DISSERTATION

MAPPING THE METABOLIC PROTEIN INTERACTOME THAT SUPPORTS ENERGY CONSERVATION AT THE LIMITS OF LIFE

Submitted by

Seré Anne Williams

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2024

Doctoral Committee:

Advisor: Thomas Santangelo

Jeffrey C. Hansen
Marinus Pilon
G. Brooke Anderson
Christopher Snow

ABSTRACT

MAPPING THE METABOLIC PROTEIN INTERACTOME THAT SUPPORTS ENERGY CONSERVATION AT THE LIMITS OF LIFE

Distinct metabolic strategies yield energetic gains from a wide variety of substrates, yet only three overarching methods of energy conservation have been defined: substrate level phosphorylation, the generation of a charged membrane, and electron bifurcation. The dominant theme of known energy conservation mechanisms suggests that energy is conserved through the selective movement and management of electrons, thus essentially all life relies on redox (reduction and oxidation) reactions. Small molecule redox cofactors (such as $NAD(P)^+$) and proteinaceous electron carriers (such as ferredoxins) are employed as electron carriers throughout the biosphere. Proteinaceous electron carriers offer the potential for selective protein-protein interactions to bridge reductive flow from catabolic reactions to the membrane, providing a “proteinaceous electron highway” for efficient electron shuttling. Specific redox protein partnerships have been shown to adapt to changing physiological conditions, suggesting that proteinaceous electron flux is tunable and provides a level of selectivity not possible with small molecule electron transport. While electron flux through a tunable and regulated system of protein interactions can offer exceptional energy conservation strategies, large gaps remain in our knowledge of how electron flux is regulated *in vivo.* Identification of *bona fide in vivo* protein assemblies – and how such assemblies dictate the totality of electron flow and thus cellular metabolism – is an important milestone to understand the regulation imposed on metabolism, energy-production, and energy conservation. Resolving the dynamic nature of nanoscale interactions in living systems is arguably the current frontier of molecular biology, and

combinatorial methods – which layer multiple *in vitro* and *in vivo* techniques with large data analysis – have come to the forefront.

This dissertation addresses energy conservation strategies of *in vivo* protein associations in a model, genetically accessible, hyperthermophilic archaeon (*Thermococcus kodakarensis*) by mapping the metabolic protein interactome using affinity purification mass spectrometry (AP-MS) and generating engineered strains where fusion proteins selectively redirect electron flux *in vivo*. Twenty-five proteins involved in distinct metabolic functions were tagged to reveal that each tagged-protein interacts with ~ thirty proteins on average. These interactions connected disparate functions suggesting catabolic and anabolic activities may occur in concert -- in temporal and spatial proximity *in vivo*. The AP-MS method also refined our understanding of previously determined stable complexes suggesting that protein complexes *in vivo* likely adapt to redox conditions. Engineered strains linking a proteinaceous electron donor to a proposed electron acceptor were viable and impacted electron flux *in vivo*. Fusion strains linking a ferredoxin to the hydrogen-generating respiratory system increased hydrogen gas output ~8% on average with one strain showing a ~45% increase over wild type. Fusion strains impacting lipid saturation were shown to inhibit saturation, and future studies aim to determine if electrons can be redirected from the vast reductant sink of lipids to the generation of hydrogen gas, a valuable biofuel.

ACKNOWLEDGMENTS

As a curious person I've found that my never-ending love of learning can in fact be satiated by designing a good, logical experiment and working through it until a new understanding has been extracted from either the outcome or reflecting on the process itself. I am grateful for both my curiosity and the people who have supported me in my journey of understanding "how science works." From my childlike perspective, I used to believe that the world had a sense or order and that most decisions were made based on reason and evidence. I now realize that logical decision making is the exception, not the rule. We are emotional creatures and rely on instinct much more than deduction, but we can learn to exercise our rational muscle and in a community of scientists this has led to the advancements we now consider commonplace – refrigeration, combustion engines, and contraception, to name a few life-changing inventions – which support my opinion that now is the best time to be alive. Science is hard work, and I would like to thank all people who have endeavored to pursue a scientific process to gain an objective understanding. Thank you for your efforts.

To my scientific mentors, I need to begin with Judy Justice, my high school AP Biology teacher who likely taught me electron transport-mediated phosphorylation when I was ~16 years old. Judy was a stellar science teacher who provided an exceptional education in molecular biology to high school students. To Dr. Steve McBryant who was my first mentor in Dr. Jeff Hansen's lab and taught me to always cross my sevens. To Dr. Jeff Hansen who was there at the beginning and "end" (read "PhD"). To Dr. Mykles, our "lab dad" and the most excellent cell biology instructor a young scientist could hope for. To Dr. Joe Covi for being a true friend in the sciences and in life. To Dr. John Walrond for teaching me how water works. To Dr. Maxwell Ware for helping me feel "normal" for heading back into the lab at 9 pm, yet again. To Dr. Prasad Kasavajhala for his dedication to his science and for always greeting me with the warmth of a big smile. To Dr. Rien Pilon for loving science in the way I aspire to love science.

To Dr. Jolanta Menert who was my first “boss” and who pushed me to improve processes that she hadn’t identified – a valuable attitude in a mentor that I’ve attempted to emulate.

To the Santangelo lab, we’re a rowdy bunch, and I’ve never felt so at home in a scientific group. I look forward to lab meetings every week and love the depth of discussion. We are forced to engage with our science and thank you for doing this with me. To Kristin, thank you for encouraging me to talk to Tom about joining the lab. To Brett, thank you for always giving me a hard time in a fun way. To Aldy, thank you for your lab cartoons and sharing your breadth of knowledge humbly. To Bree, thank you for being there “with me.” I think we led the way with each other, and I’m so grateful you were there too. To current and future lab members – please use this dissertation as an encyclopedic reference for protein purification and the AP-MS method – I wrote it for you (*use the Table of Contents as a guide*). To Alina, Alex, Gabe, Paul, Liam, Jackson, Marina, Jason, Chris, Rayia, David, Savina, and all future lab members – Tom is here to push you through your science. Remember that one of the things you can learn is how to calmly and reasonably push back. Use this opportunity. It’s important.

To my advisor, Tom, you have inspired me in many ways, and I have grown as a scientist and a professional due to your care and attention – thank you. We often focus on the science, but I want to thank you for being a teacher, too. You share your understanding of the nano-scale world in a way that leads to awe and wonder. You are a true mentor, and that is rare to find.

To the undergraduate researchers who contributed to this work and my development as a mentor – thank you to Danielle Riley, Lucy Latta, Teagan Rockwood, David Crosby, Isabella Bendorf, Kate Call, Jared LeCuyer, and Savina Miller. Dani, thank you for your stalwart attitude, your science jokes, and your hard work. Lucy, thank you for teaching me so much about mentoring. Teagan, thank you for many things, but most of all for our fun science conversations – you’ve shown me how to mix a foundation of scientific knowledge with creativity and this dynamic is one that I am grateful to have experienced and will look for going forward. David,

thank you for your incredible attitude. You've kept your chin up and kept going even when it was tough – it has been a pleasure working with you. Izzy, it was a short stint, but thanks for your help. Kate, thank you for your repetition in making media and harvesting cultures. It's valuable to be able to do the same thing over and over with consistency. Jared, thank you for truly being a quick learner and a solid scientist – your protein purification procedure will benefit many. Savina, thank you for showing up on time every time and with a smile.

To Dr. Carol Wilusz, thank you for growing the CMB program into a collaborative community. Thank you for acquiring multiple funding opportunities and encouraging me to be a strong female scientist, attend leadership workshops, teach at Colorado College, and write for KUNC for the AAAS Mass Media Fellowship. To all of CMB, thank you and keep rocking it.

More personally, I would like to thank my family and friends for being here and packing the room (in person, literally, and on Zoom) during my defense. To my housemates, Jeremy, Zach, Jack, and Harmeen – thank you for all the delicious food and living with me through the end. To Audrey, thank you for being not only my friend by my scientific comrade from the crab lab to AB InBev to now. It's been – and I'm sure will continue to be – fun. To Ruth, thank you for being the strong woman that you are and for your support on the trail, in the woods, and in life. To John (and Leonard), thank you for being here this last year to teach me so much about partnership and for your love of data – we can all use more excitement for data. To Nancy Evans, thank you for your never-ending patience and for sharing your wisdom with me. We have done a PhD together, too, and I am so grateful for it.

To my family – parents, aunts, uncles, and sisters – thank you for being here and for your support. Jo, when they make this into a movie starring Julia Roberts as Seré Williams and George Clooney as "the cell," let's watch it.

Dad, thank you for encouraging me to question the world. The sentence "99.9% of the time my opinions are right" (told by a father to his daughter, one car ride in the 1990's) led me to explore how we know what we know.

Mike, thank you for noticing how I love to wonder and showing me how to live a curious life – through blowing bubbles at the sink, pushing back when I thought you could regain energy lost from climbing a mountain on the way down, and the myriad of other wonders that exist simply by living.

To my mom, who has offered her never-ending support for me in whatever direction and path I choose. I am honored and humbled by your love and care. I will always be here for you, too – to the moon and back and then some.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 4

CHAPTER 1
INTRODUCTION

1.1 Energy in biologic systems

A foundational understanding of metabolism includes glycolysis and oxidative phosphorylation which yield predictable products, such as ATP, via the tenets of fermentation and cellular respiration. While such metabolic processes are routine within the biosphere, they only touch the surface of the vast array of processes which provide energy to biological systems. In this chapter, a handful of unique and diverse metabolisms will be described to lay the groundwork for energy strategies which support life. Next, trends within these strategies will be discussed followed by a discussion of electrons and electron flux. Understanding and then rationally directing electron flux in a model hyperthermophile is the topic of this dissertation, thus the introduction will conclude with an overview of archaeal metabolism and how this work has contributed to the field.

1.1.1 Energetic gains can be made through a variety of pathways

Less than ten biological reactions generate enough energy to produce a single molecule of ATP through substrate level phosphorylation.[1] Instead, cells have evolved to generate most of their ATP through exergonic electron transfer that is used to develop an ion gradient across a membrane driving ATP synthesis via a membrane-bound ATP synthase (Figure 1.1).[2] A third, lesser known, energy conserving mechanism involves electron bifurcation wherein an endergonic electron transfer is coupled to an exergonic transfer, which provides sufficient energy for both reactions to proceed.[3–5] These three pathways will be discussed here and followed by a discussion of the variety of substrates utilized to generate a charged membrane.

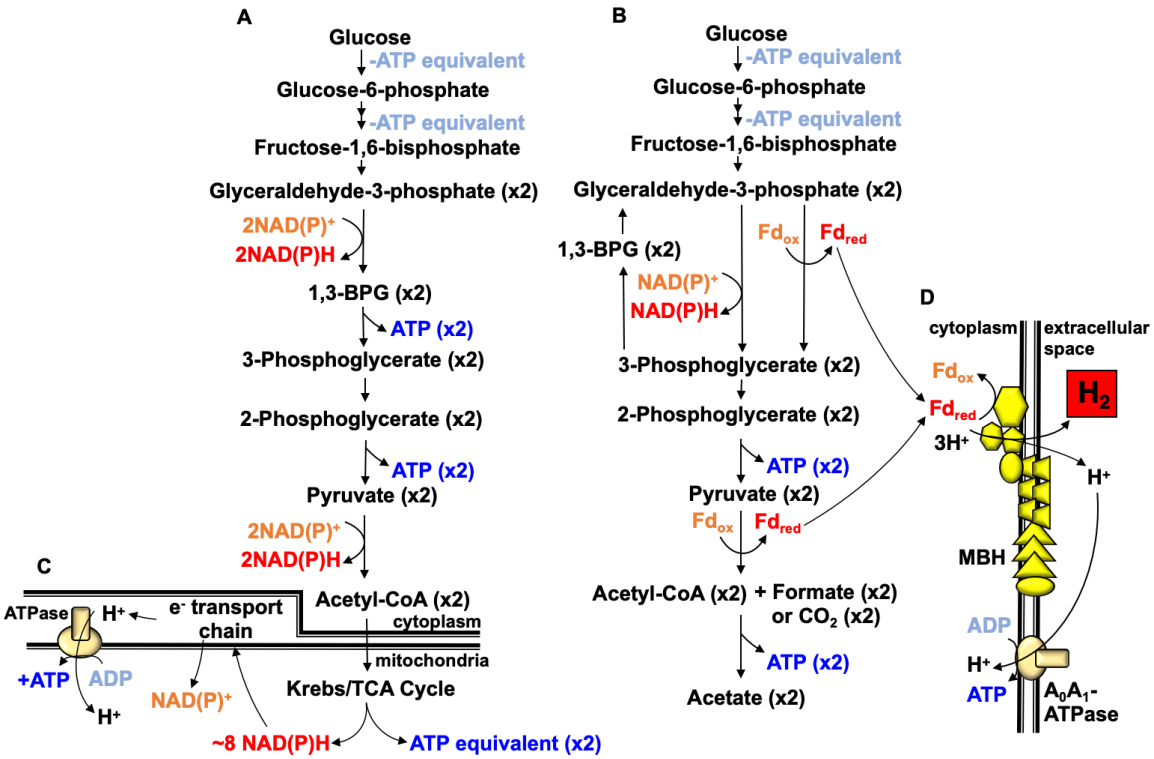


Figure 1.1: Eukaryotic and archaeal metabolisms compared. High-energy molecules such as glucose can be catabolized to generate a net gain of ATP in eukaryotic (A) and archaeal (B) glycolytic pathways. Energetic gains from substrate level phosphorylation are conserved in the form of ATP in the conversion of 2-phosphoglycerate to pyruvate in both eukaryotic and archaeal metabolisms. While eukaryotes maintain the ATP yielding reaction which converts 1,3-bisphosphoglycerate (1,3-BPG) to 3-phosphoglycerate, the conversion of glyceraldehyde-3-phosphate to 3-phosphoglycerate generates reduced co-factors (NAD(P)H or a reduced ferredoxin, Fd_{red}) in archaea. Reduced cofactors are used to fuel the generation of a proton gradient at a membrane. In eukaryotes (C), the inner mitochondrial membrane houses an efficient electron transport system fed by reduced products of the tricarboxylic acid (TCA) cycle. Archaea are prokaryotes and do not maintain organelles, thus they employ an electron transport chain at their membrane (D). Reduced cofactors are used to generate a proton gradient via a membrane bound hydrogenase which ultimately drives ATP synthesis via ATP synthase. To note, a variety of metabolisms exist; the modified Embden-Meyerhof-Parnas (EMP) glycolytic pathway is provided here as an example glycolytic pathway in archaea.[6]

Substrate level phosphorylation

The high energy bond present in the phosphorylation of ADP to ATP requires ~ $\Delta G^{\circ\prime}$ = -31 kJ mol^{-1} under standard conditions.[7,8] A handful of reactions (e.g. 1, 3-bisphosphoglycrate conversion to 3-phosphoglycerate by phosphoglycerate kinase, Figure 1.1) maintain the energy

required for the triphosphate bond to be formed in a single step, but these reactions are not plentiful enough nor use a wide enough variety of substrates to support life.[1,7]

Energy conservation via a charged membrane

A wide variety of metabolic strategies use a chemiosmotic potential across a membrane to store small bits of energy (often as ions or protons) which can be exchanged for larger amounts of energy in the form of ATP.[9] Just as a dam reserves water, thereby conserving potential energy which can be converted to electricity through a turbine, cells store charge on one side of a membrane which drives the rotation of ATP synthase (which interestingly resembles a turbine) to phosphorylate ADP to ATP, converting this energy in a high-energy covalent bond (Figure 1.2).[10–12]

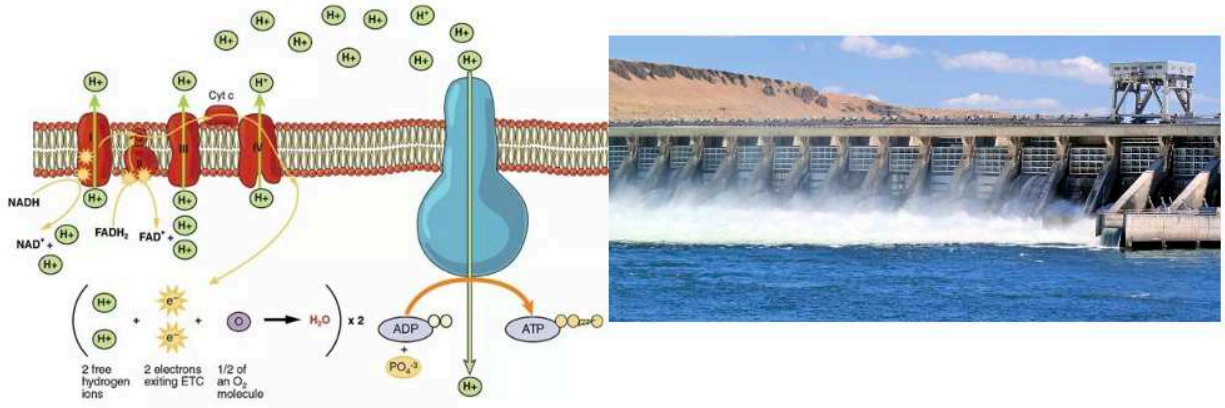


Figure 1.2: Electron transport mediated phosphorylation.[96, 97] Reduced cofactors like NADH or $FADH_2$ are oxidized at membrane-bound proton pumps (red) which provides energy to drive a proton across a membrane (left). Upon generating a sufficient gradient to drive ATP synthase (blue), protons flow through this specialized ion pump with enough energy to phosphorylate ADP to ATP, much like water in a reservoir generates electricity (right). To drive this process, electrons must ultimately reduce a terminal electron acceptor. Oxygen is an ideal terminal electron acceptor in that it maintains a large, positive reduction potential (E° = ~ +800 mV).[13] Upon reduction, oxygen combines with protons to make water.

To generate this proton gradient, energy must drive a proton pump. The energy input can come in the form of light energy or chemical energy. Either way, redox reactions with sufficient reduction potential can lead to the translocation of a proton across a membrane and

initiate a charge imbalance.[10–12] Upon binding 3-4 protons, the membrane-bound ATP synthase converts the potential energy stored in the proton motive force into a chemical bond by phosphorylating ADP to ATP.[14] Such a system requires that an organism maintain pumps which translocate protons or ions, often fueled by redox reactions from catabolic processes in the cytoplasm.[15,16] The variety of membrane-bound proteins which can provide this activity ranges from ferredoxin-dependent hydrogenases to methane-dependent methyltransferases.[17,18] Thiol chemistries can also provide redox cofactors.[9]

ATP synthase is conserved across all domains, though unique differences in the subunit size and structure lead to the efficiency of this important motor.[14] Much of the efficiency of ATP synthase is discussed in the terms of the electrochemical potential of the coupling ion (*e.g.* H^+), however the efficiency of the conversion of ADP to ATP is equally dependent on the intracellular phosphorylation potential (ΔG_p) which considers the concentrations of ADP, P_i, and ATP.[1] ΔG_p has been measured in a few model species and has since been considered a standard of ~+60 kJ mol^{-1},[11,12] but a lower ΔG_p could provide efficiency gains which are equivalent to increasing the electrical potential of the ion gradient.[1] In fact, the ΔG_p for an archaeal methanogen was estimated to be closer to ~ +45 kJ mol^{-1} suggesting that fewer protons would be required to generate an ATP in this organism.[19]

Electron bifurcation

The third energy conservation mechanism, electron bifurcation, has only been identified in the last few decades.[3–5] An electron bifurcation system allows for a two-electron transfer which couples an exergonic transfer with an endergonic transfer. Certain flavin-containing proteins are capable of electron bifurcation wherein two reduced flavin molecules are bound by a protein complex and oxidation of each flavin generates one high energy product and one lower energy product. The excess energy of the exergonic reaction drives the endergonic

reaction. This strategy was determined in *Clostridium kluyveri* where it was demonstrated that the endergonic reduction of a ferredoxin ($E^{\circ\prime}$ = -420 mV) and NADH ($E^{\circ\prime}$ = -320 mV) was coupled with the exergonic reduction of crotonyl-CoA to butyryl-CoA ($E^{\circ\prime}$ = -10 mV) (Figure 1.3B).[3] Such a technique can be considered a third mechanism of energy conservation along with substrate level phosphorylation and electron-transport linked phosphorylation.

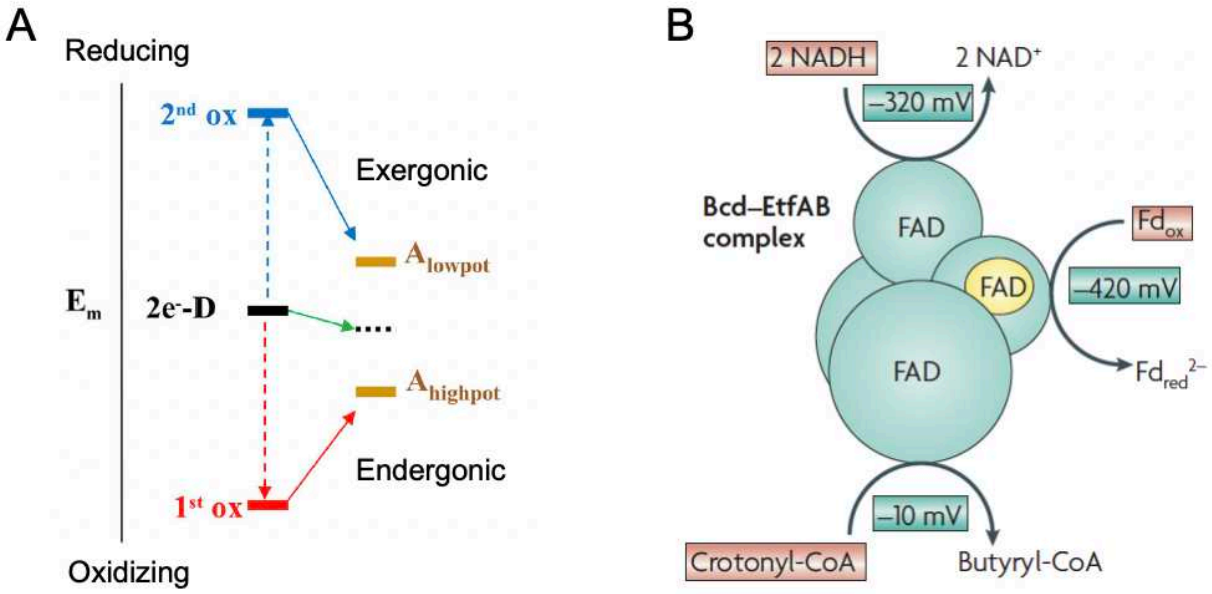


Figure 1.3: Electron bifurcation. A two-electron transfer conserves energy by coupling an exergonic transfer with an endergonic transfer. In (A), a two-electron donor is oxidized (black) and reduces a one-electron acceptor which has a high midpoint redox potential ($A_{highpot}$) and another one-electron acceptor which has a low midpoint redox potential (A_{lowpot}).[20] The reduction of $A_{highpot}$ is exergonic, driving the endergonic reduction of A_{lowpot} in a unique energy conserving mechanism. In (B), the exergonic reduction of crotonyl-CoA to butyryl-CoA is coupled to the endergonic reduction of Fd with NADH as was determined in *C. kluyveri*.[7] The enzyme complex responsible for this conversion is the butyryl-CoA dehydrogenase (Bcd) and two electron transfer flavoproteins (EtfAB). (Adapted from Nitschke *et al*. 2022 and Thauer *et al*., 2008.)

1.1.2 Diversity at the electron transport chain

ATP is widely considered the energetic currency of the cell but considering that the majority of ATP in the biosphere is generated from the potential energy at a charged membrane, we will now turn our focus to electrons and the electron transport chain which drives the generation of a charged membrane through redox chemistry. The electron transport chain is fueled by the movement of electrons from a donor molecule to a terminal electron

acceptor.[11,12,21] A variety of substrates can provide electrons, and likewise, a variety of substrates can serve as the terminal electron acceptor. Metabolisms have evolved to make use of unique substrates and acceptors in energy limited environments.[1] These lesser-known metabolisms likely provided the energy for ancient life and led to the oxygenation of the Earth's atmosphere.[22] The metabolic strategies discussed here employ a variety of substrates which provide sufficient energy to bank small gains across a charged membrane. These small gains are converted into high-energy, covalent bonds in the phosphorylation of ADP to ATP providing a long-term, stable asset. Before we transverse the variety of substrates and products which generate a charged membrane, we will review midpoint reduction potential as a foundation to the concepts that follow.

Midpoint reduction potential – a crucial concept in redox chemistry

For a charged membrane to effectively facilitate electron transfers such that a proton or ion gradient is established, the effective midpoint reduction potential of electron donors and acceptors must be understood. Bioenergetic molecules can maintain charge (or reduction potential) which can be transferred "downhill" such that a molecule with a more negative reduction potential can reduce a molecule with a more positive reduction potential (Figure 1.4).[23] Electron acceptors with high, positive midpoint reduction potentials provide the most versatility in redox reactions, which is why oxic strategies dominate a broad range of metabolisms. With a staggering reduction potential of $E^{\circ} \sim +800$ mV, oxygen is an ideal electron acceptor. Oxygen can receive an electron from a wide variety of electron donors as long as the reduction potential of the donor is lower (more negative) than that of oxygen.

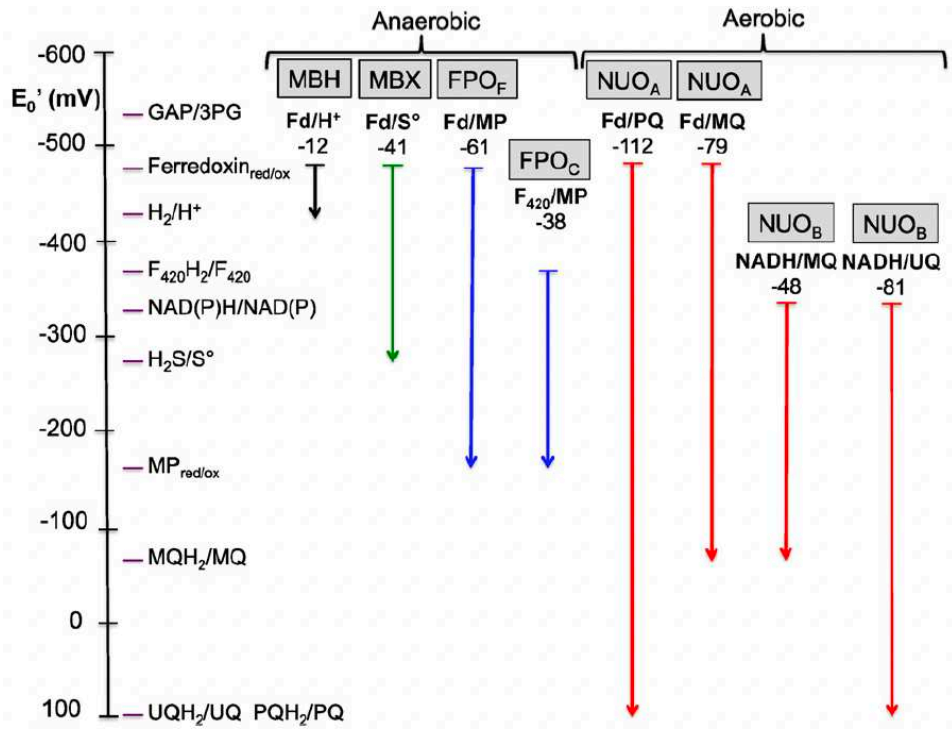


Figure 1.4: Redox potentials of select bioenergetic molecules in anaerobic and aerobic metabolisms. The redox potential of reducing agents such as Fd, F_{420}, and NADH are shown in anaerobic (left) and aerobic (right) environments as measured by the midpoint reduction potential (E_0' in mV, scale at far left). The free energy (in kJ mol^{-1}) released in each conversion is noted above the vertical line indicating the change in redox potential between the donor/acceptor pair. Fd_{red} (~ -480 mV) is one of the few molecules which maintains a low enough reduction potential to reduce a proton (E = ~ -420 mV) resulting in ΔG = ~ -12 kJ mol^{-1}. Reduction of S° generates almost three-fold energy (ΔG = ~ -41 kJ mol^{-1}), while reduction of methanophenzine (MP) by Fd_{red} which occurs in methanogenesis generates ΔG = ~ -61 kJ mol^{-1}. Aerobic energetic gains involving UQ or PQ resolve considerably greater free energy, and it should be noted that oxygen, which is not included in this scale, maintains ΔG = ~ +800 kJ mol^{-1}, thus a majority of energetic gains are driven by the large change in reduction potential of an electron donor to oxygen which readily gains two protons and forms water upon reduction of two electrons. (MBH = membrane-bound hydrogenase, MBX = membrane bound sulfane reductase, FPO = methanophenzine-reducing subunit of a methanogenic respiratory complex, Fd = ferredoxin, , MQ = menaquinone, UQ = ubiquinone, PQ = plastiquinone, GAP = glyceraldehyde-3 phosphate, 3PG = glyceraldehyde-3-phosphate, S° = elemental sulfur, and NUO = NADH quinone oxidoreductase.) (Adapted from Schut *et al*., 2016.)

Anaerobes must find alternative electron acceptors. Sulfur in the form of polysulfides or inorganic sulfur is an effective electron acceptor, when available. Like oxygen, the outer valence shell of sulfur is short two electrons, thus sulfur is a suitable terminal electron acceptor resulting in the production of H_2S (comparable to H_2O). The free energy change of the conversion of H_2 and S to HS^- and H^+ is -28 kJ mol^{-1} (Figure 1.4).[23]

In the absence of a better electron acceptor, protons are plentiful and serve as a possible terminal electron acceptor, producing H_2. Reduction of a proton requires E° ~-414 mV, but many electrons are not available at such a low reduction potential (*e.g.* NAD(P)H E° ~-350 mV). Ferredoxin proteins maintain iron-sulfur clusters and are capable of low reduction potentials (E° ~-500 mV), thus they are ideal electron donors to reduce protons (E° ~-414 mV) to H_2.[24] The presence and expression of proteinaceous electron carriers become important factors if protons are used as terminal electron acceptors.

Upon reduction, terminal products must be cycled such that an increase in their concentration does not limit electron flux.[25,26] Aerobic systems reduce oxygen to water, but some anaerobic systems generate gasses, such as H_2 or H_2S. The benefit of gaseous terminal products is that they are literally blown away, managing any buildup through simple gaseous diffusion.[1]

Canonical electron transport in photosynthesis and respiration

Perhaps the most foundational process which supports life is the energetic gain harvested from light.[27] Many photoautotrophs absorb photon energy via light-sensitive pigments in membrane-bound photosystems. Light energy catalyzes the ultimate conversion of water (the electron donor) and carbon dioxide into carbohydrates, thereby fixing inorganic carbon into glucose, an energy-dense substrate which is broadly metabolized to produce high energy gains via cellular respiration. In an elegant exchange, water disassociates into O_2 and $2H^+$ in photosynthesis while O_2 is the terminal electron acceptor in cellular respiration producing H_2O and CO_2. Thus, carbon fixation evolves O_2 while respiration consumes it in aerobic life. With a high, positive midpoint reduction potential of O_2, aerobic strategies reign supreme, however a multitude of other strategies should be considered for their ingenuity in garnering energy from substrates that are otherwise deemed inert to many organisms.[1]

One-carbon metabolisms

Methanogens have devised a way to generate energy from synthesizing methane from simple substrates such as H_2 + CO_2, but they can also consume methanol, acetate, or trimethylamines (Figure 1.5).[18,28–30] Fds are reduced by H_2-consuming hydrogenases. CO_2 is reduced by Fd_{red} in multiple steps which fully saturates the carbon molecule into CH_4, while generating a Na^+ gradient at the membrane.[7,28] A Na^+-ATPase drives the production of ATP and the terminal electron acceptor is a heterodisulfide Co-M and Co-B, catalyzed by the electron bifurcating enzyme, heterodisulfide reductase (HDR).[31] Some methanogens contain cytochrome complexes in the membrane, which allow for the generation of an ion gradient such that a non-specific ATPase is driven by binding either Na^+ or H^+.[1,32] Only two ions are translocated per mol of methane formed, leading to 0.5 mol ATP per mol of methane, thus methanogens live at the thermodynamic limit of life.[33] Methane oxidation wherein methane is converted to CO_2 or methanol produces minimal energetic gains but can be used for growth by methanotrophs.[30,33]

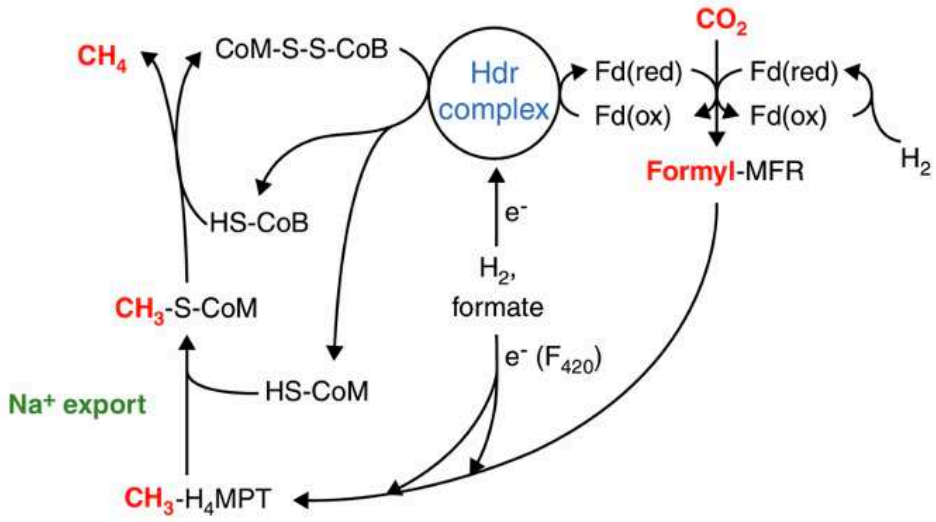


Figure 1.5: Methanogenesis. H_2 is consumed to generate Fd_{red} (top right) which reduces CO_2 to formyl-methanofuran (MFR). Upon multiple reductions, formyl-MFR is converted to CH_3-tetrahydromethanopterin (H_4MPT), and a Na^+ gradient is generated upon H^+ exchange from reduced-CoM (HS-CoM) to CH_3-S-CoM. Methane (CH_4) is generated upon a second thiol reaction wherein reduced-CoB (HS-CoB) provides a hydrogen to CH_3 and subsequently forms a heterodisulfide (CoM-S-S-CoB) which is reduced via electron bifurcation at the heterodisulfide reductase (Hdr) complex. (Figure adapted from Costa and Leigh, 2014.)

Methanogens which encode a formate dehydrogenase (FDH) can consume formate upon conversion to CO_2, however it was surprising to find that growth of the hyperthermophile, *Thermococcus onnurineus,* was observed on formate alone.[34] In *T. onnurineus,* formate oxidation to CO_2 by FDH is coupled to the membrane-bound hydrogenase (MBH),[35] thus it is proposed that formate oxidation drives the generation of the proton gradient at the membrane evolving H_2 gas in the process. The conversion of formate to H_2 and CO_2 has been measured at ~ ΔG -8 to -20 kJ mol^{-1},[34] and produces ~ 1.3 kJ mol^{-1} per oxidation and was thought to not provide enough energy or only be feasible in mix cultures with syntropic bacteria. However, the ΔG becomes negative (-2.6 kJ mol^{-1}) at 80˚C, providing for energetic gains, albeit small ones, in a hyperthermophilic lifestyle. The *Ton*-FDH:MBH complex was heterologously expressed in another hyperthermophile and while the expression did not confer sufficient energy for growth on formate alone, H_2 production was increased six-fold.[36]

Sulfate reducing metabolisms

A discussion involving diverse metabolic strategies would be remiss without including sulfate reducing prokaryotes, which can utilize a variety of substrates but employ a unique electron disposal system.[37] Sulfur maintains a broad range of oxidation states (from -2 to +6), and sulfate (SO_4^{2-}) can be used as a terminal electron acceptor in sulfate-reducing organisms.[38] Terminal electron acceptors must be dispersed else their concentration would increase to effectively halt energetic gains. In sulfate reducing organisms, sulfate is converted to sulfide (H_2S) which is a gas and can disperse without issue, but the conversion comes at a cost of one ATP (Figure 1.6). Sulfate (SO_4^{2-}) reduction to sulfite (SO_3^{2-}) requires $E^{0'}$ = ~ -516 mV, and thus even Fd_{red} ($E^{0'}$ = ~ -500) is not sufficient to perform this conversion.[23] Instead, sulfate reducing organisms hydrolyze ATP to ADP to convert sulfate to adenosine 5'-phosphosulfate (APS).[38] APS is readily converted to sulfite ($E^{0'}$ = ~ -60 mV) by NADH and further converted to HS^-.

Conversion of sulfate to sulfide consumes energy in the form of ATP to decrease the change in reduction potential.

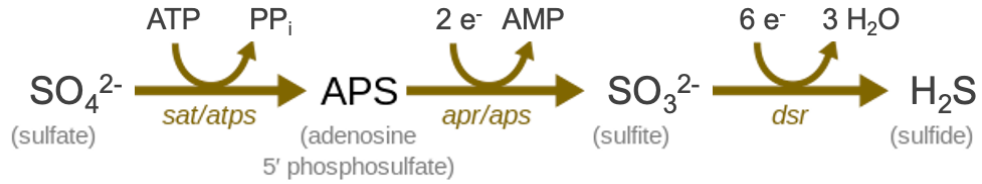


Figure 1.6: Sulfate conversion to sulfide in sulfate reducing bacteria. Sulfate is converted to APS with the hydrolysis of ATP to ADP. APS is converted to sulfite and then sulfide which is easily disposed of as a final reductant. (Figure adapted from Muyzer and Stams, 2008.)

Electron donors in the form of light

Photoexcitation at Complex I drives photosynthesis in aerobic environments, but the diversity of metabolic strategies expands in the photoautotrophic realm when one considers anoxygenic photosynthesis. Take, for example, the *Halobacteriaceae*, a salt-loving Archaea. This halophile contains a light-driven proton pump, bacteriorhodopsin, which covalently conjugates a light-sensitive retinal molecule.[39] Upon excitation by light, bacteriorhodopisin translocates a proton across the membrane which is then used to drive ATP synthase via the proton motive force.

1.1.3 Many ways to get energy but a few trends are conserved

Amongst the diversity of substrates and pathways to store energy, a few themes arise. First organisms store energy in bonds or in gradients and the energy between the two must be converted without significant loss. Second, energy is conserved in the management of electrons and protons (or ions). Lastly, most gains are not immediate but instead rely on multi-step mechanisms. Essentially all life relies on redox reactions and a charged membrane and the vast bulk of generating sufficient energy for life comes down to correctly moving electrons around.[1,22]

1.2 Electrons as energetic currency

Electrons are too electronegative and too reactive to exist in cells without a binding partner. Redox potential is a measure of the tendency of a chemical moiety to lose or acquire an electron, and a variety of small molecule and proteinaceous electron carriers are commonly employed (Figure 1.7).[40]

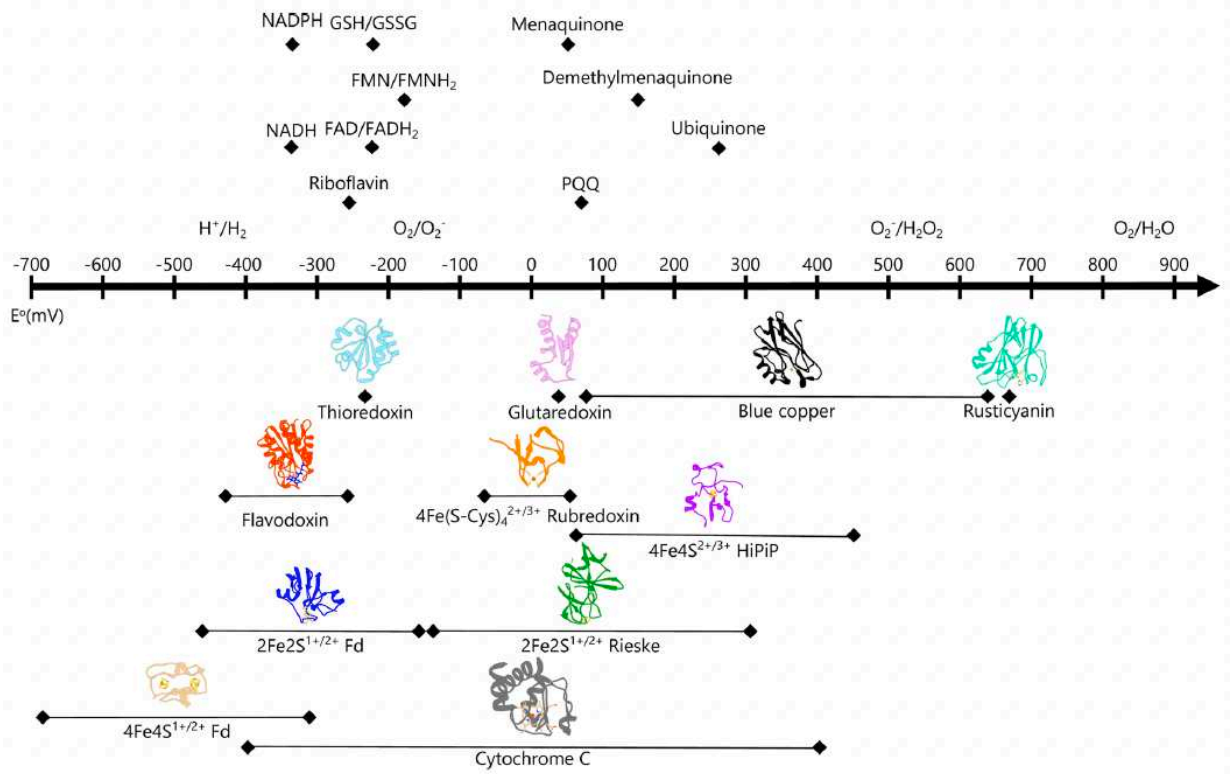


Figure 1.7: Common electron carriers and their midpoint reduction potentials. A midpoint reduction potential is shown as a horizontal line where the most negative reduction potential (E˚ -700 mV) is observed at far left and increases to the most positive reduction potential (E˚ +900 mV) at far right. Common redox-sensitive small molecules are shown above the line and placed according to their midpoint reduction potential, while proteins which maintain metal centers and are capable of electron transport are shown below the line and placed according to the range of reduction potentials they have been shown to provide (indicated by the horizontal line between two diamonds). (Adapted from Atkinson *et al.*, 2016.)

Electrons provide energy via transfer when they are passed from a carrier with a lower (more negative) reduction potential to a carrier with a higher (more positive) reduction potential, and the greater the number of passages which lead to “work” being done, the more energy is conserved. For instance, NADH (E˚ ~-325 mV) can be used to reduce O_2 to H_2O (E˚ ~+800), but

a ΔE° of 1125 mV is arguably overkill when other molecules can do the same job with less input. The movement of electrons, deemed electron flux, thus offers energy conservations strategies.

Small molecules are ubiquitous and can be regulated in the case of substrate binding sites within an enzyme, but they do not offer the regulatory potential that proteinaceous electron carriers (PECs) offer (Figure 1.8).[40,41] While the construction of a protein requires a greater energetic investment than that for a small molecule, the surface interactions of proteins provide for adaptable and regulated interactions.

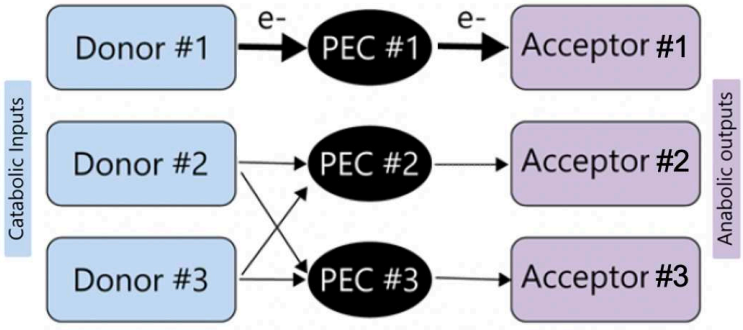


Figure 1.8: Electron transfer pathways between PECs with preferred electron donors and electron acceptors. Proteinaceous electron carriers (PECs, black ellipses at center) will associate with specific electron donors (light blue rectangles) and electron acceptors (light purple rectangles) such that electron flux may be regulated towards one anabolic outcome over another. Donor #1 will associate with PEC #1, wherein PEC#1 will become reduced and associate with Acceptor #1. Acceptor #1 will generate a unique anabolic product than either Acceptor #2 or #3. In this way, PECs provide regulated and adaptable electron transfer. Mapping specific metabolic Donor:PEC:Acceptor relationships is of interest for directing electrons to valuable bioproducts such as H_2 which is a valuable biofuel. (Adapted from Atkinson *et al*., 2016.)

Protein regulated electron flux can be thought of as any other protein-protein regulatory network, with the consideration that electrons must be passed at each interaction. Proteins which maintain metal centers, often composed of iron, nickel, and/or sulfur, provide effective electron carriers, and can provide exceptional reduction potentials (*e.g.* $E^{\circ} \sim -700$).[40] Transfer of an electron from a donor to an acceptor via a protein interface suggests that electrons tunnel through a protein medium.[42] Experimental evidence supports that electrons can transfer up to ~ 14 Å.[43]

Electron flux through a tunable and regulated system of protein-protein interactions can offer exceptional energy conservation strategies in energy limiting conditions, such as those seen in the early Earth.

1.3 Exploring electron flux in archaeal metabolisms

As their name implies, archaea are thought to be related to the prokaryotic organisms which inhabited the early Earth.[9,17,44–47] Defined by a unique isoprenoid lipid membrane and ribosomes which resemble eukaryotic machinery more so than that in the Bacteria, the archaeal domain comprises an array of unique metabolisms.[48–50] While many Archaea live an aerobic, mesophilic lifestyle,[51] they also occupy habitats that were otherwise thought to be devoid of life,[2,7,52,53] giving rise to particular metabolic strategies discussed in Section 1.1. The energy conservation mechanisms discussed above offer minimal gains in extreme environments, but they are sufficient for life.

Here we will turn our focus from the diversity of energy conservation mechanisms and focus on a single archaeal hyperthermophile, *Thermococcus kodakarensis.*[54–56] Via genetic modification and quick growth rates, this genetically tractable anaerobe has captured our attention. We can probe energy conservation mechanisms and regulated electron flux such that we reroute pathways towards the production of hydrogen gas, a valuable biofuel.

1.3.1 Metabolism of a model anaerobic, hyperthermophilic archaeon

Many hyperthermophiles are found in anaerobic environments and utilize inorganic compounds like molecular hydrogen, carbon dioxide, ferrous iron, and sulfur as their final electron acceptor.[1,2] *Thermococcus kodakarensis* is an anaerobic, hyperthermophilic archaeon that ferments amino acids and sugars and can double every ~ forty minutes in optimal conditions.[54,55] This hyperthermophilic archaeon utilizes a modified Embden-Meyerhof Parnas (EMP) glycolytic pathway (Figure 1.1B) and catabolizes amino acids into acyl-CoA moieties

which are further converted to CO_2 and acids.[57,58] Fermentation reactions produce reduced small molecules, such as NAD(P)H, or reduced PECs, such as Fd_{red} which can further transfer electrons towards pathways for growth, energy conservation strategies, or funnel electrons to reductant sinks such as fully saturated lipids (Figure 1.9).[26]

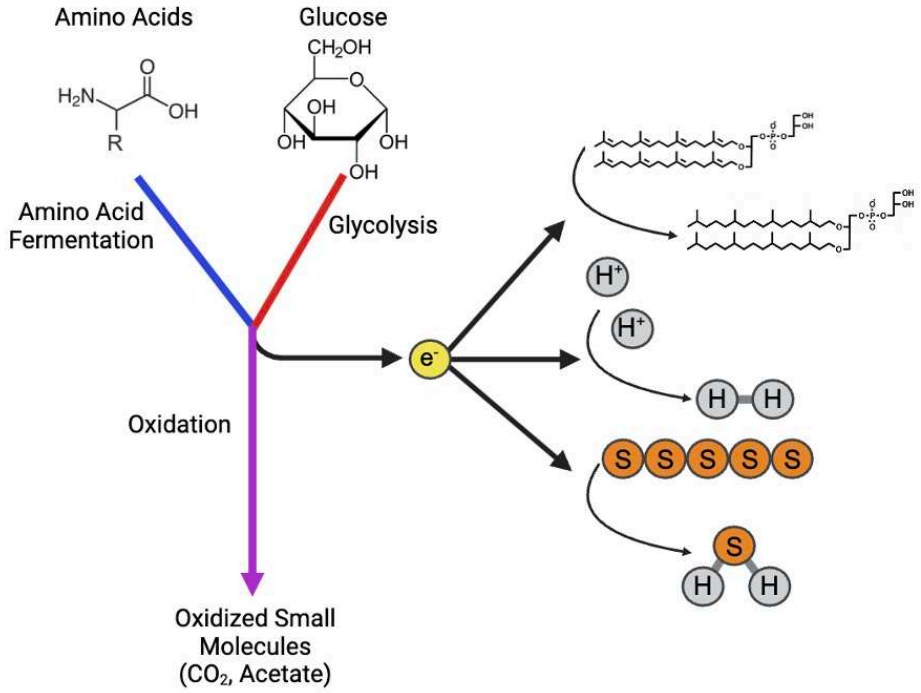


Figure 1.9: An overview of metabolism and reductant sinks in *T. kodakarensis*. *T. kodakarensis* utilizes electrons gained from the catabolic reactions of amino acid fermentation and glycolysis (top left) to generate oxidized small molecules (bottom left). Electrons are shuttled to various anabolic outcomes but ultimately, three reductant sinks are employed (right). Unsaturated isoprene lipids are reduced to fully saturated, mature lipids, protons are reduced to generated H_2 gas, and polysulfides are reduced to generate H_2S gas when inorganic sulfur is present.

