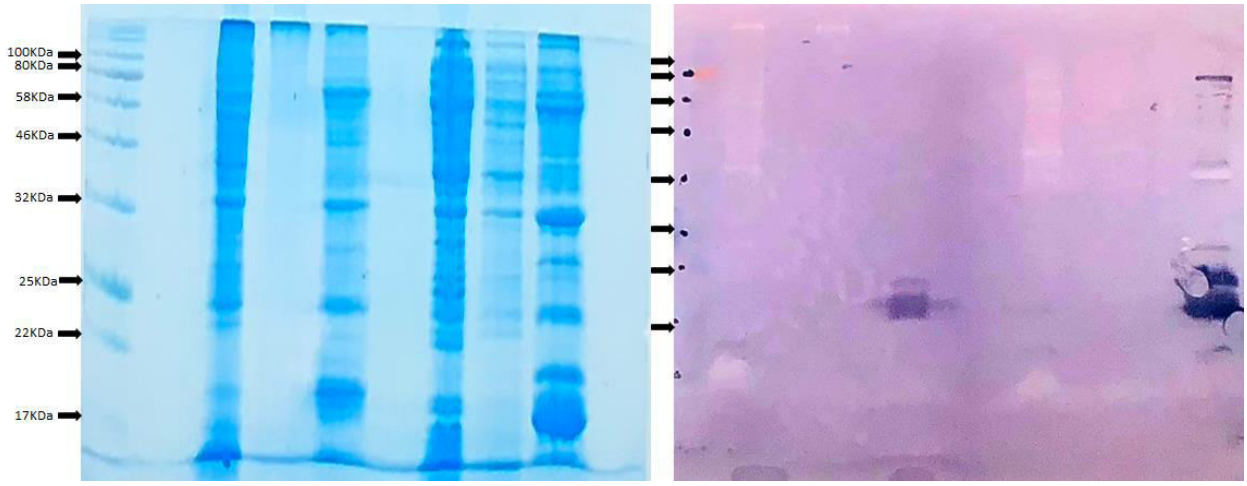


Figure S8: The SDS-PAGE gel (left) and western blot (right) from proteins expressed by the RSUMO-McrABG cell line, which contains the pETSUMO-*mcrA* and pETDuet:*mcrB-mcrG*

vectors. Lanes 1-7 are as follows: molecular weight standard; flow-through, wash, and elution from LB expressed cells; flow-through, wash, and elution from TB expressed cells.

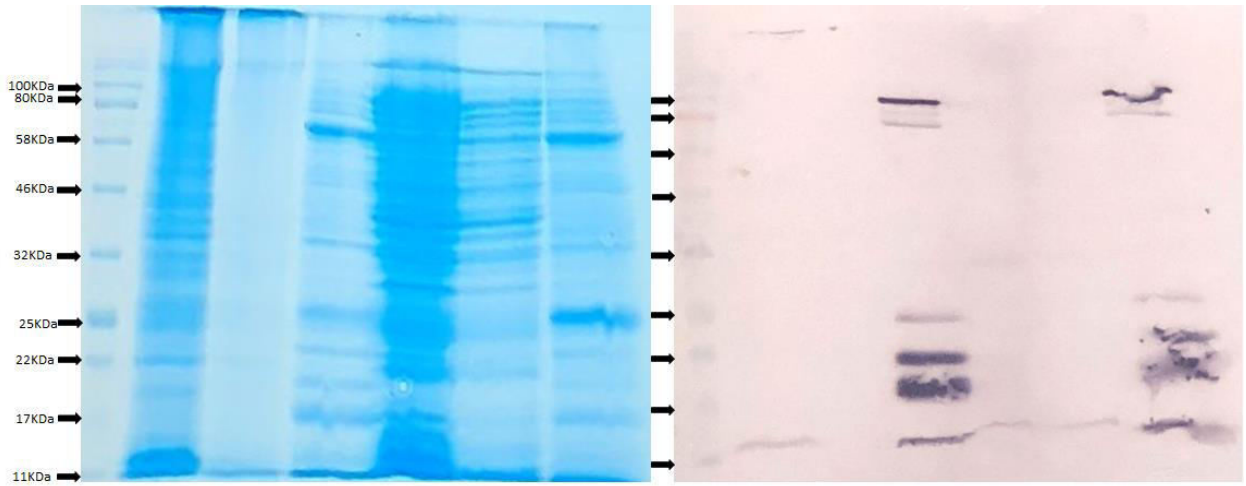


Figure S9: The SDS-PAGE gel (left) and western blot (right) from proteins expressed by the optSUMO-McrABG cell line, which contains the pETSUMO-*optmcrA* and pETDuet:*mcrB-mcrG* vectors. Lanes 1-7 are as follows: molecular weight standard; flow-through, wash, and elution from LB expressed cells; flow-through, wash, and elution from TB expressed cells.

A4. PTM MS reports

A
1 MAADIFAKFK KSMEVKFTQE YGSNKQAGGD ITGKTEKFLR LGPEQDARKQ
51 EMIKAGKEIA EKRGIAFYNP MMHMGAPLGQ RAITPYTISG TDIVAEPDDL
101 HYVNNAAMQQ MWDDIRRTCI VGLDMAHETL EKRLGKEVTP ETINHYLETL
151 NHAMPGAAVV QEMMVETHPA LVDDCYVKIF TGDDELADEI DKQYVINVNK
201 MFSEEQAAQI KASIGKTTWQ AIHIPTIVSR TTDGAQTSRW AAMQIGMSFI
251 SAYAMCAGEA AVADLSFAAK HAALVSMGEM LPARRARGPN EPGGLSFGHL
301 SDIVQTSRVS KDPAKIALEV VGAGCMLYDQ IWLGSYMSGG VGFTQYATAA
351 YTDDILDNNT YYDVDYINDK YNGAANLGTD NKVKATLDVV KDIATESTLY
401 GIETYEKFPT ALEDHFGGSQ RATVLAAASG VACALATGNA NAGLSGWYLS
451 MYVHKEAWGR LGFFGFDLQD QCGATNVLSY QGDEGLPDEL RGPNYPNYAM
501 NVGHQGGYAG IAQAAHSGRG DAFTVNPLLK VCFADELMPF NFAEPRREFG
551 RGAIREFMPA GERSLVIPAK
B
1 MAADIFAKFK KSMEVKFTQE YGSNKQAGGD ITGKTEKFLR LGPEQDARKQ
51 EMIKAGKEIA EKRGIAFYNP MMHMGAPLGQ RAITPYTISG TDIVAEPDDL
101 HYVNNAAMQQ MWDDIRRTCI VGLDMAHETL EKRLGKEVTP ETINHYLETL
151 NHAMPGAAVV QEMMVETHPA LVDDCYVKIF TGDDELADEI DKQYVINVNK
201 MFSEEQAAQI KASIGKTTWQ AIHIPTIVSR TTDGAQTSRW AAMQIGMSFI
251 SAYAMCAGEA AVADLSFAAK HAALVSMGEM LPARRARGPN EPGGLSFGHL
301 SDIVQTSRVS KDPAKIALEV VGAGCMLYDQ IWLGSYMSGG VGFTQYATAA
351 YTDDILDNNT YYDVDYINDK YNGAANLGTD NKVKATLDVV KDIATESTLY
401 GIETYEKFPT ALEDHFGGSQ RATVLAAASG VACALATGNA NAGLSGWYLS
451 MYVHKEAWGR LGFFGFDLQD QCGATNVLSY QGDEGLPDEL RGPNYPNYAM
501 NVGHQGGYAG IAQAAHSGRG DAFTVNPLLK VCFADELMPF NFAEPRREFG
551 RGAIREFMPA GERSLVIPAK

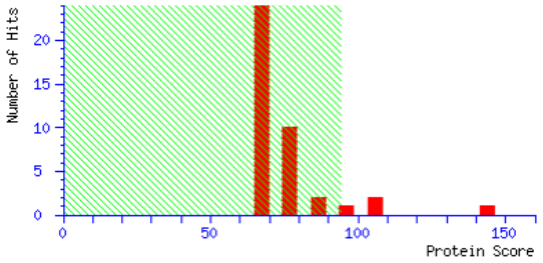
Figure S10: MS protein coverage (red) of the MCR α subunit expressed by the McrA-PrmA cell line after digestion with trypsin. Sample A contains the HAALVSMGEMLPARR peptide (underlined), which contains the 1-*N*-methylhistidine PTM. Sample B contains the LGFFGFDLQDQCGATNVLSYQGDEGLPDELR peptide (underlined), which contains the *S*-methylcysteine PTM.

MATRIX SCIENCE Mascot Search Results

```
User            : drp
Email           : drp@uga.edu
Search title    : Thil 1
Database        : NCBIprot 20171024 (135744157 sequences; 49805139192 residues)
Timestamp       : 9 Nov 2017 at 17:12:29 GMT
Top Score       : 144 for WP_011024419.1, coenzyme-B sulfoethylthiotransferase subunit alph
```

Mascot Score Histogram

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 94 are significant (p<0.05).



Concise Protein Summary Report

Format As: Concise Protein Summary | Help
Significance threshold p< 0.05 | Max. number of hits AUTO
Preferred taxonomy: All entries

Re-Search All | Search Unmatched

1. WP_011024419.1 **Mass:** 62035 **Score:** 144 **Expect:** 5.4e-07 **Matches:** 15
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina acetivorans]
WP_048174272.1 **Mass:** 62019 **Score:** 117 **Expect:** 0.00027 **Matches:** 13
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina siciliae]

2. WP_048141827.1 **Mass:** 62274 **Score:** 102 **Expect:** 0.0086 **Matches:** 12
MULTISPECIES: coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina]
WP_048131296.1 **Mass:** 62246 **Score:** 91 **Expect:** 0.12 **Matches:** 11
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina sp. 1.H.T.1A.1]

Figure S11: MASCOT search results based on the MALDI-TOF MS spectrum of the trypsin-digested MCR α subunit produced by the McrA-ThiI cell line. The search result indicate that the SDS-PAGE gel slice contains the MCR α subunit.

1 MAADIFAKFK KSMEVKFTQE YGSNKQAGGD ITGKTEKFLR LGPEQDARKQ
51 EMIKAGKEIA EKRGIAFYNP MMHMGAPLGQ RAITPYTISG TDIVAEPDDL
101 HYVNNAAMQQ MWDDIRRTCI VGLDMAHETL EKRLGKEVTP ETINHYLETL
151 NHAMPGAAVV QEMMVETHPA LVDDCYVKIF TGDDELADEI DKQYVINVNK
201 MFSEEQAAQI KASIGKTTWQ AIHIPTIVSR TTDGAQTSRW AAMQIGMSFI
251 SAYAMCAGEA AVADLSFAAK HAALVSMGEM LPARRARGPN EPGGLSFGHL
301 SDIVQTSRVS KDPAKIALEV VGAGCMLYDQ IWLGSYMSGG VGFTQYATAA
351 YTDDILDNNT YYDVDYINDK YNGAANLGTD NKVKATLDVV KDIATESTLY
401 GIETYEKFPT ALEDHFGGSQ RATVLAAASG VACALATGNA NAGLSGWYLS
451 MYVHKEAWGR LGFFGFDLQD QCGATNVLSY QGDEGLPDEL RGPNYPNYAM
501 NVGHQGGYAG IAQAAHSGRG DAFTVNPLLK VCFADELMPF NFAEPRREFG
551 RGAIREFMPA GERSLVIPAK

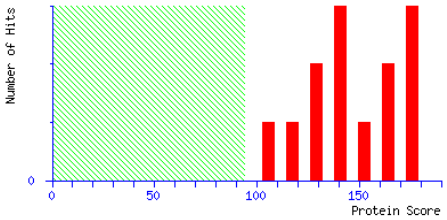
Figure S12: MS protein coverage (red) of the MCR α subunit expressed by the McrA-ThiI cell line after digestion with trypsin. The sample contains the FGFDLQDQCGATNVLS peptide (underlined), which contains the thioglycine PTM.

MATRIX SCIENCE **Mascot Search Results**

User : drp
Email : drp@uga.edu
Search title : Prma 2
Database : NCBIprot 20171205 (139213787 sequences; 51013024959 residues)
Timestamp : 10 Apr 2018 at 20:03:44 GMT
Top Score : 176 for Mixture 1, WP_052934128.1 + WP_011024419.1

Mascot Score Histogram

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 94 are significant (p<0.05).



Concise Protein Summary Report

Format As | Concise Protein Summary | Help
Significance threshold p< 0.05 | Max. number of hits AUTO

Re-Search All | Search Unmatched

1. **Mixture 1** **Total score: 176** **Expect:** 3.5e-10 **Matches:** 41
Components (only one family member shown for each component):
WP_052934128.1 **Mass:** 74311 **Score: 143** **Expect:** 7e-07 **Matches:** 22
bifunctional UDP-4-amino-4-deoxy-L-arabinose formyltransferase/UDP-glucuronic acid oxidase ArnA [Escherichia coli]
WP_011024419.1 **Mass:** 62035 **Score: 109** **Expect:** 0.0018 **Matches:** 21
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina acetivorans]

---

2. **Mixture 2** **Total score: 175** **Expect:** 4.4e-10 **Matches:** 41
Components (only one family member shown for each component):
WP_029400001.1 **Mass:** 74311 **Score: 141** **Expect:** 1.1e-06 **Matches:** 22
bifunctional UDP-4-amino-4-deoxy-L-arabinose formyltransferase/UDP-glucuronic acid oxidase ArnA [Escherichia coli]
WP_011024419.1 **Mass:** 62035 **Score: 109** **Expect:** 0.0018 **Matches:** 21
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina acetivorans]

---

3. **Mixture 3** **Total score: 173** **Expect:** 7e-10 **Matches:** 41
Components (only one family member shown for each component):
WP_000649095.1 **Mass:** 74073 **Score: 139** **Expect:** 1.8e-06 **Matches:** 22
MULTISPECIES: bifunctional UDP-4-amino-4-deoxy-L-arabinose formyltransferase/UDP-glucuronic acid oxidase ArnA [Escherichia]
WP_011024419.1 **Mass:** 62035 **Score: 109** **Expect:** 0.0018 **Matches:** 21
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina acetivorans]

---

4. **Mixture 4** **Total score: 163** **Expect:** 7e-09 **Matches:** 40
Components (only one family member shown for each component):
WP_072720808.1 **Mass:** 74182 **Score: 132** **Expect:** 8.8e-06 **Matches:** 21
bifunctional UDP-4-amino-4-deoxy-L-arabinose formyltransferase/UDP-glucuronic acid oxidase ArnA [Escherichia coli]
WP_011024419.1 **Mass:** 62035 **Score: 109** **Expect:** 0.0018 **Matches:** 21
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina acetivorans]

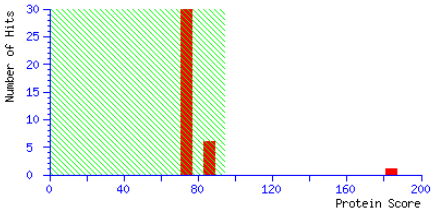
Figure S13: MASCOT search results based on the MALDI-TOF MS spectrum of the trypsin-digested MCR α subunit produced by the RSUMO-McrABG-PrmA cell line. This search result indicates that the SDS-PAGE gel slice contains the MCR α subunit.

MATRIX SCIENCE **Mascot Search Results**

User : drp
Email : drp@uga.edu
Search title : McR 4
Database : NCBIprot 20171205 (139213787 sequences; 51013024959 residues)
Timestamp : 10 Apr 2018 at 19:52:16 GMT
Top Score : 184 for WP_011024419.1, coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina acetivorans]

Mascot Score Histogram

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 94 are significant (p<0.05).



Concise Protein Summary Report

Format As: Concise Protein Summary | Help
Significance threshold p< 0.05 | Max. number of hits AUTO
Preferred taxonomy: All entries

Re-Search All | Search Unmatched

1. WP_011024419.1 **Mass:** 62035 **Score:** 184 **Expect:** 5.5e-11 **Matches:** 27
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina acetivorans]
WP_048174272.1 **Mass:** 62019 **Score:** 141 **Expect:** 1.1e-06 **Matches:** 23
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina siciliae]
AGF96667.1 **Mass:** 61903 **Score:** 110 **Expect:** 0.0014 **Matches:** 19
Methyl coenzyme M reductase alpha subunit [Methanosarcina mazei Tuc01]
WP_048141827.1 **Mass:** 62274 **Score:** 97 **Expect:** 0.03 **Matches:** 16
MULTISPECIES: coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina]
WP_011033189.1 **Mass:** 61874 **Score:** 93 **Expect:** 0.065 **Matches:** 17
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina mazei]
WP_048131296.1 **Mass:** 62246 **Score:** 86 **Expect:** 0.34 **Matches:** 15
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina sp. 1.H.T.1A.1]
WP_048037134.1 **Mass:** 61928 **Score:** 81 **Expect:** 1 **Matches:** 16
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina mazei]
WP_048037303.1 **Mass:** 61912 **Score:** 81 **Expect:** 1 **Matches:** 16
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina mazei]
WP_048048304.1 **Mass:** 61944 **Score:** 81 **Expect:** 1 **Matches:** 16
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina mazei]
WP_048044585.1 **Mass:** 61972 **Score:** 81 **Expect:** 1 **Matches:** 16
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina mazei]
WP_048050667.1 **Mass:** 61942 **Score:** 81 **Expect:** 1 **Matches:** 16
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina soligelidi]
WP_048136663.1 **Mass:** 62071 **Score:** 71 **Expect:** 11 **Matches:** 18
coenzyme-B sulfoethylthiotransferase subunit alpha [Methanosarcina horonobensis]

2. OLQ10709.1 **Mass:** 327968 **Score:** 86 **Expect:** 0.33 **Matches:** 24
Retrovirus-related Pol polyprotein from transposon TNT 1-94 [Symbiodinium microadriaticum]

Figure S14: MASCOT search results based on the MALDI-TOF MS spectrum of the trypsin-digested MCR α subunit produced by the RSUMO-McrABG-Mm-ThiI cell line. This search result indicates that the SDS-PAGE gel slice contains the MCR α subunit.

References

1. C. R. Woese, O. Kandler, M. L. Wheelis, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4576–4579 (1990).

2. C. R. Woese, G. E. Fox, Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5088–5090 (1977).

3. V. Peters, R. Conrad, Methanogenic and other strictly anaerobic bacteria in desert soil and other oxic soils. *Appl. Environ. Microbiol.* **61**, 1673–1676 (1995).

4. L. Appels, J. Baeyens, J. Degrève, R. Dewil, Principles and potential of the anaerobic digestion of waste-activated sludge. *Prog. Energy Combust. Sci.* **34**, 755–781 (2008).

5. F. Ali Shah, Q. Mahmood, M. Maroof Shah, A. Pervez, S. Ahmad Asad, Microbial Ecology of Anaerobic Digesters: The Key Players of Anaerobiosis. *Sci. World J.* **2014**, 1–21 (2014).

6. G. Lettinga, Anaerobic digestion and wastewater treatment systems. *Antonie Van Leeuwenhoek*. **67**, 3–28 (1995).

7. M. Tabatabaei *et al.*, Importance of the methanogenic archaea populations in anaerobic wastewater treatments. *Process Biochem.* **45**, 1214–1225 (2010).

8. R. K. Thauer, Biochemistry of methanogenesis: a tribute to Marjory Stephenson:1998 Marjory Stephenson Prize Lecture. *Microbiology*. **144**, 2377–2406 (1998).

9. S. W. Ragsdale, in *The Metal-Driven Biogeochemistry of Gaseous Compounds in the Environment*, P. M. H. Kroneck, M. E. S. Torres, Eds. (Springer Netherlands, Dordrecht, 2014; http://link.springer.com/10.1007/978-94-017-9269-1_6), vol. 14, pp. 125–145.

10. D. J. Aceti, J. G. Ferry, Purification and characterization of acetate kinase from acetate-grown Methanosarcina thermophila. Evidence for regulation of synthesis. *J. Biol. Chem.* **263**, 15444–15448 (1988).

11. D. R. Abbanat, J. G. Ferry, Synthesis of acetyl coenzyme A by carbon monoxide dehydrogenase complex from acetate-grown Methanosarcina thermophila. *J. Bacteriol.* **172**, 7145–7150 (1990).

12. P. A. Bertram *et al.*, Formylmethanofuran dehydrogenases from methanogenic Archaea. Substrate specificity, EPR properties and reversible inactivation by cyanide of the molybdenum or tungsten iron-sulfur proteins. *Eur. J. Biochem.* **220**, 477–484 (1994).

13. M. I. Donnelly, R. S. Wolfe, The role of formylmethanofuran: tetrahydromethanopterin formyltransferase in methanogenesis from carbon dioxide. *J. Biol. Chem.* **261**, 16653–16659 (1986).

14. D. J. Ferguson, J. A. Krzycki, Reconstitution of trimethylamine-dependent coenzyme M methylation with the trimethylamine corrinoid protein and the isozymes of methyltransferase II from Methanosarcina barkeri. *J. Bacteriol.* **179**, 846–852 (1997).

