THE UNIVERSITY OF CHICAGO

PHYLOGENETIC AND MOLECULAR DIVERSITY OF FLAVINYLATED PROTEINS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOCHEMISTRY AND MOLECULAR BIOPHYSICS

BY

SHUO HUANG

CHICAGO, ILLINOIS

DECEMBER 2024

Table of Contents

List of Publications

Huang S, Méheust R, Barquera B, Light SH. 2024. Versatile roles of protein flavinylation in bacterial extracyotosolic electron transfer. *mSystems* **9**:e00375-24. doi:10.1128/msystems.00375-24

Méheust R, Huang S, Rivera-Lugo R, Banfield JF, Light SH. 2021. Post-translational flavinylation is associated with diverse extracytosolic redox functionalities throughout bacterial life. *eLife* **10**:e66878. doi:10.7554/eLife.66878

List of Figures

List of Tables

Acknowledgements

I would like to thank my thesis advisor, Dr. Sam Light. His unwavering support and expertise have been instrumental in shaping my academic journey and the completion of this dissertation. I would also like to thank members of my thesis committee, Dr. Anthony Kossiakoff, Dr. Minglei Zhao, and Dr. Sampriti Mukherjee, who, in addition to their time, effort, and commitment to my success, provided guidance beyond the confines of academia.

Completion of this thesis would not be possible without the support from our colleagues, and I would like to thank them for their contribution. I would like to express my gratitude to Dr. Raphaël Méheust, our collaborator and a former postdoctoral fellow at the Banfield lab at University of California, Berkeley. His expertise in bioinformatic analyses is integral to the findings in this dissertation. I would also like to thank our administrative assistant, Noreen Bently; former research technician and current graduate student, Issac Younker; and research technician, Josh Stemczynski, for their invaluable assistance in ensuring smooth daily operations in the Light lab. I would like to extend my gratitude to Dr. Jerry Zhang, a postdoctoral fellow from Dr. Eric Pamer's laboratory, for providing me with various training as a colleague and offering valuable advice in my research career as a friend. The successful completion of this dissertation is, of course, owed to the staff and faculty of my graduate program in Biochemistry and Molecular Biophysics, the Department of Microbiology, and the Biological Sciences Division, with special thanks to Shani Charles and Giovanni Tolentino Ramos for their indispensable administrative assistance. And finally, I would like to thank my husband, Jun Zeng, and my family for their unconditional love and support throughout my graduate studies and beyond.

Abstract

Electron transfer across the cell envelope is a critical process in all cellular organisms. Proteins with redox-active cofactors construct insulated circuits that span the cell envelope, connecting cytoplasmic electron pools to electron acceptors in the extracytosolic space. Proteins that use covalently attached flavin mononucleotide moieties (FMN) as redox-active cofactors have been recently reported by several studies to be involved in extracytosolic electron transfer. These proteins are substrates for the FMN transferase ApbE, which post-translationally attaches FMN to its targets through a process termed flavinylation. Yet, there exists a great paucity of information on the prevalence and molecular basis of flavinylated proteins. We aimed to provide a better understanding of the biological and biochemical contexts of ApbE-mediated flavinylation. To this end, we conducted comprehensive studies with three primary goals: 1) evaluate phylogenetic diversity of flavinylated proteins, 2) determine structural and functional properties of flavinylated proteins, and 3) assess relevance of flavinylated proteins to human health.

Our work reveals that ApbE-flavinylated proteins are highly prevalent among prokaryotic organisms, with implicated roles in crucial cellular processes, such as anaerobic respiration or iron assimilation. Comparative genomics and computational structural biology further demonstrate similar versatility of flavinylated proteins in their predicted structures and functions. We identify and experimentally confirm previously unknown proteins as novel ApbE substrates. We also identify flavinylated proteins encoded by pathogenic bacterial species, confirming their FMN-binding activity and mode of electron transfer in vitro. And finally, we provide preliminary assessment of prevalence of flavinylation in publicly available human gut microbiome data. Our work thus expands the breadth of our knowledge on flavinylation-mediated extracytosolic electron transfer, revealing a new, common facet of prokaryotic redox physiology.

Chapter 1: Introduction

1.1 Mechanisms of extracytosolic electron transfer

The capability of electron transfer is crucial to all cellular lifeforms. Electron transfer occurs within multiple cellular compartments. In addition to redox reactions in the central metabolism within the cytoplasm, many cellular processes involve the transfer of electrons across the cell envelope via extracytosolic electron transfer (EET). This includes respiration, where electrons are shuttled from cytoplasmic electron donors to extracytosolic electron acceptors. EET can also be energetically coupled with other membrane systems to facilitate the uptake of external resources, such as the assimilation of iron. In Gram-negative bacteria, EET is also responsible for supplying electrons in the periplasm, aiding in proper folding of periplasmic proteins (Bertini et al., 2006; Cho and Collet, 2013; Landeta et al., 2018; Schröder et al., 2003).

Translocation of electrons into periplasm or extracytosolic space depends on insulated circuits that span across the cell envelope (**Figure 1.1A**). These circuits are composed of proteins that possess electron-transferring cofactors. Of the most well-known examples are cytochromes, a class of proteins that leverages heme cofactors for electron transfer (Bertini et al., 2006). In the Gram-negative bacteria *Shewanella oneidensis*, the MtrABC complex encodes two deca-heme cytochrome subunits that together facilitate respiration on extracytosolic ferric iron (**Figure 1.1B**) (Edwards et al., 2020). In more extreme cases such as nanowires in *Geobacter* species, hundreds of cytochrome proteins collectively form conductive pili that sequentially transfer electrons onto insoluble iron oxides (**Figure 1.1C**) (Wang et al., 2019). The complexity of these systems highlights the importance of securing paths of EET when proper electron acceptors are either distant or scarce.

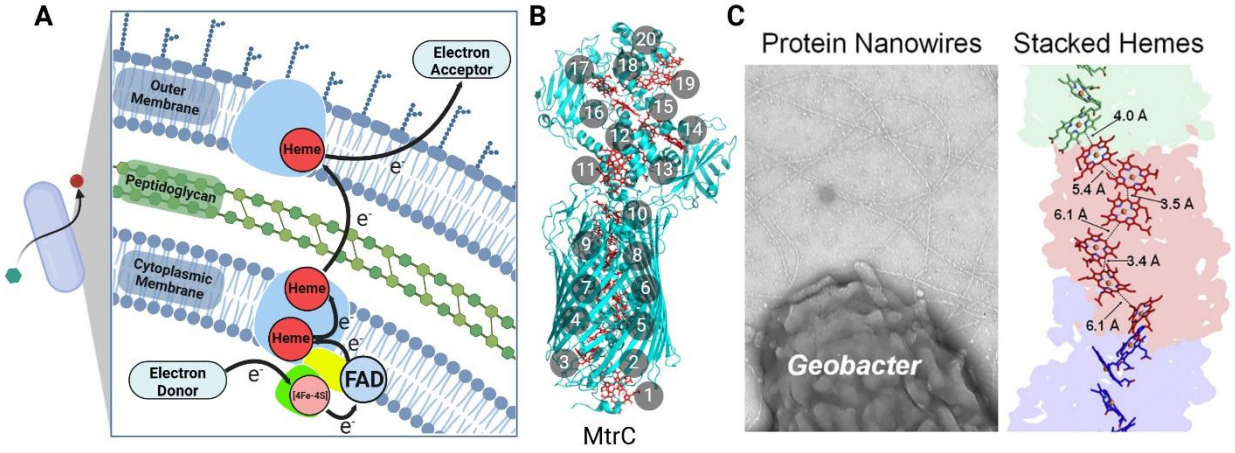


Figure 1.1 Extracytosolic electron transfer via redox-active cofactors.

(A) Schematic diagram of extracytosolic electron transfer. Circles represent electron-transferring protein cofactors. Black arrows indicate directions of electron flow. (B) Previously resolved crystal structure of MtrABC complex from *Shewanella oneidensis*. Structure visualization was adopted from Edwards et al., 2020 (PDB: 6R2Q) (Edwards et al., 2020). (C) Cryo-EM and crystal structures of nanowires in *Geobacter sulfurreducens* from Wang et al., 2019 (PDB: 6EF8) (Wang et al., 2019).

Previous efforts have investigated and exploited the molecular basis of EET. Numerous studies reported involvement of EET in biofilm formation and nutrient acquisition in pathogenic bacteria during host colonization (Keogh et al., 2018; Liu et al., 2018; Manck et al., 2020). These mechanistic insights of EET have been translated into biotechnological applications, leading to development of microbial fuel cells and bioremediation strategies (Nealson, 2017; Wang et al., 2020). Thus, better understanding of EET proteins would promote the betterment of both human and environmental health.

1.2 Flavinylation: a novel post-translational modification for electron transfer

Like heme, flavins are a group of molecules that facilitate electron transfer through transitions in their redox states. The most well-known flavin molecule is riboflavin or vitamin B2, which serves as a precursor for two other flavins, namely flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). Conserved in all flavins, a redox-reactive isoalloxazine ring structure manifests electron-transferring capabilities through transitions between one- or two-electron redox states. Both riboflavin and FAD are well known to serve as electron-transferring cofactors integral to EET proteins in a wide range of organisms (**Figure 1.2A**) (Fraaije and Mattevi, 2000).

Compared to riboflavin and FAD, much less is known on whether FMN similarly partakes in EET activities. A previous report highlighted the possible involvement of conserved FMN-binding domains in electron transfer (Bogachev et al., 2018; Zhou et al., 1999). The FMN-binding domain serves as the target for FMN flavinylation, a post-translational modification carried out by the alternative pyrimidine biosynthesis protein ApbE. During flavinylation, ApbE transfers the FMN moiety from FAD onto the second Ser residue in the conserved FMN-binding [S/T]GA[S/T] motif, generating a covalent phosphoester bond between the ribitylphosphate group of FMN and

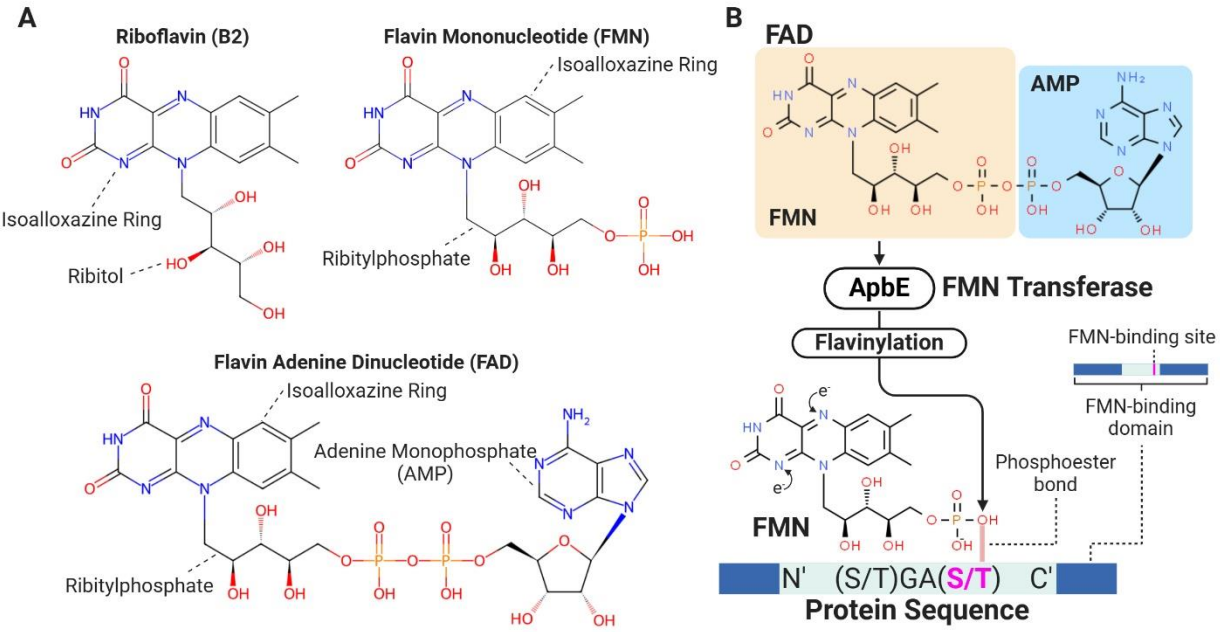


Figure 1.2 Schematic representation of ApbE-mediated flavinylation.
(A) molecular structures of flavin molecules. (B) Process of ApbE-mediated flavinylation. Serine and threonine residues colored in purple indicate the site of covalent FMN attachment. Sequence containing FMN_bind domain is highlighted in dark blue, which encompasses the conserved 5-residue FMN-binding motif colored in light blue. Position of redox-active sites of the isoalloxazine ring of FMN are highlighted by arrows connected to electrons.

side chain of Ser/Thr residue (**Figure 1.2B**) (Bertsova et al., 2013). This type of covalent linkage to amino acid side chain is unique to ApbE-mediated flavinylation and differs covalent linkage of riboflavin or FAD. For riboflavin and FAD, a covalent bond is formed on the redox-reactive isoalloxazine ring, leading to a change in redox potential that is important for the protein's activity (Macheroux et al., 2011). In addition, non-covalent FMN binding-domains have lower affinity for flavins (Backiel et al., 2008; Barquera, 2014; Borshchevskiy et al., 2015). When taken together, these insights suggest that ApbE-mediated flavinylation could secure the FMN moiety within the protein structure and enhance electron transfer.

More recent studies have demonstrated the involvement of flavinylated proteins in EET activities. In the human enteric pathogen *Listeria monocytogenes*, a double-flavinylated PplA protein was found to be responsible for ferric iron respiration (**Figure 1.3A**) (Light et al., 2018). PplA is anchored on the extracytosolic face of the outer membrane. It interacts with other EET proteins encoded in the same locus, such as the NADH dehydrogenase NDH-2, to shuttle electrons originating from cytoplasmic NADH. Interestingly, NDH-2 was previously shown to interact with other EET proteins, such as cytochromes, to facilitate reduction of different terminal electron acceptors like nitrate (Lencina et al., 2018). This suggests that flavinylated proteins are functionally similar to other EET proteins, providing alternative strategies for respiratory redox. Furthermore, disruption of PplA and NDH-2 led to reduced *L. monocytogenes* growth under anaerobic conditions and decreased colonization in mice (Light et al., 2018). These insights demonstrate that flavinylation-mediated EET activity could be crucial for fitness and virulence of bacterial species that are relevant for human health.

Flavinylated proteins are also involved in large EET complexes, facilitating electron transfer internally between different subunits. This is most evidently shown in the

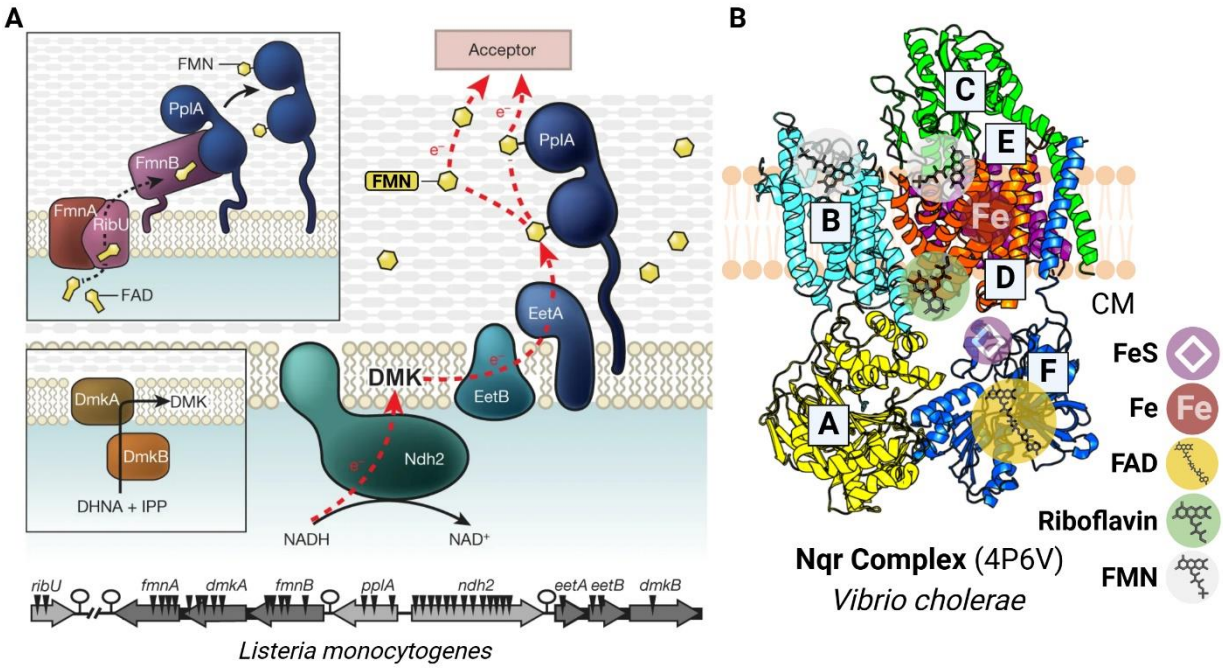


Figure 1.3 Previously characterized cellular contexts of ApbE-flavinylated proteins.
(A) Proposed mode of electron transport mediated by the double-flavinylated PplA protein in *Listeria monocytogenes* from Light et al., 2018 (Light et al., 2018). FMN molecules are presented by yellow hexagons. Ndh2: NADH dehydrogenase. DMK: demethylmenaquinone, synthesized by DmkA and DmkB. FmnA and RibU: FAD transporters. FmnB: equivalent to ApbE. EetA and EetB: uncharacterized protein with suggested roles in flavin transport (Rivera-Lugo et al., 2023). (B) Crystal structure of the NADH:ubiquinone oxidoreductase complex (Nqr). Visualization of the protein structures were derived from Kishikawa et al., 2022 (PDB: 4P6V) (Kishikawa et al., 2022). Distinct subunits are in different colors. Circular labels represent different redox-active cofactors in each subunit.

NADH:ubiquinone oxidoreductase complex, or Nqr in short, that is encoded by many Gram-negative bacteria such as the pathogen *Vibrio cholerae* (**Figure 1.3B**) (Kishikawa et al., 2022). This complex contains 6 subunits, with the B and C subunits flavinylated. The Nqr complex couples EET with translocation of sodium ions, creating a sodium membrane gradient that is required for other cellular processes. This is achieved through sequential transfer of electrons from cytoplasmic NADH to a FAD cofactor and iron-sulfur clusters on NqrF, FMN of NqrC, FMN and riboflavin of NqrB, and finally onto ubiquinone. Homologous to the Nqr complex, the ferredoxin:NAD^+ oxidoreductase (Rnf) complex found in other Gram-negative bacteria also preserved similar electron circuits (Backiel et al., 2008; Kuhns et al., 2020). These findings demonstrate that flavinylated proteins could be highly cooperative with other electron-transferring components in large EET complexes.

1.3 Understanding the prevalence and molecular basis of flavinylation

Despite these known involvements of flavinylated proteins in critical cellular activities, there remains a great paucity of information on the prevalence of flavinylation in prokaryotic organisms. Several lines of evidence suggest that flavinylation is common in a wide range of bacteria. First, previous studies have reported flavinylated proteins in bacterial species that are phenotypically and evolutionarily distinct, indicating versatile roles of flavinylation under different microbial habitats (Fraaije and Mattevi, 2000; Zhou et al., 1999). Secondly, known flavinylated proteins often interact with other membrane EET systems with great phylogenetic diversity. Members of the Nqr and NDH-2 protein families are distributed widely across major bacterial phyla, with the latter also found in archaeal and eukaryotic microorganisms (Barquera,

2014; Borshchevskiy et al., 2015; Lencina et al., 2018). These observations from pioneering research elucidated that flavinylation can situate in diverse biological contexts.

This auspicious versatility could also be reflected in both functional structural diversity of flavinylated proteins. This is evident in known EET proteins such as membrane cytochrome proteins, which could act as a monomeric protein, form complex with membrane complex, or comprise highly intricate structures like nanowires. However, only a few studies reported this molecular versatility of flavinylated proteins. While in silico prediction of protein function could be performed in a high throughput manner using functional annotation tools, comparative structural analysis of flavinylated proteins at large scale is a major rate-limiting step. Traditionally, structural analysis of proteins requires preparation of protein samples for either X-ray crystallography or cryo-electron microscopy, which could be a laborious process when the number of proteins of interest is high. Previous in silico alternatives for structural prediction are also either slow or inaccurate, making them less reliable, especially for models involving multiple proteins in complex. Addressing these limitations would promote our understanding of the molecular basis of flavinylation.

1.4 Goals of this thesis

The main objectives of this thesis are to characterize the phylogenetic and molecular versatility of flavinylated proteins. In doing so, we would be able to assess pivotal roles of flavinylation in microbial redox physiology. Understanding how flavinylated proteins structurally interact with other EET components would reveal the path of electron flow across the cell envelope. We will also be able to leverage the versatility of flavinylated proteins to identify previously unexamined EET components that are associated with flavinylation.

In **Chapter 2**, we present our bioinformatic survey of prokaryotic genomes for presence of flavinylation. This approach uses genetic markers that are specific to flavinylation, revealing a wide distribution of flavinylation in prokaryotic life. We additionally report the identification of proteins with previously unknown function as novel substrate of flavinylation. In **Chapter 3**, we describe our findings on large scale comparative analysis on both structures and functions of flavinylated proteins. Using genomic annotations in combination with the AI-powered AlphaFold for efficient structural predictions (Evans et al., 2021; Jumper et al., 2021; Mirdita et al., 2022), we reveal structural and functional diversity of flavinylated proteins. This confirms our hypothesis that phylogenetic versatility of flavinylation is reflected by molecular diversity of flavinylated proteins. In addition, we will demonstrate how interactions between flavinylated proteins and other EET components could be exploited to identify previously uncharacterized EET proteins. And finally, **Chapter 4** will provide a summary of preliminary exploration of ApbE-mediated flavinylation in different contexts, including mechanisms of flavinylation-mediated electron transfer and prevalence of flavinylation in the human gut microbiome.

In summary, our work unveils the versatility of flavinylation in a wide range of biological and molecular contexts. This work establishes a foundation for future efforts that aim to uncover mechanisms or biotechnological potentials of flavinylation-mediated electron transfer.

1.5 References

Backiel J, Zagorevski DV, Wang Z, Nilges MJ, Barquera B. 2008. Covalent Binding of Flavins to RnfG and RnfD in the Rnf Complex from *Vibrio cholerae*. *Biochemistry* **47**:11273–11284. doi:10.1021/bi800920j

Barquera B. 2014. The sodium pumping NADH:quinone oxidoreductase (Na+-NQR), a unique redox-driven ion pump. *J Bioenerg Biomembr* **46**:289–298. doi:10.1007/s10863-014-9565-9

Bertini I, Cavallaro G, Rosato A. 2006. Cytochrome *c* : Occurrence and Functions. *Chem Rev* **106**:90–115. doi:10.1021/cr050241v

Bertsova YV, Fadeeva MS, Kostyrko VA, Serebryakova MV, Baykov AA, Bogachev AV. 2013. Alternative Pyrimidine Biosynthesis Protein ApbE Is a Flavin Transferase Catalyzing Covalent Attachment of FMN to a Threonine Residue in Bacterial Flavoproteins. *Journal of Biological Chemistry* **288**:14276–14286. doi:10.1074/jbc.M113.455402

Bogachev AV, Baykov AA, Bertsova YV. 2018. Flavin transferase: the maturation factor of flavin-containing oxidoreductases. *Biochemical Society Transactions* **46**:1161–1169. doi:10.1042/BST20180524

Borshchevskiy V, Round E, Bertsova Y, Polovinkin V, Gushchin I, Ishchenko A, Kovalev K, Mishin A, Kachalova G, Popov A, Bogachev A, Gordeliy V. 2015. Structural and Functional Investigation of Flavin Binding Center of the NqrC Subunit of Sodium-Translocating NADH:Quinone Oxidoreductase from Vibrio harveyi. *PLoS ONE* **10**:e0118548. doi:10.1371/journal.pone.0118548

Cho S-H, Collet J-F. 2013. Many Roles of the Bacterial Envelope Reducing Pathways. *Antioxidants & Redox Signaling* **18**:1690–1698. doi:10.1089/ars.2012.4962

Edwards MJ, White GF, Butt JN, Richardson DJ, Clarke TA. 2020. The Crystal Structure of a Biological Insulated Transmembrane Molecular Wire. *Cell* **181**:665-673.e10. doi:10.1016/j.cell.2020.03.032

Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D. 2021. Protein complex prediction with AlphaFold-Multimer. doi:10.1101/2021.10.04.463034

Fraaije MW, Mattevi A. 2000. Flavoenzymes: diverse catalysts with recurrent features. *Trends in Biochemical Sciences* **25**:126–132. doi:10.1016/S0968-0004(99)01533-9

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. doi:10.1038/s41586-021-03819-2

Keogh D, Lam LN, Doyle LE, Matysik A, Pavagadhi S, Umashankar S, Low PM, Dale JL, Song Y, Ng SP, Boothroyd CB, Dunny GM, Swarup S, Williams RBH, Marsili E, Kline KA. 2018. Extracellular Electron Transfer Powers Enterococcus faecalis Biofilm Metabolism. *mBio* **9**:e00626-17. doi:10.1128/mBio.00626-17

Kishikawa J, Ishikawa M, Masuya T, Murai M, Kitazumi Y, Butler NL, Kato T, Barquera B, Miyoshi H. 2022. Cryo-EM structures of Na+-pumping NADH-ubiquinone oxidoreductase from Vibrio cholerae. *Nat Commun* **13**:4082. doi:10.1038/s41467-022-31718-1

Kuhns M, Trifunović D, Huber H, Müller V. 2020. The Rnf complex is a Na+ coupled respiratory enzyme in a fermenting bacterium, Thermotoga maritima. *Commun Biol* **3**:431. doi:10.1038/s42003-020-01158-y

Landeta C, Boyd D, Beckwith J. 2018. Disulfide bond formation in prokaryotes. *Nat Microbiol* **3**:270–280. doi:10.1038/s41564-017-0106-2

Lencina AM, Franza T, Sullivan MJ, Ulett GC, Ipe DS, Gaudu P, Gennis RB, Schurig-Briccio LA. 2018. Type 2 NADH Dehydrogenase Is the Only Point of Entry for Electrons into the Streptococcus agalactiae Respiratory Chain and Is a Potential Drug Target. *mBio* **9**:e01034-18. doi:10.1128/mBio.01034-18

Light SH, Su L, Rivera-Lugo R, Cornejo JA, Louie A, Iavarone AT, Ajo-Franklin CM, Portnoy DA. 2018. A flavin-based extracellular electron transfer mechanism in diverse Gram-positive bacteria. *Nature* **562**:140–144. doi:10.1038/s41586-018-0498-z

Liu MM, Boinett CJ, Chan ACK, Parkhill J, Murphy MEP, Gaynor EC. 2018. Investigating the Campylobacter jejuni Transcriptional Response to Host Intestinal Extracts Reveals the Involvement of a Widely Conserved Iron Uptake System. *mBio* **9**:e01347-18. doi:10.1128/mBio.01347-18

Macheroux P, Kappes B, Ealick SE. 2011. Flavogenomics – a genomic and structural view of flavin-dependent proteins. *The FEBS Journal* **278**:2625–2634. doi:10.1111/j.1742-4658.2011.08202.x

Manck LE, Espinoza JL, Dupont CL, Barbeau KA. 2020. Transcriptomic Study of Substrate-Specific Transport Mechanisms for Iron and Carbon in the Marine Copiotroph Alteromonas macleodii. *mSystems* **5**:e00070-20. doi:10.1128/mSystems.00070-20

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nat Methods* **19**:679–682. doi:10.1038/s41592-022-01488-1

Nealson KH. 2017. Bioelectricity (electromicrobiology) and sustainability. *Microbial Biotechnology* **10**:1114–1119. doi:10.1111/1751-7915.12834

Rivera-Lugo R, Huang S, Lee F, Méheust R, Iavarone AT, Sidebottom AM, Oldfield E, Portnoy DA, Light SH. 2023. Distinct Energy-Coupling Factor Transporter Subunits Enable Flavin Acquisition and Extracytosolic Trafficking for Extracellular Electron Transfer in Listeria monocytogenes. *mBio* **14**:e03085-22. doi:10.1128/mbio.03085-22

Schröder I, Johnson E, De Vries S. 2003. Microbial ferric iron reductases. *FEMS Microbiol Rev* **27**:427–447. doi:10.1016/S0168-6445(03)00043-3

Wang F, Gu Y, O'Brien JP, Yi SM, Yalcin SE, Srikanth V, Shen C, Vu D, Ing NL, Hochbaum AI, Egelman EH, Malvankar NS. 2019. Structure of Microbial Nanowires Reveals Stacked Hemes that Transport Electrons over Micrometers. *Cell* **177**:361-369.e10. doi:10.1016/j.cell.2019.03.029

Wang X, Aulenta F, Puig S, Esteve-Núñez A, He Y, Mu Y, Rabaey K. 2020. Microbial electrochemistry for bioremediation. *Environmental Science and Ecotechnology* **1**:100013. doi:10.1016/j.ese.2020.100013

Zhou W, Bertsova YV, Feng B, Tsatsos P, Verkhovskaya ML, Gennis RB, Bogachev AV, Barquera B. 1999. Sequencing and Preliminary Characterization of the Na^+-Translocating NADH:Ubiquinone Oxidoreductase from *Vibrio harveyi*. *Biochemistry* **38**:16246–16252. doi:10.1021/bi991664s

Chapter 2: Prevalence of flavinylation in prokaryotic life

Data included in this chapter are featured in a manuscript titled "Post-translational flavinylation is associated with diverse extracytosolic redox functionalities throughout bacterial life", which was accepted and published by *eLife*. Affiliations and contributions of authors are listed below.

Raphaël Méheust[1,2,3†], Shuo Huang[4,5†], Rafael Rivera-Lugo[6], Jillian F Banfield[1,2], Samuel H Light[4,5*]

[1]Department of Earth and Planetary Science, University of California, Berkeley, United States; [2]Innovative Genomics Institute, United States; [3]LABGeM, Génomique Métabolique, Genoscope, Institut François Jacob, CEA, France; [4]Duchossois Family Institute, University of Chicago, United States; [5]Department of Microbiology, University of Chicago, United States; [6]Department of Molecular and Cell Biology, University of California, Berkeley, United States. *Corresponding author. [†]Authors contributed equally.

R.M.: Conceptualization, Investigation, Writing - original draft; S.H.: Formal analysis, Investigation, Visualization, Writing - review and editing; R.R.L.: Investigation, Writing - review and editing; J.F.B.: Supervision, Funding acquisition, Writing - review and editing; S.H.L.: Conceptualization, Funding acquisition, Writing - original draft.