T. kodakarensis obtains a small amount of energy from the breakdown of amino acids and carbohydrates via substrate-level phosphorylation, but the primary mechanism by which it produces energy is via the generation of a chemiosmotic gradient using two different membrane-bound reductases (Figure 1.10).[16,17,44,59] The preferred of these two complexes, the membrane bound sulfane reductase (MBS), utilizes sulfur as the final electron acceptor.[44,60,61] When sulfur is unavailable, *T. kodakarensis* employs an alternative respiration complex called the membrane bound hydrogenase (MBH) which couples the generation of a proton gradient to

hydrogen production via the reduction of protons.[17] The ability to reduce protons to hydrogen gas makes the metabolic pathway of *T. kodakarensis* a desirable target for developing a comprehensive understanding of foundational biological hydrogen generation as well as offering a potential target of interest for bioengineering.

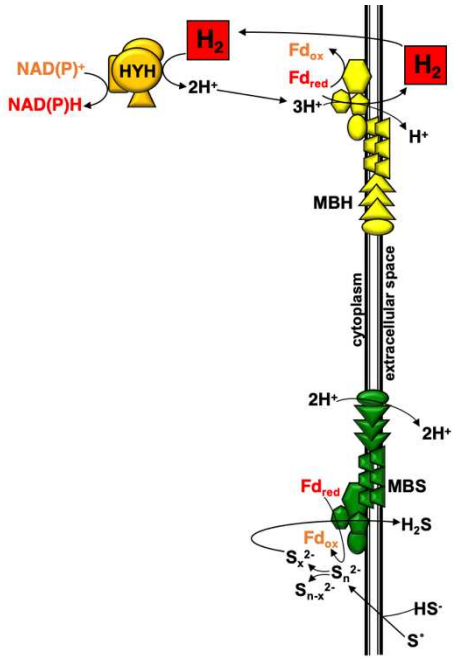


Figure 1.10: Terminal electron acceptors and the generation of a proton gradient. Membrane-bound hydrogenase (MBH, yellow) generates H_2 with the translocation of a single proton across the membrane, while membrane-bound sulfane reductase (MBS, green) reduces elemental sulfur to translocate two protons across the membrane. H_2 can freely cross the membrane and be consumed by a cytosolic hydrogenase (HYH, orange) to generate reduced NAD(P)H. Both MBH and MBS are multimeric complexes which maintain a membrane arm which includes a proton pump and a cytosolic arm which maintains Ni-Fe moieties which oxidize Fd_{red}. Both MBS and MBH respiratory systems resemble Complex I, though MBH is considered to be the ancestral respiratory system. (Figure adapted from Burkhart *et al*., 2019.)

T. kodakarensis utilizes three ferredoxins (Fds) which can reduce protons due to maintaining exceptionally low reductions potentials.[40,62] Fds are PECs of particular interest in the generation of H_2 gas, as they maintain a variety of combinations of metal centers which confer reduction potentials low enough to reduce H^+ to H_2.[42,63] Coordinated production of a specific Fd could, in principle, permit selective transport of electrons from select electron donors to specific electron acceptors. Selective electron transport could provide routes for electron flux

that favor individual reactions without demanding cellular-level shifts in redox balancing. Previous studies showed that the flow of electrons was regulated through specific and distinct protein partnerships.[41]

The goal of this work is to i) map the *in vivo* metabolic protein interactome in multiple redox conditions using a novel technique, affinity purification and protein mass spectrometry (AP-MS) and ii) redirect electron flux by fusing electron donor and acceptor proteins (a first in *T. kodakarensis*) and measuring the impact to reductant pools.

1.4 The state of the field pre and post

The Archaeal domain was identified in 1977, and the study of this group grew slowly as species from disparate environments were identified.[47,49,50,55,56,64,65] As remains the case today, only a minority of archaeal extremophiles can be grown in a lab and even fewer offer genetic techniques. *T. kodakarensis*, which was isolated in marine sediment off the coast of Kodakara Island in Japan,[55,56] is essentially a microbial "weed"; it can grow on a variety of substrates with a doubling time of ~40-minutes in optimal conditions. Genetic techniques for this exceptional model thermophile were developed in the early 2000's,[66] and evolved to allow for markerless genomic modification in just a few weeks. [54,66–70] Thus, *T. kodakarensis* became the model hyperthermophile of choice for many avenues of archaeal research.

The metabolism of a closely related hyperthermophile isolated from the Mediterranean, *Pyrococcus furiosis,*[64] was studied extensively in the early 1990's using classic *in vitro* techniques which were available at the time.[6,25,53,57,58,71–79] Proteins were natively or recombinantly purified and studied in isolation, which led to the basic understanding of protein compositions (from monomers to multimers). The limited number of substrates and cofactors that could be mixed with these purified proteins suggested certain functions, but it remains that little is known about expression patterns, physiological responses to a changing environment, or what the *in vivo* protein landscape looks like.

1.4.1 The complexity of the *in vivo* protein landscape

The simplistic nature of *in vitro* studies has suggested that enzymes act on specific substrates, produce specific products, and do so in the composition they were identified, but the complexity of the *in vivo* environment is likely much more dynamic (Figure 1.11).

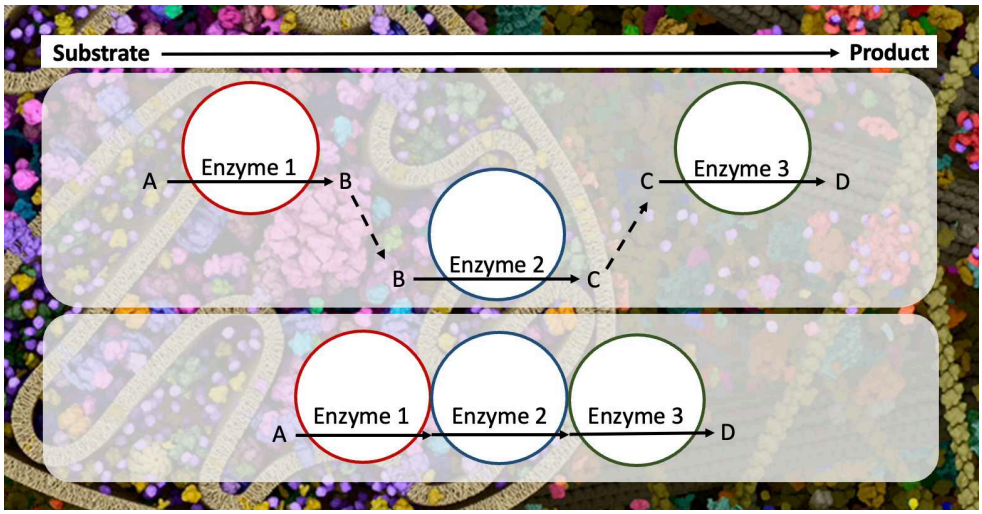


Figure 1.11: Catalysis *in vivo* is likely more dynamic than *in vitro* studies can predict. A purified protein can be seen to catalyze a reaction from a substrate (A) to a product (B) (top panel), and such activity is indicative of its function. However, these data arrive from isolated studies with limited substrates, such that they likely cannot resemble the true state of *in vivo* activities which likely function in a more dynamic way (lower panel). Here, our simplified understanding of isolated protein catalysis is shown in the top panel where each enzyme catalyzes a single reaction leading from substrate (A) to product (B) which is the substrate of Enzyme 2 (B) catalyzing the product (C) and so on. Instead, the *in vivo* protein landscape may be more complex where multiple proteins interact in a transient manner such that substrate (A) is converted to product (D) via the same enzymes in concert as opposed to in isolation.

Understanding *in vivo* macromolecular interactions is arguably the current frontier of molecular biology. Crosslinking techniques[80] have been employed for decades and genetic tools have evolved from assays developed in the late 1980's,[81] but techniques still struggle to correctly identify nanoscale interactions in complex biological environments.[82] Even words like "dynamic" or "complex" seem to be understatements when considering the four-dimensional coordination of chemistry within even a simple single-celled organism.

A few methods have moved the field forward in recent years. Microscopic assays using fluorescent labeling have shown *in vivo* interactions at the translational interface via live

imaging,[83,84] but these techniques are limited to abundant protein complexes and aerobic lifestyles which permit fluorescent imaging. Liquid-liquid phase separation has been observed via cryo-electron tomography and suggest that biomolecular condensates occur in living systems which localize and modulate chemistry through spatial and temporal coordination.[85]

Combinatorial methods – which layer multiple *in vitro* and *in vivo* techniques with large data analysis methods – have come to the forefront. While affinity purification followed by proteomic tandem mass spectrometry (AP-MS) methods have been available since the late 2000's,[86–92] only in the last year has this technique been used to map whole proteomes.[93] The AP-MS method has been used successfully in *T. kodakarensis* to map the components of the replisome,[86] and more recently to define partnerships of the three Fds.[41]

The Fd interactome supported the idea that distinct protein associations likely aid in modulating reductive flux. Each Fd was observed to have unique electron donors and electron acceptors, as proposed in Figure 1.8. Such specific interactions suggest that expression and activity of each Fd can be used to tune reductive flux. It was the goal of this project to expand on this hypothesis, which was done in two ways: i) the metabolic protein network was assessed by AP-MS to reveal protein-protein associations in the presence and absence of sulfur and ii) Fd proteins were fused to two proteins involved in unique reductant sinks to determine if we could intentionally tune reductive flux.

1.4.2 Contributions to the field

The fast growth of naturally competent microbial systems allows for us to probe metabolic flux *in vivo* using combined genetic and biochemical methods. Genomically tagged proteins with epitope and affinity sequences expressed at native levels can be generated in otherwise markerless *T. kodakarensis* parent strains in as little as three weeks.[67,70,94] Over forty new strains were constructed for this thesis, and these strains used to define and redefine metabolic partnerships and reroute electron flux *in vivo* (Figure 1.12).

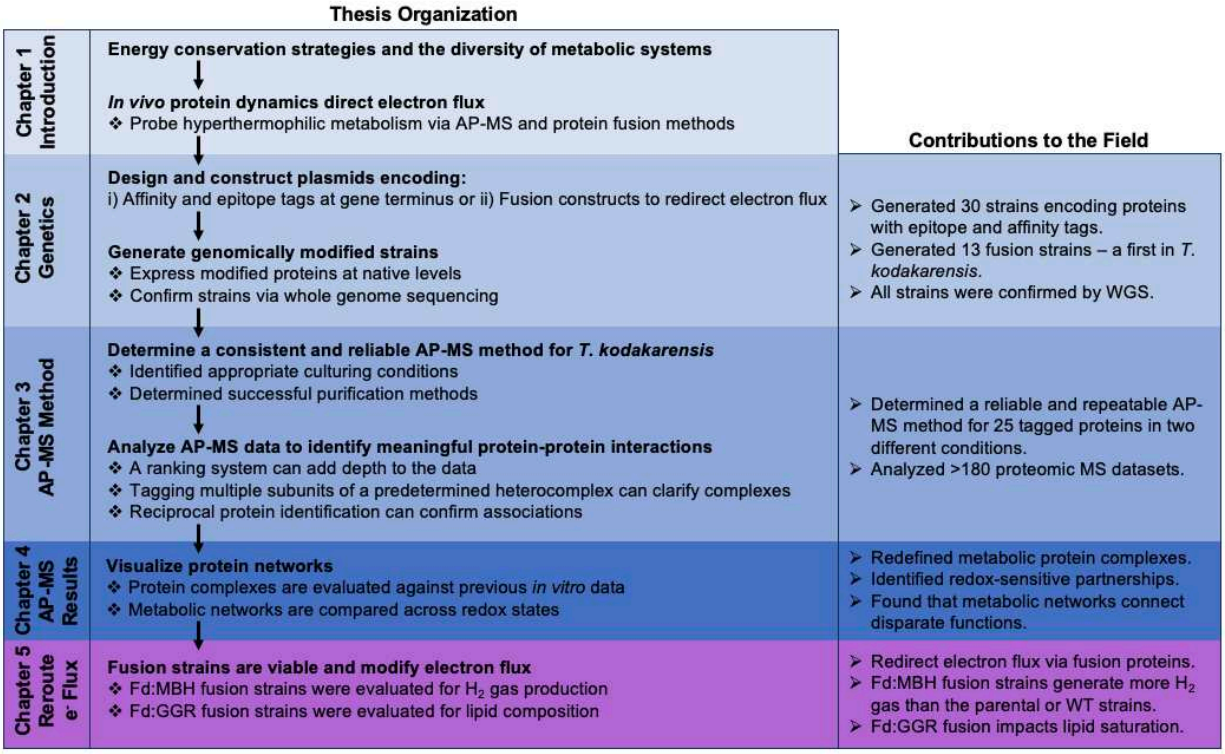


Figure 1.12: Thesis organization and contributions to the field. This thesis has been organized to introduce the reader to energy conservation strategies of diverse metabolisms in Chapter 1 (light blue, at top). Chapter 2 describes the genetic methods applicable for *T. kodakarensis* and defines the strains that were constructed to evaluate the metabolic protein landscape *in vivo* (Chapters 3 and 4) and see if electron flux could be perturbed by fusing electron donor and acceptor proteins (Chapter 5, fuchsia at bottom). Contributions to the field as organized by chapter topic are noted at right.

Applying the AP-MS method to understand the in vivo metabolic protein landscape

Affinity purification of the clarified cell lysate followed by proteomic mass spectrometry (AP-MS) is used to identify protein associations which were previously unknown by standard *in vitro* techniques.[41,86,93] In this work, we have applied AP-MS to thirty proteins of known metabolic function to map their *in vivo* associations in two redox conditions. The goal of this project was to initially map redox protein partnerships, however the large number of proteins included in this study led to refining the purification process and the AP-MS analysis method for *T. kodakarensis.* The process of determining a repeatable and successful method is the focus of Chapter 3. Additionally, large datasets generated by AP-MS require stringent analysis.

Statistical rigor and the development of a ranking system to add depth of meaning to the data are also detailed in Chapter 3 (Figure 1.13).

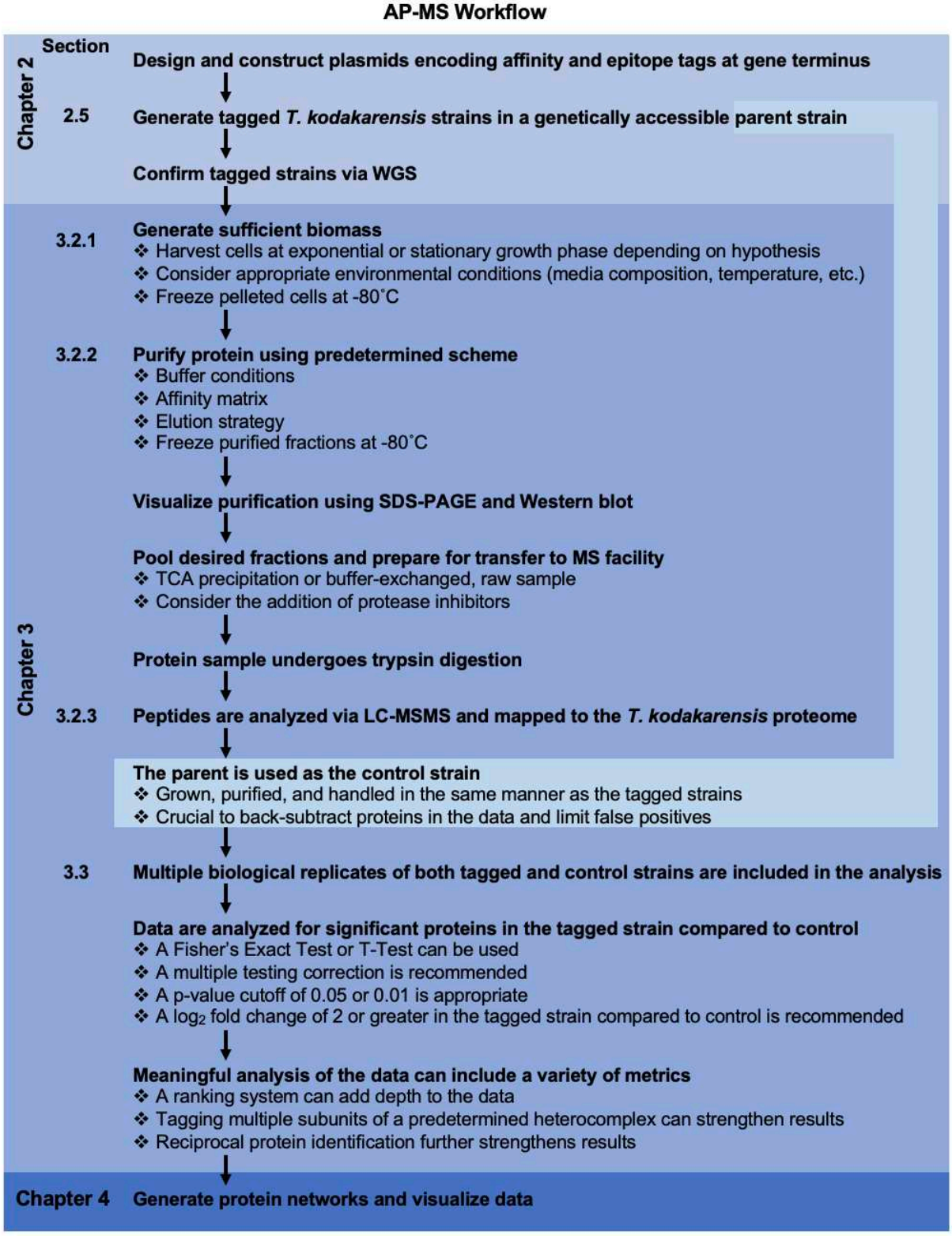


Figure 1.13: The AP-MS workflow as organized by chapter and section in this thesis. Steps for AP-MS procedure and analysis are detailed in Chapters 2 - 4. Chapter 2 (light blue, top) describes the genetic manipulation of *T. kodakarensis* to genomically encode affinity-tagged proteins expressed at native levels. Chapter 3 (blue, middle) describes the steps involved in optimizing this method for ~ thirty proteins involved in metabolism in *T. kodakarensis*. The parent strain in which all tagged-strains were generated serves as the control, and is grown, purified, and analyzed in the same manner as tagged strains (bright blue connecting Chapter 2 to Chapter 3 at far right). Chapter 4 (darker blue, bottom) details the results of the entirety of the AP-MS method.

Mapping associations of twenty-five proteins led to expected and unexpected findings which are the focus of Chapter 4. Proteins whose complex associations were previously identified by *in vitro* studies were confirmed in some cases and redefined in others. Additionally, some complexes were seen to form different associations when sulfur was present. For example, two heterotetrameric oxidoreductases (oxoisovalerate oxidoreductase, VOR and pyruvate oxidoreductase, POR) were observed to function independently in the presence of sulfur but may form a larger octameric complex when redox conditions shift and protons are the only available terminal electron acceptor. Additionally, complexes which are involved in glutamate catabolism appear to localize near the membrane-bound sulfane reductase, possibly providing a localized electron shuttle for efficient metabolisms. I will direct the reader to the results and discussion sections of Chapter 4 for a full overview of results from this broad AP-MS study. In summary, we now have an established portfolio of interactions that detail protein associations in a more dynamic state. We have described the adaptable nature of these associations in response to redox states and challenged limited existing data from *in vitro* studies. Finally, we present new functional protein-protein associations which can be used to better understand dynamic metabolic processes *in vivo.*

Redirecting electron flux

By restricting the local concentration of two proteins such that they are physically linked in the cell at all times, we aimed to redirect electron flux (Chapter 5). Fusion constructs have

never before been generated in *T. kodakarensis,* thus, this work demonstrates a first in synthetic engineering in this model hyperthermophile. Thirteen fusions strains were constructed, wherein a Fd was tethered to either the MBH or geranylgeranyl reductase (GGR), which are responsible for hydrogen evolution and lipid saturation, respectively.[17,95] Copious data suggests a strong relationship between Fd2 and GGR, thus we only generated a single fusion combination to GGR, however twelve fusion constructs were generated for MBH, as directing electrons to MBH is of particular interest in the production of H_2 gas. Each Fd was fused to either of two MBH-subunits using either of two linkers – and all twelve strains were viable. Fusion strains were evaluated for H_2 production and growth to evaluate unique phenotypes of each Fd:linker:MBH combination.

Fd:MBH fusion strains showed increased hydrogen generation compared to both parental and wild type strains, suggesting that this physical constraint can increase electron transfer as measured by H_2 production. To note, exceptional increases were observed when Fd3 was tethered to the MBH-J subunit by a short flexible linker. Fd:GGR fusion strains resulted in a change to lipid composition such that mature cultures were observed with considerably more unsaturated bonds in the fusion strain compared to the parental strain. These data suggest that such a linkage between Fd2 and GGR may inhibit electron transfer. Interestingly, while fusion strains redirected electrons as observed by a change in reductive sinks, growth defects were limited suggesting that electron flux can be redirected without significantly hampering growth.

REFERENCES

1. Mayer, F. & Müller, V. Adaptations of anaerobic archaea to life under extreme energy limitation. *FEMS Microbiol Rev* **38**, 449–472 (2014).
2. Schäfer, G., Engelhard, M. & Müller, V. Bioenergetics of the Archaea. *Microbiology and Molecular Biology Reviews* **63**, 570–620 (1999).
3. Li, F. *et al.* Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from Clostridium kluyveri. *J Bacteriol* **190**, 843–850 (2008).
4. Buckel, W. & Thauer, R. K. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation. *Biochimica et Biophysica Acta - Bioenergetics* vol. 1827 94–113 Preprint at https://doi.org/10.1016/j.bbabio.2012.07.002 (2013).
5. Lubner, C. E. *et al.* Mechanistic insights into energy conservation by flavin-based electron bifurcation. *Nat Chem Biol* **13**, 655–659 (2017).
6. Bräsen, C., Esser, D., Rauch, B. & Siebers, B. Carbohydrate metabolism in archaea: Current insights into unusual enzymes and pathways and their regulation. *JAMA Ophthalmol* **132**, 326–331 (2014).
7. Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea: Ecologically relevant differences in energy conservation. *Nature Reviews Microbiology* vol. 6 579–591 Preprint at https://doi.org/10.1038/nrmicro1931 (2008).
8. Hoehler, T. M., Alperin, M. J., Albert, D. B. & Martens, C. S. *Apparent Minimum Free Energy Requirements for Methanogenic Archaea and Sulfate-Reducing Bacteria in an Anoxic Marine Sediment*. www.fems-microbiology.org.
9. Martin, W. F. & Thauer, R. K. Energy in Ancient Metabolism. *Cell* **168**, 953–955 (2017).
10. Mitchell, P. Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *Biological reviews of the Cambridge Philosophical Society* vol. 41 445–502 Preprint at https://doi.org/10.1111/j.1469-185x.1966.tb01501.x (1966).
11. Kashket, E. R. *Stoichiometry of the H+-ATPase of Growing and Resting, Aerobic Escherichia CoW*. vol. 21 https://pubs.acs.org/sharingguidelines (1982).
12. Slater, E. C., Rosing, J. & Mol, A. *THE PHOSPHORYLATION POTENTIAL GENERATED BY RESPIRING M1TOCHONDRIA*. *Biochimica et Biophysica Acta* vol. 292 (1973).
13. Truong, A., Myerscough, D., Campbell, I., Atkinson, J. & Silberg, J. J. A cellular selection identifies elongated flavodoxins that support electron transfer to sulfite reductase. *Protein Science* **32**, (2023).
14. Grüber, G., Manimekalai, M. S. S., Mayer, F. & Müller, V. ATP synthases from archaea: The beauty of a molecular motor. *Biochim Biophys Acta Bioenerg* **1837**, 940–952 (2014).
15. Peters, J. W. *et al.* [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochim Biophys Acta Mol Cell Res* **1853**, 1350–1369 (2015).
16. Maier, R. J. *RESPIRATORY METABOLISM IN HYPERTHERMOPHILIC ORGANISMS: HYDROGENASES. SULFUR REDUCTASES. AND ELECTRON TRANSPORT FACTORS THAT FUNCTION AT TEMPERATURES EXCEEDING 100°C*. (1996).
17. Yu, H. *et al.* Structure of an Ancient Respiratory System. *Cell* **173**, 1636-1649.e16 (2018).
18. Sarmiento, F., Mrázek, J. & Whitman, W. B. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Methanococcus maripaludis. *Proc Natl Acad Sci U S A* **110**, 4726–4731 (2013).
19. Jetten, M. S. M., Stams, A. J. M. & Zehnder, A. J. B. Adenine nucleotide content and energy charge of Methanothrix soehngenii during acetate degradation. *FEMS Microbiol Lett* **84**, 313–318 (1991).

20. Nitschke, W. *et al.* Aqueous electrochemistry: The toolbox for life's emergence from redox disequilibria. *Electrochemical Science Advances* vol. 3 Preprint at https://doi.org/10.1002/elsa.202100192 (2023).
21. Schink, B. *Energetics of Syntrophic Cooperation in Methanogenic Degradation*. *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS* vol. 61 (1997).
22. Merino, N. *et al.* Living at the extremes: Extremophiles and the limits of life in a planetary context. *Front Microbiol* **10**, (2019).
23. Schut, G. J. *et al.* The role of geochemistry and energetics in the evolution of modern respiratory complexes from a proton-reducing ancestor. *Biochimica et Biophysica Acta - Bioenergetics* vol. 1857 958–970 Preprint at https://doi.org/10.1016/j.bbabio.2016.01.010 (2016).
24. Smith, E. T., Blarney, J. M., Zhou, Z. H. & Adams, M. W. W. *A Variable-Temperature Direct Electrochemical Study of Metalloproteins from Hyperthermophilic Microorganisms Involved in Hydrogen Production from Pyruvate+*. *Biochemistry* vol. 34 https://pubs.acs.org/sharingguidelines (1995).
25. Kanai, T. *et al.* Continuous hydrogen production by the hyperthermophilic archaeon, Thermococcus kodakaraensis KOD1. *J Biotechnol* **116**, 271–282 (2005).
26. Santangelo, T. J., Čuboňová, L. & Reeve, J. N. Deletion of alternative pathways for reductant recycling in Thermococcus kodakarensis increases hydrogen production. *Mol Microbiol* **81**, 897–911 (2011).
27. Krebs, M. P. & Khorana, H. G. *MINIREVIEW Mechanism of Light-Dependent Proton Translocation by Bacteriorhodopsin*. *JOURNAL OF BACTERIOLOGY* vol. 175 (1993).
28. Leahy, S. C. *et al.* The Genome Sequence of the Rumen Methanogen Methanobrevibacter ruminantium Reveals New Possibilities for Controlling Ruminant Methane Emissions. *PLoS One* **5**, (2013).
29. Tietze, M. *et al.* Redox potentials of methanophenazine and CoB-S-S-CoM, factors involved in electron transport in methanogenic archaea. *ChemBioChem* **4**, 333–335 (2003).
30. Costa, K. C. & Leigh, J. A. Metabolic versatility in methanogens. *Current Opinion in Biotechnology* vol. 29 70–75 Preprint at https://doi.org/10.1016/j.copbio.2014.02.012 (2014).
31. Laird, M. G., Adlung, N., Koivisto, J. J. & Scheller, S. Thiol-Disulfide Exchange Kinetics and Redox Potential of the Coenzyme M and Coenzyme B Heterodisulfide, an Electron Acceptor Coupled to Energy Conservation in Methanogenic Archaea. *ChemBioChem* **25**, 333–335 (2024).
32. Mayer, F., Lim, J. K., Langer, J. D., Kang, S. G. & Müller, V. Na+ transport by the A1AO-ATP synthase purified from Thermococcus onnurineus and reconstituted into liposomes. *Journal of Biological Chemistry* **290**, 6994–7002 (2015).
33. Deppenmeier, U. & Volker Müller, ·. Life Close to the Thermodynamic Limit: How Methanogenic Archaea Conserve Energy. (2007) doi:10.1007/400_2006_026/Published.
34. Kim, Y. J. *et al.* Formate-driven growth coupled with H2 production. *Nature* **467**, 352–355 (2010).
35. Yang, J. I., Lee, S. H., Ryu, J. Y., Lee, H. S. & Kang, S. G. A Novel NADP-Dependent Formate Dehydrogenase From the Hyperthermophilic Archaeon Thermococcus onnurineus NA1. *Front Microbiol* **13**, (2022).
36. Lipscomb, G. L. *et al.* Engineering hydrogen gas production from formate in a hyperthermophile by heterologous production of an 18-subunit membrane-bound complex. *Journal of Biological Chemistry* **289**, 2873–2879 (2014).
37. Zhang, Z. *et al.* A review of sulfate-reducing bacteria: Metabolism, influencing factors and application in wastewater treatment. *Journal of Cleaner Production* vol. 376 Preprint at https://doi.org/10.1016/j.jclepro.2022.134109 (2022).

38. Muyzer, G. & Stams, A. J. M. The ecology and biotechnology of sulphate-reducing bacteria. *Nature Reviews Microbiology* vol. 6 441–454 Preprint at https://doi.org/10.1038/nrmicro1892 (2008).
39. Mukohata, Y., Ihara, K., Tamura, T. & Sugiyama, Y. *Halobacterial Rhodopsins1*. *JB Review J. Biochem* vol. 125 (1999).
40. Atkinson, J. T., Campbell, I., Bennett, G. N. & Silberg, J. J. Cellular Assays for Ferredoxins: A Strategy for Understanding Electron Flow through Protein Carriers That Link Metabolic Pathways. *Biochemistry* **55**, 7047–7064 (2016).
41. Burkhart, B. W., Febvre, H. P. & Santangelo, T. J. Distinct physiological roles of the three ferredoxins encoded in the hyperthermophilic archaeon thermococcus kodakarensis. *mBio* **10**, (2019).
42. G6mez-Moreno, C. *et al. Protein-Protein Imeraction in E|ectron Transfer Reactions: The Ferredoxhfffiavodoxirdferredoxin:NADP + Reductase System from Anabaena*.
43. Page, C. C., Moser, C. C., Chen, X. & Dutton, P. L. *Natural Engineering Principles of Electron Tunnelling in Biological Oxidation±reduction*. *NATURE* vol. 402 www.nature.com (1999).
44. Yu, H. *et al.* Structure of the respiratory MBS complex reveals iron-sulfur cluster catalyzed sulfane sulfur reduction in ancient life. *Nat Commun* **11**, (2020).
45. Valentine, D. L. Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat Rev Microbiol* **5**, 316–323 (2007).
46. Schut, G. J., Boyd, E. S., Peters, J. W. & Adams, M. W. W. The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiology Reviews* vol. 37 182–203 Preprint at https://doi.org/10.1111/j.1574-6976.2012.00346.x (2013).
47. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nature Reviews Microbiology* vol. 15 711–723 Preprint at https://doi.org/10.1038/nrmicro.2017.133 (2017).
48. Woese, C. R. & Fox, G. E. *Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms (Archaebacteria/Eubacteria/Urkaryote/16S Ribosomal RNA/Molecular Phylogeny)*. vol. 74 (1977).
49. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc Natl Acad Sci U S A* **74**, 5088–5090 (1977).
50. Woese, C. R., Kandlert, O. & Wheelis, M. L. *Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya (Euryarchaeota/Crenarchaeota/Kingdom/Evolution)*. *Proc. Nati. Acad. Sci. USA* vol. 87 (1990).
51. Hulcr, J. *et al.* A Jungle in There: Bacteria in Belly Buttons are Highly Diverse, but Predictable. *PLoS One* **7**, (2012).
52. Reed, C. J., Lewis, H., Trejo, E., Winston, V. & Evilia, C. Protein adaptations in archaeal extremophiles. *Archaea* **2013**, (2013).
53. Sato, T. & Atomi, H. Novel metabolic pathways in Archaea. *Current Opinion in Microbiology* vol. 14 307–314 Preprint at https://doi.org/10.1016/j.mib.2011.04.014 (2011).
54. Atomi, H. & Reeve, J. Microbe profile: Thermococcus kodakarensis: The modehyperthermophilic archaeon. *Microbiology (United Kingdom)* **165**, 1166–1168 (2019).
55. Atomi, H., Fukui, T., Kanai, T., Morikawa, M. & Imanaka, T. Description of Thermococcus kodakaraensis sp. nov., a well studied hyperthermophilic archaeon previously reported as Pyrococcus sp. KOD1. *Archaea* **1**, 263–267 (2004).
56. Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T. & Imanaka, T. Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic Pyrococcus sp. *Appl Environ Microbiol* **60**, 4559–4566 (1994).

57. Mukund, S. & Adams, M. W. W. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon Pyrococcus furiosus. *Journal of Biological Chemistry* **270**, 8389–8392 (1995).
58. Schut~, G. J., Menon, A. L. & Adams, M. W. W. *ENZYMES OF CENTRAL METABOLISM [12]*. (2001).
59. Schicho, R. N., Ma, K., Adams, M. W. W. & Kelly3, R. M. *Bioenergetics of Sulfur Reduction in the Hyperthermophilic Archaeon Pyrococcus Furiosus*. *JOURNAL OF BACTERIOLOGY* vol. 175 https://journals.asm.org/journal/jb (1993).
60. Lipscomb, G. L. *et al.* SurR: A transcriptional activator and repressor controlling hydrogen and elemental sulphur metabolism in Pyrococcus furiosus. *Mol Microbiol* **71**, 332–349 (2009).
61. Yang, H. *et al.* SurR regulates hydrogen production in Pyrococcus furiosus by a sulfur-dependent redox switch. *Mol Microbiol* **77**, 1111–1122 (2010).
62. Amend, J. P. & Shock, E. L. Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol Rev* **25**, 175–243 (2001).
63. Nzuza, N. *et al.* Diversification of ferredoxins across living organisms. *Curr Issues Mol Biol* **43**, 1374–1390 (2021).
64. Fiala, G. & Stetter, K. O. *Pyrococcus Furiosus Sp. Nov. Represents a Novel Genus of Marine Heterotrophic Archaebacteria Growing Optimally at 100~*. *Arch Microbiol* vol. 145 (1986).
65. Hug, L. A. *et al.* A new view of the tree of life. *Nat Microbiol* **1**, (2016).
66. Sato, T., Fukui, T., Atomi, H. & Imanaka, T. Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. *J Bacteriol* **185**, 210–220 (2003).
67. Hileman, T. H. & Santangelo, T. J. Genetics techniques for Thermococcus kodakarensis. *Front Microbiol* **3**, 1–12 (2012).
68. Farkas, J. A., Picking, J. W. & Santangelo, T. J. Genetic techniques for the Archaea. *Annual Review of Genetics* vol. 47 539–561 Preprint at https://doi.org/10.1146/annurev-genet-111212-133225 (2013).
69. Santangelo, T. J., Čuboňová, L. & Reeve, J. N. Thermococcus kodakarensis Genetics: Tk1827-encoded β-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl Environ Microbiol* **76**, 1044–1052 (2010).
70. Scott, K. A., Williams, S. A. & Santangelo, T. J. Thermococcus kodakarensis provides a versatile hyperthermophilic archaeal platform for protein expression. in *Methods in Enzymology* vol. 659 243–273 (Academic Press Inc., 2021).
71. Blarney, J. M. & Adams, M. W. W. *Purification and Characterization of Pyruvate Ferredoxin Oxidoreductase from the Hyperthermophilic Archaeon Pyrococcus Furiosus*. *Biochimica et Biophysica Acta* (1993).
72. Smith, E. T., Blarney, J. M. & Adams, M. W. W. Pyruvate Ferredoxin Oxidoreductases of the Hyperthermophilic Archaeon, Pyrococcus furiosus, and the Hyperthermophilic Bacterium, Thermotoga marítima, Have Different Catalytic Mechanisms'* 1".
73. Mai, X. & Adams, M. W. W. Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon Pyrococcus furiosus. A new enzyme involved in peptide fermentation. *Journal of Biological Chemistry* **269**, 16726–16732 (1994).
74. Ma, K. & Adams, M. W. W. *Sulfide Dehydrogenase from the Hyperthermophilic Archaeon Pyrococcus Furiosus: A New Multifunctional Enzyme Involved in the Reduction of Elemental Sulfur*. *JOURNAL OF BACrERIOLOGY* vol. 176 https://journals.asm.org/journal/jb (1994).
75. Adams, M. W. W. & Kletzin', A. *OXIDOREDUCTASE-TYPE ENZYMES AND REDOX PROTEINS INVOLVED IN FERMENTATIVE METABOLISMS OF HYPERTHERMOPHILIC ARCHAEA*. (1996).

76. Mai, X. & Adams, M. W. W. *Characterization of a Fourth Type of 2-Keto Acid-Oxidizing Enzyme from a Hyperthermophilic Archaeon: 2-Ketoglutarate Ferredoxin Oxidoreductase from Thermococcus Litoralis*. *JOURNAL OF BACTERIOLOGY* vol. 178 https://journals.asm.org/journal/jb (1996).
77. Hutchins, A. M., Mai, X. & Adams, M. W. W. *Acetyl-CoA Synthetases I and II from Pyrococcus Furiosus*. (2001).
78. Schut, G. J., Menon, A. L. & Adams, M. W. W. *2-Keto Acid Oxidoreductases from Pyrococcus and Thermococcus Litoralis*. (2001).
79. Matsubara, K., Yokooji, Y., Atomi, H. & Imanaka, T. Biochemical and genetic characterization of the three metabolic routes in Thermococcus kodakarensis linking glyceraldehyde 3-phosphate and 3-phosphoglycerate. *Mol Microbiol* **81**, 1300–1312 (2011).
80. Clegg, C. & Hayes, D. Identification of Neighbouring Proteins in the Ribosomes of Escherichia coli A Topographical Study with the Cross-Linking Reagent Dimethyl Suberimidate. *Eur J Biochem* **42**, 21–28 (1974).
81. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246 (1989).
82. Elhabashy, H., Merino, F., Alva, V., Kohlbacher, O. & Lupas, A. N. Exploring protein-protein interactions at the proteome level. *Structure* vol. 30 462–475 Preprint at https://doi.org/10.1016/j.str.2022.02.004 (2022).
83. Cialek, C. A. *et al.* Imaging translational control by Argonaute with single-molecule resolution in live cells. *Nat Commun* **13**, (2022).
84. 2023_Stasevitch_LiveCellImaging_SaxtonSA2023.
85. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. *Biomolecular Condensates: Organizers of Cellular Biochemistry*.
86. Li, Z., Santangelo, T. J., Čuboňová, L., Reeve, J. N. & Kelman, Z. Affinity purification of an archaeal DNA replication protein network. *mBio* **1**, (2010).
87. Deckert, J. *et al.* Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* **26**, 5528–5543 (2006).
88. Lee, P. Y. & Low, T. Y. Identification and Quantification of Affinity-Purified Proteins with MaxQuant, Followed by the Discrimination of Nonspecific Interactions with the CRAPome Interface. in *Methods in Molecular Biology* vol. 2690 299–310 (Humana Press Inc., 2023).
89. Lee, C. M., Adamchek, C., Feke, A., Nusinow, D. A. & Gendron, J. M. Mapping protein–protein interactions using affinity purification and mass spectrometry. in *Methods in Molecular Biology* vol. 1610 231–249 (Humana Press Inc., 2017).
90. Rohila, J. S. *et al.* Protein-protein interactions of tandem affinity purified protein kinases from rice. *PLoS One* **4**, e6685 (2009).
91. Meyer, K. & Selbach, M. Quantitative affinity purification mass spectrometry: A versatile technology to study protein-protein interactions. *Frontiers in Genetics* vol. 6 Preprint at https://doi.org/10.3389/fgene.2015.00237 (2015).
92. Rohila, J. S. *et al.* Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. *Plant J* **46**, 1–13 (2006).
93. Michaelis, A. C. *et al.* The social and structural architecture of the yeast protein interactome. *Nature* (2023) doi:10.1038/s41586-023-06739-5.
94. Gehring, A., Sanders, T. & Santangelo, T. J. Markerless Gene Editing in the Hyperthermophilic Archaeon Thermococcus kodakarensis. *Bio Protoc* **7**, (2017).
95. Jain, S., Caforio, A. & Driessen, A. J. M. Biosynthesis of archaeal membrane ether lipids. *Frontiers in Microbiology* vol. 5 Preprint at https://doi.org/10.3389/fmicb.2014.00641 (2014).

96. https://teachmephysiology.com/biochemistry/atp-production/electron-transport-chain/, Accessed September 29, 2023.
97. https://graconllc.com/what-is-hydropower-engineering/, Accessed April 27, 2024.

CHAPTER 2
THERMOCOCCUS KODAKARENSIS PROVIDES A VERSATILE HYPERTHERMOPHILIC ACHAEAL PLATFORM FOR PROTEIN EXPRESSION[1]

Summary

Hyperthermophiles, typically defined as organisms with growth optima ≥ 80˚C, are dominated by the Archaea. Proteins that support life at the extremes of temperatures often retain substantial biotechnological and commercial value, but the recombinant expression of individual hyperthermophilic proteins is commonly complicated in non-native mesophilic hosts due to differences in codon bias, intracellular solutes, and the requirement for accessory factors that aid in folding or deposition of metal centers within archaeal proteins. The development of versatile protein expression and facilitated protein purification systems in the model, genetically tractable, hyperthermophilic marine archaeon *Thermococcus kodakarensis* provides an attractive platform for protein expression within the hyperthermophiles. The assortment of *T. kodakarensis* genetic backgrounds and compatible selection markers allow iterative genetic manipulations that facilitate protein overexpression and expedite protein purifications. Expression vectors that stably replicate both in *T. kodakarensis* and *Escherichia coli* have been validated and permit high-level ectopic gene expression from a variety of controlled and constitutive promoters. Biologically relevant protein associations can be maintained during protein purifications to identify native protein partnerships and define protein interaction networks. *T. kodakarensis* thus provides a versatile platform for the expression and purification of thermostable proteins.

[1] Portions of this chapter were previously published as: Scott, K. A., Williams, S. A. & Santangelo, T. J. Thermococcus kodakarensis provides a versatile hyperthermophilic archaeal platform for protein expression. in *Methods in Enzymology* vol. 659 243–273 (Academic Press Inc., 2021).

2.1 Introduction

Well-regulated and robust bacterial- and fungal-based recombinant gene expression systems are commonly employed to overproduce and, ultimately, purify proteins in large quantities for biochemical and structural characterization.[1,2] In choosing a recombinant expression system, parameters such as cost, yield and ease of purification are considered. Additional parameters, including the assembly of metal-cluster binding sites, proper post-translational modifications,[3] introduction of disulfide-linkages,[4] protein secretion,[5] post-translation N- and C-terminal processing, intein excision, the availability of specific chaperones,[6] unique genetic codes, insertion into membranes of specific compositions, and the availability of essential cofactors limit the choice of production host or demand overexpression in the native host.[7–9] Protein expression systems have been widely established and commercialized in several mesophilic species, including *Escherichia coli*, yeasts and insect cells (with baculoviruses). While these systems often permit production of cross-species, and occasionally cross-Domain recombinant protein expression, there remain instances wherein the canonical protein expression platforms fail to retain the necessary parameters for optimal protein expression. An increasing class of such proteins originate in the Archaea, wherein proteins often meant to survive the extremes of pressures, temperature, salinity, and pH have unique expression and folding requirements that are not easily replicated outside of the archaeal Domain.

The dearth of genetic systems for most archaeal lineages limits controlled protein overexpression in most archaeal clades, but a few systems have been established that permit the controlled overexpression of proteins in archaeal cell lysates or whole cells.[5,10–12] Archaeal-derived *in vitro* translation systems, often-linked to *in vitro* transcription systems,[13] provide a route to produce archaeal proteins in minute quantities but are impractical for moderate- or large-scale protein production. Many efforts have thus focused on the few genetically-tractable clades of Archaea to develop techniques to aid in protein overexpression and to facilitate rapid

and efficient purification of the desired protein products.[14] Controlled and inducible expression systems have long been established in halophilic (salt-loving) archaeal species,[14–17] but the near-saturating intracellular salinities often impair protein folding of non-halophilic proteins and impede downstream protein processing in almost all cases. Several methanogenic archaeal lineages have been manipulated to provide controlled and inducible protein expression platforms,[14,18,19] but none have been developed at scale, as the growth rates and complicated growth conditions of many methanogens limit the scale of culture growth. The remaining archaeal cell-based expression systems are found in species that grow at high (> 60˚C, thermophilic) or very high (> 80˚C, hyperthermophilic) temperatures. The thermophilic protein-expression systems established for the *Sulfolobales*[20,21] and the hyperthermophilic protein expression systems established for the *Thermococcales*[12,22–39] provide unique platforms for expression of thermostable proteins with a broad range of applications.

Hyperthermophilic archaea have significant potential as host systems to invigorate new biotechnological innovations and conversion chemistries.[25,39–44] Protein expression in hyperthermophilic archaeal systems motivates both basic and industrial efforts for greener chemistries, particularly towards the production of enzymes that survive the high temperatures common to many manufacturing processes.[45,46] Casual mesophilic, microbial contamination is reduced or effectively eliminated by carrying out enzymatic conversions at high temperatures. Increased thermostability of enzymes is necessary to maintain activity at higher temperatures that often increase processing rates and increased half-life of thermotolerant or thermostable enzymes reduces total enzyme requirements and costs.[47–49] Thermostable enzymes, particularly amylases and xylanases, are employed in bulk in biomass degradation and processing[48] as well as in pulp and fiber productions[50] and starch liquefaction processes[51]. Thermostable proteases are in high-demand for bulk addition to laundry detergents,[52,53] for leather processing,[54,55] as rennins for cheese-making,[56–59] as meat-tenderizers[60,61] and for baking and brewing platforms. Thermostable lipases are employed in high-temperature, low-

water-content solvent systems for hydrolysis activities, transesterification and ester synthesis. Thermostable enzymes are more diversely used in smaller scale molecular studies, including PCR,[62,63] next-generation sequencing techniques[64,65] and manipulations of nucleic acids. Finally, many thermostable proteins tightly fold and display reduced long-range intramolecular dynamics at ambient and/or mesophilic conditions, properties which often aid or inform structural or biophysical characterization of proteins.

Thermococcus kodakarensis has emerged as a model species for the study of biological processes at high temperatures and as an expression platform for thermostable enzymes.[23] *T. kodakarensis* is an ecologically pervasive, fast-growing, anaerobic, heterotroph first isolated from a solfatara on Kodakara island, Japan[66,67] that can be grown to high cell densities[68] with economical media requirements. The development of highly reproducible, accurate genomic and ectopic manipulation strategies for *T. kodakarensis* permits rational and iterative strain construction,[22,69,70] controlled protein overexpression,[5,10–12] and the introduction of protein tags[33,34] that aid in rapid and cost-effective protein purifications. *T. kodakarensis* thus provides a readily available and versatile platform for the expression of thermostable proteins for functional studies[14] and a relatively untapped source of thermostable enzymes. Here, we outline current methodologies for expression and purification of native and recombinant proteins which have been developed for *T. kodakarensis*. These procedures take advantage of the natural competency of *T. kodakarensis* cells, reliable culturing techniques and commercialized approaches for affinity chromatography. We first introduce the general procedures that permit genetic manipulation of *T. kodakarensis*, then detail genomic-, then plasmid-based protein expression and purification systems.