15. D. J. Ferguson, N. Gorlatova, D. A. Grahame, J. A. Krzycki, Reconstitution of dimethylamine:coenzyme M methyl transfer with a discrete corrinoid protein and two methyltransferases purified from Methanosarcina barkeri. *J. Biol. Chem.* **275**, 29053–29060 (2000).

16. S. A. Burke, J. A. Krzycki, Reconstitution of Monomethylamine:Coenzyme M methyl transfer with a corrinoid protein and two methyltransferases purified from Methanosarcina barkeri. *J. Biol. Chem.* **272**, 16570–16577 (1997).

17. K. Sauer, R. K. Thauer, Methanol:coenzyme M methyltransferase from Methanosarcina barkeri -- substitution of the corrinoid harbouring subunit MtaC by free cob(I)alamin. *Eur. J. Biochem.* **261**, 674–681 (1999).

18. U. Deppenmeier, The unique biochemistry of methanogenesis. *Prog. Nucleic Acid Res. Mol. Biol.* **71**, 223–283 (2002).

19. J. Ellermann, R. Hedderich, R. Bocher, R. K. Thauer, The final step in methane formation. Investigations with highly purified methyl-CoM reductase (component C) from Methanobacterium thermoautotrophicum (strain Marburg). *Eur. J. Biochem.* **172**, 669–677 (1988).

20. C. A. Orengo, J. M. Thornton, Alpha plus beta folds revisited: some favoured motifs. *Struct. Lond. Engl. 1993*. **1**, 105–120 (1993).

21. U. Ermler, M. Merckel, R. Thauer, S. Shima, Formylmethanofuran: tetrahydromethanopterin formyltransferase from Methanopyrus kandleri — new insights into salt-dependence and thermostability. *Structure*. **5**, 635–646 (1997).

22. U. Ermler, Crystal Structure of Methyl-Coenzyme M Reductase: The Key Enzyme of Biological Methane Formation. *Science*. **278**, 1457–1462 (1997).

23. K. Knittel, A. Boetius, Anaerobic Oxidation of Methane: Progress with an Unknown Process. *Annu. Rev. Microbiol.* **63**, 311–334 (2009).

24. J. Reimann, M. S. M. Jetten, J. T. Keltjens, in *Sustaining Life on Planet Earth: Metalloenzymes Mastering Dioxygen and Other Chewy Gases*, P. M. H. Kroneck, M. E. Sosa Torres, Eds. (Springer International Publishing, Cham, 2015; http://link.springer.com/10.1007/978-3-319-12415-5_7), vol. 15, pp. 257–313.

25. S. Shima *et al.*, Structure of a methyl-coenzyme M reductase from Black Sea mats that oxidize methane anaerobically. *Nature*. **481**, 98–101 (2011).

26. S. Mayr *et al.*, Structure of an F430 Variant from Archaea Associated with Anaerobic Oxidation of Methane. *J. Am. Chem. Soc.* **130**, 10758–10767 (2008).

27. S. Shima, R. K. Thauer, Methyl-coenzyme M reductase and the anaerobic oxidation of methane in methanotrophic Archaea. *Curr. Opin. Microbiol.* **8**, 643–648 (2005).

28. E. C. Duin, N. J. Cosper, F. Mahlert, R. K. Thauer, R. A. Scott, Coordination and geometry of the nickel atom in active methyl-coenzyme M reductase from Methanothermobacter marburgensis as detected by X-ray absorption spectroscopy. *JBIC J. Biol. Inorg. Chem.* **8**, 141–148 (2003).

29. T. Wongnate *et al.*, The radical mechanism of biological methane synthesis by methyl-coenzyme M reductase. *Science*. **352**, 953–958 (2016).

30. D. Prakash, Y. Wu, S.-J. Suh, E. C. Duin, Elucidating the Process of Activation of Methyl-Coenzyme M Reductase. *J. Bacteriol.* **196**, 2491–2498 (2014).

31. S. J. Hallam, Reverse Methanogenesis: Testing the Hypothesis with Environmental Genomics. *Science*. **305**, 1457–1462 (2004).

32. S. L. Caldwell *et al.*, Anaerobic Oxidation of Methane: Mechanisms, Bioenergetics, and the Ecology of Associated Microorganisms. *Environ. Sci. Technol.* **42**, 6791–6799 (2008).

33. G. Wegener, V. Krukenberg, D. Riedel, H. E. Tegetmeyer, A. Boetius, Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature*. **526**, 587–590 (2015).

34. S. E. McGlynn, G. L. Chadwick, C. P. Kempes, V. J. Orphan, Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature*. **526**, 531–535 (2015).

35. G. Kikuchi, A. Kumar, P. Talmage, D. Shemin, The enzymatic synthesis of delta-aminolevulinic acid. *J. Biol. Chem.* **233**, 1214–1219 (1958).

36. E. L. Bolt, L. Kryszak, J. Zeilstra-Ryalls, P. M. Shoolingin-Jordan, M. J. Warren, Characterization of the rhodobacter sphaeroides 5-aminolaevulinic acid synthase isoenzymes, HemA and HemT, isolated from recombinant Escherichia coli. *Eur. J. Biochem.* **265**, 290–299 (1999).

37. B.-X. Tian, E. Erdtman, L. A. Eriksson, Catalytic Mechanism of Porphobilinogen Synthase: The Chemical Step Revisited by QM/MM Calculations. *J. Phys. Chem. B*. **116**, 12105–12112 (2012).

38. E. Y. Levin, D. L. Coleman, The enzymatic conversion of porphobilinogen to uroporphyrinogen catalyzed by extracts of hematopoietic mouse spleen. *J. Biol. Chem.* **242**, 4247–4253 (1967).

39. A. R. Battersby, C. J. Fookes, G. W. Matcham, E. McDonald, Biosynthesis of the pigments of life: formation of the macrocycle. *Nature*. **285**, 17–21 (1980).

40. J. M. Ravel, S. F. Wang, C. Heinemeyer, W. Shive, GLUTAMYL AND GLUTAMINYL RIBONUCLEIC ACID SYNTHETASES OF ESCHERICHIA COLI W. SEPARATION, PROPERTIES, AND STIMULATION OF ADENOSINE TRIPHOSPHATE-PYROPHOSPHATE EXCHANGE BY ACCEPTOR RIBONUCLEIC ACID. *J. Biol. Chem.* **240**, 432–438 (1965).

41. B. Pontoppidan, C. G. Kannangara, Purification and partial characterisation of barley glutamyl-tRNA(Glu) reductase, the enzyme that directs glutamate to chlorophyll biosynthesis. *Eur. J. Biochem.* **225**, 529–537 (1994).

42. L. L. Ilag, D. Jahn, G. Eggertsson, D. Söll, The Escherichia coli hemL gene encodes glutamate 1-semialdehyde aminotransferase. *J. Bacteriol.* **173**, 3408–3413 (1991).

43. F. Blanche *et al.*, Purification, characterization, and molecular cloning of S-adenosyl-L-methionine: uroporphyrinogen III methyltransferase from Methanobacterium ivanovii. *J. Bacteriol.* **173**, 4637–4645 (1991).

44. D. Mauzerall, S. Granick, Porphyrin biosynthesis in erythrocytes. III. Uroporphyrinogen and its decarboxylase. *J. Biol. Chem.* **232**, 1141–1162 (1958).

45. A. A. Brindley, E. Raux, H. K. Leech, H. L. Schubert, M. J. Warren, A Story of Chelatase Evolution: IDENTIFICATION AND CHARACTERIZATION OF A SMALL 13-15-kDa "ANCESTRAL" COBALTOCHELATASE (CbiXS) IN THE ARCHAEA. *J. Biol. Chem.* **278**, 22388–22395 (2003).

46. H. K. Leech, E. Raux-Deery, P. Heathcote, M. J. Warren, Production of cobalamin and sirohaem in Bacillus megaterium: an investigation into the role of the branchpoint

chelatases sirohydrochlorin ferrochelatase (SirB) and sirohydrochlorin cobalt chelatase (CbiX). *Biochem. Soc. Trans.* **30**, 610–613 (2002).

47. V. Fresquet, L. Williams, F. M. Raushel, Mechanism of Cobyrinic Acid *a* , *c* -Diamide Synthetase from *Salmonella typhimurium* LT2 †. *Biochemistry (Mosc.)*. **43**, 10619–10627 (2004).

48. L. Debussche, D. Thibaut, B. Cameron, J. Crouzet, F. Blanche, Purification and characterization of cobyrinic acid a,c-diamide synthase from Pseudomonas denitrificans. *J. Bacteriol.* **172**, 6239–6244 (1990).

49. S. N. Dedysh, NifH and NifD phylogenies: an evolutionary basis for understanding nitrogen fixation capabilities of methanotrophic bacteria. *Microbiology*. **150**, 1301–1313 (2004).

50. N. Carrillo, E. A. Ceccarelli, Open questions in ferredoxin-NADP+ reductase catalytic mechanism. *Eur. J. Biochem.* **270**, 1900–1915 (2003).

51. G. Kurisu *et al.*, Structure of the electron transfer complex between ferredoxin and ferredoxin-NADP(+) reductase. *Nat. Struct. Biol.* **8**, 117–121 (2001).

52. V. I. Dumit, T. Essigke, N. Cortez, G. M. Ullmann, Mechanistic Insights into Ferredoxin–NADP(H) Reductase Catalysis Involving the Conserved Glutamate in the Active Site. *J. Mol. Biol.* **397**, 814–825 (2010).

53. J. Yang, X. Xie, M. Yang, R. Dixon, Y.-P. Wang, Modular electron-transport chains from eukaryotic organelles function to support nitrogenase activity. *Proc. Natl. Acad. Sci.* **114**, E2460–E2465 (2017).

54. J. A. Bertrand *et al.*, Crystal structure of UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase from Escherichia coli. *EMBO J.* **16**, 3416–3425 (1997).

55. W. Grabarse, F. Mahlert, S. Shima, R. K. Thauer, U. Ermler, Comparison of three methyl-coenzyme M reductases from phylogenetically distant organisms: unusual amino acid modification, conservation and adaptation. *J. Mol. Biol.* **303**, 329–344 (2000).

56. T. Selmer *et al.*, The Biosynthesis of Methylated Amino Acids in the Active Site Region of Methyl-coenzyme M Reductase. *J. Biol. Chem.* **275**, 3755–3760 (2000).

57. H. Demirci, S. T. Gregory, A. E. Dahlberg, G. Jogl, Multiple-site trimethylation of ribosomal protein L11 by the PrmA methyltransferase. *Struct. Lond. Engl. 1993*. **16**, 1059–1066 (2008).

58. F. N. Chang, L. B. Cohen, I. J. Navickas, C. N. Chang, Purification and properties of a ribosomal protein methylase from Eschericha coli Q13. *Biochemistry (Mosc.)*. **14**, 4994–4998 (1975).

59. A. Rios, T. L. Amyes, J. P. Richard, Formation and Stability of Organic Zwitterions in Aqueous Solution: Enolates of the Amino Acid Glycine and Its Derivatives. *J. Am. Chem. Soc.* **122**, 9373–9385 (2000).

60. F. G. Bordwell, S. Zhang, X.-M. Zhang, W.-Z. Liu, Homolytic Bond Dissociation Enthalpies of the Acidic H-A Bonds Caused by Proximate Substituents in Sets of Methyl Ketones, Carboxylic Esters, and Carboxamides Related to Changes in Ground State Energies. *J. Am. Chem. Soc.* **117**, 7092–7096 (1995).

61. Q. Zhang, W. A. van der Donk, W. Liu, Radical-Mediated Enzymatic Methylation: A Tale of Two SAMS. *Acc. Chem. Res.* **45**, 555–564 (2012).

62. M. R. Bauerle, E. L. Schwalm, S. J. Booker, Mechanistic Diversity of Radical *S* -Adenosylmethionine (SAM)-dependent Methylation. *J. Biol. Chem.* **290**, 3995–4002 (2015).

63. Y. Hayakawa *et al.*, Thioviridamide, a Novel Apoptosis Inducer in Transformed Cells from Streptomyces olivoviridis. *J. Antibiot. (Tokyo)*. **59**, 1–5 (2006).

64. Y. Hayakawa, K. Sasaki, K. Nagai, K. Shin-ya, K. Furihata, Structure of Thioviridamide, a Novel Apoptosis Inducer from Streptomyces olivoviridis. *J. Antibiot. (Tokyo)*. **59**, 6–10 (2006).

65. M. Izawa, T. Kawasaki, Y. Hayakawa, Cloning and Heterologous Expression of the Thioviridamide Biosynthesis Gene Cluster from Streptomyces olivoviridis. *Appl. Environ. Microbiol.* **79**, 7110–7113 (2013).

66. K. L. Dunbar, J. O. Melby, D. A. Mitchell, YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat. Chem. Biol.* **8**, 569–575 (2012).

67. K. L. Dunbar *et al.*, Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nat. Chem. Biol.* **10**, 823–829 (2014).

68. B. Breil, J. Borneman, E. W. Triplett, A newly discovered gene, tfuA, involved in the production of the ribosomally synthesized peptide antibiotic trifolitoxin. *J. Bacteriol.* **178**, 4150–4156 (1996).

69. C. T. Jurgenson, T. P. Begley, S. E. Ealick, The Structural and Biochemical Foundations of Thiamin Biosynthesis. *Annu. Rev. Biochem.* **78**, 569–603 (2009).

70. P. M. Palenchar, C. J. Buck, H. Cheng, T. J. Larson, E. G. Mueller, Evidence That ThiI, an Enzyme Shared between Thiamin and 4-Thiouridine Biosynthesis, May Be a Sulfurtransferase That Proceeds through a Persulfide Intermediate. *J. Biol. Chem.* **275**, 8283–8286 (2000).

71. Y. Liu *et al.*, Biosynthesis of 4-thiouridine in tRNA in the methanogenic archaeon Methanococcus maripaludis. *J. Biol. Chem.* **287**, 36683–36692 (2012).

72. S. E. Hook, A.-D. G. Wright, B. W. McBride, Methanogens: Methane Producers of the Rumen and Mitigation Strategies. *Archaea*. **2010**, 1–11 (2010).

73. H. Tian *et al.*, The terrestrial biosphere as a net source of greenhouse gases to the atmosphere. *Nature*. **531**, 225–228 (2016).

74. D. Ankel-Fuchs, R. Hüster, E. Mörschel, S. P. J. Albracht, R. K. Thauer, Structure and function of methyl-coenzyme M reductase and of factor F430 in methanogenic bacteria. *Syst. Appl. Microbiol.* **7**, 383–387 (1986).

75. S. Scheller, M. Goenrich, R. K. Thauer, B. Jaun, Methyl-Coenzyme M Reductase from Methanogenic Archaea: Isotope Effects on Label Exchange and Ethane Formation with the Homologous Substrate Ethyl-Coenzyme M. *J. Am. Chem. Soc.* **135**, 14985–14995 (2013).

76. S. Scheller, M. Goenrich, R. K. Thauer, B. Jaun, Methyl-Coenzyme M Reductase from Methanogenic Archaea: Isotope Effects on the Formation and Anaerobic Oxidation of Methane. *J. Am. Chem. Soc.* **135**, 14975–14984 (2013).

77. R. Jasso-Chávez *et al.*, Air-Adapted Methanosarcina acetivorans Shows High Methane Production and Develops Resistance against Oxygen Stress. *PLOS ONE*. **10**, e0117331 (2015).

78. J. E. Galagan, The Genome of M. acetivorans Reveals Extensive Metabolic and Physiological Diversity. *Genome Res.* **12**, 532–542 (2002).

79. R. C. Mierendorf, B. B. Morris, B. Hammer, R. E. Novy, in *Molecular Diagnosis of Infectious Diseases* (Humana Press, New Jersey, 1997; http://link.springer.com/10.1385/0-89603-485-2:257), vol. 13, pp. 257–292.

80. J. Grodberg, J. J. Dunn, ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.* **170**, 1245–1253 (1988).

81. E. Raux *et al.*, Identification and functional analysis of enzymes required for precorrin-2 dehydrogenation and metal ion insertion in the biosynthesis of sirohaem and cobalamin in Bacillus megaterium. *Biochem. J.* **370**, 505–516 (2003).

82. H. Gilles, R. K. Thauer, Uroporphyrinogen III, an intermediate in the biosynthesis of the nickel-containing factor F430 in Methanobacterium thermoautotrophicum. *Eur. J. Biochem. FEBS*. **135**, 109–112 (1983).

83. M. Hansson, L. Rutberg, I. Schröder, L. Hederstedt, The Bacillus subtilis hemAXCDBL gene cluster, which encodes enzymes of the biosynthetic pathway from glutamate to uroporphyrinogen III. *J. Bacteriol.* **173**, 2590–2599 (1991).

84. M. J. Warren *et al.*, Enzymatic synthesis of dihydrosirohydrochlorin (precorrin-2) and of a novel pyrrocorphin by uroporphyrinogen III methylase. *FEBS Lett.* **261**, 76–80 (1990).

85. S. Bali *et al.*, Molecular hijacking of siroheme for the synthesis of heme and d1 heme. *Proc. Natl. Acad. Sci.* **108**, 18260–18265 (2011).

86. M. Fontecave, S. O. de Choudens, B. Py, F. Barras, Mechanisms of iron-sulfur cluster assembly: the SUF machinery. *J. Biol. Inorg. Chem. JBIC Publ. Soc. Biol. Inorg. Chem.* **10**, 713–721 (2005).

87. R. Lill *et al.*, Mechanisms of iron-sulfur protein maturation in mitochondria, cytosol and nucleus of eukaryotes. *Biochim. Biophys. Acta*. **1763**, 652–667 (2006).

88. J. R. Cupp-Vickery, J. C. Peterson, D. T. Ta, L. E. Vickery, Crystal structure of the molecular chaperone HscA substrate binding domain complexed with the IscU recognition peptide ELPPVKIHC. *J. Mol. Biol.* **342**, 1265–1278 (2004).

89. J. R. Cupp-Vickery, L. E. Vickery, Crystal Structure of Hsc20, a J-type Co-chaperone from Escherichia coli. *J. Mol. Biol.* **304**, 835–845 (2000).

90. J. Frazzon, Formation of iron–sulfur clusters in bacteria: an emerging field in bioinorganic chemistry. *Curr. Opin. Chem. Biol.* **7**, 166–173 (2003).

91. L. Zheng, V. L. Cash, D. H. Flint, D. R. Dean, Assembly of Iron-Sulfur Clusters: IDENTIFICATION OF AN *iscSUA-hscBA-fdx* GENE CLUSTER FROM *AZOTOBACTER VINELANDII*. *J. Biol. Chem.* **273**, 13264–13272 (1998).

92. J. R. Cupp-Vickery, H. Urbina, L. E. Vickery, Crystal structure of IscS, a cysteine desulfurase from Escherichia coli. *J. Mol. Biol.* **330**, 1049–1059 (2003).

93. Y. Kakuta, T. Horio, Y. Takahashi, K. Fukuyama, Crystal Structure of *Escherichia coli* Fdx, an Adrenodoxin-Type Ferredoxin Involved in the Assembly of Iron−Sulfur Clusters †. *Biochemistry (Mosc.)*. **40**, 11007–11012 (2001).

94. J. N. Agar *et al.*, IscU as a scaffold for iron-sulfur cluster biosynthesis: sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU. *Biochemistry (Mosc.)*. **39**, 7856–7862 (2000).

95. S. Ollagnier-de-Choudens, T. Mattioli, Y. Takahashi, M. Fontecave, Iron-sulfur cluster assembly: characterization of IscA and evidence for a specific and functional complex with ferredoxin. *J. Biol. Chem.* **276**, 22604–22607 (2001).

96. Y. Takahashi, M. Nakamura, Functional Assignment of the ORF2-iscS-iscU-iscA-hscB-hscA-fdx-0RF3 Gene Cluster Involved in the Assembly of Fe-S Clusters in Escherichia coli. *J. Biochem. (Tokyo)*. **126**, 917–926 (1999).

97. J. G. Ferry, Methane: small molecule, big impact. *Science*. **278**, 1413–1414 (1997).

98. J. G. Ferry, Enzymology of one-carbon metabolism in methanogenic pathways. *FEMS Microbiol. Rev.* **23**, 13–38 (1999).

99. M. Goubeaud, G. Schreiner, R. K. Thauer, Purified methyl-coenzyme-M reductase is activated when the enzyme-bound coenzyme F430 is reduced to the nickel(I) oxidation state by titanium(III) citrate. *Eur. J. Biochem.* **243**, 110–114 (1997).

100. A. K. Jain, B. P. Briegleb, K. Minschwaner, D. J. Wuebbles, Radiative forcings and global warming potentials of 39 greenhouse gases. *J. Geophys. Res. Atmospheres*. **105**, 20773–20790 (2000).

101. K. Zheng, P. D. Ngo, V. L. Owens, X. Yang, S. O. Mansoorabadi, The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science*. **354**, 339–342 (2016).

102. M. Keb-Llanes, G. González, B. Chi-Manzanero, D. Infante, A rapid and simple method for small-scale DNA extraction inAgavaceae and other tropical plants. *Plant Mol. Biol. Report.* **20**, 299–299 (2002).