2.1 Abstract

Electron transfer across the cell envelope is a process critical to all cellular lifeforms. ApbE-mediated flavinylation post-translationally equips proteins with covalent, redox-active flavin mononucleotide cofactors for electron transfer. Several characterized electron transfer systems are associated with ApbE, but the biological prevalence of ApbE-mediated flavinylation remains unexplored. Here, we performed a comprehensive bioinformatic screen for genomic makers of ApbE-mediated flavinylation in over 30,000 prokaryotic genomes. We present that ApbE-mediated flavinylation is biologically versatile, encoded by ~50% of bacterial species. Genomic contexts of flavinylation machinery reveal associations with both characterized and unknown electron transfer systems. We identify flavinylated gene clusters that co-encode iron transporters and respiratory oxidoreductases, implying involvements of flavinylation in iron assimilation and respiratory electron transport chain. We additionally identify and characterize two

previously unknown, ApbE-associated proteins, namely DUF2271 and DUF3570, as novel flavinylated substrates for ApbE. Finally, we observed flavinylated gene clusters that functionally mimic cytochrome-associated gene clusters. This flavinylation/cytochrome modularity is also observed between multi-flavinylated proteins and multi-heme cytochromes, suggesting that higher multiplicity of FMN moieties could similarly promote electron transfer across a longer path. These findings provide mechanistic insight into an important facet of bacterial physiology and establish flavinylation as a functionally diverse mediator of extracytosolic electron transfer.

2.2 Introduction - Leveraging markers of flavinylation for comparative genomics

ApbE mediated flavinylation is found to play critical roles in electron transfer in five previously characterized flavinylated systems. These include Nqr and Rnf complexes (Backiel et al., 2008; Borshchevskiy et al., 2015; Zhou et al., 1999), which are mentioned in the previous chapter, nitrous oxide and organohalide respiratory complexes (Buttet et al., 2018; Zhang et al., 2017), and a Gram-positive extracellular electron transfer system (Light et al., 2019, 2018). These systems are widespread in microbial life, partaking in different aspects of cellular respiration and redox metabolism. Nonetheless, a systematic understanding of the prevalence of flavinylation is currently lacking.

Comparative genomics offers a path to address this gap in our knowledge. A substantial body of previous research demonstrate comparative genomics as a robust tool for high throughput identification and analysis of prokaryotic protein domain (Burstein et al., 2017; Crits-Christoph et al., 2018; Doron et al., 2018). Using known protein markers, one could efficiently query all functionally or genetically associated proteins encoded in publicly available genome sequences at a large scale. Downstream comparisons of protein functions, sequences, and phylogenetic

distributions could reveal the evolutionary history of protein domains of interest, identification of virulent factors in pathogenic species, or characterization of novel therapeutic targets.

Here, we applied comparative genomics to determine the prevalence of flavinylation among prokaryotic organisms. Leveraging the two known markers of flavinylation, namely the FMN transferase ApbE and the conserved FMN-binding domain, we bioinformatically surveyed over 30,000 prokaryotic genomes. We present evidence that flavinylation is widespread in prokaryotic life. Functional analysis of genetic elements co-encoded with flavinylation markers reveal that flavinylated proteins would participate in diverse EET activities, including respiration and iron assimilation. From genomes that encode ApbE but lack identifiable FMN-binding domains, we also characterized two domains with unknown functions as novel ApbE substrates. Taken together, these findings widened the breadth of our knowledge on flavinylation as a common redox strategy in prokaryotic organisms.

2.3 Results

2.3.1 Extracytosolic flavinylation is widespread in prokaryotic genomes

To identify flavinylated systems, we searched for genes encoding the ApbE domain (Pfam accession: PF02424) and/or the FMN-binding domain (Pfam accession: PF04205) in 31,910 prokaryotic genomes (**Figure 2.1**) (Parks et al., 2020, 2018). We found that 18,965 of bacterial and 238 of archaeal genomes encode either the ApbE enzyme and/or a protein containing the FMN-bidning domain. We additionally limited our search to proteins involved in extracytosolic transfer by removing proteins that lack a predicted signal peptide or lipidation site, which are required for membrane secretion. This results in 15,095 bacterial and 63 archaeal genomes that are positive for extracytosolic flavinylation. Phylogenetic analysis of these genomes reveals a wide

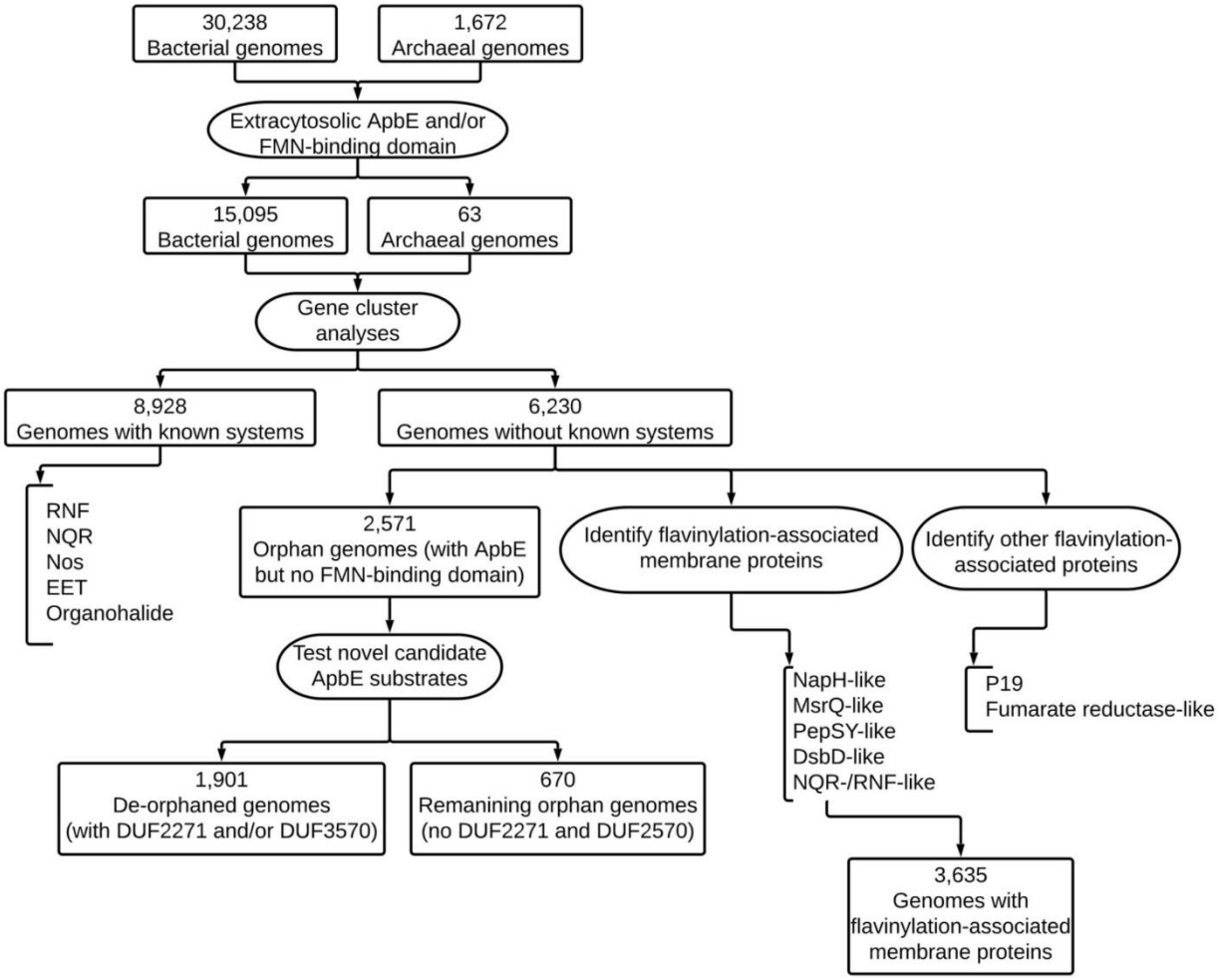


Figure 2.1 Bioinformatic pipeline for identifying flavinylated gene clusters.

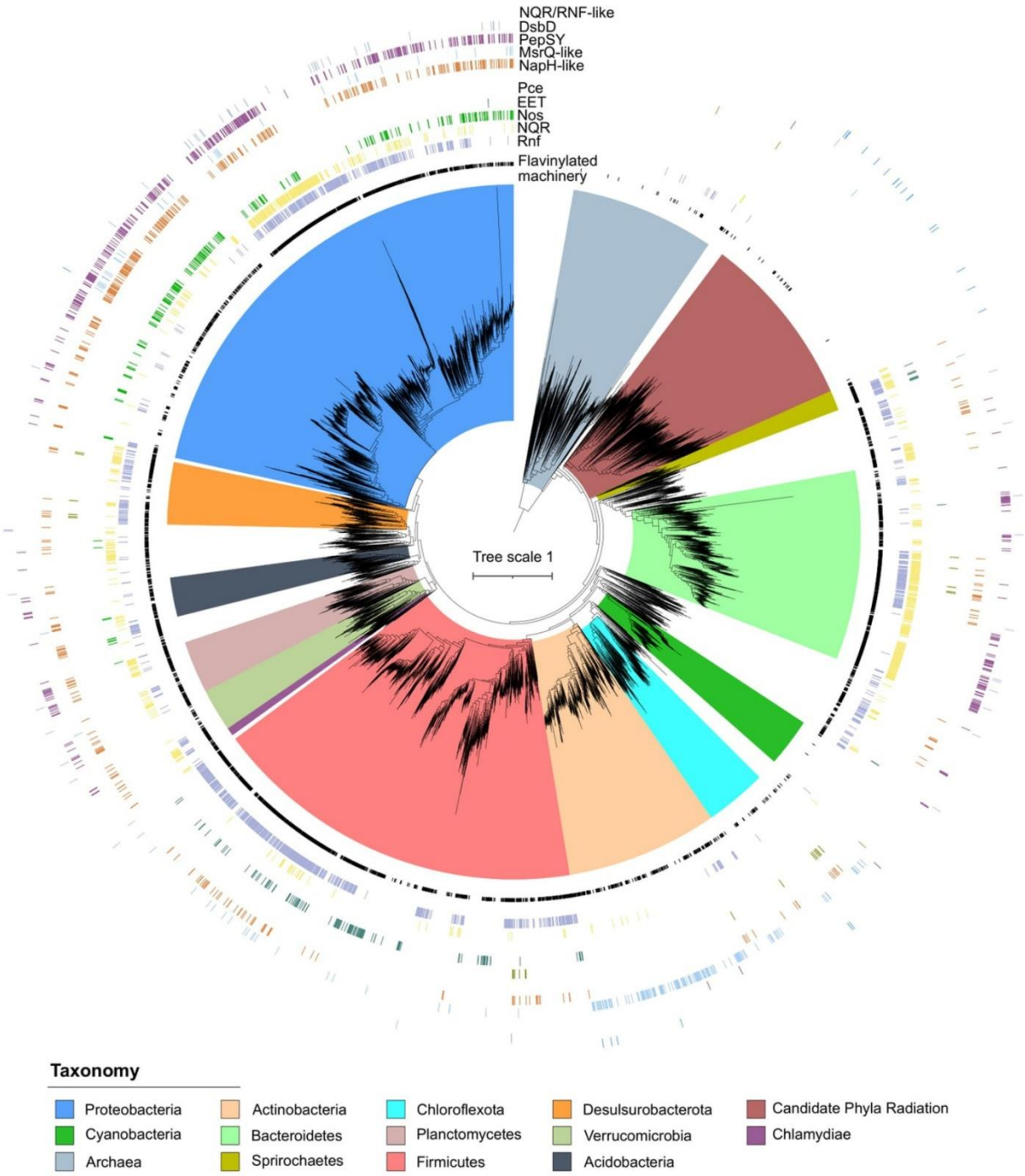


Figure 2.2 Phylogenetic distribution of genetic marks of ApbE-mediated flavinylation. Phylogenetic reconstruction of the evolutionary history of 9,152 genomes representing 97% of the diversity available at the genus level in the GTDB (9,428 distinct genera) (Parks et al., 2020). The maximum likelihood tree was constructed based on a concatenated alignment of 14 ribosomal proteins under an LG + I + G4 model of evolution (2,850 amino acid sites). The inner to outer rings represent the presence of flavinylation markers or electron-transfer systems. The scale bar indicates the mean number of substitutions per site.

distribution of flavinylation in major bacterial phyla, including Proteobacteria, Firmicutes, and Bacteroidetes that contain members closely associated with human health (**Figure 2.2**).

We also examined functional annotations of genes that co-encode in proximity with flavinylation markers in identified genomes. We found that in 8,929 genomes, flavinylation makers are co-encoded with a previously characterized electron transfer system. However, we observed that flavinylation markers are encoded in the absence of known systems in 6,230 genomes (**Figure 2.3A**). This suggests that a large proportion of flavinylated systems are unexamined by previous studies.

2.3.2 Characterization of DUF2271 and DUF3570 as novel ApbE substrates

In our initial bioinformatic screen, we used the presence of identifiable FMN-binding domain as the only criteria for selecting flavinylated protein candidates. However, we observed that 2,571 genomes included in our analysis exclusively encode only the FMN transferase ApbE but not FMN-binding domain (**Figure 2.3B**). This observation implies that some targets of ApbE-mediated flavinylation may lack a canonical FMN-binding domain.

To test this hypothesis, we examined genomic elements encoded in proximity to *apbE* genes in these genomes—hereinafter referred to as "orphan" apbE genes for their lack of identifiable flavinylation substrate. This led to the identification of a subset of genomes that encode an orphan *apbE* gene cluster associated with DUF2271, a ~135 amino acid domain of unknown function. Closer examination of sequences of DUF2271 domains reveals the presence of the conserved [S/T]GA[S/T] FMN-binding site, despite that gene segments containing this binding site are not recognized as a canonical FMN-binding domain (**Figure 2.4A**).

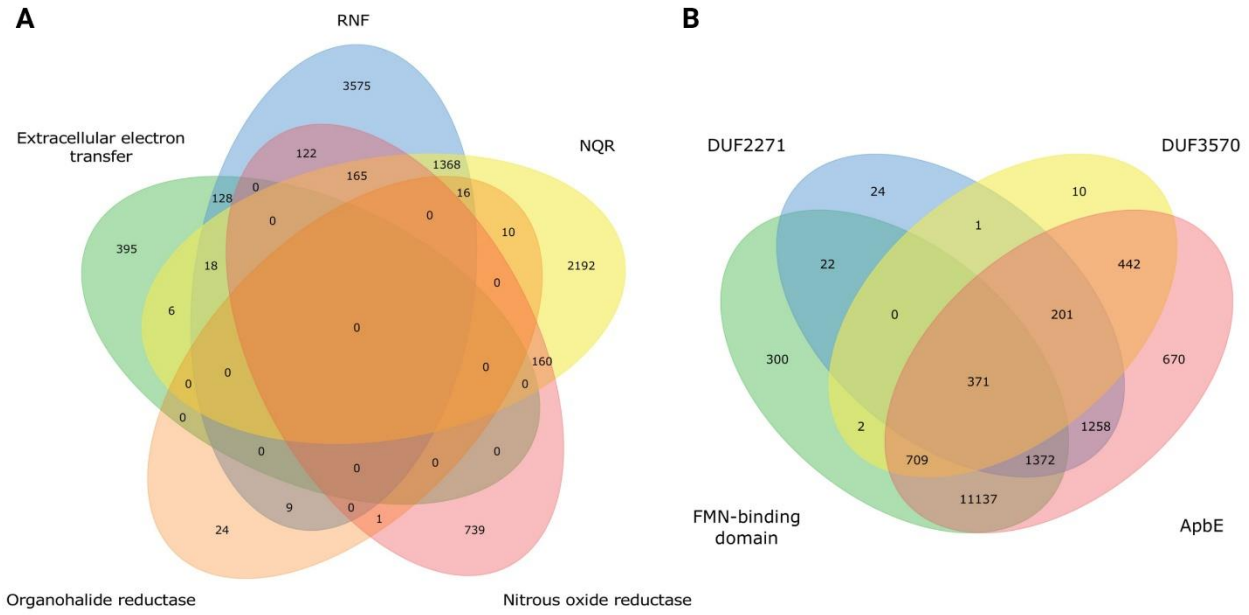


Figure 2.3 Distribution of extracytosolic electron transfer systems in prokaryotic genomes. (A) Venn diagram showing the number of flavinylated gene clusters in 8,928 prokaryotic genomes that encode previously characterize extracytosolic electron transport systems. (B) Venn diagram showing numbers of gene clusters encoding known flavinylation makers and putative ApbE-substrates, DUF2271 and DUF3570.

We next aimed to determine whether DUF2271 serves as a substrate for flavinylation by ApbE. To this end, we heterologously expressed a DUF2271 protein from *Amantichitinum ursilacus* in *E. coli* and found that it is flavinylated when the cognate ApbE is co-expressed. We additionally introduced a point mutation to convert the presumed FMN-binding Thr residue to Ala and found that this mutation depleted flavinylation signal (**Figure 2.4C**). These findings suggest that, consistent with our hypothesis, DUF2271 is indeed a novel substrate for ApbE.

We similarly observed another subset of genomes encoding orphan *apbE* gene clusters that associate with another unknown domain, namely DUF3570. DUF3570 is a ~420 amino acid domain containing two conserved [S/T]GA[S/T] FMN-binding motif, while similarly lacking identifiable FMN-binding domains (**Figure 2.4B**). We conducted similar co-expression experiment on a DUF3570 protein encoded by *Chlorobium luteolum* and confirmed flavinylation in the presence of a cognate ApbE. Individual point mutations at either presumed FMN-binding Ser residues did not alter flavinylation signal, whereas double mutations at both sites concurrently depleted flavinylation (**Figure 2.4D**). These results present DUF3570 as a novel, double-flavinylated ApbE substrate.

In light of these findings, we additionally included DUF2271 and DUF3570 as screening criteria for flavinylated proteins. This decreased the number of remaining genomes encoding orphan *apbE* clusters to 670, which is ~4% of all genomes included in analysis. This ascertained a more comprehensive examination of flavinylated systems in our downstream analysis.

2.3.3 Electron-transferring systems associated with flavinylation markers

We next evaluated the roles of flavinylation in previously unknown extracytosolic contexts. We examined genomic elements of flavinylated gene clusters that encode

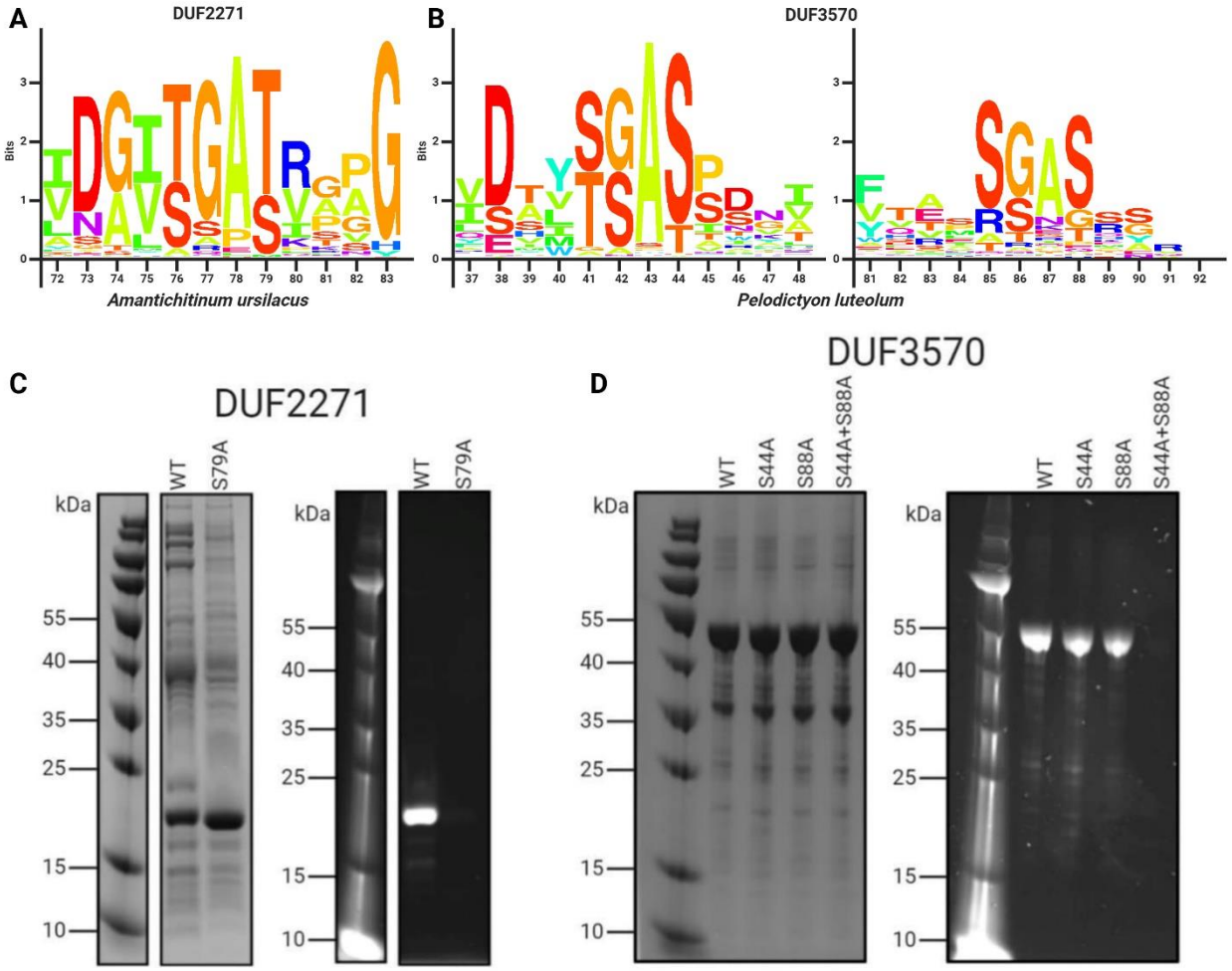


Figure 2.4 DUF2271 and DUF3570 are novel substrates for ApbE-mediated flavinylation. (A&B) Conserved sequence motif within 282 flavinylation-associated DUF2271 proteins (A) and 228 flavinylation-associated DUF3570 proteins (B). Letter size (bits) is proportional to amino acid frequency at each position in the sequence alignment. Amino acid numbering corresponds to the *Amantichitinum ursilacus* DUF2271 sequence or the *Chlorobium luteolum* DUF3570 sequence. (C&D) SDS-PAGE of purified *A. ursilacus* DUF2271 variants and *C. luteolum* DUF3570 variants co-expressed in *Escherichia coli* with their cognate ApbE. The gel is shown with coomassie stain (C or D, left) and under ultraviolet illumination (C or D, right). Bright band under UV indicates covalent attachment of FMN resulted from UV resonance property of the isoalloxazine ring on FMN.

uncharacterized electron-transferring systems. Using ApbE, FMN-binding, DUF2271, and DUF3570 domains as targets of our query, we reviewed genes encoded in genomic segments both upstream and downstream of these flavinylation markers. This led to the identification of five putative electron-transferring membrane components involved in previously uncharacterized flavinylated systems. Unsing functional annotations, genomic contexts, and predictions of protein topology, we described possible interactions between flavinylated proteins and other components in these systems.

2.3.3.1 NapH-like systems

We found that 2,465 flavinylated gene clusters from 2,153 genomes encode an iron-sulfur cluster-binding protein (Pfam accession: PF12801). This protein is homologous to NapH, which is the quinone-binding subunit of periplasmic nitrate reductase widely distributed in major bacterial phyla but primarily found in Proteobacteria species (**Figure 2.5A**) (Brondijk et al., 2002). For 1,926 of identified gene clusters, the NapH-like protein encodes and extracytosolic FMN-binding domain at its N-terminus, while in the remaining 539 gene clusters the FMN-binding domain is encoded by a separate gene.

Insights from pioneering studies suggest that NapH-like proteins function with respiratory oxidoreductases. Previously characterized NapH-like proteins such as PceC and NosZ are flavinylated and are encoded in gene clusters that are responsible for organohalide and nitrous oxide reduction, respectively (Buttet et al., 2018; Zhang et al., 2017). These characterized NapH-like systems are encoded by 1,197 genomes in our dataset. Additionally, we discovered NapH-like gene clusters that co-encode in the same gene clusters with an extracytosolic nitrite reductase in 133 genomes and an ethanol oxidase in 172 genomes. Although the functions of most NapH-like

systems remain undetermined, this capability of interaction with either reductases or oxidases manifest modularity of NapH-oxidoreductase during electron transfer. This modularity is further supported by phylogenetic analysis that reveals convergent evolution of NapH-like associations with the nitrite reductase *nirS* and ethanol dehydrogenase *exaA*.

2.3.3.2 MsrQ-like systems

In 1,797 flavinylated gene clusters encoded by 1,468 genomes, we observed proteins that share homology to the MsrQ (Pfam accession: PF01794), the quinone-binding subunit of periplasmic methionine sulfoxide reductase. Genomes encoding these MsrQ-like loci primarily belong to the Actinobacteria phylum (**Figure 2.5B**). MsrQ-like proteins are predicted to possess two pairs of conserved histidine residues that serve as binding sites for heme groups (Gennaris et al., 2015). These proteins are also distantly related to eukaryotic proteins that function in transmembrane electron transfer, including NADPH oxidase and STEAP iron reductases (Zhang et al., 2013).

We found that in 1,437 of these gene clusters, MsrQ-proteins commonly co-encode with the FMN transferase ApbE and a protein possessing the FMN-binding domain. MsrQ-like proteins in these clusters also contain a NAD-binding domain at their C-termini (Pfam accession: PF00175). In another 31 clusters, we found that MsrQ-like proteins co-encode with a protein that is homologous to the NAD-binding subunit of the respiratory complex I, NuoF (Pfam accession: PF10589). These observations suggest that many flavinylated MsrQ-like gene clusters interact with cytosolic NAD(P)H, which acts as an electron donor. In addition, we identified 153 MsrQ-like gene clusters that encode a protein that shares homology with the eukaryotic ferrous iron

transporter VIT1 (Pfam accession: PF01988), implying that some MsrQ-like systems could couple with assimilatory iron reductases and facilitate iron update through VIT1.

2.3.3.3 PepSY-like systems

In 3,220 flavinylated gene clusters found in 3,040 genomes, we identified proteins that possess a PepSY domain (Pfam accession: PF03929 and PF16357). These PepSY-like proteins are primarily encoded by Gram-negative bacteria, with most analyzed genomes from the Proteobacteria phylum (**Figure 2.5C**). Similarly, these flavinylated PepSY-like gene clusters commonly contain the FMN transferase ApbE. Of these clusters, 2,833 of them also encode the novel ApbE substrate DUF2271 (Pfam accession:PF10029). In 1,500 clusters, we also observed a secreted protein possessing the domain with unknown function DUF4198 (Pfam accession: PF10670). Consistent with this association between PepSY-like proteins and proteins with previously unexamined domains, little is known on the activity of PepSY-like proteins. The only previously characterized PepSY-like homolog, VciB, is encoded by the pathogen *Vibrio cholerae* (Peng and Payne, 2017). Although the *vciB* locus lacks *abpE* or genes encoding the FMN-binding domain, VciB is reported to have extracytosolic iron reductase activity. In addition, PepSY-like proteins in 1,077 analyzed genomes possess a NAD-binding domain (pFam accession: PF00175). These observations suggest that these flavinylated PepSY-like systems may similarly utilize cytosolic NADH as an electron donor and partake in transmembrane electron transfer.

Several lines of evidence also led us to consider that PepSY-like proteins may also facilitate iron reduction and assimilation, similar to MsrQ-like proteins. First, we found that PepSY-like proteins and MsrQ-like proteins co-encoded in 108 gene clusters. Second, we observed that PepSY-like proteins in 167 gene clusters also possess a homologous domain of the eukaryotic

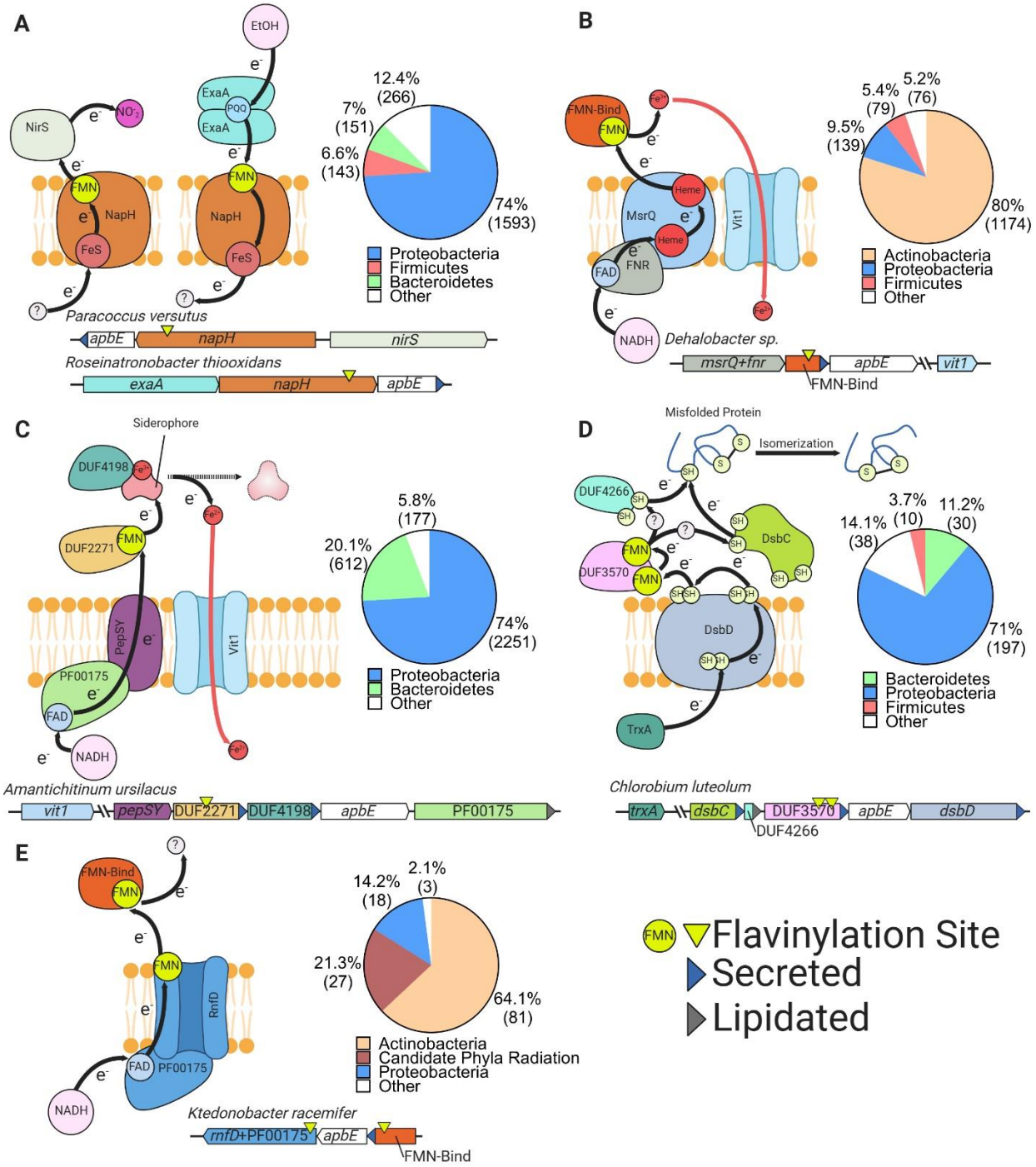


Figure 2.5 Uncharacterized flavinylation-associated extracytosolic electron transfer systems. (A-E) Taxonomic distribution, genomic context, and hypothesized mode of electron transfer for previously uncharacterized flavinylation-associated gene clusters co-encoding flavinylation markers and extracytosolic electron transfer systems, including NapH-like systems (A) (Moreno-Vivián et al., 1999), MsrQ-like systems (B) (Juillan-Binard et al., 2017), PepSY (C) (Yeats et al., 2004), DsbD (D) (Bushweller, 2020), and RNF/NQR-like (E) (Steuber et al., 2014). 'Secreted' and 'lipidated' refer to the presence of computationally predicted signal peptides and lipidation sites, respectively.

ferrous iron transporter VIT1 (Pfam accession: PF01988). Third, previous studies demonstrated that PepSY-like gene clusters are repressed by Fur, a transcription regulator that responds to iron limitation in the Gram-negative bacteria *Shewanella oneidensis* and *Caulobacter crescentus* (Da Silva Neto et al., 2013; Wan et al., 2004). And finally, 231 PepSY-like clusters also encode a TonB receptor (Pfam accession PF03544), which is associated with outer membrane siderophore transporters. Taken together, these insights imply that PepSY-like clusters are also likely involved in iron uptake (Liu et al., 2018; Manck et al., 2020).