2.2 Genetic systems for *Thermococcus kodakarensis*

The *Thermococcales* family is distributed globally and represents a well-studied branch within the *Euryarchaeota* that includes a growing list of species divided among three genera: the *Pyrococcus, Paleococcus,* and *Thermococcus*.[71] *T. kodakarensis* was initially classified as *Pyrococcus kodakaraensis* strain KOD1 but was reclassified as *T. kodakarensis* strain KOD1 after genome re-sequencing and phylogenetic analysis of the 16S ribosomal RNA (note that the original species designation was also changed from *kodakaraensis* to *kodakarensis*).[72] *T. kodakarensis* is naturally competent and, through homologous recombination, readily integrates exogenous DNA into the genome.[69,73] The natural competency of *T. kodakarensis* contributes to its utility as a platform for protein expression. Transformation procedures are not arduous, and clonal populations can be selected and maintained on a variety of media using a range of selectable markers.[14] Specialized strains (Table 2.1) permit selections based on the restoration of prototrophy[22,69,70,74] and resistance to toxic nucleotide analogs[22,69] (Table 2.2). Complementary counter-selective pressures permit unlimited reuse of selective markers and iterative strain constructions, inclusive of both genomic modifications and plasmid retention.[28,69] While *T. kodakarensis* is not known to harbor any naturally-occurring plasmids,[75–77] cryptic mini plasmids from related members of the *Thermococcales* have been modified to stably replicate in *T. kodakarensis*.[12,35] The development of genetic techniques has led to the emergence of *T. kodakarensis* as a versatile host for protein expression.

Table 2.1: *T. kodakarensis* strains and compatible selection strategies.

Strain	Genotype	Medium requirements	Available selection strategies	Reference(s)
KOD1	wild type	rich or minimal medium	HMG-CoA reductase and citrulline.	Morikawa *et al*. 1994, *Fukui et al*. 2005
KU216	Δ*pyrF*	rich or minimal medium with uracil	HMG-CoA reductase and citrulline. Transformation with DNA encoding *pyrF* confers uracil prototrophy.	Sato *et al*. 2005
KUW1	Δ*pyrF;* Δ*trpE*	minimal medium with uracil and tryptophan	HMG-CoA reductase and citrulline. Transformation with DNA encoding *pyrF* or *trpE* confers uracil or tryptophan prototrophy, respectively.	Sato *et al.* 2005
KUWH1	Δ*pyrF;* Δ*trpE;* Δ*hisD*	minimal medium with uracil, tryptophan and histidine	HMG-CoA reductase and citrulline. Transformation with DNA encoding *pyrF*, *trpE*, or *hisD* confers uracil, tryptophan, or histidine prototrophy, respectively.	Sato *et al.* 2005
KW128	Δ*pyrF;* Δ*trpE::pyrF*	rich medium or minimal medium with tryptophan	HMG-CoA reductase and citrulline. Compatible with MAE using tryptophan (selection) and 5-FOA (counter-selection).	Sato *et al*. 2005
TS517	Δ*pyrF;* Δ*trpE::pyrF;* Δ*HPRT*	rich medium or minimal medium with tryptophan	HMG-CoA reductase and citrulline. Compatible with MAE using tryptophan (selection) and 6MP (counter-selection).	Santangelo *et al.* 2010
TS559	Δ*pyrF;* Δ*trpE::pyrF;* Δ*pdaD;* Δ*HPRT*	rich medium with agmatine or minimal medium with tryptophan and agmatine	HMG-CoA reductase and citrulline. Compatible with MAE using agmatine (selection) and 6MP (counter-selection).	Santangelo *et al*. 2010
TS900	Δ*pyrF;* Δ*trpE;* Δ*pdaD;* Δ*HPRT*	rich medium with agmatine or minimal medium with tryptophan and agmatine	HMG-CoA reductase and citrulline. Compatible with MAE using (i) agmatine (selection) and 6MP (counter-selection); (ii) uracil (selection) and 5-FOA (counter-selection).	Unpublished

Table 2.2: Selectable markers for *T. kodakarensis.*

Selection	Genetic background	Selection requirements	Compatible strains	Reference(s)
tryptophan	*ΔtrpE* (ΔTK0254)	Minimal medium lacking tryptophan.	KUW1, KUWH1, KW128	Sato *et al*. 2003, Sato *et al.* 2005
simvastatin, mevinolin	None	None. Requires overexpression of HMG-CoA reductase (PF1848).	Any strain	Matsumi *et al.* 2007, Santangelo *et al*. 2008
agmatine	Δ*pdaD* (ΔTK0149)	Minimal or rich medium lacking agmatine supplementation.	TS559, TS900	Fukuda *et al*. 2008; Santangelo *et al.* 2008
histidine	Δ*hisD* (ΔTK0244)	Minimal medium lacking histidine.	KUWH1	Sato *et al.* 2005
6-methyl purine (6MP)	Δ*HPRT* (ΔTK0664)	Minimal medium containing 6MP.	TS517, TS559, TS900	Santangelo *et al.* 2010
uracil, 5-fluororatic acid (5-FOA)	*ΔpyrF* (ΔTK2276)	Minimal medium lacking uracil and supplemented with 5-FOA, respectively.	KU216, KW128, TS517, TS559, TS900	Sato *et al.* 2003
citrulline, aspartate	+*argG* (PF0207), +*argH* (PF0208)	Minimal medium lacking citrulline or aspartate.	Any strain	Atomi *et al*. 2004

2.2.1 Genetic backgrounds and selectable markers

The choice of selection strategy and host genotype dictate initial plans for protein expression and purification within *T. kodakarensis*. When using *T. kodakarensis* strains with WT or near WT genotypes, only two validated selection markers exist that do not require specialized genotypic backgrounds. The first is based on hydroxymethylglutaryl coenzyme A (HMG-CoA) reductase activity that is essential for archaeal isoprenoid-based lipid biosynthesis.[78] HMG-CoA is Inhibited with the statins simvastatin or mevinolin (a simvastatin analog),[79] and overexpression of the *P. furiosus* HMG-CoA reductase *via* a constitutive promoter provides *T. kodakarensis* strains resistance to ~8 μM simvastatin[79] or >30 μM mevinolin[80]. The second universal selective pressure is based on amino acid auxotrophies. Unlike many other *Thermococcus* species, *T. kodakarensis* lacks an arginine biosynthetic pathway and requires

arginine supplementation for growth.[67,74] The *P. furiosus* genes PF0207 and PF0208 encode argininosuccinate synthase and argininosuccinate lyase, respectively, which together catalyze the synthesis of arginine from citrulline or aspartate.[14,81] *T. kodakarensis* strains that incorporate and express donor DNA carrying both PF0207 and PF0208 can grow with citrulline or aspartate in the absence of arginine in minimal medium.[81] These selection markers are compatible with all known strains; however, counter selection strategies independent of genotypic background have not yet been developed.

It is far more common to direct protein expression in one of several strains (Table 2.1) that have been constructed to allow auxotrophic–prototrophic selection of genomic integration of donor DNA by homologous recombination or retention of autonomously replicating plasmids.[12,31,35] Auxotrophic strains that incorporate donor DNA restoring prototrophy provide positive selection strategies. KU216 lacks the orotidine-5′ -phosphate decarboxylase (*pyrF)* gene and displays uracil auxotrophy in minimal medium.[70] KU216 also confers resistance to 5-fluoroorotic acid (5-FOA), allowing counter selection for cells that have excised *pyrF* encoded donor DNA. Building from KU216, stains KUW1 and KUWH1 additionally lack anthranilate synthase (TK0254), encoded by the *trpE* locus, conferring tryptophan auxotrophy.[70] KUWH1 further lacks *hisD,* providing histidine auxotrophy. Strain KW128 lacks *pyrF* and the *trpE* locus is replaced with *pyrF*, leading to tryptophan auxotrophy.[70] The absence of the amino acid corresponding to the induced auxotrophy provides a positive selection strategy based on the restoration of prototrophies provided in the donor DNA.

Multiple strains derived from KU216 were constructed to allow for sequential selection and counter selection of genomically integrated DNA. Such strains are useful for genetic manipulations *via* markerless allelic exchange (MAE) and are the backbone of genetic techniques in *T. kodakarensis*. Strain TS517 is genotypically similar to KW128 but with an additional deletion of hypoxanthine/guanine phosphoribosyltransferase (TK0664), a purine

scavenging protein.[22] Strain TS559 additionally lacks the pyruvoyl-dependent arginine decarboxylase (TK0149) which leads to agmatine auxotrophy.[22] Integration of donor DNA encoding TK0149 results in agmatine prototrophy. Similar to TS517, ejection of donor DNA carrying TK0664 is selected for by isolating mutants resistant to 6-methylpurine (6MP).[31] Lastly, strain TS900 was derived from TS559 and is cleanly deleted for *pyrF* (unpublished). Selection and counter selection of transformants in TS900 are identical to that of TS559. The strains described here and their compatible selection strategies allow targeted mutagenesis of chromosomal DNA.

2.2.2 Targeted mutagenesis of chromosomal DNA

T. kodakarensis is naturally competent for DNA uptake and incorporates donor DNA into its genome by homologous recombination,[69,70] allowing targeted chromosomal mutations for increased protein expression,[5] placement of extracellular secretion signals,[5] gene deletions,[82–85] epitope- or affinity-purification,[33,34] or allelic substitutions.[86] Genome manipulations are currently possible using MAE where transformation with plasmid DNAs that carry a positive selection marker (*i.e.*: TK0254 (*trpE*), TK2276 (*pyrF*) or TK0149 (*pdaD*)) are integrated into the genome resulting in tryptophan, uracil, or agmatine prototrophic transformants, respectively (Figure 2.1).[22,31,70] After selection and confirmation of the intermediate genotype, growth of confirmed intermediate-genome containing cells in the presence of tryptophan, uracil, or agmatine permits a small percentage of cells to spontaneously remove the integrated plasmid sequences *via* homologous recombination.[69] These unique cells containing the desired final genotypes are now able to survive in the presence of 5-FOA or 6MP due to the loss of the counter-selectable marker (the activity of which will result in death of all cells still containing the plasmid-encoded counter-selectable marker).[69,80] MAE, sometimes referred to as pop-in/pop-out, permits a tailored approach to expression and purification of proteins in a hyperthermophilic system.

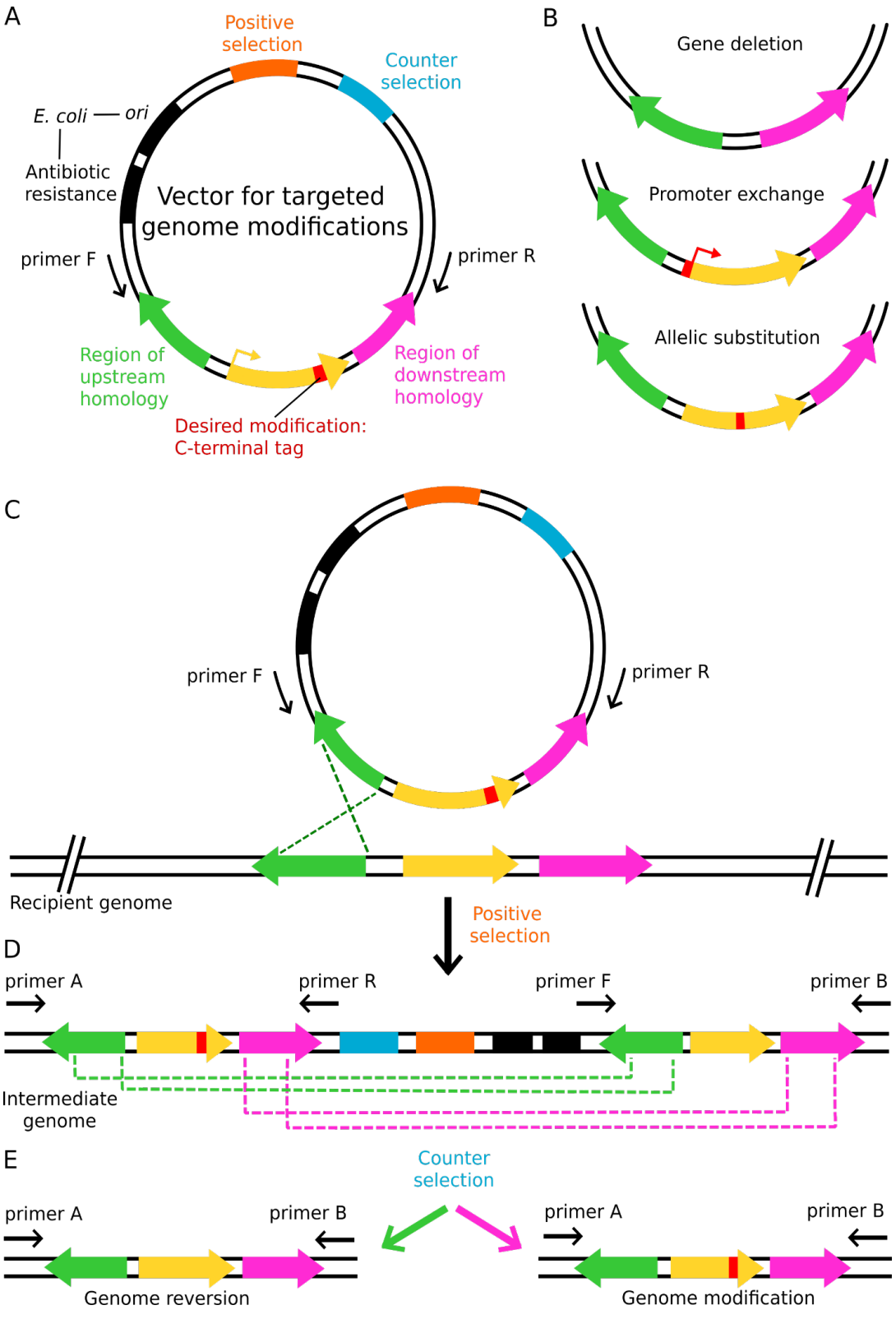


Figure 2.1: Sequential homologous recombination events permit modifications of chromosomal DNA. A vector designed for genomic integration (A) includes replication and selection elements for *E. coli* (black) but only selection (orange) and counter selection (blue)

elements for *T. kodakarensis*. Regions of up (green) and downstream (pink) homology to the target locus sequences (yellow, with red modification) permit the vector to integrate adjacent to the target locus. Possible genomic modifications including (B) gene deletion(s), promoter exchange and allelic substitutions can be generated using identical pop-in/pop-out techniques. Genomic integration is mediated by either the region of upstream or downstream homology (C), and the first recombination event (pop-in) is selected for using a positive selection marker (D). In this example, the region with upstream homology (green) is the recombination point. The intermediate genome is verified using primers pairs that have homology to the genome and the plasmid separately (primers A and F and primers R and B). The second recombination event, wherein the plasmid is ejected from the genome (pop-out), is selected for using a counter selectable marker (E); there is an equal probability of recombination at the green or pink loci. In this example, recombination between the green loci yields the parent genome while recombination between the pink loci results in the desired modification. Depending on the modification, the final genome can be verified using PCR and amplicon sequencing or whole genome sequencing.

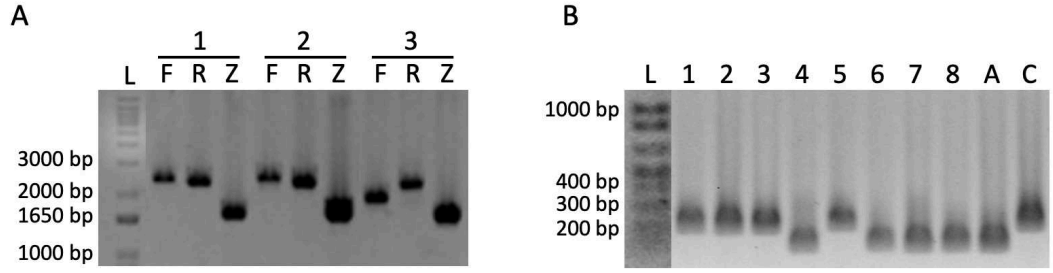


Figure 2.2: The presence and size of PCR generated amplicons provides confidence in the genomic modifications retained in intermediate and final strains of *T. kodakarensis*. A) Agarose gel electrophoresis separates amplicons, visualized with ethidium bromide staining, generated with distinct primer pairs using genomic DNA isolated from four (1-4) strains of *T. kodakarensis* that were selected as intermediate strains. Lanes denoted by an F employ primers A and F from Figure 2.1; lanes denoted by an R employ primers B and R from Figure 2.1. The lanes denoted Z use a pair of primers complementary to a region of the genome far from the site of predicted plasmid incorporation; amplification of the ~1700 bp product yields confidence in the quality of the DNA preparation from each strain. The amplicons generated within lanes F and R are consistently sized when DNA from strains 1 and 2 is used as the DNA templates, suggesting that each of these strains represents the same intermediate genome configuration. In contrast, the distinctly different amplicon sizes generated from the same primer pairs when DNA from strain 3 is used as a template support a different intermediate genome configuration of strain 3 versus strains 1 or 2. Strains 1 and 2 indicate a larger amplicon for the F reaction, suggesting the crossover event occurred as demonstrated in Figure 2.1D. Strain 3 generates a smaller F amplicon suggesting that the crossover event occurred downstream of the modification as would occur if the crossover event occurred in the pink region depicted in Figure 2.1C (note that the resultant intermediate genome from a downstream crossover event is not shown in Figure 2.1D). B) Agarose gel electrophoresis separates amplicons, visualized with ethidium bromide staining, generated with primer pairs bracketing the genomic locus that was targeted for insertion of 45 bp encoding epitope and affinity tags. Lanes 1-8 include amplicons generated with DNAs purified from *T. kodakarensis* strains following excision of the integrated plasmid (so called “final” strains). Lanes A and C include amplicons generated using plasmid DNA as the template, with either the A-plasmid encoding the unmodified *T. kodakarensis* gene sequence or the C-plasmid encoding the modified *T. kodakarensis* gene sequence and tag serving as controls. Amplification of a small length (~150 vs ~200 bp) amplicon visualized in a 2% agarose gel can identify the presence (~200 bp) or absence (~150 bp) of the 45 bp

sequences encoding the HA and His_6 tag. Larger bands observed in lanes 1-3, 5, and C identify the presence of the modified sequence with the 45 bp insertion, while a smaller band in lanes 4, 6-8, and A is indicative of the wildtype genotype. A 1 Kb plus DNA ladder is used to confirm amplicon size (Invitrogen, 10787026).

Targeted chromosomal alterations are achieved by transforming a vector containing the desired modification flanked by several hundred bases of up- and downstream homology (Figure 2.1A).[31] Routine modifications including N- or C-terminal epitope- or affinity-tag additions, promoter exchange, gene deletions and allelic substitutions (Figure 2.1A, B) are encoded on a vector that has both a selectable and counter selectable marker compatible with MAE in *T. kodakarensis* (Figure 2.1A and Table 2.2). Components for autonomous replication and selection in *E. coli* are included; however, no *T. kodakarensis* origin of replication (*ori*) is necessary, as the plasmid is expected to integrate into the genome. After vector transformation, either the up- or downstream regions homologous to the target loci will integrate into the genome (Figure 2.1C). Prototrophic selection of plasmid integration (pop-in) *via* homologous recombination is mediated by the positive selectable marker (Table 2.2). The intermediate genome - inclusive of the plasmid and modified gene - is confirmed *via* PCR using two primer sets (primers A/R and primers F/B, Figure 2.1D, 2.2A). Primers A and B are external to the site of plasmid integration and are not homologous to any region on the integrative vector, while primers F and R have homology to the plasmid. Two intermediate genotypes are possible depending on if the recombination event occurs upstream or downstream of the modification. Figure 2.1D shows the intermediate genotype generated when integration of the plasmid occurs upstream of the modification. PCR using primers A and R will result in a larger amplicon in this situation as demonstrated in intermediate strains 1 and 2 in Figure 2.2A. If recombination of the plasmid into the genome occurs downstream of the modification, PCR using DNAs isolated from such strains with primers A and R will instead be smaller, while PCR using primers F and B will be larger as observed in intermediate strain 3 (Figure 2.2A).

Spontaneous recombination of the plasmid out of the genome (pop-out) is selected for using a counter selectable marker (Figure 2.1E and Table 2.2) and has a ~50% chance of reverting to the parent genome (Figure 2.1E, green; Figure 2.2B, lanes 4, 6-8) or including the modification (Figure 2.1E, pink; Figure 2.2B, lanes 1-3, 5) assuming the genomic modification incurs a neutral effect on fitness. Confirmation of the final genotype can be achieved through PCR analysis and/or Sanger sequencing of the target loci (Figure 2.2B).[28] Sequencing is recommended for small modifications (i.e. single nucleotide polymorphisms), while larger modifications can be visualized via differences in the sizes of PCR amplicons. For example, incorporation of sequence encoding both an HA and epitope tag necessitates addition of 45 bp to the genome. Such a small insertion is at the lower limit for visualization via PCR amplification and agarose gel electrophoresis. Figure 2.2B demonstrates amplification of a small sequence (<300 bp) wherein a 45 bp difference can be observed by running a dense gel at a low voltage for an increased length of time to resolve such a small difference in size (*e.g.* 2% agarose gel, ran at 80V, for 2.5 hours). As predicted, 50% of transformants reverted to wildtype indicated by the smaller band in lanes 4 and 6-8, while 50% maintained the modification in the genome indicated by the larger band in lanes 1-3 and 5 (Figure 2.2B). Whole genome sequencing should always be considered, as off target recombination events are possible and random mutational variations will likely exist. While modification of the chromosomal coding DNA sequence retains native protein levels, in instances where high level protein expression is desired, ectopic shuttle vectors which stably replicate in *T. kodakarensis* and *E. coli* are often more practical.

2.3 Autonomously replicating plasmids

Plasmids that autonomously replicate and express selectable markers in both *T. kodakaraensis* and *E. coli* can be easily manipulated and confirmed in *E. coli*, then transferred to *T. kodakaraensis* where manipulation of plasmids is more complicated. Shuttle vectors, when

modified to encode expression cassettes under the control of a variety of constitutive promoters (Table 2.3 and Figure 2.3, purple), provide a mechanism for high levels of ectopic expression. The cell surface glycoprotein (*csg*, TK0895) is a highly expressed gene in *T. kodakarensis*, and, therefore, the P_{csg} promoter, among others, can direct high level expression when inserted into the multiple cloning site (Figure 2.3, green).[87] Where an exogenous promoter is preferred, the P_{hmtB} promoter from *Methanothermobacter thermautotrophicus* is compatible with *T. kodakarensis* expression machinery.[80] In the related species, *Pyrococcus furiosus,* some level of inducible expression can be achieved from the P_{cipA} promoter.[88] With the exception of a riboswitch-mediated expression system, no gene-controlled expression system currently exists in *T. kodakarensis*.[32] The endogenous P_{fbpase} promoter can be activated by the addition of elemental sulfur (S˚) but, like the P_{cipA} promoter, such large scale changes to growth conditions alter the expression of many genes.[89] Select changes to the ribosome binding site (RBS, 5′ - AGGTGA) can act as another point to control protein expression levels.[12]

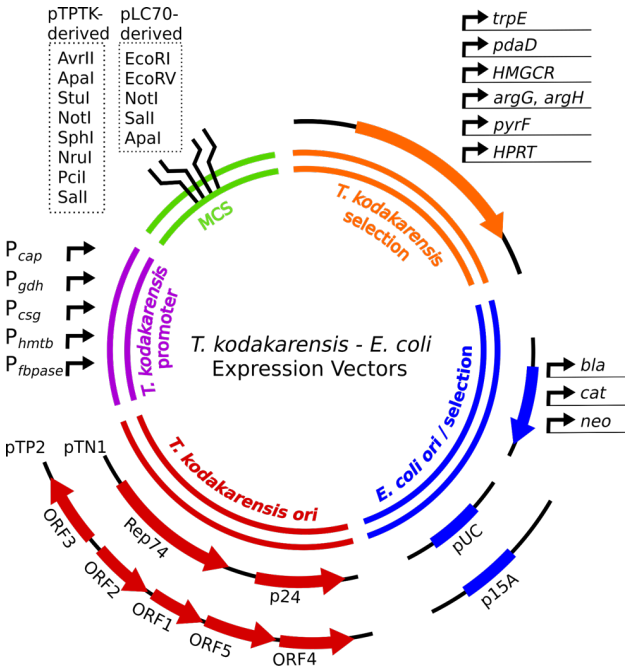


Figure 2.3: Autonomously replicating expression vectors permit stable replication and gene expression in *T. kodakarensis* and *E. coli*. Constitutive or inducible promoters (purple) drive high levels of ectopic gene expression of sequences cloned into the multiple cloning site (MCS, green). Selections in *T. kodakarensis* (orange) rely on restorations of prototrophies (agmatine, tryptophan, arginine, uracil), statin resistance (mevinolin, simvastatin) or resistance

to toxic nucleoside analogues (6MP, 5-FOA). Expression vectors retain either a pUC or p15A origin (ori) and an antibiotic selection cassette(s) for replication in *E. coli* (blue): bla (ampicillin resistance); cat (chloramphenicol resistance); neo (kanamycin resistance). The entire pTN1 or pTP2 plasmid sequences (red) permit autonomous replication in *T. kodakarensis.*

Table 2.3: Promoters compatible with gene expression in *T. kodakarensis.*

Promoter	Gene ID	Origin	Description	Reference(s)
P_{hmtB}	TM0254	*Methanothermobacter thermautotrophicus*	Constitutive; Histone B.	Santangelo *et al*. 2007
P_{cap}	TK2279	*Thermococcus kodakarensis*	Constitutive; CDP-alcohol phosphatidyltransferase.	Rodrigues *et al*. 2007
P_{gdh}	TK1431	*Thermococcus kodakarensis*	Constitutive; Glutamate dehydrogenase.	Santangelo *et al*. 2008
P_{csg}	TK0895	*Thermococcus kodakarensis*	Constitutive; Cell surface glycoprotein.	Takemasa *et al*. 2011
P_{fbpase}	TK2164	*Thermococcus kodakarensis*	Inducible; Fructose-1,6-bisphosphatase. Involvement in central metabolism hinders tight regulation.	Sato *et al*. 2004
P_{cipA}	PF0190	*Pyrococcus furiosus*	Inducible; membrane-bound glycoprotein, cold-induced protein A. Induction *via* heat/cold shock or other stress has significant biological effects.	Basen *et al*. 2012

No naturally occurring plasmids exist in *T. kodakarensis*.[75] All *T. kodakarensis-E. coli* shuttle vectors (Table 2.4) rely on one of two plasmids isolated from close relatives in the *Thermococcales* for replication (Figure 2.3, red). The cryptic mini plasmid pTN1 isolated from *T. nautilus* encodes a rolling‐circle replication initiator protein (Rep74) and a p24 protein, although p24 is not essential.[76] The fusion of pTN1 to *E. coli* vectors provides the minimal replicative machinery or both species and selectable markers for *E. coli*, but not *T. kodakarensis*.[12] Analogous to pTN1, pTP2 from *Thermococcus prieurii* contains five putative open reading frames, one of which encodes a rolling circle replication gene that also supports plasmid replication in *T. kodakarensis.*[35,90] Both pTN1- and pTP2-based plasmids can be maintained within the same cell, providing two distinct and simultaneous means of selection. *T.*

kodakarensis is polyploid, retaining ~7-19 genome copies.[91] pTN1-based plasmids have been measured to exist at a copy number of ~3 copies per chromosome[12] while pTP2-based plasmids have been observed at ~50 copies per cell[90] (Table 2.4). Stable replication of plasmids in *E. coli* is dependent on a pUC or p15A origin (Figure 2.3, blue), with a copy number of ~500-700 and ~20-30 per cell, respectively (Table 4).[12,35,92,93]

Naturally occurring replicative plasmids (*i.e.*: pTP2 and pTN1; Figure 2, red) have been modified to contain cassettes for expression of selection markers in both *T. kodakaraensis* and *E. coli* (Table 2.4). Ampicillin, chloramphenicol and kanamycin resistance genes (*bla*, *cat* and *neo*, respectively; Figure 2.3, blue) have previously been used for selection in *E. coli*, while a variety of prototrophic and statin-resistance strategies have been validated in *T. kodakarensis* (Figure 2.3, orange).[12,35] The pLC plasmid series (*e.g.*: pLC64, pLC70, pLC71, and pTS543) can provide statin resistance or restore tryptophan or agmatine protoprophy to transformants.[10,12,22] The pLC64 plasmid was constructed by ligating together pTN1 and the commercially available pCR2.1-TOPO. pLC64 contains the trpE gene under control of the constitutive promoter, P_{cap}.[11] Plasmids pLC70 and then pLC71 were derived by adding HMG-CoA from *P. furiosus* under the *T. kodakarensis* P_{gdh} promoter, providing an additional mechanism for statin resistance. Several restriction enzyme sites are available for digestion at the MCS (Figure 2.3, green).

The pTPTK plasmid series (*e.g.* pTPTK1, pTPTK2, pTPTK3) are a result of the ligation of pTP2 and pBAD33 and have been modified to enable statin-resistance or restore tryptophan or agmatine prototrophy under a native or constitutive promoter (Figure 2.3, purple).[35] Restriction enzyme recognition sites are present in the multiple cloning site; however, *Stu*I and *Sal*I digestion should be avoided in pTPTK2 as their recognition sequences are present in the *trpE* cassette. With the exception of agmatine prototrophy, all other prototrophic selection strategies require cells to be grown in minimal medium. Use of all prototrophic selective markers requires a mutant genetic background that induces auxotrophy for the selection nutrient.[22,69,70,80]

Table 2.4: *T. kodakarensis* - *E. coli* shuttle vectors.

Plasmid	Size (bp)	*T. kodakarensis* copy number	*E. coli* copy number	*T. kodakarensis* selection(s)	*E. coli* selection(s)	Description	Genetic background	Reference
pLC64	9034	~3/chromosome	~500-700/cell	tryptophan	ampicillin, kanamycin	pTN1 + pCR2.1-TOPO::P_{cap}-TrpE	Δ*trpE* (ΔTK0254)	Santangelo *et al*. 2008
pLC70	10495	~3/chromosome	~500-700/cell	statins, tryptophan	ampicillin, kanamycin	pTN1 + pCR2.1-TOPO::P_{cap}-TrpE; P_{gdh}-HMG-CoA	Δ*trpE* (ΔTK0254)	Santangelo *et al*. 2008
pLC71	9190	~3/chromosome	~500-700/cell	statins, tryptophan	ampicillin, kanamycin	pTN1::Δp24 + pCR2.1-TOPO::P_{cap}-TrpE; P_{gdh}-HMG-CoA	Δ*trpE* (ΔTK0254)	Santangelo *et al*. 2008
pTS543	8196	~3/chromosome	~500-700/cell	agmatine	ampicillin, kanamycin	pTN1 + pCR2.1-TOPO::P_{hmtB}-PdaD	Δ*pdaD* (ΔTK0149)	Santangelo *et al*. 2010
pTPTK1	5489	~50/cell	~20-30/cell	statins	chloramphenicol	pTP2 + pBAD33::P_{gdh}-HMG-CoA	Any strain	Catchpole *et al*. 2018
pTPTK2	5444	~50/cell	~20-30/cell	tryptophan	chloramphenicol	pTP2 + pBAD33::P_{cap}-TrpE	Δ*trpE* (ΔTK0254)	Catchpole *et al*. 2018
pTPTK3	4710	~50/cell	~20-30/cell	agmatine	chloramphenicol	pTP2 + pBAD33::P_{pdaD}-PdaD	*ΔpdaD (ΔTK0149)*	*Catchpole et al. 2018*

2.4 Purifying proteins from an archaeal host

Recombinant protein expression in a genetically accessible, native hyperthermophile such as *T. kodakarensis* provides access to protein production and research applications in a novel setting. Hyperthermophilic protein expression can produce thermostable proteins in their native state, with biologically relevant associations and modifications (*e.g.* in complex and/or containing post-translational modifications). Protein purifications *via* liquid chromatography (LC) can be applied i) to purify proteins at industrial scale (typically from ectopic vectors) or ii) to describe native-level protein expression or identify protein–protein partnerships at research scale.

2.4.1 Protein production and purification

Ectopic expression vectors permit easy introduction and expression of heterologous genes and are commonly used for applications involving large-scale protein production. *T. kodakarensis* cultures grow rapidly at 85˚C, doubling every ~40 minutes and achieve a final cell density of ~1 x 10^{8-9} cells/mL in rich media. Optical density (OD) at 600 nm provides a convenient method to monitor cell growth with cultures reaching ~1.0-1.3 absorbance units at maximum cell densities. *T. kodakarensis* strains harboring an expression vector(s) may be cultured at a range of temperatures (~65-95˚C) and in media containing the necessary nutritional and selectable components using batch fermentation to yield large quantities of biomass. Growth conditions (*e.g.* temperature) and media additives should be selected for the specific strain (Table 2.1) and the compatible genetic markers (Table 2.2). Industrial-scale protein purifications typically require greater culture volumes, and pilot trials at small scale are recommended for process development. Larger culture volumes (>10 L) are ideally grown within a bioreactor. Protein production in minimal medium is possible, although culture yields are dramatically reduced resulting in minimal biomass. Cultures that have reached late exponential

or early stationary growth phases are rapidly cooled and harvested, and the biomass is processed for protein purification. Cell pellets can be stored at -80˚C indefinitely prior to processing.

Cell lysis to separate cytoplasmic from membrane-bound or extracellular products is achieved by sonication in physiologically relevant buffers followed by centrifugation to pellet membranes, while retaining cytoplasmic proteins in the supernatant. Purification of the desired protein can be obtained by liquid chromatography, and conditions specific to the protein of interest should be considered (*e.g.* a protein bound strongly to DNA may require elution at higher salinity or unique pHs). Reverse phase, ion exchange, size exclusion or affinity chromatography can be employed individually or in tandem for a variety of purification techniques. Affinity purification using a divalent cation bound to a chelating column is a common method for purification using a His_6 tag which have a natural affinity for divalent cations such as Ni^{2+}, Co^{2+}, Cu^{2+}, and Zn^{2+} (Figure 2.4). Such a purification uses two buffers: Buffer A (10 mM Tris HCl pH 8.0, 500 mM NaCl, 10% w/v glycerol), often called the lysis buffer, that maintains salt and pH characteristics of the cytoplasm and functions to preserve native confirmations and interactions once the cell is lysed; Buffer B, often referred to as the elution buffer, mimics Buffer A but includes imidazole at 100-500 mM and a reduced salt concentration (100 mM NaCl) to maintain charge characteristics with the addition of the ionic solute. Harvested cell pellets are lysed in Buffer A and clarified by centrifugation to produce a clarified cell lysate (CCL). A charged chelating column is prepared by first binding the divalent cation of choice, then saturating the column with Buffer B (Figure 2.2C) before washing and equilibrating the column with Buffer A (Figure 2.2D). The CCL flows over a charged chelating column and proteins with affinity for the charged substrate will bind while the remaining cytoplasmic contents remain mobile and are collected in the flow through and wash (Figure 2.4E and F). Imidazole is the sidechain of histidine, and thus also has affinity for the charged cation. With increasing concentration of Buffer B using an isocratic or linear gradient (the latter of which is depicted in

Figure 2.4G), imidazole will outcompete binding of the His_6 tag, moving proteins that were originally bound and immobile into the mobile phase for collection. Collecting small (often 1-2 mL) fractions during the elution will capture proteins as they release from the column as they are out competed by an increasing imidazole concentration. Many organisms maintain proteins with metal centers, thus there will often be native (un-tagged) proteins that have a natural affinity for the charged column. Such proteins often elute mid-way through the gradient at a great enough concentration to be detected by UV_{280}, producing a visible peak on the chromatogram (Figure 2.4H). Often, a His_6 tagged protein will elute after the majority of proteins with natural affinity elute, appearing after the UV peak, but with no UV signature of its own at native expression levels. Elution of over-expressed proteins should generate a visible peak. The chelating column can be renewed by removal of any protein contaminates with 6M Guanadine HCl, removal of the divalent cation with treatment of EDTA (Figure 2.2A), and recharging the column with fresh metal cation, such as 100 mM $NiSO_4$ (Figure 2.2B).

Purification efficacy can be assessed by SDS-PAGE and Coomassie or silver staining (Figure 2.5A-C) or by Western blotting if respective antibodies are available (Figure 2.5D). The addition of an epitope tag, such as the hemagglutinin epitope (HA) to a protein of interest can be utilized for immunodetection if protein-specific antibodies are not available. It is recommended that epitope tags are encoded between the gene of interest and the His_6 tag sequence to allow for the greatest accessibility of the His_6 tag during affinity purification. Including a positive control, pellet, CCL, flow through (FT), and wash samples can aid in troubleshooting if the protein of interest is not purified as expected. If the protein of interest appears in the pellet, the protein has not been separated from the membrane sufficiently. The protein of interest should appear in the CCL, but if it appears in the FT or wash, the protein did not bind to the column. If the affinity tag sequence has been confirmed, the protein may not be expressed or the tag is cleaved, in which case generating an N-terminal tagged strain is advised.

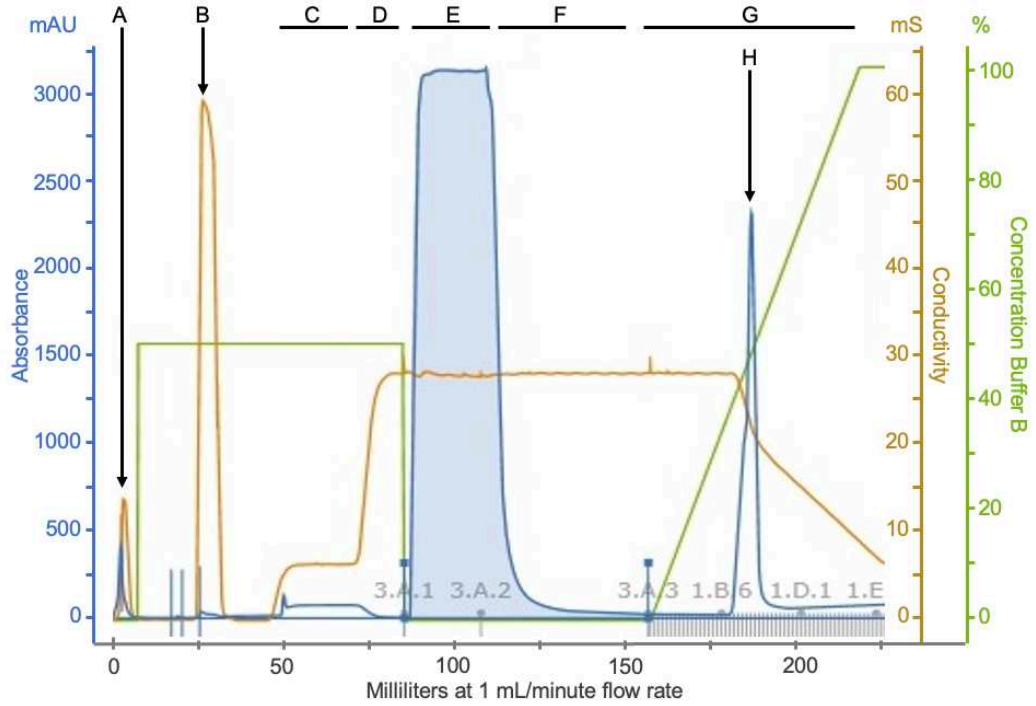


Figure 2.4: Affinity purification scheme. Affinity tagged proteins encoding His_6 sequences at the C-terminus are purified from a CCL using a metal-charged chelating column. Columns are prepared by assuring all protein is removed by flowing 2 column volumes of 6 M Guanidine HCl across the column (not shown on the chromatogram, but large spike in conductivity, orange, will be seen) and all metal is removed by flowing 2 column volumes of 100 mM EDTA (A) across the column. Columns are then charged with a divalent cation, of which Ni^{2+} is a common choice. 2 column volumes of 100 mM $NiSO_4$ sufficiently saturates the column matrix (B). The column is then prepared by flowing 5 column volumes of Buffer B (C) followed by 5 column volumes of Buffer A (D) before flowing the filtered, CCL across the column (E). In this scheme, Buffer B contains 200 mM imidazole and 100 mM NaCl which produces a UV_{280} signal of ~100 mAU (blue) and conductivity signal of ~5 mS (orange), respectively (C). Buffer A contains 0 mM imidazole and 500 mM NaCl, which is observed as the mAU returns to 0 and the conductivity rises to ~27 mS, respectively (D). Protein in the CCL that does not bind to the column (commonly referred to as the "flow through") is observed in the wide, flat UV_{280} peak >3,000 mAU and ~15 mLs are collected (fraction 3.A.1). The column is washed (F) with Buffer A and ~15-20 mLs are collected (fraction 3.A.2) until the UV_{280} signal returns to baseline. Proteins are eluted (G) in this scheme using a linear gradient of increasing Buffer B concentration (from 0-100%), however an isocratic or combination of isocratic and linear gradients can be employed. During elution, fifty-eight (58) 1 mL fractions are collected (fraction 3.A.3 – 1.E.12). Native proteins with a natural affinity for the Ni^{2+}-charged matrix are eluted en masse at ~50% Buffer B (~100 mM imidazole), however proteins which possess a His_6 tag are often of greater affinity, eluting at higher concentrations (55-80% Buffer B) but are often not sufficiently abundant to produce a notable UV_{280} signal. The tagged protein of interest will likely elute in fractions 1.B.5 – 1.D.6, of which 900 µL are transferred to 1.7 mL tubes and stored at -80˚C to reduce protein degradation while the remaining 100 µL is stored at 4˚C and used for SDS-PAGE and Western blotting. Flow rate remains consistent at 1 mL/minute throughout the purification. Concentration of Buffer B (green) is at 50% from ~5-80 mL because both flow paths are rinsed with Buffers B and then A to assure that tubing from the Buffer B flow path contains Buffer A until the elution gradient begins. If the Buffer B flow path is not prepared as described, the elution would begin

with ~5 mL of 50% Buffer B before decreasing to ~5% Buffer B and increasing as desired. Such a jump in Buffer B concentration will negatively impact protein elution. Flow paths are submerged in their respective buffers before the CCL is flowed over the column and remain here until the elution ends. In full, Buffer A is composed of 10 mM Tris HCl pH 8.0, 500 mM NaCl, and 10% w/v glycerol. Buffer B is composed of 10 mM Tris HCl pH 8.0, 100 mM NaCl, 200 mM imidazole, and 10% w/v glycerol.

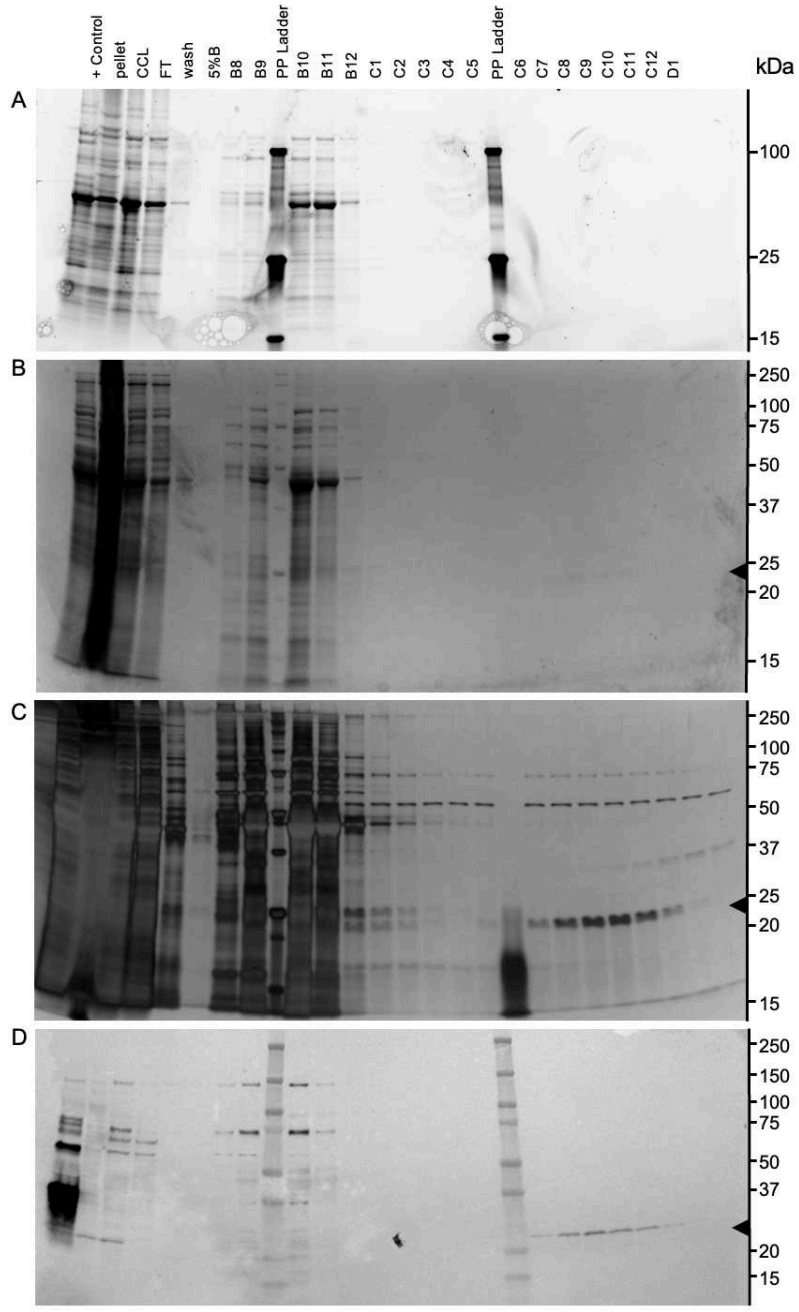


Figure 2.5: Visualization of affinity purified HA and His_6 tagged protein via SDS-PAGE Stain-free gel, Coomassie staining, silver staining, and Western blot. CCL from C-terminal tagged TK0826 (22.6 kDa) purified via affinity chromatography is denatured and proteins are separated by polyacrylamide gel electrophoresis. An SDS-PAGE 4-20% Criterion TGX Stain-Free gel (A) is visualized under UV light. At 22.6 kDa, the natively expressed, tagged protein is not visible within the sensitivity of the “strain-free” system. Denatured proteins run on a 12%

Tris-Glycine gel are stained with Coomassie blue dye (B) which binds to proteins indiscriminately and repeated methanol rinses remove sufficient stain to weakly visualize low concentrations of purified, tagged-protein (lanes C8-10) at 22.6 kDa. Silver staining (C) of denatured proteins separated on a 12% Tris-Glycine gel provides exceptional sensitivity to visualize proteins at low concentrations. Silver (Ag^+) ions bind to the carboxylic acid group on Asp, Glu, or His, the sulfhydryl group on Cys, and the amine group on Lys. The purified, tagged protein is visible at 22.6 kDa along with other proteins that elute within different fraction groups at ~75, 60, 37, and 18 kDa are all easily identified. Denatured proteins separated by SDS-PAGE can be transferred to 0.2 µm polyvinylidene fluoride (PVDF) membrane and visualized via Western blotting employing the appropriate 1˚ and 2˚ antibodies. Here anti-HA 1˚ antibodies made in mouse bind to the HA epitope encoded on tagged TK0826 protein and an anti-mouse conjugated alkaline phosphatase 2˚ antibody is used for chemiluminescent detection. Tagged TK0826 is visible in lanes C6-11, however the greatest concentration is observed in lanes C7-10. CCL from tagged TK1984 (36 kDa) in lane 1 serves as the positive control. “5% B” in lane 6 refers to an isocratic step of 5% Buffer B used in this purification scheme. The remaining elution employed a linear gradient from 5-100% Buffer B. Precision Plus Dual Color Standard protein ladder (BioRad, 1610374) is used in lane 9 on (A-D) and lane 18 on (A) and (D), but not (B) or (C). No sample was loaded in lane 18 on (B) and 1 µg of recombinant-TK1694 was loaded in lane 18 on (C).

2.4.2 Native protein-expression and retention of *in vivo* protein-partnerships

Purification of tagged-protein complexes and retention of native protein interactions established *in vivo* permits identification of associated protein partnerships. Chromosomal MAE techniques can be used to introduce sequences to the genome that encode affinity- (His_6) and epitope- (HA) tags to protein targets and result in retention of native promoters and native levels of target proteins within *T. kodakarensis*. Steady-state protein levels retain protein-partnerships that can be identified through the rapid, gentle, co-purification of the tagged-target protein in association with its natural protein partners.

To maintain the greatest number of protein interactions, cultures should be chilled rapidly in an ice-water bath to preserve as many interactions in their native conformational state as possible. Subjecting cultures to dry ice-mediated cooling is an alternative when exceptionally delicate interactions (*e.g.* protein–RNA associations) are of interest. Pelleted and resuspended cells can be sonicated briefly to disturb membranes while maintaining protein–protein interactions. The downstream chromatography must be completed as rapidly as possible to preserve the maximum number of protein partnerships. Purified fractions can be assessed for

the presence of the tagged protein by SDS-PAGE and Western blotting. Fractions containing the protein of interest are typically pooled, quantified for total protein and trypsin digested for peptide analysis on liquid chromatography tandem mass spectrometry (LC-MS/MS). The high-confidence peptide identification and low-input requirements offered by LC- MS/MS provides a simple and economical route to establish protein interactions in complex protein mixtures.[33,34]

In contrast to large-scale protein production strategies, a single liter or less of *T. kodakarensis* culture grown in rich medium to mid-exponential phase is typically sufficient to identify protein–protein interaction networks. Meaningful interpretation of proteomics data requires reproducibility and biochemical validation. Repeated identification of target and co-purifying proteins, with a minimum of two unique peptides per protein, in multiple biological replicates provides a high level of confidence in *bona fide in vivo* partnerships. Reciprocal pulldowns or known protein associations (*e.g.* established protein complexes) can be used to confirm interactions *in vivo*. Given that a minor number of *T. kodakarensis* proteins have a natural affinity for some purification medias, care should be taken to ensure that co-purifying proteins are not present in identical purifications from near isogenic control strains (*e.g.* the parental strain grown and processed under the same conditions).