103. J. Kahnt *et al.*, Post-translational modifications in the active site region of methyl-coenzyme M reductase from methanogenic and methanotrophic archaea: Methanogenic archaea methyl-coenzyme M reductase. *FEBS J.* **274**, 4913–4921 (2007).

104. B. J. Burkhart, C. J. Schwalen, G. Mann, J. H. Naismith, D. A. Mitchell, YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. *Chem. Rev.* **117**, 5389–5456 (2017).

105. B. J. Rauch, J. Klimek, L. David, J. J. Perona, Persulfide Formation Mediates Cysteine and Homocysteine Biosynthesis in *Methanosarcina acetivorans*. *Biochemistry (Mosc.)*. **56**, 1051–1061 (2017).

106. M. A. Stamnes, S. L. Rutherford, C. S. Zuker, Cyclophilins: a new family of proteins involved in intracellular folding. *Trends Cell Biol.* **2**, 272–276 (1992).

107. G. Fischer, F. X. Schmid, The mechanism of protein folding. Implications of in vitro refolding models for de novo protein folding and translocation in the cell. *Biochemistry (Mosc.)*. **29**, 2205–2212 (1990).

108. R. T. Hay, SUMO. *Mol. Cell*. **18**, 1–12 (2005).

109. Z. Wang *et al.*, Human SUMO fusion systems enhance protein expression and solubility. *Protein Expr. Purif.* **73**, 203–208 (2010).

110. D. Deobald, L. Adrian, C. Schöne, M. Rother, G. Layer, Identification of a unique Radical SAM methyltransferase required for the sp3-C-methylation of an arginine residue of methyl-coenzyme M reductase. *Sci. Rep.* **8** (2018), doi:10.1038/s41598-018-25716-x.

111. V. W. C. Soo *et al.*, Reversing methanogenesis to capture methane for liquid biofuel precursors. *Microb. Cell Factories*. **15** (2016), doi:10.1186/s12934-015-0397-z.

112. T. J. Mueller *et al.*, Methane oxidation by anaerobic archaea for conversion to liquid fuels. *J. Ind. Microbiol. Biotechnol.* **42**, 391–401 (2015).

113. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*. **6**, 343–345 (2009).

Additional Primary Efforts

A version of this manuscript was published in a handbook associated with the *Encyclopedia of Inorganic and Bioinorganic Chemistry*.

S. O. Mansoorabadi, K. Zheng, P. D. Ngo, Biosynthesis of Coenzyme F430 and the Posttranslational Modification of the Active Site Region of Methyl-Coenzyme M Reductase, in *Metalloprotein Active Site Assembly* (Johnson, M. K. and Scott, R. A., Eds.), John Wiley & Sons, Ltd, West Sussex, UK.

These findings expose a clear demarcation between the realms of chance and necessity at different hierarchical levels. At the level of biochemical phenotype, and even at the level of functional mechanism, evolutionary changes are highly predictable. At the amino acid level, in contrast, predictability breaks down.

In addition to the many-to-one mapping of genotype to phenotype, the phylogenetic distribution of affinity-enhancing parallel substitutions suggests another possible explanation for the limited contribution of such substitutions to convergent functional changes in the Hbs of distantly related species. The most striking functional parallelism at the amino acid level was concentrated in the hummingbird clade. Replicated G83S substitutions contributed to convergent increases in Hb-O_2 affinity in multiple high-altitude hummingbird species (table S5 and fig. S4) (*16*), and a convergent substitution at the same site (N83S) occurred in one other (nonhummingbird) high-altitude species: the black-throated flowerpiercer, *Diglossa brunneiventris*. One possible explanation for this phylogenetically concentrated pattern of parallelism is that the mutation's phenotypic effect is conditional on genetic background, so the same mutation produces different effects in different species.

To test this hypothesis, we used ancestral sequence reconstruction in combination with site-directed mutagenesis to test the effect of β83 substitutions in a set of distinct genetic backgrounds. We first resurrected HbA of the common ancestor of hummingbirds ("Anc hummingbird") (figs. S5 to S7), and we confirmed that G83S has a significant affinity-enhancing effect on this ancestral genetic background (Fig. 3A). This result is consistent with the affinity-enhancing effect of G83S in numerous descendant lineages of high-altitude hummingbirds (table S5 and fig. S4). In similar fashion, we resurrected HbA of the common ancestor of the high- and low-altitude flowerpiercers ("Anc flowerpiercer") to test the effect of N83S (fig. S7). Hbs of the two flowerpiercers differed at two sites because of substitutions in the *D. brunneiventris* lineage (V67A in α^A-globin, in addition to N83S in β^A-globin; Fig. 1). We therefore synthesized a total of four recombinant Hb mutants, representing each possible genotypic combination of the two substituted sites, to measure the relative contributions of V67A and N83S to the evolved increase in Hb-O_2 affinity in *D. brunneiventris* (table S2 and fig. S4). The tests showed that both mutations increased Hb-O_2 affinity in an additive fashion (Fig. 3B). We then engineered the same N83S mutation into resurrected ancestral Hbs representing two far more ancient nodes in the avian phylogeny: the reconstructed common ancestor of Neoaves ("Anc Neoaves") and the common ancestor of all extant birds ("Anc Neornithes") (Fig. 3C and figs. S5, S7, S8, and S9). In contrast to the highly significant effects of N/G83S in hummingbird and flowerpiercer Hbs, N83S produced no detectable effect in Anc Neoaves or Anc Neornithes (Fig. 3D and table S6). The ancestral hummingbird and flowerpiercer Hbs contained 18 and 32 amino acid states, respectively, that were not present in Anc Neornithes (fig. S7), representing net sequence differences that accumulated over a ~100-million-year time period. The context-dependent effects of N/G83S indicate that lineage-specific substitutions in the ancestry of hummingbirds and flowerpiercers produced a genetic background in which mutations at β83 could contribute to an adaptive increase in Hb-O_2 affinity. This adaptive solution was apparently not an option in the deeper ancestry of birds and may also represent a precluded possibility in contemporary, high-altitude members of other avian lineages.

These findings reveal a potentially important role of contingency in adaptive protein evolution. In different species that are adapting to the same selection pressure, the set of possible amino acids at a given site that have unconditionally beneficial effects may be contingent on the set of antecedent substitutions that have independently accumulated in the history of each lineage. Consequently, possible options for adaptive change in one species may be foreclosed options in other species.

REFERENCES AND NOTES

1. D. L. Stern, V. Orgogozo, *Science* **323**, 746–751 (2009).
2. J. B. Losos, *Evolution* **65**, 1827–1840 (2011).
3. D. L. Stern, *Nat. Rev. Genet.* **14**, 751–764 (2013).
4. J. F. Storz, *Nat. Rev. Genet.* **17**, 239–250 (2016).
5. R. E. Weber, *Respir. Physiol. Neurobiol.* **158**, 132–142 (2007).
6. J. F. Storz, G. R. Scott, Z. A. Cheviron, *J. Exp. Biol.* **213**, 4125–4136 (2010).
7. M. F. Perutz, *Mol. Biol. Evol.* **1**, 1–28 (1983).
8. Materials and methods are available as supplementary materials on *Science* Online.
9. M. T. Grispo *et al.*, *J. Biol. Chem.* **287**, 37647–37658 (2012).
10. Z. A. Cheviron *et al.*, *Mol. Biol. Evol.* **31**, 2948–2962 (2014).
11. S. C. Galen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13958–13963 (2015).
12. C. Natarajan *et al.*, *PLOS Genet.* **11**, e1005681 (2015).
13. J. C. Opazo *et al.*, *Mol. Biol. Evol.* **32**, 871–887 (2015).
14. J. F. Storz, A. M. Runck, H. Moriyama, R. E. Weber, A. Fago, *J. Exp. Biol.* **213**, 2565–2574 (2010).
15. C. Natarajan *et al.*, *Mol. Biol. Evol.* **32**, 978–997 (2015).
16. J. Projecto-Garcia *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20669–20674 (2013).

ACKNOWLEDGMENTS

This work was funded by grants from the U.S. NIH (HL087216), the U.S. NSF (IOS-0949931, MCB-1517636, and MCB-1516660), and the Danish Council for Independent Research (10-084-565 and 4181-00094). We thank E. Petersen, H. Moriyama, and A. Kumar for assistance in the laboratory and C. Meiklejohn and K. Montooth for helpful suggestions. All experimental data are tabulated in the supplementary materials, and sequence data are archived in GenBank under accession numbers KX240692 to KX241466.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6310/336/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S6
References (*17*–*33*)

19 April 2016; accepted 20 July 2016
10.1126/science.aaf9070

ENZYMOLOGY

The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea

Kaiyuan Zheng, Phong D. Ngo, Victoria L. Owens, Xue-peng Yang,* Steven O. Mansoorabadi†

Methyl-coenzyme M reductase (MCR) is the key enzyme of methanogenesis and anaerobic methane oxidation. The activity of MCR is dependent on the unique nickel-containing tetrapyrrole known as coenzyme F430. We used comparative genomics to identify the coenzyme F430 biosynthesis (*cfb*) genes and characterized the encoded enzymes from *Methanosarcina acetivorans* C2A. The pathway involves nickelochelation by a nickel-specific chelatase, followed by amidation to form Ni-sirohydrochlorin *a*,*c*-diamide. Next, a primitive homolog of nitrogenase mediates a six-electron reduction and γ-lactamization reaction before a Mur ligase homolog forms the six-membered carbocyclic ring in the final step of the pathway. These data show that coenzyme F430 can be synthesized from sirohydrochlorin using Cfb enzymes produced heterologously in a nonmethanogen host and identify several targets for inhibitors of biological methane formation.

Methanogenic archaea are a major player in the global carbon cycle, producing nearly 1 billion metric tons of methane annually (*1*, *2*). The terminal step of methanogenesis is catalyzed by methyl-coenzyme M reductase (MCR) and involves the conversion of coenzyme B (CoB-SH) and methyl-coenzyme M (MeS-CoM) to the mixed heterodisulfide CoB-S-S-CoM and methane (*3*) (Fig. 1). MCR uses the unique nickel-containing tetrapyrrole coenzyme F430 to carry out its catalytic function (*4*) (Fig. 1). Recently, anaerobic methanotrophic archaea (ANME) have been shown

Department of Chemistry and Biochemistry, Auburn University, Auburn, AL 36849, USA.
*Present address: School of Food and Biological Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China.
†Corresponding author. Email: som@auburn.edu

to contain a homolog of MCR and catalyze the anaerobic oxidation of methane (AOM) (*5*). AOM is thought to operate, at least in part, as the reverse of methanogenesis, with MCR catalyzing the critical first step in the pathway, the activation of methane with CoB-S-S-CoM (*6*).

There is great interest in strategies to convert methane to liquid fuel or other more easily transported commodity chemicals. The development of a bioconversion process for methane that uses AOM is an attractive solution (*7*); however, efforts to engineer industrially viable anaerobic methanotrophic strains are hindered by the lack of genetic and biochemical information about the biosynthesis of coenzyme F430 and the formation of holo-MCR.

At present, the genomes of >60 species of methanogenic archaea have been sequenced. We searched these genomes for homologs of known chelatase genes, whose products are responsible for metal ion insertion into tetrapyrrolic cofactors (*8*). Analysis of the genomic contexts of the chelatase homologs, along with knowledge of the chemistry required for the conversion of known precursors of C2 and C7 methylated tetrapyrroles to coenzyme F430 (*9*), led to the identification of five genes, conserved in all methanogens, that are potentially involved in coenzyme F430 biosynthesis (Fig. 1 and fig. S1). These genes are also present in the genome of an ANME-2d strain (*Candidatus Methanoperedens nitroreducens*) (fig. S1) (*10*).

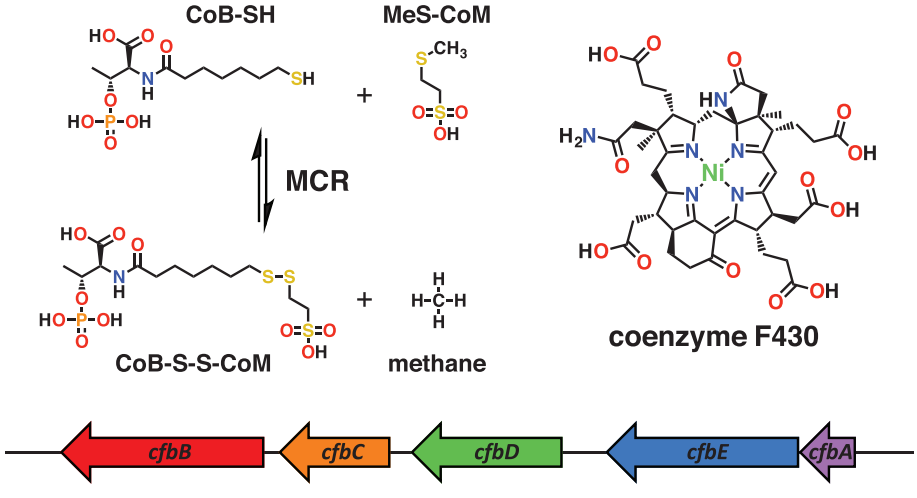
Included among these genes are homologs of the genes for sirohydrochlorin cobaltochelatase (*cbiX*^{S}) and cobyrinic acid *a*,*c*-diamide synthetase (*cbiA2*), which are involved in the biosynthesis of cobalamin (*11*). Also present are homologs of the nitrogenase genes *nifD* and *nifH*, which (together with *nifK*) encode subunits of the two-component metalloenzyme responsible for the adenosine triphosphate (ATP)–dependent reduction of dinitrogen to ammonia (nitrogen fixation) (*12*). Nitrogenase is structurally and functionally related to the dark-operative protochlorophyllide oxidoreductase (DPOR), which is involved in chlorophyll and bacteriochlorophyll biosynthesis (*13*). However, methanogens are not photosynthetic microorganisms, and not all methanogens are diazotrophic (i.e., fix nitrogen). The presence of the *nifD* (methanogenesis marker 13) and *nifH* homologs in all methanogens was noted previously, and these genes were found to be constitutively expressed and the encoded proteins shown to associate with one another (*14*, *15*). Methanogenesis markers are found in prokaryotic genomes if, and only if, the species is an archaeal methanogen. The fifth gene is homologous to *murD*, a gene that encodes an ATP-dependent Mur ligase (uridine diphosphate *N*-acetylmuramoyl-L-alanine:D-glutamate ligase) involved in bacterial cell wall biosynthesis (*16*). Each of these genes (except for *cbiA2*) was targeted in a genome-wide transposon mutagenesis experiment in the methanogen *Methanococcus maripaludis* and were all found to be essential (*17*).

We cloned the identified genes from *Methanosarcina acetivorans* C2A and ligated them into expression vectors for heterologous production of the encoded enzymes in *Escherichia coli*. The enzymes were then purified as N-terminal His_6-tagged fusion proteins and systematically tested for activity (supplementary materials and fig. S2).

We tested the "small" sirohydrochlorin cobaltochelatase (CbiX^S) homolog, which we designate as CfbA, for nickelochelatase activity with enzymatically prepared sirohydrochlorin and dihydrosirohydrochlorin (precorrin 2) (*18*). Precorrin 2 is the immediate biosynthetic precursor of sirohydrochlorin and is two-electrons more reduced than the latter (*19*). Because coenzyme F430 is a highly reduced tetrapyrrole, we reasoned that precorrin 2 might be the substrate of CfbA. However, no nickel chelation activity was observed with precorrin 2 under any of the assay conditions tested. We therefore turned our attention to sirohydrochlorin. We observed that, unlike other divalent transition metal ions (e.g., Fe^{2+}, Co^{2+}, and Zn^{2+}), there is no evidence of rapid, nonenzymatic insertion of Ni^{2+} into sirohydrochlorin under our assay conditions. However, in the presence of both Ni^{2+} and CfbA (and only if the His_6-tag of CfbA was first removed by thrombin cleavage), the reaction mixture changed from the bright magenta color characteristic of sirohydrochlorin to a deep purple (Fig. 2 and fig. S3). Analysis of the reaction mixtures by reversed-phase high-performance liquid chromatography (HPLC) showed the near-complete conversion of sirohydrochlorin (which has a retention time of 16.1 min) to a new compound that elutes at 20.2 min (Fig. 2). The ultraviolet (UV)–visible absorption properties [wavelengths of maximum light absorption (λ_{max}) = 386 and 590 nm] and mass spectrum of this compound are consistent with those of Ni-sirohydrochlorin [calculated mass-to-charge ratio (*m*/*z*) of the protonated molecule ($[M + H]^+_{calc}$) = 919.22 *m*/*z*] (*20*). Thus, CfbA is a sirohydrochlorin nickelochelatase.

The addition of the cobyrinic acid *a*,*c*-diamide synthetase homolog (CfbB) to the reaction mixture enhanced the yield of Ni-sirohydrochlorin (fig. S4). Intermediates in tetrapyrrole biosynthesis often remain tightly bound to their cognate enzyme and are thought to be transferred to the next enzyme in the pathway by substrate channeling (*21*). Therefore, tetrapyrrole biosynthetic enzymes often exhibit substantial product inhibition, which can be relieved by the addition of the subsequent pathway enzyme in vitro. This suggests that CfbB acts next in the pathway and will accept Ni-sirohydrochlorin as a substrate.

CbiA is a glutamine amidotransferase that catalyzes the ATP-dependent amidation of the *a*- and *c*-carboxylic acid moieties of cobyrinic acid in the cobalamin biosynthetic pathway (*22*). Coenzyme F430 also has amide functional groups at these positions, and we reasoned that CfbB was a Ni-sirohydrochlorin *a*,*c*-diamide synthetase. Indeed, addition of both ATP and glutamine to the reaction mixture led to the formation of a new intermediate with a nearly identical



gene	locus	annotation
cfbA	MA3631	sirohydrochlorin cobaltochelatase (*cbiX*^{S})
cfbB	MA3626	cobyrinic acid *a*,*c*-diamide synthetase (*cbiA2*)
cfbC	MA3627	nitrogenase iron protein (*nifH*)
cfbD	MA3628	nitrogenase molybdenum-iron protein (*nifD*)
cfbE	MA3630	UDP-*N*-acetylmuramoylalanine:D-glutamate ligase (*murD*)

Fig. 1. The MCR-catalyzed reaction, the structure of coenzyme F430, and the identified coenzyme F430 biosynthesis (*cfb*) gene cluster from *M. acetivorans* C2A. Colored arrows indicate the relative size and orientation of each *cfb* gene.

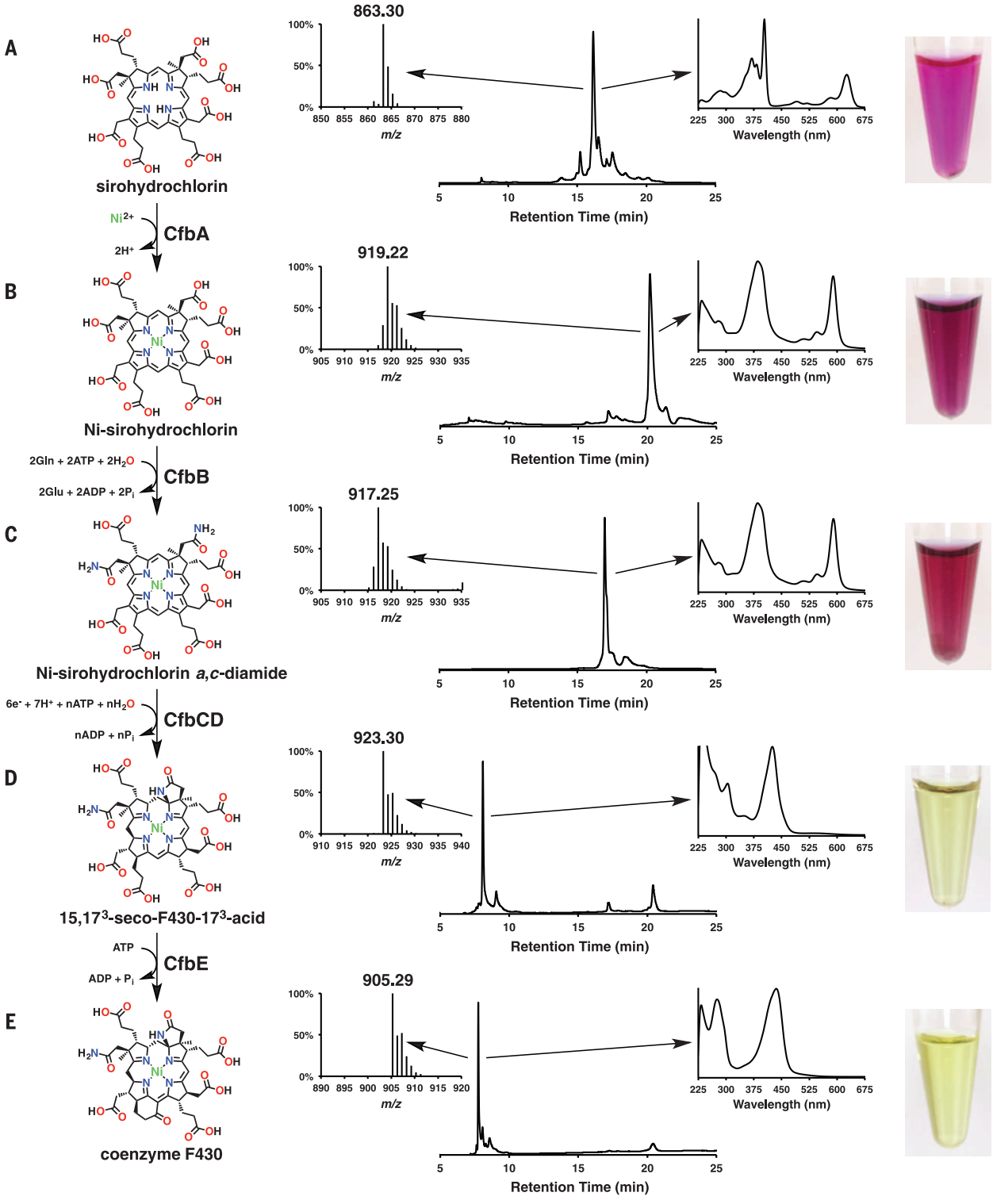


Fig. 2. In vitro activity assays of the coenzyme F430 biosynthesis enzymes. Reversed-phase HPLC traces, liquid chromatography–mass spectrometry (LC-MS) data, and UV-visible spectra for each of the biosynthetic reactions are shown. (**A**) Sirohydrochlorin prepared from porphobilinogen by using HemC, HemD, SirA, SirC, *S*-adenosyl-L-methionine, and nicotinamide adenine dinucleotide (phosphate). (**B**) Ni-sirohydrochlorin prepared by adding CfbA and $NiCl_2$ to the sirohydrochlorin reaction (along with CfbB to alleviate product inhibition). (**C**) Ni-sirohydrochlorin *a,c*-diamide prepared by adding CfbB, glutamine, ATP, and an ATP regeneration system (PEP and PK) to the Ni-sirohydrochlorin reaction (along with CfbCD to alleviate product inhibition). (**D**) 15,17³-seco-F430-17³-acid prepared by adding CfbCD, sodium dithionite, ATP, and an ATP regeneration system (PEP and PK) to the Ni-sirohydrochlorin *a,c*-diamide reaction. (**E**) Coenzyme F430 prepared by adding CfbE, ATP, and an ATP regeneration system (PEP and PK) to the 15,17³-seco-F430-17³-acid reaction (along with McrD to alleviate product inhibition). Further experimental details can be found in the supplementary materials. ADP, adenosine diphosphate; P_i, inorganic phosphate.