2.3.3.4 DsbD systems

From 275 genomes, we observed 285 flaviynlated gene clusters that encode a DsbD protein (Pfam accession: PF2683). These genomes primarily belong to Gram-negative bacteria in the Proteobacteria and Bacteroidetes phyla (**Figure 2.5D**). Several previous studies characterized DsbD family proteins as transmembrane thioredoxin-like proteins (Krupp et al., 2001; Missiakas et al., 1995). Leveraging thiol-disulfide exchange chemistry from a pair of conserved cysteine residues, DsbD family proteins shuttle electrons to the extracytosolic space and promote redox-dependent activities, such as oxidative protein folding (Cho and Collet, 2013).

Among these flavinylated DsbD gene clusters, we observed a common pattern where the same clusters also encode the FMN transferase ApbE, an unknown protein with the novel ApbE substrate DUF3570 domain (227 clusters; Pfam accession: PF12094), a thioredoxin-like protein (183 clusters; Pfam accession: PF13899), and another unknown protein with the DUF4266 domain (221 clusters; Pfam accession: PF14086). Of note, DUF4266 proteins are small transmembrane proteins that possess a conserved pair of cysteine residues in a CXC motif, implying its possible involvement in electron transfer through similar thiol-disulfide exchange chemistry.

The combination of ApbE, DUF3570, and DUF4266 is also found in other 316 flavinylated gene clusters that lack a DsbD protein. This observation suggests that the DUF3570/DUF4266 systems are modular, with a hybrid thioredoxin-like/flavinylation-based system receiving electrons from DsbD in some but not all bacteria. While mechanisms underlying DUF3570/DUF4266-mediated electron transfer remain uncertain, we observed the VIT1 ferrous iron transporter VIT1 in 25 of analyzed gene clusters. This finding implies that DUF3570/DUF4266 systems may similarly partake in iron assimilation.

2.3.3.5 Nqr/Rnf-like systems

From 127 genomes, we identified 127 flavinylated gene clusters encoding components of the Nqr/Rnf systems, which are described in earlier sections. These genomes primarily belong to species under the Actinobacteria phylum (**Figure 2.5E**). Common to complete Nqr and Rnf systems, a core transmembrane apparatus is responsible for bidirectional electron transfer paths across the cell envelope: a first path takes electrons from the cytosol to the extracytosolic FMN-binding domain, while a second path takes electrons from the FMN-binding domain back to a cytosolic substrate. However, partial, Nqr/Rnf-like gene clusters identified in our analysis only encode components that establish one electron transfer pathway for a unidirectional electron flow (Juárez et al., 2010; Steuber et al., 2014).

Identified Nqr/Rnf-like gene clusters encode a protein with an N-terminal membrane domain that is homologous to the flavinylated NqrB/RnfD and a cytosolic C-terminal domain homologous to the NAD-binding domain NqrF (Pfam accession PF00175). Given their homology to components in complete Nqr/Rnf systems, Nqr/Rnf-like systems may also utilize cytosolic NAD(P)H as a cytosolic electron donor and shuttle electrons to extracytosolic space through

NqrB/RnfD domains (Juárez et al., 2010). While it is unclear whether Nqr/Rnf-like systems are also involved in iron reduction and assimilation, we identified 6 Nqr/Rnf-like gene clusters that also encode VIT1, suggesting that these systems may also aid in iron transport despite the absence of other components in the complete Nqr/Rnf complexes.

2.3.4 Flavinylated proteins associated with iron assimilation and respiration

In addition to flavinylated gene clusters that contain a core electron-transferring membrane component, some clusters identified in our analysis do not encode a transmembrane apparatus. This led us to consider that these clusters are likely to interact with other systems located in a distant genomic location to complete the circuit for electron transfer. In this section, we describe two examples of such clusters.

2.3.4.1 P19 clusters

We identified 260 flavinylated gene clusters that encode a ferrous iron-binding protein, P19 (Pfan accession: PF10634), as well as an iron transporter FTR1 (Pfam accession: PF03239). Both P19 and FTR1 family proteins were previously reported to be involved in iron assimilation (Chan et al., 2010). These clusters are primarily found in Gram-positive bacteria from the Actinobacteria and Firmicutes phyla (**Figure 2.6A**). Interestingly, many P19 gene clusters from Gram-negative bacteria do not contain a flavinylated protein. Instead, Gram-negative P19 clusters commonly encode an additional thioredoxin-like protein (**Figure 2.6B**) (Liu et al., 2018). This observation suggests that the role of flavinylated proteins in iron uptake in P19 gene clusters could functionally mimic thioredoxin-like proteins in Gram-negative P19 clusters.

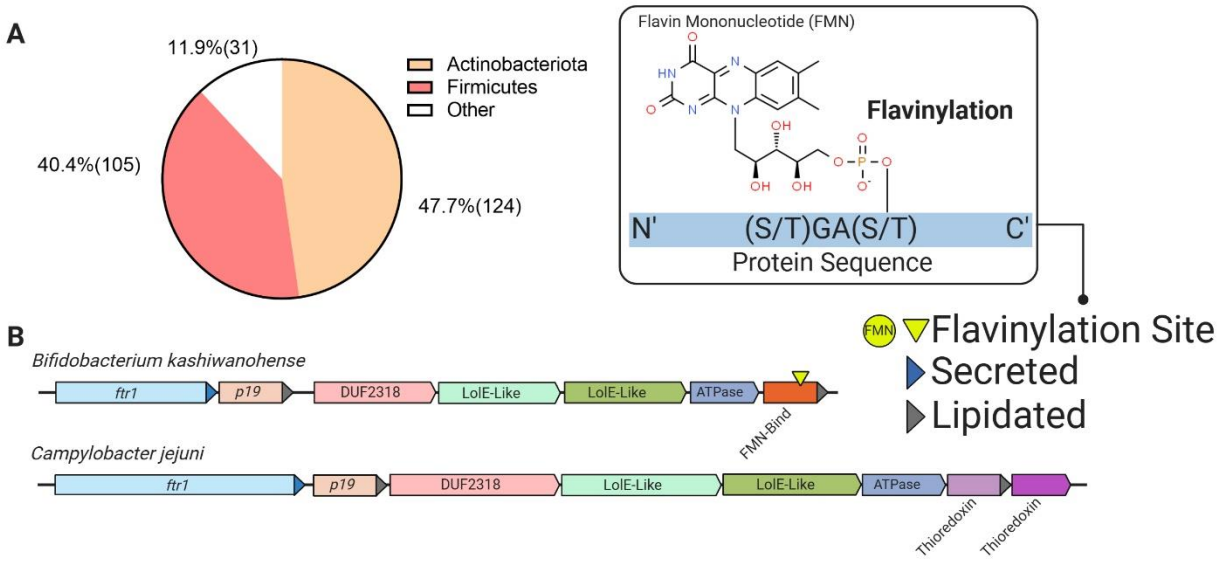


Figure 2.6 Flavinylation and thioredoxin-like proteins encoded in P19 gene clusters.
(A) Phylum-level distribution of genomes encoding flavintlation-associated gene clusters that contain the ferrous iron-binding protein P19. (B) Representative P19 gene clusters with thioredoxin-like or FMN-binding proteins. The *Campylobacter jejuni* cluster includes a mechanistically uncharacterized iron transporter and a thioredoxin-like protein (Liu et al., 2018). P19 clusters in other genomes, such as *Bifidobacterium kashiwanohense*, lack the thioredoxin-like gene, but contain a gene with an FMN-binding domain.

2.3.4.2 Fumarate reductase-like oxidoreductase clusters

We identified 3,070 fumarate reductase-like enzymes (Pfam accession: PF00890) with an FMN-binding domain in 2,979 gene clusters from 1,236 genomes. The fumarate reductase-like enzyme superfamily entails a wide collection of evolutionarily related redox-reactive proteins, but the activity and substrate of many members in this superfamily remains unknown (Jardim-Messeder et al., 2017; Light et al., 2019). Characterized proteins in this family, fumarate, urocanate, and methacrylate reductases could facilitate extracytosolic electron transfer and function in respiration (Bogachev et al., 2018; Light et al., 2019; Mikoulinskaia et al., 1999). Fumarate reductases in *Listeria monocytogenes* and *Shewanella oneidensis* were previously reported to be flavinylated (Bogachev et al., 2018; Light et al., 2019), with presumed interactions with components of the electron transport chain encoded in a different genomic location (Kees et al., 2019; Light et al., 2019). This insight suggests that flavinylated proteins could be associated with a wide range of reductases that are involved in diverse respiratory redox activities.

2.3.5 Functional modularity of respiration systems

In our analysis, we observed a common pattern in some gene clusters that share similar electron transfer components with a subset of flavinylated gene clusters but, in place of a flavinylated protein, encode a cytochrome protein. For example, we identified a flavinylated NapH-like protein that is well situated to mediate electron transfer from the extracytosolic alcohol oxidase to the electron transport chain, where a cytochrome c protein has been shown to play this role in other microbes (**Figure 2.7A**) (Schobert and GBrisch, 1999).

A similar example is found for a fumarate reductase cluster in *S. oneidensis*, which was mentioned in the previous section. In this cluster, the fumarate reductase possesses multiple heme

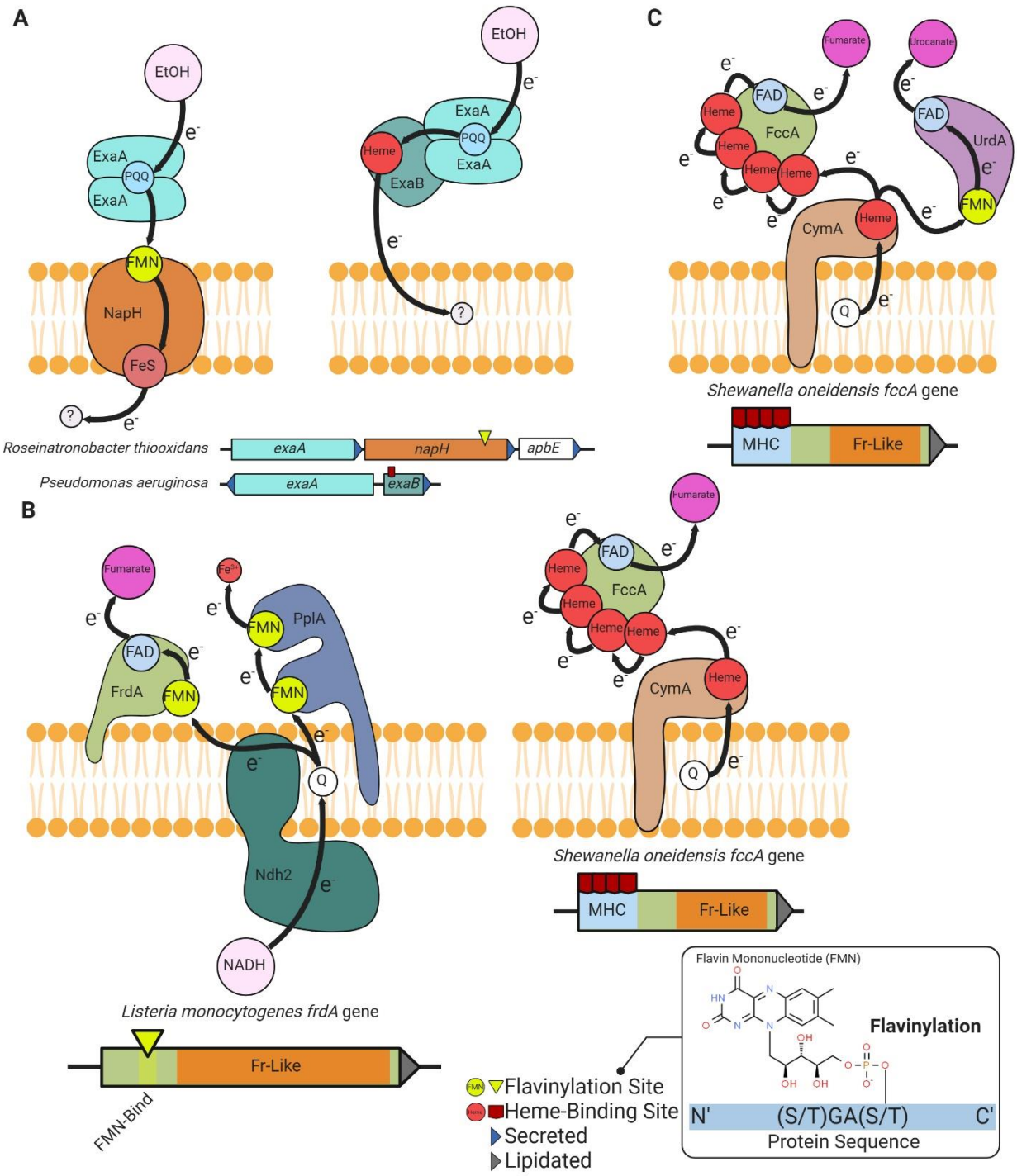


Figure 2.7 Modular flavinylation/cytochrome usage in extracytosolic electron transfer.
(A) Models of the characterized *Pseudomonas aeruginosa* (Görisch, 2003; Matsutani and Yakushi, 2018) and uncharacterized *Rhiobacillus thiooxidans* exaA alcohol oxidase gene clusters. (B) Models of *Listeria monocytogenes* (Light et al., 2019) and *Shewanella oneidensis* fumarate reductases (FrdA and FccA, respectively) (Kees et al., 2019). (C) Characterized electron transfer pathways for the *S. oneidensis* fumarate reductase-like enzymes, fumarate reductase (FccA), and urocanate reductase (UrdA) (Kees et al., 2019).

cytochrome c domains in its N-terminus (Pfam accession: PF14537). However, in a similar fumarate reductase cluster in *L. monocytogenes*, the fumarate reductase gene encodes an FMN-binding domain instead (**Figure 2.7B**) (DiChristina and DeLong, 1994; Light et al., 2019). This substitution suggests great functional modularity of FMN-binding domains under different microbiological contexts.

We also found that this functional modularity is a common scheme for fumarate reductases. We identified 147 gene clusters from 108 genomes that encode separate fumarate reductase-like and cytochrome proteins and 879 genes within 360 genomes that encode a single protein with both fumarate reductase-like and cytochrome domains. Many genomes contain multiple paralogs of fumarate reductase-like genes, with 99 genomes encoding enzymes associated with both cytochrome and flavinylation. This is exemplified by *S. oneidensis*, which encode both a cytochrome-associated fumarate reductase and a flavinylated urocanate reductase with respiratory function (**Figure 2.7C**) (Bogachev et al., 2018). These findings suggest that both flavinylation- and cytochrome-based extracytosolic electron transfer mechanisms could coexist in some microbes.

2.3.6 Flavinylated proteins with greater multiplicity of FMN

In our analysis, we identified 2,081 proteins from 1,530 genomes that possess more than 2 FMN-binding domains, with the highest multiplicity of FMN moiety being 13 in a single protein. Most of these genomes belong to Gram-positive bacteria under Firmicutes and Actinobacteria phyla (**Figure 2.8A&B**). The presence of multiple co-factors is well known in other electron-transferring proteins. For instance, many cytochrome proteins possess multiple heme groups to secure a longer path of electron transfer in the extracytosolic space (Blumberger, 2018). In some

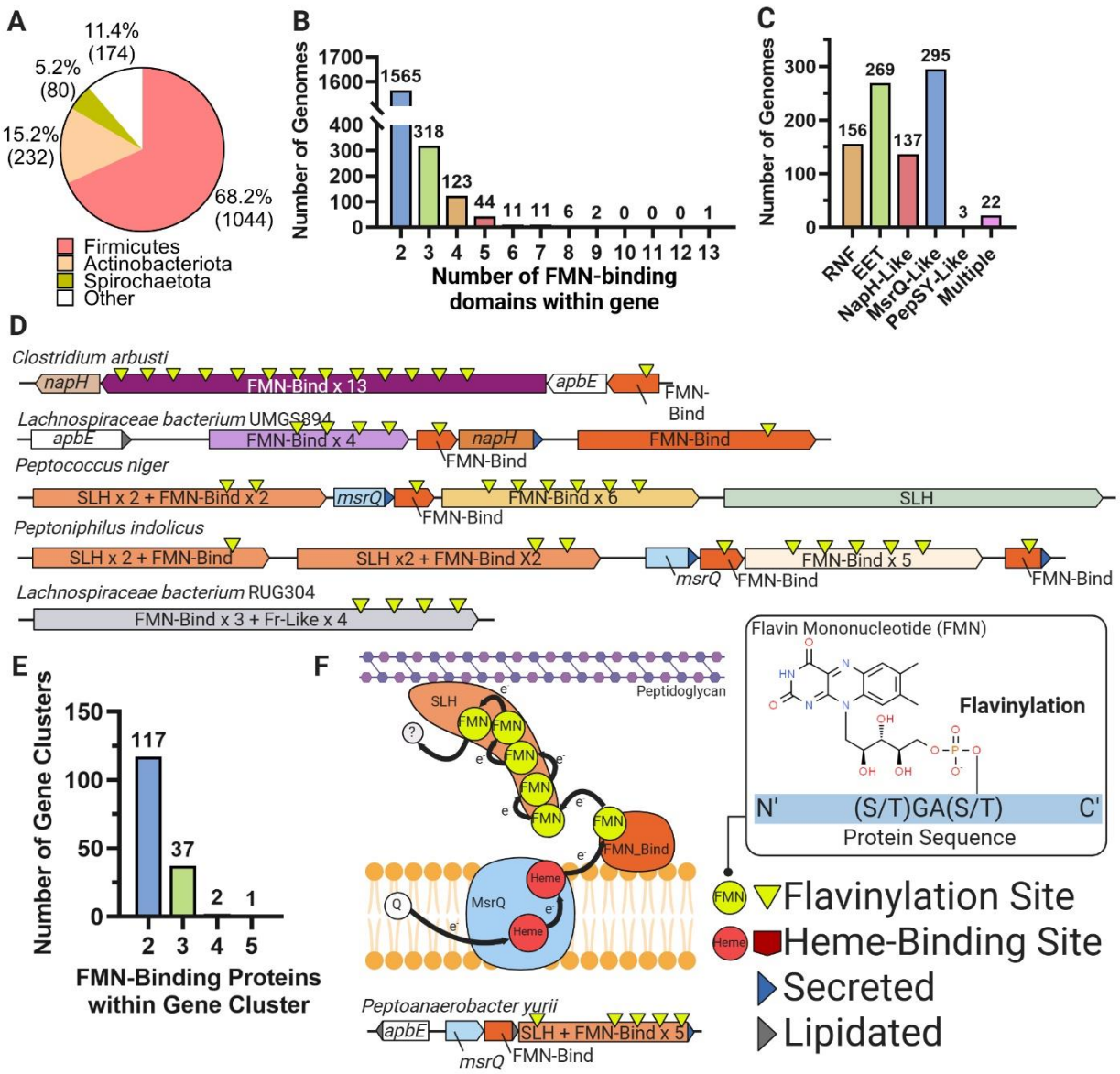


Figure 2.8 Multi-flavinylated proteins may possess novel electron transfer properties.
(A) Phylum-level distribution of multi-flavinylated proteins (i.e., proteins with >1 FMN-binding domain). (B) Histogram showing the number of FMN-binding domains within identified multi-flavinylated proteins. (C) Histogram showing the number of multi-flavinylated protein gene clusters that encode one of the transmembrane electron transfer systems. (D) Examples of multi-flavinylated protein gene clusters. Fumarate reductase-like domains are abbreviated as Fr-like. (E) Histogram showing the number of proteins containing at least one FMN-binding domain within multi-flavinylated protein gene clusters. (F) Model of a possible electron transfer path encoded by a representative multi-flavinylated gene cluster that includes a protein with a cell wall-binding domain.

cases, multi-heme cytochromes are also responsible for shuttling electrons across the cell envelope to reduce insoluble extracytosolic electron acceptors (El-Naggar et al., 2010; Wang et al., 2019). This led us to consider that multi-flavinylated proteins may be functionally similar to multi-heme cytochromes, serving as another parallel between flavinylation- and cytochrome-based electron transfer.

We next examined genomic contexts of these multi-flavinylated gene clusters and found that most multi-flavinylated proteins co-encode with known electron transfer components (**Figure 2.8C**). This suggests that multi-flavinylated proteins may receive electrons through canonical mechanisms. In addition, these clusters also commonly encode unknown proteins that possess cell wall-binding domains, such as SLH (Pfam accession: PF00395), Rib (Pfam accession: PF08428), FIVAR (Pfam accession: PF07554), and LysM (Pfam accession: PF01476) domains. In combination of their prevalence in Gram-positive bacteria, these multi-flavinylated gene clusters thus likely participate in redox reactions in the cell wall or at the cell surface (**Figure 2.8D&F**). In 157 gene clusters, we also observed multi-flavinylated proteins that co-encode with multiple mono-/multi-flavinylated proteins (**Figure 2.8E**). When taken together, these observations suggest that some multi-flavinylated proteins are part of multi-step electron transfer pathways and may form elaborate multi-subunit complexes that span the cell wall.

The function of these multi-flavinylated systems is unknown. In a subset of clusters, we observed proteins with multiple fumarate reductase-like domains that likely establish an unusual multi-functional reductase platform (**Figure 2.9A**). We also identified several clusters that encode multi-flavinylated RnfG, which is the extracytosolic subunit of the Rnf complex, that may facilitate electron transfer to alternative electron acceptors (**Figure 2.9B**). Such Rnf-mediated alternative electron pathway has been previously documented in a methanogen Rnf system

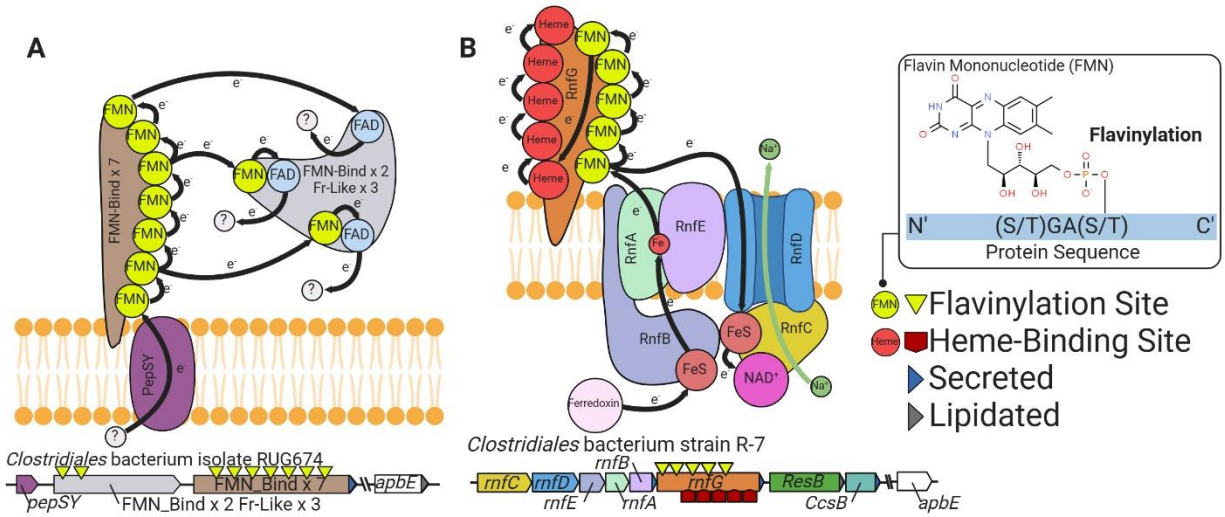


Figure 2.9 Electron transfer in multi-flavinylated PepSY and Rnf systems.
(A) Model of a possible electron transfer path encoded by a representative multi-flavinylated gene cluster that includes multiple Fr-like domains. (B) Model of possible bifurcated electron transfer pathway encoded by an RNF complex with a multi-flavinylated/multi-heme cytochrome RnfG.

(Holmes et al., 2019). Interestingly, we also identified a subset of multi-flavinylated RnfG proteins in the family Christensenellales that also contain a multi-heme cytochrome domain, which implies involvement of muti-flavinylated domains in cytochrome-mediated electron transfer.

2.4 Discussion

Comparative genomic analysis of prokaryotic gene clusters provides high throughput examination of gene functions, facilitating efficient discovery of novel microbial proteins and protein systems. Our study extends this approach to genetic markers associated with ApbE-mediated flavinylation, revealing flavinylation as a prominent feature of bacterial redox physiology. Our scrutinization of flavinylated gene clusters contextualizes flavinylated proteins in a wide range of cellular processes that take place in the extracytosolic space. Modularity of flavinylated proteins in different identified systems further manifests versatile integration of flavinylation in various biochemical processes. These findings suggest that flavinylation plays critical roles in extracytosolic electron transfer and bacterial physiology, establishing a foundation for future research that aims to mechanistically uncover biological significance of flavinylated proteins.

The functional interchangeability of three mechanistically distinctive protein classes gives rise to questions on environmental or physiological determinants that govern differential preferences of each system. However, the breadth of taxonomic distribution of flavinylated proteins makes it difficult to identify unique features specific to flavinylation-encoding microbes. Nevertheless, the functional interchangeability of flavinylation- and cytochrome-based electron transfer is interesting. This phenomenon resembles a similar functional interchangeability that is present between flavodoxins and ferredoxins. Both flavodoxins and ferredoxins are families of

cytosolic redox-active proteins that utilize flavin and iron as their electron-transferring cofactors, respectively (Yoch and Valentine, 1972). Previous studies reported that some microbes are capable of switching from ferredoxin to flavodoxin usage under iron-poor environments, which mitigates demands for iron and enables better conservation of iron for other cellular processes (Knight et al., 1966; Smillie, 1965). This precedence provides a plausible explanation for the biological significance of functional interchangeability of flavinylation- and cytochrome-based electron transfer. When environmental iron is low, flavinylation-based electron transfer may similarly provide an alternative electron circuit that alleviates cellular demands for iron. Thus, such functional interchangeability may reflect divergent but efficient resource management strategies in response to environmental fluctuations.

This resemblance of flavinylation- and cytochrome-based electron transfer is also present for proteins that possess greater multiplicity of cofactors. Multi-flavinylated proteins functionally mimic multi-heme proteins, facilitating electron transfer across a longer circuit. In the extraordinary case of nanowires, multi-heme cytochrome proteins form elongated structures that extend far from the cell body and promote electron transfer onto distant terminal electron acceptors. This unique property of multi-heme cytochrome proteins led to emergence of potential biotechnological applications for novel bioremediation approaches and bioelectricity generation (Blumberger, 2018; Liu et al., 2020). Thus, the functional resemblance of multi-flavinylated proteins to multi-heme cytochromes reveals new implications for the development of novel, flavinylation-based biotechnological tools that represent an interesting subject for future studies.

2.5 Methods

2.5.1 Collection of prokaryotic genomes and functional annotations of protein domains

The 30,238 bacterial and 1,672 archaeal genomes from the GTDB (release 05-RS95 of July 17, 2020) were downloaded with the taxonomy and the predicted protein sequences of the genomes (Parks et al., 2020).

Protein sequences were functionally annotated based on the accession of their best Hmmsearch match (version 3.3) (E-value cut-off 0.001) (Eddy, 1998) against the KOfam database (downloaded on February 18, 2020) (Aramaki et al., 2020). Domains were predicted using the same Hmmsearch procedure against the Pfam database (version 33.0) (Mistry et al., 2021). SIGNALP (version 5.0) was run to predict the putative cellular localization of the proteins using the parameters -org arch in archaeal genomes and -org gram +in bacterial genomes (Almagro Armenteros et al., 2019). Prediction of transmembrane helices in proteins was performed using TMHMM (version 2.0) (default parameters) (Krogh et al., 2001).