2.5 Generation of a tagged library of strains to reveal the metabolic protein interactome.

T. kodakarensis generates hydrogen gas and is thus a potentially valuable biological agent for the production of biohydrogen. To date, closely related species are known to produce 5-10% H_2 gas depending on conditions, however these species are not genetically tractable.[87] Basic parameters of *T. kodakarensis* metabolism are known, yet the complexity of the *in vivo* protein associations has not been fully described. Such in-depth knowledge will allow for regulated reductive flux in a genetically accessible, fast-growing archaeon. To these ends, thirty-three strains have been constructed wherein epitope and affinity tags were added to genes of known metabolic function. Each strain was grown in triplicate and harvested at mid-

exponential growth phase to capture and identify protein interactions in actively dividing cells. Tagged proteins were purified via affinity chromatography and visualized in subsequent fractions via SDS-PAGE and Western blotting. Fractions containing the tagged protein of interest were TCA precipitated and shipped to a mass spectrometry collaborator where proteins underwent tryptic digestion and proteomic mass spectrometry. Data were analyzed against samples from a near isogenic control strain processed in the same manner to determine significant protein associations in the tagged strain. A metabolic network was generated for each known complex and for the entirety of the tagged metabolic protein network. What follows is an in-depth description for each step in generating strains for affinity purification mass spectrometry (AP-MS) analysis to connect the *in vivo* metabolic protein network of thirty-three genes of interest.

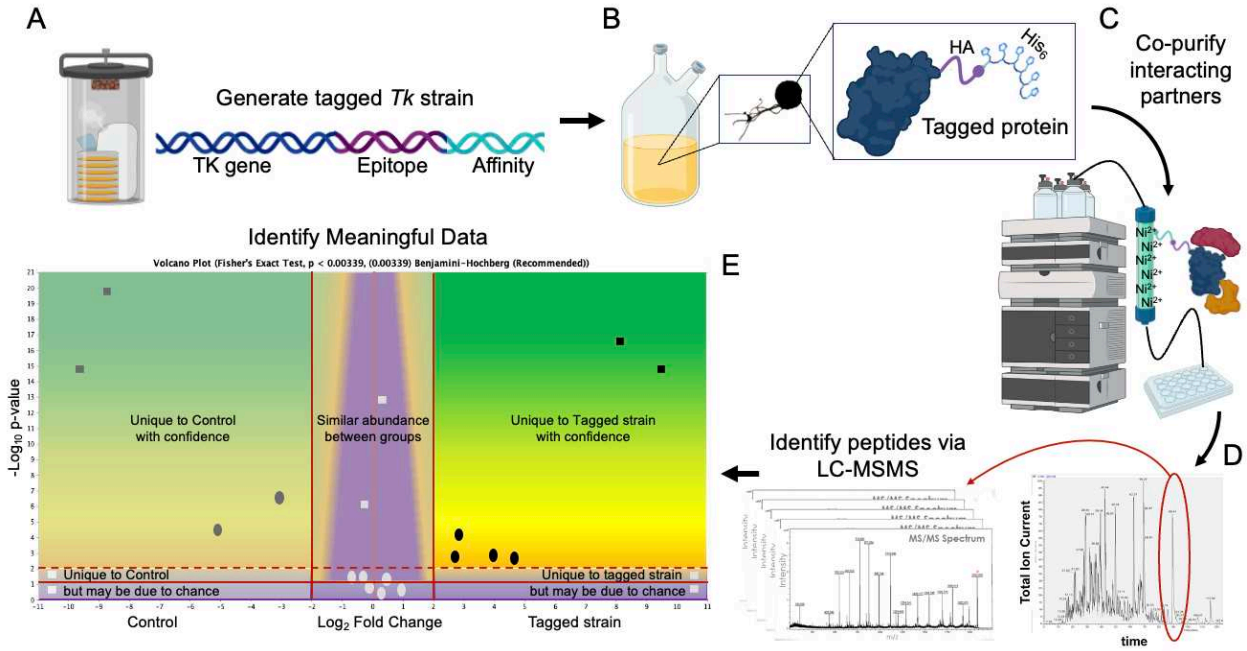


Figure 2.6 Overview of steps involved in generating a metabolic protein network using Affinity Purification Mass Spectrometry. *T. kodakarensis* strains are modified to encode a genomic epitope and affinity tag at the 3' end of the gene sequence (A) using targeted mutagenesis (Sections 2.5.2 and 2.5.3) and selection strategies (Section 2.2.2). The remainder of the steps for AP-MS will be described in detail in Chapter 3, but in brief, sequence confirmed strains are cultured to produce biomass in biological replicates (B). Actively dividing cells are pelleted before being lysed and clarified. The clarified cell lysate (CCL) is purified over a prepared Ni^{2+}-charged affinity column (C) and fractions containing the tagged protein of interest and co-purifying partners are identified by Western blotting (Figure 2.5D, lanes C7-9). Fractions

are TCA precipitated in preparation for sending for proteomic mass spectrometry analysis where they are tryptically digested before peptides are analyzed by tandem mass spectrometry (D). Identified peptides are mapped to the *T. kodakarensis* proteome and tagged samples are compared to samples from the parent strain (a near iso-genic control) to determine statistically significant, meaningful differences in the tagged strain as compared to the parental strain (E). Data from three biological replicates of each tagged protein of interest are used to map a metabolic protein network.

2.5.1 Key genes involved in metabolic processes are tagged.

The *T. kodakarensi*s ~2 Mbp genome encodes glycolytic and amino acid fermentation pathways which generate oxidized small molecules, a small amount of ATP, reduced small molecules (such as NADH), and reduced proteinaceous electron carriers, such as ferredoxins (Figure 2.7). Membrane-bound respiratory complexes are reduced to generate a proton, and then subsequent sodium gradient, to drive the production of ATP via an A_1A_0 ATPase. The generation of such a gradient at the membrane requires that reductive flux must be directed to these complexes by electron carriers such as small molecules, like NADH, or proteins with low reduction potentials. Due to a hyperthermophilic lifestyle, NAD^+ naturally denatures into ADP-ribose and nicotinamide, suggesting an associated expense of using these small molecules for electron transfer.[95] Instead, *T. kodakarensis* encodes three ferredoxins (Fds), iron-sulfur containing proteins with low reduction potential (~ -410 E_0' mV) which serve as proteinaceous electron carriers.[96, 97] Fds are reduced by catabolic enzymes and transfer these valuable electrons to specific anabolic complexes.

It was determined that the three ferredoxins encoded in the *T. kodakarensis* genome maintain specific and distinct protein partnerships suggesting that the metabolic protein network is highly regulated. Using the method of genomic modification via a protein epitope and affinity tag described in this chapter, protein associations of the three *Tk*Fds were mapped revealing a network of catabolic and anabolic partnerships.[34] To expand this metabolic protein network, twenty-five proteins that were found to interact with at least one Fd and 5 proteins that were not shown to directly interact with a Fd but are known for important metabolic roles (TK1215,

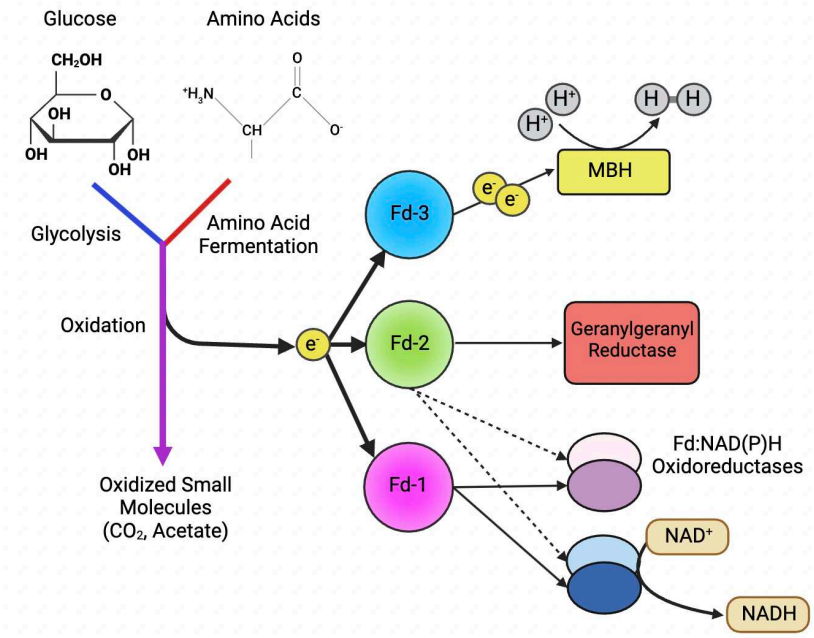


Figure 2.7: Generalized view of metabolism in *T. kodakarensis* involves amino acid and glycolytic fermentation pathways and the distribution of electrons to distinct anabolic complexes by three ferredoxins. Glucose and amino acids are oxidized to generate reduced proteinaceous electron carriers such as the ferredoxin (Fd) proteins.[98] The three Fds encoded in the *T. kodakarensis* genome are proposed to transfer electrons from distinct electron donors to distinct electron acceptors.[34] Fd1 (pink) is primarily involved in the generation of reduced small molecules, such as NADH. Fd2 (green) is involved with GGR (red) and lipid maturation,[99] while Fd3 is thought to primarily donate electrons to MBH in the generation of H_2 gas at the membrane, driving the production of ATP via ATP synthase.[100]

TK1218, TK1222, TK2093, and TK2163) were chosen to generate an AP-MS metabolic network from thirty proteins that were found to interact with at least one Fd and five proteins that were not shown to directly interact with a Fd but are known for important metabolic roles (TK1215, TK1218, TK1222, TK2093, and TK2163) were chosen to generate an AP-MS metabolic network from thirty proteins (Table 2.5). Initial results from two complexes, pyruvate oxidoreductase (POR) and oxoisovalerate oxidoreductase (VOR), indicated that two rubrerythrin-related proteins (TK0650, RBR1 and TK0826, RBR2) that were not originally included in the analysis were clear interacting partners both POR and VOR. Upon further reading, it was realized that RBR1 was found to interact with Fd-1,[34] thus both RBR1 and RBR2 were added to the network. Lastly, Fd-3 (TK2012) was shown to interact with four subunits of the membrane-bound hydrogenase (MBH) complex. Fd-3 can be deleted from the genome if sulfur is present. (Fd-1

and -2 are essential in all conditions.) As Fd-3 is clearly implicated to interact with the MBH complex, we wanted to know see how MBH-partnerships changed in a Fd-3 deletion strain, thus we tagged TK2093 (MBH-N) in ΔTK2012. A total of 33 strains were planned to generate a comprehensive metabolic protein network (Table 2.5).

The process and protocols for generating a sequence confirmed, tagged strain and analyzing this strain in replicate follows in this chapter. Troubleshooting this method to result in reproducible and reliable protein interactions is described in Chapter 3, and the results of the metabolic protein network are described in Chapter 4.

Table 2.5: *T. kodakarensis* proteins found in association with ferredoxins and tagged for generating the metabolic protein interactome.

Gene	Protein	Associated with		
		Fd1	Fd2	Fd3
TK0135	Indolepyruvate-Fd oxidoreductase, β subunit	X		
TK0136	Indolepyruvate-Fd oxidoreductase, α subunit	X	X	X
TK0650	Rubrerythrin Related Protein	X		
TK0826	Rubrerytherin Mn-Catalase			
TK1088	Geranylgeranyl hydrogenase		X	
TK1123	2-Oxoacid:ferredoxin oxidoreductase, γ subunit		X	X
TK1129	2-Oxoacid:ferredoxin oxidoreductases, β subunit		X	
TK1130	2-Oxoacid:ferredoxin oxidoreductase, α subunit		X	X
TK1215	Membrane-bound sulfane reductase, NiFe-hydrogenase large subunit 2			
TK1218	Membrane-bound sulfane reductase, MbxM			
TK1222	Membrane-bound sulfane reductase, MbxF			
TK1325	Ferredoxin:NADH oxidoreductase, α subunit	X	X	
TK1326	Ferredoxin:NADH oxidoreductase, β subunit	X	X	
TK1684	Ferredoxin:NADP oxidoreductase, α subunit	X	X	
TK1685	Ferredoxin:NADP oxidoreductase, β subunit	X	X	
TK1978	2-Oxoisovalarate-ferredoxin oxidoreductase, γ subunit	X		X
TK1979	2-Oxoisovalarate-ferredoxin oxidoreductase, δ subunit	X		X
TK1980	2-Oxoisovalarate-ferredoxin oxidoreductase, α subunit	X		
TK1981	2-Oxoisovalarate-ferredoxin oxidoreductase, β subunit	X		X
TK1982	Pyruvate:ferredoxin oxidoreductase, δ subunit	X		X
TK1983	Pyruvate:ferredoxin oxidoreductase, α subunit	X	X	X
TK1984	Pyruvate:ferredoxin oxidoreductase, β subunit	X	X	X
TK2075	4Fe-4S cluster-binding protein associated with FDH		X	X
TK2076	Formate dehydrogenase, α subunit		X	X
TK2077	Formate dehydrogenase, 4Fe-4S cluster-binding protein		X	X
TK2078	4Fe-4S cluster-binding protein associated with FDH		X	
TK2088	Membrane-bound hydrogenase, MbhI			X
TK2089	Membrane-bound hydrogenase, MbhJ			X
TK2090	Membrane-bound hydrogenase, MbhK			X
TK2091	Membrane-bound hydrogenase, MbhL		X	X
TK2093	Membrane-bound hydrogenase, MbhN			
TK2093: ΔTK2012	Membrane-bound hydrogenase, MbhN; Ferredoxin 3 deletion			
TK2163	Glyceraldehyde-3-phosphate:ferredoxin oxidoreductase			

2.5.2 Generation and confirmation of plasmid vectors

TS559 serves as the parental strain used for markerless genome editing in *T. kodakarensis*. TS559 contains a deletion of TK0149 (Pyruvoyl-dependent arginine decarboxylase) which allows for auxotrophic selection using agmatine. By adding TK0149 to a plasmid, transformants can be selected by their ability to survive without agmatine provided in the media. TS559 is also deleted for TK0664 (Hypoxanthine guanine phosphoribosyl-transferase) which allows for a counter-selection step so the plasmid can be excised from the genome leaving only the desired mutation. The presence of TK0664 on a plasmid induces a phenotype where cells cannot survive in the presence of 6-methylpurine. With the plasmid integrated, cells will not survive in the presence of 6-methylpurine and only cells that excise the plasmid survive, leaving cells that either revert to TS559 or contain the desired mutation only. TS559 also lacks TK0254 (trpE) allowing for the selection of transformants by tryptophan auxotrophy.

pTS700 complements the selection methods of TS559 by containing expression cassettes for TK0149 and TK0664. Sequences that can replace native genomic sequences are added to pTS700 at a SwaI cut site. A pTS700 plasmid containing an unmodified *T. kodakarensis* sequence is referred to as an “A-plasmid” which can then be modified further using site-directed mutagenesis (QuikChange XL Site-Directed Mutagenesis Kit, Agilent 200517).

To insert an HA (epitope) and His_6 (affinity) tag sequence at the 3’ end of a gene, primers homologous to the 3’ end of the gene and surrounding sequence yet which encode the codon optimized HA (TACCCATACGACGTTCCGGACTACGCA) and His_6 (CATCACCATCACCATCAC) sequences are cloned into the A-plasmid using site-directed mutagenesis before the stop codon. Dpn1 digestion of the mutagenized product degrades the original, non-modified plasmid. If the 3’ end overlaps with a neighboring gene, the overlap must be repeated following the 3’ end of the gene of interest such that the tag does not interfere with

the following gene. Commercially-synthesized "gene blocks" can be used in place of primers as longer sequences are often more cost effective to purchase as gene blocks.

The modified plasmid, termed the "C-plasmid" as the tag is encoded on the C-terminus of the protein, is transformed into competent *E. coli* cells, such as Stellar Competent Cells (Takara 636763). The 45-nucleotide insertion and surrounding gene sequence is confirmed by PCR amplification and gel purification followed Sanger sequencing by Azenta (formerly GeneWiz) or by whole plasmid sequencing by Plasmidsaurus. Plasmids can be purified (Qiagen 27104) from fresh *E. coli* culture, and *E. coli* stocks can be maintained in 50% v/v glycerol at -20˚C.

If a C-terminal tag is not viable or cleaved in post-translational processing, an N-terminal tag can be used. To construct a plasmid vector for an N-terminal tag, insert the codon optimized His_6 and then HA sequence following the start codon to generate an "N-plasmid." The methionine codon can be repeated following the HA tag if desired. Again, if the 3' end of the previous gene overlaps with the 5' end of the modified gene, sequences must be repeated to assure the full gene sequences are present.

Six out of 31 C-plasmids necessary for the metabolic protein interactome were previously generated (pTK1685C, pTK1979C, pTK1981C, pTK1982C, pTK1984C, and pTK2088C) by members of the Santangelo Lab and were available for transformation into Stellar cells. The majority of the remaining plasmids were cloned as described above, and six C- or N-plasmids were synthesized by Twist Bioscience (pTK0650C, pTK0826C, pTK129N, pTK1215C, pTK1222N, and pTK2089N).

2.5.3 Generation and confirmation of tagged *T. kodakarensis* strains

Transformation of TS559 with a C- or N-plasmid is performed as described in Section 2.2 with specific and unique features of the process noted here.

TS559 strains that have been transformed with a C-plasmid and contain the entire plasmid integrated into the genome are designated as intermediate strains. PCR of intermediate genomes using primers A/R or F/B (Figure 2.1) cannot distinguish which intermediate genotype is present because a 45-nucleotide difference cannot be identified in a >1,500 bp PCR product and primer pairs must include one primer which binds to the plasmid sequence and one primer which binds the *T. kodakarensis* genome, thus amplicon products are minimally the size of the gene sequence plus the 500-700 bp upstream or downstream required for homologous recombination. Gel purification of the amplified product followed by Sanger sequencing can be used to identify which intermediate genome is present in each strain, however, this is not recommended as there is only modest benefit gained by determining which intermediate genotype is present. Instead, intermediate PCR will allow for identification of bands that are generally the correct size, providing sufficient evidence that the plasmid has integrated into the genome at the anticipated position. Carrying up to ten intermediate strains through to finals selection is often sufficient to provide enough genetic opportunity to realize the desired final strain.

Confirmation of a tagged strain which has excised the plasmid sequence yet maintained the tag can be done via PCR amplification and restriction enzyme digestion or by PCR amplification of a short sequence (< 400 bp) which can permit visual interpretation of a difference in amplicon size of 45 bp (the size of the HA and His_6 tag) (Figure 2.8B). The HA and His_6 sequence contains a BspE1 digestion site, thus PCR amplification, isopropanol precipitation of the PCR product, and treatment with BspE1 enzyme followed by visualization on agarose gel electrophoresis will identify two smaller fragments if the tag is present and one larger fragment if the tag is not present (Figure 2.8A). If BspE1 digest is the method of choice, note that the native *T. kodakarensis* genomic DNA may retain a cut site within the region of amplification, resulting in two bands when the tag is not incorporated and three bands when the tag is has been incorporated. NEB Cutter 3.0 (https://nc3.neb.com/NEBcutter/) can assist in

predicting digestion patterns. If PCR amplification of a small sequence (< 400 bp) is the desired method to confirm presence or absence of the tag sequence, gel electrophoresis using 2% w/v agarose ran for 2-3 hours at low voltage (80-100 mV) will assist in discerning a 250 bp product containing the tag from a 200 bp product which does not contain the tag (Figure 2.2B).

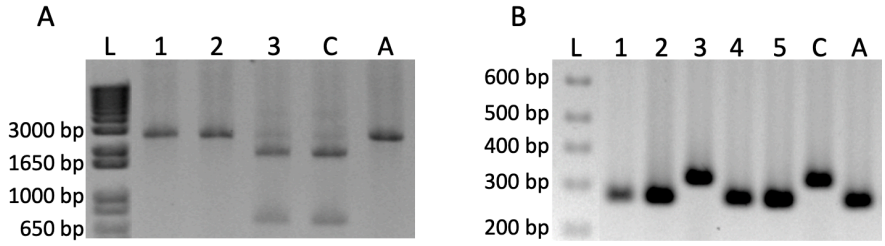


Figure 2.8: Confirmation of the presence or absence of the tag sequence in final, transformed genomic DNA via PCR and enzymatic digest or PCR alone. PCR and agarose gel electrophoresis of purified DNA from isolated colonies of "finals" strains can identify the genotype of interest. The BspE1 restriction enzyme cleaves following the T in the 5'-TCCGGA-3' sequence which occurs in the HA tag sequence (TACCCATACGACGTTCCGGACTACGCA). PCR amplification of the ~2 kb sequence followed by isopropanol precipitation and BspE1 digestion will produce two smaller DNA products if the tag sequence is present (A, lane 3)) and one larger product if the strain reverted back to the wildtype genotype (A, lanes 1 & 2). DNA from TK1088 putative tagged strains has been amplified with primers (007-1087 and 002-1090) generating a 2800 bp product when intact. BspE1 digestion cleaves this single product into two products: 1-2,000 bp product and 1-800 bp product (lanes 3 and C). DNA from strain #3 contains the tag, while strains #1 and 2 have reverted back to wild type. Alternatively, PCR and agarose gel electrophoresis of a small sequence of DNA can reveal the 45-nucleotide difference seen in the tagged strain (B, lane 3 and C). Primers (007-2090 and 008-2089) bracketing the 3' end of TK2090 will amplify a 315 bp product if the tag sequence remains in the genome (B, lane 3) or a 270 bp product if the genome has reverted back to wild type (B, lanes 1, 2, 4, and 5). Amplification and visualization of the C- and A- plasmid can serve as positive and negative controls, respectively, in either approach.

Confirmation of the desired final strain of choice by whole genome sequencing (WGS) should be performed to confirm that no off-target or deleterious mutations have been incorporated during the transformation process (Figure 2.9). Before the lab acquired a MinION (Oxford Nanopore Technologies) for in-house WGS in Fall 2021, WGS was not readily available. Instead, PCR amplification of 1-2 kb bracketing the tagged gene followed by gel purification and Sanger sequencing by Azenta (formerly Genewiz) confirmed the presence of the tag and surrounding gene region. WGS is ideal as sequencing the entire genome can assist in identifying any mutations that may have occurred while passaging intermediate and final

strains in the transformation process. Strains constructed before July 2022 (~twenty-three strains) were confirmed by Sanger sequencing of 1-2 kb surrounding the site of modification. These strains were later confirmed via WGS. In-house WGS using the MinION platform was implemented beginning December 2021 and became the primary means of strain confirmation until ~April 2023 when it was determined that Plasmidsaurus bacterial genome sequencing was available for an affordable $90 per sample. As of Fall 2023, we have been taking advantage of Plasmidsaurus DNA extraction from a cell pellet stabilized in DNA/RNA Shield (Zymo Research, R1100-50) for $15 per sample.

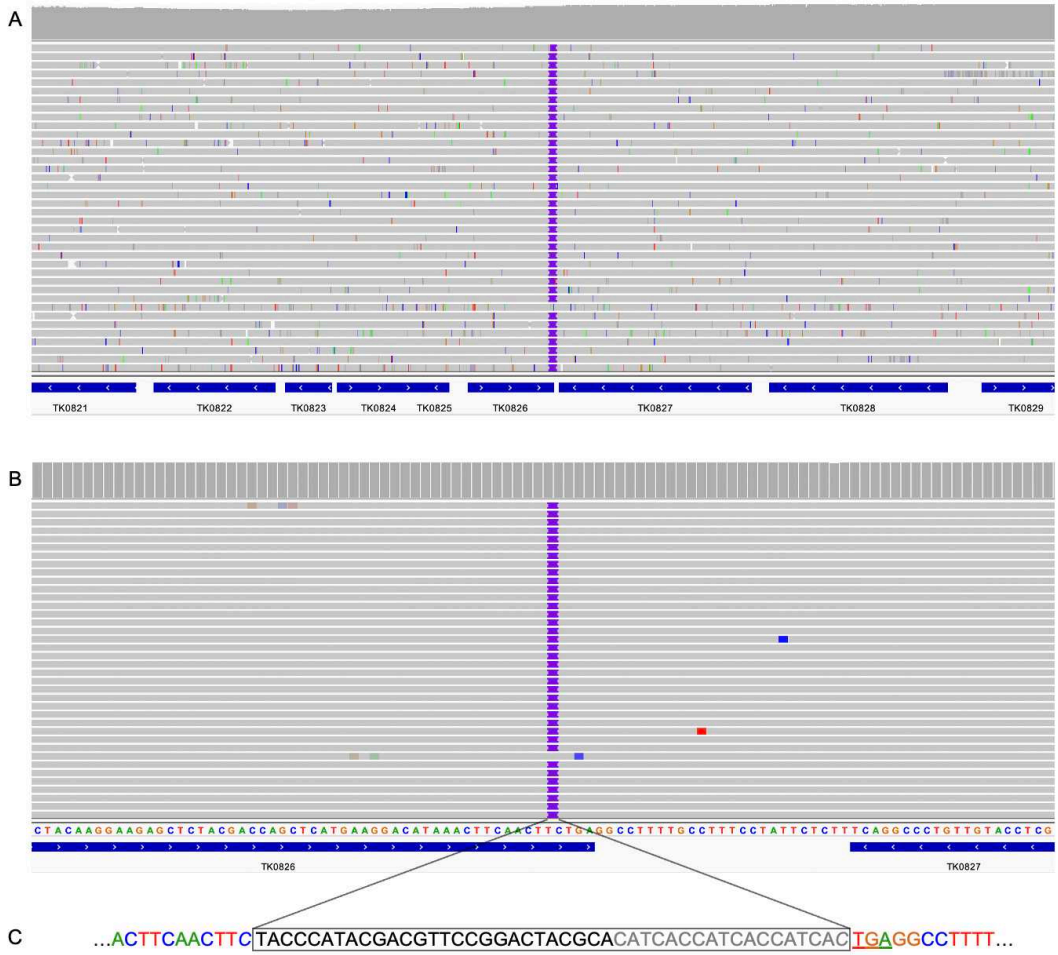


Figure 2.9: Whole genome sequencing of *T. kodakarensis* strain encoding an epitope and affinity tag sequence at the 3' end of the gene. Binary alignment map (.bam) files generated by Plasmidsaurus' sequencing platform from DNA extracted from TK0826C, encoding an HA and His_6 tagged Rubreryther in Mn-catalyze (RBR2), are mapped to the TS559 (parental) genome and viewed using the Integrative Genomics Viewer (IGV). Standard IGV settings identify reads horizontally as grey bars and depth of coverage by stacking reads vertically.

While only ~40X coverage is shown, the genome was sequenced at an average of ~700X coverage. Nucleotide scale differences compared to the reference genome are shown in red (thymine, T), green (adenine, A), blue (cytosine, C), or orange (guanine, G). Insertions are shown as purple bars with the grey reads and deletions are indicated by the absence of a grey bar and a thinner black line in its place. The annotated reference genome (.gtf file) is shown in dark blue at the bottom. Mutations are observed as colored lines throughout the TK0821-TK0829 sequence (A). While multiple single point mutations appear scattered throughout, the the 45-nucleotide insertion (purple bar) indicates that the HA (black) and His_6 (grey) sequences have been inserted at the 3' end of TK0826 (B) immediately before the stop codon (TGA) (C). Note that the insertion sequence (purple bar) vertically aligns before the 2nd and 3rd position of the last codon (TTC) of TK0826, however note also that a cystine residue occurs at the 3' end of both the His_6 tag sequence and the TK0826 gene preceding the stop codon. The mapping algorithm has aligned the *C* (italicized) of the final TK0826 codon (TT*C*) as the insertion and the 3' C of the insertion as the final nucleotide in TK0826. Such an error is due to the alignment and is not a cause for concern.

Long-read sequencing is error prone, however depth of coverage can compensate for high error rates. In-house MinION sequencing yielded 3-30X coverage depending on the number of samples in the library, the number of available pores on the flow cell, and run time (fewer samples, more pores, and longer run times increase read per run). While 3-30X coverage is not ideal, it was considered sufficient to identify the accurate integration of an HA and His_6 tag sequence. Plasmidsaurus WGS returns > 200X coverage consistently and has reached up to 1000X coverage (Figure 2.9). Read depth using Plasmidsaurus DNA extraction yields ~100-200X coverage consistently.

2.5.4 Limitations to genetic techniques

Site specific mutagenesis of certain gene sequences is sometimes not permissible. The most widely understood limitation is observed with essential genes (source). Of the thirty-three` genes involved in this study, thirteen are considered essential (Table 6). Twenty-nine C-terminal tagged strains encoding HA and His 6 sequences at the 3' end of the gene of interest were generated. One C-tagged strain (TK1130C) could not be visualized via Western blot. As the tag may have been cleaved, an N-terminal tagged strain (TK1130N) was generated encoding a His_6 and HA tag sequence at the 5' end of the gene. Genetic addition of a C-terminal tag was only

limited in four strains (TK1129, TK1218, TK1222, TK2089) of which, an N-termina tag was possible in two (TK1129N and TK2089N). TK1218 and TK1222 encode the M and F subunits of the membrane-bound sulfane reductase (MBS) complex. While both genes are non-essential, over ten attempts to generate a tagged strain at the 5' or 3' end failed. Two out of ten attempts with each gene at both termini produced viable intermediates, however all final strains reverted back to wild type. One MBS subunit, TK1215C the NiFe-hydrogenase large subunit 2 of MBS, could be tagged and confirmed.

Table 2.6: *T. kodakarensis* proteins essentiality and tagging outcomes.

Gene	Protein	Essential ?	C-Tag	N-Tag
TK0135	Indolepyruvate-Fd oxidoreductase, β	Yes	X	
TK0136	Indolepyruvate-Fd oxidoreductase, α	Yes	X	
TK0650	Rubrerythrin Related Protein	?	X	
TK0826	Rubrerytherin Mn-Catalase	?	X	
TK1088	Geranylgeranyl hydrogenase	Yes	X	
TK1123	2-Oxoacid:ferredoxin oxidoreductase, γ	Yes	X	
TK1129	2-Oxoacid:ferredoxin oxidoreductases, β	Yes	Not viable	X
TK1130	2-Oxoacid:ferredoxin oxidoreductase, α	Yes	X	X
TK1215	Membrane-bound sulfane reductase, NiFe-hydrogenase large subunit 2	No	X	
TK1218	Membrane-bound sulfane reductase, MbxM	No	Not viable	Not viable
TK1222	Membrane-bound sulfane reductase, MbxF	No	Not viable	Not viable
TK1325	Ferredoxin:NADH oxidoreductase, α	No	X	
TK1326	Ferredoxin:NADH oxidoreductase, β	No	X	
TK1684	Ferredoxin:NADP oxidoreductase, α	No	X	
TK1685	Ferredoxin:NADP oxidoreductase, β	No	X	
TK1978	2-Oxoisovalarate-ferredoxin oxidoreductase, γ	Yes	X	
TK1979	2-Oxoisovalarate-ferredoxin oxidoreductase, δ	Yes	X	
TK1980	2-Oxoisovalarate-ferredoxin oxidoreductase, α	Yes	X	
TK1981	2-Oxoisovalarate-ferredoxin oxidoreductase, β	Yes	X	
TK1982	Pyruvate:ferredoxin oxidoreductase, δ	Yes	X	
TK1983	Pyruvate:ferredoxin oxidoreductase, α	Yes	X	
TK1984	Pyruvate:ferredoxin oxidoreductase, β	Yes	X	
TK2075	4Fe-4S cluster-binding protein associated with FDH	No	X	
TK2076	Formate dehydrogenase, α subunit	No	X	
TK2077	Formate dehydrogenase, 4Fe-4S cluster-binding protein	No	X	
TK2078	4Fe-4S cluster-binding protein associated with FDH	No	X	
TK2088	Membrane-bound hydrogenase, MbhI	No	X	
TK2089	Membrane-bound hydrogenase, MbhJ	No	Not viable	X

TK2090	Membrane-bound hydrogenase, MbhK	No	X
TK2091	Membrane-bound hydrogenase, MbhL	No	X
TK2093	Membrane-bound hydrogenase, MbhN	No	X
TK2093: ΔTK2012	Membrane-bound hydrogenase, MbhN; Ferredoxin 3 deletion	No	X
TK2163	Glyceraldehyde-3-phosphate:ferredoxin oxidoreductase	No	X

2.6 Discussion

T. kodakarensis has emerged as a model species for the study of biological processes at high temperatures and as an expression platform for thermostable enzymes, owing mainly to the development of highly reproducible, accurate genomic and ectopic manipulation strategies.[23] While numerous advantages of a native hyperthermophilic expression system are evident, the *T. kodakarensis* platform could be improved.

Although genomic manipulation strategies in *T. kodakarensis* are simple and sufficiently straightforward to allow many researchers to generate large numbers of unique strains, genome modification systems that use CRISPR/Cas are not yet available for *T. kodakarensis*. Most archaeal species (including the *Thermococcales*) encode CRISPR elements,[72,94] and efforts to establish a thermostable CRISPR/Cas system for use in *T. kodakarensis* may accelerate strain construction. The development of additional cost-effective selection strategies, particularly non-prototrophic selections that could be universally deployed regardless of host genotype, would speed strain construction and add to the versatility of genetic techniques for *T. kodakarensis*. One of the most important tools for exploiting *T. kodakarensis* as an expression platform will be the development of a tightly-controlled, inducible promoter system. Available riboswitch-mediated regulatory expression systems are useful and are commonly employed within the *T. kodakarensis* research community [32,86] but lack the dynamic range desired for rapid and efficient control of protein expression.

Despite the challenges imposed by high temperature environments, the rapid development of tools for genetic manipulation for *T. kodakarensis* have provided the scientific community with a versatile protein expression platform conducive to the production and study of thermostable proteins. The combination of autonomously replicating ectopic expression systems, controlled protein overexpression and the ability to introduce genomic modifications makes *T. kodakarensis* a versatile platform for the expression of thermostable proteins and the evaluation of *in vivo* protein partnerships.

Thirty-one strains have been generated to evaluate metabolic protein-protein interactions in the hydrogen-generating, genetically accessible *T. kodakarensis* using AP-MS. Two studies using AP-MS in *T. kodakarensis* have been published.[33, 34] The first tagged and purified nineteen components thought to be involved in archaeal DNA replication, thereby generating a conclusive view of the DNA replication protein network.[33] The second study may be considered a prequel to this current work in that the three ferredoxins encoded by *T. kodakarensis* were tagged to reveal specific and distinct interacting partners, alluding to the potential for rational, regulated electron flux through these three proteinaceous electron carriers. The following chapter will first outline AP-MS as a method, and then dive deep into each step of the process to provide the necessary foundation for a statistically significant, reproducible *in vivo* metabolic protein network.

2.7. Protocols

2.7.1 Targeted mutagenesis of chromosomal DNA

Markerless-modifications of *T. kodakarensis* chromosomal DNA typically rely on integration of a non-replicative plasmid harboring the desired modification into the genome, followed by excision of most plasmid sequences from the genome to yield a strain with a modified genomic sequence. Detailed instructions for the construction of integrative plasmids

have been provided elsewhere;[31] herein we focus on the steps for transformation and confirmation of genomic modifications. All procedures are carried out anaerobically. For simplicity, we do not detail the specific selective and counter-selective procedures nor the exact media formulations of solid or liquid mediums, as the multitude of genetic techniques for *T. kodakarensis* provide many options depending on the host strain employed (Tables 2.1 and 2.2).

Transformation and selection for plasmid integration into the genome

Grow an MAE-compatible *T. kodakarensis* strain (Table 2.1) at 85˚C for 12 hours in ASW-YT medium supplemented with S˚, KOD1 vitamins and any necessary additives. Harvest *via* centrifugation and resuspend cells in 3 mL of 0.8X ASW, transfer to a microcentrifuge tube, and place on ice. To transform, add ~2 µg of plasmid DNA (in < 15 µL) to a 200 µL aliquot of concentrated cell suspension and incubate on ice for 30 minutes. Heat-shock the cell and plasmid mixture at 85˚C for 45 seconds using a heating block, then immediately incubate on ice for an additional 30 minutes. Plate transformed cells on a selection medium and grow anaerobically at 85˚C for ~2-5 days.

Confirmation of appropriate plasmid integration and intermediate genome formation

Colonies resulting from transformations are anticipated to have integrated the plasmid sequences into the genome. Cells from distinct colonies should be picked and inoculated in 5 mL ASW-based medium containing KOD1 vitamins while ensuring maintenance of the positive selection(s) employed. 5 mL cultures are grown overnight anaerobically at 85˚C to yield sufficient biomass for 1 mL genomic DNA preparations following established procedures.[31] Perform diagnostic PCRs using genomic DNA. Amplicons generated from primers pairs complementary to genomic and integrated plasmid sequences permit diagnostic confirmation of the desired intermediate genotype(s) (Figure 2.2A). It is imperative that the primer sets used in

the diagnostic PCRs include one primer that is complementary to genomic DNA sequences and a second primer complementary to sequences of the integrated plasmid (Figure 2.1D). Such pairing ensures that resulting amplicons originate from a genomically-integrated plasmid. Depending on how the plasmid integrated into the genome (either upstream or downstream to the modification site) two intermediate genomes are possible.[28] Whenever possible, identification of separate strains containing each potential intermediate genotype is desirable as each intermediate strain can yield the desired final genotype.

Counter selection, confirmation of plasmid excision and final genome confirmation

5 mL ASW-based medium grown cultures of confirmed intermediate strains, supplemented to permit spontaneous excision of the selective and/or counterselective makers, should be grown overnight, and the resulting cells spread on solid-medium containing the appropriate counter-selective pressures and allowed ~4 days of anaerobic growth at 85˚C. Resulting colonies are anticipated to have spontaneously excised the plasmid from their genomes and either retained the desired modification or reverted back to the parental genome (Figure 2.1E; 2.2B). Cells from distinct colonies should be picked, inoculated in 5 mL ASW-based medium and grown anaerobically overnight at 85˚C to yield sufficient biomass for 1 mL genomic DNA preparations following established procedures.[31] Amplicons generated from primers pairs complementary to genomic sequences permit diagnostic confirmation of the desired final genotype (Figure 2.2B). Depending on the desired modification, PCR followed by gel electrophoresis can be used for size discrimination of the target region. If the modification produces a size difference that is difficult to discern *via* PCR (*i.e.*: short sequences), differential restriction digestion patterns of the PCR amplicon originating from the region of interest may assist in identifying desired genotypes. When diagnostic PCRs indicate the presence of the desired modification, genotypes should be confirmed *via* Sanger sequencing of PCR

amplicon(s) that span the target loci or *via* whole genome sequencing (*e.g.*: Oxford Nanopore Sequencing).

Preparing high quality DNA for whole genome sequencing

In-house, WGS using the MinION from Oxford Nanopore Technologies requires 50 ng of high quality genomic DNA in ~9 µL. Cells from ~30 mL of freshly grown culture can be pelleted at 15,000 x *g* for 15 minutes before being resuspended in 400 µL of cold 1X PBS, pH 7.23. Thorough pipetting aids in cell lysis before adding 1 µL of Proteinase K (20 mg/mL), vortexing, and incubating at 55˚C for >1 hour to degrade proteins. 3 µL of RNase A (10 ng/mL), is added before incubating at 37˚C for 30 minutes to degrade RNA. Solution is separated by centrifugation at 15,000 x *g* for 10 minutes and the supernatant is transferred to a new tube. Add 400 µL Monarch gDNA Binding Buffer (New England BioLabs, T3014-2), vortex, and transfer into Monarch gDNA Purification Column (New England BioLabs, T3017-1), before adding 500 µL Monarch gDNA Wash Buffer (New England BioLabs, T3015-2). Centrifuge 15,000 x *g* and dispose of flow through. Place column in a new collection tube and add 35 µL of heated (~55˚C) Monarch gDNA Elution Buffer (New England BioLabs, T3016-2). Incubate elution buffer on the column at ambient temperature for > 1 minute before eluting DNA into the collection tube by centrifugation (15,000 x *g* for 1 minute). Procedure should yield 200-600 ng/µL DNA.

In-house whole genome sequencing using the Oxford Nanopore MinION

Prepare libraries using the Rapid Barcode Sequencing Kit (SQK-RBK110.96, Oxford Nanopore Technologies [ONT]) by preparing 50 ng high quality DNA in 9 µL and incubate with 1 µL barcode (provided in 96-well plate with ONT kit) at 30˚C for 2 minutes followed by 80˚C for 2 minutes. Pool samples and add 1:1 volume of well-mixed SPRI beads (AMPure XP, Beckman Coulter, A63881) and incubate at ambient temperature for 5 minutes before migrating beads

with a magnet and removing liquid. Suspend beads in 1.5 mL 80% ethanol before migrating beads, removing ethanol, and repeating ethanol wash once. Allow beads to dry before incubating in 15 µL Elution Buffer for 10 minutes at ambient temperature. Migrate beads and remove 11 µL into a new 1.7 mL tube and add 1 µL Rapid Adapter F (RAPF), mix gently, and incubate at ambient temperature for 5 minutes. Prepare the flow cell and confirm sufficient pores are available and in good condition. Make Priming Solution by adding 30 µL Flush Tether (FLT) and 1,170 µL Flush Buffer to a new 1.7 mL tube. Load the sample onto the flow cell and initiate sequencing. A typical sequencing run will produce 100,000 – 400,000 reads depending on samples, available pores, and time.

FASTQ files are aligned to the reference genome of choice (previously downloaded onto the MinION hard drive) by selecting Analysis > Alignment > data/Users/, navigating to the appropriate folder where fastq.pass files from the respective sequencing run are stored. Select files and continue to output where an output folder is selected. Navigate to the folder where reference genomes are stored and select the reference genome of choice (*e.g.* KOD1 or TS559 are recommended). Continue to summary and start alignment to generate binary alignment map (BAM) files for sample in the sequencing library. Transfer BAM files to a data repository of choice.

Whole genome sequencing using Plasmidsaurus Bacterial Genome Sequencing service

1 µg of high-quality DNA is shipped overnight to Plasmidsaurus for ONT long-read sequencing. Alternatively, Plasmidsaurus offers DNA extraction from cell pellets. Cells from 20-30 mL of freshly growing *T. kodakarensis* culture grown in ASW-YT media with 5 g/L pyruvate in exponential or early stationary phase (OD_{600} 0.8) are pelleted at 9,000 x *g* for 15 minutes. Avoid using S˚ in cultures as it will be difficult to isolate the cell pellet. Spent media is removed and the cell pellet (~30-50 mg) is resuspended in 1 mL 1X PBS before transfer to a 1.7 mL microcentrifuge tube and centrifuging at 15,000 x *g* for 5 minutes. Remove supernatant and

resuspend the final pellet in 500 µL of Zymo 1X DNA/RNA shield (Zymo Research, R1100-250). Transfer to a 2 mL screw cap tube and wrap cap with parafilm before shipment.

Plasmidsaurus will email a link to download a zipped file containing a summarized report (.html), and Analysis, Annotation, and Reads folders. A zipped file of FASTQ reads (raw_reads.fastq.gz) are accessed in the Reads folder. These data can also be found on the Plasmidsaurus Dashboard when logged in.

Generating merged and indexed alignment files to view WGS using IGV

Merged, indexed alignment files can be generated in a number of ways. Here, UseGalaxy.Org is employed. Upload BAM files or zipped raw reads in fastq format (*e.g.* raw_reads.fastq.gz, zipped files do not need to be unzipped) and the respective reference genome to UseGalaxy.Org. Use the SamToolsMerge tool (tools can be searched using the search bar in the Tools panel at left) to merge and index read files. Download both the merged file (.bam) and the index (.bai) and save them with the same name in the same folder (only the file extension, .bam or .bai, will differ). Aligned sequences can be viewed using the Integrative Genomics Viewer (IGV.org) application. It is recommended to construct an annotated genome using Gene Transfer Format or General Feature Format (.gtf or .gff) file for the reference genome to easily navigate aligned sequencing reads in IGV.

2.7.2 Transformation of *T. kodakarensis - E. coli* shuttle vectors

Grow an MAE-compatible *T. kodakarensis* strain (Table 2.1) at 85˚C for 12 hours in ASW-YT medium supplemented with S˚, KOD1 vitamins and any necessary additives. Harvest *via* centrifugation and resuspend cells in 3 mL of 0.8X ASW, transfer to a microcentrifuge tube, and place on ice. Add ~2 µg of plasmid DNA (in < 15 µL) to a 200 µL aliquot of concentrated cell suspension and incubate on ice for 30 minutes. Heat-shock the cell and plasmid mixture at 85˚C for 45 seconds using a heating block, then immediately incubate on ice for an additional 30

minutes. Spread transformed cells on a solid-medium compatible with the selection marker(s) (see Table 2.2). Incubate the plated cells anaerobically for ~2-6 days at 85˚C. Pick cells from a discrete colony and inoculate into 10 mL of the appropriate ASW-based medium required for the chosen selectable marker(s). Allow the culture to grow anaerobically at 85˚C overnight. Maintenance of a retained plasmid can be confirmed by growth on selective media, PCR, or by plasmid purification and transformation back into *E. coli*. Methodologies for generating a *T. kodakarensis–E. coli* shuttle vector with selectable phenotypes and replicative origins are outlined elsewhere.[12,35]

2.7.3 Ectopic protein expression

Inoculate 1 L of ASW-based medium supplemented with S˚, KOD1 vitamins, and any necessary additives (*e.g.* 1 mM Agmatine sulfate) with a 10 mL freshly grown, overnight culture. Cultures should be started with a 1:100 inoculum. For expression reliant on constitutive promoters, cells should be harvested *via* centrifugation immediately prior to entry into the stationary phase to maximize cell density. For expression reliant on inducible promoters, culture conditions should be altered to permit protein expression when the optical density has reached $OD_{600} \approx 0.2\text{-}0.3$, and cells should be harvested *via* centrifugation prior to stationary phase. Secreted protein targets can be recovered from the culture media. Large scale protein expression and purification can be done using affinity chromatography as described in 2.5.4. Additional purification of the desired protein product may be desired. Size exclusion following affinity purification is recommended.

2.7.4 Native expression to analyze protein-protein interactions

Affinity purification followed by tandem mass spectrometry to reveal native level *in vivo* protein associations is reliant on genetically modified strains encoding affinity and epitope tags on the protein of interest. Six-histidines (His_6; affinity-tag) at the N- or C-terminus allow for

protein purification using a chelating column charged with nickel. A C-terminal tag assures complete translation and is ideal, however, in the case that a C-terminal tag is not genetically viable or the tag is cleaved in post-translational processing, an N-terminal tag can be used. Bound proteins are eluted using an imidazole gradient. A hemagglutinin sequence (HA; epitope-tag) is encoded between the His_6 and protein coding sequence to allow for easy identification of the target protein via Western blotting. The following procedures should be done at 4˚C.

Generation of Vectors for the Genetic Modification of T. kodakarensis

The pTS700 vector is linearized and incubated with T4 DNA polymerase and dCTP to produce unique sticky ends. “A”-plasmids which contain an unmodified *T. kodakarensis* genomic sequence are constructed by amplification of the target site of modification +500-700 bp upstream and downstream using 13 bp tailed primers. Incubation with T4 DNA polymerase and dGTP produces non-complimentary sticky-ends. The amplicon is directionally inserted into pTS700 via ligation-independent cloning.

Codon optimized His_6 (CATCACCATCACCATCAC) and hemagglutinin, HA, (TACCCATACGACGTTCCGGACTACGCA) sequences are integrated before the stop codon of the 3’ end of the gene of interest, present on the pTS700 A-plasmid. Primers homologous to the 3’ end of the target gene and adjacent sequence but which contain an HA and His_6 sequence are used in QuikChange reaction to produce a “C”-plasmid. If the 3’ end of the gene of interest overlaps with the 5’ end of the following gene, be sure to repeat overlapping sequences before or after the tag and stop codon as needed so that neither final protein sequence is affected. If the C-terminus cannot be modified, the His6 and HA sequence can be integrated at the 5’ end of the gene following the start codon. Note that the affinity tag should always be on the most terminal position. Plasmid modifications are confirmed by Sanger sequencing by GeneWiz (now Azenta) or by whole plasmid sequencing by Plasmidsaurus.