UV-visible spectrum to that of Ni-sirohydrochlorin (λ_{max} = 386 and 590 nm)—although its HPLC retention time was shorter by 3 min (17.0 min) and its observed mass was lighter by 1.97 atomic mass units, with a m/z identical to the $[M + H]^+_{calc}$ for Ni-sirohydrochlorin *a,c*-diamide (Fig. 2 and fig. S5). The yield of Ni-sirohydrochlorin *a,c*-diamide could again be enhanced by the addition of the subsequent enzyme in the pathway (CfbCD, a complex of CfbC and CfbD), without the reductant required for its activity (vide infra), to alleviate product inhibition (fig. S6). The inclusion of an ATP regeneration system [phosphoenolpyruvate (PEP) and pyruvate kinase (PK)] also helped to drive the CfbB reaction forward (fig. S6).

Two distinctive structural features of coenzyme F430 are the presence of the γ-lactam E ring and the carbocyclic F ring, which form from the *c*-acetamide and *g*-propionate side chains of Ni-sirohydrochlorin *a,c*-diamide, respectively. The high degree of similarity between the UV-visible spectra of Ni-sirohydrochlorin and Ni-sirohydrochlorin *a,c*-diamide indicates that the product of the CfbB reaction lacks the γ-lactam ring and contains the free amide (Fig. 2).

When the purified CfbCD complex was included in the CfbB reaction, along with the reductant sodium dithionite, ATP, and an ATP regeneration system (all of which were required for activity), the solution changed from the deep purple color characteristic of Ni-sirohydrochlorin *a,c*-diamide to a pale yellow. Analysis of the reaction mixture by HPLC showed the disappearance of the 17.0-min peak and the formation of a new peak with a retention time of 8.1 min (Fig. 2 and fig. S7). The UV-visible and mass spectra of this new intermediate are indistinguishable from those of the only previously identified intermediate that is unique to the coenzyme F430 biosynthetic pathway, 15,17^3-seco-F430-17^3-acid ($[M]^+_{calc}$ = 923.30 m/z) (*23*) (Fig. 2). This intermediate was identical in structure to coenzyme F430, except for the presence of the *g*-propionate side chain instead of the F ring. Thus, CfbCD effects both the six-electron reduction of the tetrahydroporphyrin ring system of Ni-sirohydrochlorin *a,c*-diamide and the γ-lactamization of its *c*-acetamide side chain to form the E ring.

As noted above, CfbC is homologous to the Fe protein (NifH) and CfbD to the NifD subunit of the MoFe protein (NifDK) of nitrogenase, which catalyzes an eight-electron reduction of N_2 to NH_3 and two protons to H_2 (*12*). The *nifD* and *nifK* genes are proposed to have arisen from the paralogous gene duplication and divergence of an ancient shared precursor (*24*). The nitrogenase homolog DPOR is an analogous two-component system (BchH and BchNB) with similar structural topology and catalyzes a two-electron reduction of the C17=C18 double bond of protochlorophyllide to form chlorophyllide *a* in the chlorophyll biosynthetic pathway (*13*). Unlike nitrogenase and DPOR, the Ni-sirohydrochlorin *a,c*-diamide reductive cyclase contains a homomeric MoFe protein homolog (CfbD) and is thus representative of an early-branching lineage of this enzyme family. A study of the molecular phylogeny of nitrogenase homologs placed the ancestral *cfbC* and *cfbD* genes in the last common ancestor of modern organisms and positioned them basal to the emergence of the groups involved in nitrogen fixation and the biosynthesis of photosynthetic pigments (*14*).

The last enzyme encoded by the *cfb* cluster, CfbE, is homologous to an ATP-dependent Mur ligase. Mur ligases use ATP to activate a carboxylic acid group as an acyl-phosphate for nonribosomal peptide bond formation during the biosynthesis of peptidoglycan (*25*). We reasoned that CfbE could use similar chemistry to activate the *g*-propionate side chain for intramolecular C–C bond formation to produce the carbocyclic F ring and thus function as a coenzyme F430 synthetase. As expected, addition of CfbE to reaction mixtures containing 15,17^3-seco-F430-17^3-acid resulted in the production of a new compound in low yield, which had an identical HPLC retention time (7.6 min), UV-visible spectrum, and isotopic mass distribution to authentic coenzyme F430 (fig. S8).

We hypothesized that the low yield of coenzyme F430 was due to product inhibition of the CfbE reaction. The *mcr* gene cluster encoding the α, β, and γ subunits of MCR, which has been identified previously and is distinct from the *cfb* cluster, contains two genes, *mcrC* and *mcrD*, of unknown function (*26*). Recently, McrC was identified as a component of a large reductase complex capable of reducing coenzyme F430 to the Ni^{1+} form, and thus it may play a role in MCR activation (*27*). McrD has been shown to physically interact with MCR through coprecipitation experiments, though it is not required for in vitro MCR activity (*28*). We postulated that McrD may function as a chaperone protein that could bind coenzyme F430 and deliver it to apo-MCR. We therefore cloned *mcrD* from *M. acetivorans* C2A and expressed and purified the encoded protein to determine whether it was capable of accepting coenzyme F430 from CfbE and alleviating the observed inhibition. Consistent with this expectation, nearly full conversion of 15,17^3-seco-F430-17^3-acid to coenzyme F430 was observed when McrD was included in the reaction mixtures (Fig. 2 and figs. S9 and S10).

Each of the identified coenzyme F430 biosynthetic enzymes represents a new target for inhibitors of methanogenesis. The data show that these enzymes are sufficient for the synthesis of coenzyme F430 from the common tetrapyrrolic intermediate sirohydrochlorin and can be produced in an active form in *E. coli*. Furthermore, if McrD is confirmed as a coenzyme F430–binding protein that chaperones the coenzyme to MCR, this protein will also be required for the heterologous production of holo-MCR. Taken together, these findings set the stage for metabolic engineering efforts using MCR for anaerobic methane conversion.

REFERENCES AND NOTES

1. R. K. Thauer, A. K. Kaster, H. Seedorf, W. Buckel, R. Hedderich, *Nat. Rev. Microbiol.* **6**, 579–591 (2008).
2. J. G. Ferry, *Science* **278**, 1413–1414 (1997).
3. T. Wongnate *et al.*, *Science* **352**, 953–958 (2016).
4. G. Diekert, R. Jaenchen, R. K. Thauer, *FEBS Lett.* **119**, 118–120 (1980).
5. K. Knittel, A. Boetius, *Annu. Rev. Microbiol.* **63**, 311–334 (2009).
6. S. Scheller, M. Goenrich, R. Boecher, R. K. Thauer, B. Jaun, *Nature* **465**, 606–608 (2010).
7. T. J. Mueller *et al.*, *J. Ind. Microbiol. Biotechnol.* **42**, 391–401 (2015).
8. H. L. Schubert, E. Raux, K. S. Wilson, M. J. Warren, *Biochemistry* **38**, 10660–10669 (1999).
9. H. Mucha, E. Keller, H. Weber, F. Lingens, W. Trösch, *FEBS Lett.* **190**, 169–171 (1985).
10. M. F. Haroon *et al.*, *Nature* **500**, 567–570 (2013).
11. J. R. Roth, J. G. Lawrence, M. Rubenfield, S. Kieffer-Higgins, G. M. Church, *J. Bacteriol.* **175**, 3303–3316 (1993).
12. H. Schindelin, C. Kisker, J. L. Schlessman, J. B. Howard, D. C. Rees, *Nature* **387**, 370–376 (1997).
13. N. Muraki *et al.*, *Nature* **465**, 110–114 (2010).
14. J. Raymond, J. L. Siefert, C. R. Staples, R. E. Blankenship, *Mol. Biol. Evol.* **21**, 541–554 (2004).
15. C. R. Staples *et al.*, *J. Bacteriol.* **189**, 7392–7398 (2007).
16. F. Pratviel-Sosa, D. Mengin-Lecreulx, J. van Heijenoort, *Eur. J. Biochem.* **202**, 1169–1176 (1991).
17. F. Sarmiento, J. Mrázek, W. B. Whitman, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4726–4731 (2013).
18. A. G. Smith, M. Witty, *Heme, Chlorophyll, and Bilins: Methods and Protocols* (Humana Press, 2002).
19. M. J. Warren *et al.*, *FEBS Lett.* **261**, 76–80 (1990).
20. H. K. Leech, E. Raux-Deery, P. Heathcote, M. J. Warren, *Biochem. Soc. Trans.* **30**, 610–613 (2002).
21. E. Deery *et al.*, *Nat. Chem. Biol.* **8**, 933–940 (2012).
22. V. Fresquet, L. Williams, F. M. Raushel, *Biochemistry* **43**, 10619–10627 (2004).
23. A. Pfaltz, A. Kobelt, R. Hüster, R. K. Thauer, *Eur. J. Biochem.* **170**, 459–467 (1987).
24. R. Fani, R. Gallo, P. Liò, *J. Mol. Evol.* **51**, 1–11 (2000).
25. I. Kouidmi, R. C. Levesque, C. Paradis-Bleau, *Mol. Microbiol.* **94**, 242–253 (2014).
26. M. Bokranz, A. Klein, *Nucleic Acids Res.* **15**, 4350–4351 (1987).
27. D. Prakash, Y. Wu, S.-J. Suh, E. C. Duin, *J. Bacteriol.* **196**, 2491–2498 (2014).
28. B. A. Sherf, J. N. Reeve, *J. Bacteriol.* **172**, 1828–1833 (1990).

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy Advanced Research Projects Agency–Energy (ARPA-E) (grants DE-AR0000428 and DE-AR0000433). S.O.M., P.D.N., and K.Z. are inventors on patent application 62332658 submitted by Auburn University, which covers the synthesis of coenzyme F430 using the *cfb* gene products for the production of holo MCR. The data reported in this paper are available in the supplementary materials. S.O.M. provided the scientific direction and overall experimental design for the studies. K.Z. cloned the *sirC* gene from *M. acetivorans*, expressed and purified the coenzyme F430 biosynthetic enzymes, and performed the in vitro enzyme assays. P.D.N. cloned the *mcrD* gene from *M. acetivorans*, expressed and purified the encoded enzyme, and aided in the initial HPLC and LC-MS experiments. V.L.O. cloned the *hemC* and *hemD* genes from *E. coli*, expressed and purified the encoded enzymes, and aided in the initial synthesis of sirohydrochlorin. X.-p.Y. cloned the *sirA* and *cfbA*, *-B*, *-C*, *-D*, and *-E* genes; developed the initial purification procedures for the encoded enzymes; and confirmed the activity of SirA. S.O.M., K.Z., P.D.N., and V.L.O. wrote and edited the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6310/339/suppl/DC1
Materials and Methods
Figs. S1 to S10
Table S1

5 June 2016; accepted 7 September 2016
10.1126/science.aag2947

Editor's Summary

Enzymes for making (or breaking) methane

The last enzymatic step of microbial methanogenesis, and the first step of microbial methane oxidation, relies on the nickel-containing tetrapyrrole coenzyme F430. The successful metabolic engineering of any organism to enzymatically consume methane thus also needs the appropriate machinery to synthesize this compound. Using comparative genomics, Zheng *et al.* identified several candidate genes responsible for coenzyme F430 biosynthesis. Cloning and expression of all the subsequent proteins in *Escherichia coli* confirmed the complete in vitro conversion of sirohydrochlorin into mature F430.

Science, this issue p. 339

www.sciencemag.org/content/354/6310/339/suppl/DC1

Supplementary Materials for

The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea

Kaiyuan Zheng, Phong D. Ngo, Victoria L. Owens, Xue-peng Yang, Steven O. Mansoorabadi*

*Corresponding author. Email: som@auburn.edu

Published 21 October 2016, *Science* **354**, 339 (2016)
DOI: 10.1126/science.aag2947

This PDF file includes:

Materials and Methods

Plasmid construction

The *hemCD* genes were amplified by polymerase chain reaction (PCR) from the genomic DNA of *Escherichia coli* BL21 (DE3) (New England Biolabs). The *sirAC, cfbABCDE, and mcrD* genes were amplified from the genomic DNA of *Methanosarcina acetivorans* C2A (DSM-2834). Primers were synthesized by Sigma-Aldrich and their sequences are provided in Table S1. Phusion High-Fidelity DNA Polymerase (New England Biolabs) was utilized for all PCR reactions in accordance with the manufacturer's protocol.

The aforementioned PCR products (with the exception of *cfbD* and the *cfbC* PCR product obtained with reverse primer 2) were digested using the appropriate restriction enzymes from New England Biolabs and cloned into the pET-28b(+) vector (Novagen) for heterologous expression in *E. coli*. Each of the recombinant proteins thus produced contained a thrombin-cleavable His_6-tag incorporated at the N-terminus for purification using immobilized metal ion affinity chromatography (IMAC). The *cfbD* and *cfbC* genes were ligated into the 1st and 2nd multiple cloning sites (MCSs) of pRSFDuet-1 (Novagen), respectively, without the incorporation of affinity tags. The *cfbC* gene (obtained with reverse primer 1) was then subcloned from pET-28b(+) into the 1st MCS of pCDFDuet-1 (Novagen) using the NcoI and HindIII restriction enzymes, allowing the coexpression of CfbC containing a cleavable N-terminal His_6-tag with untagged CfbD (and CfbC) for IMAC purification of the entire CfbCD complex. The sequences of the cloned genes within each of the constructed plasmids were verified by either the Genomics and Sequencing Laboratory (GSL) at Auburn University or Eurofins Scientific.

Protein expression and purification

The HemC, HemD, SirA, SirC, CfbA, CfbB, CfbE, and McrD proteins were prepared by transforming *E. coli* BL21 (DE3) with the appropriate pET-28b(+) vector and culturing the resulting cells in Luria-Bertani (LB) medium supplemented with 50 μg/mL kanamycin at 37 °C in an incubator shaker. The CfbCD complex was obtained by sequentially transforming *E. coli* BL21 (DE3) with the pRSFDuet-1 and pCDFDuet-1 vectors described above, along with the pDB1282 vector containing the iron-sulfur cluster (*isc*) biosynthetic gene cluster from *Azotobacter vinelandii* (a generous gift from Prof. Dennis R. Dean, Virginia Polytechnic Institute and State University). The resulting cells were then propagated in LB medium containing kanamycin (50 μg/mL), spectinomycin (25 μg/mL), and ampicillin (100 μg/mL) at 37 °C in an incubator shaker.

For the purification of HemC, the temperature of the culture was lowered to 15 °C after the culture reached an OD_{600} ~ 0.5. After incubating for an additional 1 h at 15 °C, the cells were induced with 40 μM isopropyl β-D-thiogalactoside (IPTG), supplemented with 10 μM 5-aminolevulinic acid (ALA) (Ark Pharm, Inc.), and incubated for an additional 8 h. The production of HemD was also induced with 40 μM IPTG once the culture reached an OD_{600} ~ 0.6, after which the culture was incubated for an additional 8 h at 18 °C. The production of SirA, SirC, CfbB, and CfbE was induced with 100 μM IPTG when each of the cultures reached an OD_{600} ~ 0.5. Similarly, CfbA was induced with 400 μM IPTG when the cultures reached an OD_{600} ~ 0.5. For the production of the CfbCD complex, the *isc* operon was first induced with 3.0 g/L L-(+)-arabinose and the culture was supplemented with 3.0 mM each of $FeSO_4$ and L-cysteine. After incubation for 3 h, expression of the CfbCD complex was induced with 300 μM IPTG. The cultures were then allowed to incubate with shaking for 12 h (at 18 °C for SirA, SirC,

CfbA, and the CfbCD complex, and 25 °C for CfbB and CfbE). For the production of McrD, cells were induced with 400 μM IPTG when the culture reached an OD_{600} ~ 0.6. The culture was then grown for an additional 16 h at 25 °C.

Cells from each of the cultures were harvested by centrifugation at 15,970 × g and 4 °C. The remaining steps of the purification for HemC and the CfbCD complex were carried in a Coy anaerobic chamber with degassed buffers, while those for the rest of the enzymes were carried out aerobically. Cells were resuspended in lysis buffer consisting of 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 5 mM imidazole, lysozyme (1 mg/ml), and Ameresco's Protease Inhibitor Cocktail. The cell suspension was then sonicated and centrifuged at 104,600 × g for 20 min at 4 °C. The supernatant was applied to a Bio-Rad Econo-Pac column packed with Profinity IMAC Ni-Charged Resin. The column was then washed with 50 mM sodium phosphate (pH 8.0) buffer containing 300 mM NaCl and 5 mM imidazole. All of the proteins (except for CfbA and CfbB) were then eluted with a 50 mM sodium phosphate (pH 8.0) buffer containing 300 mM NaCl and 500 mM imidazole. The columns containing CfbA and CfbB were washed with 100 mM Tris-HCl (pH 8.0) buffer and then thrombin (80 units/mL of IMAC resin) was applied to the columns to cleave off the N-terminal His_6-tags. The columns were capped at both ends and incubated at 25 °C with shaking for 16 h. CfbA and CfbB were then eluted from their respective columns with 100 mM Tris-HCl (pH 8.0) buffer and the eluates were applied to columns containing Benzamidine Sepharose 4 Fast Flow (GE Healthcare) to remove the thrombin. The buffers of all of the proteins were then exchanged with 100 mM Tris-HCl (pH 8.0) containing 16% glycerol.

Activity assays of coenzyme F430 biosynthesis enzymes

All coenzyme F430 biosynthetic reactions were carried out in a MBraun LABmaster Glove Box Workstation under a N_2 atmosphere containing <0.1 ppm O_2. All chemicals, unless otherwise noted, were obtained from Sigma-Aldrich. In a typical reaction, sirohydrochlorin was synthesized by incubating porphobilinogen (PBG) (0.88 mM) (Frontier Scientific) with HemC (0.06 mg/mL), HemD (0.06 mg/mL), SirA (0.12 mg/mL), SirC (0.36 mg/mL), *S*-adenosyl-L-methionine (SAM) (1.0 mM) (Carbosynth), $NAD(P)^+$ (1.0 mM), and $MgCl_2$ (4 mM) in 100 mM Tris-HCl buffer (pH 8.0) at 37 °C for 12 hours unless otherwise noted.

The sirohydrochlorin nickelochelatase was assayed by including CfbA (0.09 mg/mL) and $NiCl_2$ (200 μM) in a reaction otherwise identical to the above for sirohydrochlorin. An identical reaction (1 mL) was also prepared with the addition of 13 μL of a 3.6 mg/mL solution of CfbB (the subsequent enzyme in the pathway), but without the cosubstrates (i.e., glutamine, ATP) required for its activity, in order to help drive the CfbA reaction forward by alleviating any potential product inhibition. In this reaction, the molar ratio of potential product (Ni-sirohydrochlorin) to CfbA and CfbB was ~200:4:1.