2.5.2 Identification of flavinylated systems

The five genes downstream and upstream of an ApbE, FMN-binding domain, DUF3570 or DUF2271 encoding genes were first collected. Only gene clusters with at least one signal peptide or lipidation site predicted in one of the four target genes were considered for further analysis and were referred to as "flavinylation-associated gene clusters." The flavinylation-associated gene clusters were then assigned to 1 of the 10 flavinylated systems based on the presence of key genes. The RNF system was considered present if a Na+-translocating ferredoxin:NAD+ oxidoreductase subunit B (RnfB, KEGG accession K03616) was encoded within a flavinylation-associated gene cluster. The nitrous oxide reduction system (Nos) was considered present if a nitrous oxide reductase (NosZ, KEGG accession K00376) was encoded within a flavinylation-associated gene

cluster. The organohalide respiration system was defined by the presence of a PceA enzyme (Pfam accession PF13486) encoded within a flavinylation-associated gene cluster. The extracellular electron transfer system was considered present if a NADH dehydrogenase (KEGG accession K03885) with a transmembrane helix was encoded within a flavinylation-associated gene cluster. The NQR was defined by the Na+-transporting NADH:ubiquinone oxidoreductase subunit F (NqrF, KEGG accession K00351) encoded within a flavinylation-associated gene cluster. The NapH-like was considered present if an 'NapH-like' iron-sulfur cluster-binding protein (Pfam accession PF12801 or KEGG accession K19339) was encoded within a flavinylation-associated gene cluster. NapH-like gene clusters that encoded a NirS (KEGG accession K15864) were identified as containing a nitrite reductase. NapH-like gene clusters that encoded an ExaA enzyme (KEGG accession K00114) were identified as containing an alcohol oxidase. The MsrQ-like system was defined by the presence of 'MsrQ-like' (Pfam accession PF01794) gene within a flavinylation-associated gene cluster. The PepSY-like system was defined by a PepSY-like (Pfam accessions PF03929 and PF16357) protein encoded within a flavinylation-associated gene cluster. The DsbD system was considered present if a DsbD protein (Pfam accession PF02683) was encoded within a flavinylation-associated gene cluster. Finally, the NQR/RNF-like system was considered present if a flavinylation-associated gene cluster encoded a protein with an N-terminal membrane domain that is homologous to NqrB/RnfD (Pfam accession PF03116) and a cytosolic C-terminal domain homologous to the NAD-binding domain NqrF (Pfam accession PF00175).

2.5.3 Phylogenetic analyses of the 'NapH-like' iron-sulfur cluster-binding protein sequences

The 'NapH-like' iron-sulfur cluster-binding protein tree was built as follows. Sequences were aligned using MAFFT (version 7.390) (–auto option) (Katoh and Standley, 2016). The alignment was further trimmed using Trimal (version 1.4.22; –gappyout option) (Capella-

Gutiérrez et al., 2009). Tree reconstruction was performed using IQ-TREE (version 1.6.12) (Nguyen et al., 2015), using ModelFinder to select the best model of evolution (Kalyaanamoorthy et al., 2017) and with 1000 ultrafast bootstrap (Hoang et al., 2018).

2.5.4 Concatenated 16 ribosomal proteins phylogeny

A maximum-likelihood tree was calculated based on the concatenation of 14 ribosomal proteins (L2, L3, L4, L5, L6, L14, L15, L18, L22, L24, S3, S8, S17, and S19). Homologous protein sequences were aligned using MAFFT (version 7.390; --auto option) (Katoh and Standley, 2016) and alignments refined to remove gapped regions using Trimal (version 1.4.22; --gappyout 570 option) (Capella-Gutiérrez et al., 2009). The protein alignments were concatenated with a final alignment of 9,152 genomes and 2,850 positions. Tree reconstruction was performed using IQ-TREE (version 1.6.12) (Nguyen et al., 2015). A LG + I + G4 model of evolution was selected using ModelFinder (Kalyaanamoorthy et al., 2017) and 1000 ultrafast bootstraps were performed (Hoang et al., 2018)

2.5.5 DUF2271 and DUF3570 sequence analyses

Sequences of flavinylation-associated DUF3570 and DUF2271 proteins were aligned using EMBL-EBI Clustal Omega Multiple Sequence Alignment (Madeira et al., 2019). Sequence logos of the flavinylation sites were generated in R using the 'ggseqlogo' package (Wagih, 2017).

2.5.6 DUF2271 and DUF3570 overexpression and purification

A synthetic construct of the signal peptide-truncated *A. ursilacus* IGB-41 DUF2271 gene (NCBI accession WP_053936890.1) was subcloned into the pMCSG53 vector. A second construct contained a ribosome-binding site and the signal peptide-truncated cognate *apbE* (NCBI accession WP_053936888.1) just downstream of the DUF2271 gene. A similar cloning strategy was used for the *C. luteolum* DSM 273 DUF3570 (NCBI accession ABB24424.1) and its cognate *apbE*

(NCBI accession ABB24423.1). Point mutations of the DUF3570- and DUF2271-encoding genes were generated using the NEB Q5 Site-Directed Mutagenesis Kit. Briefly, overlapping primers containing mutated sequences were used in PCR using pMCSG53::DUF3570 or pMCSG53::DUF2771 as DNA template to generate expression vectors containing respective mutant sequences. Plasmids containing wild-type protein sequences were removed using digestion with the DpnI enzyme, which only acts on methylated DNA sequences.

Sequence-verified plasmids were transformed into *E. coli* BL21 cells (Rosetta DE3, Novagen). A single colony of each expression strain was isolated on Luria-Bertani (LB) agar supplemented with carbenicillin (100 μg/mL) and inoculated into 15 mL of LB. Following overnight growth, cultures were diluted in 500 mL of brain heart infusion broth to a final OD600 of 0.1. When the OD600 reached 0.7–1 protein overexpression was induced by adding isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 mM. The culture was then incubated overnight at 25°C with aeration and collected by centrifugation (7000 × g for 15 min). After removing the supernatant, cells were washed in 30 mL of lysis buffer (5:1 v/weight of pellet; 300 mM NaCl, 1 mM dithiothreitol, 10 mM imidazole, 1 mM ethylenediaminetetraacetic acid (EDTA), and 50 mM Tris-HCl pH = 7.5). Pelleted cells were stored at −80°C overnight, resuspended in lysis buffer, lysed by sonication (8 × 30 s pulses), and cleared by centrifugation (40,000 × g for 30 min).

For the purification of *A. ursilacus* DUF2271, cell lysate was collected and loaded onto a 5 mL HisTrapTM column (GE Healthcare) using the ÄKTA Pure FPLC. Protein was eluted using an imidazole concentration gradient with a maximal concentration of 500 mM. Protein concentrations of eluted fractions containing His6-tagged DUF2271 were measured on a DeNovix

DS-11 FX+ Spectrophotometer based on protein molar mass and extinction coefficient and standardized to 0.4 mg/mL.

Initial observations revealed that the majority of expressed *C. luteolum* DUF3570 was present in the lysed cell pellet. To purify *C. luteolum* DUF3570, the cell pellet was washed with wash buffer (10:1 v/weight of pellet; 100 mM Tris-HCl pH = 7.5, 300 mM NaCl, 2 mM 2-mercaptoethanol, 1 M guanidine-HCl, 1 mM EDTA, and 2% w/v Triton X-100) and centrifuged at 40,000 × g for 30 min. The supernatant was then discarded, and this washing step was repeated until the supernatant was clear. The cell pellet was then resuspended in wash buffer (10:1 v/weight of pellet) without guanidine-HCl and Triton X-100 and centrifuged (40,000 × g for 15 min) to remove guanidine-HCl and Triton X-100. After supernatant was discarded, the pellet was resuspended in extraction buffer (4:1 v/weight of pellet; 100 mM Tris-HCl pH = 7.5, 300 mM NaCl, 10 mM imidazole, 6 M guanidine-HCl, 2 mM 2-mercaptoethanol, and 1 mM EDTA) and protein was denatured by overnight rotator mixation. Supernatant containing denatured DUF3570 was collected by centrifugation (20,000 × g for 30 min).

Subsequent *C. luteolum* DUF3570 purification steps were conducted on a Ni-NTA column. Specifically, 4 mL of Ni-NTA slurry (Nuvia IMAC Resin, 25 mL) were added into a glass chromatography column (Econo-Column, 1.0 × 10 cm2). The column was prepared with 10 mL of column wash buffer and the samples containing denatured DUF3570 were then loaded. Bound protein was eluted using 5 mL of modified column wash buffer containing 50, 100, 200, or 500 mM of imidazole. To prevent guanidine-HCl from forming precipitates with SDS in following steps, eluted samples were mixed with 100% ethanol (9:1 v/v) and incubated at −20°C for 10 min. After centrifugation at 21,100 × g for 5 min and removal of the supernatant, the pelleted protein was washed once with 90% ethanol. The sample was centrifuged again, and the pellet was

resuspended in diH2O. Protein concentrations of eluted fractions containing His6-tagged DUF3570 were measured as described above and standardized to 1.2 mg/mL.

2.5.7 DUF2271 and DUF3570 flavinylation analyses

Purified and normalized DUF2271 and DUF3570 were loaded and separated on a 12% Bis-Tris gel. Prior to gel staining, flavinylated bands were visualized under UV due to the UV resonance of the flavin molecule. Visualizations of Coomassie blue-stained protein were captured with an iBright 1500 gel imager.

2.6 References

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. doi:10.1038/s41587-019-0036-z

Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36:2251–2252. doi:10.1093/bioinformatics/btz859

Backiel J, Zagorevski DV, Wang Z, Nilges MJ, Barquera B. 2008. Covalent Binding of Flavins to RnfG and RnfD in the Rnf Complex from *Vibrio cholerae*. *Biochemistry* 47:11273–11284. doi:10.1021/bi800920j

Blumberger J. 2018. Electron transfer and transport through multi-heme proteins: recent progress and future directions. *Current Opinion in Chemical Biology* 47:24–31. doi:10.1016/j.cbpa.2018.06.021

Bogachev AV, Baykov AA, Bertsova YV. 2018. Flavin transferase: the maturation factor of flavin-containing oxidoreductases. *Biochemical Society Transactions* 46:1161–1169. doi:10.1042/BST20180524

Borshchevskiy V, Round E, Bertsova Y, Polovinkin V, Gushchin I, Ishchenko A, Kovalev K, Mishin A, Kachalova G, Popov A, Bogachev A, Gordeliy V. 2015. Structural and Functional Investigation of Flavin Binding Center of the NqrC Subunit of Sodium-Translocating NADH:Quinone Oxidoreductase from Vibrio harveyi. *PLoS ONE* 10:e0118548. doi:10.1371/journal.pone.0118548

Brondijk THC, Fiegen D, Richardson DJ, Cole JA. 2002. Roles of NapF, NapG and NapH, subunits of the Escherichia coli periplasmic nitrate reductase, in ubiquinol oxidation: Roles of

NapF, G and H in ubiquinol oxidation in E. coli. *Molecular Microbiology* 44:245–255. doi:10.1046/j.1365-2958.2002.02875.x

Burstein D, Harrington LB, Strutt SC, Probst AJ, Anantharaman K, Thomas BC, Doudna JA, Banfield JF. 2017. New CRISPR–Cas systems from uncultivated microbes. *Nature* 542:237–241. doi:10.1038/nature21059

Bushweller JH. 2020. Protein Disulfide Exchange by the Intramembrane Enzymes DsbB, DsbD, and CcdA. *Journal of Molecular Biology* 432:5091–5103. doi:10.1016/j.jmb.2020.04.008

Buttet GF, Willemin MS, Hamelin R, Rupakula A, Maillard J. 2018. The Membrane-Bound C Subunit of Reductive Dehalogenases: Topology Analysis and Reconstitution of the FMN-Binding Domain of PceC. *Front Microbiol* 9:755. doi:10.3389/fmicb.2018.00755

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. doi:10.1093/bioinformatics/btp348

Chan ACK, Doukov TI, Scofield M, Tom-Yew SAL, Ramin AB, MacKichan JK, Gaynor EC, Murphy MEP. 2010. Structure and Function of P19, a High-Affinity Iron Transporter of the Human Pathogen Campylobacter jejuni. *Journal of Molecular Biology* 401:590–604. doi:10.1016/j.jmb.2010.06.038

Cho S-H, Collet J-F. 2013. Many Roles of the Bacterial Envelope Reducing Pathways. *Antioxidants & Redox Signaling* 18:1690–1698. doi:10.1089/ars.2012.4962

Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. 2018. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* 558:440–444. doi:10.1038/s41586-018-0207-y

Da Silva Neto JF, Lourenço RF, Marques MV. 2013. Global transcriptional response of Caulobacter crescentus to iron availability. *BMC Genomics* 14:549. doi:10.1186/1471-2164-14-549

DiChristina TJ, DeLong EF. 1994. Isolation of anaerobic respiratory mutants of Shewannella putrefaciens and genetic analysis of mutants deficient in anaerobic growth on Fe3+. *J Bacteriol* 176:1468–1474. doi:10.1128/jb.176.5.1468-1474.1994

Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359:eaar4120. doi:10.1126/science.aar4120

Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics* 14:755–763. doi:10.1093/bioinformatics/14.9.755

El-Naggar MY, Wanger G, Leung KM, Yuzvinsky TD, Southam G, Yang J, Lau WM, Nealson KH, Gorby YA. 2010. Electrical transport along bacterial nanowires from *Shewanella oneidensis* MR-1. *Proc Natl Acad Sci USA* 107:18127–18131. doi:10.1073/pnas.1004880107

Gennaris A, Ezraty B, Henry C, Agrebi R, Vergnes A, Oheix E, Bos J, Leverrier P, Espinosa L, Szewczyk J, Vertommen D, Iranzo O, Collet J-F, Barras F. 2015. Repairing oxidized proteins in the bacterial envelope using respiratory chain electrons. *Nature* 528:409–412. doi:10.1038/nature15764

Görisch H. 2003. The ethanol oxidation system and its regulation in Pseudomonas aeruginosa. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1647:98–102. doi:10.1016/S1570-9639(03)00066-9

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35:518–522. doi:10.1093/molbev/msx281

Holmes DE, Ueki T, Tang H-Y, Zhou J, Smith JA, Chaput G, Lovley DR. 2019. A Membrane-Bound Cytochrome Enables *Methanosarcina acetivorans* To Conserve Energy from Extracellular Electron Transfer. *mBio* 10:e00789-19. doi:10.1128/mBio.00789-19

Jardim-Messeder D, Cabreira-Cagliari C, Rauber R, Turchetto-Zolet AC, Margis R, Margis-Pinheiro M. 2017. Fumarate reductase superfamily: A diverse group of enzymes whose evolution is correlated to the establishment of different metabolic pathways. *Mitochondrion* 34:56–66. doi:10.1016/j.mito.2017.01.002

Juárez O, Morgan JE, Nilges MJ, Barquera B. 2010. Energy transducing redox steps of the Na^+-pumping NADH:quinone oxidoreductase from *Vibrio cholerae*. *Proc Natl Acad Sci USA* 107:12505–12510. doi:10.1073/pnas.1002866107

Juillan-Binard C, Picciocchi A, Andrieu J-P, Dupuy J, Petit-Hartlein I, Caux-Thang C, Vivès C, Nivière V, Fieschi F. 2017. A Two-component NADPH Oxidase (NOX)-like System in Bacteria Is Involved in the Electron Transfer Chain to the Methionine Sulfoxide Reductase MsrP. *Journal of Biological Chemistry* 292:2485–2494. doi:10.1074/jbc.M116.752014

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. doi:10.1038/nmeth.4285

Katoh K, Standley DM. 2016. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* 32:1933–1942. doi:10.1093/bioinformatics/btw108

Kees ED, Pendleton AR, Paquete CM, Arriola MB, Kane AL, Kotloski NJ, Intile PJ, Gralnick JA. 2019. Secreted Flavin Cofactors for Anaerobic Respiration of Fumarate and Urocanate by Shewanella oneidensis: Cost and Role. *Appl Environ Microbiol* 85:e00852-19. doi:10.1128/AEM.00852-19

Knight E, D'Eustachio AJ, Hardy RWF. 1966. Flavodoxin: A flavoprotein with ferredoxin activity from Clostridium pasteurianum. *Biochimica et Biophysica Acta (BBA) - Enzymology and Biological Oxidation* 113:626–628. doi:10.1016/S0926-6593(66)80025-5

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *Journal of Molecular Biology* 305:567–580. doi:10.1006/jmbi.2000.4315

Krupp R, Chan C, Missiakas D. 2001. DsbD-catalyzed Transport of Electrons across the Membrane ofEscherichia coli. *Journal of Biological Chemistry* 276:3696–3701. doi:10.1074/jbc.M009500200

Light SH, Méheust R, Ferrell JL, Cho J, Deng D, Agostoni M, Iavarone AT, Banfield JF, D'Orazio SEF, Portnoy DA. 2019. Extracellular electron transfer powers flavinylated extracellular reductases in Gram-positive bacteria. *Proc Natl Acad Sci USA* 116:26892–26899. doi:10.1073/pnas.1915678116

Light SH, Su L, Rivera-Lugo R, Cornejo JA, Louie A, Iavarone AT, Ajo-Franklin CM, Portnoy DA. 2018. A flavin-based extracellular electron transfer mechanism in diverse Gram-positive bacteria. *Nature* 562:140–144. doi:10.1038/s41586-018-0498-z

Liu MM, Boinett CJ, Chan ACK, Parkhill J, Murphy MEP, Gaynor EC. 2018. Investigating the Campylobacter jejuni Transcriptional Response to Host Intestinal Extracts Reveals the Involvement of a Widely Conserved Iron Uptake System. *mBio* 9:e01347-18. doi:10.1128/mBio.01347-18

Liu Xiaomeng, Gao H, Ward JE, Liu Xiaorong, Yin B, Fu T, Chen J, Lovley DR, Yao J. 2020. Power generation from ambient humidity using protein nanowires. *Nature* 578:550–554. doi:10.1038/s41586-020-2010-9

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 47:W636–W641. doi:10.1093/nar/gkz268

Manck LE, Espinoza JL, Dupont CL, Barbeau KA. 2020. Transcriptomic Study of Substrate-Specific Transport Mechanisms for Iron and Carbon in the Marine Copiotroph Alteromonas macleodii. *mSystems* 5:e00070-20. doi:10.1128/mSystems.00070-20

Matsutani M, Yakushi T. 2018. Pyrroloquinoline quinone-dependent dehydrogenases of acetic acid bacteria. *Appl Microbiol Biotechnol* 102:9531–9540. doi:10.1007/s00253-018-9360-3

Mikoulinskaia O, Akimenko V, Galouchko A, Thauer RK, Hedderich R. 1999. Cytochrome *c* -dependent methacrylate reductase from *Geobacter sulfurreducens* AM-1. *European Journal of Biochemistry* 263:346–352. doi:10.1046/j.1432-1327.1999.00489.x

Missiakas D, Schwager F, Raina S. 1995. Identification and characterization of a new disulfide isomerase-like protein (DsbD) in Escherichia coli. *The EMBO Journal* 14:3415–3424. doi:10.1002/j.1460-2075.1995.tb07347.x

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Research* 49:D412–D419. doi:10.1093/nar/gkaa913

Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F. 1999. Prokaryotic Nitrate Reduction: Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases. *J Bacteriol* 181:6573–6584. doi:10.1128/JB.181.21.6573-6584.1999

Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 32:268–274. doi:10.1093/molbev/msu300

Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 38:1079–1086. doi:10.1038/s41587-020-0501-8

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004. doi:10.1038/nbt.4229

Peng ED, Payne SM. 2017. Vibrio cholerae VciB Mediates Iron Reduction. *J Bacteriol* 199. doi:10.1128/JB.00874-16

Schobert M, GBrisch H. 1999. Cytochrome c550 is an essential component of the quinoprotein ethanol oxidation system in Pseudornonas aeruginosa :cloning and sequencing of the genes encoding cytochrome c550 and an adjacent acetaldehyde dehydrogenase. *Microbiology* 145:471–481.

Smillie I. 1965. Vol. 20, No. 5, 1965 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS* 20.

Steuber J, Vohl G, Casutt MS, Vorburger T, Diederichs K, Fritz G. 2014. Structure of the V. cholerae Na+-pumping NADH:quinone oxidoreductase. *Nature* 516:62–67. doi:10.1038/nature14003

Wagih O. 2017. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* 33:3645–3647. doi:10.1093/bioinformatics/btx469

Wan X-F, VerBerkmoes NC, McCue LA, Stanek D, Connelly H, Hauser LJ, Wu L, Liu X, Yan T, Leaphart A, Hettich RL, Zhou J, Thompson DK. 2004. Transcriptomic and Proteomic Characterization of the Fur Modulon in the Metal-Reducing Bacterium *Shewanella oneidensis*. *J Bacteriol* 186:8385–8400. doi:10.1128/JB.186.24.8385-8400.2004

Wang F, Gu Y, O'Brien JP, Yi SM, Yalcin SE, Srikanth V, Shen C, Vu D, Ing NL, Hochbaum AI, Egelman EH, Malvankar NS. 2019. Structure of Microbial Nanowires Reveals Stacked Hemes that Transport Electrons over Micrometers. *Cell* 177:361-369.e10. doi:10.1016/j.cell.2019.03.029

Yeats C, Rawlings ND, Bateman A. 2004. The PepSY domain: a regulator of peptidase activity in the microbial environment? *Trends in Biochemical Sciences* 29:169–172. doi:10.1016/j.tibs.2004.02.004

Yoch DC, Valentine RC. 1972. Ferredoxins and Flavodoxins of Bacteria. *Annu Rev Microbiol* 26:139–162. doi:10.1146/annurev.mi.26.100172.001035

Zhang L, Trncik C, Andrade SLA, Einsle O. 2017. The flavinyl transferase ApbE of Pseudomonas stutzeri matures the NosR protein required for nitrous oxide reduction. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1858:95–102. doi:10.1016/j.bbabio.2016.11.008

Zhang X, Krause K-H, Xenarios I, Soldati T, Boeckmann B. 2013. Evolution of the Ferric Reductase Domain (FRD) Superfamily: Modularity, Functional Diversification, and Signature Motifs. *PLoS ONE* 8:e58126. doi:10.1371/journal.pone.0058126

Zhou W, Bertsova YV, Feng B, Tsatsos P, Verkhovskaya ML, Gennis RB, Bogachev AV, Barquera B. 1999. Sequencing and Preliminary Characterization of the Na^+-Translocating NADH:Ubiquinone Oxidoreductase from *Vibrio harveyi*. *Biochemistry* 38:16246–16252. doi:10.1021/bi991664s

Chapter 3: Structural and functional diversity of flavinylated proteins

Data included in this chapter are featured in a manuscript titled "Versatile roles of protein flavinylation in bacterial extracytosolic electron transfer", which was accepted and published by *mSystems*. Affiliations and contributions of authors are listed below.

Shuo Huang[1,2], Raphaël Méheust[3], Blanca Barquera[4,5,6], Samuel H. Light[1,2]*

[1]Duchossois Family Institute, University of Chicago, Chicago, Illinois, USA; [2]Department of Microbiology, University of Chicago, Chicago, Illinois, USA; [3]Génomique Métabolique, CEA, Genoscope, Institut François Jacob, Université d'Évry, Université Paris-Saclay, CNRS, Evry, France; [4]Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, New York, USA; [5]Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, Troy, New York, USA; [6]Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York, USA. *Corresponding author.

S.H.: Conceptualization, Investigation, Formal analysis, Visualization, Writing - original draft; R.M.: Investigation, Formal analysis, Writing - review and editing; B.B.: Formal analysis, Writing - review and editing; S.H.L.: Conceptualization, Funding acquisition, Writing - original draft.

3.1 Abstract

ApbE-mediated flavinylation is a common redox strategy, facilitating extracytosolic electron transfer in a wide range of bacteria. Despite this phylogenetic versatility of ApbE-flavinylated proteins, the molecular context of flavinylation is largely unexplored. Here, we combined comparative genomics, computational structural biology, and in vitro assays to provide a comprehensive examination of functional and structural diversity of flavinylated proteins. We observe that both mono- and multi-flavinylated proteins adopt diverse tertiary structures, demonstrating involvement of flavinylated domains in diverse molecular contexts. We also identify and confirm flavinylation in two classes of proteins under protein families that were largely known for non-covalent FMN-binding, suggesting possible evolution events through which flavinylation evolved from non-covalent FMN attachment. Furthermore, we report a previously unknown protein, DUF4405, as a novel membrane cytochrome in association with flavinylation. Extending our finding on DUF4405, we further identified FezC, a previously

characterized protein involved ferrosome iron storage, as a DUF4405 homolog. These results demonstrate that biological versatility of flavinylation is reflected by structures, functions, and cellular contexts of flavinylated proteins.

3.2 Introduction – Versatility of flavinylation reflected by protein structure and function

In our previous work, we leveraged ApbE and/or FMN-binding domains as genomic markers of flavinylation and conducted a comprehensive bioinformatic screen for flavinylated extracytosolic electron transfer systems in over 30,000 prokaryotic genomes. This analysis revealed that extracytosolic flavinylation machineries are present in approximately 50% of the analyzed genomes, establishing flavinylation as a common redox strategy utilized by a wide range of prokaryotic organisms. Furthermore, we examined the genomic contexts of the identified flavinylated proteins. Our findings suggest that flavinylated proteins often co-localize with other electron transfer components, both known and previously uncharacterized. These co-localized components together form complete electron circuits, connecting a variety of cytosolic electron donors with extracytosolic electron acceptors. These observations present a great molecular versatility of flavinylated proteins. The involvement of flavinylation in diverse redox-dependent cellular processes of a wide spectrum of phenotypically distinct bacterial species led us to hypothesize that this biological versatility of flavinylation is reflected in the structural and functional characteristics of flavinylated proteins.

Assessing this hypothesis faces a critical limitation in throughput of comparative structural analysis. Large-scale structural analysis of flavinylated proteins is rate limiting. Previous approaches for structural analysis commonly relied on experimentally resolved structures resulting from X-ray crystallography or cryo-electron microscopy. While these approaches have high

resolution and are highly reliable, the procedures required for sample preparation could be laborious, precluding the possibility of structural analysis when the number of proteins of interest is high. As in silico alternatives, structural prediction tools such as PHYRE2 and Rosetta were also commonly used (Kelley et al., 2015; Rohl et al., 2004). However, the accuracy of predicted structures from these tools is dependent on the availability of previously resolved protein structures that share structural homology.

The recent development of AlphaFold, an artificial intelligence-powered protein structure prediction tool, has alleviated this limitation (Evans et al., 2021; Jumper et al., 2021). Unlike pre-existing tools such as PHYRE2 or Rosetta, PHYRE2 or Rosetta, AlphaFold employs sophisticated neural network architectures and training methods that leverage spatial data from protein sequence alignments and pairwise residue interactions. This approach is less dependent on previously resolved structures of protein homologs, allowing AlphaFold to achieve high accuracy in structural predictions even for proteins that lack previously known homologous structures.

Here, we leveraged AlphaFold for large-scale comparative structural analysis of flavinylated proteins. By coupling structural insights with genomic data curated from our previous bioinformatic analysis, we reveal that the biological versatility of flavinylation is indeed captured in diverse structural and functional contexts in which flavitnlated proteins situate. Our data demonstrates that AlphaFold-derived protein models complement gene co-localization inference of protein function and reveal multiple facets of flavinylation-based electron transfer throughout prokaryotic life.

3.3 Results

3.3.1 ApbE-flavinylated proteins are structurally diverse

To understand canonical structural properties of ApbE-mediated flavinylation, we first examined structural models of previously identified ApbE substrates derived from either experimentally resolved or Alphafold predicted structures. This includes a family of small monomeric protein FMN-bind, the family containing B subunit of the Nqr complex NqrB, and the novel ApbE substrate identified in our previous work DUF2271 (**Figure 3.1**). The resulting collection of structures reveals a diverse context of flavinylation that varies across the different classes of flavinylated proteins. Despite the common possession of the conserved FMN-binding site, FMN-bind, NqrB, and DUF2271 all have unique overall protein structures. Both FMN-bind and DUF2271 are small soluble proteins, whereas NqrB is a transmembrane protein that interacts with other electron-transfer components as a subunit in Nqr or Nqr-like systems (Hayashi et al., 2001). These insights provide evidence for variable contexts for ApbE-mediated flavination and reveal involvement of flavinylation proteins in different cellular compartments.

We additionally observed that structural variations also distinguish different proteins categorized under the same class. For example, the most common flavinylated domain FMN-bind entails two structurally distinct subtypes. The more common subtype, referred to as FMN-bind$_1$, comprises a compact ~120-amino acid structural core, whereas the less common subtype FMN-bind$_2$ is ~160 amino acid in size and sometimes found in multiple copies within proteins (**Figure 3.1A&B**). Interestingly, proteins with multiple FMN-bind$_2$ domains often possess an even number of these domains, implying possible structural symmetry of multi-FMN-bind$_2$ proteins. Indeed, AlphaFold prediction of an *Erysipelotrichaceae bacterium* double flavinylated protein possessing 2 FMN-bind$_2$ domains provides evidence for this implication. By structurally aligning FMN-bind$_1$

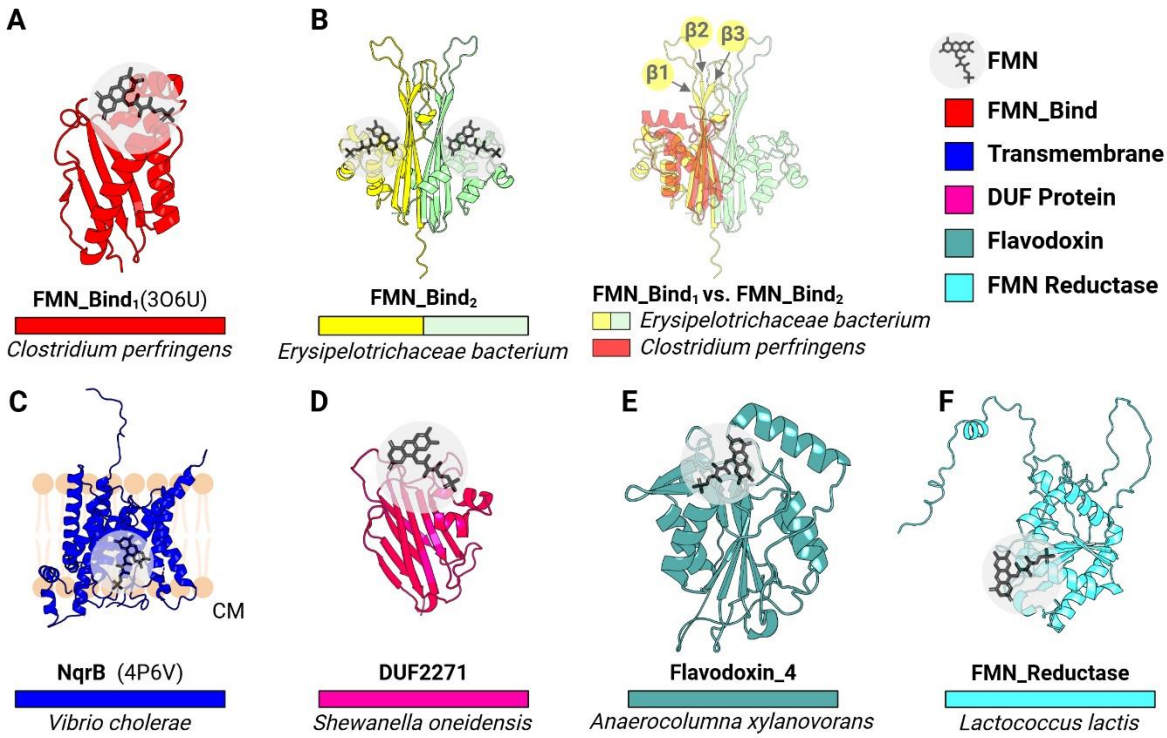


Figure 3.1 Structural contexts of ApbE-flavinylation sites.
(A) Previously resolved crystal structure of a flavinylated monomeric protein from *C. perfringens* (FMN-Bind$_1$, PDB: 3O6U). (B) AlphaFold model of a double-flavinylated protein from *Erysipelotrichaceae bacterium* containing 2 FMN-bind$_2$ domains (left) and its structural alignment with FMN-bind$_1$ (right). Arrows highlight β1, β2, and β3 strands. (C) Previously resolved crystal structure of the B subunit from the Nqr complex (PDB: 4P6V) (Kishikawa et al., 2022). (D) AlphaFold model of a *Shewanella oneidensis* protein with a flavinylated DUF2271 domain. (E&F) AlphaFold models of flavinylated Flavodoxin_4 (E) and flavinylated FMN reductase (F). Gray circles highlight the position of the flavinylated amino acid.

and FMN-bind$_2$ domains, we observed that FMN-bind$_2$ domains have additional insertions between β1-/β2-strands and the β3-strand/α1-helix that are absent in FMN-bind$_1$ domains. These insertions extend the β-sheet face of the domain, serving as an interface with another FMN-bind$_2$ domain that is spatially pseudo-symmetrical (**Figure 3.1B, right**). These distinctions between the FMN-bind$_1$ and FMN-bind$_2$ domains thus establish unique one- and two-flavinylated structural units, respectively.