Generation of tagged strains

Transformation follows as previously described in Section 2.5.1. The *T. kodakarensis* parental strain TS559 lacks TK0149 and TK0644; thus, TS559 requires agmatine supplement for growth and can survive in the presence of 6-methyl purine. The sequence confirmed C-plasmid is transformed into TS559 via heat shock and integrates into the genome via homologous recombination before or after the target gene. Transformed cultures are grown without agmatine to select for incorporation of the plasmid (which confers agmatine prototrophy). Two intermediate genome options may be produced. Intermediates are grown with agmatine and counter-selected using 6-methyl purine to encourage excision of the vector through homologous recombination. Excising the plasmid will either restore the original genome or retain the tag which is verified by diagnostic PCR. Genomic DNA of final strains are confirmed via amplification of the modified region and Sanger sequencing by GeneWiz (now Azenta) or whole genome sequencing using the MinION or by Plasmidsaurus.

Culture and Harvest of Biomass

Sequence confirmed, tagged strains are grown in ASW-YT +5 g/L pyruvate at 85˚C. When appropriate, 2g/L S˚ was added. 1 L cultures were inoculated with freshly growing starter cultures at 1:100 dilution. Cultures grown to mid-exponential phase (OD_{600} 0.4-0.7) are rapidly cooled on ice and equal volumes are distributed into 3-350 mL centrifuge tubes to generate three technical replicates per 1 L culture. Cultures are spun at 9,000 x *g* for 20 minutes at 4˚C in in a JA10 rotor in a Beckman-Coulter floor centrifuge. Spent media is discarded and cell pellets are transferred to 15 mL conical tubes and stored at -80˚C. Each 1 L culture is considered a biological replicate. Three biological replicates are minimally recommended to ensure statistically relevant and reproducible results.

Cell lysis

Cell pellets are thawed and resuspended with ~3 mL per g of biomass in lysis buffer (Buffer A). Pellet from 330 mL of culture can be expected to be resuspended in ~12 mL. Sonicate the cell suspension on ice at half-maximal output for 10 seconds on and 30 seconds off for a total of ~5 minutes. Suspension will be grey and homogenous when cells are sufficiently lysed. To retrieve membrane-bound proteins, cells may be sonicated for up to 60 minutes but avoid unnecessary sonication as this will disturb native protein associations. Generate a clarified cell lysate (CCL) by centrifugation (9,000 x *g* for 20 minutes) to pellet membranes while retaining cytoplasmic proteins in the supernatant. To assure no debris remains, filter the CCL over 0.45 µm sterile filter and purify immediately. A pellet from 330 mL of culture at OD_{600} of 0.4 should result in 10-11 mL of filtered CCL.

Purification via affinity chromatography (Figure 2.3)

Prepare the ÄKTA chromatography unit (or similar) by ensuring all tubing is flushed and a fraction collector is prepared. Purifications should be done at 4˚C. Charge HiTrap chelating column with 100 mM $NiSO_4$. Clean flow paths and HiTrap column with 5 column volumes of Buffer B (elution buffer) followed by 5 column volumes of Buffer A (lysis buffer). Load CCL over the column. The His_6 tagged protein will bind to the Ni^{2+}-charged column. Collect the flow through (FT) in a 50 mL conical. Wash the column with 5 column volumes of Buffer A. Collect the wash in a 50 mL conical. The tagged protein should not be in the wash. Optionally wash the column with 5% of Buffer B and collect the 5% B wash in a 50 mL conical. Elute the target protein with an imidazole gradient and collect the eluant in small (~1 mL) fractions. Transfer 90% of each fraction (900 µL if 1 mL fraction) to 1.7 mL tubes and freeze at -80˚C as soon as elution is complete. Store the remaining 10% of fraction at 4˚C for visualization using SDS-PAGE and Western blotting. Flush the entire chromatography system with 5 column volumes of Buffer B, followed by 10 column volumes of ultra-pure H_2O. To regenerate the chelating column,

first remove any remaining proteins with 5 column volumes of 6 M guanidine HCl, then flush the column with H_2O, and finally strip the bound Ni^{2+} from the column with 2 column volumes 0.5 M EDTA, then flush with H_2O. Store the chromatography apparatus and HiTrap column in 20% ethanol.

Visualization and identification of tagged proteins in purified fractions (Figure 2.4)

Aliquots of each fraction can be resolved *via* SDS-PAGE and/or Western blotting to identify fractions containing the target protein. It is recommended to run the following samples to include both positive and negative controls: protein ladder (Precision Plus Stained protein ladder, BioRad, 1610375) pellet, CCL, flow through, wash, and the 5% B wash (Figure 2.3). The chromatogram from the purification should include a spike in UV during the elution (Figure 2.4). It is recommended to visualize ~18 fractions beginning at this UV spike and continuing past the spike. Heat treat 10ul of sample with 5X SDS-PAGE gel-loading buffer (Cold Spring Harbor Protocols) at 99˚C for 5 minutes before loading onto Criterion TGX Stain-Free 4-20% tris-glycine pre-cast gel (BioRad, 5678095). Denatured samples are separated at 150mV for 50 minutes before being imaged (BioRad, GelDoc). Proteins are transferred to 0.2 µM Trans-Blot Turbo Midi PVDF Transfer Pack and the BioRad transfer system (BioRad, 1704157 and 1704150) using standard settings (25 V, 2,500 mA, 7 min.). PVDF membrane is incubated in 5% BSA in 1X TBST for 10 minutes at room temperature before replacing the BSA with 10 mL of 1:10,000 Anti-HA monoclonal primary antibody (Invitrogen, 26183-1MG) in 1X TBST. Blots are rocked in 1˚ AB overnight at 4˚C overnight and washed with 1X TBST 3 times before rocking with 1:1000 anti-mouse polyclonal alkaline phosphatase secondary antibody (LGC Clinical Diagnostics, 5220-0312) in 1X TBST for 1 hour at room temperature. Blots are washed with Wash Buffer (0.3% Tween-20, 100 mM Maleic Acid, 150 mM NaCl, pH 7.5) twice for 15 minutes before being incubated with Detection Buffer (1 M Tris-HCl, 1 M NaCl, pH 9.5) for 5 minutes followed by development with NBT-BCIP (ThermoScientific, 34042) for ~5-50 minutes in the dark until

bands were visible. Blots are dried in ambient conditions and imaged (BioRad, GelDoc). Denatured, tagged proteins appeare in 3-10 fractions, at the correct size, and are not observed in the un-tagged, parental strain (TS559).

TCA precipitation for sample storage and shipping

The 3-4 fractions clearly showing the greatest amount of tagged protein are thawed on ice, pooled, and quantified using the Qubit4 Fluorometer (ThermoFisher Scientific). Volume for 20 µg of purified protein is precipitated by the addition of 4:1 100% (w/v) TCA. Samples are incubated on ice for 10 minutes before precipitated proteins are centrifuged at 15,000 x *g* for 5 minutes at 4˚C. The supernatant is removed and the pellet is washed with 200 µL cold acetone before being centrifuged at 15,000 x *g* for 5 minutes at 4˚C. The supernatant is removed and the acetone wash is repeated once. The supernatant is removed once more and the pellet is dried at 95˚C for 5-10 minutes. TCA precipitated proteins are stored at -80˚C before being shipped on ice overnight to a mass spec facility of choice. The Ohio State University Campus Chemical Instrument Center (CCIC) is recommended.

Tryptic digestion and proteomic mass spectrometry for protein identification

Samples (20 µg) are resuspended in 100 µL 50 mM ammonium bicarbonate. 5 µL of DTT (5 µg/µL in 50 mM ammonium bicarbonate) is added and the sample is incubated at 65˚C for 15 minutes. After the incubation, 5 µL of iodoacetamide (15 mg/ml in 50 mM ammonium bicarbonate) is added and the sample is kept in the dark at room temperature for 30 minutes. Sequencing grade-modified trypsin (Promega, Madison WI) prepared in 50 mM ammonium bicarbonate is added to the sample with a 1:20 enzyme-substrate ratio and the reaction is carried on at 37˚C overnight. The reaction is quenched the next morning by adding acetic acid (50 µL, 0.2%) for acidification. Sample is dried by speedvac and resuspended with 40 µL of 0.1% formic acid.

Liquid chromatography and tandem mass spectrometry

Liquid chromatography-nanospray tandem mass spectrometry (LC/MS/MS) of protein identification is performed on a Thermo Scientific orbitrap Fusion mass spectrometer equipped with an EASY-Spray™ Sources operated in positive ion mode. Samples (1 µg or 2 µL) are separated on an easy spray nano column (Pepmap™ RSLC, C18 3µ 100A, 75µm X 250 mm Thermo Scientific) using a 2D RSLC HPLC system from Thermo Scientific. Each sample is injected into the µ-Precolumn Cartridge (Thermo Scientific) and desalted with 0.1% Formic Acid in water for 5 minutes. The injector port is then switched to injection mode and the peptides are eluted off of the trap onto the column. Mobile phase A is 0.1% Formic Acid in water and acetonitrile (with 0.1% formic acid) is used as mobile phase B. Flow rate is set at 300 nL/minute. Mobile phase B was first kept at 2% for 5 minutes and then increased from 2% to 16% in 50 min before increased again from 16% 50 25%% in 10 minutes. Mobile phase B is increased again from 25%-85% in 1 minute and then kept at 85% for another 1 minute before being brought back quickly to 2% in 1 minute. The column is equilibrated at 2% of mobile phase B (or 98% A) for 12 minutes before the next sample injection. MS/MS data is acquired with a spray voltage of 1.6 KV and a capillary temperature of 305 °C is used. The scan sequence of the mass spectrometer is based on the preview mode data dependent TopSpeed™ method: the analysis is programmed for a full scan recorded between *m/z* 200-2000 and a MS/MS scan to generate product ion spectra to determine amino acid sequence in consecutive scans starting from the most abundant peaks in the spectrum in the next 3 seconds. To achieve high mass accuracy MS determination, the full scan is performed at FT mode and the resolution is set at 120,000 with internal mass calibration. The AGC Target ion number for FT full scan is set at 4×10^5 ions, maximum ion injection time is set at 50 ms and micro scan number is set at 1. MSn is performed using HCD in ion trap mode to ensure the highest signal intensity of MSn spectra. The HCD collision energy is set at 32%. The AGC Target ion number for ion trap MSn scan is set at 3.0E4 ions, maximum ion injection time is set at 35 ms and micro scan number is set at 1. Dynamic

exclusion is enabled with a repeat count of 1 within 20s and a low mass width and high mass width of 10ppm.

Data are searched using Mascot Daemon by Matrix Science version 2.7.0 (Boston, MA) via ProteomeDiscoverer (version 2.4 Thermo Scientific,) and the database searched against the most recent Uniprot databases. The mass accuracy of the precursor ions is set to 10 ppm, accidental pick of 1 ^{13}C peaks was also included into the search. The fragment mass tolerance is set to 0.5 Da. Carbamidomethylation (Cys) is used as a fixed modification and considered variable modifications are oxidation (Met) and deamidation (N and Q). Four missed cleavages for the enzyme are permitted. A decoy database is also searched to determine the false discovery rate (FDR) and peptides are filtered according at 1% FDR. Proteins identified with at least one unique peptide are considered as reliable identification.

REFERENCES

1. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **5**, 172 (2014).
2. Kim, H., Yoo, S. J. & Kang, H. A. Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Res.* **15**, 1–16 (2015).
3. Cao, J., Wang, T., Wang, Q., Zheng, X. & Huang, L. Functional insights into protein acetylation in the hyperthermophilic archaeon *Sulfolobus islandicus*. *Mol. Cell. Proteomics* **18**, 1572–1587 (2019).
4. Gąciarz, A. *et al.* Efficient soluble expression of disulfide bonded proteins in the cytoplasm of *Escherichia coli* in fed-batch fermentations on chemically defined minimal media. *Microb. Cell Fact.* **16**, 108 (2017).
5. Takemasa, R., Yokooji, Y., Yamatsu, A., Atomi, H. & Imanaka, T. *Thermococcus kodakarensis* as a host for gene expression and protein secretion. *Appl. Environ. Microbiol.* **77**, 2392–2398 (2011).
6. Gao, L., Hidese, R. & Fujiwara, S. Function of a thermophilic archaeal chaperonin is enhanced by electrostatic interactions with its targets. *J. Biosci. Bioeng.* **124**, 283–288 (2017).
7. Duong-Ly, K. C. & Gabelli, S. B. Explanatory chapter: troubleshooting recombinant protein expression: general. *Methods Enzymol.* **541**, 209–229 (2014).
8. Peleg, Y. & Unger, T. Resolving bottlenecks for recombinant protein expression in *E. coli*. *Methods Mol. Biol.* **800**, 173–186 (2012).
9. Sørensen, H. P. & Mortensen, K. K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* **115**, 113–128 (2005).
10. Santangelo, T. J., Cubonová, L., James, C. L. & Reeve, J. N. TFB1 or TFB2 is sufficient for *Thermococcus kodakaraensis* viability and for basal transcription *in vitro*. *J. Mol. Biol.* **367**, 344–357 (2007).
11. Rodrigues, M. V. *et al.* Bifunctional CTP:inositol-1-phosphate cytidylyltransferase/CDP-inositol:inositol-1-phosphate transferase, the key enzyme for di-myo-inositol-phosphate synthesis in several (hyper)thermophiles. *J. Bacteriol.* **189**, 5405–5412 (2007).
12. Santangelo, T. J., Cubonová, L., & Reeve, J. N. Shuttle vector expression in Thermococcus kodakaraensis: contributions of cis elements to protein synthesis in a hyperthermophilic archaeon. *Appl. Environ. Microbiol.* **74**, 3099–3104 (2008).
13. French, S. L., Santangelo, T. J., Beyer, A. L. & Reeve, J. N. Transcription and translation are coupled in Archaea. *Mol. Biol. Evol.* **24**, 893–895 (2007).
14. Farkas, J. A., Picking, J. W. & Santangelo, T. J. Genetic techniques for the archaea. *Annu. Rev. Genet.* **47**, 539–561 (2013).
15. Haque, R. U., Paradisi, F. & Allers, T. *Haloferax volcanii* for biotechnology applications: challenges, current state and perspectives. *Appl. Microbiol. Biotechnol.* **104**, 1371–1382 (2020).
16. Martínez-Espinosa, R. M. Heterologous and homologous expression of proteins from haloarchaea: denitrification as case of study. *Int. J. Mol. Sci.* **21**, (2019).
17. Pohlschroder, M. & Schulze, S. Haloferax volcanii. *Trends Microbiol.* **27**, 86–87 (2019).
18. Guss, A. M., Rother, M., Zhang, J. K., Kulkarni, G. & Metcalf, W. W. New methods for tightly regulated gene expression and highly efficient chromosomal integration of cloned genes for Methanosarcina species. *Archaea* **2**, 193–203 (2008).
19. Rother, M. & Metcalf, W. W. Genetic technologies for Archaea. *Curr. Opin. Microbiol.* **8**, 745–751 (2005).
20. Schocke, L., Bräsen, C. & Siebers, B. Thermoacidophilic *Sulfolobus* species as source for extremozymes and as novel archaeal platform organisms. *Curr. Opin. Biotechnol.* **59**, 71–

77 (2019).
21. Peng, N., Han, W., Li, Y., Liang, Y. & She, Q. Genetic technologies for extremely thermophilic microorganisms of *Sulfolobus*, the only genetically tractable genus of crenarchaea. *Sci. China Life Sci.* **60**, 370–385 (2017).
22. Santangelo, T. J., Cubonová, L., & Reeve, J. N. *Thermococcus kodakarensis* genetics: TK1827-encoded beta-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl. Environ. Microbiol.* **76**, 1044–1052 (2010).
23. Atomi, H. & Reeve, J. Microbe Profile: *Thermococcus kodakarensis*: the model hyperthermophilic archaeon. *Microbiology* **165**, 1166–1168 (2019).
24. Farkas, J., Chung, D., DeBarry, M., Adams, M. W. W. & Westpheling, J. Defining components of the chromosomal origin of replication of the hyperthermophilic archaeon *Pyrococcus furiosus* needed for construction of a stable replicating shuttle vector. *Appl. Environ. Microbiol.* **77**, 6343–6349 (2011).
25. Zeldes, B. M. *et al.* Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Front. Microbiol.* **6**, 1209 (2015).
26. Lipscomb, G. L. *et al.* Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. *Appl. Environ. Microbiol.* **77**, 2232–2238 (2011).
27. Rashid, N. & Aslam, M. An overview of 25 years of research on *Thermococcus kodakarensis*, a genetically versatile model organism for archaeal research. *Folia Microbiol.* **65**, 67–78 (2020).
28. Hileman, T. H. & Santangelo, T. J. Genetics Techniques for *Thermococcus kodakarensis*. *Front. Microbiol.* **3**, 195 (2012).
29. Walker, J. E. & Santangelo, T. J. Analyses of in vivo interactions between transcription factors and the archaeal RNA polymerase. *Methods* **86**, 73–79 (2015).
30. Santangelo, T. J. & Reeve, J. N. Deletion of switch 3 results in an archaeal RNA polymerase that is defective in transcript elongation. *J. Biol. Chem.* **285**, 23908–23915 (2010).
31. Gehring, A. M., Sanders, T. J. & Santangelo, T. J. Markerless gene editing in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Bio Protoc* **7**, (2017).
32. Speed, M. C., Burkhart, B. W., Picking, J. W. & Santangelo, T. J. An Archaeal fluoride-responsive riboswitch provides an inducible expression system for hyperthermophiles. *Appl. Environ. Microbiol.* **84**, (2018).
33. Li, Z., Santangelo, T. J., Cuboňová, L., Reeve, J. N. & Kelman, Z. Affinity purification of an archaeal DNA replication protein network. *MBio* **1**, e00221-10 (2010).
34. Burkhart, B. W., Febvre, H. P. & Santangelo, T. J. Distinct physiological roles of the three ferredoxins encoded in the hyperthermophilic archaeon *Thermococcus kodakarensis*. MBio 10, e02807-18 (2019).
35. Catchpole, R., Gorlas, A., Oberto, J. & Forterre, P. A series of new *E. coli-Thermococcus* shuttle vectors compatible with previously existing vectors. *Extremophiles* **22**, 591–598 (2018).
36. Endoh, T., Kanai, T. & Imanaka, T. Effective approaches for the production of heterologous proteins using the *Thermococcus kodakaraensis*-based translation system. *J. Biotechnol.* **133**, 177–182 (2008).
37. Endoh, T., Kanai, T. & Imanaka, T. A highly productive system for cell-free protein synthesis using a lysate of the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. *Appl. Microbiol. Biotechnol.* **74**, 1153–1161 (2007).
38. Endoh, T. *et al.* Cell-free protein synthesis at high temperatures using the lysate of a hyperthermophile. *J. Biotechnol.* **126**, 186–195 (2006).
39. Counts, J. A. *et al.* Physiological, metabolic and biotechnological features of extremely

thermophilic microorganisms. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **9**, (2017).
40. Adams, M. W. & Kelly, R. M. Finding and using hyperthermophilic enzymes. *Trends Biotechnol.* **16**, 329–332 (1998).
41. Crosby, J. R. *et al.* Extreme thermophiles as emerging metabolic engineering platforms. *Curr. Opin. Biotechnol.* **59**, 55–64 (2019).
42. Straub, C. T. *et al.* Biotechnology of extremely thermophilic archaea. *FEMS Microbiol. Rev.* **42**, 543–578 (2018).
43. Jin, M., Gai, Y., Guo, X., Hou, Y. & Zeng, R. Properties and applications of extremozymes from deep-sea extremophilic microorganisms: A Mini Review. *Mar. Drugs* **17**, (2019).
44. Gomes, J. & Steiner, W. The biocatalytic potential of extremophiles and extremozymes. *Food Technol. Biotechnol.* **42**, 223–225 (2004).
45. Laksanalamai, P., Pavlov, A. R., Slesarev, A. I. & Robb, F. T. Stabilization of Taq DNA polymerase at high temperature by protein folding pathways from a hyperthermophilic archaeon, Pyrococcus furiosus. *Biotechnol. Bioeng.* **93**, 1–5 (2006).
46. Han, H. *et al.* Improvements of thermophilic enzymes: From genetic modifications to applications. *Bioresour. Technol.* **279**, 350–361 (2019).
47. Suresh, A. *et al.* Recent advancements in the synthesis of novel thermostable biocatalysts and their applications in commercially important chemoenzymatic conversion processes. *Bioresour. Technol.* **323**, 124558 (2021).
48. Kumar, S., Dangi, A. K., Shukla, P., Baishya, D. & Khare, S. K. Thermozymes: Adaptive strategies and tools for their biotechnological applications. *Bioresour. Technol.* **278**, 372–382 (2019).
49. Rekadwad, B. & Gonzalez, J. M. Multidisciplinary involvement and potential of thermophiles. *Folia Microbiol.* **64**, 389–406 (2019).
50. Bhardwaj, N., Kumar, B. & Verma, P. A detailed overview of xylanases: an emerging biomolecule for current and future prospective. *Bioresources and Bioprocessing* **6**, 1–36 (2019).
51. de Souza, P. M. & de Oliveira Magalhães, P. Application of microbial α-amylase in industry - A review. *Braz. J. Microbiol.* **41**, 850–861 (2010).
52. Sellami-Kamoun, A. *et al.* Stability of thermostable alkaline protease from *Bacillus licheniformis* RP1 in commercial solid laundry detergent formulations. *Microbiol. Res.* **163**, 299–306 (2008).
53. Stoner, M. R. *et al.* Protease autolysis in heavy-duty liquid detergent formulations: effects of thermodynamic stabilizers and protease inhibitors. *Enzyme Microb. Technol.* **34**, 114–125 (2004).
54. Madhavi, J., Srilakshmi, J., Rao, M. V. R. & Rao, K. Efficient leather dehairing by bacterial thermostable protease. *International Journal of Bio-Science and Bio-Technology* **3**, 11–26 (2011).
55. Shrinivas, D. & Naik, G. R. Characterization of alkaline thermostable keratinolytic protease from thermoalkalophilic *Bacillus halodurans* JB 99 exhibiting dehairing activity. *Int. Biodeterior. Biodegradation* **65**, 29–35 (2011).
56. Preetha, S. & Boopathy, R. Purification and characterization of a milk clotting protease from *Rhizomucor miehei*. *World J. Microbiol. Biotechnol.* **13**, 573–578 (1997).
57. Larsen, M. D., Kristiansen, K. R. & Hansen, T. K. Characterization of the proteolytic activity of starter cultures of *Penicillium roqueforti* for production of blue veined cheeses. *Int. J. Food Microbiol.* **43**, 215–221 (1998).
58. Maheshwari, R., Bharadwaj, G. & Bhat, M. K. Thermophilic fungi: their physiology and enzymes. *Microbiol. Mol. Biol. Rev.* **64**, 461–488 (2000).
59. Andrade, V. S. *et al.* Production of extracellular proteases by *Mucor circinelloides* using D-glucose as carbon source / substrate. *Braz. J. Microbiol.* **33**, 106–110 (2002).
60. Kim, J.-M. *et al.* Effects of crude proteases extracted from *Bacillus polyfermenticus* on

tenderizing pork meat. *Food Science of Animal Resources* **31**, 491–496 (2011).
61. Bekhit, A. A., Hopkins, D. L., Geesink, G., Bekhit, A. A. & Franks, P. Exogenous proteases for meat tenderization. *Crit. Rev. Food Sci. Nutr.* **54**, 1012–1031 (2014).
62. Fujiwara, S. Extremophiles: developments of their special functions and potential resources. *J. Biosci. Bioeng.* **94**, 518–525 (2002).
63. Rothschild, L. J. & Mancinelli, R. L. Life in extreme environments. *Nature* **409**, 1092–1101 (2001).
64. Chambers, C. R. & Patrick, W. M. Archaeal nucleic acid ligases and their potential in biotechnology. *Archaea* **2015**, 170571 (2015).
65. Tanabe, M., Ishino, Y. & Nishida, H. From structure-function analyses to protein engineering for practical applications of DNA ligase. *Archaea* **2015**, 267570 (2015).
66. Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T. & Imanaka, T. Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic Pyrococcus sp. *Appl. Environ. Microbiol.* **60**, 4559–4566 (1994).
67. Atomi, H., Fukui, T., Kanai, T., Morikawa, M. & Imanaka, T. Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. *Archaea* **1**, 263–267 (2004).
68. Liman, G. L. S., Hulko, T., Febvre, H. P., Brachfeld, A. C. & Santangelo, T. J. A linear pathway for mevalonate production supports growth of *Thermococcus kodakarensis*. *Extremophiles* **23**, 229–238 (2019).
69. Sato, T., Fukui, T., Atomi, H. & Imanaka, T. Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* **185**, 210–220 (2003).
70. Sato, T., Fukui, T., Atomi, H. & Imanaka, T. Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Appl. Environ. Microbiol.* **71**, 3889–3899 (2005).
71. Schut, G. J. *et al.* The order thermococcales and the family thermococcaceae. *The Prokaryotes: Other Major Lineages of Bacteria and the Archaea* 363–383 (2014).
72. Fukui, T. *et al.* Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* **15**, 352–363 (2005).
73. Čuboňováa, L. *et al.* An archaeal histone is required for transformation of *Thermococcus kodakarensis. J. Bacteriol.* **194**, 6864–6874 (2012).
74. Fukuda, W., Morimoto, N., Imanaka, T. & Fujiwara, S. Agmatine is essential for the cell growth of *Thermococcus kodakaraensis*. *FEMS Microbiol. Lett.* **287**, 113–120 (2008).
75. Krupovic, M., Gonnet, M., Hania, W. B., Forterre, P. & Erauso, G. Insights into dynamics of mobile genetic elements in hyperthermophilic environments from five new *Thermococcus* plasmids. *PLoS One* **8**, e49044 (2013).
76. Soler, N. *et al.* The rolling-circle plasmid pTN1 from the hyperthermophilic archaeon *Thermococcus nautilus*. *Mol. Microbiol.* **66**, 357–370 (2007).
77. Gorlas, A., Alain, K., Bienvenu, N. & Geslin, C. *Thermococcus prieurii* sp. nov., a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* **63**, 2920–2926 (2013).
78. Cabrera, J. A., Bolds, J., Shields, P. E., Havel, C. M. & Watson, J. A. Isoprenoid synthesis in *Halobacterium halobium*. Modulation of 3-hydroxy-3-methylglutaryl coenzyme a concentration in response to mevalonate availability. *J. Biol. Chem.* **261**, 3578–3583 (1986).
79. Matsumi, R., Manabe, K., Fukui, T., Atomi, H. & Imanaka, T. Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host-marker system based on antibiotic resistance. *J. Bacteriol.* **189**, 2683–2691 (2007).
80. Santangelo, T. J. *et al.* Polarity in archaeal operon transcription in *Thermococcus*

kodakaraensis. *J. Bacteriol.* **190**, 2244–2248 (2008).
81. Santangelo, T. J. & Reeve, J. N. Genetic tools and manipulations of the hyperthermophilic heterotrophic Archaeon *Thermococcus kodakarensis*. in *Extremophiles Handbook, ed. K. Horikoshi (Tokyo: Springer), 567–582* 567–582 (2010).
82. Gehring, A. M. *et al.* Genome replication in *Thermococcus kodakarensis* Independent of Cdc6 and an origin of replication. *Front. Microbiol.* **8**, 2084 (2017).
83. Santangelo, T. J., Cuboňová, L., & Reeve, J. N. Deletion of alternative pathways for reductant recycling in *Thermococcus kodakarensis* increases hydrogen production. *Mol. Microbiol.* **81**, 897–911 (2011).
84. Ishino, S. *et al.* Biochemical and genetical analyses of the three mcm genes from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Genes Cells* **16**, 1176–1189 (2011).
85. Hidese, R., Inoue, T., Imanaka, T. & Fujiwara, S. Cysteine desulphurase plays an important role in environmental adaptation of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Mol. Microbiol.* **93**, 331–345 (2014).
86. Sanders, T. J. *et al.* FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat Microbiol* **5**, 545–553 (2020).
87. Kanai, T. *et al.* Overproduction of the membrane-bound [NiFe]-hydrogenase in *Thermococcus kodakarensis* and its effect on hydrogen production. *Front. Microbiol.* **6**, 847 (2015).
88. Basen, M., Sun, J. & Adams, M. W. W. Engineering a hyperthermophilic archaeon for temperature-dependent product formation. *mBio* **3**, e00053–12 (2012).
89. Mattiroli, F. *et al.* Structure of histone-based chromatin in Archaea. *Science* **357**, 609–612 (2017).
90. Gorlas, A., Krupovic, M., Forterre, P. & Geslin, C. Living side by side with a virus: characterization of two novel plasmids from *Thermococcus prieurii*, a host for the spindle-shaped virus TPV1. *Appl. Environ. Microbiol.* **79**, 3822–3828 (2013).
91. Spaans, S. K., van der Oost, J. & Kengen, S. W. M. The chromosome copy number of the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1. *Extremophiles* **19**, 741–750 (2015).
92. Lin-Chao, S., Chen, W. T. & Wong, T. T. High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. *Mol. Microbiol.* **6**, 3385–3393 (1992).
93. Del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M. & Díaz-Orejas, R. Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **62**, 434–464 (1998).
94. Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J. & Reeve, J. N. Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* **15**, 684 (2014).
95. Hachisuka, S. I., Sato, T. & Atomi, H. Metabolism dealing with thermal degradation of NAD+ in the hyperthermophilic archaeon Thermococcus kodakarensis. J Bacteriol 199, 1–14 (2017).
96. Schut, G. J. et al. The role of geochemistry and energetics in the evolution of modern respiratory complexes from a proton-reducing ancestor. Biochimica et Biophysica Acta - Bioenergetics vol. 1857 958–970 (2016).
97. Atkinson, J. T., Campbell, I., Bennett, G. N. & Silberg, J. J. Cellular Assays for Ferredoxins: A Strategy for Understanding Electron Flow through Protein Carriers That Link Metabolic Pathways. Biochemistry 55, 7047–7064 (2016).
98. Schut, G. J., Menon, A. L. & Adams, M. W. W. Enzymes of Central Metabolism. [12]. (2001).
99. Jain, S., Caforio, A. & Driessen, A. J. M. Biosynthesis of archaeal membrane ether lipids.

Frontiers in Microbiology vol. 5 Preprint at https://doi.org/10.3389/fmicb.2014.00641 (2014).
100. Yu, H. et al. Structure of an Ancient Respiratory System. Cell 173, 1636-1649.e16 (2018).

CHAPTER 3
ESTABLISHING AFFINITY PURIFICATION - MASS SPECTROMETRY PROCEDURES PERMITTING REPRODUCIBLE AND HIGH-CONFIDENCE IDENTIFICATIONS OF *IN VIVO* PROTEIN INTERACTIONS [2]

Summary

The majority of enzymatic activities have been established from *in vitro* assays which provide valuable insights into the conditions required for optimal protein function, including parameters such as K_m, K_d, pH, temperature, substrate preference, or turnover-rates. While *in vitro* investigations of enzyme activities reveal key parameters under defined conditions, *in vitro* studies alone often fail to reveal dynamic, regulatory, and key protein-protein interactions (*i.e*. protein-assemblies) that dictate overall enzyme function, efficiency, location, and availability within a complex cellular environment. Identification of *bona fide in vivo* protein assemblies -- and how such assemblies dictate the totality of cellular metabolism -- is thus an important milestone to understand the regulation imposed on metabolism, energy-production, and energy. This chapter details the comprehensive efforts to establish a reproducible technique to preserve and identify *bona fide in vivo* protein partnerships of twenty-nine known and predicted dominant proteins in the central metabolism and energy-conserving pathways of the model marine hyperthermophilic species *Thermococcus kodakarensis*.[1–8] We establish the optimal procedures to gently isolate protein assemblies, to identify – with high statistical confidence – the proteins retained within unique assemblies, LC-MSMS techniques and analyses pathways which yield the greatest confidence metrics for protein identification, what techniques and methods permit a rationale for removal of insignificant protein-interactions spuriously identified within our procedures, and how new metrics can be applied to quantize and provide confidence metrics in the validity of newly established protein-protein interactions *in vivo.*

[2] Data presented in this chapter in part represents the efforts of talented students and/or colleagues under the tutelage of Seré Williams including: Danielle Riley, Teagan Rockwood, David Crosby, Kate Call, Jared LeCuyer, Lucy Latta, and Isabella Bendorf.

3.1 Introduction

Resolving the dynamic nature of nanoscale interactions in living systems is arguably the current frontier of molecular biology. A multitude of non-, semi-, and rigorously-quantitative methods have attempted to capture biologically accurate, *in vivo* protein-protein interactions.[9–28] Combinatorial methods prevail -- melding classic *in vitro* methods, such as affinity purification, with genetic strategies that allow for markerless, genomic modification of proteins at native expression levels in combination with evolving algorithms to process large data sets, rigorous experimental design, use of controls, and statistics – reproducible, *in vivo* networks are coming to life. Affinity purification followed by proteomic mass spectrometry (AP-MS) is one such combinatorial method which has the potential to identify intracellular partnerships *en masse* quickly and accurately aiding in the ability to understand protein partnerships in a manner more closely resembling the complexity found in biological systems.[12,27–30] Using AP-MS, Michaelis et al., (2023) recently mapped the entirety of the yeast proteome and resolved a highly connected network which maintained 3X more protein interactions than reported in currently published interactome maps.[31] Such methods may be able to explain previously unexplainable findings such as the metabolic efficiency of certain microbes. But first, the power of such a method must be met with rigorous and reproducible validation at each step to limit false positives and false negatives and requires considerable preparation including the generation of genetically modified and confirmed strains, protein expression at native levels, brisk yet repeatable purification schemes, and detection via proteomic mass spectrometry.

Here, in-depth troubleshooting of each step will be described for the generation of a metabolic protein interactome in the presence and absence of elemental sulfur (the preferred terminal electron acceptor discussed in Chapter 1, +S˚ or -S˚) of the archaeal, hyperthermophile *Thermococcus kodakarensis.*[32,33] The generation of twenty-nine unique strains for this analysis is detailed in Chapter 2, while results of this broad protein network will be detailed in Chapter 4.

3.1.1 Overview of the AP-MS method and how it is employed to generate a protein interactome

The AP-MS method relies on genetics, protein purification, and proteomic mass spectrometry techniques (Figure 3.1). When combined, these techniques can identify wholistic protein networks representing the complexity of *in vivo* protein associations.

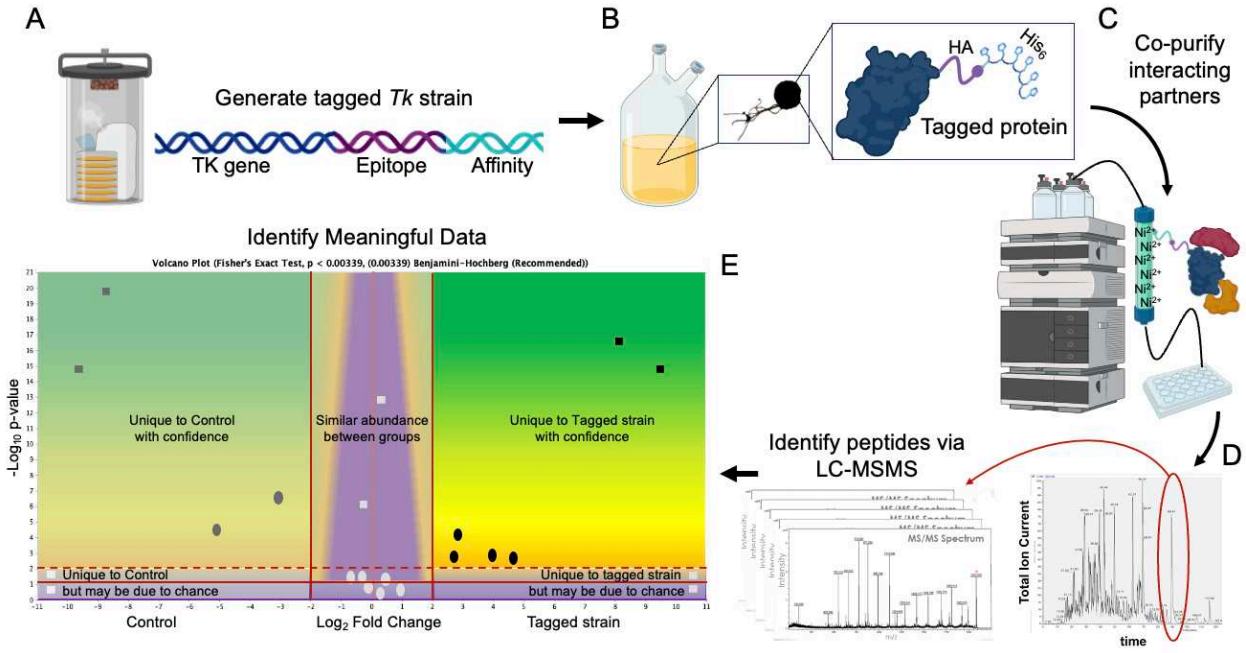


Figure 3.1: Overview of steps involved in generating a metabolic protein network using Affinity Purification Mass Spectrometry. *T. kodakarensis strains are modified to encode a genomic epitope and affinity tag at the end of the gene sequence (A) using targeted mutagenesis (Sections 2.5.2 and 2.5.3) and selection strategies (Section 2.2.2). Sequence confirmed strains are cultured to produce biomass in biological replicates (B, Section 3.2.1). Actively dividing cells are pelleted before being lysed and clarified. The clarified cell lysate (CCL) is purified over a prepared Ni^{2+}-charged affinity column (C, Section 3.2.2) and fractions containing the tagged protein of interest and co-purifying partners are identified by Western blotting (Figure 3.2D). Fractions are pooled and proteins are TCA precipitated in preparation for shipment to the MS facility. Peptides from trypsin-digested proteins are separated by liquid chromatography followed by tandem mass spectrometry (LC-MSMS, D, Section 3.2.3). Identified peptides are mapped to the T. kodakarensis proteome. Proteins identified in tagged samples are compared to those identified in the parent strain (a near iso-genic control, TS559, described in Chapter 2) to determine statistically significant, meaningful differences in the tagged strain as compared to the parental strain (E, Section 3.3). Data from three biological replicates of each tagged protein of interest are used to map a metabolic protein network.*

As described in Chapter 2, a genetically accessible strain of *T. kodakarensis*, such as TS559,[34] allows for markerless, site-directed genetic modification via homologous

recombination at the genome (also referred to as “markerless allelic exchange,” or MAE; Figure 3.1A). A vector containing native *T. kodakarensis* sequence and the modification can be inserted into the genome using auxotrophic selection. The vector can likewise be excised from the genome and the resulting strain will either revert to the parental genotype or maintain the modification of interest, verified first by PCR and then by whole genome sequencing.

First a collection of genes which encoded protein of minimally putative functions can be determined as useful in generating a protein network. Here, thirty-one genes encoding proteins with known metabolic function were chosen to evaluate the metabolic network in *T. kodakarensis*. Only twenty-nine genes could be successfully tagged due to genetic limitations of two proteins, TK1218 and TK1222. Genetic limitations are discussed in Chapter 2.5.4.

Genetic modification for AP-MS includes the addition of two genomically encoded sequences. First, histidines (typically in sequences of threes such as His_6) have a natural affinity for divalent cations.[35] The inclusion of the His_6 tag onto the terminus of a protein of interest allows for simple affinity purification of the protein out of a complex cell mixture, such as a cell lysate (Figure 3.1B). Proteins expressed at native expression levels are not easily identified from affinity purification schemes without specific antibodies for the protein of interest. While the purification is sure to separate proteins with a natural affinity to the divalent cation, the protein of interest may elute in any of ~5-50 fractions (depending on the purification scheme). SDS-PAGE and subsequent Western blotting with an effective primary antibody against an epitope may be used to visualize the purified proteins (Figure 3.2).

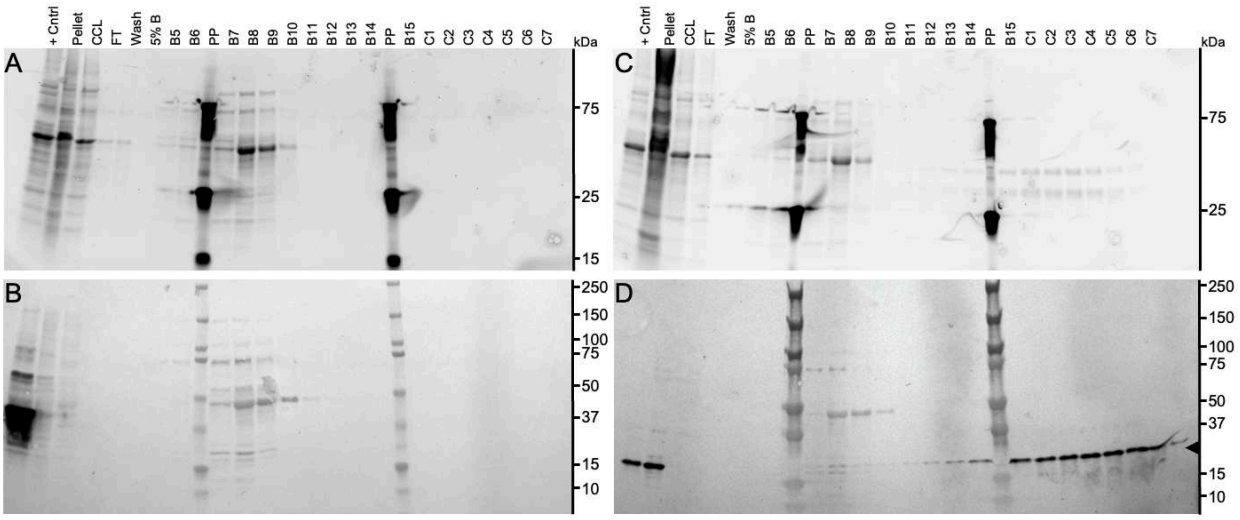


Figure 3.2: Affinity purification of the parental strain (TS559) compared to a tagged strain (TK1978C) at native expression levels visualized by SDS-PAGE and Western blot. Purification of a near isogenic control is necessary to compare purification of a tagged-protein. While the HA and His_6 tags employed are known for having a neutral impact to physiology, such null impact must be verified. Additionally, the anti-HA 1˚ antibody may bind to off-target epitopes naturally occurring in the T. kodakarensis proteome, thus comparative identification of the tagged protein in the parental and tagged-strains is recommended. Here heat and SDS-denatured protein samples from the TS559 parent (A) and TK1978C (the tagged γ subunit of pyruvate-ferredoxin oxidoreductase, POR, and 2-oxoisovalerate-ferredoxin oxidoreductase, VOR) (C) are loaded onto a 4-20% Criterion TGX stain-free gel (BioRad, 5678095) and proteins are separated in a tris-glycine buffer (1X Laemmli) at 150 V for 50 minutes. CCL from TK1984C (36 kDa) is used as the positive control (+ Cntrl), and samples of the pellet, CCL, FT, wash, and 5% Buffer B collection steps are loaded before loading 18 fractions from the beginning of peak 1 elution (B5) through 195 mM imidazole (97.5% Buffer B, fraction C7). Precision Plus Stained Dual-Color (PP) protein ladders are loaded in the 9th and 18th wells of this 26-well gel to aid in identifying protein size (BioRad, 1610375). For this ladder, only 75 and 25 kDa bands are visible via SDS-PAGE. Benchmark Protein Ladder (Thermo Fisher, 10747012) is recommended for use with SDS-PAGE when Western blotting is not preformed. Many proteins of various sizes are observed in +Cntrl, pellet, and CCL, with fewer proteins observed in the FT, Wash, and 5% B lanes for both parental (A) and tagged (C) strains. Elution of peak 1 is visible in fractions B5 – B9 from the parental (A) and tagged (C) strains. Bands in lanes B13 – C5 observed in only the tagged-strain SDS-PAGE (C) are contamination (likely human keratin protein of ~ 44 and 66 kDa) often seen in SDS-PAGE. Proteins separated via SDS-PAGE are transferred to 0.2 µm PVDF via a TransBlot Turbo transfer system (BioRad), blocked with 5% bovine serum albumin (BSA) and blotted with anti-HA 1˚Ab and anti-mouse 2˚ AP Ab (Invitrogen, 26183-1MG; LGC Clinical Diagnostics, 5220-0312) (B and D). A strong band at 36 kDa in the left-most lane on (B) identifies TK1984C, the positive control which did not transfer from SDS-PAGE to Western blot on (D). The tagged protein of interest (TK1978C, 19.7 kDa) is clearly visible in the pellet, CCL, and fractions B12 – C7 from the tagged-strain (D), while not visible in the FT, Wash, 5% B, or fractions B5 – B9 indicating that the tagged-protein bound to the Ni^{2+}-chelating column and was eluted following peak 1. TK1978C eluted in ~11 fractions (from ~145 – 195 mM imidazole or 72 – 97.5 % Buffer B) with the peak of the tagged protein eluting in C3 – 5 (at 175 – 185 mM imidazole, 87.5 – 92.5% Buffer B). A smattering of proteins in peak 1 bind the 1˚ or 2˚ anti-body non-specifically, producing a mild signal (lanes B7 – B10) for both (B) and (D).

While anti-His_6 antibodies exist and are effective epitopes in some species, this short sequence is not easily identified in *T. kodakarensis* cell lysates. Instead, a second tag encoding the broadly used hemagglutinin (HA) sequence is an effective epitope.[36,37] Primary antibodies are readily available for this short, nine amino acid tag and there is little off target biding of native proteins in *T. kodakarensis*. Additionally, the *T. kodakarensis* codon optimized sequence contains a restriction enzyme (*Bsp*E1) cut site. Digestion of the gene sequence at this cut site can aid in genetic techniques to confirm the presence of the tag sequence (Figure 2.7). Genomic addition of the affinity and epitope tag sequences at the 3' end of the gene of interest assures that the full protein is expressed. A collection of strains wherein an epitope and affinity tag sequence are expressed on a single of protein of interest in each strain can provide the means to generate a metabolic protein network via AP-MS.

Modified strains are cultured to generate biomass such that sufficient protein from purified fractions is available for proteomic mass spectrometry. Cultures growing at early- to mid-exponential are harvested and cells are pelleted (Figure 3.1B). Cell pellets are resuspended in a biologically neutral buffer before cells are lysed to generate a clarified cell lysate (CCL) containing a complex protein mixture resembling cytoplasmic contents.

Proteins are purified at native expression levels from the CCL in an unbiased approach using affinity purification where weakly interacting partners that may have a transient relationship or may link two more tightly held complexes together can be maintained (Figure 3.1C, Figure 3.3).[38] Affinity-tagged proteins bind to a charged column and are eluted with an increasing concentration gradient of imidazole which out competes binding to the divalent cation bound to the column matrix. The eluent is collected throughout the purification, and proteins in the eluent are visualized using SDS-PAGE and Western blotting. Anti-HA primary antibody binds to the epitope tagged protein and identifies specific fractions containing the protein of interest. Fractions containing the greatest abundance of tagged protein are pooled and quantified (Figure 3.2C-D).

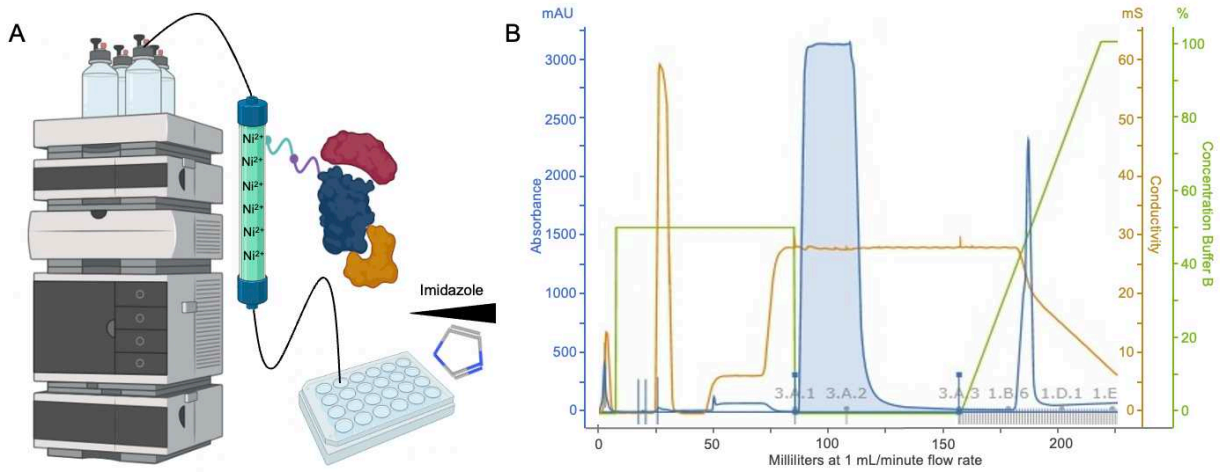


Figure 3.3: Affinity purification of a His_6-tagged protein on a Ni^{2+}-chelating column eluted using an imidazole gradient. An AKTA liquid chromatography system (A) (Cytiva Life Sciences) employs a regulated flow rate of various solutions while measuring pressure (not shown), conductivity (B, orange/brown), and absorbance (B, blue) at a specific wavelength. Here, UV_{280} is used to identify aromatic residues on amino acid side chains, but UV_{205} can be used to identify the peptide backbone if a specific protein lacking aromatic residues is of interest. A real-time chromatographic output (B) displays these measures at the specified flow rate. Purification of a His_6-tagged protein (A, dark blue protein with a purple epitope and teal affinity tag sequence) from a complex mixture such as a clarified cell lysate (CCL) can be performed using a Ni^{2+}-chelating column. The CCL flows across the column and the histidine sequence of the tagged-protein binds to this divalent cation. A buffer (Buffer B) containing imidazole is used in an increasing linear gradient (B, green) and will elute the tagged protein and associated proteins (A, red and orange proteins). 1 mL fractions of the eluent are collected and analyzed using SDS-PAGE and Western blot for the epitope tag, thereby identifying the protein of interest which is separated from the complex mixture of proteins in the CCL but maintains relevant protein associations.