Ni-sirohydrochlorin-*a*,*c*-diamide synthetase was assayed by adding 50 μL of a 3.6 mg/mL solution of CfbB, 23 μL of a 50 mM solution of L-glutamine, 2.5 μL of a 200 mM solution of ATP, 4.0 μL of a 500 mM solution of phosphoenol pyruvate (PEP), and 4 units of (1 unit/μL) *Bacillus stearothermophilus* pyruvate kinase (PK) to 250 μL of the completed Ni-sirohydrochlorin reaction and incubating for 12 hours at 37 °C. The last two components were used to regenerate ATP to help drive the CfbB reaction forward. An identical reaction was also prepared with 10 μL of a 4.8 mg/mL solution of the CfbCD complex (which catalyzes the next step in the pathway, and was once again added without a necessary component for activity, the

reductant sodium dithionite) to help alleviate any product inhibition. In this reaction, the molar ratio of potential product (Ni-sirohydrochlorin-*a*,*c*-diamide) to CfbB and the CfbCD complex was ~400:30:1.

The Ni-sirohydrochlorin-*a*,*c*-diamide reductive cyclase was assayed by incubating 200 μL of the completed Ni-sirohydrochlorin-*a*,*c*-diamide reaction with 30 μL of a 4.8 mg/mL solution of the CfbCD complex, 6.0 μL of a 1.0 M solution of sodium dithionite, 2.0 μL of a 200 mM solution of ATP, 8 μL of a 500 mM solution of PEP, and 4 units of PK for 12 hours at 37 °C. This reaction, which produces the known coenzyme F430 biosynthetic intermediate 15,17^3-seco-F430-17^3-acid, could not be driven in the same way by the addition of CfbE, since the cosubstrate required for its activity (ATP) could not be omitted from the reaction. The molar ratio of potential product (15,17^3-seco-F430-17^3-acid) to the CfbCD complex was ~30:1.

Finally, coenzyme F430 synthetase was assayed by adding 20 μL of a 1.4 mg/mL solution of CfbE, 1.0 μL of a 200 mM solution of ATP, 2.5 μL of a 500 mM solution of PEP, and 2 units of PK to 80 μL of the completed 15,17^3-seco-F430-17^3-acid reaction and incubating for 12 hours at 37 °C. The CfbE reaction was also driven forward by the inclusion of 20 μL of a 2.6 mg/mL solution of McrD (a putative coenzyme F430-binding protein/MCR chaperone) in the assay mixture. In this reaction, the molar ratio of potential product (coenzyme F430) to CfbE and McrD was approximately 20:1:5.

After completion of each of the assays described above, the reaction mixtures were quenched with an equal volume of methanol and centrifuged at 6,153 × *g* for 20 min. The pellet containing precipitated proteins and insoluble material was discarded and the supernatant was subjected to chromatographic analysis. An authentic coenzyme F430 standard was extracted in

an identical manner from MCR purified from *Methanothermobacter marburgensis* (a generous gift from Prof. Eduardus C. Duin, Auburn University).

High-performance liquid chromatography (HPLC)

Reversed-phase HPLC analysis was performed on an Agilent 1260 Infinity Quaternary LC System equipped with a Diode Array Detector (DAD) VL+ and an Agilent Poroshell 120 EC-C18 (4.6 × 150 mm, 2.7 μm) column. The Agilent OpenLAB ChemStation Edition software was used for data analysis. The chromatographic method utilized for characterization of the coenzyme F430 biosynthetic reactions consisted of the following gradient of water (solvent A) and acetonitrile (solvent B), each containing 0.5% formic acid: 0% B for 2 min, 0-20% B over 3 min, 20% B for 5 min, 20-25% B over 5 min, 25% B for 5 min, 25-30% B over 5 min, 30-100% B over 5 min. The flow rate was 1.0 mL/min and the chromatogram was acquired with detection at 400 nm.

Mass spectrometry (MS)

LC-MS analysis was performed on a Waters Acquity UPLC/Q-TOF Premier Mass Spectrometer equipped with an identical Agilent Poroshell 120 EC-C18 column. The Waters MassLynx MS software was used for data analysis. The LC method consisted of the same solvent system and gradient as described above. The electrospray ionization (ESI) mass detector was configured to positive ion mode with scanning between 0-1100 *m*/*z*. The inline Tunable UV (TUV) detector was set to 400 nm to match the peaks observed in the mass chromatograms to those observed by HPLC analysis.

Table S1. Forward and reverse primers utilized in PCR reactions. The underlined sequences indicate the restriction sites (for BamHI, BspHI, FseI, HindIII, NdeI, PciI, or XhoI) incorporated into the PCR products.

Gene	Primer
hemC	Forward: 5′-GCGGCCATATGTTAGACAATGTTTTAAGAATTGCC-3′ Reverse: 5′-TATAACTCGAGTCATGCCGGAGCGTC-3′
hemD	Forward: 5′-TGGGCCATATGAGTATCCTGGTC-3′ Reverse: 5′-TAGGACTCGAGTTATTGTAATGCCCG-3′
sirA	Forward: 5′-CGGCGCATATGTCAGAAAATTACGG-3′ Reverse: 5′-ATGAGCTCGAGTCAGAAATCCTTTCCTGC-3′
sirC	Forward: 5′-GAGGACATATGATGGCTGAAACAAATAATTTTC-3′ Reverse: 5′-TAGGACTCGAGTTATTCGAGCTTATCCGAG-3′
cfbA	Forward: 5′-GGCACCATATGACTGAGAAACTCGG-3′ Reverse: 5′-ATTACGGATCCTTACAGGGCTTCCTG-3′
cfbB	Forward: 5′-CCACACATATGTCCCACAGCAAACAATC-3′ Reverse: 5′-ATTAAGGTACCCTACCGGGGAGCCC-3′
cfbC	Forward: 5′-CGCTGCATATGAAAAAACAAAAGATCGTTGC-3′ Reverse 1: 5′-CCGCGAAGCTTTTATTTTGTCATTTCCC-3′ Reverse 2: 5′-ATTATGGCCGGCCTTATTTTGTCATTTCCC-3′
cfbD	Forward: 5′-CGCCGTCATGACTCAAAAAGAGATCTC-3′ Reverse: 5′-ATCACAAGCTTTCAGGCTTCTTTTGCAAC-3′
cfbE	Forward: 5′-GACACCATATGGACCTGTTCCGG-3′ Reverse: 5′-CGCACCTCGAGTTAACGGAAACATTTC-3′
mcrD	Forward: 5′-AATCTCATATGTCAGACTCTGCTTCAAACACG-3′ Reverse: 5′-GCTCTCTCGAGTCACTCATCTTTATCAGTGTC-3′

Fig. S1. Coenzyme F430 biosynthesis (*cfb*) genes from representative methanogens and ANME identified by comparative genomics.

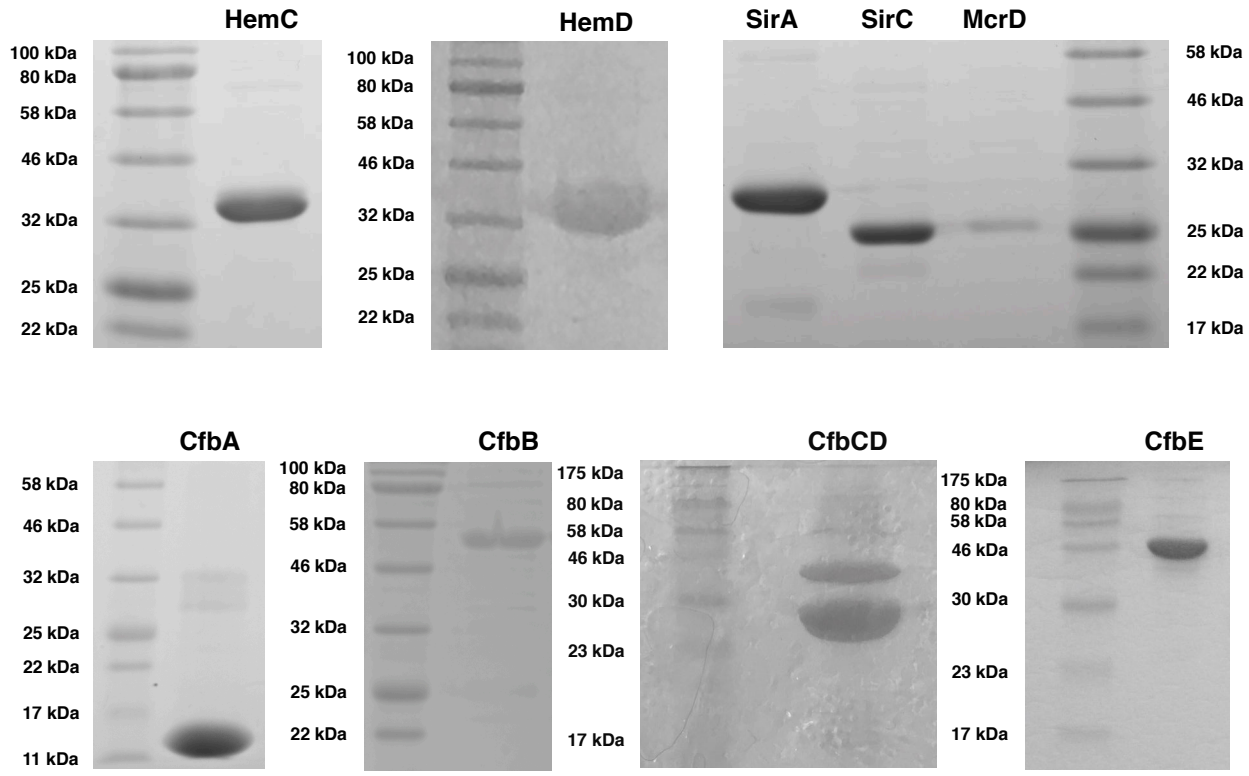


Fig. S2. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the purified enzymes used in the coenzyme F430 biosynthetic reactions.

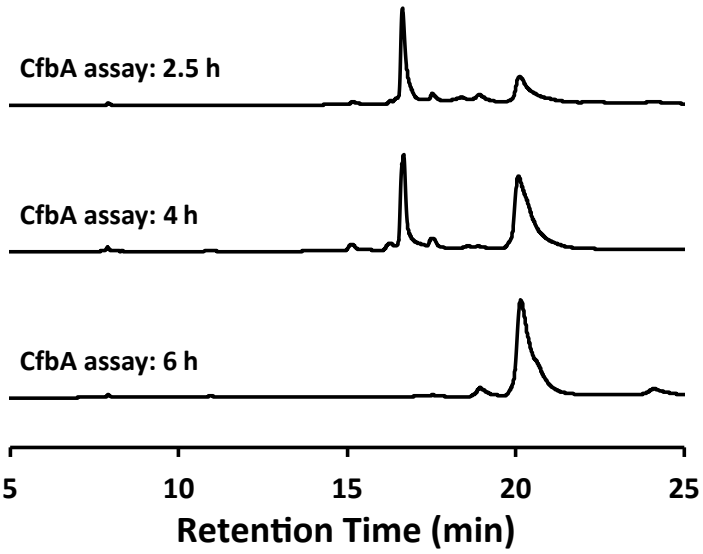


Fig. S3. HPLC assays showing the time course of the CfbA-catalyzed reaction. The reaction mixtures contain CfbA, $NiCl_2$, and enzymatically prepared sirohydrochlorin (along with CfbB to alleviate product inhibition).

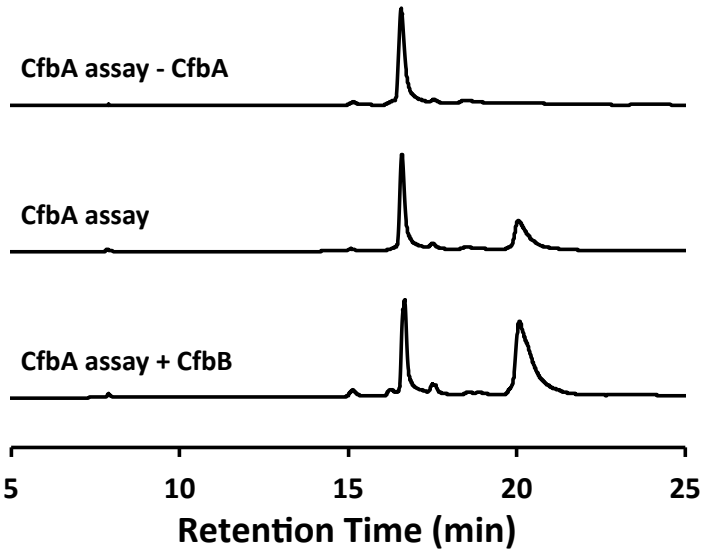


Fig. S4. HPLC assays of the CfbA-catalyzed reaction showing the enzyme/substrate requirements and the effect of the subsequent enzyme in the pathway (CfbB) on product yield. Each of these reactions were quenched after a 4 h incubation.

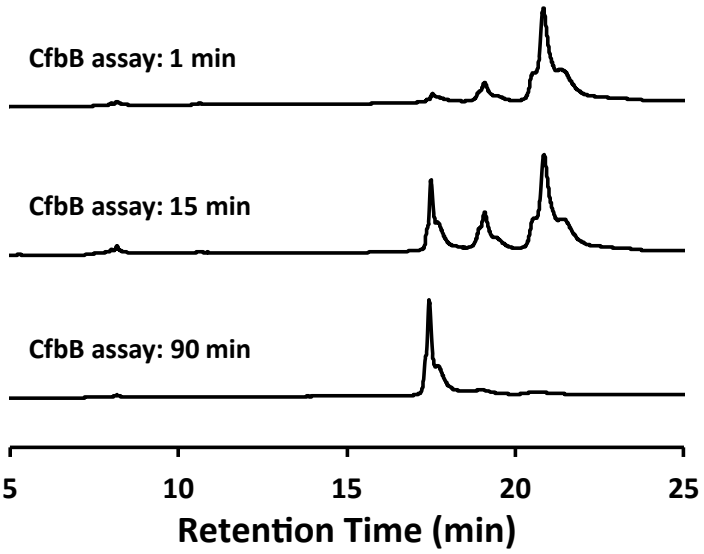


Fig. S5. HPLC assays showing the time course of the CfbB-catalyzed reaction. The reaction mixtures contain CfbB, glutamine, ATP, an ATP regeneration system (PEP and PK), and enzymatically prepared Ni-sirohydrochlorin (along with CfbCD to alleviate product inhibition).

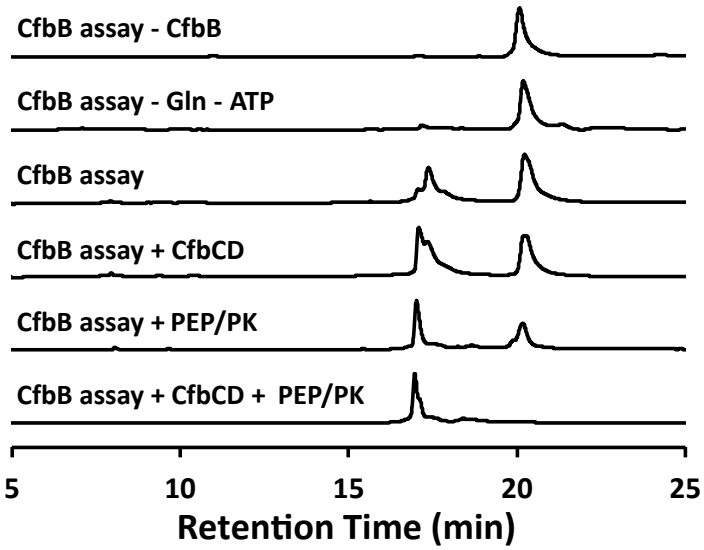


Fig. S6. HPLC assays of the CfbB-catalyzed reaction showing the cosubstrate (glutamine and ATP) requirements and the effect of an ATP regeneration system (PEP and PK) and the subsequent enzyme in the pathway (CfbCD) on product yield. Each of these reactions were quenched after a 12 h incubation.

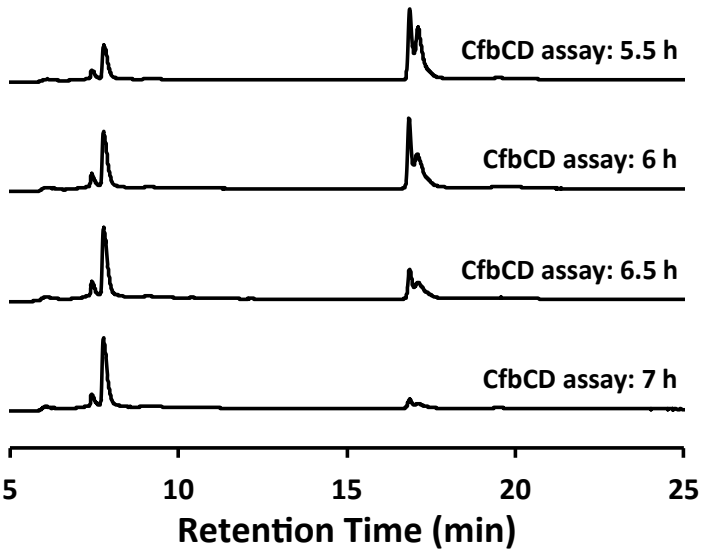


Fig. S7. HPLC assays showing the time course of the CfbCD-catalyzed reaction. The reaction mixtures contain CfbCD, sodium dithionite, ATP, an ATP regeneration system (PEP and PK), and enzymatically prepared Ni-sirohydrochlorin *a*,*c*-diamide.

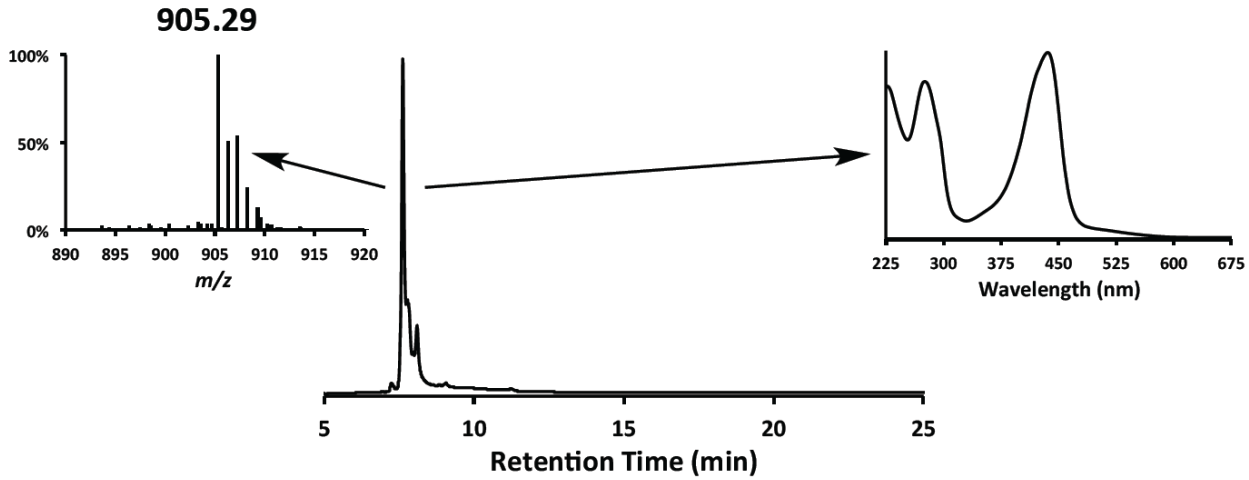


Fig. S8. HPLC, LC-MS, and UV-visible spectrophotometric analysis of authentic coenzyme F430 extracted from *M. marburgensis* MCR.

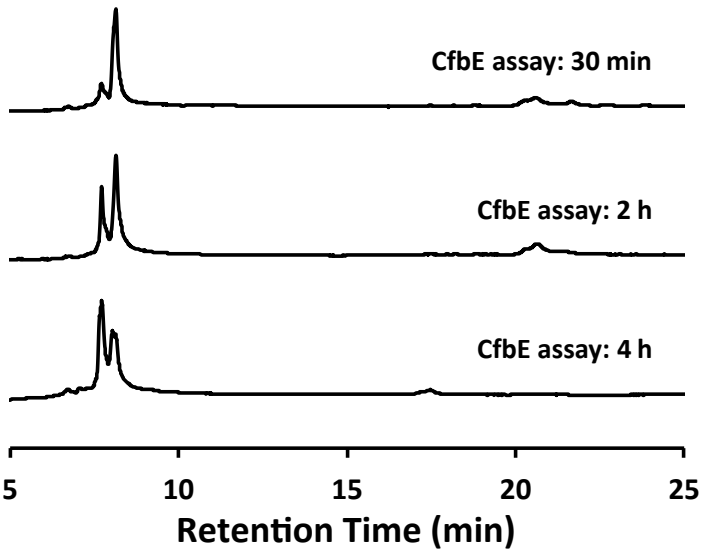


Fig. S9. HPLC assays showing the time course of the CfbE-catalyzed reaction. The reaction mixtures contain CfbE, ATP, an ATP regeneration system (PEP and PK), and enzymatically prepared 15,17^3-seco-F430-17^3-acid (along with McrD to alleviate product inhibition).