3.3.2 Evolution connection between non-covalent flavoproteins and ApbE flavinylation

In our genomic analysis, we identified two groups of ApbE-mediated flavinylation substrates that are related to characterized flavoproteins possessing non-covalent flavins as their redox-active cofactors. The first group of these flavoproteins is associated with the Flavodoxin_4 protein family (Pfam accession: PF12682; **Figure 3.2A-C**). The Flavodoxin_4 domains of these proteins encompass an insertion containing a flavinylation motif-like sequence. The second group of flavoproteins associates with the FMN_red protein family (Pfam accession: PF03358; **Figure 3.3A-C**). Proteins in this FMN_red family, in turn, encode a C-terminal region containing one to two flavinylation motif-like sequences. These identified Flavodoxin_4 and FMN_red gene clusters that associate with ApbE-mediated flavinylation are commonly found in members of the Firmicutes phylum (**Figure 3.2E and Figure 3.3E**).

The co-localization of ApbE, which is responsible for covalent attachment of FMN, and members of flavoproteins families that largely bind flavins non-covalently is a counterintuitive observation. However, our earlier insights led us to consider the possibility that structural variations may similarly distinguish novel structural subtypes of flavoproteins in their corresponding families that are associated with ApbE. To test this, we compared AlphaFold-

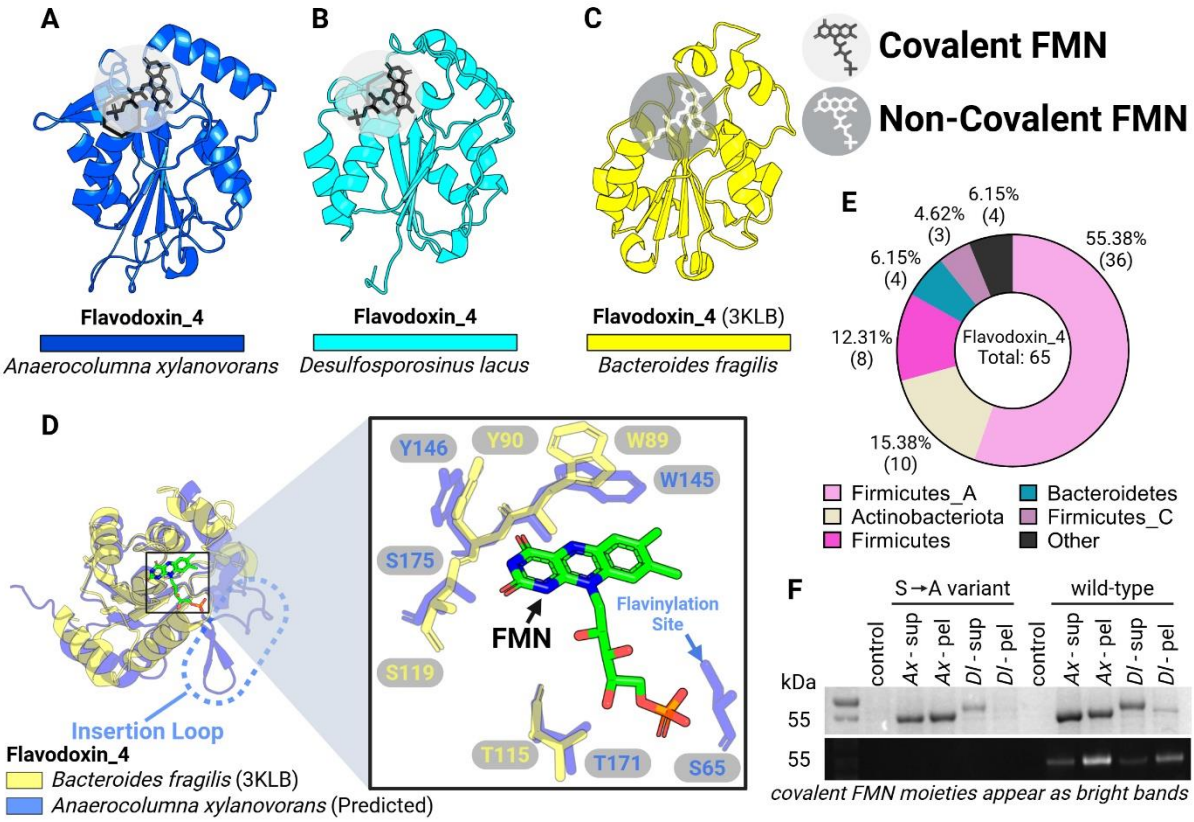


Figure 3.2 ApbE flavinylation evolved from non-covalent Flavodoxin_4.
(A&B) AlphaFold models for flavinylated Flavodoxin_4 proteins from *Anaerocolumna xylanovorans* (A) and *Desulfosporosinus lacus* (B). (C) Previously resolved crystal structure of a Flavodoxin_4 with non-covalently bound FMN (PDB: 3KLB). (D) Structural alignments of Flavodoxin_4 proteins with and without predicted flavinylation site. Dashed oval highlights an inserted loop that is lacking in 3KLB (left). Right panel shows zoom-in view of non-covalent FMN molecule from 3KLB and surrounding residues. Arrow indicates the serine residue in flavinylated Flavodoxin_4 responsible for covalent FMN binding. (E) Taxonomic distribution of Flavodoxin_4 proteins in bacteria. (F) SDS-PAGE gel image of purified flavinylated Flavodoxin_4 from *A. xylanovorans* and *D. lacus*, visualized under UV. Covalent FMN moieties appear as bright bands on SDS-PAGE visualized under UV.

predicted structures of ApbE-associated Flavodoxin_4 and FMN_red candidates to previously experimentally resolved structures of their corresponding homologs. For both Flavodoxin_4 and FMN_red structures, we observed that the core flavin-binding domain is structurally similar irrespective of putative flavinylation status (**Figure 3.2D and Figure 3.3D, left**). AlphaFold-derived models of the ApbE-associated Flavodoxin_4 proteins from *Anaerocolumna xylanovorans* and *Desulfosporosinus lacus* resemble a crystal structure of a homologous non-covalent flavin-binding protein (PDB: 3KLB) from *Bacteroides fragilis*. However, both ApbE-associated Flavodoxin_4 proteins possess an additional insertion between the 1β-strand and the 1α-helix that contains a serine residue serving as the predicted flavinylation site (**Figure 3.2D, right**). Consistently, we observed similar patterns for ApbE-associated FMN_red proteins from *Lactococcus lactis* and *Lactococcus taiwanensis*. We compared AlphaFold models of ApbE-associated FMN_red candidates to a previously resolved structure of FMN_red protein from *E. coli* (PDB: 4PTZ). While all proteins present symmetric binding sites at the dimerization interface, ApbE-associated FMN_red proteins additionally encode an unstructured C-terminal extension that posit the flavinylation sites in spatial proximity to the conserved flavin-binding sites (**Figure 3.3D, right**). When taken together, these observations suggest a structural congruity between ApbE-associated Flavodoxin_4/FMN_red proteins flavinylation motif-like sequences and structurally conserved flavin-binding sites.

To further ascertain this observation, we next aim to determine if ApbE-associated Flavoprotein_4 and FMN_red proteins are truly ApbE substrates that bind to FMN covalently. To this end, we designed an in vitro flavinylation assay to confirm covalent flavinylation in candidate proteins. We heterologously co-expressed each of the ApbE-associated Flavodoxin_4 and FMN_red candidates with their corresponding cognate ApbE in *E. coli*. SDS-PAGE analyses of

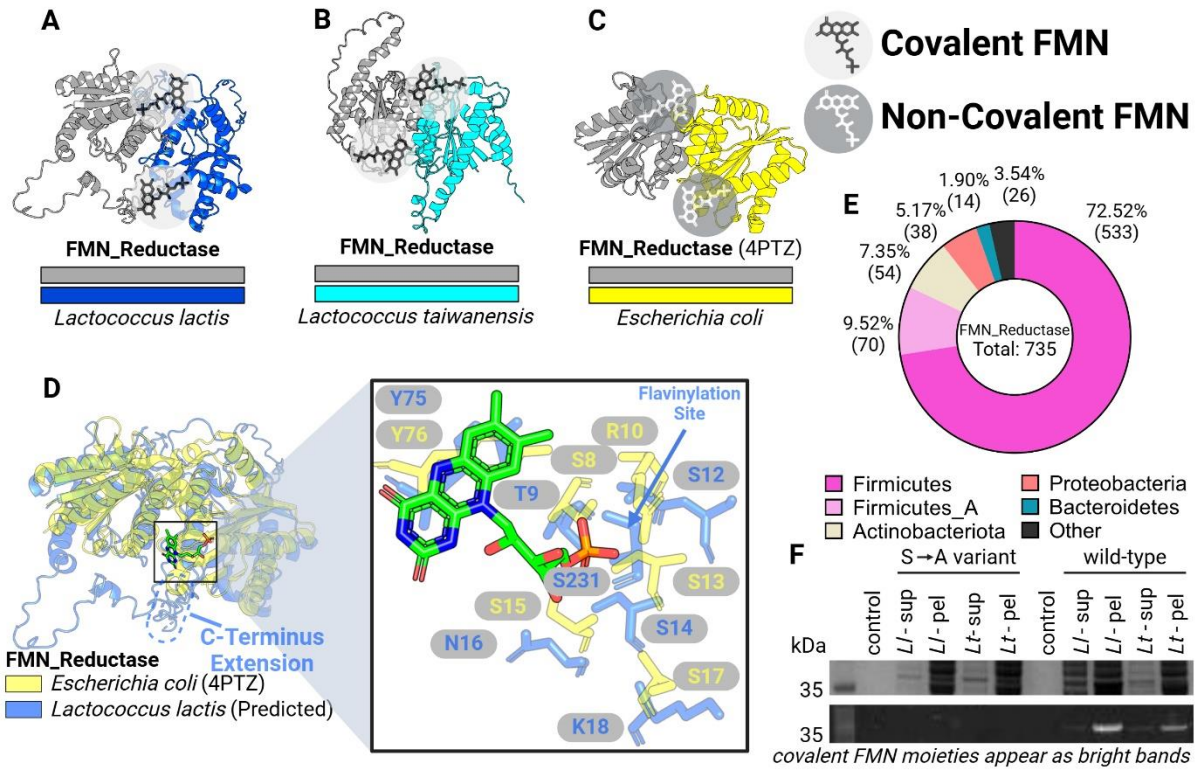


Figure 3.3 ApbE flavinylation evolved from non-covalent FMN_red.
(A&B) AlphaFold models for flavinylated FMN reductases from *Lactococcus lactis* and *Lactococcus taiwanensis*. (C) Previously resolved crystal structure of a FMN reductase with non-covalently bound FMN (PDB: 4PTZ). (D) Structural alignments of FMN reductase proteins with and without covalent FMN binding. Dashed oval highlights a C-terminus extension that is lacking in 4PTZ (left). Right panel shows zoom-in view of non-covalent FMN molecule from 4PTZ and surrounding residues. Arrow indicates the serine residue in flavinylated FMN reductase responsible for covalent FMN binding. (E) Taxonomic distribution of FMN reductases. (F) SDS-PAGE gel image of purified flavinylated Flavodoxin_4 proteins from *L. lactis* and *L. taiwanensis*, visualized under UV. Covalent FMN moieties appear as bright bands on SDS-PAGE visualized under UV.

expressed proteins exhibited strong flavinylation signal, indicated by UV resonance of FMN moieties. We additionally introduced point mutations in each candidate protein, resulting in serine to alanine mutations at their predicted flavinylation sites. This leads to mutant variant of candidate proteins that are depleted of flavinylation signal (**Figure 3.2F and Figure 3.3F**). These data suggest that, despite belonging to protein families that are largely associated with non-covalent flavin binding, ApbE-associated Flavodoxin_4 and FMN_red proteins are flavinylation substrates with covalent FMN-binding activity. Because of their structural homologies to other members of the Flavodoxin_4 and FMN_red families, our findings suggest possible evolution events by which flavinylated proteins evolved through the acquisition of a flavinylation motif at a non-covalent flavin-binding site, securing FMN moieties in extracytosolic space.

3.3.3 Structural diversity in multi-flavinylated proteins

In our previous bioinformatic analysis, we identified over 2,000 proteins that encode 2 or more FMN-binding domains, with putative multiplicity of FMN moieties being as high as 13 in some cases. While the overall protein functions for these multi-flavinylated proteins are largely unknown, this high number of functionally redundant FMN-binding domains and the presence of extracytosolic secretion signal sequences highlight the possible involvement of multi-flavinylated proteins in stepwise electron transfer across a longer electron circuit across the cell envelope (Berthomieu et al., 2022; Bewley et al., 2013). In multi-cytochrome proteins such as the A and C subunits of the MtrABC complex, the heme cofactors are positioned in close contact with each other to facilitate inter-heme electron transfer (Edwards et al., 2020). To determine if the FMN-binding domains within the same protein sequence functionally mimic heme cofactors in multi-

heme cytochromes and similarly spatially interact with each other, we examined AlphaFold-derived structural predictions for selected candidates.

We first noticed that many of these multi-flavinylated proteins possess a "beads-on-a-string" structure, where multiple FMN-bind$_1$ domains are encoded consecutively within the coding sequence but with minimal interactions predicted among these domains (**Figure 3.4A**). In other cases, however, we observed other multi-flavinylated proteins that adopted a different structural arrangement. In a double-flavinylated example, we observed FMN-bind$_1$ domains that are circularly permuted (**Figure 3.4B**). The full sequence encompassing the second FMN-bind$_1$ domain is "inserted" between the N-terminal beta-sheets and the C-terminal helix of the first FMN-bind$_1$ domain, presenting incorporation of FMN-bind$_1$ domains. In another triple-flavinylated example, the three FMN-bind$_1$ domains are predicted to form a three-fold symmetry pattern, resulting in a structure that resembles a three-petaled flower (**Figure 3.4C**). In addition, the triple-flavinylated protein encodes a dimerization domain at its N-terminus, allowing formation of a homodimer. A final example of multi-flavinylated proteins entail a group of proteins that contain a coiled-coil core, serving as the central manifold for up to 8 FMN-bind$_1$ domains (**Figure 3.4D**). These findings illustrate highly intricate structural arrangements in multi-flavinylated proteins.

3.3.4 Flavinylated proteins in diverse transmembrane electron transfer mechanisms

Our previous bioinformatic screen described flavinylated proteins that putatively interact with other known or uncharacterized membrane components. These membrane components adopted various mechanisms for extracytosolic electron transfers, suggesting high modularity and diverse interactions of FMN-binding domains in protein complexes. However, structural properties of these flavinylated quaternary structures is largely unexamined, with the exception for

previously mentioned complexes whose structures were previously resolved, namely the Nqr, Rnf, and PepSY systems (Josts et al., 2021; Kishikawa et al., 2022; Steuber et al., 2014; Vitt et al., 2022; Zhang and Einsle, 2022). To generate initial insights for flavinylated proteins in this context, we leveraged AlphaFold-Multimer to predicted structures for flavinylated membrane complexes.

In the previously resolved structure for the Nqr complex, the six subunits form a semi-circular path for electron transfer, shuttling electron from cytosolic NADH to the flavinylated NqrC subunit, then to the quinone terminal electron acceptor on the cytosolic side of the membrane (Kishikawa et al., 2022; Steuber et al., 2014). This process is energetically coupled with transport of ions, maintaining an electromotive gradient across the membrane (**Figure 3.5A**). As a complex that shares homology to the Nqr complex, the Rnf compelx preserved similar patterns for complex structures and electron transfer mechanisms (**Figure 3.5B**) (Vitt et al., 2022; Zhang and Einsle, 2022).

We previously identified a group of flavinylated complexes that are partially homologous to Nqr/Rnf complexes. These Nqr/Rnf-like complexes only contain two subunits: an extracytosolic subunit that is homologous to NqrC or RnfG, as well as a second subunit whose N-terminal region encode for a NqrB/RnfD domain while the N-terminal region is predicted as an NAD-binding domain (**Figure 3.5C&E**). In addition, due to the absence of other Nqr/Rnf subunit equivalents, we previously proposed that these Nqr/Rnf-like complexes are only capable of unidirectional electron transfer, likely using cytosolic NAD(P)H as an electron donor and depositing electrons onto extracytosolic electron acceptors. We particularly examined the AlphaFold-Multimer-derived structure for one of these Nqr/Rnf-like structures encoded by *Ktedonbacter racemifer* (**Figure 3.5C**). This reveals that the two subunits are in close spatial proximity with each other. Interestingly, such intimate interaction between their Nqr/Rnf homologs (i.e., between NqrC/RnfG

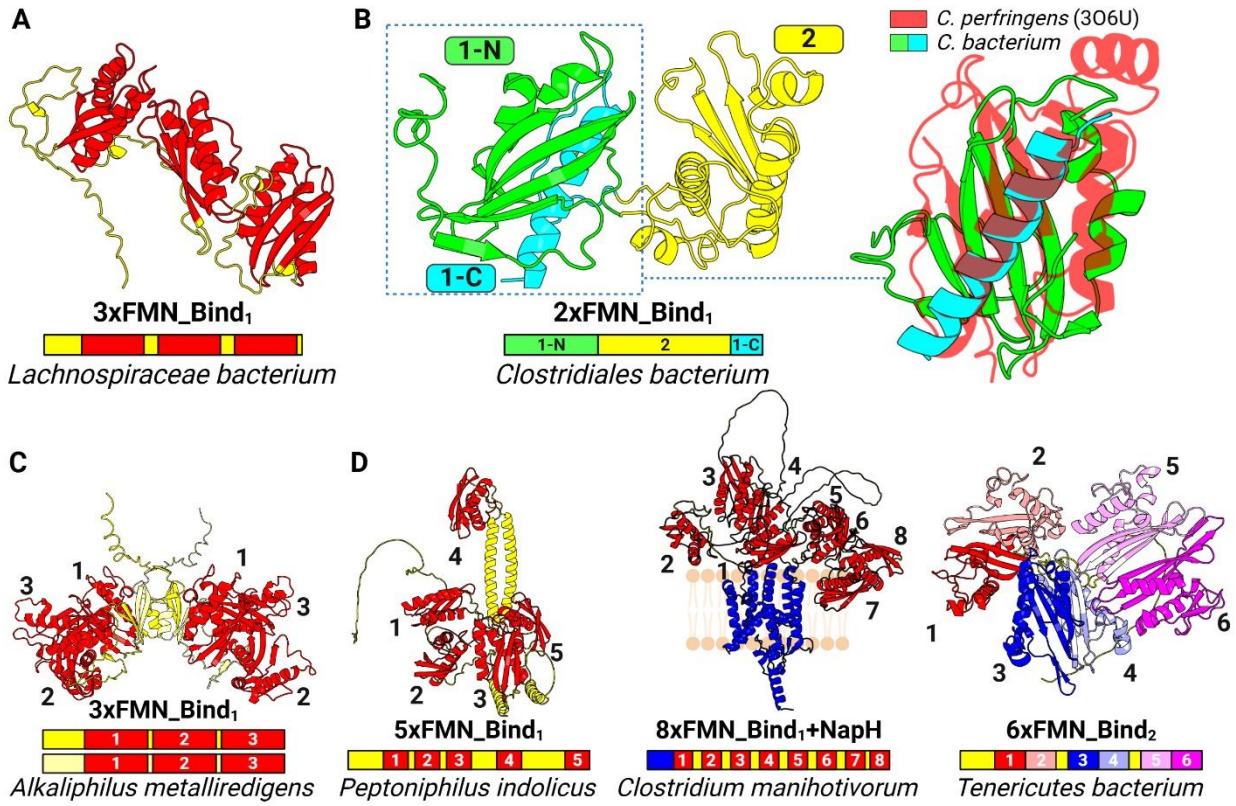


Figure 3.4 AlphaFold models of multi-flavinylated proteins.
(A) Predicted structure of a "beads-on-a-string-like" 3×FMN-bind$_1$ protein. (B) Predicted structure of a circularly permuted double-flavinylated protein. The N-terminus and C-terminus regions predicted to form the first FMN-bind$_1$ domain are green and cyan, respectively, whereas the second FMN-bind$_1$ domain is highlighted in yellow (left). Circularly permuted FMN-bind$_1$ domain is structurally aligned with previously resolved crystal structure of a stand-alone FMN-bind1 protein (PDB: 3O6U; right). (C) Predicted structure of a dimerized 3×FMN-bind$_1$ protein. (D) Predicted structures of flavinylated proteins with more than three FMN-bind$_1$ or FMN-bind$_2$ domains.

and NqrB/RnfD) is absent in the original Nqr and Rnf complexes. Furthermore, the NqrB/RnfD-like N-terminal domain of the transmembrane subunit also lacks a flavinylation site that is otherwise conserved in NqrB and RnfD family proteins (**Figure 3.5G**). These observations led us to propose that the unique interaction between the two subunits in the Nqr/Rnf-like complex brings the extracytosolic subunit's flavinylation site in close proximity to the apparent non-covalent flavin-binding site in the transmembrane subunit, implying that flavin covalently bound to the extracytosolic subunit may conformationally sample the non-covalent flavin-binding site in the transmembrane domain (**Figure 3.5D&F**).

3.3.5 Membrane cytochromes that are in association with extracytosolic flavinylation

We identified that the interaction between extracytosolic flavinylated domains and transmembrane cytochrome proteins is a common scheme among flavinylated membrane complexes. Two prominent examples we previously described are the PepSY-like and MsrQ-like systems (**Figure 3.6A&B**). In these systems, the membrane component is predicted to possess conserved pairs of histidine residues, which serve as binding sites for heme. In the recently resolved crystal structure of a PepSY-like system encoded by *Pseudomonas aeruginosa*, the membrane component possess 2 pairs of highly conserved histidine residues, allowing binding for 2 heme cofactors (Josts et al., 2021). The MsrQ-like system in turn shares structural homology to the PepSY-like system in this aspect. While MsrQ-like and PepSY-like membrane components share low sequence similarity, the putative heme-binding positions are structurally conserved (**Figure 3.6C**). These two systems exemplify a group of flavinylation-associated electron transfer systems that utilize a transmembrane cytochrome as a core apparatus.

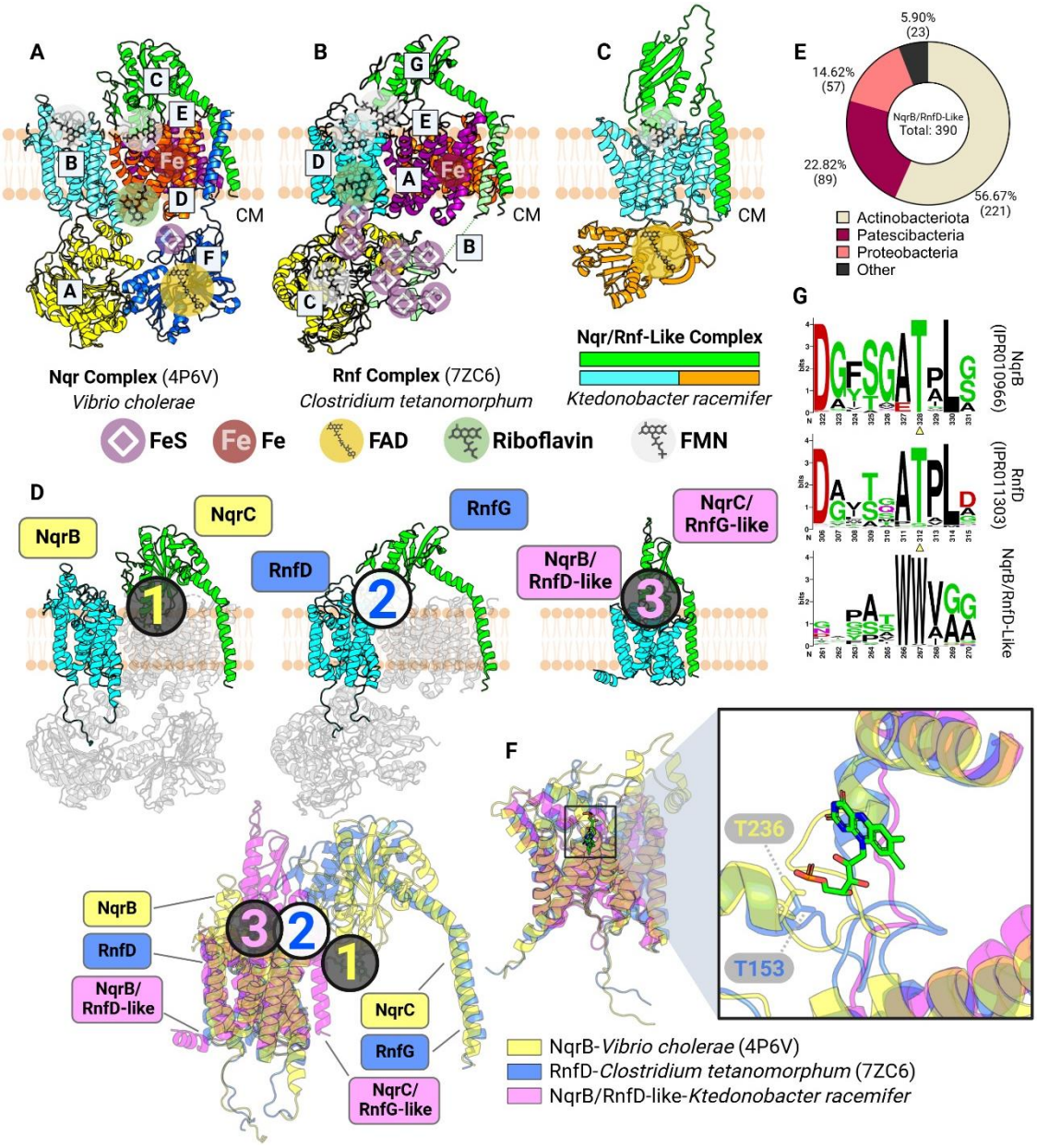


Figure 3.5 Structural heterogeneity of flavinylation-associated transmembrane proteins. (A&B) Previously resolved crystal structures of the Nqr complex (A; PDB: 4P6V) and the Rnf complex (B; PDB: 7ZC6). The cytoplasm (CM) is indicated. (C) AlphaFold-multimer model of the Nqr/Rnf-like complex containing a transmembrane subunit and a membrane-anchored extracytosolic subunit that share homology with corresponding subunits in Nqr and Rnf systems. (D) Structural alignment of Nqr, Rnf, and Nqr/Rnf-like complexes based on transmembrane subunits homologous across the three complexes (cyan). Subunits that do not share homology to subunits of the Nqr/Rnf-like complex are transparent (top). Predicted or confirmed FMN moieties are highlighted by circled numbers, color-coded by their corresponding complex (bottom). (E) Taxonomic distribution of the NqrB/RnfD-like complexes in bacteria. (F) Structural alignment of NqrB, RnfD, and NqrB/RnfD-like subunits (left) with zoom-in view showing the FMN moiety from NqrB and the Thr residue in NqrB or RnfD responsible for FMN-binding (right). (G) HMM logos showing conserved Thr residues in NqrB and RnfD but not in the NqrB/RnfD-like subunit.

Given the prevalence of this flavinylation-cytochrome interaction in our database, we considered the possibility of exploiting this common scheme for identification of novel membrane cytochromes. We tackled this question with a two-step approach. First, we examined flavinylation-associated gene clusters that did not co-encode with a known transmembrane component. These gene clusters would thus partake in extracytosolic transfer through uncharacterized mechanisms. Second, we determined if genes in these clusters possess conserved histidine pairs, which could serve as putative binding sites for heme. This led us to identify proteins that possess a domain of unknown function, DUF4405 (**Figure 3.7A, yellow**). Our bioinformatic screen showed that ApbE-asscoiated DUF4405 proteins are frequently found in Proteobacteria and Firmicutes species (**Figure 3.7B**). These gene clusters also commonly encode known electron transfer domains, such as flavinylated Flavodoxin_4 domains or a previously characterized transporter responsible for flavin trafficking in *Listeria monocytogene* (**Figure 3.7C**) (Rivera-Lugo et al., 2023). These results suggest that DUF4405 proteins are likely involved in flavinylation-based electron transfer.

We next examined the AlphaFold-predicted structures for a selected DUF4405 candidates found in *Oscillospiraceae bacterium*. This protein possesses a six-transmembrane structure containing 4 conserved histidine residues in the center, which is similar to our observations for PepSY-like and MsrQ-like proteins. Consistently, the sequence of DUF4405 proteins shares low consensus with PepSY-like or MsrQ-like proteins, but the two heme-binding sites are structurally conserved across the three proteins (**Figure 3.7A&D**). These data suggest the previously uncharacterized DUF4405 domain may function as a novel membrane cytochrome protein.