Pooled proteins are TCA precipitated before shipment to a mass spectrometry collaborator. Upon arrival, proteins are resuspended and trypsin digested. Peptides are separated via reverse phase liquid chromatography before being ionized and flown through tandem MS detectors where fragmented peptide mass to charge ratios are compared to an unfragmented analysis (LC-MSMS), allowing for peptides to be sequenced and mapped to the *T. kodakarensis* proteome (Figure 3.1D).

Protein samples from the parent strain are prepared and analyzed in the same manner such that proteins identified in the parental strain are back subtracted from the tagged strain to reveal only statistically significant protein partners of the protein of interest (Figure 3.1E). Data

undergoes stringent analysis to identify co-purified protein partnerships. Strains encoding tagged proteins of known heterocomplexes can be used to validate results. Additionally, reciprocal protein identifications demonstrate exceptionally robust associations, adding confidence to the data.

To generate a metabolic protein interactome, twenty-nine strains were generated in which a gene was genomically modified to encode an epitope and affinity tag resulting in native expression levels of the respective tagged protein. Of these twenty-nine proteins, 15 are involved in catabolic reactions including glycolysis, amino acid fermentation, or the breakdown of small molecules into acids, twelve are involved in anabolic reactions including the production of reduced small molecules (*e.g.* NADH), mature lipids, or the generation of energy via respiratory complexes at the membrane. Additionally, two proteins, the rubrerythrins (RBR1, TK0650 and RBR2, TK0826), were included following preliminary data from POR and VOR complexes showing abundant interactions with RBRs.

Discussion of optimizing and troubleshooting the AP-MS method follows. It should be noted that while each step is discussed sequentially, the validity of all steps is dependent on the outcome of the LC-MSMS data, thus it may be helpful to the reader to become familiar with all steps and return to various optimization and troubleshooting sections as desired.

3.1.2 Strengths and limitations of AP-MS

AP-MS offers specific advantages in a genetically accessible organism with a small genome. A genetically modified strain which poses no challenge to fitness can be generated and sequenced confirmed in as few as ~three weeks.[36,37,39] *T. kodakarensis* doubles every ~40 minutes, thus cultures inoculated at 1:100 with freshly growing culture will generate sufficient biomass at mid-exponential phase in ~9-12 hours.[33,40] Genetically encoded epitope and affinity tagged proteins of interest are expressed from the genome at native expression levels. With a small 2 Mbp genome, a small proteome of ~2,300 proteins, and no known post-translational

modifications, mapping peptides from proteomic tandem mass spectrometry data is simple and reliable.

While the combinatorial steps of the AP-MS method can be applied to a strong majority of genes and proteins of interest, there are some inherent downsides to AP-MS. Here, a brief list of challenges (and their affected genes/proteins) will be described. These challenges include genetic accessibility, growth of the modified strain, purification of the tagged protein, identification of the purified protein, and identification of the tagged protein as significant compared to the control strain in proteomic MS data.

All genes of interest must be genetically modified, but not all regions of the genome permit modification (Section 2.5). While many strains were generated in the first transformation attempt, some strains required multiple attempts. Strains that could not be generated after ten attempts tagging the 3' (C-terminus) or ten attempts tagging the 5' (N-terminus) end of the gene were considered unviable, which was the case with two out of thirty-one strains (TK1218 and TK1222).

The addition of the tag must not hamper growth. Growth of a tagged strain compared to the parental strain (TS559) can identify if the modification poses an impact on fitness (Section 3.2.1). The majority of tagged strains grew as expected in both +S˚ and -S˚ conditions. Only one tagged strain (TK2091C) posed a fitness challenge as the culture reached early-exponential after ~>18 hours when the parental strain grew to early exponential in ~6 hours.

Purification requires that the tag is not cleaved or embedded in the protein (Section 3.2.2). One strain (TK1130C) was genetically confirmed, but the protein could not be identified via Western blot. Another strain was generated wherein the 5' end of the TK1130 gene was tagged (TK1130N). The protein encoding the N-terminal tag could be identified in purified fractions.

Some proteins could not be purified using a Ni^{2+} cation and the tagged protein of interest was found in the flow through or could not be identified on the Western blot (Section 3.2.2, *A*

choice of divalent cations for affinity purification). If the HA-tag could not be identified easily via Western blot, a stronger divalent cation (Cu^{2+}) was employed.[41]

Proteomic mass spectrometry detects peptides based on a mass to charge ratio (Section 3.2.3). Proteins which digest into peptides which are too small, too large, or maintain too much positive charge when ionized cannot be detected.[42,43] The abundance of proteomic data returned must be rationally curated. The application of appropriate statistical methods aids in resolving reliable data, but blind processing is not advised.

Lastly, it should be clearly noted that while some proteomic mass spec methods can provide quantitative data in a limited fashion, [29,44,45] the data independent acquisition (DIA) proteomic mass spec method employed here is not quantitative (Section 3.3).[43] Proteins which pass statistical thresholds in biological replicate when compared to controls must be interpreted using a presence or absence metric. Due to the biased nature of MS data collection, quantitative interpretation of peptide or protein abundance cannot be conflated with meaningful protein associations. The abundance of a peptide detected in a tagged sample compared to the control only suggests that the protein is present, not that this protein is highly associated with the tagged-protein of interest. Again, quantitative proteomic MS methods exists,[29,44,45] such as SILAC, but such methods are not appropriate for AP-MS.

The majority of genes/proteins applied to the AP-MS method in this study (25/29) were successful and returned data that easily met statistical thresholds and method validation parameters. The outliers which posed unresolved challenges include two genes of the MBS complex and four genes of the MBH complex. It should be noted that these complexes constitute large, membrane bound proton pumps which are responsible for the respiratory activity of *T. kodakarensis.*[46,47] Employing an alternative method for the MBH and MBS complexes is of interest and discussed at the close of this chapter.

3.2 Variables in the AP-MS method

Section 3.2 will focus on describing modifications and troubleshooting techniques for the physical parameters of all steps of the AP-MS method, excluding genetic techniques which are described in detail in Chapter 2. Using WGS confirmed strains, this section will walk the reader through generating sufficient biomass, purification strategies, preparing protein for shipment, and method options for optimal LC-MSMS performance.

As the validity of all steps relies on effective LC-MSMS data, the reader is encouraged to grasp all steps of the method by focusing on Section 3.1 in detail or skimming the entirety of this chapter before diving into the variables of the method discussed in Section 3.2

3.2.1 Generating biomass

Proteomic tandem mass spectrometry requires ~1 µg of trypsin digested protein, however, the MS collaborator asked for 20 µg of protein to be shipped providing ample sample for repeated MS runs, if necessary. Ideally, 20 µg of protein could be isolated from < 1 mL of affinity purified protein. Fortuitously, ~333 mL of culture grown to OD_{600} 0.4-0.6 provided sufficient protein for 20 µg in 200 – 600 µL using the purification and elution scheme described in Section 3.3.2.

Biological replicates

Growth of cultures in biological replicate is necessary to provide confidence in the AP-MS analysis. While greater numbers of replicates will provide greater statistical confidence, three biological replicates were considered sufficient for this study. Minimally, three biological replicates of all strains in +S˚ and -S˚ conditions were grown. WGS confirmed strains were cultured in 1 L batches, which were each considered a biological replicate. Each 1 L biological replicate was divided into three technical replicates of ~333 mL each before cells were pelleted and stored at -80˚C. A cell pellet from a single technical replicate provided 20 µg protein in <1

mL of purified protein. Three technical replicates per biological replicate (nine cell pellets from a single strain in a single condition [S° or non-S°] in total) often provided sufficient biomass to troubleshoot purification, SDS-PAGE, Western blot, and LC-MSMS of each strain without having to regrow biomass.

Technical replicates

Sample variability is a possible source of error, thus we wanted to confirm that technical replicates of a single tagged-protein that was purified and sent for LC-MSMS would provide similar data. To determine the consistency of technical replicates, two samples of TK1088C +S°, geranylgeranyl reductase (GGR), were independently pooled and TCA precipitated before being sent for analysis. Technical replicates were analyzed by AP-MS and compared to one another. If there is no difference between these two samples, there should be no significant difference between the proteins identified in each sample (Table 3.1).

Table 3.1: AP-MS results of significant proteins compared between technical replicates of TK1088C in sulfur conditions.

Protein	Molec. Weight	p-value	Quantitative Value (Total Spectra) Tech. Rep. 1	Tech. Rep. 2
Vitamin B12-dependent ribonucleotide reductase, TK1736	200 kDa	0.033	28	65
Phosphoglucomutase/ phosphomannomutase, TK1108	50 kDa	0.048	0	6

We would expect for technical replicates to show minimal differences, suggesting that the technique is repeatable producing the same results multiple times. Between the two technical replicates, only two were identified as significant: Vitamin B12-dependent ribonucleotide reductase (TK1736) and Phosphoglucomutase/phosphomannomutase (TK1108), both are large (200 and 50 kDa, respectively), thus many peptides will map to each protein, allowing for their frequent identification. False positive results will be discussed in further detail in Section 3.3.3, but overall technical replicates proved to be highly reproducible.

Refining the inoculation procedure

T. kodakarensis microbial cultures involve populations of individuals that double every ~40 minutes.[40] Mutation frequencies for such biological systems are high,[48] thus maintaining a single stock culture is prudent. A strain is passaged from a long-term stock culture to a starter culture and the freshly grown starter culture is used to inoculate the 1 L culture that will be harvested. The Santangelo lab has historically used a freshly grown over-night (~12 hours) starter culture to inoculate a 1 L culture at 1:100. After attempting to grow and harvest multiple strains at once and noticing a considerable difference in the length of time it took a 1 L culture to reach an OD_{600} of 0.4, the method required attention and improvement. *T. kodakarensis* doubles every ~40 minutes in optimal conditions (strictly anaerobic, with rich media, in the presence of inorganic sulfur).[40] As stock cultures vary in the number of viable cells, starter cultures vary in the number of cells grown in 12 hours, thus a 1:100 inoculation from a freshly grown starter could have twice the number of actively growing cells if the starter grew to an OD_{600} of 1.0 (typically assumed to equate to 1.0×10^8 cells) or 0.5 (1.0×10^4 cells). To improve the method, the optical density of a freshly growing starter culture was observed, and the volume used to inoculate a 1 L culture for harvest was adjusted. For example, a starter culture grown to OD_{600} 0.5 should be used at 1:50 to inoculate a 1 L culture for harvest to account for the difference in cells. Note that the final volume of this culture will be 1,020 mL instead of 1,010 mL if inoculated at 1:100. A 10 mL difference in 1,020 or 1,010 mL is ~1% and thus negligible.

3.2.2 Purification considerations

Extensive troubleshooting of purification conditions was appropriate and necessary. As this study follows Burkhart et al., 2019 in which the three Fds were tagged and purified to generate a Fd network, similar conditions were considered a sound starting point. As will be described here, sonication conditions to generate a clarified cell lysate, buffer composition, affinity matrix using various divalent cations, the number of fractions to pool, as well as the

differences in various liquid chromatography systems were tested and considered. Each variable and the optimally determined method will be discussed here.

Resuspending cells and resolving a clarified cell lysate

A whole or total cell lysate (WCL or TCL) would consist of complete cellular contents, inclusive of membranes, however, as all proteins tagged are either known to reside in the cytoplasmic fraction or are cytoplasmic subunits of respective membrane-bound complexes (membrane bound sulfane reductase, MBS, and membrane bound hydrogenase, MBH), and membrane fragments can clog a packed column matrix, this purification procedure utilizes a clarified cell lysate (CCL) where cells are lysed and membrane fragments are pelleted to isolate suspended cytoplasmic contents. Cell lysis via detergent was not considered a viable option for AP-MS as amphipathic solvents disrupt protein associations, thus sonication was the lysis method of choice. Pelleted cells from ~333 mL of harvested biomass (Section 3.2.1) are stored at -80˚C until being thawed on ice and resuspended in a buffer resembling the cytoplasm: 25 mM Tris HCl pH 8.0, 500 mM NaCl, and 10% w/v glycerol (commonly called the lysis buffer, and henceforth referred to as lysis buffer or Buffer A). Such a neutral buffer should maintain protein-protein interactions closely resembling *in vivo* states. Originally, the cell pellet was resuspended in Buffer A by various rounds of thaw at 85˚C and freeze at -80˚C over two hours, but such steps were determined to be unnecessary. Instead, the cell pellet was retrieved from the -80˚C freezer and put on ice, and ~ 10 mL of room temperature or 4˚C Buffer A was added. A p1000 pipette with multiple (~ two) stacked 1000 µL tips could be used to scrape the pellet into suspension in about 5 – 10 minutes on ice in the original 15 mL conical tube the cell pellet was stored. Stacking multiple tips onto the pipette provides sufficient length that the pipette will not encounter the CCL and remain clean. Historically in the Santangelo Lab, cells were lysed on ice using a sonicator at 50% power “on” for 10 seconds followed by “off” for 30 seconds for ~50-60 minutes. Such an extended sonication regime was tested against a shorter method of 5 minutes

at the same settings. Both methods resolved proteins which could be purified sufficiently, however the same tagged-protein eluted at a lower imidazole concentration when sonicated for the longer time period. (TK2077C was found to elute at 96-136 mM imidazole when sonicated for 50 minutes, when the same protein eluted at 120 – 200 mM when sonicated for only 5 minutes.) It can be inferred that an increase in sonication time may allow for greater accessibility of the His_6 tag to bind to the column, but it should be noted that this was not tested in replicate or with multiple tagged strains. As 5 minutes is sufficient for cell lysis and a tagged protein can be effectively purified from the CCL, a 5 minute sonication regime was retained to both speed the purification process and retain as many bona fide interactions as possible.

Overview of affinity purification and elution schemes

Various proteins are known to require specific elution protocols, however it was the goal of this project to find a single repeatable scheme to minimize variability in the data. In the case of affinity purification and a His_6-tagged protein within a CCL, the column is charged with a divalent cation (Ni^{2+} is a common choice).[41] Proteins with no affinity for Ni^{2+} or the tagged protein will flow through the column and be collected in the “flow through” (FT) or “wash” steps (Figure 3.4 E and F). The His_6 residues on the tagged protein binds to the charged matrix. Proteins with natural affinity for Ni^{2+} will also bind (Figure 3.4, H) and both His_6- and natural affinity binding proteins require elution based on a competitive solvent. The buffers and elution scheme require that the tagged protein be resolved in fractions that are separate from the proteins with natural affinity for the matrix. Ideally, the tagged protein will have greater affinity, thus proteins with natural affinity will elute first and the tagged protein will elute later, following the UV spike during elution, termed “peak 1” (Figure 3.4H).

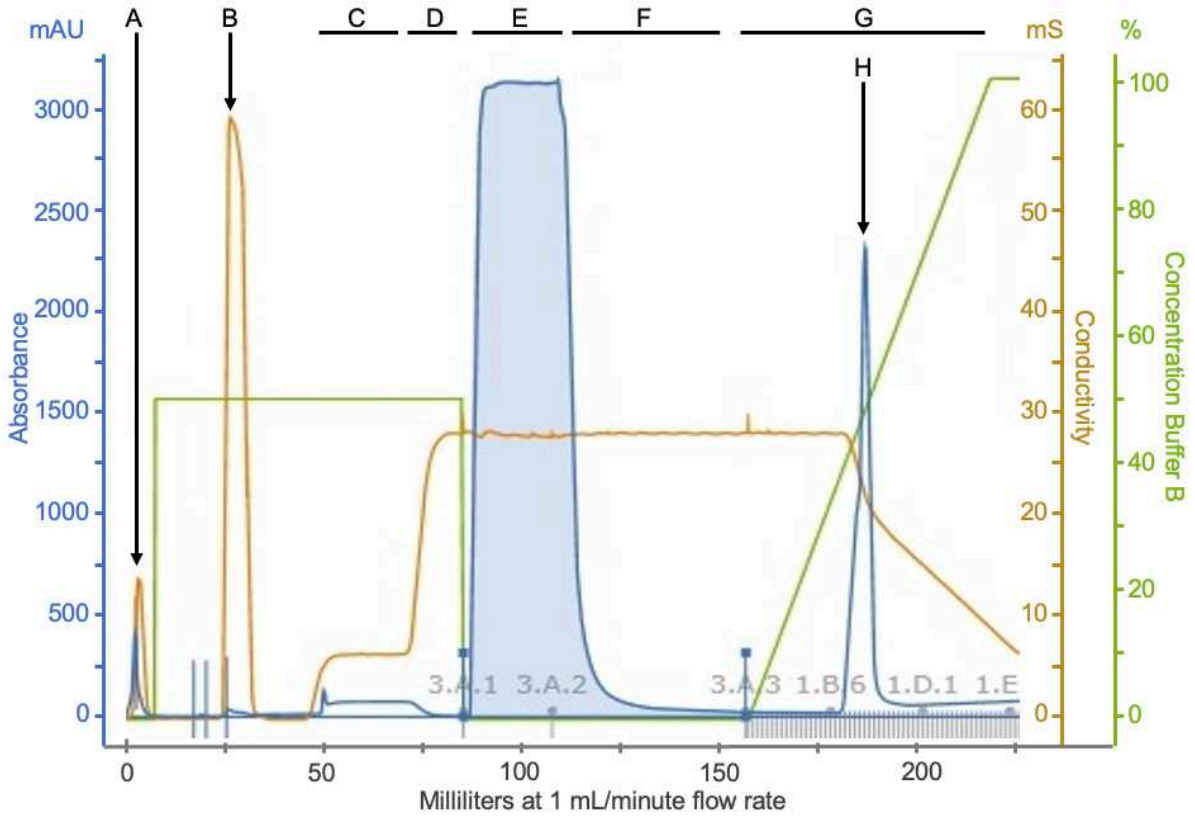


Figure 3.4: Affinity purification scheme. Affinity tagged proteins encoding His_6 sequences are purified from a CCL using a Ni^{2+}-charged chelating column. Columns are prepared by assuring all protein is removed by flowing two column volumes of 6 M Guanidine HCl across the column (not shown on the chromatogram, but a large spike in conductivity, orange/brown, will be seen). All metal is removed by flowing two column volumes of 100 mM EDTA (A) across the column. Columns are then charged with Ni^{2+}; two column volumes of 100 mM $NiSO_4$ sufficiently saturates the column matrix (B). The column is then prepared by flowing five column volumes of Buffer B (C) followed by five column volumes of Buffer A (D) before flowing the 0.2 µm filtered, CCL across the column (E). In this scheme, Buffer B contains 200 mM imidazole and 100 mM NaCl which produces a UV_{280} signal of ~100 mAU (blue) and conductivity signal of ~5 mS (orange/brown), respectively (C). Buffer A contains 0 mM imidazole and 500 mM NaCl, which is observed as the mAU returns to 0 and the conductivity rises to ~27 mS, respectively (D). Protein in the CCL that does not bind to the column (commonly referred to as the "flow through," FT) is observed in the wide, flat UV_{280} peak >3,000 mAU and ~15 mL are collected (fraction 3.A.1). The column is washed (F) with Buffer A and ~15-20 mL are collected (fraction 3.A.2) until the UV_{280} signal returns to baseline. Proteins are eluted (G) in this scheme using a linear gradient of increasing Buffer B concentration (from 0-100%), however an isocratic or combination of isocratic and linear gradients can be employed. During elution, fifty-eight 1 mL fractions are collected (fraction 3.A.3 -1.E.12). Native proteins with a natural affinity for the Ni^{2+}-charged matrix are eluted *en masse* at ~50% Buffer B (~100 mM imidazole), however proteins which possess a His_6 tag are often of greater affinity, eluting at higher concentrations (55-80% Buffer B) but are often not sufficiently abundant to produce a notable UV_{280} signal. The tagged protein of interest will likely elute in fractions 1.B.5 - 1.D.6, of which 900 µL are transferred to 1.7 mL tubes and stored at -80˚C to reduce protein degradation while the remaining 100 µL is stored at 4˚C and used for SDS-PAGE and Western blotting. Flow rate remains consistent at 1 mL/minute throughout the purification. Concentration of Buffer B (green) is set at 50% from ~5-

80 mL because both flow paths are rinsed with Buffers B and then A to assure that tubing from the Buffer B flow path contains Buffer A until the elution gradient begins. If the Buffer B flow path is not prepared as described, the elution would begin with ~5 mL of 50% Buffer B before decreasing to ~5% Buffer B. Such a jump in Buffer B concentration will impact protein elution. Flow paths are submerged in their respective buffers before the CCL is flowed over the column and remain here until the elution ends. In full, Buffer A is composed of 10 mM Tris HCl pH 8.0, 500 mM NaCl, and 10% w/v glycerol. Buffer B is composed of 10 mM Tris HCl pH 8.0, 100 mM NaCl, 200 mM imidazole, and 10% w/v glycerol. Note i: This figure does not depict the actual purification scheme used for the purification of all tagged proteins in this study. Tagged proteins in this study were washed with 5% Buffer B for 15 mL (collected) after the wash step and before the elution. Tagged proteins were eluted from 5 - 100% Buffer B in 38 mL (not 58 mL, as depicted in this figure). Note ii: Unicorn software used for analyzing the ÄKTA chromatographic profiles has poorly integrated the area under the peaks, shown in light blue. These data, while useful for measuring protein abundance in other purification scenarios, are unimportant in preparing samples for AP-MS analysis and can be ignored.

A wide variety of buffers are used for affinity purification. In the case of maintaining native protein-protein associations, it is desirable to minimally impact solvent conditions while effectively eluting the tagged protein of interest. As described above, Buffer A intends to mimic the density and charge properties of the cytoplasm. Buffer B, the elution buffer, should maintain similar conditions, but must contain an agent to outcompete the affinity of the His_6 tag to the divalent cation on the column matrix. Imidazole is the R-group of histidine; thus, an increasing concentration of imidazole will provide a competitive environment. Various concentrations of imidazole ranging from ~100 – 500 mM are often used to outcompete His_6 affinity for divalent cations.

The elution scheme refers to the way in which the concentration of the two buffers is applied after the CCL is loaded over the column and the UV signal returns to 0 mAU (Figure 3.4 E-G). Up until the beginning of the elution (Figure 3.4G), the system is flowing with Buffer A. The green line in Figure 3.4 depicts a linear gradient of increasing concentration of Buffer B during the elution, however isocratic steps whereby the concentration of Buffer B increases by 10% every 10 mL (or two column volumes if using a 5 mL column), for example, are also common.

Imidazole content in the elution buffer

200 mM imidazole in Buffer B was used to purify the tagged Fds (Burkhard et al., 2019) and initial trials followed this method, however we wanted to determine which strategy would be ideal for multiple proteins in this study. Two tagged proteins (TK1685C and TK2077C in +S˚) were tested with linear elution schemes using 200 mM imidazole and four tagged proteins and the parental strain (TK1325C, TK1326C, TK1685C, TK1979C, and TS559 in +S˚ or -S˚) were tested with various elution schemes using 500 mM imidazole in Buffer B. Only TK1685C in +S˚ conditions was trialed in both imidazole concentrations, and interestingly, peak 1 of TK1685C +S˚ CCL began eluting at ~38% Buffer B with 200 mM imidazole (76 mM imidazole) after 21 mL and ~36% Buffer B with 500 mM imidazole (180 mM imidazole) after 26 mL suggesting that the concentration of imidazole in Buffer B is negligible above a certain lower limit (not tested below 200 mM imidazole). By increasing the rate at which Buffer B concentration increases (*e.g.* 0 - 50% Buffer B in 25 or 35 mL), peak 1 will elute earlier the faster that Buffer B increases. If the elution begins with Buffer B at 50%, peak 1 will begin to elute in ~10 mL.

It is ideal to reduce the number of fractions in which peak 1 or the tagged protein elutes. In all purification schemes trialed here, peak 1 eluted in ~ seven fractions (or 7 mL as the flow rate is 1 mL/minute). Interestingly, the tagged protein of interest eluted after peak 1 when 200 mM imidazole was used but generally eluted within peak 1 when 500 mM imidazole was used. As it is crucial to separate the tagged protein of interest from the bulk of proteins in the CCL, 200 mM imidazole was chosen as the desired concentration of Buffer B for this study.

Linear versus isocratic elution schemes

Linear and isocratic step elution schemes were tested at 200 mM imidazole Buffer B. Two tagged proteins (TK1685C and TK2077C) were purified using a linear gradient of 0 - 200 mM imidazole (0 - 100% Buffer B). TK1685C CCL eluted peak 1 in ~ six fractions at ~76 mM imidazole (38% Buffer B). On the same day, using the same buffers, TK2077C CCL eluted peak

1 in ~ six fractions at a lower concentration of imidazole (~58 mM imidazole, ~29% Buffer B). To test this pattern, an isocratic step gradient was used wherein the concentration of imidazole increased from 0 mM to 50 mM and maintained 50 mM for 10 minutes, then increased to 60 mM and maintained at 60 mM for 10 minutes, then increasing 70 mM and maintained at 70 mM for 10 minutes before employing a linear gradient from 70 mM to 200 mM imidazole. Peak 1 for both TK1684C and TK2077C resolved clearly during the 60 mM imidazole step in ~6 fractions.

In both the linear and step schemes described above, TK1684C tagged protein eluted following peak 1 at 100-136 mM imidazole, however TK2077C tagged protein eluted beginning in peak 1 and extending past from ~60 – 136 mM imidazole (38 – 68% Buffer B). It is clear that different tagged proteins will elute from the column at different concentrations of imidazole.

Ideal affinity purification method and elution expectations of His_6 tagged-proteins

Ultimately it was determined that a step of 5% Buffer B (10 mM imidazole) would be used for 15 mL following the wash step to release loosely held proteins from the Ni^{2+}-charged matrix. A linear gradient of 5 – 100% Buffer B containing 200 mM imidazole would be used to elute protein over 38 mL providing an increase of 5 mM imidazole for each 1 mL fraction. Using this scheme for the twenty-nine strains in both +S˚ and -S˚, peak 1 consistently eluted off the column in ~5 mL beginning at 80 mM imidazole (~ 40% Buffer B). Tagged proteins eluted at varying concentrations of Buffer B. Some eluted within peak 1 at ~ 80-90 mM imidazole (TK1980C +S˚, TK2089C +/-S˚, TK2090C +/-S˚, and TK2093C +/-S˚) and two eluted near the end of the elution at ~ 180-190 mM imidazole (TK0650C +/-S˚ and TK1981C +/-S˚). The majority of tagged proteins eluted between 130-170 mM imidazole (~65-85% Buffer B).

A choice of divalent cations for affinity purification

Some proteins could not be purified using a Ni^{2+}-charged column. In such a case, the tagged protein remained in the flow through suggesting that either the His_6 tag had been

cleaved or that there was not sufficient affinity to Ni^{2+} for the tagged protein to bind. HiTrap Chelating HP columns (GE Healthcare, 17-0409-03) contain a highly crosslinked spherical agarose substrate with an iminodiacetic acid chelating group which can bind divalent cations of which four are commonly used in affinity chromatography with varying strengths and purification outcomes; greater affinity results in less purity. From the strongest affinity force and least pure separation to the weakest and purist, Cu^{2+}, Ni^{2+}, Zn^{2+}, and Co^{2+} can be used (Figure 3.5 upper right inset).

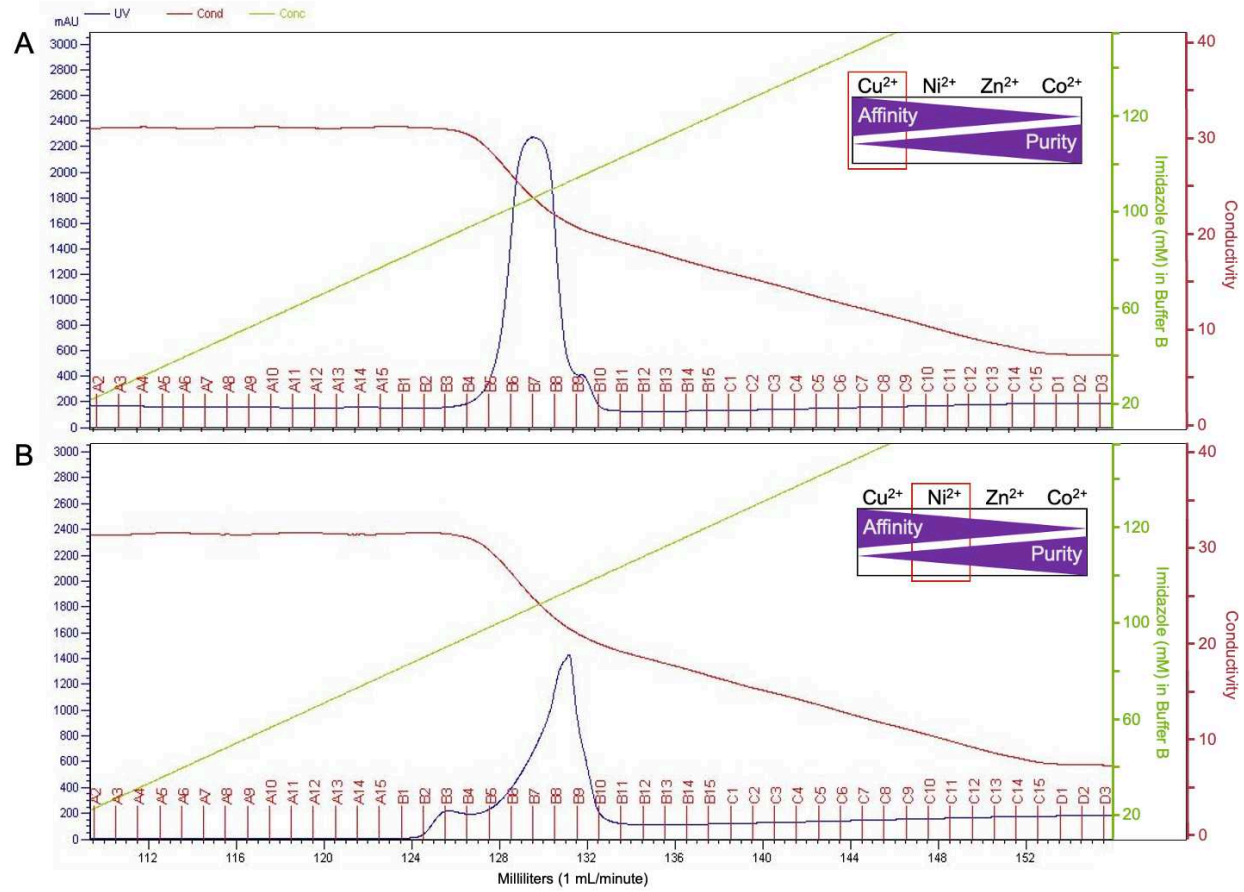


Figure 3.5: Elution patterns for a His_6-tagged protein when using Cu^{2+} or Ni^{2+} affinity matrices. Copper provides an alternative to nickel as a divalent cation that can be used for affinity liquid chromatography on an ÄKTA purification system. A clarified cell lysate (CCL) flows over the metal-chelated column at 1 mL per minute in Buffer A and proteins with affinity for the divalent cation bind, while proteins with no affinity flow past and are collected in the wash (not shown). A linear elution gradient of 200 mM imidazole in Buffer B outcompetes binding for the divalent cation with increasing concentration of imidazole (green line) based on a percentage of Buffer B vs. Buffer A flowing over the column. Buffer B contains 200 mM imidazole, thus 50% Buffer B contains 100 mM imidazole. Proteins are eluted from 5-100% Buffer B in 38-1 mL fractions (red, A2 - C8), resulting in a 5 mM imidazole or 2.5% increase in Buffer B for each

subsequent fraction. Fractions C9 - D3 show elution at 100% Buffer B. Ultraviolet light at 280 nm, which is absorbed by amino acids with aromatic rings, provides a signal for protein concentration in milli-absorbance units (mAU, blue line). The bulk of proteins which have a natural affinity for Cu^{2+} (A) elute beginning at ~95 mM imidazole, peaking at fractions B7 and B8 (red) where the UV_{280} signal reaches ~2250 mAU. Proteins bound to the Ni^{2+} column (B) begin eluting ~85 mM imidazole (fraction B2), however a wider elution pattern for peak 1 does not show the majority of proteins eluting until ~120 mM imidazole (fraction B8 - B9). These UV traces correspond nicely to protein concentrations observed in Figure 3.4C. Purifications shown here were performed on the "old" ÄKTA chromatography system.

Experimentation with Cu^{2+} provided a comparison in purity and elution characteristics with multiple divalent cations (Figures 3.5 and 3.6). When tagged proteins that remained in the flow through using a Ni^{2+} affinity agent (Figure 3.6B, FT) were purified using Cu^{2+} (Figure 3.5A), they could be identified in the fractions during elution (Figure 3.6A, B9-C1). Even when loading a more concentrated CCL onto a Ni^{2+}-chelating column (Figure 3.6C, CCL), less protein binds a Ni^{2+}-chelating column than a Cu^{2+}-chelating column (Figure 3.6C, FT). Additionally, the greater amount of bound protein observed in peak 1 was eluted ~1-2 fractions earlier from the Cu^{2+} matrix (Figure 3.6C, B6 and B7) than it was from the Ni^{2+} matrix (Figure 3.6C, B8). Neither Zn^{2+} nor Co^{2+} substrates were tested as they provided even less affinity than Ni^{2+}, which was not desirable for the parameters of this project but can be kept in mind if purifications of greater purity are desired.

As depicted in Figure 3.6B, some tagged proteins could not be identified using a Ni^{2+} purification scheme and were better resolved using Cu^{2+}. These proteins include all tagged-proteins of the membrane bound hydrogenase (MBH) complex, TK2088C, TK2089N, TK2090C, TK2091C, TK2093C, and glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR), TK2163C which were purified using Cu^{2+}, with the exception of TK2088C which could not be identified via Western blot from Ni^{2+} or Cu^{2+} purifications. An explanation of why TK2088C could not be identified follows in the discussion (Section 3.5, *Alternative approaches to determine MBS and MBH protein networks*).

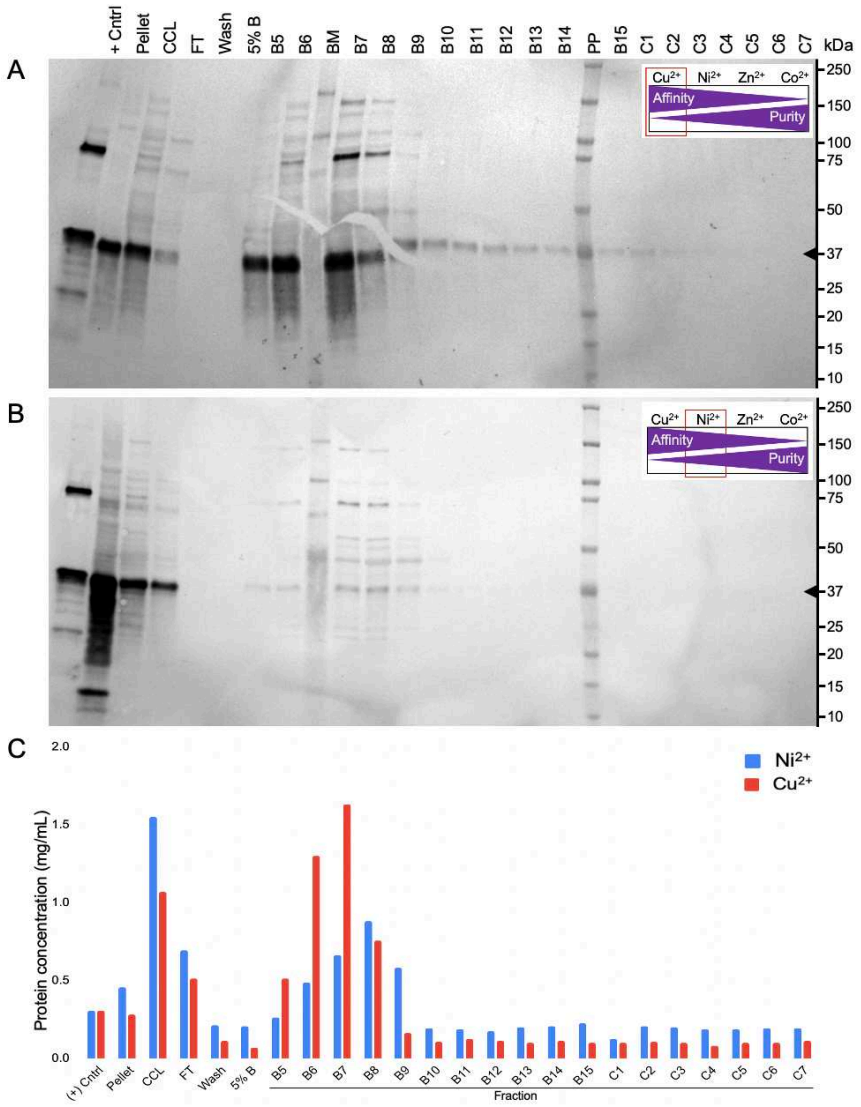


Figure 3.6: His_6-tagged protein purified on an Cu^{2+}- or Ni^{2+}-chelating column visualized via Western blot and corresponding protein concentration. A His_6-tagged protein (~37 kDa) is purified via liquid chromatography on an ÄKTA purification system using Cu^{2+}- (A) or Ni^{2+}- (B) chelating columns. Cu^{2+} provides the greatest affinity, but lowest purity within a protein purification (inset, upper right). 10 μL of sample is denatured by SDS and heat-treatment and loaded into each well in a 4 - 20% Criterion TGX Stain-Free gel. Proteins are separated in a standard tris-glycine buffer (1X Laemmli) for 50 minutes at 150 V (not shown). Proteins separated by SDS-PAGE are transferred to 0.2 μm polyvinylidene fluoride (PVDF) membrane using standard settings in a Turbo Blot Transfer system (BioRad) and visualized via Western blotting employing the appropriate 1˚ and 2˚ antibodies. Here anti-HA 1˚ antibodies made in mouse bind to the HA epitope encoded on tagged 35N protein and an anti-mouse conjugated alkaline phosphatase 2˚ antibody is used for chemiluminescent detection. The positive control (+ Cntrl, left-most lane) is a 1:1 mixture of purified TK1984C (36 kDa) and TK2076C (80 kDa). The tagged protein appears at 37 kDa in the FT in (B) when purified with Ni^{2+}, but this band is considerably fainter in (A) when purified with Cu^{2}, suggesting that the protein binds more strongly to the Cu^{2+} cation than to the Ni^{2+} cation. The ~37 kDa protein of interest is clearly purified in fractions B10 – C1 in (A) while these bands are difficult to distinguish in (B). Protein abundance (C) as measured by the Qubit 4 Fluorometer (Thermo Fisher Scientific) shows that

while a greater amount of protein was loaded over the Ni^{2+} column (C, CCL), more protein was found in the FT, wash, and 5% B steps and less in the peak 1 elution (B5 - B9) for Ni^{2+}, supporting that Cu^{2+} provides greater attractive force. Interestingly, proteins in the Cu^{2+} purification elute at a slightly lower concentration of imidazole seen by the apex of peak 1 occurring at fraction B7 for Cu^{2+} while the apex of peak 1 occurs at fraction B8 for Ni^{2+} (C). 10 mL (two column volumes) of 20 mM $CuSO_4$ or 500 mM $NiSO_4$ was used to chelate the 5 mL column for respective purification schemes.

Avoid water in flow paths during purification

Protein structure and associations are dependent on chemical parameters of the environment, such as the ionic strength of the solvent, thus, 500 mM NaCl is used to maintain appropriate ionic conditions in Buffer A. Imidazole maintains charge, thus to account for the charge contributions of imidazole in Buffer B, only 100 mM NaCl is added. In preparation for a purification, the flow paths are fully saturated first with Buffer B and then with Buffer A. It is crucial that the injection port be flushed with Buffer A as a decrease in conductivity due to a few microliters of water left in the injection valve can affect the outcome of the purification (Figure 3.7).

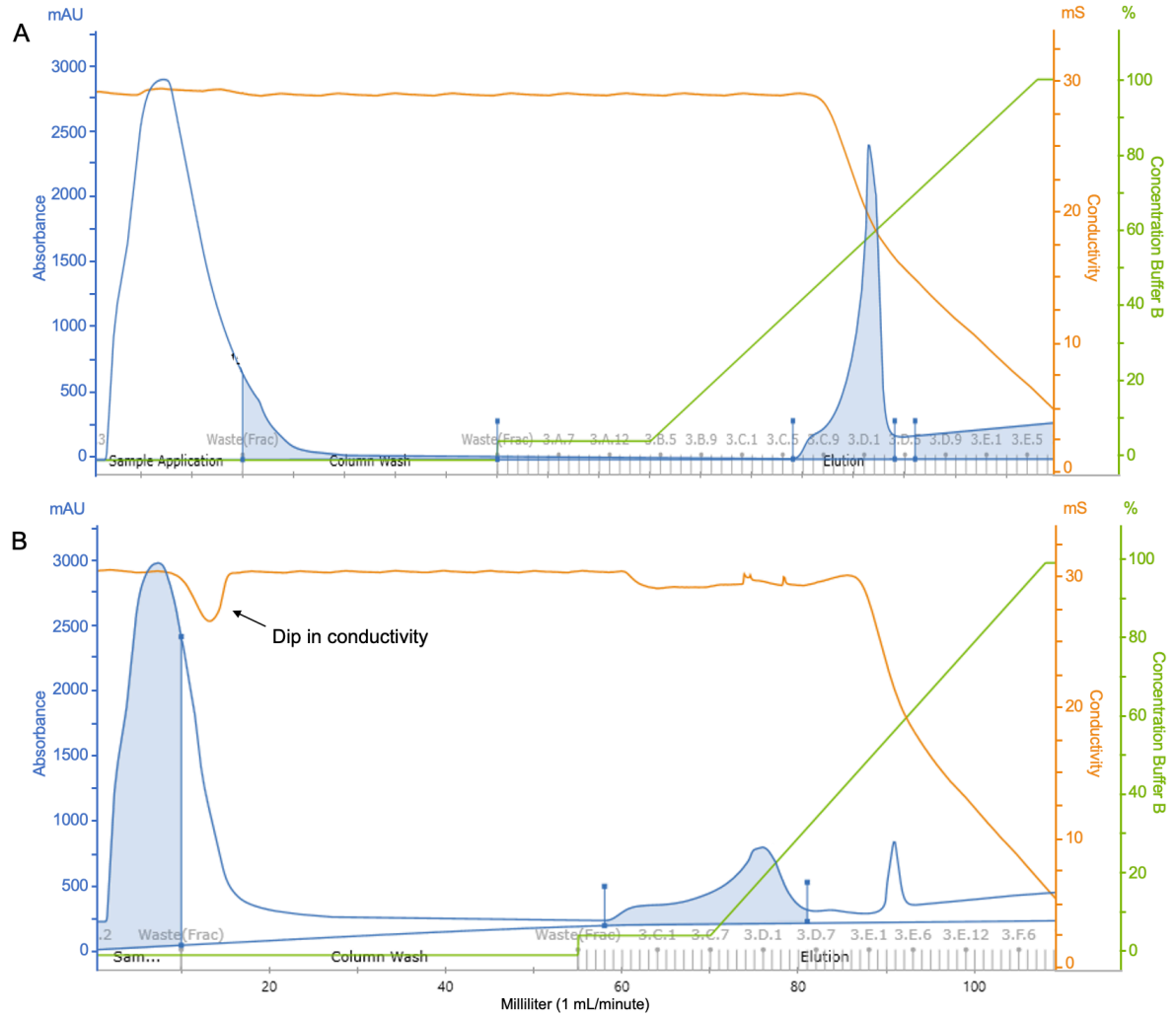


Figure 3.7: Affinity purification of a His_6-tagged protein and the elution pattern when water is in the flow path during the wash step. Affinity purification scheme of a His_6-tagged protein (TK0136C) using a Ni^{2+}-chelating column. The column and flow paths are saturated with Buffer A when the CCL is loaded over the column (left-most peak). The column is typically washed with 25 mL of Buffer A (here ~40-50 mL of Buffer A was used in both (A) and (B)). Note that a dip in conductivity from ~30 mS to ~27 mS is visible in (B) while conductivity remains steady at ~30 mS in (A) until peak 1 elution (~80 mL). Also note that the elution profile of peak 1 occurs at 40-60% Buffer B in a standard single peak, while the elution of the same tagged protein in (B) results in a wide peak that begins during the 5% B wash step and a second small peak follows later (B). This dip in conductivity can be attributed to water entering the system. When the inlet valve was switched from “inject” back to “load,” it is likely that the “load” flow path contained water instead of Buffer B. While the TK1036C protein was purified sufficiently in (B) as visualized by Western blot (not shown), the AP-MS data from (B) deviated significantly from the two biological replicates that did not experience a dip in conductivity. AP-MS data from (B) did not identify TK0136C protein as significant compared to control. These samples were discarded, and the purification was repeated. Note: Unicorn software used for analyzing the ÄKTA chromatographic profiles has poorly integrated the area under the peaks, shown in light blue. These data, while useful for measuring protein abundance in other purification scenarios, are unimportant in preparing samples for AP-MS analysis and can be ignored.

Pooling fractions and matching experimental and control samples

Most His_6-tagged proteins in this study eluted in ~3 - 12 fractions. To limit identifying extraneous proteins and realizing that sufficient protein for LC-MSMS (20 µg) was present in a single 1 mL fraction, we chose to pool three fractions (on average) in which identification of the tagged protein on the Western blot was the strongest (Figure 3.2D, fractions C10-12). Control samples were pooled in such a way as to represent the same fractions that were pooled in experimental samples. For example, samples from C1 - C3 from both the control and experimental sample fractions were compared. Similarly, if the tagged-protein eluted later, as in the case for TK1978C in Figure 3.2, C10 - 12 from both the experimental and control purifications were compared.

Peaks of a bimodal purification do not produce equivalent results

The majority of tagged-proteins eluted after peak 1 in a Gaussian distribution (Figure 3.2D), however one tagged-protein, a 4Fe-4S cluster-binding protein associated with formate dehydrogenase ($FDH\gamma_2$, TK2075C -S˚) eluted in a unique bimodal distribution (Figure 3.8) with an abundance of tagged-protein within peak 1 (fractions B7 - 9) and an abundance of tagged-protein eluting after peak 1 (fractions B13 - C1) with a visual decrease of tagged-protein in between (fractions B11 - 12) these bimodal regions of abundance occurred for all three biological replicates, but was only observed in -S˚ conditions and was not observed in +S˚ (nor with any other purification of a tagged protein in this study). Such a unique elution pattern allowed for us to test if purifying a protein within peak 1 was sufficient or if separation from peak 1 was preferred. TK2075C -S˚ samples with respective fractions from a TS559 -S˚ control were analyzed and it was determined that purification outside of peak 1 was required to identify TK2075C -S˚ with statistical relevance against a TS559 -S˚ control. It should be noted that TK2075C +S˚ eluted in fractions B14 - C1.