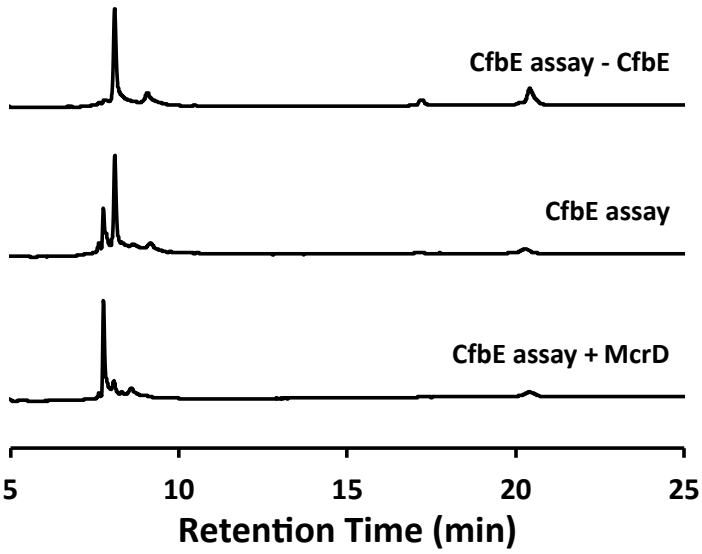


Fig. S10. HPLC assays of the CfbE-catalyzed reaction showing the enzyme requirement and the effect of McrD on the yield of coenzyme F430. Each of these reactions were quenched after a 12 h incubation.

F430 Biosynthesis and Insertion

Steven O. Mansoorabadi, Kaiyuan Zheng & Phong D. Ngo
Auburn University, Auburn, AL, USA

1 INTRODUCTION

Nearly 1 billion tons of methane, a potent greenhouse gas with ~30 times the global warming potential (GWP) of carbon dioxide, is produced annually by methanogenic archaea.[1–3] Different species of methanogens are capable of producing methane from a wide range of electron acceptors, including carbon dioxide, acetate, methanol, and methylamines.[4] The key methane forming enzyme in all of these pathways is methyl-coenzyme M reductase (MCR), which catalyzes the conversion of coenzyme B (CoB-SH) and methyl-coenzyme M (MeS-CoM) to the mixed heterodisulfide CoB-S-S-CoM and methane (also see ***Nickel Enzymes & Cofactors***; ***Methyl-Coenzyme M Reductase***) (Scheme 1).[5]

MCR is an ~270 kDa $(\alpha\beta\gamma)_2$ heterohexamer containing two active sites (Figure 1), each of which binds the unique nickel tetrahydrocorphin, coenzyme F430.[6,7] Coenzyme F430 is the most highly reduced tetrapyrrole found in Nature and contains two exocyclic rings: a γ-lactam E ring and a carbocyclic F ring (Scheme 2).[8] Additionally, the active site region of MCR contains several unusual posttranslational modifications (PTMs), including 1-*N*-methylhistidine, 5-(*S*)-methylarginine, 2-(*S*)-methylglutamine, thioglycine, and *S*-methylcysteine residues.[9]

In the active form of MCR (MCR_{red1}), coenzyme F430 is in the Ni(I) oxidation state.[10] MCR catalysis is thought to proceed via a radical mechanism with formation of Ni(II)-thiolate and methyl radical intermediates (Scheme 3).[11] This reaction is highly exergonic with a biochemical standard Gibbs free energy change of $\Delta G^{\circ\prime} = -30 \pm 10\,\mathrm{kJ\,mol^{-1}}$.[12] Regardless, MCR from *Methanothermobacter marburgensis* was shown to catalyze the oxidation of methane with CoB-S-S-CoM with a specific activity of ~11.4 nmol min^{-1} mg^{-1} at 60 °C and 1 bar CH_4 (corresponding to ~1 mM dissolved CH_4).[13]

Recently, anaerobic methanotrophic archaea (ANME) have been shown to contain a homolog of MCR and to catalyze the anaerobic oxidation of methane (AOM).[14] AOM is thought to operate, at least in part, as the reverse of methanogenesis, with MCR catalyzing the first step, the oxidation of methane with CoB-S-S-CoM.[15] Reverse methanogenesis is an endergonic process, and the electrons derived from the oxidation of methane must be coupled to the reduction of a favorable electron acceptor, such as sulfate.[16] Consequently, ANME are often found as part of syntrophic consortia with sulfate-reducing bacteria (SRB).[17] Several groups of ANME are known, including those that are phylogenetically related to methanomicrobiales (ANME-1) and methanosarcinales (ANME-2 and ANME-3).[18] SRB that participate in the AOM include *Desulfosarcina* and *Desulfococcus* spp. (associated with ANME-1 and ANME-2) and *Desulfobulbus* spp. (associated with ANME-3).[18] The interactions between ANME and their associated SRB are proposed to involve syntrophic coupling via direct interspecies electron transfer (DIET).[19]

The structure of MCR from an ANME-1 was solved and found to be nearly identical to that of methanogenic MCR (r.m.s.d. of ~1 Å).[20] However, there were several noteworthy differences, including a modified coenzyme, 17^2-methylthio-F430 (Scheme 2), and a distinct set of active site PTMs (1-*N*-methylhistidine, 7-hydroxytryptophan, thioglycine, and *S*-oxymethionine).[20,21] A comparison of the active site regions of MCR from methanogens and ANME is given in Figure 2.

In this chapter, current progress in the understanding of coenzyme F430 biosynthesis (cfb) and the PTM of the active site region of MCR is reviewed. In addition, differences in the structures of methanogenic and methanotrophic MCR are highlighted and discussed in terms of their dichotomous biochemical roles.

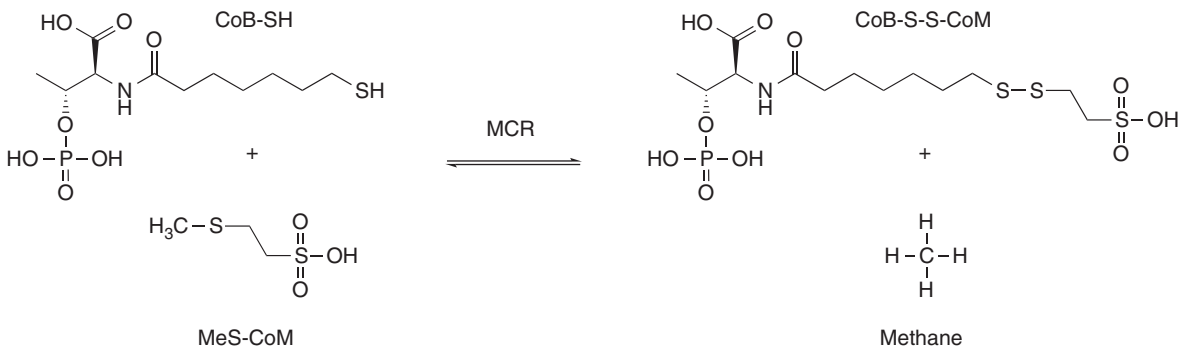
2 COENZYME F430 BIOSYNTHESIS

There are two main biosynthetic routes for the formation of tetrapyrroles, starting from either glycine or L-glutamate.[22] The pathways converge at 5-aminolevulinate and proceed with formation of uroporphyrinogen III, the last common precursor of all tetrapyrroles.[22] Uroporphyrinogen III can then be decarboxylated to coproporphyrinogen III (to produce tetrapyrroles such as heme or chlorophyll) or methylated at C2 and C7 to generate precorrin 2 (and tetrapyrroles such as cobalamin and siroheme).[22]

In archaea, tetrapyrroles are biosynthesized from L-glutamate.[22] Coenzyme F430 is a C2- and C7-methylated

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488



Scheme 1 Reaction catalyzed by methyl-coenzyme M reductase (MCR)

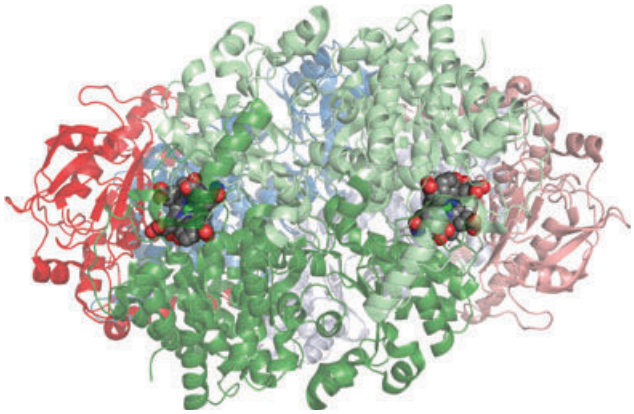
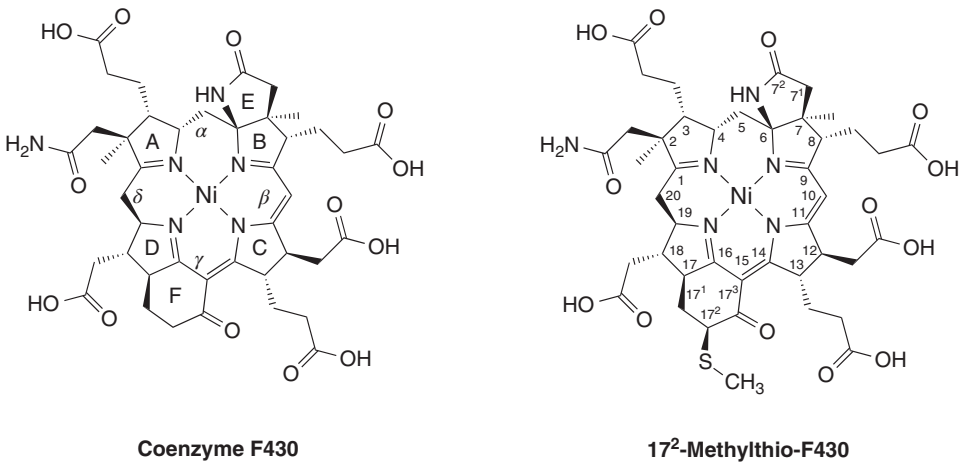


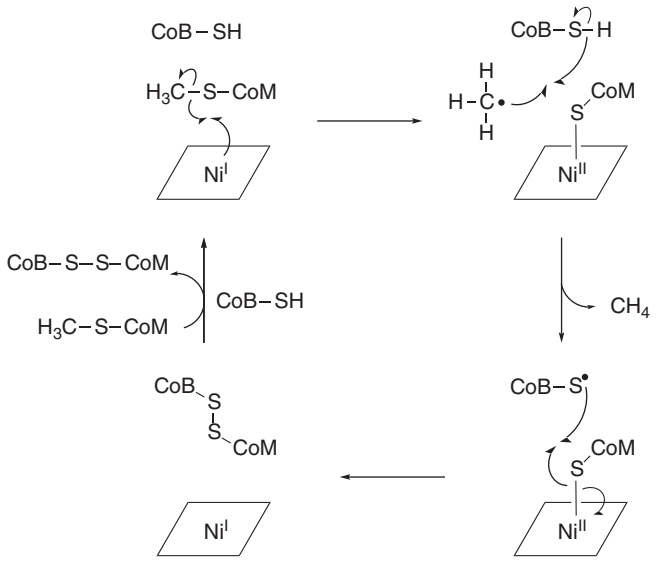
Figure 1 Quaternary structure of MCR. The α-, β-, and γ-subunits are colored in shades of green, blue, and red, respectively. Coenzyme F430 is shown in each active site as a space-filling model. (PDB ID: 1hbm)



Scheme 2 Structures of coenzyme F430 and the 17^2-methylthio derivative from ANME-1

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488



Scheme 3 Proposed radical mechanism of MCR

tetrapyrrole and thus precorrin 2 (dihydrosirohydrochlorin) is an intermediate in its biosynthesis.[22] Indeed, Mucha *et al*. showed that isotopically labeled sirohydrochlorin [a 2-electron oxidized form of precorrin 2 that is formed by a nicotinamide adenine dinucleotide (phosphate) ($NAD(P)^+$)-dependent dehydrogenase] could be metabolized by cell-free extracts of *Methanobacterium thermoautotrophicum* into coenzyme F430.[23] Likewise, precorrin 2 and sirohydrochlorin are intermediates in the biosynthesis of cobalamin, siroheme, heme d_1, and heme (via the recently discovered alternative heme biosynthetic pathway; see ***Heme Biosynthesis and Insertion*** and ***Siroheme Assembly and Insertion into Nitrite and Sulfite Reductase***).[22,24]

In 1987, Thauer and coworkers identified the first tetrapyrrolic intermediate unique to the coenzyme F430 biosynthetic pathway.[25] This intermediate accumulated in methanogens that were initially grown in Ni-depleted media and then transferred while in the exponential phase to fresh media supplemented with $NiCl_2$ and 5-aminolevulinate.[25] Nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry (MS), and comparisons with model compounds identified the intermediate as 15,17^3-seco-F430-17^3-acid, which differs from coenzyme F430 by the presence of a propionate side chain at C17 and the absence of the carbocyclic F ring.[25] Assays using cell-free extracts, as well as isotope tracer and feeding experiments, demonstrated that 15,17^3-seco-F430-17^3-acid could be converted directly to coenzyme F430 and thus was a true biosynthetic intermediate.[25]

Very recently, Mansoorabadi and coworkers used comparative genomics to identify the *cfb* genes.[26] The *cfb* genes, which are present in all methanogens, are found clustered together only in the sequenced genomes of members of the methanomicrobia and thermoplasmata.[26] Included among these genes are homologs of *cbiX*^{S} (small sirohydrochlorin cobaltochelatase) and *cbiA* (cobyrinic acid *a*,*c*-diamide synthase), two genes involved in the cobalamin biosynthetic pathway (see ***Cobalamin Biosynthesis and Insertion***); *nifH* (nitrogenase Fe protein) and *nifD* (nitrogenase MoFe protein, α-subunit), two components of the key enzyme

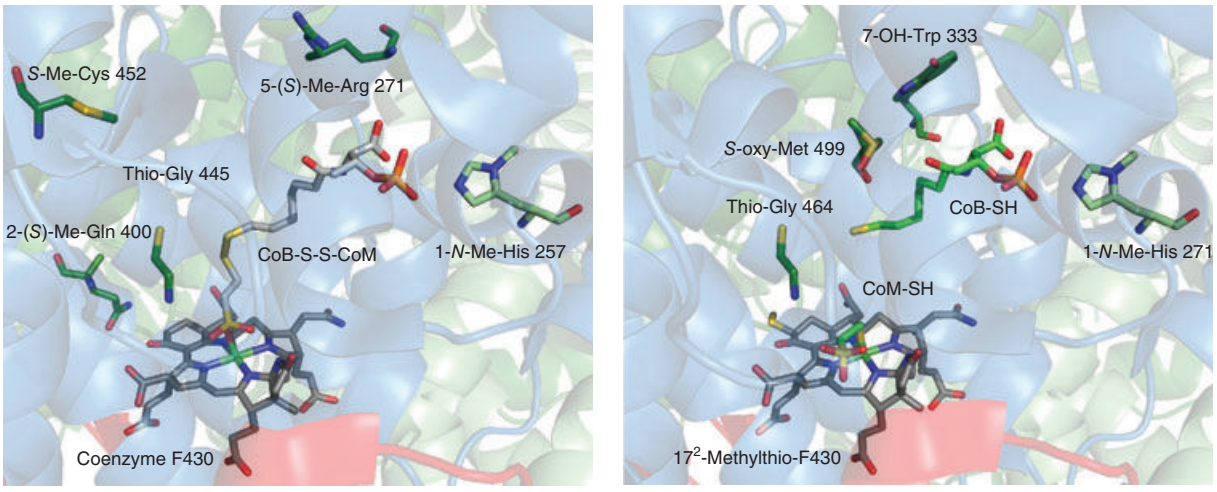
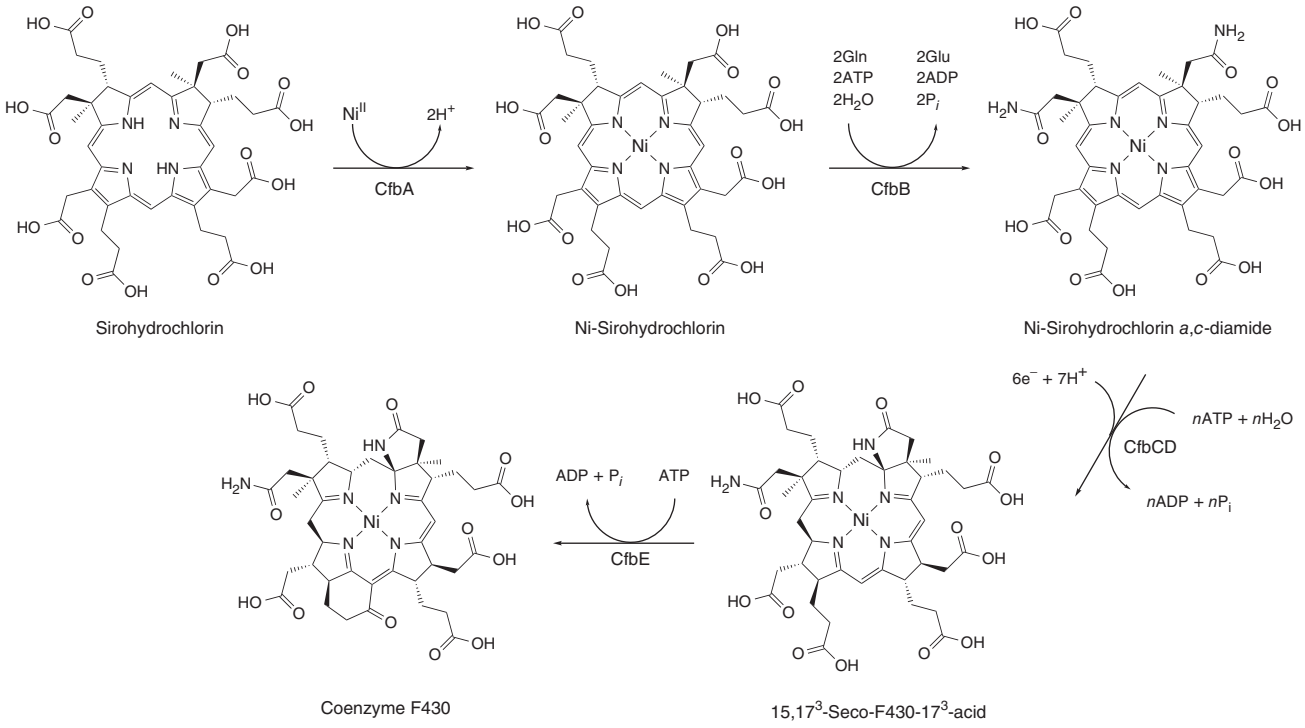


Figure 2 Comparison of the active site regions of MCR from methanogens (a) and ANME (b). The unique posttranslational modifications (PTMs) present in each active site are highlighted. (PDB IDs: 1hbm and 3sqg, respectively)

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488



Scheme 4 The biosynthetic pathway of coenzyme F430

of nitrogen fixation (see ***FeS Cluster Assembly: NIF System in Nitrogen-Fixing Bacteria*** and ***Assembly of Nitrogenase FeMoco and P Cluster***); and *murD*, an adenosine 5′-triphosphate (ATP)-dependent ligase involved in peptidoglycan biosynthesis.[27–31]

The *cfb* genes from *Methanosarcina acetivorans* were cloned and heterologously expressed in *Escherichia coli*.[26] The recombinant enzymes together were shown to be capable of synthesizing coenzyme F430 from sirohydrochlorin in vitro (Scheme 4).[26] The sirohydrochlorin cobaltochelatase homolog, CfbA, was found to catalyze the specific Ni-chelation of sirohydrochlorin while no chelation activity was observed with precorrin 2.[26]

Cobyrinic acid *a*,*c*-diamide synthase (CbiA) catalyzes the ATP-dependent amidation of the *a*- and *c*-acetate side chains of cobyrinate using glutamine as an ammonia source.[28] The CbiA homolog, CfbB, catalyzes an analogous reaction, the amidation of the *a*- and *c*-acetate side chains of Ni-sirohydrochlorin.[26]

The product of the CfbB reaction, Ni-sirohydrochlorin *a*,*c*-diamide, is then converted to 15,17^3-seco-F430-17^3-acid by the CfbCD complex in the key step of the *cfb* pathway.[25,26] This remarkable transformation involves both a six-electron reduction of the isobacteriochlorin ring system and the cyclization of the *c*-acetamide side chain to form the γ-lactam E ring.[26] As eluded to above, CfbC and CfbD are homologous to the Fe protein (NifH) and the NifD subunit of the MoFe protein (NifDK) of nitrogenase.[29,30,32] Nitrogenase couples the hydrolysis of 16 ATPs to the eight-electron reduction of dinitrogen and two protons in the pathway of biological nitrogen fixation.[33]