To test this hypothesis, we heterologously expressed a recombinant DUF4405 protein in *E. coli* and subsequently purified the protein. Interestingly, the eluent sample containing overexpressed DUF4405 appeared to have a brown/red-ish color, which is typically seen for

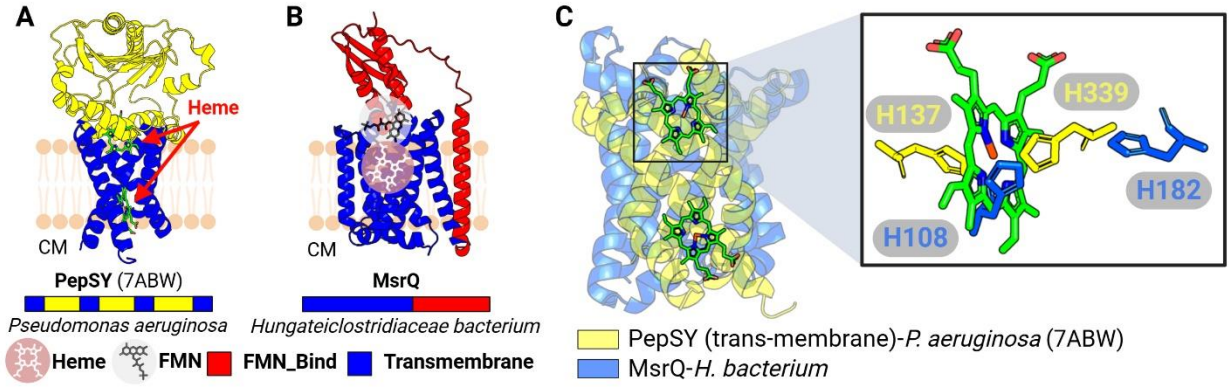


Figure 3.6 Membrane cytochromes associated with flavinylated proteins.
(A) Previously resolved crystal structure of the PepSY complex (PDB: 7ABW). The cytoplasm (CM) is indicated. (B) AlphaFold model of a protein containing the transmembrane cytochrome MsrQ (red) and an FMN-Bind$_1$ domain (red). (C) Structural alignment between transmembrane segments of PepSY and MsrQ (left) with zoom-in view on the heme and axial histidine residues from PepSY and MsrQ.

purified cytochrome protein (**Figure 3.7E, top-left**). We additionally repeated the experimental procedure for a mutant variant of DUF4405 with all 4 conserved histidine residues replaced with alanine. Purified protein samples for this mutant appeared to be transparent, which suggests that the conserved histidine pairs of DUF4405 are indeed responsible for binding of heme (**Figure 3.7E, bottom-left**).

To further ascertain our findings, we performed spectroscopic analysis for purified DUF4405 samples. Previous studies have reported a characteristic absorption pattern for heme B cytochromes, characterized by elevated absorption at the γ peak (hemoprotein soret peak) and α/β peaks (indicative of porphyrin ring transition) under a reduced form. Consistently, we observed this characteristic absorption pattern for DUF4405 (**Figure 3.7F, top**). When taken together, the results demonstrate that DUF4405 proteins as novel membrane cytochrome proteins that likely interact with extracytosolic flavinylated domains for electron transfer.

We further extended our findings on DUF4405 proteins and aimed to identify other proteins that share homology with DUF4405. In our previous analysis, we set the presence of the *apbE* gene as a screening criterion for identification of flavinylation-associated cytochromes. Here, we removed this requirement and sought to identify other novel DUF4405-like cytochrome proteins that may not necessarily associate with flavinylation. Interestingly, this led us to FezC (**Figure 3.7A, red**). Previous report on a FezC protein from *Desulfovibrio magneticus* describe FezC as a small protein involved in iron trafficking in ferrozomes, which are mrmbane-enclosed organelles that serve as intracellular iron reservoirs (Grant et al., 2022; Pi et al., 2023). We found that FezC shares sequence similarity with DUF4405. AlphaFold-derived structural prediction revealed that FezC adopted a helical structure containing 2 conserved histidine residues, but it only shares structural homology with DUF4405 in its dimerized form where the 2 histidine residues are

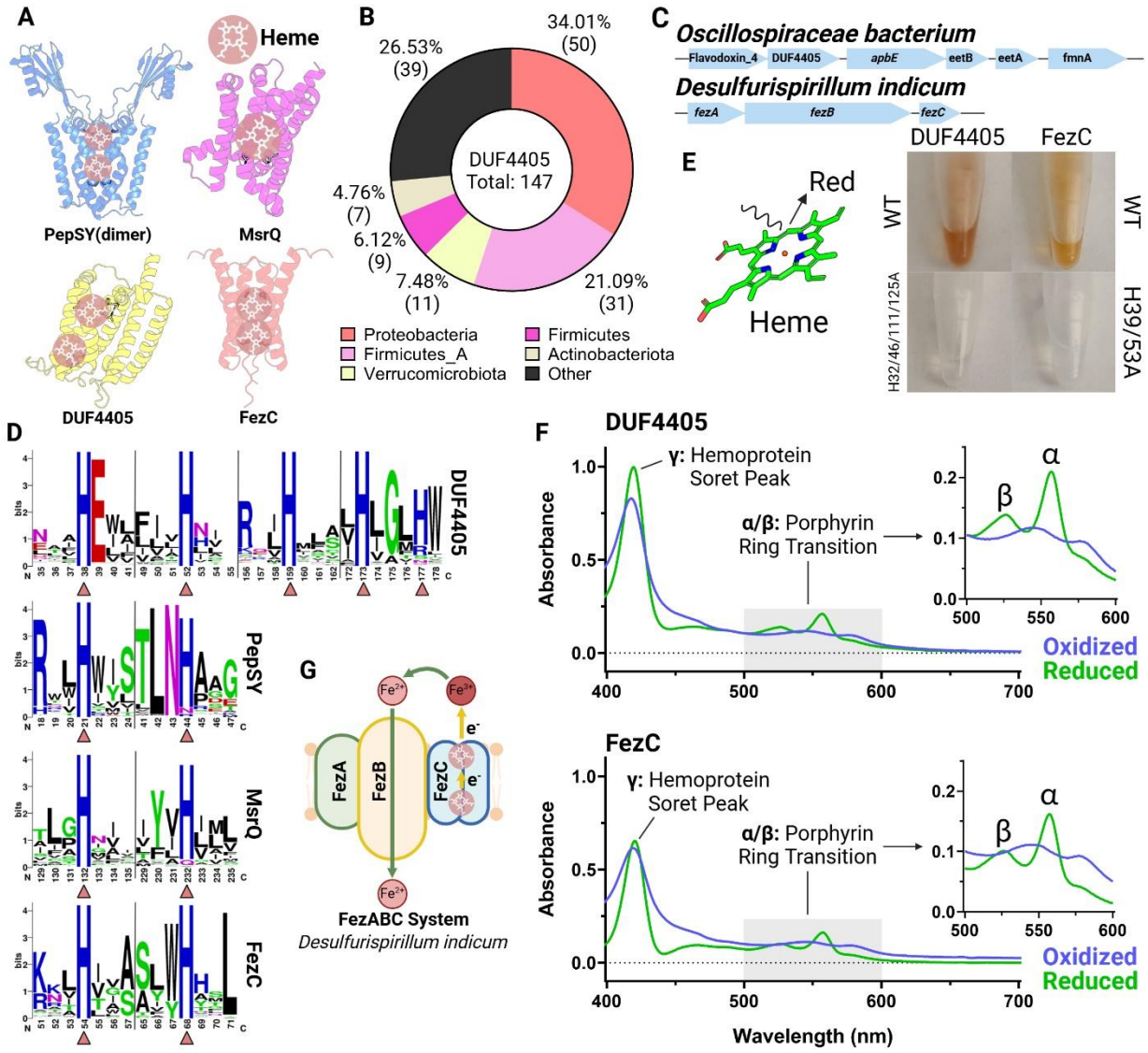


Figure 3.7 DUF4405 and FezC are novel cytochromes.
(A) AlphaFold models of PepSY (homodimer), MsrQ, DUF4405, and FezC (homodimer). Histidine residues (His) responsible for heme binding are highlighted. (B) Taxonomic distribution of DUF4405 genes contained in a cluster with apbE. (C) Gene clusters in an *Oscillospiraceae bacterium* or *Desulfurispirillum indicum* encoding DUF4405 or FezC, respectively. (D) HMM logos showing conserved heme-binding His residues in DUF4405, PepSY, MsrQ, and FezC. Triangles highlight His residues shown in panel A. (E) Purified DUF4405 and FezC proteins. (F) UV spectrums of DUF4405 (top) and FezC (bottom) showing absorption peaks characteristic of heme B binding. (G) Possible role of FezC in iron transport within ferrosomes.

spatially paired with the other 2 histidine residues on the second subunit. Repeating similar in vitro procedures generated results that mimic that of DUF4405: the protein eluent containing wild-type FezC appeared to have a brown/red-ish color, whereas the histidine-to-alanine mutant variant appeared to be transparent (**Figure 3.7E, right**). Spectroscopic analysis also demonstrated a consistent—though weaker—absorption pattern that is characteristic of heme B cytochromes (**Figure 3.7F, bottom**). These results collectively demonstrate that the FezC homodimer is a ferrsome-associated novel cytochrome protein.

3.4 Discussion

Our work described in this chapter serves as an extension to our previous bioinformatic screen of flavinylation, with an emphasis on structural and functional characterizations of flavinylated proteins and flavintlation-associated proteins. This work relies on the combination of comparative genomics and structural predictions powered by AlphaFold. Our work serves as an example for how such combination could be leveraged for novel protein discovery. Our results offer novel insights into the molecular basis of flavinylation, revealing diverse cellular contexts in which flavinylated proteins situate.

Our findings on DUF4405 and FezC as a novel membrane cytochrome protein demonstrated that our approach could be modified to enable identification of novel proteins that are indirectly associated with our initial screening criteria. This result highlights that large scale structural predictions could be incorporated with pre-existing genomic tools as a new facet for protein discovery. In addition, while FezC is a previously studied protein prior to our analysis, our data further suggest that FezC dimer may similarly function as a membrane cytochrome. This

finding further demonstrates that the combination of structural and genomic analysis could enhance the resolution of our understanding of characterized electron-transferring proteins.

These contributions of our work are also reflected by our findings on flavinylated Flavodoxin_4 and FMN_red proteins. Proteins in these families are largely known for non-covalent flavin-binding activities, but our results demonstrate that flavinylated members from these families adopted addition structures encompassing the covalent FMN-binding site. This led to the implication for evolution events, through which non-covalent FMN-binding in Flavodoxin_4 and FMN_red proteins evolved to possess the flavinylation site, which secures FMN under dynamic extracytosolic environments.

In summary, this study provides an example of how large-scale structural analysis could improve protein discovery and functional characterizations. In our focus on flavinylated proteins, we revealed diverse structural, functional, and cellular contexts of flavinylation, presenting a great versatility of flavinylated proteins. Future research is required to uncover the underlying mechanisms involved in flavinylation-based electron transfer as well as its contribution to bacterial physiology, particularly in species that are more pertinent for human health and disease.

3.5 Methods

3.5.1 Identification of flavinylated protein candidates

The flavinylation-associated proteins FMN_red, DUF4405, and flavodoxin were identified by searching their Pfam accession numbers (PF03358, PF14358, and PF12682, respectively) in the proteomes from 47,894 functionally annotated bacterial and archaeal genomes from the GTDB (release 202) (Parks et al., 2018). FMN-bind$_2$ domains were identified through BLAST searches. Briefly, protein sequences were functionally annotated based on the Pfam accession number (Pfam

database v.33.0) (Mistry et al., 2021) of their best match using Hmmsearch (v.3.3.2, E value cutoff of 0.001) (Eddy, 1998). The five genes downstream and upstream of genes FMN_red, DUF4405, or flavodoxin were collected for further analyses. InterPro accession numbers, taxonomic assignments, and amino acid sequences of flavinylated candidates presented in this study are included in **Table 1**.

3.5.2 Protein model prediction by AlphaFold2

Predicted three-dimensional models for selected flavinylated protein monomers or complexes were generated using AlphaFold2 and AlphaFold2-multimer (ColabFold v.1.5.2) (Evans et al., 2021; Jumper et al., 2021; Mirdita et al., 2022). Model metrics are provided in **Figure 3.8**. PDB files containing predicted structures were visualized, examined, or aligned using PyMOL Molecular Graphics System, Version 3.0 (Schrödinger, LLC).

3.5.3 In vitro confirmation of flavodoxin and FMN reductase flavinylation

3.5.3.1 *E. coli* expression strains

DNA fragments containing wild-type or point mutant ORFs of flavodoxins (*Anaerocolumna xylanovorans*, NCBI accession number SHO45324.1; *Desulfosporosinus lacus*, NCBI accession WP_073032509.1) and FMN reductases (*Lactococcus lactis*, NCBI accession WP_021723379.1; *Lactococcus taiwanensis*, NCBI accession WP_205872264.1) were synthesized using IDT gBlocks (Integrated DNA Technologies). Primers with overhanging sequences homologous to either the 5’ or 3’ end of target gene fragments were used to linearize pMCSG53 expression vectors at the multiple cloning sites through PCR reactions (Q5 High-Fidelity 2X Master Mix, New England Biolabs). Amplicons were subsequently gel-extracted (Wizard SV Gel and PCR Clean-Up System, Promega), quantified, and combined with corresponding gene inserts in Gibson reactions (NEBuilder HiFi DNA Assembly Master Mix, New

Species	Name	NCBI Accession	PDB Code	InterPro ID
Clostridium perfringens	FMN_Bind$_1$	3O6U_A	3O6U	IPR007329
Erysipelotrichaceae bacterium	FMN_Bind$_2$	NTW96865.1		
Vibrio cholerae	NqrB	4P6V_B	4P6V	IPR010966
Shewanella oneidensis	DUF2271	AAN54259.1		IPR014469
Anaerocolumna xylanovorans	Flavodoxin_4	SHO45324.1		IPR008254
Lactococcus lactis	FMN_Reductase	WP_021723379.1		IPR005025
Desulfosporosinus lacus	Flavodoxin_4	WP_073032509.1		IPR008254
Bacteroides fragilis	Flavodoxin_4	3KLB_A	3KLB	IPR008254
Lactococcus taiwanensis	FMN_Reductase	WP_205872264.1		IPR005025
Escherichia coli	FMN_Reductase	4PTZ_A	4PTZ	IPR005025
Lachnospiraceae bacterium	3xFMN_Bind$_1$	ETP72910.1		
Clostridiales bacterium	2xFMN_Bind$_1$	MBS6398205.1		
Alkaliphilus metalliredigens	3xFMN_Bind1	WP_012062870.1		
Peptoniphilus indolicus	5xFMN_Bind$_1$	WP_115312035.1		
Hujiaoplasma nucleasis	6xFMN_Bind$_2$	WP_312031774.1		
Clostridium manihotivorum	8xFMN_Bind$_1$	WP_128212410.1		
Vibrio cholerae	NqrA	4P6V_A	4P6V	IPR008703
Vibrio cholerae	NqrC	4P6V_C	4P6V	IPR010204
Vibrio cholerae	NqrD	4P6V_D	4P6V	IPR011292
Vibrio cholerae	NqrE	4P6V_E	4P6V	IPR010967
Vibrio cholerae	NqrF	4P6V_F	4P6V	IPR010205
Clostridium tetanomorphum	RnfA	7ZC6_A	7ZC6	IPR011293
Clostridium tetanomorphum	RnfB	7ZC6_B	7ZC6	IPR010207
Clostridium tetanomorphum	RnfC	7ZC6_C	7ZC6	IPR010208
Clostridium tetanomorphum	RnfD	7ZC6_D	7ZC6	IPR011303
Clostridium tetanomorphum	RnfE	7ZC6_E	7ZC6	IPR010968
Clostridium tetanomorphum	RnfG	7ZC6_G	7ZC6	IPR010209
Ktedonobacter racemifer	NqrB/RnfD-Like	WP_007920854.1		IPR004338, IPR001433
Ktedonobacter racemifer	NqrC/RnfG-Like	WP_007920858.1		
Pseudomonas aeruginosa	PepSY	7ABW_A	7ABW	IPR005625
Hungateiclostridiaceae bacterium	MsrQ	ANU53938.2		
Oscillospiraceae bacterium	DUF4405	MBD5117352.1		IPR025517
Desulfurispirillum indicum	FezC	WP_013506634.1		

Table 1 List of analyzed proteins in Chapter 3.

England Biolabs) to allow integration of targeted genes. Expression constructs were then transformed into *E. coli* BL21, and successful transformants were selected on lysogeny broth (LB) agar containing 100 mg/mL of carbenicillin. LB cultures of transformant colonies were supplemented with 15% (wt/vol) glycerol and stored in −80°C until use.

3.5.3.2 Purification of FMN transferase ApbE from Listeria monocytogenes

To ensure consistent flavinylation activity, we developed an in vitro flavinylation assay using the previously characterized FMN transferase ApbE protein encoded by *Listeria monocytogenes* 10403S (lm_ApbE) (Light et al., 2018). *E. coli* BL21 expression strains containing pMCSG53::*lm_apbE* expression constructs were generated through steps similar to those mentioned above with 6xHis-tag at the N-terminus. To purify lm_ApbE, overnight cultures of the expression strain were diluted to an optical density at 600 nm (OD600) of 0.05 in 1 L of LB and incubated at 37°C with aeration. After 2 h of incubation, a final concentration of 1 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) was added to allow induction of lm_ApbE expression at 30°C overnight. Cell pellet was then collected through centrifugation at 7,000 × g for 15 min and frozen at −80°C overnight. The cell pellet was resuspended in a solution containing 50 mM of Tris-HCl, pH = 7.5, 300 mM of NaCl, and 10 mM of imidazole at a volume that is five times the weight of the cell pellet. The resulting mixture was lysed through sonication (8 × 30 s pulses) and cleared by centrifugation at 40,000 × g for 30 min. Supernatants of cell lysates were passed through a nickel bead column (Profinity IMAC Ni-Charged Resin, Bio-Rad) to allow binding of lm_ApbE-6×His, which was then eluted with 500 mM imidazole. Successful elution of lm_ApbE-6×His was confirmed through 12% SDS-PAGE. Filtrate samples were then purified using a ÄKTA pure chromatography FPLC system (Cytiva). Elution fractions containing lm_ApbE-6×His were subsequently concentrated (4,000 × g; Pierce Protein Concentrators PES,

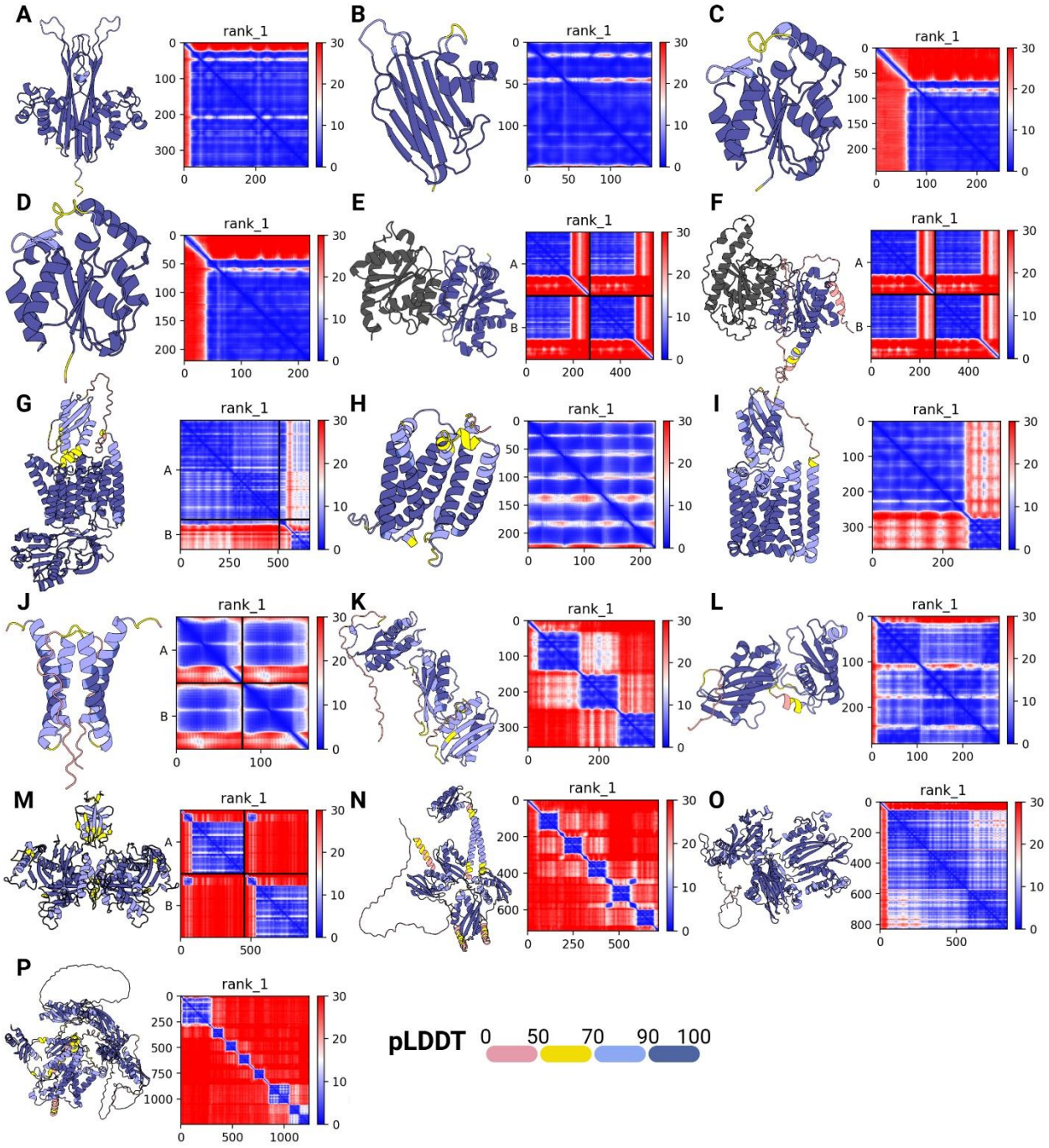


Figure 3.8 Quality of AlphaFold-predicted structures.
(A-P) AlphaFold-predicted structures colored in predicted local distance difference test scores (pLDDT; left) as well as heatmaps of predicted aligned error (PAE; right). (A) FMN_Bind2; (B) DUF2271; (C) Flavodoxin_4; (D) FMN_Reductase; (E) Flavodoxin_4 DL; (F) FMN_Reductase LT; (G) Nqr/Rnf-Like; (H) DUF4405; (I) MsrQ; (J) FezC; (K) 3xFMN_Bind1; (L) 2xFMN_Bind1; (M) 3xFMN_Bind1; (N) 5xFMN_Bind1; (O) 8xFMN_Bind1; (P) 6xFMN_Bind2.

10K MWCO, Thermo Scientific) and quantified using a spectrophotometer (DS-11 FX+ spectrophotometer, DeNovix).

3.5.3.3 In vitro expression and flavinylation of flavodoxin and FMN reductase candidates

To confirm in vitro covalent binding of FMN on target candidate proteins, overnight cultures of *E. coli* BL21 strains containing the corresponding expression vectors mentioned above were reinoculated in 3 mL of LB and grown in the presence of 1 mM IPTG with aeration at 30°C overnight. Overnight cultures were then diluted to an optical density of OD600 = 0.5 and centrifuged for 1 min at 21,100 × g. Resulting cell pellets were resuspended in 100 μL of lysis buffer (500 μg/mL of lysozyme, 300 mM of NaCl, and 10 mM of imidazole in 50 mM of Tris-HCl, pH = 7.5) and incubated on ice for 30 min. Cell lysates were then combined with 0.3 μM of lm_ApbE, 1 mM of FAD, and 5 mM of MgSO4 and incubated at 4°C overnight with rotation to enable flavinylation. Reaction mixtures were then separated into aqueous or solid phases by centrifugation at 21,100 × g for 1 min and subsequently incubated at 98°C for 10 min. Both aqueous and solid portions (resuspended in 100 μL of lysis buffer) were then run on 12% SDS-PAGE. To confirm successful flavinylation, we leveraged the UV resonance property of the isoalloxazine ring of the FMN moiety, which led to a bright band at the expected molecular weight for targeted proteins when the SDS-PAGE gel is visualized under UV (iBright 1500 imaging system, Invitrogen).

3.5.4 In vitro confirmation of heme-binding activity in FezC and DUF4405

Cloning, expression, and purification of FezC (*Desulfurispirillum indicum*, WP_013506634.1) or DUF4405 (*Oscillospiraceae bacterium*, MBD5117352.1) were done in similar procedures as lm_ApbE, except that the solid phase of cell lysates was used for downstream purification because FezC and DUF4405 are membrane proteins. Proteins in the solid phase of cell

lysates were solubilized using a previously published protocol (Kupke et al., 2020). Briefly, pelleted cell lysates were resuspended in a solution containing 50 mM of Tris-HCl, pH = 7.5, 300 mM of NaCl, 10 mM of imidazole, and 1% (wt/vol) lauryldimethylamine oxide (LDAO) and were subsequently purified as previously described using nickel bead column and FPLC (eluted with 500 mM imidazole + 0.1% [wt/vol] LDAO). Heme-binding activity of purified FezC or DUF4405 was confirmed using a previously published protocol for pyridine hemochromagen assay (Barr and Guo, 2015). Briefly, samples containing 1 mg/mL of purified FezC or DUF4405 were mixed with a solution containing 0.2 M NaOH, 40% (vol/vol) pyridine, and 500-µM potassium ferricyanide to oxidize protein samples. Oxidized proteins were then measured for their absorbance at 300–700 nm. Samples were then combined with a reducing solution containing 0.5 M sodium dithionite in 0.5 M NaOH to acquire reduced FezC or DUF4405, which were then similarly examined for its absorbance at the same range of wavelength.

3.6 References

Barr I, Guo F. 2015. Pyridine Hemochromagen Assay for Determining the Concentration of Heme in Purified Protein Solutions. *BIO-PROTOCOL* **5**. doi:10.21769/BioProtoc.1594

Berthomieu R, Pérez-Bernal MF, Santa-Catalina G, Desmond-Le Quéméner E, Bernet N, Trably E. 2022. Mechanisms underlying Clostridium pasteurianum's metabolic shift when grown with Geobacter sulfurreducens. *Appl Microbiol Biotechnol* **106**:865–876. doi:10.1007/s00253-021-11736-7

Bewley KD, Ellis KE, Firer-Sherwood MA, Elliott SJ. 2013. Multi-heme proteins: Nature's electronic multi-purpose tool. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1827**:938–948. doi:10.1016/j.bbabio.2013.03.010

Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics* **14**:755–763. doi:10.1093/bioinformatics/14.9.755

Edwards MJ, White GF, Butt JN, Richardson DJ, Clarke TA. 2020. The Crystal Structure of a Biological Insulated Transmembrane Molecular Wire. *Cell* **181**:665-673.e10. doi:10.1016/j.cell.2020.03.032

Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D. 2021. Protein complex prediction with AlphaFold-Multimer. doi:10.1101/2021.10.04.463034

Grant CR, Amor M, Trujillo HA, Krishnapura S, Iavarone AT, Komeili A. 2022. Distinct gene clusters drive formation of ferrosome organelles in bacteria. *Nature* **606**:160–164. doi:10.1038/s41586-022-04741-x

Hayashi M, Nakayama Y, Yasui M, Maeda M, Furuishi K, Unemoto T. 2001. FMN is covalently attached to a threonine residue in the NqrB and NqrC subunits of Na^+-translocating NADH-quinone reductase from *Vibrio alginolyticus*. *FEBS Letters* **488**:5–8. doi:10.1016/S0014-5793(00)02404-2

Josts I, Veith K, Normant V, Schalk IJ, Tidow H. 2021. Structural insights into a novel family of integral membrane siderophore reductases. *Proc Natl Acad Sci USA* **118**:e2101952118. doi:10.1073/pnas.2101952118

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. doi:10.1038/s41586-021-03819-2

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**:845–858. doi:10.1038/nprot.2015.053

Kishikawa J, Ishikawa M, Masuya T, Murai M, Kitazumi Y, Butler NL, Kato T, Barquera B, Miyoshi H. 2022. Cryo-EM structures of Na+-pumping NADH-ubiquinone oxidoreductase from Vibrio cholerae. *Nat Commun* **13**:4082. doi:10.1038/s41467-022-31718-1

Kupke T, Klare JP, Brügger B. 2020. Heme binding of transmembrane signaling proteins undergoing regulated intramembrane proteolysis. *Commun Biol* **3**:73. doi:10.1038/s42003-020-0800-0

Light SH, Su L, Rivera-Lugo R, Cornejo JA, Louie A, Iavarone AT, Ajo-Franklin CM, Portnoy DA. 2018. A flavin-based extracellular electron transfer mechanism in diverse Gram-positive bacteria. *Nature* **562**:140–144. doi:10.1038/s41586-018-0498-z

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nat Methods* **19**:679–682. doi:10.1038/s41592-022-01488-1

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**:D412–D419. doi:10.1093/nar/gkaa913

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**:996–1004. doi:10.1038/nbt.4229

Pi H, Sun R, McBride JR, Kruse ARS, Gibson-Corley KN, Krystofiak ES, Nicholson MR, Spraggins JM, Zhou Q, Skaar EP. 2023. Clostridioides difficile ferrosome organelles combat nutritional immunity. *Nature* **623**:1009–1016. doi:10.1038/s41586-023-06719-9

Rivera-Lugo R, Huang S, Lee F, Méheust R, Iavarone AT, Sidebottom AM, Oldfield E, Portnoy DA, Light SH. 2023. Distinct Energy-Coupling Factor Transporter Subunits Enable Flavin Acquisition and Extracytosolic Trafficking for Extracellular Electron Transfer in Listeria monocytogenes. *mBio* **14**:e03085-22. doi:10.1128/mbio.03085-22

Rohl CA, Strauss CEM, Misura KMS, Baker D. 2004. Protein Structure Prediction Using RosettaMethods in Enzymology. Elsevier. pp. 66–93. doi:10.1016/S0076-6879(04)83004-0

Steuber J, Vohl G, Casutt MS, Vorburger T, Diederichs K, Fritz G. 2014. Structure of the V. cholerae Na+-pumping NADH:quinone oxidoreductase. *Nature* **516**:62–67. doi:10.1038/nature14003

Vitt S, Prinz S, Eisinger M, Ermler U, Buckel W. 2022. Purification and structural characterization of the Na+-translocating ferredoxin: NAD+ reductase (Rnf) complex of Clostridium tetanomorphum. *Nat Commun* **13**:6315. doi:10.1038/s41467-022-34007-z

Zhang L, Einsle O. 2022. Architecture of the NADH:ferredoxin oxidoreductase RNF that drives Biological Nitrogen Fixation. doi:10.1101/2022.07.08.499327

Chapter 4: Preliminary insights on multi-flavinylated proteins and their relevance to health

4.1 Abstract

Multi-cofactor proteins involved in extracytosolic electron transfer establish a longer insulated circuit across the cell envelope. Multi-flavinylated proteins are prevalently found in major bacterial phyla with previously reported relevance to human health. Here, we report results from our preliminary characterizations of multi-flavinylated proteins. We identify a mono-flavinylated protein that may serve as an ancestral "template" for multi-flavinylated homologs, which suggests possible evolution events where multi-flavinylated proteins evolved through duplication of pre-existing FMN-binding domains. In addition, we present experimental evidence for stepwise electron transfer between FMN-binding domains in the PplA protein from the enteric pathogen *Listeria monocytogenes.* We further characterize a PplA homolog in the opportunistic intestinal pathogen *Enterococcus faecalis* that share similarity in sequence, structure, and function with *Listeria* PplA, serving as another evidence for involvement of multi-flavinylated proteins in species related to human gut health. And finally, we determined the prevalence and abundance of flavinylation markers in previously published human gut microbiome metagenomic data and found that both ApbE and FMN-binding domain are commonly encoded by gut microbes, detected in over 90% of analyzed human stool samples. These results offer preliminary insights into the biological and molecular contexts of multi-flavinylation, which establish a foundation for future research that aims to provide mechanistic understandings of multi-flavinylated proteins.