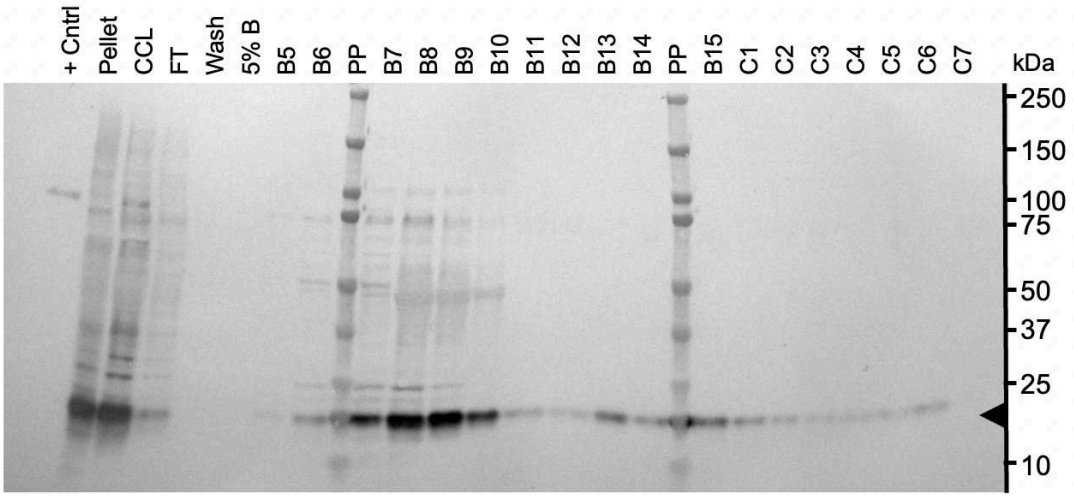


Figure 3.8: Unique purification patterns show bimodal elution of the tagged protein in a single purification. The CCL of a tagged-strain is affinity purified using a Ni^{2+}- chelating column and an imidazole gradient as described in detail in Figure 3.4. Purified fractions are denatured and separate in a 4 - 20% Criterion TGS stain-free gel (BioRad, 5678095) in a tris-glycine buffer (1X Laemmli) at 150 V for 50 minutes. Proteins separated via SDS-PAGE are transferred to 0.2 µm PVDF via a TransBlot Turbo transfer system (BioRad), blocked with 5% bovine serum albumin (BSA) and blotted with anti-HA 1˚Ab and anti-mouse 2˚ AP Ab (Invitrogen, 26183-1MG; LGC Clinical Diagnostics, 5220-0312). Precision Plus Stained Dual-Color (PP) protein ladders are loaded in the 9th and 18th wells of this 26-well gel to aid in identifying protein size (BioRad, 1610375). Purification of the tagged 4Fe-4S cluster binging subunit associated with FDH (TK2075C, ~18 kDa) protein has eluted in a bimodal distribution. The tagged protein is visible in fractions B7 - 9 and again in fractions B13 - 15, with less tagged protein apparent in lanes B11 - 12. Such a discrepancy could be due to inconsistent transfer from SDS-PAGE to the PVDF membrane, but this pattern was observed in all three biological replicates for this tagged protein in -S˚ conditions. This tagged protein did not follow this pattern in +S˚ conditions where the tagged protein was only visible in fractions past peak 1 (B13 - 15, not shown). Fractions from both areas of tagged-protein abundance were individually pooled and sent for LC-MSMS analysis and the tagged-protein was only found to be significant in the fractions which eluted past peak 1 (B13 -15).

Unique purification patterns between sulfur and non-sulfur conditions

With the exception of TK2075C (discussed above) and TK1980C (discussed here), all tagged-proteins in +S˚ and -S˚ conditions displayed generally concurrent purification profiles with the tagged-protein of interest eluting in similar fractions in each condition. TK1980C, the α subunit of VOR, was the exception showing a different elution profile in +S˚ and -S˚ conditions (Figure 3.9). TK1980C +S˚ eluted as a dimer in peak 1 (fractions B7 - 9), while TK1980C -S˚ eluted primarily as a monomer eight fractions later (C2 - 4) at ~40 mM greater imidazole. AP-MS

results showed that the TK1980C appeared significant compared to respective TS559 control samples in both +S˚ and -S˚ indicating that some proteins can be purified within the bulk of protein in peak 1 successfully.

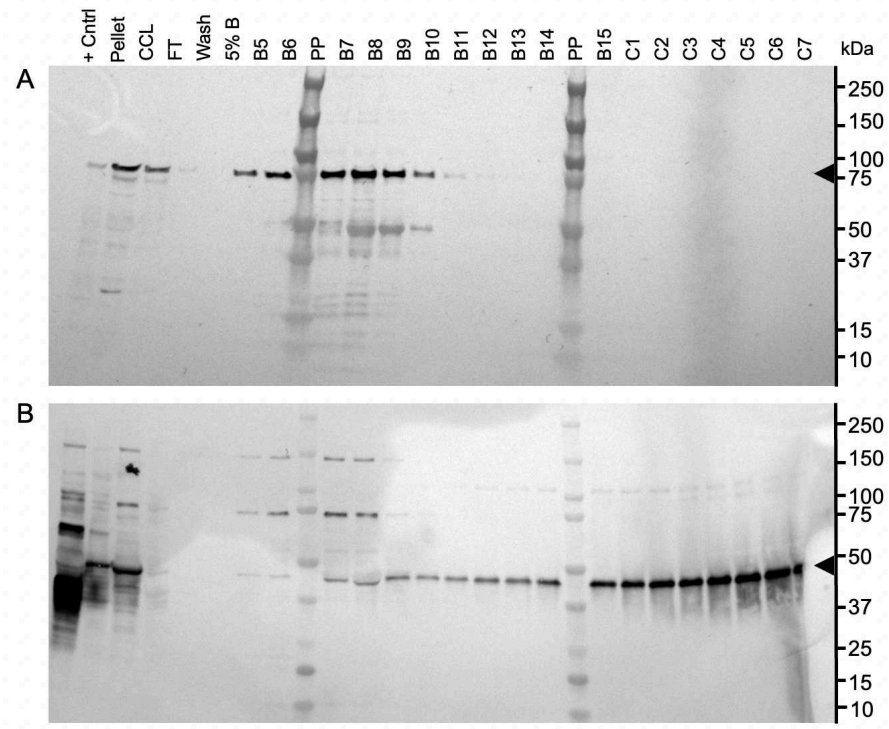


Figure 3.9: Purification of the same tagged protein in +S˚ and -S˚ produces different purification patterns. The CCL of a tagged-strain is affinity purified using a Ni^{2+}-chelating column and an imidazole gradient as described in detail in Figure 3.4. Purified fractions are denatured and separate in a 4 - 20% Criterion TGS stain-free gel (BioRad, 5678095) in a tris-glycine buffer (1X Laemmli) at 150 V for 50 minutes. Proteins separated via SDS-PAGE are transferred to 0.2 µm PVDF via a TransBlot Turbo transfer system (BioRad), blocked with 5% bovine serum albumin (BSA) and blotted with anti-HA 1˚Ab and anti-mouse 2˚ AP Ab (Invitrogen, 26183-1MG; LGC Clinical Diagnostics, 5220-0312). Precision Plus Stained Dual-Color (PP) protein ladders are loaded in the 9th and 18th wells of this 26-well gel to aid in identifying protein size (BioRad, 1610375). Here, tagged-VORα (TK1980C, 44 kDa) grown in +S˚ conditions (A) is purified as a dimer (~88 kDa) and elutes much earlier in the purification (with peak 1, fractions B5 - 9). Tagged- VORα grown in -S˚ conditions (B) appears as a monomer (~44 kDa) predominantly in fractions B15 - C7. Pool protein sent for LC-MSMS analysis returned TK1980C as significant in the respective fractions, suggesting both purifications were effective. Note that proteins of larger sizes are eluted and recognized by the anti-HA antibody in (B) at ~110 and ~130 kDa in the -S˚ purification only. While it can be concluded that these bands are not dimers, the band at ~130 kDa could be a trimer of the tagged-protein, but this does not explain the double banding pattern at these high kDa sizes fully. Instead, it is supposed that strongly associated interacting partners remain in association with the tagged protein even under denaturing conditions. Such purification patterns (where larger bands purified with the tagged-protein) were observed for other proteins such as IORα, TK0136C. However, only VORα showed such a disparate purification between +S˚ and -S˚ conditions.

One to four biological replicates compared

The original framework for this project proposed that three biological replicates of control and experimental samples would be compared. As strains were generated, proteins were successfully purified, and LC-MSMS data was returned, it became apparent that pooled fractions from various stages in the elution profile were not comparable (as explained in Figure 3.9 with TK2075C -S˚). Thus, it was prudent to pool equivalent fractions of the control samples to match fractions pooled for the experimental sample which significantly increased the number of control purifications and the number of samples sent for LC-MSMS analysis. As each LC-MSMS sample run cost $280, it was determined to be valuable to ascertain the benefit of one, two, or three control replicates. To test this, PORδ (TK1982C +S˚) was chosen. Four fractions where the tagged-protein eluted in the greatest abundance were pooled (B10 - 13), and three control samples of three fractions each which overlapped the tagged-protein samples (B9 - 11 or B12 - 14) were prepared. It may be noted that each control sample included one fraction that was not included in the tagged-protein sample (B9 or B14). As explained in detail in Section 3.3.3, every identified protein requires an independent test, thus AP-MS data will return different proteins if different samples are included. The question becomes, does an analysis with one control differ significantly from an analysis with three controls? In the case of TK1982C +S, when three experimental replicates were compared to three control replicates, twenty-two proteins were identified as significant. When control replicate #1 was used, nine proteins were found to be significant and all nine proteins were included in the twenty-two identified with all three control replicates. When control replicate #2 was used, fifteen proteins were found to be significant, eight of which were identified with the three replicates and seven of which were novel. When control replicate #3 was used, the exact same twenty-two proteins were identified as when all three replicates were included. While replicate #1 and #2 excluded some proteins and #2 included some additional proteins, the data with only a single replicate still provided meaningful information (inference is discussed in Section 3.4). Additionally, TK1982 is part of a

heterotetramer. All four subunits of this complex were identified as significant in all tests (explained in detail in Section 3.4.2). While multiple control samples will remain the gold standard, some samples were compared to a single control replicate when more control samples would require considerably more time or money. Out of sixty-one comparisons preformed, thirty-five comparisons included three control replicates, seven comparisons included four or two control replicates, and twelve comparisons included one control replicate. Four replicates were included if control samples for nearby fractions had already been sent for analysis.

Liquid chromatography instrument variability

Proteins purified using various ÄKTA liquid chromatography systems (GE Healthcare) may return different results. The breadth of this project required over 200 usable purifications (not including purifications to troubleshoot the technique). The Santangelo lab was equipped with two ÄKTA chromatography units until late-summer of 2023 when two ÄKTA GO units were purchased. The two ÄKTA units used for the bulk of this work consist of an ÄKTA Pure and an older AKTA FPLC (a system that demands creative repair work). It seemed prudent to determine whether purifications on different liquid chromatography systems yielded similar AP-MS results or if replicates of a certain group should be performed on a single unit. Not including the ÄKTA systems purchased in 2023, the two systems used will henceforth be referred to as the Pure and old ÄKTA systems, respectively. TK1979C +S, the δ subunit of 2-oxoisovalerate-ferredoxin oxidoreductase (OGOR) was purified on both the Pure and old systems and no meaningful differences were seen in the chromatogram or Western blots. Differences in AP-MS data are bound to occur (which will be discussed in detail in Section 3.3.3), but in general, protein purifications preformed on different ÄKTA systems returned similar results with minimal variability. Analyzed against a TS559 control purified on the old system, TK1979C +S˚ returned similar results from experimental samples purified on the old system or the Pure. When

TK1979C +S˚ purified on the old system was compared to the parental control purified on the old system, 12 - 18 proteins were identified as significant between the three experimental replicates. Sixteen proteins were identified as significant when TK1979C +S˚ was purified on the Pure system, with 10/16 identified in all other replicates, 4/16 identified in two other replicates, and only two proteins identified as significant that didn't appear as significant in any other sample. This is well within the range of experimental variability; thus, it is assumed that the various purification systems preform the same.

It should be noted that the different purification systems maintain different configurations for fraction collection (Table 3.2). The old ÄKTA system is connected to a fraction collection unit which collects individual tubes placed in rows of fifteen tubes per row. The ÄKTA Pure and GO units collect fractions in blocks which hold twelve fractions per row. Thus, fraction numbers on chromatograms, SDS-PAGEs, and Western blots vary depending on which ÄKTA a sample was purified. Table 3.2 lists the corresponding fractions of the three purification systems. Fractions were chosen based on where Peak 1 eluted, shown in dark blue, sometimes extending a fraction before and after (shown in light blue). For ease and consistency, all SDS-PAGE and Western blot images follow the numbering scheme for fractions purified on the Old ÄKTA.

3.2.3 Troubleshooting MSMS detection limitations

While tandem-MS is a cutting-edge method to sequence and identify proteins,[43] detection constraints (described here) introduce inherent limitations in the technology. A majority of the proteins tagged in this study were found to interact with the Fds when each Fd was tagged via the same method (Burkhart et al., 2019). Fd-1 was not identified in the entirety of the present study. Fd-2 was found solely in association with GGR. Fd-3 was found in association with 8 tagged-proteins (details of which are explained in depth in Chapter 4). Reciprocal identification of respective tagged-proteins can be expected for strong associations,

Table 3.2: Fraction collection from each of the three ÄKTA systems used for SDS-PAGE and Western blotting analysis and corresponding fractions of peak 1 elution.

Old ÄKTA	B5	B6	Ladder	B7	B8	B9	B10	B11	B12	B13	B14	Ladder	B15	C1	C2	C3	C4	C5	C6	C7	C8	C9
ÄKTA Pure	B8	B9		B10	B11	B12	C1	C2	C3	C4	C5		C6	C7	C8	C9	C10	C11	C12	D1	D2	D3
ÄKTA GO	B5	B4		B3	B2	B1	C1	C2	C3	C4	C5		C6	C7	C8	C9	C10	C11	C12	D12	D11	D10
Peak 1																						

and a lack of reciprocal identification could weaken the validity of data (discussed here and further in Section 3.4.3). Here we discuss attempts to troubleshoot the AP-MS method to allow for the greatest possibility of identification of the Fds, especially Fd-1 and Fd-2.

Protein precipitation, shipment, and handling

Fractions of purified proteins are transferred to 1.7 mL microcentrifuge tubes and stored at -80˚C immediately after purification to maintain sample integrity and limit protease degradation. Once a purification has been analyzed by SDS-PAGE and Western blot and desired fractions are chosen, specific fractions are thawed on ice and pooled. Pooled fractions are quantified and volume containing ~20 µg is moved into another 1.7 mL microcentrifuge tube. To prepare proteins for shipment to the MS facility, proteins are classically precipitated using trichloroacetic acid (TCA), however such precipitation procedures are known for losing small proteins, such as Fd-1 and -2 which are 6.5 and 7.2 kDa, respectively. After preliminary data did not identify Fd-1 or Fd-2, it was suggested that these small proteins could be lost during precipitation. To test if Fd-1 and/or Fd-2 were lost in TCA precipitation, we sent raw protein samples in a mixture of Buffers A and B to the MS collaborator. Samples containing glycerol must undergo a buffer exchange before trypsin digestion which can be accomplished using the S-Trap micro spin column digestion protocol (ProtiFi).

The TS559 control and one tagged subunit, TK1984C, of pyruvate oxidoreductase (POR) were used to compare TCA-precipitated vs. raw sample handling. TK1984C was previously found to interact with all three Fds,[49] but initial data from TCA precipitated samples did not identify Fds. Non-precipitated, raw samples of TK1984C did not identify the Fds in +S˚ or -S˚ either. Further, when looking at AP-MS identification of small proteins (< 9 kDa) from both handling methods, more small proteins were identified using TCA precipitation in TK1984C and TS559 (Table 3.3), thus TCA precipitation was determined to be the optimal method for protein shipment as fewer small proteins were lost using TCA precipitation than from raw handling.

Table 3.3: Proteins from 6 - 8 kDa identified by tandem MS from Raw and TCA precipitated samples of TK1984C and TS559.

			Quantitative Value (Normalized Total Spectra)									
			TK1984C						TS559			
			+S˚			-S˚			+S˚		-S˚	
Identified Protein	Alt. ID	Molec. Weight (kDa)	Raw	TCA 1	TCA 2	Raw	TCA 1	TCA 2	Raw	TCA	Raw	TCA
50S ribosomal protein L40e	rpl40e	6	0	0	0	0	0	0	0	0	0	1
DNA-directed RNA polymerase subunit K	rpoK	6	0	0	1	0	0	0	0	0	0	0
DNA-directed RNA polymerase subunit P	rpoP	6	0	2	1	0	1	1	0	2	0	1
Uncharacterized protein	TK0138	6	0	0	0	0	0	0	0	0	0	2
Uncharacterized protein	TK0543	6	0	2	0	0	1	1	0	0	0	1
Archaeal histone A	hpkA	7	0	10	0	0	10	6	0	0	3	4
Archaeal histone B	hpkB	7	1	14	6	3	11	9	1	21	4	4
Ribosome biogenesis protein Nop10	nop10	7	0	0	3	0	4	5	0	4	0	3
30S ribosomal protein S14 type Z	rps14	7	0	0	1	0	1	1	0	2	0	1
30S ribosomal protein S27ae	rps27ae	7	0	0	0	0	0	0	0	0	2	1
30S ribosomal protein S27e	rps27e	7	1	2	4	1	1	1	1	4	1	7
Predicted zinc-ribbon RNA-binding protein involved in translation	TK0111	7	0	2	1	0	1	1	0	4	0	4
Uncharacterized protein	TK0478	7	2	5	6	0	5	6	0	2	8	9
RNA-binding protein, containing KH domain	TK1112	7	0	2	0	0	0	0	0	0	0	0
50S ribosomal protein L24e	rpl24e	8	0	0	0	0	0	1	0	0	0	1
DNA-directed RNA polymerase subunit N	rpoN	8	0	0	1	0	0	0	0	2	0	1
30S ribosomal protein S17e	rps17e	8	3	3	1	2	0	4	4	8	8	2
30S ribosomal protein S28e	rps28e	8	3	7	4	3	5	6	5	8	7	7
Putative snRNP Sm-like protein	TK0976	8	3	24	17	8	23	24	11	54	3	3
PRC domain-containing protein	TK1270	8	0	2	1	0	0	1	0	2	0	0

Size and charge detection parameters for MSMS

Proteins for LC-MSMS analysis undergo tryptic digestion before separation by reverse-phase LC, ionization, and being flown through the MSMS detector. Samples which arrive to the MS facility undergo tryptic digestion where trypsin (purified from pig) cleaves proteins at the carboxyl side of every lysine (K) or arginine (R) unless followed by a proline (P). While other digestion options are available, tryptic digestion is by far the most common as the distribution of K and R residues allows for peptides of a detectable size. ~1 µg of trypsin digested peptides (in 2 µL) are separated via a 1 hour gradient of reverse phase liquid chromatography before being ionized (Figure 3.10A).

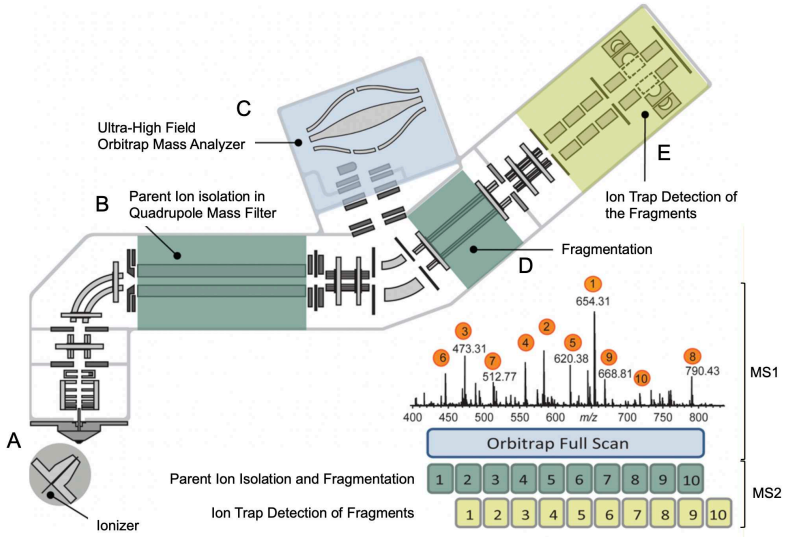


Figure 3.10: Tandem mass spectrometer analyzes charged peptide fragments to determine amino acid sequence. A mass spectrometer for proteomic mass spectrometry includes three main parts: an ion source (A), a mass filter (B), and a detector (C and E). This MS instrument is equipped with two mass detectors which can perform in tandem allowing for tandem MS/MS. A complex protein sample is trypsin digested to produce peptides. Peptides are separated by reverse phase liquid chromatography before entering the mass spectroscopy apparatus depicted here. Peptides are ionized using a single voltage at the inlet such that primary amines become protonated (A) before being directed to the first mass detector, a Quadrupole Mass Filter where parent ions are isolated (B). Peptides are then analyzed by two additional mass detectors in tandem. The Orbitrap Mass Analyzer (C) analyzes the masses of unfragmented peptides, records their mass to charge ratio (m/z), and shares this data with the second mass detector. The peptide backbone fragments in a predictable way such that each amino acid can be expected to disassociate from the whole peptide at the peptide bond (D). Fragmented peptides are detected by the Ion Trap (E). The mass to charge ratio of the intact peptides (MS1) are calculated against mass to charge ratio detected from the fragmented peptides (MS2). The mass to charge ratio of a nine amino acid peptide can be compared to the mass to charge ratio of an eight amino acid peptide, seven amino acid peptide, six amino acid

peptide, etc. allowing for the peptide to be sequenced. Peptide sequences are mapped to the proteome, thereby identifying the proteins in the complex mixture. Protein samples were analyzed by The Ohio State University Campus Chemical Instrument Center (OSU's CCIC) using a Thermo Scientific Oribitrap Fusion mass spectrometer. Adapted from Senko *et al*., 2013.[50]

Classical ionization uses a single voltage to ionize peptides, however High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) uses three voltages (-50, -65, and -80 V) to ionize peptides. Certain peptides may be ionized more effectively with different voltages, thus FAIMS is commonly used. The downside of using FAIMS is that three voltages dilutes the sample effectively by the same amount. Because an abundance of peptides exists in 1 µg, it is not considered a problem to dilute the sample with multiple voltage treatments.

Charged peptides are filtered through a quadrupole mass filter (Figure 3.10B) before flying through two mass detection apparatuses, an Orbitrap and an Ion trap that talk to one another (Figure 3.10C-E). The Oribitrap Mass Analyzer measures the mass of a peptide (MS1) and, in tandem, an Ion trap measures the mass of fragmented peptides (MS2). Peptides fragment in predictable ways, thus the masses of a peptide with ten amino acids and the same peptide fragmented to have one less amino acid will both be measured. Each amino acid maintains a unique mass, thus a peptide can be sequenced in this way. MS ideally detects peptides between 375 – 1400 Da which retain +1-3 charges. Peptides that are < 375 Da, > 1800 Da, or retain > +4 charge are lost in flight, thus undetectable. The mass detection range can be widened to 200 - 1800 Da, but stretching the limit dilutes the sensitivity of the mass detector.

At 6.5 and 7.2 kDa, a tryptic digest of Fd-1 and Fd-2 produces 3 and 5 peptides, respectively (Table 3.4). Based on size alone, only one 4-amino acid peptide (MAWK, 535 Da) from Fd-1 is detectable even within a widened mass detection range of 200 - 1800 Da. Such a small peptide is too ubiquitous to confidently map to a specific protein even within a small proteome, thus not a single peptide from Fd-1 meets the necessary criteria for detection. Three

peptides from Fd-2 fall into the detectable range for MSMS, one of which can be confidently mapped.

MSMS analysis can undergo various detection parameters such as Data Independent Acquisition (DIA) or Targeted Acquisition (TA).[51] When recognition of specific proteins in a complex mixture is required, such as in AP-MS, DIA is appropriate. TA is used if identification of a protein or a small set of proteins is desired and requires a curated list of expected, detectable peptide masses, inclusive of any anticipated modification, such as a phosphorylation or methylation. TA is highly sensitive, as the detector is programmed for a limited sample, which is beneficial in some situations. The MSMS instrument must be set to DIA or TA; the instrument cannot do DIA and TA in the same run.

In an attempt to find the Fds, a TA run was performed without FAIMS to maximize signal and by providing a list of peptides with expected, respective modifications as determined by ProteinProspector (UCSF, v6.4.2). Peptides mapping to Fd-2 and Fd-3 were found. Fd-1 was not identified until spectra were manually searched; Fd-1 was identified with an unexpected deamination. While running every sample twice (using DIA and TA) was not within the scope of this project, the mass of the deaminated Fd-1 peptide was added to the search database and future samples were run without FAIMS to provide the best possible, albeit fruitless, chance of identifying Fd-1.

In interpreting these data in conjunction with previously published studies, the lack of identifiable Fd-1 can only suggest a limitation in the method and not that previously published interactions do not exist, even though without a thorough understanding of how MSMS works, this could be easily assumed. We suggest that it is safe to assume the while Fd-1 was not found, such a result can be considered a false-negative due to the practical limitations of AP-MS. In sum, small proteins that digest into too few peptides, that are either too large or too small to detect, and that retain > +4 charge when ionized may not be accurately identified using this technique. We must assume that observed interactions, while valid, are not wholly inclusive.

Table 3.4: Ferredoxin proteins and their trypsin-digested peptides.

Protein	Amino Acid Sequence	Protein Mass (kDa)	Position	Peptide Sequence	Mass (Da)
Fd-1	MAWKVSVDVDTCIGDAICASLCP DVFEMGDDGKAHPVVETTDLDC AQEAAEACPVGAITLEEA	6.528	34 - 63	AHPVVETTDLDCAQEAAEAC PVGAITLEEA	3053.4077
			5 - 33	VSVDVDTCIGDAICASLCPDVFEMGDDGK	2974.2824
			1 - 4	MAWK	535.2697
Fd-2	MPEKIKVVVNEDRCYLCGGCAG VCPTLAIEVHSTGWEFLQDKCIS CRICINACPVGALSAKPLEVSE	7.199	14 - 42	CYLCGGCAGVCPTLAIEVHSTGWEFLQDK	3100.4035
			48 - 67	ICINACPVGALSAKPLEVSE	2014.0347
			7 - 13	VVVNEDR	830.4366
			43 - 47	CISCR	581.2534
			1 - 4	MPEK	504.2486
Fd-3	MADVKAPVIGRDALGREVKDLSV IPWWGVDRKEIEWYPKINYSVCA RCGLCFITCGRRVFDWDTEEGK PVVARPYNCMVGCNTCAILCPC NAIEFPPKEYVKKLVIEHGIIRKAF EITKPLTKKKEESTNGAESVFNP	15.611	59 - 98	VFDWDTEEGKPVVARPYNCMVGCNTCAILCPCNAIEFPPK	4430.0244
			20 - 31	DLSVIPWWGVDR	1442.7426
			126 - 138	EESTNGAESVFNP	1380.5913
			114 - 123	AFEITKPLTK	1147.6721
			48 - 57	CGLCFITCGR	1072.4736
			104 - 112	LVIEHGIIR	1049.6465
			33 - 39	EIEWYPK	964.4774
			40 - 47	INYSVCAR	925.456
			6 - 11	APVIGR	612.3827
			1 - 5	MADVK	563.2857
			99 - 102	EYVK	538.2871
			12 - 16	DALGR	531.2885

3.3 Variables in the analysis

The appropriate use of controls, method parameters, statistical analysis, and data curation by rational processing (each of which will be discussed in detail in this section) can confirm previous protein partnerships and lead to a greater understanding of the complexity of protein networks *in vivo*. Section 3.3 will focus on variables in analyzing proteomic MS data including the value of biological replicates, various statistical tests, and relevant thresholds for determining significance. Validation of the data will be discussed in Section 3.4.

Overview of Scaffold analysis software and databases

LC-MSMS data produces spectral counts of peptide masses which are mapped to the *T. kodakarensis* (KOD1) proteome available on Uniprot. The proteome is available as a FASTA file and can be modified to include proteins not originally in the database. For example, if a C-terminal tagged protein is not identified in the data and *in silico* trypsin digest shows that the only peptide that will fall within the mass detection parameters is the C-terminal peptide (which now includes a HA and His_6 sequence), the proteome can be updated to include the tagged protein sequence. Such steps were not required for this project, but are possible, and may influence data outcomes.

The Mascot Daemon search engine (Matrix Science v2.7.0, Boston, MA) is used to search spectral data through ProteomeDiscoverer (PD, ThermoScientific v2.4) against the KOD1 Uniprot database. *T. kodakarensis* encodes ~2,300 proteins, thus mapping spectral counts to this small proteome requires less statistical validation than other larger, more complex proteomes (such as the human proteome). Data are additionally searched against a contamination database which can identify if a sample contains an abundance of common contamination proteins. Trypsin protease isolated from pig appears in the contamination database (sp|TRYP_PIG|, 24 kDa) as proteins are digested by trypsin to generate peptides (Section 3.2.3). Lastly, data are searched against a decoy database (tr|*|*_THEKO-DECOY)

which includes ~six proteins from *T. kodakarensis* which are used to determine the false discovery rate (FDR). The addition of a decoy database allows for peptides to be assessed using Percolator, a semi-supervised algorithm for peptide evaluation.[52]

Data from the MS collaborator is provided in two file types (Figure 3.11). The first is a Mascot via PD file for every sample which includes % coverage, # peptides, # unique peptides, and a Mascot score for each protein identified. An embedded layer of data shows all peptides that mapped to that protein including their mass to charge values (Da), retention time, ion score, and Percolator analysis values (readable using *.XLSX). The Mascot via PD file was not used in this analysis and will not be discussed in detail.

The second file is a comparison of datafiles from multiple samples using Scaffold 5[53] (Proteome Software, Inc., v5.3.0, Portland, OR) organized into an experimental framework of test samples and control samples (Figure 3.11B). Such a comparison can only be performed using licensed Scaffold 5 software (MS facilities pay > $1k for annual licenses). A free version of Scaffold 5 can be downloaded from ProteomeSoftware.com which will allow for basic analysis of the data as formatted by the MS collaborator, but unlicensed software cannot be used to modify samples or the experimental setup (*e.g.* if an experimental sample has been labeled as a control, this error will need to be updated using licensed software). The experimental set-up can be viewed in Scaffold by navigating to Experiment (in the title bar) and choosing Edit BioSample. Alternatively, hovering over a sample name will display all sample information, including which group (Experimental or Control) a sample has been labeled. The Scaffold 5 user manual is available from ProteomeSoftware.com.

A

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Checked	Protein FDR C	Master	Accession	Description	Exp. q-value: (	Sum PEP Scor	Coverage [%]	# Peptides	# PSMs	# Unique Pept	# AAs	MW [kDa]	calc. pI	Score Mascot:	# Peptides (by	# Protein Groups
2	FALSE	Medium	Master Protei	O07836	Indolepyruvat	0.027	1.576	4	1	1	1	202	21.9	6.19	0	1	1
5	FALSE	High	Master Protei	O08467	Glutamine syr	0	3.831	5	2	2	2	443	50.3	5.27	37	2	1
9	FALSE	High	Master Protei	O59650	Glutamate de	0	27.502	21	6	9	6	421	47	6.24	169	6	1
10		Checked	Confidence	Identifying Nc	PSM Ambiguit	Annotated Se	Modifications	# Proteins	Master Protei	Protein Acces	# Missed Clea	Charge	DeltaScore	DeltaCn	Rank	Search Engine	m/z [Da]
11		FALSE	High	Mascot (A11)	Unambiguous	GPAFGVITGKF		1	O59650	O59650	0	3	1	0	1	1	617.36063
12		FALSE	High	Mascot (A11)	Unambiguous	IVAEVAnGPVT	N7(Deamidate	1	O59650	O59650	0	3	1	0	1	1	744.71692
13		FALSE	High	Mascot (A11)	Unambiguous	VAVVDLPYGG		1	O59650	O59650	0	2	1	0	1	1	587.82586
14		FALSE	High	Mascot (A11)	Unambiguous	DAAYVVAVSR		1	O59650	O59650	0	2	0.8462	0	1	1	525.7831
15		FALSE	High	Mascot (A11)	Unambiguous	GGIYNPDGLPI		1	O59650	O59650	0	2	1	0	1	1	877.951
16		FALSE	High	Mascot (A11)	Unambiguous	IVAEVANGPVT		1	O59650	O59650	0	3	1	0	1	1	744.38928
17		FALSE	High	Mascot (A11)	Unambiguous	VAVVDLPYGG		1	O59650	O59650	0	2	1	0	1	1	587.82586
18		FALSE	High	Mascot (A11)	Unambiguous	GGIIVDPK		1	O59650	O59650	0	2	0.7812	0	1	1	399.73917
19		FALSE	High	Mascot (A11)	Unambiguous	DAAYVVAVSR		1	O59650	O59650	0	2	1	0	1	1	525.78433

MH+ [Da]	Theo. MH+ [D	DeltaM [ppm]	Deltam/z [Da]	Activation Ty	MS Order	Isolation Inte	Ion Inject Tim	RT [min]	First Scan	Spectrum File	File ID	Ions Score	# Protein Gro	Identity Strict	Identity Relax	Expectation V	Percolator q-\	Percolator PEP
1850.06734	1850.06468	1.44	0.00089	HCD (High En	MS2	39.2382	35	66.3718	49696	022023_500	F1	39	1	19	13	0.00010409	0.0004092	0.000001006
2232.1362	2232.13942	-1.44	-0.00107	HCD (High En	MS2	75.16846	29.343	64.6821	47763	022023_500	F1	21	1	22	15	0.01352607	0.0004092	0.00002238
1174.64445	1174.64665	-1.88	-0.0011	HCD (High En	MS2	18.21924	23.861	57.1349	39732	022023_500	F1	62	1	22	15	1.1085E-06	0.0004092	0.000001953
1050.55892	1050.55783	1.03	0.00054	HCD (High En	MS2	15.54661	35	48.0284	29766	022023_500	F1	52	1	20	13	8.4369E-06	0.0004092	0.00006837
1754.89473	1754.89594	-0.69	-0.0006	HCD (High En	MS2	52.02864	35	67.9488	51629	022023_500	F1	19	1	22	15	0.02168303	0.0006167	0.004369
2231.15329	2231.1554	-0.95	-0.0007	HCD (High En	MS2	30.19505	35	63.6954	46715	022023_500	F1	14	1	24	17	0.10383978	0.0006167	0.003726
1174.64445	1174.64665	-1.88	-0.0011	HCD (High En	MS2	28.00488	26.229	57.136	39733	022023_500	F1	18	1	22	15	0.03095662	0.0006167	0.006527
798.47106	798.47198	-1.15	-0.00046	HCD (High En	MS2	40.4023	35	39.1637	20630	022023_500	F1	32	1	13	13	0.00011325	0.0006167	0.006436
1050.56139	1050.55783	3.38	0.00178	HCD (High En	MS2	37.72338	35	47.6717	29346	022023_500	F1	21	1	20	13	0.00780536	0.003649	0.06523

B

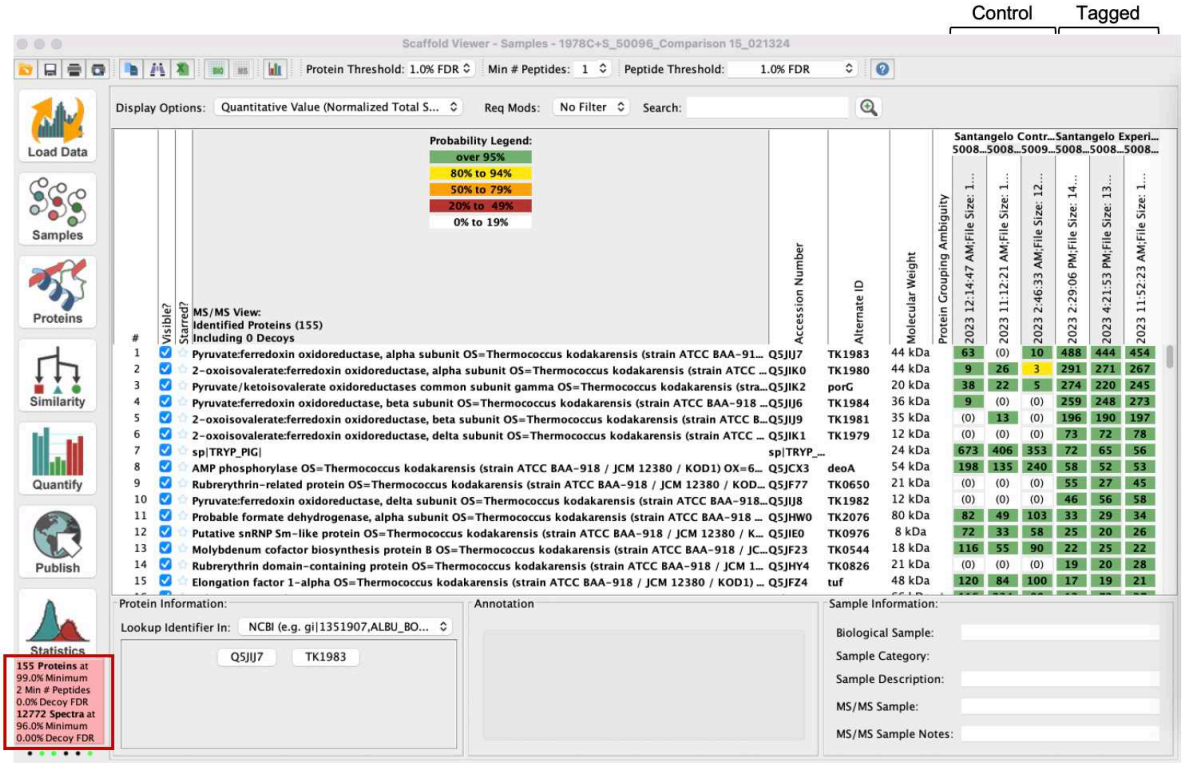


Figure 3.11: Data is returned from Proteomic MS analysis in raw and processed forms. Specific data from a single sample run is returned as an Excel-readable format (.xlsx) showing all proteins that were identified (A). Each identified protein is listed in a row (blue) including a variety of metrics which describe the data identifying this protein. Each protein is identified by at least one peptide. Peptide data is embedded in a layer and can be viewed by expanding the data for each protein by clicking on the (+) at left. Peptides are listed in rows (light pink/orange) with specific data for each peptide listed such as charge, m/z (Da), and retention time (RT). These data for each sample run are provided in a Mascot via PD folder. Samples are combined into a comparison including data from both control and tagged samples and these data are viewed using the free version of Scaffold 5,[53] Scaffold Viewer (B). Proteins that appear in these combined data sets appear as a list and the Quantitative Value (Normalized Total Spectra) is displayed for each Control or Tagged sample (labeled at top right). Three control and three experimental (tagged) samples are included in this comparison. An overview of the data is provided at the bottom left (red box). Here, 155 proteins were identified from 12,772 spectra between the six samples included in this comparison.

Proteomic MS data is qualitative not quantitative

While proteomic MS analysis provides data in a quantitative format (such as the number of peptides mapped to a protein in each sample), these data must not be interpreted quantitatively. Quantitative values reported are used for statistical comparison between experimental samples and control samples to determine the presence or absence of a protein. The presence or absence of a protein is qualitative data and must remain as such. For example, two proteins pass all statistical thresholds and are both found to be significant in an experimental sample compared to a control sample. The first is reported with 300 mapped peptides while the second is reported with 50 mapped peptides. While a different number of peptides are detected, both proteins generate equally valid results compared to the control. Analysis algorithms must account for the size of the protein. While 6X more peptides were detected for protein #1 than protein #2, if protein #1 is 10X the size of protein #2, it should be detected with 500 mapped peptides. A discussion of such logical fallacies in data interpretation continues in Section 3.3.3, *Employing a multiple testing correction*.

3.3.1 Limiting false positives using a near-isogenic control strain

In the case of AP-MS data, a false positive result can be interpreted as a protein that does not interact with the protein of interest but appeared as significant in the data. Proteins that are identified when no biologically relevant interaction exists could contaminate the data with false positives. LC-MSMS is incredibly sensitive. For example, each run of pooled, purified protein identified ~100 - 700 proteins out of the ~2,300 proteins in the *T. kodakarensis* proteome. Such data must be filtered to identify specific interactions between the tagged protein of interest and all other proteins in the proteome. Thus, the non-tagged, parent strain (TS559) was grown, harvested, and purified in the same manner as the tagged strains to serve as a near-isogenic control. In theory, only 45 bp differ in the 2 Mbp genome between the parental and tagged strains. Mutational differences will certainly arise, thus a comparative analysis of

WGS data can provide a detailed analysis of any notable genomic deviations. *Breseq* is a computational pipeline recommended for comparing mutational differences between WGS data of *T. kodakarensis* strains compared to a reference sequence such as KOD1.[54] Corresponding pooled fractions of the control were sent for LC-MSMS to back-subtract any proteins pulled down in these fractions independent of the tag. Control samples allow for sample comparisons to determine which of the identified proteins is significant in the tagged strain compared to the control. (Statistical significance is discussed in depth in Section 3.3.3.) While it is not necessary, replicate control samples may further strengthen the data and were employed in the majority of comparisons in this data set.

As discussed in Section 3.2.3, some proteins are difficult to detect based on size and charge characteristics, while other proteins are highly detectable. Such detectable false positives should be equally detectable in the control strain, thus employing multiple biological replicates of the parent strain should aid in minimizing false positives that pass statistical thresholds.

3.3.2 Limiting false negatives using rational statistical thresholds

In the case of AP-MS data, a false negative result can be interpreted as a protein that interacts with the protein of interest but did not appear as significant in the data. Such a result can occur because the protein in not detectable (as discussed in Section 3.2.3) or because statistical thresholds are too stringent. Statistical thresholds, while not arbitrary, are determined to identify meaningful data. Such thresholds can maintain a degree of flexibility until appropriate parameters are established. Exploration of these parameters is discussed thoroughly in Section 3.3.3.

3.3.3 Statistical analysis of AP-MS data

As is often the case with methods that produce large data sets, sound statistical methods must be employed to identify meaningful data while limiting false positive results. Because of the limitations of MS as a technique (discussed in Section 3.2.3), the data will never be wholly inclusive, but hopefully it will provide novel insights that move our understandings of *in vivo* proteins associations forward.

Raw LC-MSMS data are returned as spectral counts of peptide masses and peptide fragment masses that allow for peptide sequence identification. Sequenced peptides are then mapped to a proteome. Spectral data that is effectively mapped to a supplied proteome is provided from the MS collaborator for each sample in a MASCOT via PD file (.xlsx). Proteins identified are listed in rows and their respective spectral counts are embedded in additional levels of rows. Analysis can be performed from these semi-interpreted data files, however a free version of Scaffold 5, while imperfect, is available for proteomic MS data analysis with many provided functions.

Scaffold software is specifically built for analyzing and interpreting proteomic MS data.[53] The Scaffold software is not ideal, as there are many built-in errors that one must navigate or make do with. The alternative to using Scaffold is to use the sample data provided in the MASCOT via PD files (.xlsx). It should be noted that automatic functions in Scaffold, such as calculating the false discovery rate (FDR) using decoy protein values, will need to be performed manually or alternative data validation steps should be employed.

The following discussion of data analysis utilizes the free version of Scaffold 5 (v5.3.0, 2021 - 2023) and .sf3 files provided by the MS collaborator. Scaffold 5 allows for various analyses of the data including modifying the protein and peptide thresholds as well as preforming basic statistical tests including Coefficient of Variation, Analysis of Variance (ANOVA), T-Test, or Fisher's Exact Test (FET); though not all tests were available for all files.

While Scaffold provides the foundation of analysis for this project, a hybrid approach to data analysis is described here. Total spectral counts of peptides mapped to each protein per sample are analyzed using Scaffold's built-in quantitative analysis software. The outcome is exported (as an .xlsx file) and further sorted to identify significant proteins in each comparison based on a p-value and $\log_2$ fold change between experimental and control samples.

Analyzing the data using statistical testing

Scaffold allows for the adjustment of several parameters including Protein Threshold, Peptide Threshold, Minimum # Peptides, and the displayed value for each identified protein in each sample (Figure 3.12). A decoy database was included in the analysis, which provides a baseline to calculate false discovery rate (FDR), thus both Protein Threshold and Peptide Threshold were set to 1% FDR. As we were interested in detecting all possible interacting partners and some previously determined partners, such as the Fds, are small proteins of which only one peptide may be detected, one is the threshold for minimum number of peptides.

While proteomic MS data can be intimidating, it may be helpful to recognize that mean spectral counts for each protein are compared between experimental (tagged) and control samples one at a time. While many proteins are identified in a single sample run (~100 - 700 proteins), the significance of each protein in the data can be interpreted using a test that compares the mean spectral counts in biological replicates between experimental and control samples; if the average total spectral counts for a protein in the experimental sample is higher than expected, we can consider it significant at a predetermined threshold (*e.g.* p-value of < 0.05). This protein would be expected to be significant in the experimental sample compared to the control in 95% of tests. The null hypothesis states that the mean spectral counts between two groups are not different, thus, to falsify the null and support the hypothesis, the mean spectral counts between two groups must be different. This test is repeated for every protein identified in an analysis. An analysis consists of all experimental and all control replicates being

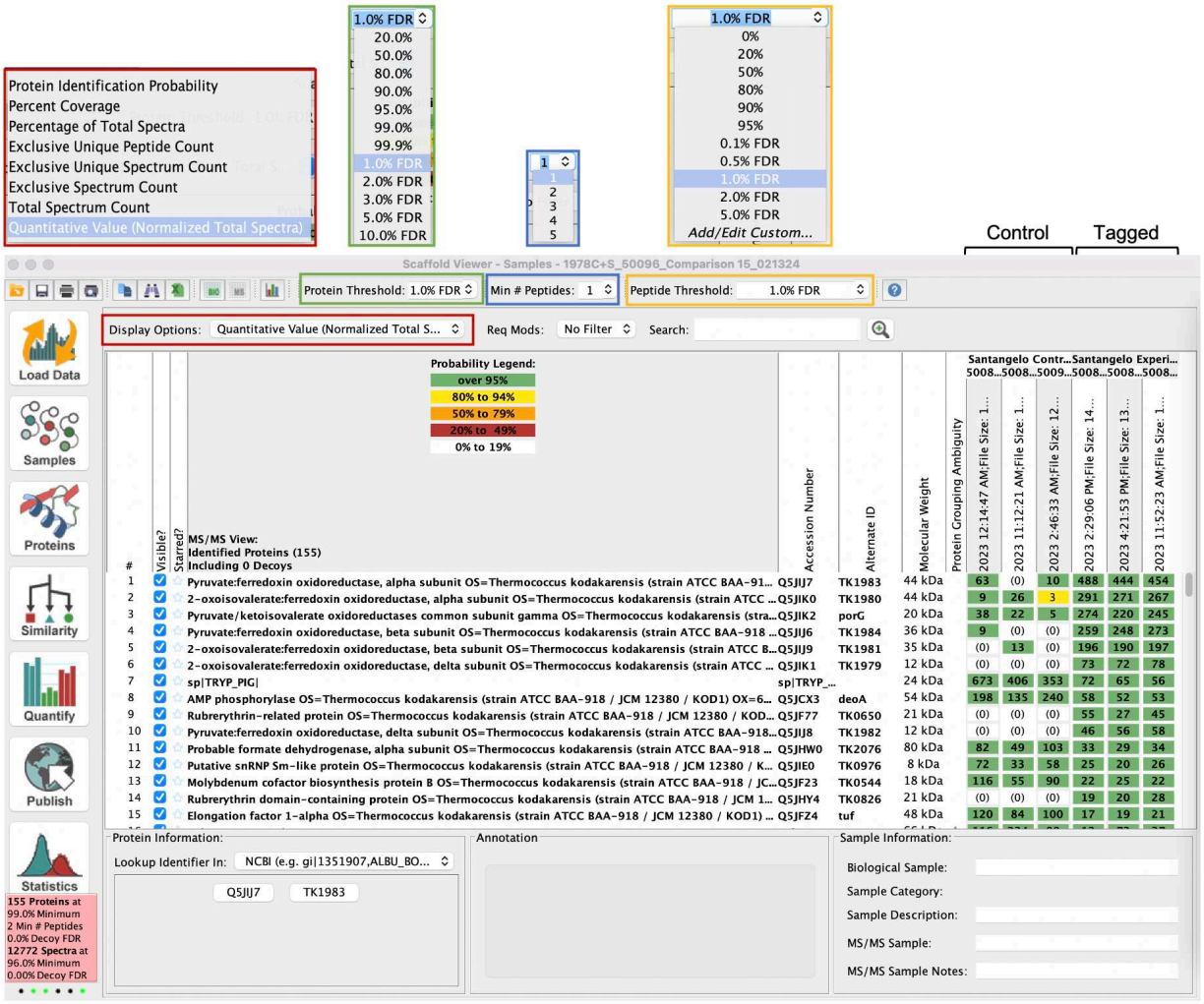


Figure 3.12: Scaffold Viewer displays proteins identified in samples from LC-MSMS and provides statistical thresholds for interpreting data. The main display pane for viewing proteins using Scaffold includes processed data from multiple control and tagged samples. All proteins identified in these combined LC-MSMS runs are listed. The analysis is run against a decoy database which allows for the calculation of a false discovery rate (FDR). Both protein and peptide thresholds have been set at 1% FDR, but a wide range of options for the protein threshold (green box) and peptide threshold (yellow box) exist. 1% FDR was chosen for each protein and peptide thresholds. The minimum number of peptides which map to each protein can be chosen (blue box). As we were interested in identifying proteins which may only have a single detectable peptide, the lowest threshold was set. A numerical value for each protein in each sample is listed in green at right. The green refers to over 95% probability (as calculated by Scaffold), while the yellow represents 80% to 94% probability. The right-most three columns show values corresponding to proteins identified in experimental (tagged) samples. The three columns to the left of the first three show values corresponding to proteins identified in control samples. The numerical value shown is described by Scaffold as the “Quantitative Value (Normalized Total Spectra).” Display Options (red box) provides a choice of values displayed for each protein in each sample. Quantitative Value (Normalized Total Spectra) is chosen as the value displayed. This value calculates the total spectra associated with each peptide (and then mapped to each protein) and adjusts this value considering the size of the protein, as larger proteins will generate more spectra than smaller proteins. It must be noted here again that these

values should not be compared between proteins; these values should only be compared for the same protein between samples. The other display setting which is recommended is “Total Spectrum Count” where the spectrum count has not been adjusted for protein size. Total Spectra is used when a statistical test has been applied. After a FET or T-Test has been calculated, the “Quantitative Value” option converts to “Quantitative Value (Total Spectra)” and “Normalized Total Spectra” is no longer visible until the file is closed and reopened.

compared to one another. Data from typically six sample runs are combined into an analysis, commonly referred to as a comparison.