There are several homologs of nitrogenase that are known, including two catalyzing the same reaction but utilizing different metals (Fe-only and V-containing nitrogenase) and two others involved in the biosynthesis of the photosynthetic pigments chlorophyll and bacteriochlorophyll.[34] The latter group includes the dark-operative protochlorophyllide *a* oxidoreductase (DPOR) and chlorophyllide *a* oxidoreductase (COR), which catalyze the ATP-dependent two-electron reduction of the C17═C18 double bond of protochlorophyllide *a* and the C7═C8 double bond of chlorophyllide *a*, respectively.[35,36] Each of these homologs contains heteromeric MoFe protein-like components comprised of subunits arising from a putative paralogous gene duplication and divergence.[34,37] In contrast, the MoFe protein-like component of the Ni-sirohydrochlorin *a*,*c*-diamide reductive cyclase is homomeric. Thus, the CfbCD complex is representative of a more primitive lineage of the nitrogenase superfamily.[26] Indeed, a molecular phylogenic analysis of the nitrogenase homologs placed the ancestor of the Ni-sirohydrochlorin *a*,*c*-diamide reductive cyclase in the last common ancestor (LCA) of modern organisms and basal to the emergence of the lineages involved in nitrogen fixation and photosynthesis.[38]

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488

The final step in the biosynthesis of coenzyme F430 is catalyzed by the Mur-ligase homolog, CfbE, which carries out the ATP-dependent cyclization of the *g*-propionate side chain to generate the carbocyclic F ring.[26,31] The yield of coenzyme F430 in the CfbE reaction was substantially enhanced by the addition of McrD, a protein of unknown function from the *mcr* gene cluster.[26,39] McrD is known to physically interact with MCR, but is not required for its in vitro activity.[40] Therefore, the positive effect of McrD on the yield of coenzyme F430 was interpreted as the alleviation of product inhibition of the CfbE reaction, which suggests that McrD may serve as a coenzyme F430-binding protein/chaperone that can deliver the coenzyme to *apo* MCR.[26]

ANME-1 contain coenzyme F430 in addition to 17^2-methylthio-F430, suggesting that the former may be a biosynthetic precursor of the latter.[21] Biochemical and genomic information of ANME-1 is currently limited, so the gene(s) and corresponding enzyme(s) responsible for the incorporation of the 17^2-methylthio group have yet to be identified. White and coworkers suggested that a member of the radical *S*-adenosyl-L-methionine (SAM) superfamily could be responsible for this transformation.[41] Radical SAM enzymes utilize a $[4Fe–4S]^{1+}$ cluster to effect the reductive cleavage of SAM to generate a 5′-deoxyadenosyl radical that initiates catalysis.[42] Radical SAM enzymes then utilize this highly reactive primary radical to catalyze a diverse array of reactions, including sulfur incorporation, methylation, and methylthiolation.[43]

White and coworkers also detected several other coenzyme F430 derivatives in the cell-free extracts of methanogens and/or ANME in addition to 17^2-methylthio-F430 (F430-2).[41] Included among these were a F430 derivative proposed to contain a 3-mercaptopropionate moiety bound as a cyclic thioether (F430-3), 12,13-didehydro-F430 (F430-4), vinyl-F430 (F430-5), vinyl-17^2-methylthio-F430 (F430-6), 17^2-methylsulfoxide-F430 (F430-7), 17^2-keto-F430 (F430-8), 17^2-hydroxy-F430 (F430-9), and 17^1-17^2-dihydroxy-F430 (F430-10) (Scheme 5).[41] The routes by which these F430 derivatives are formed and their physiological relevance are unclear, but their identification suggests the possibility that derivatives of F430 may have additional biochemical roles in Nature.

3 MCR POSTTRANSLATIONAL MODIFICATIONS

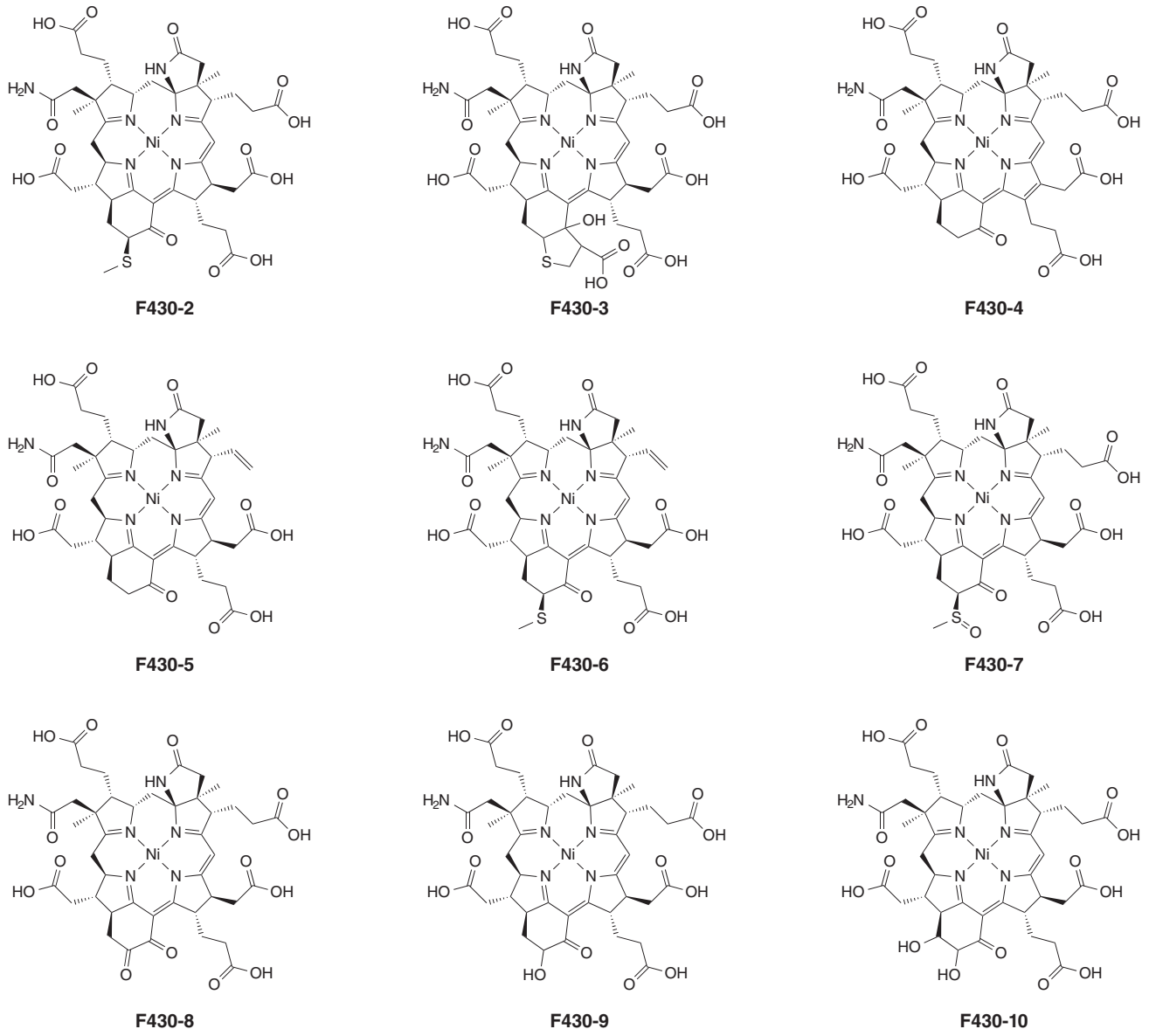
The first high-resolution crystal structure of MCR was determined in 1997 by Ermler and Grabarse.[6] This study led to the surprising observation that the active site region of MCR from *M. thermoautotrophicum* was extensively posttranslationally modified. As noted above, these modifications included 1-*N*-methylhistidine, 5-(*S*)-methylarginine, 2-(*S*)-methylglutamine, thioglycine, and *S*-methylcysteine residues (Scheme 6).[6] Not all of the PTMs are invariably present in the MCRs from all methanogens.[44] In particular, the *S*-methylcysteine residue is absent in MCR from *Methanopyrus kandleri* and *Methanocaldococcus jannaschii,* and the 2-(*S*)-methylglutamine is absent in *Methanosarcina barkeri*.[44] The absence of the 2-(*S*)-methylglutamine modification in *M. barkeri* is accompanied by a compensating alanine-to-threonine substitution in a nearby residue, which positions an ordered water molecule near the site occupied by the 2-methyl group.[44] Compensatory changes in the active site region due to the absence of the *S*-methylcysteine were less obvious in the structure of *M. kandleri*, due in part to the lower resolution (2.7 Å) of this structure.[44] Very recently, another novel PTM, didehydroaspartate, was detected adjacent to the thioglycine residue in MCRs from *M. barkeri* and *M. marburgensis*, although it was not present in MCR from *Methanothermobacter wolfeii*.[45]

The active site region of MCR from ANME-1 also contains several PTMs.[9,20] The 1-*N*-methylhistidine and thioglycine residues are conserved in the ANME-1.[9,20] In place of the 5-(*S*)-methylarginine modification, a nearby tryptophan residue is oxidized to 7-hydroxytryptophan, with the 7-hydroxyl group occupying the position of the 5-(*S*)-methyl moiety.[20] The glutamine residue that is modified in some methanogenic MCRs is replaced with a smaller valine residue in ANME-1, which removes a steric clash that would otherwise occur with the 17^2-methylthio modification of F430.[20] MCRs from the ANME-2 and ANME-3 retain glutamine at this position, suggesting that they utilize unmodified coenzyme F430 as a prosthetic group.[20] There is some evidence that MCR from certain ANME-1 may contain the *S*-methylcysteine residue; however, this modification is absent in the crystal structure.[20] The unmethylated cysteine and an *S*-oxymethionine residue are located near the invariant thioglycine, the latter of which forms a hydrogen bonding network with the thioglycine sulfur atom via an intervening water molecule.[20]

The genes and corresponding enzymes responsible for the PTMs in the active site region of MCR from both methanogens and ANME are unknown. However, the biosynthetic origin of the methylated amino acids in MCRs from methanogenic archaea (specifically *M. thermoautotrophicum*) was examined by Thauer and coworkers.[46] Labeling experiments using L-[CD_3]-methionine and MS showed incorporation of the label into each of the four methylated amino acids (1-*N*-methylhistidine, 5-(*S*)-methylarginine, 2-(*S*)-methylglutamine, and *S*-methylcysteine).[46] As L-methionine is a biosynthetic precursor of the ubiquitous methyl donor SAM, these data suggest that these PTMs are carried out by SAM-dependent methyltransferases.[46] As the methyl group of methionine was not found to be a substrate for methanogenesis, this effectively rules out SAM acting as an indirect methyl donor via MeS-CoM (or other intermediate methyl

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488



Scheme 5 Proposed structures of coenzyme F430 variants found in methanogens and ANME

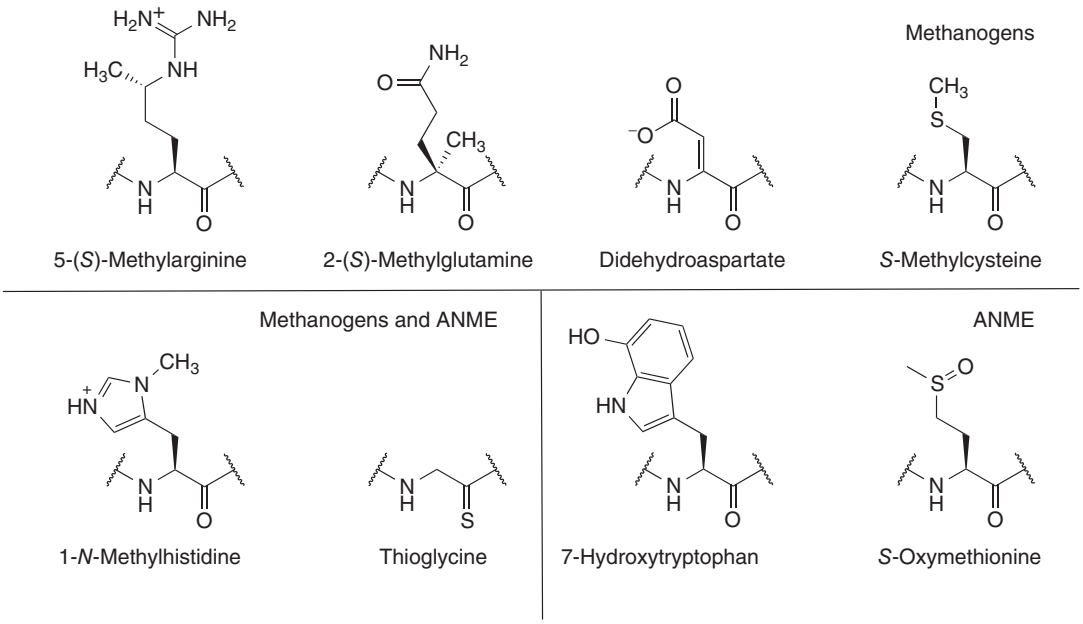
carriers, such as cobalamin or tetrahydromethanopterin (THMPT), that can be used to synthesize MeS-CoM).[47]

The biosynthesis of the 1-*N*-methylhistidine and *S*-methylcysteine modifications is likely to proceed via methyl cation transfer from SAM to the activated N and S nucleophiles of the corresponding deprotonated amino acids, respectively.[46] In contrast, the 2-(*S*)-methylglutamine and 5-(*S*)-methylarginine modifications involve C-methylation and are uniquely found in MCR.[46] If the 2-(*S*)-methylglutamine modification is to be biosynthesized in a fashion analogous to that described above, the α-proton must first be removed to generate an enolate anion that accepts the methyl cation from SAM with net retention of configuration. However, the pK_a of the α-proton of an amino acid in a peptide has been estimated to be as high as 32 and higher than that of the corresponding free amino acid, due to the lack of the positively charged ammonium group and the weaker electron-withdrawing effects of a carbonyl from an amide versus a carboxylic acid.[48,49] This pK_a is close to the maximum value known for enzymatic proton transfer reactions.[50] The δ-protons of arginine are expected to be similarly unactivated, and thus the *C*-methyl modifications are unlikely to be introduced via polar mechanisms.

To date, the enzymatic methylation of non-nucleophilic substrates has only been observed by members of the radical SAM superfamily.[51] There exist several classes

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488



Scheme 6 Posttranslational modifications found in the active site region of MCR from methanogens and ANME

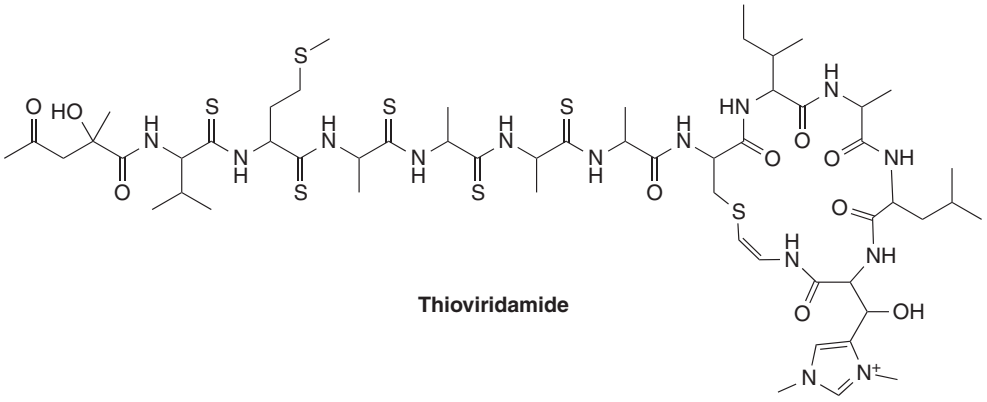
of radical SAM methyltransferases, which differ in their protein architecture, cofactor requirements, and proposed mechanisms.[51,52] Radical SAM methyltransferases capable of methylating sp^3-hybridized centers, such as the α-carbon of glutamine or the δ-carbon of arginine, belong to class B and D.[52] The class B radical SAM methyltransferases contain a methylcobalamin-binding domain in addition to the canonical CxxxCxxC motif that coordinates the [4Fe–4S] cluster that binds and reduces SAM.[52] In contrast, the class D methyltransferases contain a second auxiliary iron–sulfur cluster and a N^5,N^{10}-methylenetetrahydrofolate-binding site.[52] The methylcobalamin and N^5,N^{10}-methylenetetrahydrofolate cofactors are proposed to serve as the methyl donor for each class. It is therefore possible that a member of one of these classes of methyltransferases is responsible for the 2-(*S*)-methylglutamine and/or 5-(*S*)-methylarginine PTM. However, the radical SAM superfamily is exceedingly large and diverse with over 113,000 members (of which more than 10,000 are thought to serve as methyltransferases), so it is entirely possible that a member of this superfamily utilizing novel methyltransferase chemistry is responsible for one of these PTMs.[52] In particular, a member of the radical SAM superfamily (methanogenesis marker 10) is conserved in all methanogens and is often adjacent to, and divergently transcribed from, the *mcr* gene cluster. Sequence analysis of methanogenesis marker 10 suggests that the encoded enzyme contains a N-terminal radical SAM domain that binds an auxiliary iron–sulfur cluster with a conserved CxxCxPxxxGCxxC (or CxxCxxxxGCxYC) motif and a C-terminal domain of unknown function (DUF512) with weak homology to glutathione reductase.

The thioglycine residue, like the 2-(*S*)-methylglutamine and 5-(*S*)-methylarginine modifications, has not yet been observed in any enzyme other than MCR.[46] However, similar modifications are seen in thiopeptide-containing natural products, such as thioviridamide (tva), which contains thiovaline, thiomethionine, and three thioalanine residues (Scheme 7).[53,54] The *tva* biosynthetic gene cluster from *Streptomyces olivoviridis* was identified by Hayakawa and coworkers.[55] The cluster consists of 15 genes, including a *tfuA*-associated *ycaO* homolog, which has been implicated in the formation of the thioamide modifications.[55]

Certain members of the YcaO superfamily have recently been shown to bind ATP with a novel motif and use it to activate the carbonyl oxygen of peptide backbones of microcins for nucleophilic attack by an adjacent cysteine/serine residue and the formation of thiazole/oxazole rings.[56,57] TfuA was found to be important for the production of the thiazole-containing peptide antibiotic trifolitoxin, though its function remains unclear.[58] Another genetic marker of methanogenesis (methanogenesis marker 1) encodes a member of the YcaO superfamily and is typically found clustered with a *tfuA* homolog in the genomes of methanogens. It is therefore possible that these genes are involved in the formation of the thioglycine modification of MCR using a similar mechanism, namely, the activation of the carbonyl oxygen of glycine with ATP and the subsequent incorporation of a sulfur atom.

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488



Scheme 7 Structure of the thioamide-containing natural product thioviridamide

Much is now known about the biosynthesis of sulfur-containing cofactors and nucleosides. In the canonical thiamine biosynthetic pathway, sulfur is obtained from cysteine via a cysteine desulfurase, which generates a cysteine persulfide that is transferred to the sulfur-carrier protein, ThiI.[59] The adenylyltransferase ThiF then catalyzes the transfer of the ThiI-bound sulfur atom to a secondary sulfur-carrier protein, ThiS, as a C-terminal thiocarboxylate.[59] ThiS then serves as the immediate sulfur donor for thiazole ring formation.[59] Similar strategies are utilized for the biosynthesis of other thio-cofactors, including iron–sulfur clusters and molybdenum cofactor.[60] ThiI is also implicated as the immediate sulfur donor for the formation of the conserved 4-thiouridine modification at position 8 in the tRNAs of bacteria and archaea.[61] ThiI from the methanogen *Methanococcus maripaludis* was shown to utilize a CxxC motif to generate a cysteine persulfide using hydrogen sulfide as a source of sulfur for the biosynthesis of 4-thiouridine, but not for thiamine.[62] Moreover, several methanogens do not contain an obvious *thiS* homolog in their genomes and thus may utilize an alternative route of sulfur transfer for the biosynthesis of thiamine.[62] One of these systems could also serve to provide sulfur for the biosynthesis of the thioglycine modification in MCR from methanogens and ANME.

The biosynthetic origin of the didehydroaspartate modification is also unknown. Shima and coworkers speculate on the basis of comparative genomics that a homolog of pyrroline-5-carboxylate reductase (which catalyzes the terminal step in proline biosynthesis) may be responsible for this modification.[45,63] Didehydroaspartate is found as a PTM in a group of hexapeptide mycotoxins, the phomopsins.[64] Very recently, the phomopsin (*phom*) biosynthetic gene cluster from *Phomopsis leptostromiformis* was identified by Zhang and van der Donk.[65] The *phom* cluster encodes an oxidoreductase and several proteins of unknown function, but the enzyme responsible for the didehydroaspartate modification remains unclear.[65] However, none of the *phom* genes have homologs in the genomes of both *M. barkeri* and *M. marburgensis*. The didehydroaspartate residue could be converted to aspartate by the addition of a strong reducing agent (zinc granules) and thus may be formed auto-catalytically (e.g., as an intermediate during the reductive activation of MCR).[45]

4 MCR CATALYSIS IN ANME

In utilizing reverse methanogenesis for the AOM, a large thermodynamic barrier ($\Delta G^{\circ\prime} = +30\,\mathrm{kJ\,mol^{-1}}$) must be overcome in the initial step of the pathway, the oxidation of methane with CoB-S-S-CoM to generate CoB-SH and MeS-CoM.[66] This uphill barrier necessitates that the kinetics of the MCR-catalyzed reaction will be slow, and indeed the estimated in vivo specific rate of AOM with sulfate is only $70\,\mathrm{nmol\,min^{-1}\,mg^{-1}}$ of MCR at 12 °C and 14 bar methane.[13] The sluggish kinetics and modest driving force of sulfate-dependent AOM ($\Delta G^{\circ\prime} = -17\,\mathrm{kJ\,mol^{-1}}$) is consistent with the slow growth rates of ANME-SRB consortia (which have a doubling time on the order of 3–7 months).[13,67,68]

One strategy that could be used by ANME to enhance the rate of the AOM is to increase the cellular concentration of MCR. Consistent with this hypothesis, MCR constitutes ~10% of the protein extracted from ANME-containing microbial mats.[69] Another approach that might be utilized by ANME is to replace the overall MCR-catalyzed reaction with one that is more thermodynamically and kinetically feasible, for example, by coupling the oxidation of methane to an exergonic process such as ATP hydrolysis or to the movement of ions down an electrochemical gradient. If MCR catalyzes this alternative reaction, the altered reactivity would likely be reflected in structural differences between methanogenic and methanotrophic MCR.