4.2 Introduction - Biological contexts of multi-flavinylated proteins

In our bioinformatic screening for flavinylation markers, we identified over 2,081 proteins from 1,530 genomes that encode 2 or more FMN-binding domains, with the highest observed

multiplicity of putative FMN-binding domains being 13. Taxonomic distribution of these genomes reveals that over 68% of identified genomes (1,044/1530) belong to species under the Firmicutes phylum. This observation exposed several gaps in our current knowledge on the biological contexts of multi-flavinylated proteins. First, acquiring a higher number of FMN-binding domains could be an intricate evolution process. Given that multiple FMN-binding domains within the same coding region could be functionally redundant, the necessity of multiple FMN-binding domains gives rise to questions on the mechanisms through which multi-flavinylated proteins acquired additional domains. Second, while we hypothesized that multi-flavinylated proteins may functionally mimic multi-heme cytochromes and shuttles electrons sequentially (Edwards et al., 2020), it is unclear whether multiple FMN moieties in the same proteins facilitate electron transfer in a cooperative, stepwise manner or function independently of each other. And third, the high prevalence of multi-flavinylated proteins in the Firmicutes phylum implies possible relevance of these proteins in bacterial species that are closely associated with human health, as many of the members in the Firmicutes phylum reside in the human gut microbiome (Hou et al., 2022; Sweeney and Morton, 2013; Wang et al., 2022).

The previous finding on the double-flavinylated PplA protein in *Listeria monocytogenes* offered pioneering answers to these questions. As an iron-reducing protein anchored on the extracytosolic face of the membrane, PplA is hypothesized to transfer electrons between cytosolic NADH and external soluble ferric irons. The topology of the PplA protein posits the first FMN-binding domain in proximity to the membrane while presenting the second FMN-binding domain in a position further away from the cell surface. This suggests that PplA could transfer electrons sequentially between its two FMN moieties, as the first FMN-binding domain is closer to cytoplasm-derived electrons, whereas the second FMN-binding domain is more accessible to

extracytosolic iron. Furthermore, as an enteric pathogen, *L. monocytogenes* also belongs to the Firmicutes phylum. Genetic manipulations that disrupted the *pplA* gene cluster reduced *L. monocytogenes* colonization in mice, which further suggests that multi-flavinylated proteins could play critical roles in the physiology of bacterial species with greater relevance to human health or disease (Light et al., 2018).

In this chapter, we describe our preliminary efforts in uncovering biological contexts of multi-flavinylated proteins. We report that multi-flavinylated proteins likely acquired additional FMN-binding domains through duplication of genomic sequences of FMN-binding domains from pre-existing mono-flavinylated proteins. In addition, we scrutinized the mode of electron transfer in PplA protein using genetic approaches, and subsequently identified a PplA-homolog in another pathogenic species. This data further emphasized the significant involvement of multi-flavinylated proteins in health-associated species. And finally, we screened for genomic markers of flavinylation in publicly available human gut microbiome metagenomic data (Mehta et al., 2018), revealing the prevalence of flavinylation in a biological context that is highly relevant to human health.

4.3 Results

4.3.1 Duplication of FMN-binding domains in multi-flavinylated proteins

To determine whether multi-flavinylated proteins acquired their additional FMN-binding domains simply through duplication of pre-existing domains, we first extracted from all 2,081 proteins the peptide sequences encompassing the identifiable FMN-binding domains and performed multiple sequence alignment. This allowed us to examine the sequence similarity

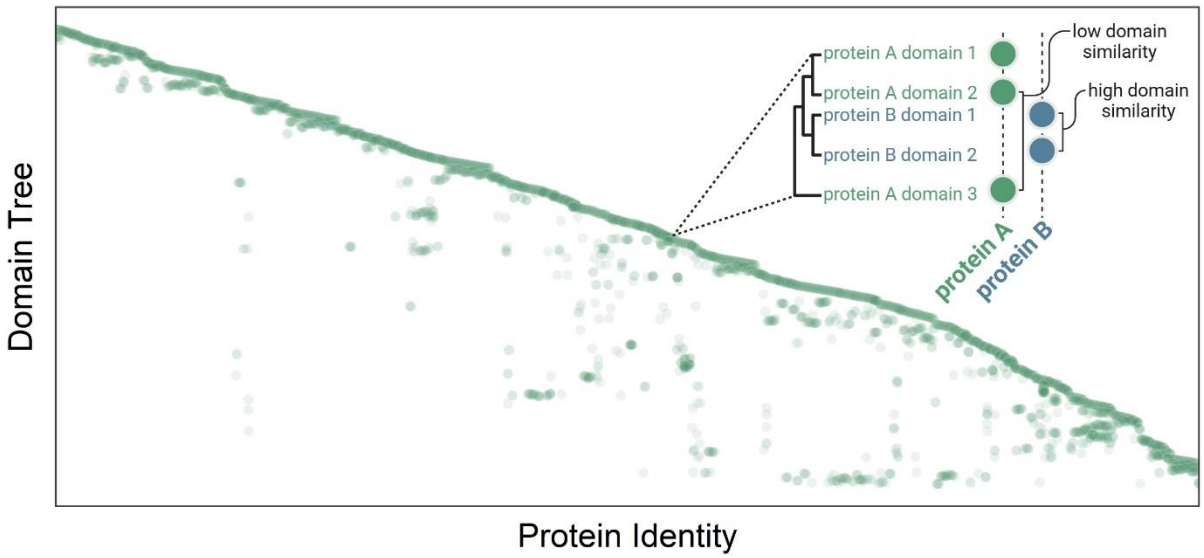


Figure 4.1 FMN-binding domains in multi-flavinylated proteins share high similarity. Matrix of FMN-binding domains from 2,081 multi-flavinylated proteins visualized by their grouping on a maximum likelihood tree derived from their sequence alignment (y-axis) and the identity of their corresponding proteins (x-axis). Each point represents a domain. Rows represent the appearing order of domains in the distance tree, where domains closer to each other on y-axis share higher sequence similarity and those more distant from each other are more dissimilar. Columns represent corresponding multi-flavinylated proteins, and domains from the same protein share identical x-coordinates.

between domains from the same proteins as well as across different proteins. We visualized the sequence alignments in a maximum likelihood tree format, where shorter distances between nodes indicate higher sequence homology and FMN-binding domains that share high sequence identity are more likely to be located in the same "monophyletic" group. We additionally distributed each domain by another categorical axis that is simply the identity of the protein encoding the corresponding domain. This visualization positioned most FMN-domains close to the diagonal line in the two-dimension coordinate: FMN-binding domains from the same protein are more likely to be located close to each other in the maximum likelihood tree, and increase in distances between domains generally happen when protein identity is changed (**Figure 4.1**). This result provided preliminary support for the hypothesis that additional FMN-binding domains in multi-flavinylated proteins are duplicated from other pre-existing domains within the same coding sequence. We also observed higher sequence dissimilarity between domains from the same proteins in less common cases, which suggested that there may be other mechanisms through which multi-flavinylated proteins acquired additional domains.

In earlier chapters, we described the biological versatility of ApbE-flavinylated proteins. Flavinylated proteins exhibit diverse structural and functional properties, which are likely outcomes of adaptation of flavinylation as a versatile redox strategy in a wide range of bacteria. This led us to consider that, if multi-flavinylated proteins acquired additional FMN-binding domains through sequence duplication, then sequence similarity of FMN-binding domains should also correspond to the sequence homology and/or taxonomic distance of the full protein. To this end, we repeated similar analysis but distributed each domain across another distance tree generated from multiple sequence alignment of overall sequences of the same proteins. This led to the identification of several distinct clusters of domains that are co-localized by the taxonomic

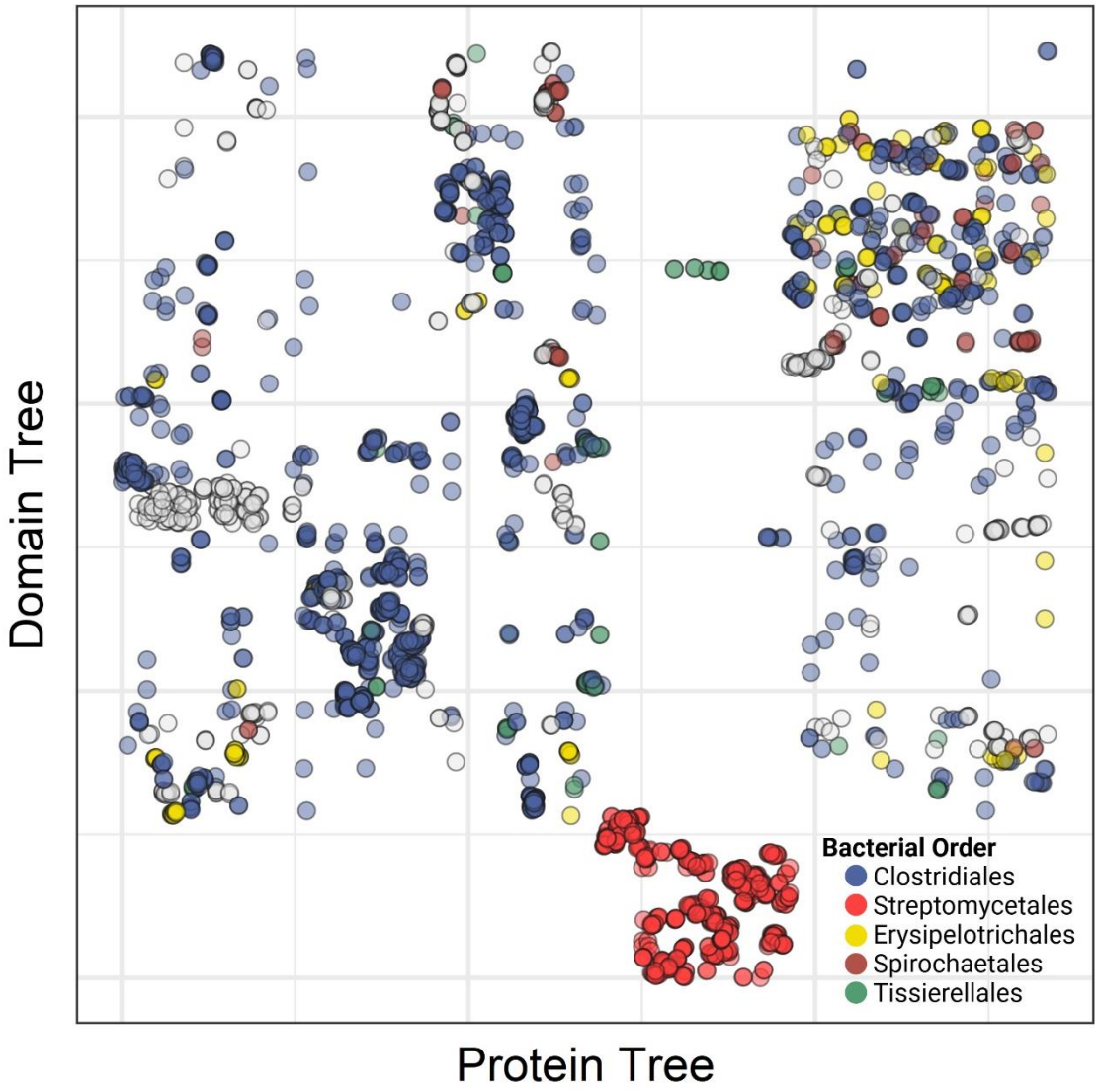


Figure 4.2 FMN-binding domains cluster by protein similarity and taxonomic assignment. Domain maximum likelihood tree (y-axis) visualized across a similar distance tree derived from multiple sequence alignments of their corresponding proteins. Proteins that share high similarity in overall protein sequences are located closer to each other on x-axis, whereas proteins that are more dissimilar with each other are located more distantly from each other. Domains belonging to the top 5 most common bacterial orders in dataset are colored. Blue: Clostridiales. Red: Streptomycetales. Yellow: Erysipelotrichales. Brown: Spirochaetales. Green: Tissierellales.

assignment of their corresponding protein (**Figure 4.2**). Of note, we observed a particularly large and distinct cluster of FMN-binding domains belonging to the Streptomycetales order. We scrutinized sequence homology of FMN-binding domains in this cluster and noted that, unlike domains from other bacterial orders, these Streptomycetales FMN-binding domains on average share higher sequence identity, implying that duplication of FMN-binding domains are more prevalent in these species.

We hypothesized that additional FMN-binding domains in multi-flavinylated proteins are duplicated from a pre-existing domain from an ancestral protein that is mono-flavinylated. To identify such ancestral proteins, we first determined ancestral domain sequences for all sets of cognate FMN-binding domains using ProtASR2 (Arenas and Bastolla, 2020), which is a tool used to reconstruct sequences of probable ancestral proteins and protein domains. To demonstrate our findings, we chose a double-flavinylated Streptomycetales protein, hereinafter referred to as SMF_4, as our example. SMF_4 contains 2 FMN-binding domains that share 82.4% sequence identity with each other. ProtASR2-predicted sequence of a possible ancestral FMN-binding domain shares 83.1% and 87.3% sequence identity with the first and the second domains of SMF_4, respectively. To determine if this ancestral FMN-binding domain is present as the sole domain in a pre-existing mono-flavinylated protein, we performed NCBI BLAST using this ancestral domain sequence as the query. This led to the identification of a mono-flavinylated protein encoded by *Streptomyces mirabilia*, abbreviated here as Sm_FMN-bind (NCBI accession: WP_212727691.1). Sm_FMN-bind shares sequence homology with SMF_4 in the ~148 residue N-terminal region containing the signal peptide, a ~39 residue unstructured loop, and the first FMN-binding domain. However, unlike SMF_4, Sm_FMN-bind lacks a C-terminal ~129 residue region that encodes another FMN-binding domain present in SMF_4 (**Figure 4.3A**). Structural alignments of both

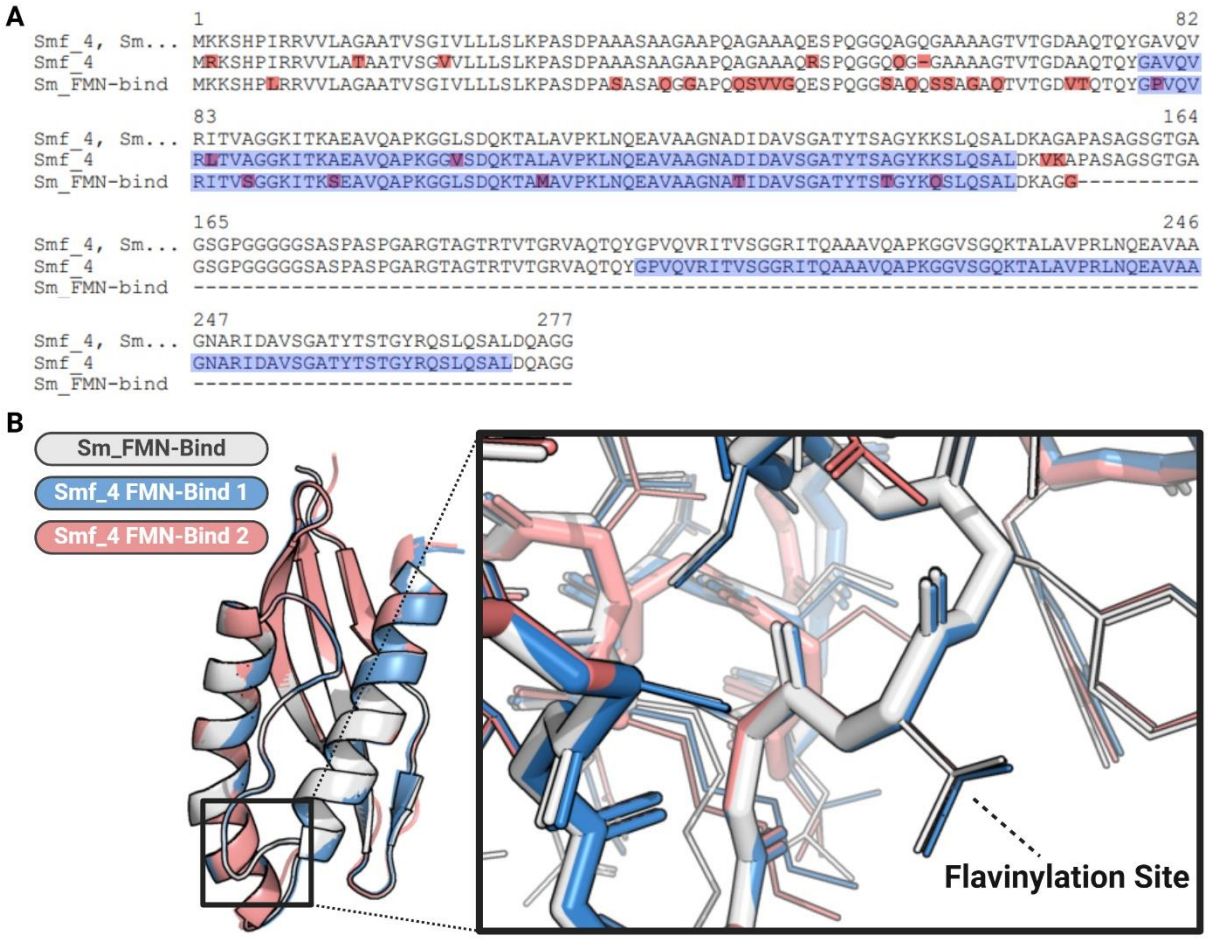


Figure 4.3 Sm_FMN-bind shares partial homology with SMF_4.
(A) Pairwise sequence alignment between SMF_4 and Sm_FMN-bind. Mismatched residues are colored in red. Segments colored in blue indicate conserved FMN_bind domains. Dashes indicate missing residues. (B) Structural alignment between the AlphaFold-derived structures of FMN-binding domains from Sm_FMN-bind, color in gray, and the two FMN-binding domains from SMF_4, colored in blue and red. Zoom-in view highlight the extend of overlap between threonine residues from the three domains, which serve as the flavinylation sites.

domains from SMF_4 and the domain from Sm_FMN-bind showed highly consistent structural arrangements across the three domains, with the predicted FMN-binding sites superimposed on each other (**Figure 4.3B**). This observation suggests that the double-flavinylated SMF_4 protein could have evolved from a Sm_FMN-like ancestral mono-flavinylated protein and acquired its second FMN-binding domain through domain duplication.

4.3.2 Stepwise electron transport in multi-flavinylated proteins

We next sought to determine whether multiple FMN-binding domains in multi-flavinylated proteins are responsible for stepwise electron transfer. In our previous analysis, we revealed that multi-flavinylated proteins could structurally mimic multi-heme cytochromes. In addition, many extracytosolic flavinylated proteins are also found in similar genomic context as cytochrome proteins. These observations imply that multi-flavinylated proteins could also functionally resemble multi-heme proteins like MtrA/C and facilitate stepwise electron transfer through sequential redox interactions between multiple FMN cofactors in the proteins (Edwards et al., 2020). However, in some cases, multi-flavinylated proteins could present highly intricate structures with no apparent spatial interactions between FMN-binding domains.

To experimentally determine the mode of electron transport in multi-flavinylated proteins, we used the double flavinylated PplA protein in *Listeria monocytogenes* as our model (Light et al., 2018). We leveraged the ferric iron-reducing activity of PplA—which can be readily measured through colorimetric methods—and performed a series of complementation assays. We found that iron-reducing activity of PplA is significantly reduced but not depleted in *L. monocytogenes* mutants that are deficient of the entire *pplA* locus, suggesting that *L. monocytogenes* is capable of reducing ferric iron through other mechanisms. We next complemented this mutant strain with

non-chromosomal variants of the *pplA* locus. We found that the iron-reducing activity is only rescued when the wild-type segment of the *pplA* locus is complemented. Complementation of a truncated *pplA* mutant with an early stop mutation did not promote recovery of iron-reducing activity, indicating that iron-reducing activity of PplA is dependent on the FMN-binding domains. In addition, complementation of *pplA* mutants that lacks either one or both of the FMN-binding sites similarly failed to rescue the iron-reducing activity (**Figure 4.4A**). This result indicates that both FMN-binding domains of PplA are required for electron transfer, which is consistent with the hypothesis that the two FMN-binding domains of PplA transfer electrons between each other, formulating a stepwise electron circuit (**Figure 4.4B**). This data provide experimental evidence that FMN-binding domains in multi-flavinylated proteins construct a linear, sequential electron path.

4.3.3 Health relevance of multi-flavinylated proteins

Our characterizations of PplA in the enteric pathogen *Listeria monocytogenes* further highlighted the known involvement of flavinylation-mediated electron transfer in physiology of bacterial species that are closely related to human health. In addition, a high proportion of multi-flavinylated proteins are found in bacterial species classified under the Firmicutes phylum. Many members in the Firmicutes phylum such as species from the *Clostridium*, *Ruminococcus*, *Faecalibacterium*, and *Enterococcus* genera are known to play critical roles in human health and disease (Hou et al., 2022; Sweeney and Morton, 2013; Wang et al., 2022). These insights suggest greater health relevance of multi-flavinylated proteins that were previously unexplored.

Consistently, we extended our finding on PplA and identified a PplA homolog in a clinically isolated strain of the opportunistic intestinal pathogen *Enterococcus faecalis*

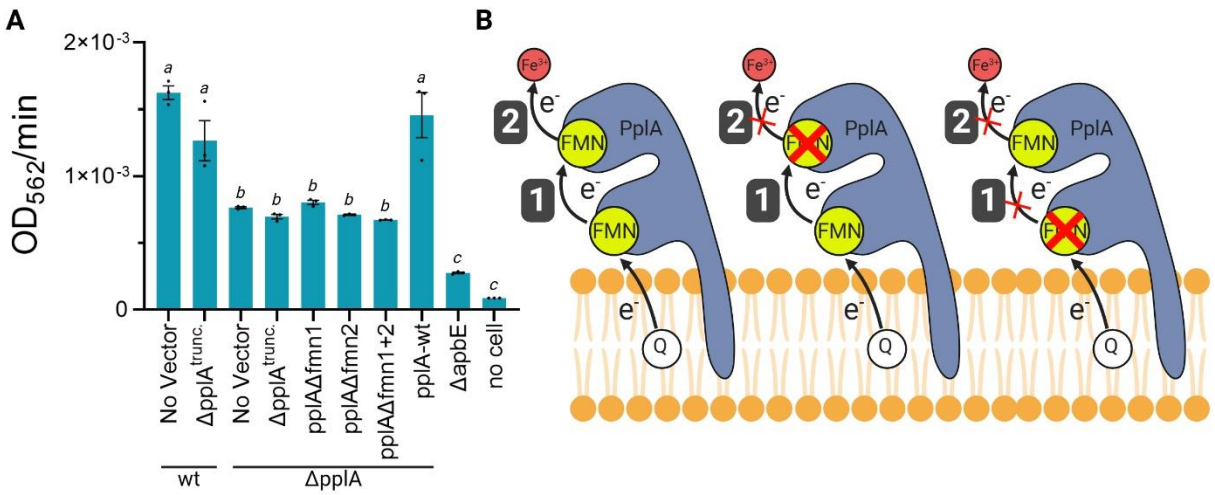


Figure 4.4 Stepwise electron transfer between FMN-binding domains of PplA.
(A) Complementation assays demonstrate stepwise electron transfer in PplA in *L. monocytogenes*. Rate of ferric iron is measured in absorbance at 562 nm of ferrous iron-bound ferrozine divided by reaction time. Δ*pplA*: mutant strain of *L. monocytogenes* with an early stop mutation introduced at the G72 residue of the chromosomal copy of *pplA*, preventing translation of both FMN-binding sites. Δ*pplA*$^{trunc.}$: identical to Δ*pplA* in sequence but complemented as an extrachromosomal copy. *pplA*Δfmn1/fmn2/fmn1+2: *pplA* mutant with point mutations in the first, second, or both FMN-binding domain(s) that converts the flavinylated threonine residue to alanine. Δ*apbE*: mutant *L. monocytogenes* strain lacking ApbE. When applicable, genotypes indicated above a line represent extrachromosomal gene variants introduced during complementation assays, in which case genotype labels below the line represent genotypes of the background strain. (B) schematic diagram for mode of stepwise electron transfer in the wild-type *pplA* (left), *pplA*Δfmn2 mutant (middle), and *pplA*Δfmn1 mutant (right).

(**Figure 4.5A**), which is known as a leading cause of nosocomial infection worldwide (Fisher and Phillips, 2009; Murray, 1990). Using liquid chromatography-mass spectrometry analysis (LC-MS) and similar colorimetric assays, we experimentally confirmed that the *Enterococcus* PplA-like protein is similarly double flavinylated and is capable of reducing ferric iron as a terminal electron acceptor (**Figure 4.5B&C**). AlphaFold prediction also reveals a similar structural arrangement of this protein, where an unstructured N-terminal transmembrane region anchors the two pseudo-symmetrical FMN-binding domains protein body onto the extracytosolic face of the membrane (**Figure 4.5A**). These observations serve as another example that demonstrates the use of flavinylation in redox physiology of pathogenic species.

4.3.4 Prevalence of ApbE-associated flavinylation markers in the human gut microbiome

As one of the most dense and dynamic microbial ecosystems, the human gut microbiome is known to contrite to health and diseases of the human host through intricate host-microbiome interactions (Hou et al., 2022). Micro-environments of the intestine underwent frequent fluctuations due to rapid changes in host physiology, diet, and drug use, and the capability to rapidly adapt to these fluctuations is pertinent to persistent survival for bacterial species in the gut microbiome (Nichols et al., 2019; Weersma et al., 2020). Both *Listeria* and *Enterococcus* are bacterial genera associated with intestinal infection (Fisher and Phillips, 2009; Low and Donachie, 1997). This observation suggests that flavinylation-mediated redox reactions are common among bacterial species that reside in the human gut microbiome.

Our previous bioinformatic survey screened ApbE-associated flavinylation markers in all available prokaryotic genomes without any selection criteria on specific microbial habitats. To comprehensively evaluate the prevalence of flavinylation in species residing in the human gut

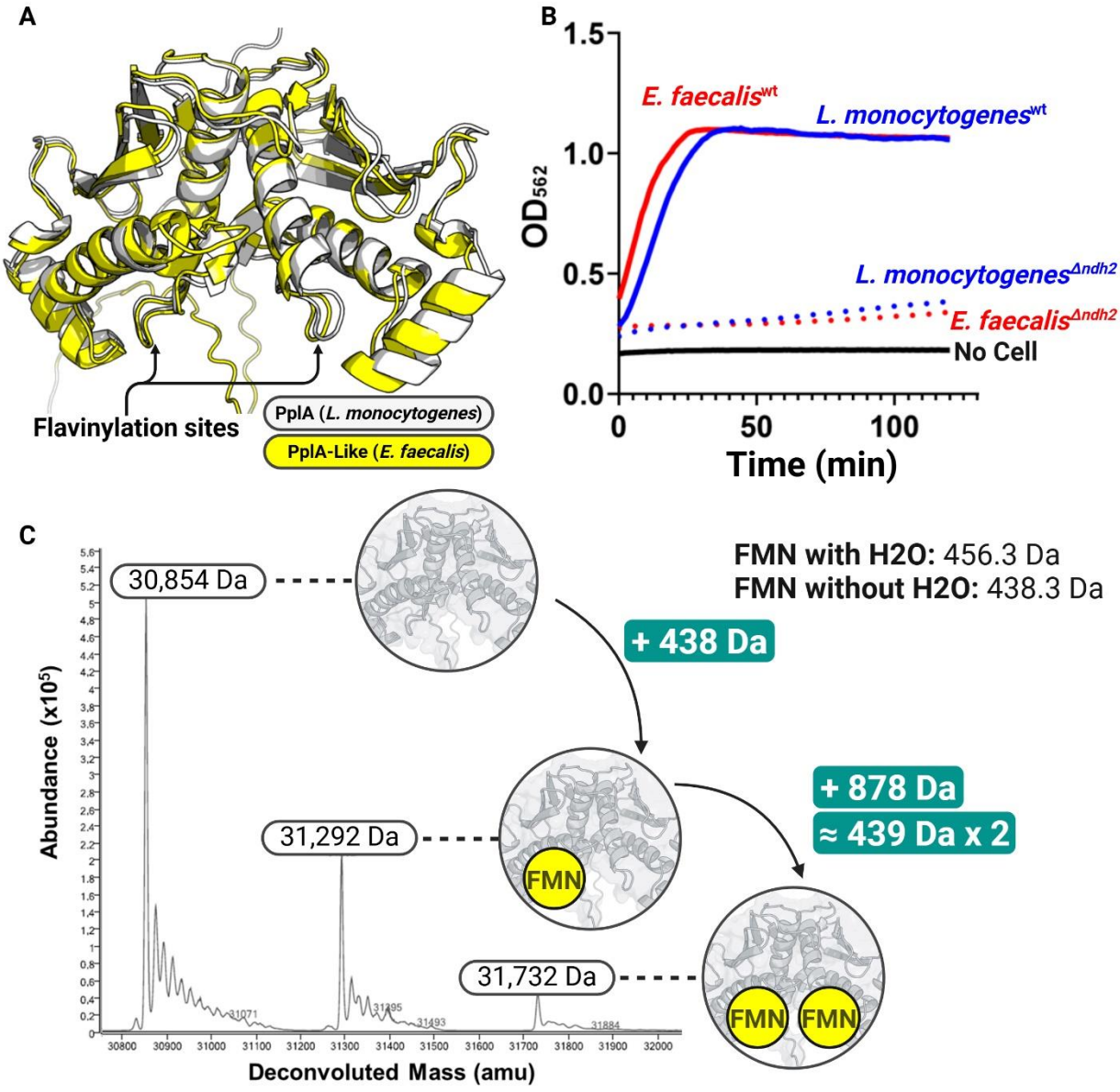
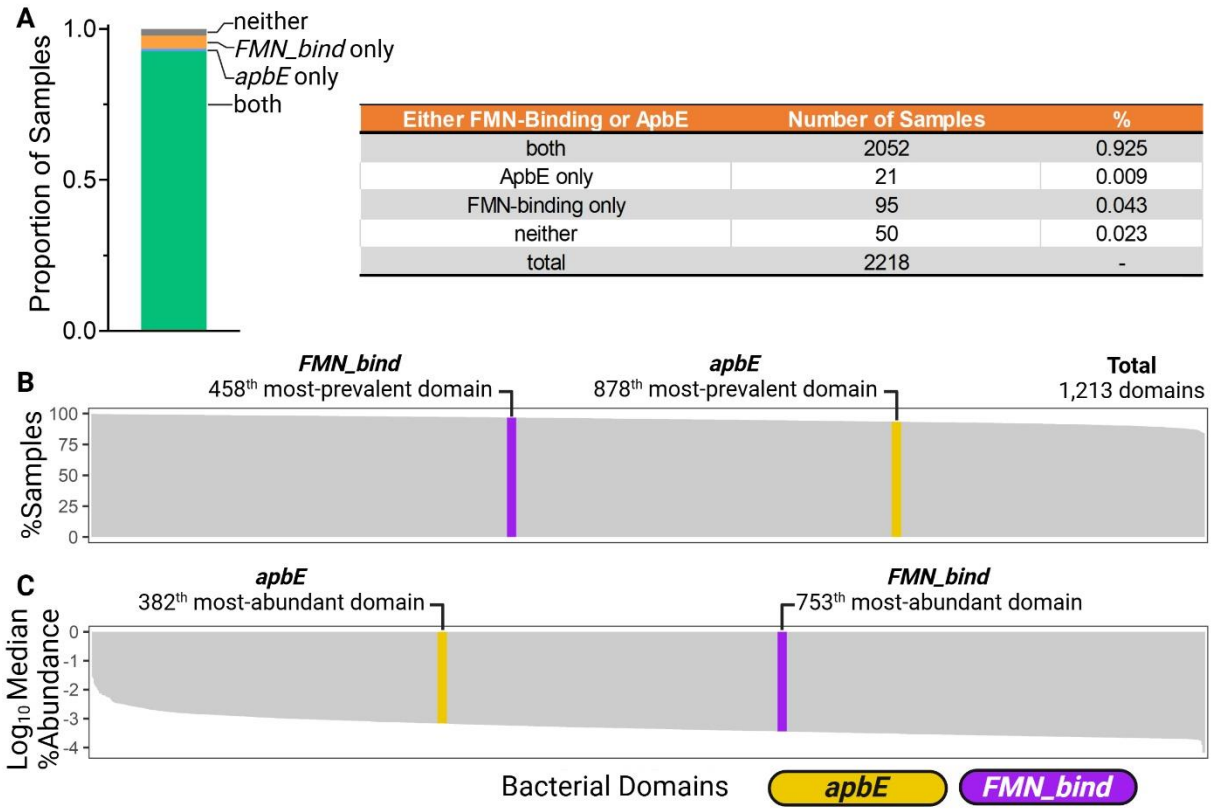


Figure 4.5 PplA homolog in *E. faecalis* reduces ferric iron and is double-flavinylated.
(A) Structural alignment of AlphaFold-predicted structures of PplA from *Listeria monocytogenes* (white) and the PplA-like protein from *Enterococcus faecalis* (yellow). Arrows indicate the two conserved FMN-binding sites in each of the proteins. (B) Ferric iron reduction assays for PplA and PplA-like. (C) Chromatogram of LC-MS analysis of purified PplA-like protein.



Either FMN-Binding or ApbE	Number of Samples	%
both	2052	0.925
ApbE only	21	0.009
FMN-binding only	95	0.043
neither	50	0.023
total	2218	-

Figure 4.6 Genomic markers of ApbE-flavinylation are prevalent in the gut microbiome. (A) Stacked bar plot and numbers showing proportions and numbers of samples that contain both, either one, or neither of the markers of flavinylation, namely the ApbE domain and the FMN-binding domain. (B) All bacterial domains are ranked by the percentage of samples in which the corresponding domain is detected. (C) All bacterial domains ranked by the $\log_{10}$ values of their median relative abundance across all samples. Yellow bars highlight the ranking of the ApbE domain, and purple bars highlight the ranking og the FMN-binding domain. All data are derived from human fecal metagenomic data (MGnify study MGYS00003733) (Mehta et al., 2018), with a total of 1,213 bacterial domains detected from 2,218 human stool samples.

microbiome, we leveraged publicly available metagenomic data that are derived from 2,218 fecal samples of a large human cohort (Mehta et al., 2018). We first determined the number of samples containing the two genomic markers of flavinylation, namely the ApbE domain and the FMN-binding domains. In agreement with our previous finding, we found that over 97% of the microbiome samples (2,168/2,218) contain genomes that encode ApbE and/or FMN-binding domain, with over 92% of the samples (2,052/2,218) containing genomes that encode both markers (**Figure 4.6A**).

We next examined the prevalence of ApbE and FMN-binding domains relative to other bacterial domains identified in the metagenomic dataset. To this end, we enumerated detection frequency (i.e., number of samples containing the corresponding protein domain) and abundance (median metagenomic relative abundance of the corresponding domain in which the domain is detected). We found that ApbE is more frequently detected in samples than ~28% of protein domains, ranking as the 878th most prevalently detected domain from a total of 1,213 domains. The FMN-binding domain is more frequently detected in samples than ~62% of protein domains, ranking as the 458th most prevalent domain (**Figure 4.6B**). In addition, we noted that the abundance of ApbE in ApbE-positive samples is higher than ~62% of protein domains, whereas the abundance of FMN-binding domain in FMN-bind-positive samples is higher than ~39% of protein domains (**Figure 4.6C**). Altogether, these data suggest that ApbE-mediated flavinylation is similarly versatile in bacterial species that are residing in the human gut microbiome. Future effort is required to determine whether multi-flavinylation in pathogen species like *E. faecalis* and *L. monocytogenes* could provide competitive advantage over gut commensals that similarly implement flavinylation-based redox strategies.

4.4 Discussion

In this chapter we described our findings from preliminary works that aim to uncover explore biological significance of multi-flavinylated proteins. Using multiple sequence alignments of FMN-binding domains from multi-flavinylated proteins, we observed evidence for prevalent domain duplication events. This result implies that previously mono-flavinylated proteins could evolve to encode additional FMN-binding sites through these domain duplication events and enable electron transfer across a longer path. This proposition is further supported by our findings on the Sm_FMN-bind protein, which shares homology with the ancestral sequence of the two FMNb-binding domains of the double-flavinylated SMF_4 protein. These data suggest that greater multiplicity of FMN may have resulted from duplication of a domain in pre-existing mono-flavinylated proteins. Future effort is required to comprehensively examine this hypothesis, potentially by using more stringent bioinformatic tools that allow classification of DNA repeats or evolution events of recurrent sequences.

While FMN-binding domains from the same proteins shared high sequence similarity in most cases, some cognate FMN-binding domains exhibited greater dissimilarities. This suggests that additional FMN-binding domains could be acquired through a different, less common mechanism. Interestingly, pioneering studies on evolution of multi-heme cytochromes have reported gene fusion and modular evolution as possible mechanisms for acquisition of additional heme-binding motifs (Soares et al., 2022). These events are characterized by fusion or exchange of heme-binding domains from smaller, mono-heme cytochromes. In our previous analysis, we observed a high prevalence of small, soluble flavinylated proteins whose entire coding sequence is comprised of an FMN-bind domain. These insights suggest multi-flavinylated proteins may also mimic multi-heme proteins in their path of evolution.

Our experimental work on characterization of stepwise electron transfer in the PplA protein shed light on the path of electron flow between multiple FMN moieties in the double-flavinylated protein. This finding suggests that consistent with our hypothesis, greater multiplicity of FMN cofactors enables further extension of the electron circuit beyond the cell envelope, allowing deposition of electrons onto a more distant electron acceptor. However, due to restraints in methods for efficient genetic characterizations, it is challenging to assess whether this stepwise mode of electron transport is similarly utilized by other multi-flavinylated proteins with an even greater number of FMN moieties. For multi-heme proteins, stepwise transport is also commonly reported as the electron transfer mechanism, though exceptions do exist. For example, the multi-heme protein OcwA found in *Thermincola spp.* is reported to encode its nine heme cofactors in different structural contexts. This unique property allowed OcwA to simultaneously interact with different electron acceptors, including nitrite, hydroxylamine, and iron oxide (Costa et al., 2019; Faustino et al., 2021). It is thus possible that some multi-flavinylated proteins may also manifest similar functional and structural properties that enable greater modularity in identity of electron acceptors.

Our results on PplA further reveal possibly greater involvement of multi-flavinylated proteins in health-relevant contexts. The iron-reducing PplA itself is encoded by the enteric pathogen *Listeria monocytogenes*. In addition, we identified a PplA homolog in *Enterococcus faecalis*, which is a known opportunistic pathogen responsible for many cases of nosocomial intestinal infection. Both species are facultative anaerobes, suggesting that the use of flavinylation-based reduction of ferric iron is a common alternative for respiration when oxygen is depleted in the intestinal environment. This hypothesis is also consistent with the observation that most multi-flavinylated proteins are encoded by Firmicutes species, many of which reside in the human gut

microbiome. Consistently, our preliminary examination of bacterial domains in previous metagenomic data derived from human stool samples also revealed that the FMN transferase ApbE and the FMN-binding domain are prevalently detected in human stool samples and at ample abundance within each sample. However, many commensal species of the gut microbiome harness energy through fermentation of host- or host diet-derived nutrients (Oliphant and Allen-Vercoe, 2019). Because fermentation is less effective than anaerobic respiration in energy generation (Kelly et al., 2014; Tejedor-Sanz et al., 2022), flavinylation-based anerobic respiration thus may serve as a more efficient strategy for redox metabolism, supporting rapid outgrowth of *L. monocytogenes* or *E. faecalis* during infection. This provides new directions for future research that aims to determine novel therapeutic mechanisms targeting flavinylation-based electron transfer in specific invading facultative pathogens.

4.5 Methods

4.5.1 Maximum likelihood tree of FMN-binding domains and multi-flavinylated proteins

Multi-flavinylated proteins were identified from a previously curated dataset (Chapter 2) with a selection criteria for more than 1 identifiable FMN-binding domain. Sequences of FMN-binding domains from multi-flavinylated proteins were extracted using residue coordinates identified by NCBI Conserved Domains. Multiple sequence alignment of domain sequences was performed using EMBL-EBI Clustal Omega Multiple Sequence Alignment (Madeira et al., 2019), and the alignment results were used to construct a maximum likelihood tree. The appearing order of each domain on the tree is then used as a surrogate measure of relative sequence homology, where domains located close to each other share greater sequence homology. Similar analysis was

performed for full protein sequences. Taxonomic assignments of multi-flavinylated proteins were determined through NCBI Entrez using their protein accession numbers.

4.5.2 Assessment of domain duplication in SMF_4 from Sm_FMN-bind

The ancestral sequence of FMN-binding domains from the double-flavinylated Streptomycetales SMF_4 protein was predicted using ProtASR2 (Arenas and Bastolla, 2020). The putative ancestral sequence is used to identify Sm_FMN-bind (NCBI accession: WP_212727691.1) using NCBI Protein BLAST. Pairwise sequence alignment between SMF_4 and Sm_FMN-bind was performed similarly using EMBL-EBI Clustal Omega Multiple Sequence Alignment (Madeira et al., 2019). Structures of SMF_4 and Sm_FMN-bind were predicted using AlphaFol2 (ColabFold v.1.5.2) (Evans et al., 2021; Jumper et al., 2021; Mirdita et al., 2022) and structures for all three FMN-binding domains were extracted and aligned in PyMol.

4.5.3 Stepwise electron transfer in PplA

4.5.3.1 Construction of constructs for complementation assay

Wild-type *pplA* sequence or mutant variants of *pplA* (Δ*pplA*^{trunc}/*pplA*Δfmn1/fmn2/fmn1+2) were completed into wild-type or Δ*pplA Listeria monocytogenes* 10403S using the pPL2x vector. *L. monocytogenes* Δ*pplA* mutant stain was derived from Xayarath et al., 2015 (Xayarath et al., 2015) and contains an early stop mutation introduced at its G72 residue. Complemented extrachromosomal copies of wild-type or mutant *pplA* sequences were synthesized using IDT gBlocks (Integrated DNA Technologies) and subsequently cloned into PCR-linearized pPL2x vectors at the multiple cloning sites through Gibson reactions (NEBuilder HiFi DNA Assembly Master Mix, New England Biolabs). Constructs with corresponding gene inserts where then transformed into chemically competent *E. coli* S17 cells through heat shock at 42 °C for 45 seconds.

Transformants were then selected by resistance to chloramphenicol (10 ug/mL) conferred by the pPL2x vector and confirmed by PCR and Sanger sequencing.

4.5.3.2 Delivery of constructs into *L. monocytogenes* 10403S

To transform pPL2x constructs into *L. monocytogenes* 10403S, corresponding *E. coli* S17 transformants were used as plasmid donors in conjugation following a protocol adopted from Lauer et al., 2002 (Lauer et al., 2002). Briefly, 3 mL of overnight cultures of *E. coli* S17 transformants (LB + 10 ug/mL chloramphenicol) were centrifugated (9,000 RPM for 2 minutes) and resuspend in 1 mL BHI for 3 times to wash out residual antibiotics. A final resuspension of cells in 1 mL BHI was used to resuspend cell pellets collected from 3 mL of *L. monocytogenes* overnight culture (BHI). This cell mixture is then centrifugated again and resuspend in 20 uL of BHI, then incubated on a nitrocellulose membrane placed on BHI agar for overnight at 30 °C. Successful *L. monocytogenes* conjugants were selected on BHI agar by resistance to chloramphenicol (10 ug/mL), which is derived from the pPL2x vector, and intrinsic resistance of *L. monocytogenes* to streptomycin (200 ug/mL). Selected colonies of conjugants were then confirmed by PCR and Sanger sequencing. Bacterial cultures were grown at 37 °C aerobically unless specified.

4.5.3.3 Measurement of ferric iron-reducing activity of *L. monocytogenes* strains

Ferric iron reduction of *L. monocytogenes* strains was measured by ferrozine assays (Riemer et al., 2004). Overnight cultures in BHI of *L. monocytogenes* strains were measured for their optical density at 600 nm and subsequently diluted in BHI to OD600 = 0.5. Cell pellets were prepared from 1 mL of diluted cultures and washed in PBS three times (9,000 RPM for 2 minutes for centrifugation). Cell pellets were then resuspended in 100 uL of 2X ferrozine buffer (4 mM ferrozine and 110 mM glucose dissolved in PBS) and seeded in 96-well plate in anaerobic chamber.

Samples were then mixed with 100 uL of 2X reaction buffer (50 mM ferric ammonium citrate dissolved in PBS), which provides electron acceptor. Measurements of ferric iron reduction activity were determined by light absorption at 562 nm, which indicates interactions between reduced iron (ferrous iron) and ferrozine. Samples were incubated at 37 °C anaerobically for 2 hours and measurements of OD562 were taken at 2-minute intervals. Rate of iron reduction was calculated by dividing baseline-substracted OD562 values by time in minutes.

4.5.4 Examination of PplA-like protein in *Enterococcus faecalis*

4.5.4.1 Identification of the PplA-like protein

The PplA-like protein in *Enterococcus faecalis* V583 was identified through NCBI Protein BLAST using sequence of PplA in *Listeria monocytogenes* as the query. Similarly, pairwise alignment between PplA and PplA-like was performed using EMBL-EBI Clustal Omega Multiple Sequence Alignment (Madeira et al., 2019) and structures of the two proteins were predicted using AlphaFold2 (ColabFold v.1.5.2) (Evans et al., 2021; Jumper et al., 2021; Mirdita et al., 2022). Structural alignment of the two proteins were performed in PyMol.

4.5.4.2 Confirmation of ferric iron-reducing activity of PplA-like

Ferric iron-reducing activity of PplA-like was measured through procedures identical to methods described in 4.5.3.3.

4.5.4.3 LC-MC analysis of PplA-like

Cloning, expression, and purification of the PplA-like protein in *E. faecalis* V583 was performed in steps identical to 2.5.6. To prepare PplA-like for LC-MS analysis, purified protein samples were dialyzed in a dialysis cassettes with 10 kDa MWCO (Thermo Fisher) in 1 L of 0.1% formic acid for 24 hours (the formic acid solution was replaced every 6 hours). Protein samples were then quantified on a DeNovix DS-11 FX+ Spectrophotometer based on molar mass of PplA-

like and extinction coefficient and subsequently standardized to 1 mg/mL. Prepared samples were transferred into a glass vial (Agilent) and submitted to the Integrated Molecular Structure Education and Research Center (IMSERC) at Northwestern University for LC-MS analysis.

4.5.5 Preliminary screen for flavinylation markers in human gut metagenomic data

Human gut metagenomic data was derived from a previous study containing 1,213 annotated bacterial domains detected from 2,218 human stool samples (MGnify study MGYS00003733) (Mehta et al., 2018). Prevalence of ApbE or FMN-binding domain was measured as percentage of human stool samples containing the corresponding domain. Absolute read counts in each sample were converted into relative abundance, and the $\log_{10}$ values of the median relative abundance of ApbE or FMN-binding domain across all samples were used to evaluate abundance of the corresponding domain. For visualization purposes, all domains were ranked by their prevalence or abundance.

4.6 References

Arenas M, Bastolla U. 2020. ProtASR2: Ancestral reconstruction of protein sequences accounting for folding stability. *Methods Ecol Evol* 11:248–257. doi:10.1111/2041-210X.13341

Costa NL, Hermann B, Fourmond V, Faustino MM, Teixeira M, Einsle O, Paquete CM, Louro RO. 2019. How Thermophilic Gram-Positive Organisms Perform Extracellular Electron Transfer: Characterization of the Cell Surface Terminal Reductase OcwA. *mBio* 10:10–1128.

Edwards MJ, White GF, Butt JN, Richardson DJ, Clarke TA. 2020. The Crystal Structure of a Biological Insulated Transmembrane Molecular Wire. *Cell* 181:665-673.e10. doi:10.1016/j.cell.2020.03.032

Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D. 2021. Protein complex prediction with AlphaFold-Multimer. doi:10.1101/2021.10.04.463034

Faustino MM, Fonseca BM, Costa NL, Lousa D, Louro RO, Paquete CM. 2021. Crossing the Wall: Characterization of the Multiheme Cytochromes Involved in the Extracellular Electron

Transfer Pathway of Thermincola ferriacetica. *Microorganisms* 9:293. doi:10.3390/microorganisms9020293

Fisher K, Phillips C. 2009. The ecology, epidemiology and virulence of Enterococcus. *Microbiology* 155:1749–1757. doi:10.1099/mic.0.026385-0

Hou K, Wu Z-X, Chen X-Y, Wang J-Q, Zhang D, Xiao C, Zhu D, Koya JB, Wei L, Li J, Chen Z-S. 2022. Microbiota in health and diseases. *Sig Transduct Target Ther* 7:135. doi:10.1038/s41392-022-00974-4

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. doi:10.1038/s41586-021-03819-2

Kelly DJ, Hughes NJ, Poole RK. 2014. Microaerobic Physiology: Aerobic Respiration, Anaerobic Respiration, and Carbon Dioxide Metabolism In: Mobley HLT, Mendz GL, Hazell SL, editors. Helicobacter Pylori. Washington, DC, USA: ASM Press. pp. 111–124. doi:10.1128/9781555818005.ch10

Lauer P, Chow MYN, Loessner MJ, Portnoy DA, Calendar R. 2002. Construction, Characterization, and Use of Two *Listeria monocytogenes* Site-Specific Phage Integration Vectors. *J Bacteriol* 184:4177–4186. doi:10.1128/JB.184.15.4177-4186.2002

Light SH, Su L, Rivera-Lugo R, Cornejo JA, Louie A, Iavarone AT, Ajo-Franklin CM, Portnoy DA. 2018. A flavin-based extracellular electron transfer mechanism in diverse Gram-positive bacteria. *Nature* 562:140–144. doi:10.1038/s41586-018-0498-z

Low JC, Donachie W. 1997. A review of Listeria monocytogenes and listeriosis. *The Veterinary Journal* 153:9–29. doi:10.1016/S1090-0233(97)80005-6

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 47:W636–W641. doi:10.1093/nar/gkz268

Mehta RS, Abu-Ali GS, Drew DA, Lloyd-Price J, Subramanian A, Lochhead P, Joshi AD, Ivey KL, Khalili H, Brown GT, DuLong C, Song M, Nguyen LH, Mallick H, Rimm EB, Izard J, Huttenhower C, Chan AT. 2018. Stability of the human faecal microbiome in a cohort of adult men. *Nat Microbiol* 3:347–355. doi:10.1038/s41564-017-0096-0

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nat Methods* 19:679–682. doi:10.1038/s41592-022-01488-1

Murray BE. 1990. The life and times of the Enterococcus. *Clin Microbiol Rev* 3:46–65. doi:10.1128/CMR.3.1.46

Nichols RG, Peters JM, Patterson AD. 2019. Interplay Between the Host, the Human Microbiome, and Drug Metabolism. *Hum Genomics* 13:27. doi:10.1186/s40246-019-0211-9

Oliphant K, Allen-Vercoe E. 2019. Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome* 7:91. doi:10.1186/s40168-019-0704-8

Riemer J, Hoepken HH, Czerwinska H, Robinson SR, Dringen R. 2004. Colorimetric ferrozine-based assay for the quantitation of iron in cultured cells. *Analytical Biochemistry* 331:370–375. doi:10.1016/j.ab.2004.03.049

Soares R, Costa NL, Paquete CM, Andreini C, Louro RO. 2022. A New Paradigm of Multiheme Cytochrome Evolution by Grafting and Pruning Protein Modules. *Molecular Biology and Evolution* 39:msac139. doi:10.1093/molbev/msac139

Sweeney TE, Morton JM. 2013. The Human Gut Microbiome: A Review of the Effect of Obesity and Surgically Induced Weight Loss. *JAMA Surg* 148:563. doi:10.1001/jamasurg.2013.5

Tejedor-Sanz S, Stevens ET, Li S, Finnegan P, Nelson J, Knoesen A, Light SH, Ajo-Franklin CM, Marco ML. 2022. Extracellular electron transfer increases fermentation in lactic acid bacteria via a hybrid metabolism. *eLife* 11:e70684. doi:10.7554/eLife.70684

Wang S, Song F, Gu H, Shu Z, Wei X, Zhang K, Zhou Y, Jiang L, Wang Z, Li J, Luo H, Liang W. 2022. Assess the diversity of gut microbiota among healthy adults for forensic application. *Microb Cell Fact* 21:46. doi:10.1186/s12934-022-01769-6

Weersma RK, Zhernakova A, Fu J. 2020. Interaction between drugs and the gut microbiome. *Gut* 69:1510–1519. doi:10.1136/gutjnl-2019-320204

Xayarath B, Alonzo F, Freitag NE. 2015. Identification of a Peptide-Pheromone that Enhances Listeria monocytogenes Escape from Host Cell Vacuoles. *PLoS Pathog* 11:e1004707. doi:10.1371/journal.ppat.1004707

Chapter 5: Conclusion and Summary

The objectives of this thesis are to provide a better understanding of phylogenetic and molecular versatility of ApbE-flavinylated proteins. As another redox strategy, flavinylation resembles other known redox-active protein modifications, such as heme in cytochrome proteins, in many aspects, from critical metabolic activities to industrial applications in bioelectricity generation (Kumar, 2010). Thus, fulfilling these objectives would offer novel insights into flavinylation-based bacterial redox physiology, while providing brand new solutions for biotechnology. This chapter summarizes our progress, limitations, and future directions on achieving each of these goals.

5.1 Mechanistic validation of flavinylation-mediated electron transfer

Our bioinformatic screen of ApbE-based flavinylation makers reveals that flavinylation is highly prevalent in prokaryotic life, widely encoded by a diverse range of bacterial and archaeal species. To move beyond bioinformatic analysis, we attempted to further ascertain whether selected flavinylated proteins are indeed ApbE substrates. This led us to develop an in vitro flavinylation assay that could rapidly confirm covalent attachment of FMN in heterologously expressed candidate proteins. Using this approach we successfully confirmed over 30 putative flavinylated proteins as ApbE substrates. In addition, for flavinylated candidates such as PplA that utilize extracytosolic ferric iron as terminal electron acceptors, we were able to further confirm their electron transfer activity using previously developed colorimetric assays.

These approaches facilitate high throughput characterization of putative flavinylated proteins. However, several bottlenecks exist for these approaches. First, while flavinylation could be efficiently confirmed in vitro, similar characterization in vivo is prohibited by availability of

tools for genetic manipulation in specific bacterial species. Unlike *Listeria* or *Enterococcus* species whose genetic tools are readily available, many bacterial species encoding flavinylation machineries identified in our analysis are understudied. Thus, physiological characterizations of flavinylation in these species are challenging. In addition, the identity of electron acceptors for many identified flavinylated systems is unknown. Functional examinations of these systems would also require large scale screening of redox activities on a wide array of possible electron acceptors, which would serve as another rate-limiting step. Future advancement in genetic and functional tools would thus be complementary to our in vitro flavinylation assay.

5.2 Molecular basis of electron transfer in flavinylated proteins

We made several attempts in uncovering the mode of electron transfer in flavinylated proteins in hope to provide a mechanistic understanding for molecular basis of flavinylation. These attempts led to our findings on possible evolution events in Flavodoxin_4 and FMN_red proteins, where proteins with non-covalent FMN moieties may have evolved to encode additional structures containing a covalent FMN-binding site, which in turn secures the FMN cofactor within the conserved overall protein structures. We next examined the diverse cellular contexts of flavinylated proteins through predicted structural models. This reveals spatial interactions between flavinylated proteins and other electron-transferring subunits in larger membrane protein complexes, including transmembrane cytochromes that commonly co-localize with membrane-anchored extracytosolic FMN-binding domains.

Because most of these findings leveraged structural models predicted by AlphaFold, they also share the same limitations applicable to structural prediction models. Our observations on spatial interactions of FMN moieties depend on accurate predictions of 3D positions of FMN

molecules, which are not directly provided by AlphaFold. We instead relied on the position of the FMN-binding site to infer possible positions of FMN moieties, leading to uncertainties on how the "T-shaped" FMN molecules orient in the 3D space. We previously attempted an alternative method for more reliable predictions of FMN-binding in candidate proteins. This entails the use of tools like AlphaFill (Hekkelman et al., 2023), which superimpose predicted structures with pre-existing, co-factor-bound structures to transplant co-factors of interest. However, this approach suffers from its requirement for pre-existing protein structures, which is absent for many flavinylated candidates identified in our bioinformatic analysis. A more recent update to AlphaFold, namely AlphaFold3, may serve as a promising solution (Abramson et al., 2024). AlphaFold3 includes a new feature for prediction of cofactor binding. Future efforts are required to determine whether tools like AlphaFold3 could provide accurate predictions for 3D orientations of FMN moieties.

5.3 Flavinylated proteins in health-relevant contexts

A substantial body of recent research has highlighted contributions of the human gut microbiome to its human host in both healthy and diseased context (Hou et al., 2022; Sweeney and Morton, 2013; Wang et al., 2022). In light of our findings on PplA and PplA homologs in the two enteric pathogens *L. monocytogenes* and *E. faecalis*, we determined the prevalence of flavinylation in human gut metagenome data and observed that flavinylation-associated domains are frequently encoded by gut commensals. This suggests that flavinylation-mediated electron transfer could contribute to the intricate interactions between the human host and the gut microbiome. For example, anaerobic respiration in the gut microbiome could be affected by host diet (Oliphant and Allen-Vercoe, 2019). Many terminal electron acceptors implemented by extracytosolic electron

transfer systems have known dietary sources, such as nitrate from green-leaf vegetables (Guadagnin et al., 2005), sulfate from eggs (Réhault-Godbert et al., 2019), or ferric iron from red meats (Czerwonka and Tokarz, 2017). Alterations in human diet thus may serve as a modifiable mechanism for manipulating microbiome compositions by promoting species that are capable of respiration on available nutrients. Future efforts are required to examine whether flavinylation-based respiration is an important modulator of composition and function of the gut microbiome.

5.4 References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung C-C, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Žídek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493–500. doi:10.1038/s41586-024-07487-w

Czerwonka M, Tokarz A. 2017. Iron in red meat–friend or foe. *Meat Science* 123:157–165. doi:10.1016/j.meatsci.2016.09.012

Guadagnin SG, Rath S, Reyes FGR. 2005. Evaluation of the nitrate content in leaf vegetables produced through different agricultural systems. *Food Additives and Contaminants* 22:1203–1208. doi:10.1080/02652030500239649

Hekkelman ML, De Vries I, Joosten RP, Perrakis A. 2023. AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nat Methods* 20:205–213. doi:10.1038/s41592-022-01685-y

Hou K, Wu Z-X, Chen X-Y, Wang J-Q, Zhang D, Xiao C, Zhu D, Koya JB, Wei L, Li J, Chen Z-S. 2022. Microbiota in health and diseases. *Sig Transduct Target Ther* 7:135. doi:10.1038/s41392-022-00974-4

Kumar S. 2010. Engineering cytochrome P450 biocatalysts for biotechnology, medicine and bioremediation. *Expert Opinion on Drug Metabolism & Toxicology* 6:115–131. doi:10.1517/17425250903431040

Oliphant K, Allen-Vercoe E. 2019. Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome* 7:91. doi:10.1186/s40168-019-0704-8

Réhault-Godbert S, Guyot N, Nys Y. 2019. The Golden Egg: Nutritional Value, Bioactivities, and Emerging Benefits for Human Health. *Nutrients* 11:684. doi:10.3390/nu11030684

Sweeney TE, Morton JM. 2013. The Human Gut Microbiome: A Review of the Effect of Obesity and Surgically Induced Weight Loss. *JAMA Surg* 148:563. doi:10.1001/jamasurg.2013.5

Wang S, Song F, Gu H, Shu Z, Wei X, Zhang K, Zhou Y, Jiang L, Wang Z, Li J, Luo H, Liang W. 2022. Assess the diversity of gut microbiota among healthy adults for forensic application. *Microb Cell Fact* 21:46. doi:10.1186/s12934-022-01769-6