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488

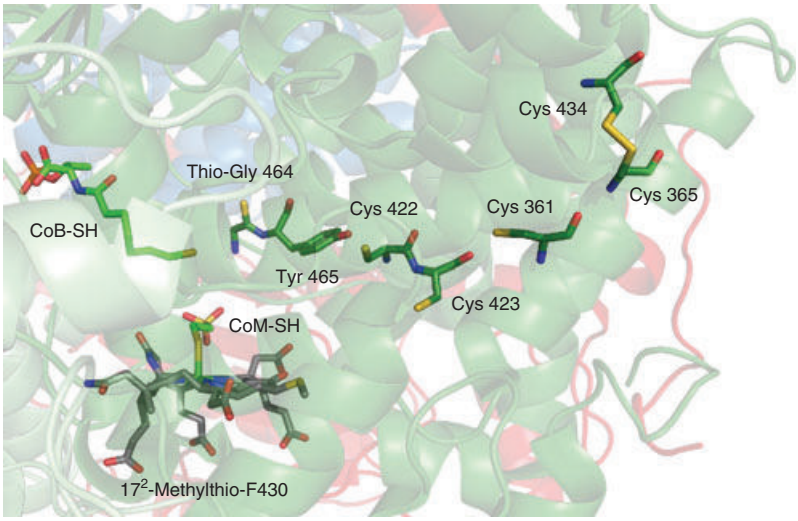


Figure 3 Cysteine-rich path leading from the active site to the protein surface in MCR from ANME-1. (PDB ID: 3sqg)

One of the most noteworthy differences in the structure of MCR from ANME-1 is the presence of a cysteine-rich patch that leads from the active site to the protein surface (Figure 3).[20] At the proximal end of the patch is the conserved thioglycine modification. It has been proposed that the thioglycine residue could play a redox role during catalysis and/or as part of an electron transfer path for the reduction of coenzyme F430 to the active Ni(I) form by an exogenous reductase complex.[70]

The cysteine-rich patch could allow the ANME-1 MCR to catalyze an alternative reaction by coupling methane oxidation to the reduction of an exogenous electron acceptor (Scheme 8). In this scenario, the initial stages of catalysis proceed as the microscopic reverse of the methanogenesis reaction, with the formation of MeS-CoM and CoB-SH.[11] However, instead of the catalytic cycle completing with the release of both products, CoM-SH replaces MeS-CoM in the active site and the CoB-S-S-CoM heterodisulfide is regenerated by thiol-disulfide exchange or proton-coupled electron transfer through the thioglycine residue and cysteine-rich patch to an external electron acceptor (A_{ox}). Depending on the identity of A_{ox}, the critical methane-activating step could operate near equilibrium or even with favorable thermodynamics. The preparation of ANME-1 MCR in an active form (which has yet to be achieved) and/or the identification of the putative electron acceptor (A_{ox}) would provide a means of testing this alternative reaction hypothesis of methanotrophic MCR.

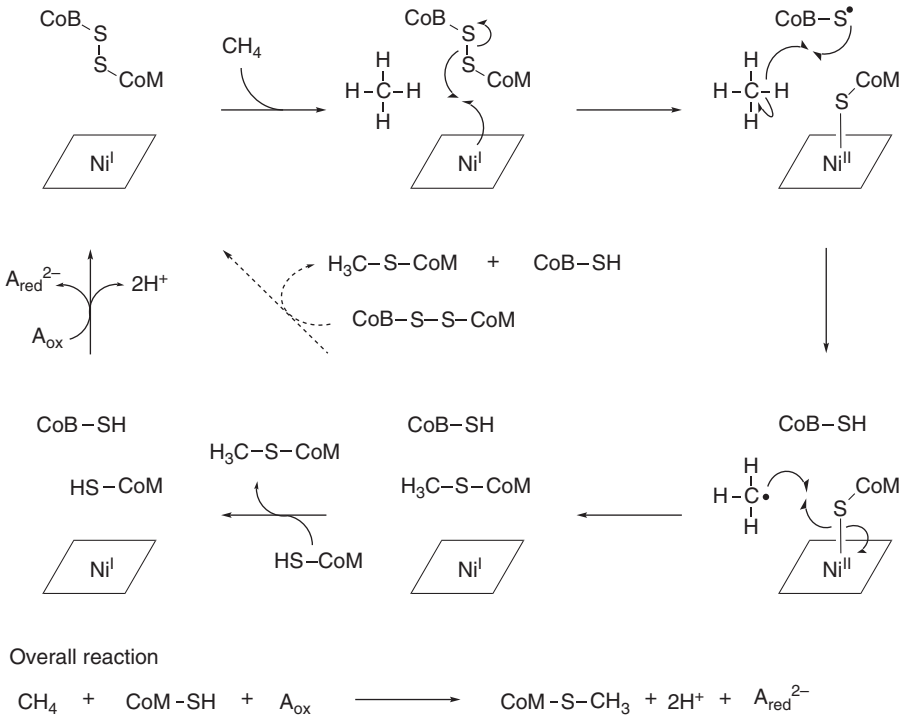
5 CONCLUSIONS

With the discovery of vast natural gas reserves, the rapid increase in the production of natural gas from unconventional sources (e.g., shale formations), and the growing concern about the environmental impact of the flaring/venting of methane, there is great current interest in biological methane conversion technologies.[71,72] Given the ability of ANME to utilize MCR to catalyze the AOM, this process holds much promise for use in a bio-based gas-to-liquid (GTL) conversion strategy.[73] The direct use of ANME in such processes will require major advances in the understanding of the microbiology of these organisms, which cannot currently be obtained in pure culture or grown at industrially viable rates. However, significant recent progress has been made in the understanding of the mechanisms of MCR activation and catalysis by the Duin and Ragsdale groups, respectively, which may be exploited for the development of biomimetic catalysts for methane activation.[11,74]

Alternatively, a metabolic engineering approach can be envisioned in which the critical enzymes of the AOM pathway are introduced into a more industrially viable strain. The feasibility of such a scenario was recently tested by Wood and coworkers, who cloned the ANME-1 MCR into *M. acetivorans* and demonstrated the anaerobic growth of this strain on methane using $FeCl_3$ as an external electron acceptor.[75] However, to extend this strategy beyond methanogen hosts will require significant advances in the understanding of the genetics and biochemistry of ANME and the formation of *holo* MCR. A major step toward this goal was achieved by the elucidation of the cfb pathway by Mansoorabadi and coworkers.[26] However, much

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488



Scheme 8 Potential alternative reaction catalyzed by methanotrophic MCR

work remains before MCR can be produced heterologously in an active form in a nonmethanogen host, particularly with regard to understanding how coenzyme F430 and the PTMs are introduced into the active site of MCR.

6 RELATED ARTICLES

FeS Cluster Assembly: NIF System in Nitrogen-Fixing Bacteria; Assembly of Nitrogenase FeMoco and P Cluster; Heme Biosynthesis and Insertion; Siroheme Assembly and Insertion into Nitrite and Sulfite Reductase; Cobalamin Biosynthesis and Insertion. Nickel Enzymes & Cofactors; Methane–to–Methanol Conversion; Methyl-Coenzyme M Reductase.

7 ABBREVIATIONS AND ACRONYMS

ANME	anaerobic methanotrophic archaea
AOM	anaerobic oxidation of methane
ATP	adenosine 5′-triphosphate
CoB-SH	coenzyme B
CoM-SH	coenzyme M
COR	chlorophyllide *a* oxidoreductase
cfb	coenzyme F430 biosynthesis
DIET	direct interspecies electron transfer
DPOR	dark-operative protochlorophyllide *a* oxidoreductase
GTL	gas-to-liquid
GWP	global warming potential
LCA	last common ancestor
MCR	methyl-coenzyme M reductase
MeS-CoM	methyl-coenzyme M
MS	mass spectrometry
$NAD(P)^+$	nicotinamide adenine dinucleotide (phosphate)
NMR	nuclear magnetic resonance
PTMs	posttranslational modifications
phom	phomopsin
r.m.s.d.	root-mean-square deviation
SAM	*S*-adenosyl-L-methionine
SRB	sulfate-reducing bacteria
THMPT	tetrahydromethanopterin
tva	thioviridamide

8 REFERENCES

1. R. K. Thauer, A. K. Kaster, H. Seedorf, W. Buckel and R. Hedderich, *Nat. Rev. Microbiol.*, 2008, **6**, 579.
2. J. G. Ferry, *Science*, 1997, **278**, 1413.

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488

3. Core Writing Team, R. K. Pachauri, and L. A. Meyer, IPCC, 2014: Climate Change 2014: Synthesis Report, IPCC, Geneva, Switzerland, 2014.
4. J. G. Ferry, *FEMS Microbiol. Rev.*, 1999, **23**, 13.
5. J. Ellermann, R. Hedderich, R. Böcher and R. K. Thauer, *Eur. J. Biochem.*, 1988, **172**, 669.
6. U. Ermler and W. Grabarse, *Science*, 1997, **278**, 1457.
7. G. Diekert, R. Jaenchen and R. K. Thauer, *FEBS Lett.*, 1980, **119**, 118.
8. H. Won, M. F. Summers, K. D. Olson and R. S. Wolfe, *J. Am. Chem. Soc.*, 1990, **112**, 2178.
9. J. Kahnt, B. Buchenau, F. Mahlert, M. Krüger, S. Shima and R. K. Thauer, *FEBS J.*, 2007, **274**, 4913.
10. M. Goubeaud, G. Schreiner and R. K. Thauer, *Eur. J. Biochem.*, 1997, **243**, 110.
11. T. Wongnate, D. Sliwa, B. Ginovska, D. Smith, M. W. Wolf, N. Lehnert, S. Raugei and S. W. Ragsdale, *Science*, 2016, **352**, 953.
12. S. Shima and R. K. Thauer, *Curr. Opin. Microbiol.*, 2005, **8**, 643.
13. S. Scheller, M. Goenrich, R. Boecher, R. K. Thauer and B. Jaun, *Nature*, 2010, **465**, 606.
14. K. Knittel and A. Boetius, *Annu. Rev. Microbiol.*, 2009, **63**, 311.
15. S. J. Hallam, N. Putnam, C. M. Preston, J. C. Detter, D. Rokhsar, P. M. Richardson and E. F. DeLong, *Science*, 2004, **305**, 1457.
16. S. L. Caldwell, J. R. Laidler, E. A. Brewer, J. O. Eberly, S. C. Sandborgh and F. S. Colwell, *Environ. Sci. Technol.*, 2008, **42**, 6791.
17. G. Wegener, V. Krukenberg, D. Riedel, H. E. Tegetmeyer and A. Boetius, *Nature*, 2015, **526**, 587.
18. K. Knittel, T. Lösekann, A. Boetius, R. Kort and R. Amann, *Appl. Environ. Microbiol.*, 2005, **71**, 467.
19. S. E. McGlynn, G. L. Chadwick, C. P. Kempes and V. J. Orphan, *Nature*, 2015, **526**, 531.
20. S. Shima, M. Krueger, T. Weinert, U. Demmer, J. Kahnt, R. K. Thauer and U. Ermler, *Nature*, 2011, **481**, 98.
21. S. Mayr, C. Latkoczy, M. Krüger, D. Günther, S. Shima, R. K. Thauer, F. Widdel and B. Jaun, *J. Am. Chem. Soc.*, 2008, **130**, 10758.
22. M. J. Warren and A. G. Smith, 'Tetrapyrroles - Birth, Life and Death', Springer-Verlag, New York, 2009.
23. H. Mucha, E. Keller, H. Weber, F. Lingens and W. Trösch, *FEBS Lett.*, 1985, **190**, 169.
24. S. Bali, A. D. Lawrence, S. A. Lobo, L. M. Saraiva, B. T. Golding, D. J. Palmer, M. J. Howard, S. J. Ferguson and M. J. Warren, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 18260.
25. A. Pfaltz, A. Kobelt, R. Hüster and R. K. Thauer, *Eur. J. Biochem.*, 1987, **170**, 459.
26. K. Zheng, P. D. Ngo, V. L. Owens, X.-p. Yang and S. O. Mansoorabadi, *Science*, 2016, **354**, 339.
27. A. A. Brindley, E. Raux, H. K. Leech, H. L. Schubert and M. J. Warren, *J. Biol. Chem.*, 2003, **278**, 22388.
28. V. Fresquet, L. Williams and F. M. Raushel, *Biochemistry*, 2004, **43**, 10619.
29. M. M. Georgiadis, H. Komiya, P. Chakrabarti, D. Woo, J. J. Kornuc and D. C. Rees, *Science*, 1992, **257**, 1653.
30. O. Einsle, F. A. Tezcan, S. L. Andrade, B. Schmid, M. Yoshida, J. B. Howard and D. C. Rees, *Science*, 2002, **297**, 1696.
31. F. Pratviel-Sosa, D. Mengin-Lecreulx and J. van Heijenoort, *Eur. J. Biochem.*, 1991, **202**, 1169.
32. C. R. Staples, S. Lahiri, J. Raymond, L. Von Herbulis, B. Mukhophadhyay and R. E. Blankenship, *J. Bacteriol.*, 2007, **189**, 7392.
33. B. M. Hoffman, D. Lukoyanov, Z. Y. Yang, D. R. Dean and L. C. Seefeldt, *Chem. Rev.*, 2014, **114**, 4041.
34. Y. Hu and M. W. Ribbe, *J. Biol. Inorg. Chem.*, 2015, **20**, 435.
35. N. Muraki, J. Nomata, K. Ebata, T. Mizoguchi, T. Shiba, H. Tamiaki, G. Kurisu and Y. Fujita, *Nature*, 2010, **465**, 110.
36. J. Nomata, T. Mizoguchi, H. Tamiaki and Y. Fujita, *J. Biol. Chem.*, 2006, **281**, 15021.
37. R. Fani, R. Gallo and P. Lió, *J. Mol. Evol.*, 2000, **51**, 1–11.
38. J. Raymond, J. L. Siefert, C. R. Staples and R. E. Blankenship, *Mol. Biol. Evol.*, 2004, **21**, 541.
39. M. Bokranz and A. Klein, *Nucleic Acids Res.*, 1987, **15**, 4350.
40. B. A. Sherf and J. N. Reeve, *J. Bacteriol.*, 1990, **172**, 1828.
41. K. D. Allen, G. Wegener and R. H. White, *Appl. Environ. Microbiol.*, 2014, **80**, 6403.
42. H. J. Sofia, G. Chen, B. G. Hetzler, J. F. Reyes-Spindola and N. E. Miller, *Nucleic Acids Res.*, 2001, **29**, 1097.
43. R. U. Hutcheson and J. B. Broderick, *Metallomics*, 2012, **4**, 1149.
44. W. Grabarse, F. Mahlert, S. Shima, R. K. Thauer and U. Ermler, *J. Mol. Biol.*, 2000, **303**, 329.
45. T. Wagner, J. Kahnt, U. Ermler and S. Shima, *Angew. Chem. Int. Ed.*, 2016, **55**, 10630.
46. T. Selmer, J. Kahnt, M. Goubeaud, S. Shima, W. Grabarse, U. Ermler and R. K. Thauer, *J. Biol. Chem.*, 2000, **275**, 3755.
47. R. Jaenchen, G. Diekert and R. K. Thauer, *FEBS Lett.*, 1981, **130**, 133.
48. A. Rios, T. L. Amyes and J. P. Richard, *J. Am. Chem. Soc.*, 2000, **122**, 9373.
49. F. G. Bordwell, S. Zhang, X.-M. Zhang and W.-Z. Liu, *J. Am. Chem. Soc.*, 1995, **117**, 7092.
50. J. A. Gerlt and P. G. Gassman, *Biochemistry*, 1993, **32**, 11943.
51. Q. Zhang, W. A. van der Donk and W. Liu, *Acc. Chem. Res.*, 2012, **45**, 555.

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488

52. M. R. Bauerle, E. L. Schwalm and S. J. Booker, *J. Biol. Chem.*, 2015, **290**, 3995.
53. Y. Hayakawa, K. Sasaki, H. Adachi, K. Furihata, K. Nagai and K. Shin-ya, *J. Antibiot.*, 2006, **59**, 1.
54. Y. Hayakawa, K. Sasaki, K. Nagai, K. Shin-ya and K. Furihata, *J. Antibiot*, 2006, **59**, 6.
55. M. Izawa, T. Kawasaki and Y. Hayakawa, *Appl. Environ. Microbiol.*, 2013, **79**, 7110.
56. K. L. Dunbar, J. R. Chekan, C. L. Cox, B. J. Burkhart, S. K. Nair and D. A. Mitchell, *Nat. Chem. Biol.*, 2014, **10**, 823.
57. K. L. Dunbar, J. O. Melby and D. A. Mitchell, *Nat. Chem. Biol.*, 2012, **8**, 569.
58. B. Breil, J. Borneman and E. W. Triplett, *J. Bacteriol.*, 1996, **178**, 4150.
59. C. T. Jurgenson, T. P. Begley and S. E. Ealick, *Annu. Rev. Biochem.*, 2009, **78**, 569.
60. K. A. Black and P. C. Dos Santos, *Biochim. Biophys. Acta*, 2015, **1853**, 1470.
61. P. M. Palenchar, C. J. Buck, H. Cheng, T. J. Larson and E. G. Mueller, *J. Biol. Chem.*, 2000, **275**, 8283.
62. Y. Liu, X. Zhu, A. Nakamura, R. Orlando, D. Söll and W. B. Whitman, *J. Biol. Chem.*, 2012, **287**, 36683.
63. M. E. Smith and D. M. Greenberg, *Nature*, 1956, **177**, 1130.
64. C. C. J. Culvenor, A. B. Beck, M. Clarke, P. A. Cockrum, J. A. Edgar, J. L. Frahn, M. V. Jago, G. W. Lanigan, A. L. Payne, J. E. Peterson, D. S. Petterson, L. W. Smith and R. R. White, *Aust. J. Biol. Sci.*, 1977, **30**, 269.
65. W. Ding, W. Q. Liu, Y. Jia, Y. Li, W. A. van der Donk and Q. Zhang, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 3521.
66. R. K. Thauer, *Curr. Opin. Microbiol.*, 2011, **14**, 292.
67. P. R. Girguis, V. J. Orphan, S. J. Hallam and E. F. DeLong, *Appl. Environ. Microbiol.*, 2003, **69**, 5472.
68. K. Nauhaus, M. Albrecht, M. Elvert, A. Boetius and F. Widdel, *Environ. Microbiol.*, 2007, **9**, 187.
69. M. Krüger, A. Meyerdierks, F. O. Glöckner, R. Amann, F. Widdel, M. Kube, R. Reinhardt, J. Kahnt, R. Böcher, R. K. Thauer and S. Shima, *Nature*, 2003, **426**, 878.
70. R. K. Thauer, *Microbiology*, 1998, **144**, 2377.
71. U.S. Department of Energy, 'Modern Shale Gas Development in the United States: A Primer', Office of Fossil Energy, National Energy Technology Laboratory, 2009.
72. C. A. Haynes and R. Gonzalez, *Nat. Chem. Biol.*, 2014, **10**, 331.
73. T. J. Mueller, M. J. Grisewood, H. Nazem-Bokaee, S. Gopalakrishnan, J. G. Ferry, T. K. Wood and C. D. Maranas, *J. Ind. Microbiol. Biotechnol.*, 2015, **42**, 391.
74. D. Prakash, Y. Wu, S. J. Suh and E. C. Duin, *J. Bacteriol.*, 2014, **196**, 2491.
75. V. W. Soo, M. J. McAnulty, A. Tripathi, F. Zhu, L. Zhang, E. Hatzakis, P. B. Smith, S. Agrawal, H. Nazem-Bokaee, S. Gopalakrishnan, H. M. Salis, J. G. Ferry, C. D. Maranas, A. D. Patterson and T. K. Wood, *Microb. Cell Fact.*, 2016, **15**, 11.

Encyclopedia of Inorganic and Bioinorganic Chemistry, Online © 2011–2017 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Inorganic and Bioinorganic Chemistry* in 2017 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781119951438.eibc2488