In the Quantitative Analysis Setup pane (Figure 3.13), Scaffold provides the following tests: Coefficient of Variation, T-Test, and the FET. (Analysis of Variance, ANOVA, is visible, but not an option as indicated by being greyed-out, which was also observed for the T-Test option in some files.) Coefficient of variation is a measure of dispersion and is not recommended. A T-Test is a common method to compare two means in continuous data. (A chi-square test is used to compare two means from discrete or categorical data.) A T-Test is ideal for large sample sizes (typically >three samples per group). Here, three experimental and one to three control biological replicates are provided in each comparison. A FET is ideal for small sample sizes, though it can be used for larger sample sizes as well.

As Scaffold code is proprietary and the user’s manual is limited in describing how the code is implemented, it was determined that iterative experimentation with multiple data sets and various combinations of settings would be beneficial. Protein and peptide threshold were minimally tested, but a FDR of 1% seemed appropriate and was not further probed. The minimum peptide threshold was never increased from one to provide the greatest opportunity of identifying small proteins with few peptides. The T-Test and FET were compared extensively. The T-Test (Table 3.5, bottom) generally provided less conservative results than the FET (Table 3.5, top) and often identified proteins as significant when very few (1 - 4) spectral counts were observed in the experimental samples compared to 0 counts in the control samples (Table 3.5, bottom grey). Table 3.5 shows a comparison of a FET and T-Test analysis preformed on PORβ (TK1984C) in +S˚ conditions. While both tests identified all subunits of the POR complex

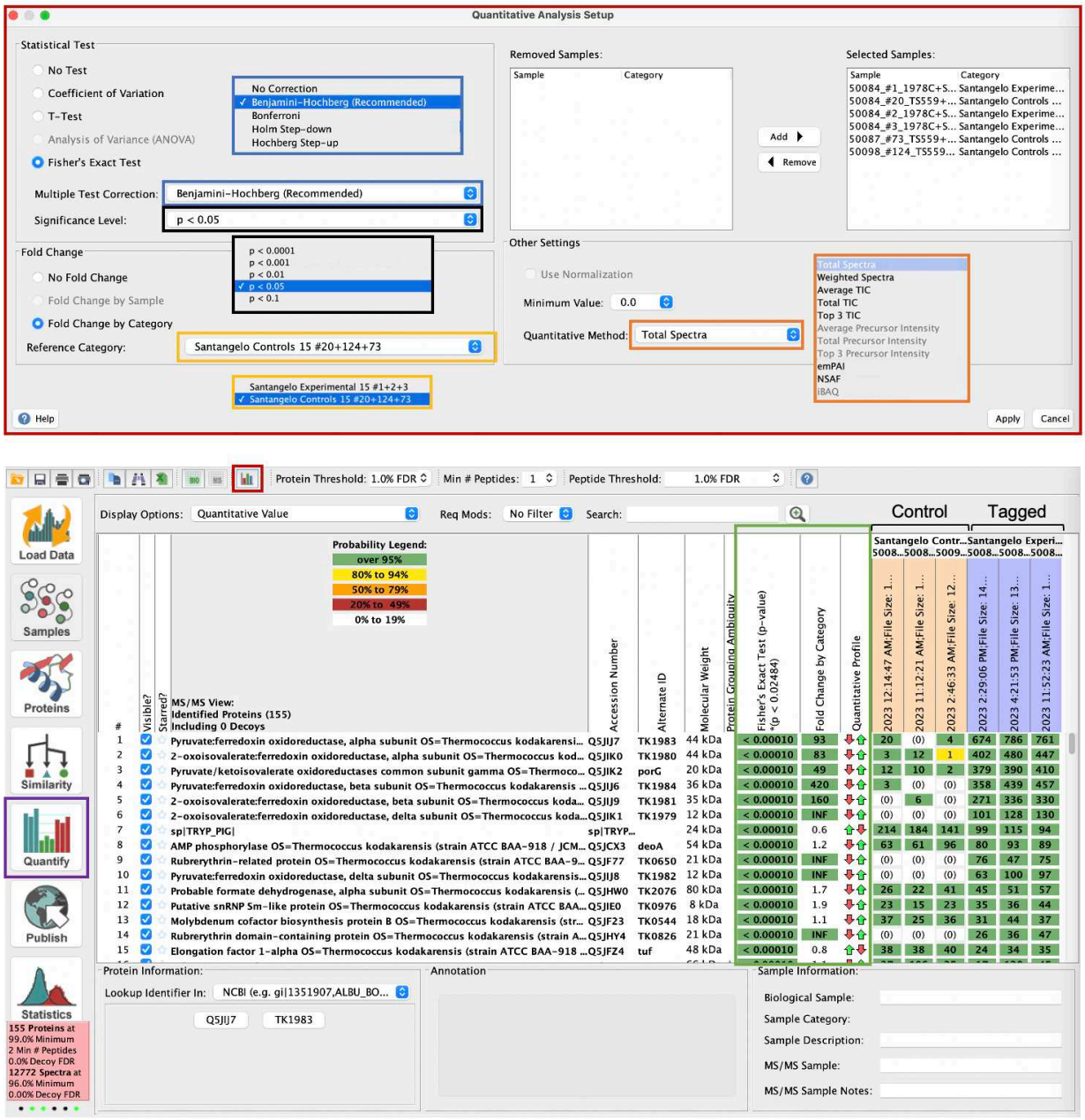


Figure 3.13: Statistical analysis options are offered in Scaffold and provide a p-value and fold change by category value. Scaffold provides built-in statistical testing software. The Quantitative Analysis Setup pane (above in red) is accessed by clicking the bar-graph icon (red box) at the top left of the main Scaffold window. The upper left quadrant of the Quantitative Analysis Setup provides a choice of statistical tests including T-Test and FET. Note that Analysis of Variance (ANOVA) is greyed out and not available in this setup. Four multiple test correction methods are available (blue box). Benjamini-Hochberg, recommended test, has been chosen for this analysis. A p-value < 0.05 has been chosen as the Significance Level (black box), though more relaxed and more stringent cutoffs are available. The upper right quadrant provides a choice for which samples will be included in the comparison. Here, three tagged samples (TK1978C+S) and three control samples (TS559+S) have been placed into one of two categories ("Santangelo Controls" or "Santangelo Experimental") and are included. The lower left quadrant allows for Fold Change to be calculated based on Category (mustard box). It

should be noted that sample category is a feature of the licensed version of Scaffold and must be modified before analyzing using Scaffold Viewer. As we are interested in a difference in the tagged sample compared to control, the Santangelo Control category is chosen as the Reference Category. In the lower right quadrant of the Quantitative Analysis Setup, the minimum value is 0.0 and the Quantitative Method chosen is Total Spectra (orange box), though other options are available. After a Quantitative Analysis has been applied, three columns appear in the main Scaffold window: “Fisher’s Exact Test (p-value),” “Fold Change by Category,” and “Quantitative Profile” (green box). [Note: These statistical analysis values can be exported from Scaffold as a tab separated value file (TSV) by clicking “Export: in the top bar and selecting “Samples Report.” It is recommended to export Total Spectrum Count of Proteins (not Proteins with Isoforms) as a TSV and then convert to a “.xlsx” file format using Excel.] The p-value and Fold Change by Category value are used to determine significance at thresholds discussed in Section 3.3.3. The data can be further visualized in by clicking the “Quantify” option in the left-column (purple box). Total Spectral Counts for control samples (light salmon) and tagged samples (periwinkle) are shown in columns at the right. This data set has been sorted by the greatest spectral counts (in descending order) for the left-most experimental sample.

(TK1978 and TK1981 - TK1984) as significant (yellow), the FET identified twenty-one proteins as significant, while the T-Test identified twenty-seven. While both tests identified fifteen similar proteins, FET identified six proteins that were not identified by the T-Test (grey) and the T-Test identified eleven proteins that were not identified by the T-Test (grey). In observing the total spectral counts of proteins identified in one test and not the other it becomes apparent that the T-Test is sensitive to low numbers of spectral counts in experimental samples which are considered significant compared to zero spectral counts in the control samples. Such sensitivity is not desirable for AP-MS data where false-positive results require filtering with stringent statistical methods.

The FET is designed to handle small sample sizes. The degrees of freedom are reduced in the FET allowing for it to be applied when sample sizes are limited. Greater sample sizes increase the number of degrees of freedom, and greater degrees of freedom raises the threshold for significance. Degrees of freedom are limited in a FET which is why it is used when sample sizes are restricted.

Table 3.5: Fisher's Exact Test analysis compared to T-Test analysis of PORβ (TK1984C) in +S° conditions.

Fisher's Exact Test (p-value): *(p < 0.00679)

Identified Proteins	Gene No.	p-value	Fold Change	Experimental Bio. Reps. 1	2	3	Control Bio. Reps. 1	2	3
Xanthine/uracilpermease	TK0157	0.037	INF	5	2	1	0	0	0
GMP synthase [glutamine-hydrolyzing] subunit B	TK0193	0.00001	8.8	13	23	17	6	0	0
Phosphoribosylamine--glycine ligase	TK0204	0.016	INF	4	3	3	0	0	0
Formate-dependent phosphoribosylglycinamide formyltransferase	TK0207	0.0021	9.3	7	10	11	3	0	0
Amidophosphoribosyltransferase	TK0211	0.00001	24	86	82	46	8	1	0
Phosphoribosylaminoimidazole-succinocarboxamide synthase	TK0432	0.00001	INF	46	47	24	0	0	0
Adenylosuccinate lyase	TK0561	0.009	5.1	13	12	11	5	0	2
Rubrerythrin domain-containing protein	TK0826	0.00001	9.9	24	30	25	8	0	0
Aldehyde ferredoxin oxidoreductase	TK0844	0.0033	5.4	25	11	7	5	1	2
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1325	0.041	7.5	8	4	3	1	0	1
Ferredoxin:NADP oxidoreductase, beta subunit	TK1326	0.025	INF	5	2	2	0	0	0
Archaeal histone A	TK1413	0.0021	INF	0	5	10	0	0	0
Cell division protein FtsZ 1	TK1421	0.0014	21	2	11	8	1	0	0
Predicted dehydrogenase	TK1557	0.00001	3.9	240	157	73	90	16	16
Glutamine synthetase	TK1796	0.00001	8	30	29	13	6	1	2
Uncharacterized protein	TK1958	0.011	INF	6	3	2	0	0	0
Pyruvate/ketoisovalerate oxidoreductases common subunit gamma	TK1978	0.00001	16	388	380	305	44	14	11
Pyruvate:ferredoxin oxidoreductase, delta subunit	TK1982	0.00001	51	51	56	45	3	0	0
Pyruvate:ferredoxin oxidoreductase, alpha subunit	TK1983	0.00001	8.9	920	832	789	255	26	5
Pyruvate:ferredoxin oxidoreductase, beta subunit	TK1984	0.00001	29	526	492	460	49	2	0
Iron-molybdenum cofactor-binding protein	TK2016	0.011	INF	5	3	3	0	0	0

T-Test (p-value): *(p < 0.00078)

Identified Proteins	Gene No.	p-value	Fold Change	Experimental Bio. Reps. 1	2	3	Control Bio. Reps. 1	2	3
GMP synthase [glutamine-hydrolyzing] subunit B	TK0193	0.015	8.6	13	23	17	6	0	0
Phosphoribosylformylglycinamidine synthase subunit PurL	TK0197	0.0018	INF	2	3	2	0	0	0
ATP-grasp domain-containing protein	TK0203	0.015	INF	2	1	2	0	0	0
Phosphoribosylamine--glycine ligase	TK0204	0.00015	INF	4	3	3	0	0	0
Formate-dependent phosphoribosylglycinamide formyltransferase	TK0207	0.019	9.3	7	10	11	3	0	0
Amidophosphoribosyltransferase	TK0211	0.0011	17	86	82	46	8	1	0
Phosphoribosylaminoimidazole-succinocarboxamide synthase	TK0432	0.0014	INF	46	47	24	0	0	0
PDDEXK_1 domain-containing protein	TK0446	0.012	INF	1	2	1	0	0	0
Flavin prenyltransferase UbiX	TK0509	0.041	9	3	2	4	1	0	0
Adenylosuccinate lyase	TK0561	0.014	3.2	13	12	11	5	0	2

Rubrerythrin-related protein	TK0650	0.012	INF	1	2	1	0	0	0
Rubrerythrin domain-containing protein	TK0826	0.0035	9.6	24	30	25	8	0	0
Uncharacterized protein	TK0883	0.0057	INF	2	1	1	0	0	0
Rubrerythrin-related protein	TK1056	0.00045	INF	1	1	1	0	0	0
Translation initiation factor 6	TK1321	0.00045	INF	1	1	1	0	0	0
Ferredoxin:NADP oxidoreductase, beta subunit	TK1326	0.019	INF	5	2	2	0	0	0
Uncharacterized protein	TK1459	0.0057	INF	2	1	1	0	0	0
tRNA uridine(34) acetyltransferase	TK1574	0.029	7.7	2	4	2	1	0	0
Ferredoxin:NADP oxidoreductase, beta subunit	TK1685	0.013	5.3	4	4	3	2	0	0
Glutamine synthetase	TK1796	0.01	4.2	30	29	13	6	1	2
Uncharacterized protein	TK1958	0.018	INF	6	3	2	0	0	0
Pyruvate/ketoisovalerate oxidoreductases common subunit gamma	TK1978	0.00001	7.8	388	380	305	44	14	11
Pyruvate:ferredoxin oxidoreductase, delta subunit	TK1982	0.00001	49	51	56	45	3	0	0
Pyruvate:ferredoxin oxidoreductase, alpha subunit	TK1983	0.0012	6.6	920	832	789	255	26	5
Pyruvate:ferredoxin oxidoreductase, beta subunit	TK1984	0.00012	25	526	492	460	49	2	0
Iron-molybdenum cofactor-binding protein	TK2016	0.0013	INF	5	3	3	0	0	0
Deoxycytidylate deaminase	TK2257	0.035	4.7	2	2	1	1	0	0

0

A two-sided test is appropriate for this analysis; the null hypothesis of a two-sided test assumes that there is no difference between two means ($h_0 = 0$), while a one-sided test assumes direction ($h_0 < 0$ or $h_0 > 0$). Such direction is determined for every test and considering that every protein requires its own test, the direction would need to be hypothesized *a prori*, which was not the case. In the case of AP-MS, a two-sided test is appropriate.

With the assistance of Dr. Sean Cascarina (and later confirmed by Scaffold's technical assistance), it was determined that the FET preforms a one-sided test in both directions and combines the results as if a two-sided test was performed. Such a method is not ideal (as noted by Scaffold technical assistance), but foundational code in Scaffold limits the capacity of the software from calculating an actual two-sided test. (Since this email exchange, Scaffold has updated their software. The newest version of Scaffold v5.3.3 no longer includes FET as an option.)

In some .sf3 files, the T-Test option was greyed-out and it was not determined how to make this option available for all files. It is assumed that this feature is dependent on the experimental set-up, which requires a licensed version to modify. The MS collaborator maintains a license, but such analytical assistance to troubleshoot Scaffold software settings was beyond the expectations of our relationship.

Thus, the options for analysis were as follows: 1) preform a T-Test with relaxes statistical thresholds resulting likely in an abundance of false positives in the data, 2) preform a FET as allowed by Scaffold (as two, one-sided tests) and assume these finding are sufficiently valid for analysis, or 3) attempt to analyze data outside of Scaffold and miss out on built-in benefits of Scaffold such as calculating the FDR using decoy proteins or other features that are unknown. It was determined that option (2) would provide for consistent analyses with appropriate statistical rigor while benefiting from specific proteomic analysis factors in Scaffold that would otherwise be challenging to replicate.

As a final note, while the Scaffold 5 code is proprietary, it seems that Scaffold has built-in specific controls so that the Quantitative Value (Normalized Total Spectra) cannot be viewed once a quantitative analysis has been applied. For instance, after executing a FET, the values displayed are described as Quantitative Value (Total Spectra) which are the same values for Total Spectra and the values observed for Quantitative Value (Normalized Total Spectra) are no longer visible unless the file is closed and reopened without a quantitative analysis applied.

Employing a multiple testing correction

A common logical fallacy when interpreting large proteomic data sets is to compare numerical results for multiple proteins to one another. Such interpretation must be avoided as the results for every protein must be treated as an independent test only comparing counts of the same protein between control and experimental samples. As the result for every protein is treated as an independent statistical test, some are bound to show significance simply due to a multitude of tests being performed. Scaffold offers a handful of multiple testing correction options under the Quantitative Analysis Setup pane. While Benjamini-Hochberg is recommended (and employed in this analysis), Bonferroni, Holm Step-down, and Hochberg Step-up corrections are available. Multiple testing corrections further aid in reducing false positive results. It should be noted that analyses employing the Benjamini-Hochberg multiple testing correction did not noticeably deviate from analyses lacking this multiple testing correction, thus it cannot be confirmed how this step is applied in Scaffold, however, because a multiple testing correction is appropriate, it was included in all analyses.

Optimal settings in Scaffold 5, in summary

Scaffold allows for the adjustment of many parameters including Protein Threshold, Peptide Threshold, Minimum # Peptides, and the displayed value for each identified protein in each sample. A decoy database was included in the analysis, which provides a baseline to

calculate false discovery rate (FDR), thus both Protein Threshold and Peptide Threshold were set to 1% FDR. As we were interested in detecting all possible interacting partners and some previously determined partners, such as the Fds, are small proteins of which only 1 peptide may be detected, 1 is the threshold for minimum number of peptides. Proteins were identified as significant in the tagged strain compared to the control using a Fisher's Exact Test with a Benjamini-Hochberg multiple testing correction, a p-value cutoff of 0.05, and a $\log_2$ fold change of 2 or greater in the experimental compared to control.

Different values for each protein detected per sample can be displayed under Display Options (Figure 3.13, red box). Total Spectrum Count is recommended and will list all spectra from all peptides that map to a particular protein. Quantitative Value (Normalize Total Spectra) adjusts for protein size, as larger proteins will inherently have more peptides and more spectra than smaller proteins. While this value may seem tempting to use, it is not recommended as two proteins should not be compared as there are vast differences in detectability for different proteins; only values for the same protein should be compared between samples.

Setting significance thresholds

Significance as measured by the Fisher's Exact Test for every protein in a comparison and returns a p-value and $\log_2$ fold change value. The p-value provides a quantitative metric for the likelihood of observing the difference in spectral counts of a protein between sample conditions (experimental vs. control). If the likelihood of the difference is due to chance, the p-value is high and the if the likelihood of the difference is not due to chance to p-value is low. Statisticians must arbitrarily determine a threshold for a p-value and a 95% confidence interval is standard in the field, thus a protein was considered significant (likelihood of observing the difference between sample conditions is not due to chance) if the p-value < 0.05. This value can be observed by plotting $-\log_{10}$ of the p-value on a y-axis (i.e. $-\log_{10}(0.05) = 1.3$ or $-\log_{10}(0.01) = 2$). The $\log_2$ fold change value measures the difference in abundance of a protein between the

experimental and control samples. A $\log_2$ fold change of 2 suggests that a protein is 4X more abundant in one sample condition (tagged) compared to the other sample condition (control).

As described in 3.3.1, it is crucial to employ a near-isogenic control sample to identify proteins that interact with the tagged-protein from proteins that appear in the same fractions of the purified CCL (Figure 3.14).

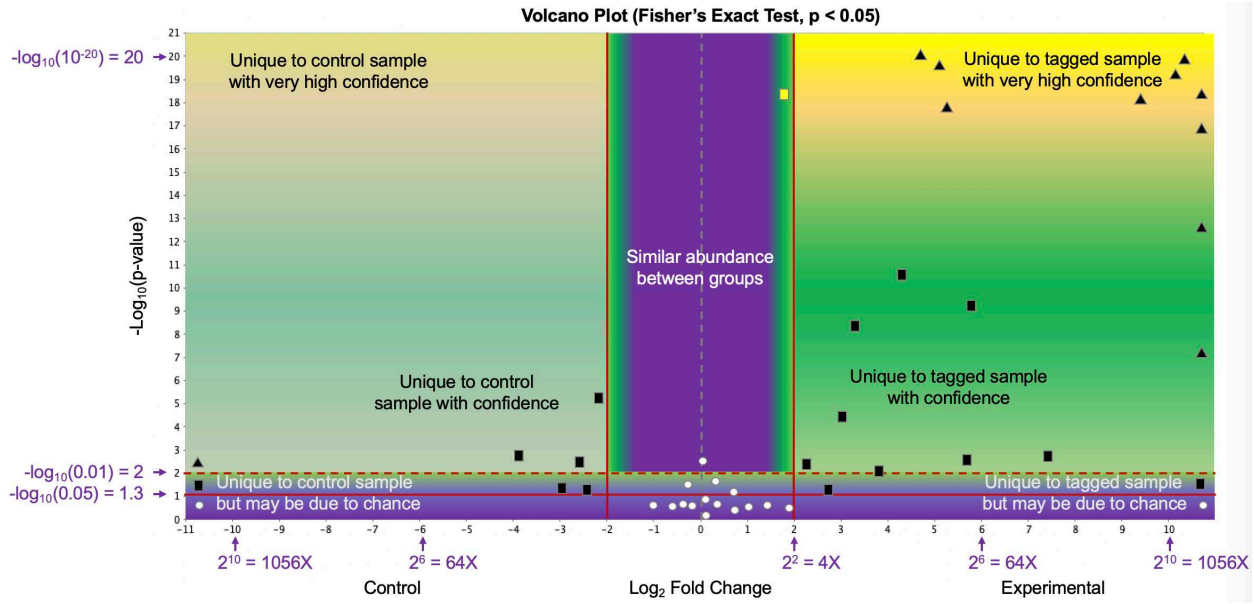


Figure 3.14: A volcano plot comparing the results of a Fisher's Exact Test between control and experimental proteomic MS samples. A volcano plot is a common pictural for displaying the results of a Fisher's Exact Test between two sample groups. The X-axis measures the $\log_2$ fold change in abundance for a protein between sample groups. Proteins are represented as white circles, black squares, or black triangles on the plot. As will be described, proteins represented by white circles are not considered significant. Proteins which appear as black squares or black triangles are considered significant or highly significant, respectively. A protein that returned data at ~equal amounts between groups would appear ~0 (dashed grey line, center of plot), while a protein that returned data 64X greater in the experimental sample compared to the control sample would appear ~ 6 ($\log_2$ of 6 is 64). Proteins that are identified with greater abundance in the experimental (tagged-protein) sample appear on the right side of the plot, while proteins identified with greater abundance in the control (parental, TS559) sample appear on the left side of the plot. Proteins on the far left or far right (> $\log_2$ fold change |10|) were likely not identified in the other sample and thus appear at infinite abundance in the respective sample. The Y-axis represents the likelihood that the difference in spectral counts for a given protein between sample groups is due to chance and is displayed as the $-\log_{10}$ of the p-value. If the p-value is 0.01, the $-\log_{10}$ of 0.01 is 2. A protein that returned a p-value of 0.2 will appear at the bottom of the plot while a protein that returned a p-value of 0.00000001 will appear higher up along the y-axis. Proteins that were identified in both samples but with no difference between samples appear below the horizontal red line demarking the $-\log_{10}$ of 0.05 (1.3) on the y-axis and between the red lines demarking a $\log_2$ fold change of 2 or -2 on the x-axis. Proteins which appear between a $\log_2$ fold change of -2 and 2 are observed to have similar abundance between groups (white circles). Proteins which appear with a $\log_2$ fold change > 2

may be unique to the sample group, but if the p-value is > 0.05 ($-\log_{10}$(p-value) < 1.3), this difference could be due to chance and the protein is not considered significant (white circle). Proteins which appear at a $\log_2$ fold change > 2 and a p-value <0.05 (-log10(0.05) is 1.3) denoted by the horizontal red line and vertical red lines, are significant in the data (black squares and triangles). The dashed horizontal red line indicates a p-value threshold of < 0.01 (-log10(0.01) is 2), if a more stringent cutoff is desired. Proteins which are identified in the upper right quadrant are significant in the tagged sample (black squares and triangles). Proteins which are identified in the upper left quadrant appeared in the control sample but did not appear in the experimental sample. While such a result could indicate that the tag is interfering with these proteins from interacting with their typical partners, the chances of this are low. Proteins that appear significant in the control sample are not of interest in this study. As significance cutoff values are necessary but arbitrary, when a protein appears with a highly significant p-value but a slightly lower abundance than a $\log_2$ fold change of 2 (yellow square), it is manually curated and deemed significant.

For the purposes of this analysis, it was determined that a protein with a p-value of < 0.05 and a $\log_2$ fold change ≥ 2 in experimental conditions was considered significant. After determining standard statistical parameters, only one protein, TK2076, was frequently narrowly excluded from datasets (by occurring with a $\log_2$ fold change ~1.8, Figure 3.14 white square). In the rare case that the tagged protein fell just outside of the threshold of significance (*e.g.* $\log_2$ fold change of 1.8), the threshold was lowered to include the tagged protein and all proteins within this adjusted range in the respective sample comparison; thus, TK2076C was rescued when appropriate.

Tagged-MBH proteins did not pass significance thresholds

The AP-MS method relies on affinity purification techniques to separate the tagged protein sufficiently from most proteins such that the tagged-protein and it's interacting partners can be identified (described in detail in Section 3.2.2). The tagged-protein must appear as significant in experimental samples compared to control samples. One tagged-MBH subunit (MBH-I, TK2088) could not be identified using Ni^{2+} or Cu^{2+}. MBH-I links the membrane-bound portion of MBH to the cytoplasmic arm, thus while MBH-I could be genetically tagged I, it could not be purified sufficiently using affinity purification techniques described here. All other MBH-tagged-proteins were visualized via Western blot and respective fractions were pooled, not all

tagged-proteins appeared significant. Consistently, tagged-proteins of the MBH complex (TK2089, TK2090, TK2091, and TK2093) were not identified as statistically significant in the data. Multiple subunits of the 14-subunit complex were identified, however none were found to be significant in the tagged samples. For example, tagged-MBH-J (TK2089D) was identified a p-value of 0.4 and a $\log_2$ fold change of 1. Had these proteins been identified closer to predetermined thresholds of significance an argument for modifying thresholds would have been appropriate, however, this was not the case.

In troubleshooting this finding, two factors were noted. First, all MBH-tagged proteins were not identified via Western blot when purified using Ni^{2+}, however employing a stronger cation (Cu^{2+}) provided a means to identify the tagged-protein. A Cu^{2+} elution does not restrict a protein from being identified as significant; GAPOR (TK2163) was purified using Cu^{2+} and was identified as significant compared to control samples. Second, when MBH-tagged proteins were successfully purified using Cu^{2+}, three of the four tagged-MBH subunits (TK2089, TK2090, and TK2093) were identified in peak 1 while TK2091 eluted just past peak 1 (fractions B12 - 15 in +S˚ and B11 - 14 in -S˚). Elution in peak 1 does not restrict a protein from being identified as significant; VORα (TK1980C) +S˚ was purified within peak 1 (fractions B7 - 9) and was identified as significant compared to control samples (Figure 3.9).

As the MBH complex is responsible for the production of hydrogen gas, further studies of this complex are warranted but beyond the scope of this project. Alternatives are discussed at the end of this chapter in the Discussion.

3.4 Meaningful interpretation of complex data and method validation

AP-MS is a novel method attempting to capture *in vivo* protein interactions by analyzing large proteomic MS datasets which require stringent statistical parameters to identify interacting proteins and reduce false positive results. Such a method requires validation which can be applied using three techniques, described here.

Of the twenty-nine strains tested, twenty-five returned data that met statistical thresholds (Section 3.3.3, *Tagged-MBH proteins did not pass significance thresholds.*) Each AP-MS comparison identified ~515 proteins on average (range 112 – 822), of which six percent on average (range 0.5 – 20 %) were found to be significant at the thresholds described in section 3.3.3. Previous AP-MS data suggests that each protein maintains ~ sixteen associations *in vivo* (Michaelis et al., 2023). Six percent of 515 suggests an average of ~thirty associations per protein, thus the chosen analysis parameters appear reasonable, if slightly relaxed. A third layer of method validation has been implemented in the development of a ranking system so that high confidence data can be identified from low confidence data (which still passes statistical thresholds). As explained in 3.4.1, the ranking system will provide a means to interpret data with conservative limitations without losing meaningful data.

In generating and analyzing twenty-five tagged-proteins, a protein tagged in one sample may identify a protein tagged in another sample. If tagged-protein X identifies protein Y and tagged-protein Y identifies protein X, proteins have been identified reciprocally, bolstering the suggestion that these data are representative of *in vivo* protein relationships within a cell. Reciprocal findings can be used to validate data in two ways.

First, such relationships should be present in previously determined heterocomplexes. For instance, if a single subunit of a heterotetramer is tagged, the three additional subunits of this complex will be identified as significant in the data. If multiple subunits of a complex are tagged and each tagged subunit identifies all other subunits, such findings provide confidence that the data generated using AP-MS is sound. The twenty-five tagged proteins compose seven supposed heterocomplexes, six of which maintain significant *in vitro* data supporting these associations. Reciprocal associations within the known heterocomplexes are considered the second means of method validation and is discussed in Section 3.4.2.

Lastly, reciprocal interactions beyond known heterocomplexes can support the data. In such a case, one complex identifies a protein (or proteins) from another complex, and when

proteins from the second complex are tagged, the first complex is identified. While such findings provide novel data, they can also provide a third level of method validation and are discussed in Section 3.4.3.

Section 3.4 discusses various ways in which the data can be validated including the application of a ranking system, identification of known heterocomplex subunits, and reciprocal protein identifications.

3.4.1 A ranking system to engender confidence

It was the goal of the analysis to identify as many valid protein associations as possible while limiting false positives. Even with stringent thresholds for analysis, an abundance of significant data remains. A ranking system has been developed to provide a metric which represents the confidence of the data. The ranking system considers all three experimental biological replicates as well as the two statistical thresholds discussed in Section 3.3.3: the p-value and the $\log_2$ fold change value. As p-values are arbitrarily set (most often by a standard practice in the field), a p-value of > 0.05 was chosen, however a p-value of < 0.01 provides increased confidence that the protein did not appear significant by chance. A protein with a p-value of ≤ 0.01 earns a point in the ranking system. Likewise, a $\log_2$ fold change of 2 threshold has been set, but such a value (4X greater in the experimental samples compared to control) is stronger the larger it is. A protein with a $\log_2$ fold change ≥ 4 will be 16X greater in experimental samples compared to control and earn a point in the ranking system. Lastly, with the Fisher's Exact Test, a protein can appear significant in the data if it appears in a single replicate. Logically, a protein which appears in all three experimental replicates would suggest stronger data, thus a protein earns a point for every replicate it appears in. A protein that appears in all three experimental replicates, at a p-value ≤ 0.01, and a $\log_2$ fold change of ≥ 4 would earn a rank score of 5. A protein that falls short in any one category would earn a rank score of 4. A protein that falls short in any two categories would earn a rank score of 3 and so forth.

With this ranking system, ~34% of significant proteins scored a 5, ~44% scored a 4, 17% scored a 3, 3% scored a 2, and 0.1% scored a 1 (Table 3.6, Figure 3.15). A protein with a rank of 5 (~34% of proteins identified) can also demonstrate what the data would look like if a more stringent statistical threshold would have been set (requiring every protein to appear in all three replicates, at a p-value of < 0.01 and a $\log_2$ fold change ≥ 4). Such a cutoff would have generated data where each tagged-protein would maintain an average of ~twelve associations, which is more conservative than that found for the entire yeast proteome (Michaelis et al, 2023). The ranking system can aid in rationally deciphering a confidence metric for each result which can provided for an added value of interpreting meaningful data.

Table 3.6 Number of significant proteins identified for each gene and a count of significant proteins by rank.

	(+) S˚						(-) S˚					
	Significant	Count					Significant	Count				
Gene	Proteins	Rank 5	Rank 4	Rank 3	Rank 2	Rank 1	Proteins	Rank 5	Rank 4	Rank 3	Rank 2	Rank 1
TK0135C	36	17	15	4	0	0	13	4	5	3	1	0
TK0136C	50	25	21	3	1	0	16	3	4	7	2	0
TK0650C	60	35	22	3	0	0	11	2	5	4	0	0
TK0826C	71	24	44	3	0	0	51	21	28	2	0	0
TK1088C	36	14	17	5	0	0	18	9	4	2	2	1
TK1123C	57	13	20	11	13	0	53	19	17	17	0	0
TK1129N	95	25	49	21	0	0	49	15	22	12	0	0
TK1130N	19	3	9	7	0	0	16	7	6	3	0	0
TK1215C	53	11	19	16	6	1	49	16	22	9	2	0
TK1325C	7	4	2	1	0	0	3	3	0	0	0	0
TK1326C	25	7	7	10	1	0	8	1	2	2	3	0
TK1684C	15	3	7	5	0	0	38	11	22	5	0	0
TK1685C	31	7	12	7	5	0	67	14	28	24	1	0
TK1978C	14	10	4	0	0	0	10	9	1	0	0	0
TK1979C	30	15	15	0	0	0	26	11	15	0	0	0
TK1980C	30	9	15	5	1	0	24	15	5	4	0	0
TK1981C	7	5	1	1	0	0	8	7	1	0	0	0
TK1982C	22	6	11	5	0	0	74	27	43	4	0	0
TK1983C	6	3	3	0	0	0	12	7	3	2	0	0
TK1984C	21	13	8	0	0	0	19	12	3	4	0	0
TK2075C	21	10	10	1	0	0	63	17	32	14	0	0
TK2076C	63	11	30	20	2	0	26	4	18	4	0	0
TK2077C	21	5	9	6	1	0	42	2	24	12	4	0
TK2078C	63	24	28	11	0	0	15	5	6	4	0	0
TK2163C	13	4	6	3	0	0	14	5	8	1	0	0
TK2163C-P	4	2	0	2	0	0	NA	0	0	0	0	0
Total	**870**	**305**	**384**	**150**	**30**	**1**	**725**	**246**	**324**	**139**	**15**	**1**
% of All Significant Proteins		**35.1**	**44.1**	**17.2**	**3.4**	**0.1**		**33.9**	**44.7**	**19.2**	**2.1**	**0.1**

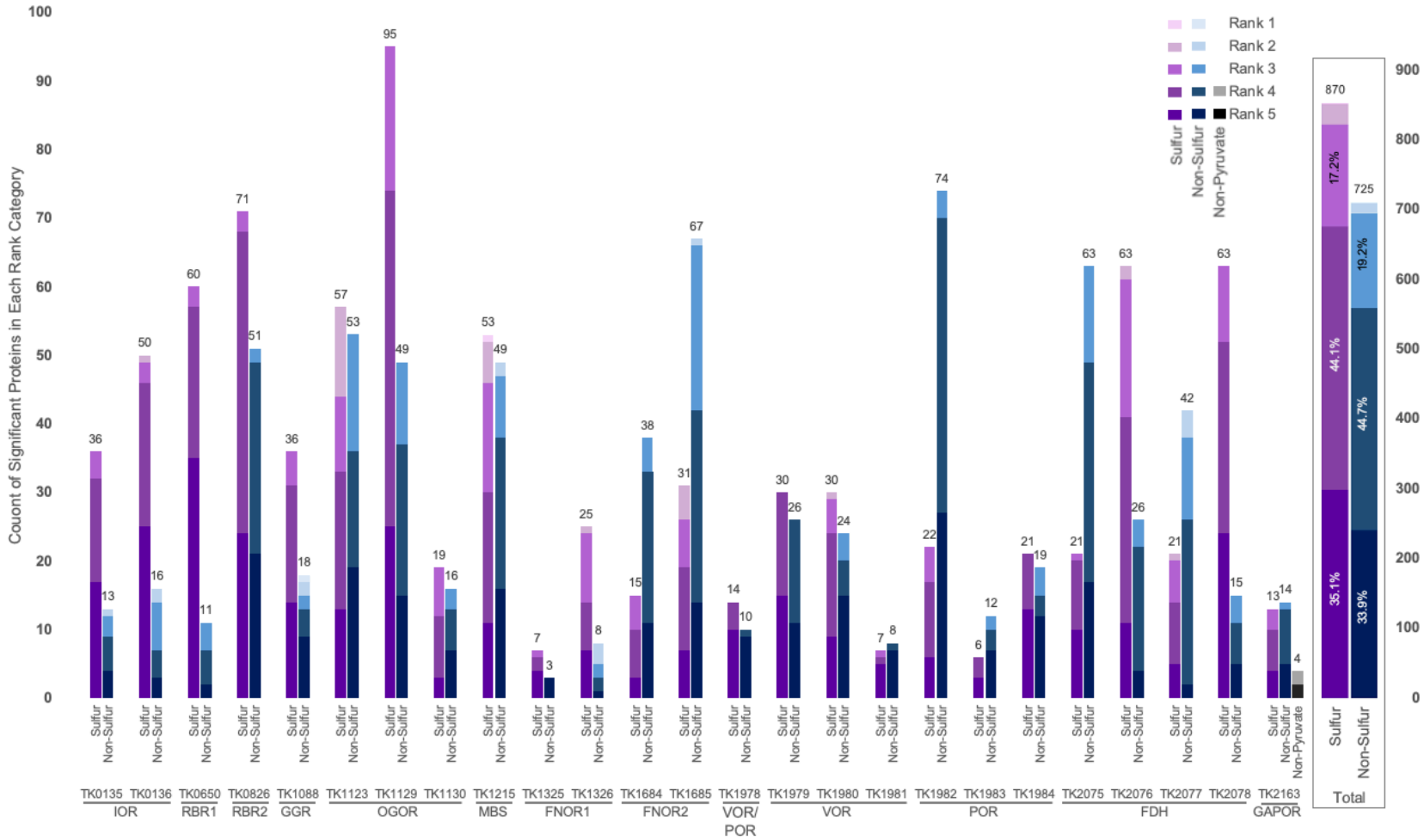


Figure 3.15: Count of significant proteins identified for each tagged protein comparison by rank score. Proteins in the AP-MS data that are identified as significant based on thresholds discussed in Section 3.3.3 are ranked with a score of 5-1 (darker shades for higher rank scores). Here, a count of significant proteins at each rank is shown for each tagged-protein comparison in +S˚ (purple) and -S˚ (blue) conditions with the total number of significant proteins shown at the top of each bar. Rankings of all significant proteins in +S˚ or -S˚ are shown at far right with a modified y-axis to include counts from 0 to 900. The percent of significant proteins which ranked 5, 4, and 3 are shown within the stacked bars for total counts in +S˚ and -S˚ at right which indicate that the majority of significant proteins identified with a rank of 4 (44.1% for +S˚ and 44.7% for -S˚), followed by a rank of 5, 3, 2 and then 1.

3.4.2 Tagging multiple subunits of a heterocomplex provides a means for method validation

A tagged-protein that exists as a subunit of a larger complex can identify their respective subunits in the AP-MS data, and thus, can serve as a simple means of method validation. Here, validation of the method by the identification of known heteromeric complexes is discussed. The majority of data supporting previously determined complexes comes from *in vitro* studies with *T. kodakarensis* or a close archaeal hyperthermophilic species such as, *Pyrococcus furiosis*. The twenty-five proteins in which AP-MS data was successfully collected make up nine complexes while four function as monomers (GGR and GAPOR) or homodimers (RBR1 and RBR2) (Table3.6).[55,56] Five complexes are supported by copious *in vitro* data and all subunits were tagged in this study: VOR, POR, IOR, FNOR1, and FNOR2.[1,3,4,57–59] These complexes are ideal for method validation. Only three out of seven subunits of OGOR were tagged, but these subunits can still provide useful data.[60] The full functional structure of FDH remains elusive and while four subunits were tagged, FDH is not ideal for method validation due to the lack of data identifying specific subunit composition in *T. kodakarensis.*[8] AP-MS data for one complex, MBH, was not successful (Section 3.3.3), and only one subunit of MBS was tagged, thus neither MBH nor MBS cannot be used to validate the method using reciprocal data.[46,47]

In brief, all subunits of VOR, POR, and IOR were identified with every tagged-subunit of the respective complex. FNOR1 and 2 results partially support previous findings and will be noted here but discussed in depth in Chapter 4. OGOR displayed reciprocal data for 2/3 subunits tagged and all seven components of OGOR were identified from the three tagged proteins.FDH showed reciprocal interactions for 3/4 subunits tagged, though it will be noted again that FDH composition has not been extensively confirmed.

Table 3.6. *T. kodakarensis* protein complexes tagged for generating the metabolic protein interactome.

Gene	Protein	kDa	Complex or Acronym	Subunits Tagged/Subunits in Complex	Complex Composition	Source	AP-MS Data Significant
TK0135	Indolepyruvate-Fd oxidoreductase, β subunit	21.89	IOR	2/2	Heterotetramer: $2\alpha2\beta$	Mai and Adams, 1994	Yes
TK0136	Indolepyruvate-Fd oxidoreductase, α subunit	70.9					
TK0650	Rubrerythrin Related Protein	22.46	RBR1	1/1	Likely homodimer	Coulter *et al*., 1999	Yes
TK0826	Rubrerytherin Mn-Catalase	22.53	RBR2	1/1	Likely homodimer	Coulter *et al*., 1999	Yes
TK1088	Geranylgeranyl reductase	44.16	GGR	1/1	Monomer	Masuchi *et al*., 1998	Yes
TK1123	2-Oxoacid:ferredoxin oxidoreductase, γ subunit	20.25	OGOR	3/8	Possible hetero-octamer: $2\alpha2\beta2\delta2\ \gamma$	Mai and Adams, 1996	Yes
TK1129	2-Oxoacid:ferredoxin oxidoreductases, β subunit	31.52					
TK1130	2-Oxoacid:ferredoxin oxidoreductase, α subunit	44.03					
TK1215	Membrane-bound sulfane reductase, NiFe-hydrogenase large subunit 2	45.28	MBS	1/13	13-subunit, membrane-bound complex	Yu *et al*., 2018	Yes
TK1218	Membrane-bound sulfane reductase, MbxM	32.63					
TK1222	Membrane-bound sulfane reductase, MbxF	26.05					
TK1325	Ferredoxin:NADH oxidoreductase, α subunit	53.28	FNOR1	2/2	Heterodimer: $\alpha\beta$	Ma and Adams, 1994;	Yes
TK1326	Ferredoxin:NADH oxidoreductase, β subunit	32.52					
TK1684	Ferredoxin:NADP oxidoreductase, α subunit	52.68	FNOR2	2/2	Heterodimer: $\alpha\beta$	Ma and Adams, 1994;	Yes
TK1685	Ferredoxin:NADP oxidoreductase, β subunit	31.98					
TK1978	2-Oxoisovalarate-ferredoxin oxidoreductase, γ subunit	19.7	VOR & POR	4/4	Heterotetramer: $\alpha\beta\delta\gamma$	Heider *et al*., 1996	Yes
TK1979	2-Oxoisovalarate-ferredoxin oxidoreductase, δ subunit	11.83	VOR				
TK1980	2-Oxoisovalarate-ferredoxin oxidoreductase, α subunit	44.4					
TK1981	2-Oxoisovalarate-ferredoxin oxidoreductase, β subunit	32.61					

TK1982	Pyruvate:ferredoxin oxidoreductase, δ subunit	12.02	POR	4/4	Heterotetramer: $\alpha\beta\delta\gamma$	Blarney and Adams, 1993; Schut *et al.,* 2001	Yes
TK1983	Pyruvate:ferredoxin oxidoreductase, α subunit	43.87					
TK1984	Pyruvate:ferredoxin oxidoreductase, β subunit	36.06					
TK2075	4Fe-4S cluster-binding protein associated with FDH, FDHγ_2	18.34	FDH	4/~4	Possible heterotetramer	Yang *et al*., 2022	Yes
TK2076	Formate dehydrogenase, α subunit	80.1					
TK2077	Formate dehydrogenase, 4Fe-4S cluster-binding protein, FDHγ_3	18.68					
TK2078	4Fe-4S cluster-binding protein associated with FDH, FDHγ_4	17.84					
TK2088	Membrane-bound hydrogenase, MbhI	13.49	MBH	5/14	14-subunit, membrane-bound complex	Yu *et al*., 2020	No
TK2089	Membrane-bound hydrogenase, MbhJ	19.13					
TK2090	Membrane-bound hydrogenase, MbhK	21.44					
TK2091	Membrane-bound hydrogenase, MbhL	48.18					
TK2093	Membrane-bound hydrogenase, MbhN	18.09					
TK2163	Glyceraldehyde-3-phosphate:ferredoxin oxidoreductase	74.16	GAPOR	1/1	Monomer	Makund and Adams, 1995	Yes

Pyruvate oxidoreductase (POR) and 2-oxoisovalerate oxidoreductase (VOR)

VOR (TK1978 - TK1981) and POR (TK1978 and TK1982 - TK1984) are both heterotetramers and share the same γ subunit, TK1978. All four subunits of VOR and all four subunits of POR were tagged and all subunits identified all respective complex subunits in both +S˚ and -S˚ conditions (Figure 3.16 and 3.17). Such extensive identification of all members of these two heterotetramers supports that biomass culturing, harvest, purification procedure, protein handling, LC-MSMS parameters, and analysis thresholds are effective.

Figure 3.16: Results for PORγ (TK1978C +S˚) in Scaffold Viewer using the Quantify Pane. Total spectra from replicate samples of tagged-TK1978C, the PORγ subunit, is displayed in Scaffold Viewer Samples Pane (A) where three biological replicates of the tagged sample (periwinkle) are compared to three biological replicates of the control sample (light salmon). A Fisher's Exact Test with a Benjamini-Hochberg multiple testing correction identifies significant proteins at a p-value threshold of 0.05. Control samples are identified as the reference category and the fold change is reported. Data are sorted in descending order by Total Spectra in the left-most column of the experimental samples. The tagged protein appears in row three and is labeled as "Pyruvate/ketoisovalerate oxidoreductases common subunit gamma." PORγ is a component of both POR and VOR complexes (VOR: TK1978 - 1981, POR: TK1978 & TK1982 - 1984). Here, six out of the seven subunits of POR and VOR are listed in the top six rows (green box) and the final subunit, TK1982, appears in row 10 (blue box). The trypsin protease isolated

from pig appears in row 7 “sp|TRYP_PIG|” (red box). These data are viewed using the Quantify pane (B), accessed by clicking “Quantify” on the left column, which provides visualizations of the analysis. The top left quadrant shows a bar graph where values for total spectra are reported for a protein of choice. Here, the tagged protein, TK1978C, is chosen and total spectra are displayed for control (pink) and experimental (periwinkle) samples. The top right quadrant, “Quantitative Scatterplots” provides a variety of plot options. Here, a Volcano Plot is selected which displays proteins based on significance as determined by the p-value (-log10(p-value)) on the y-axis and the $\log_2$ fold change in abundance between the control (x-axis, left) and experimental (x-axis, right) samples. Based on a p-value < 0.05 (horizontal dashed red line), proteins which are significant are shown as green squares, proteins where are highly significant are shown as green triangles, and proteins which do not pass the p-value cut off are shown as grey circles and triangles at bottom. Note that no fold change cut off to identify significance based on abundance in this volcano plot. The tagged protein is highlighted in yellow in the top right quadrant of the volcano plot. Other proteins which are identified as significant also appear in this region (red circle). The lower left quadrant shows a Venn-diagram of proteins which appear in both samples (115, center), the control sample only (21, left), or the tagged sample only (19, right). The lower right quadrant allows for analysis of the data using Annotation Charts such those applied in a gene ontology (GO), but none are applied here.

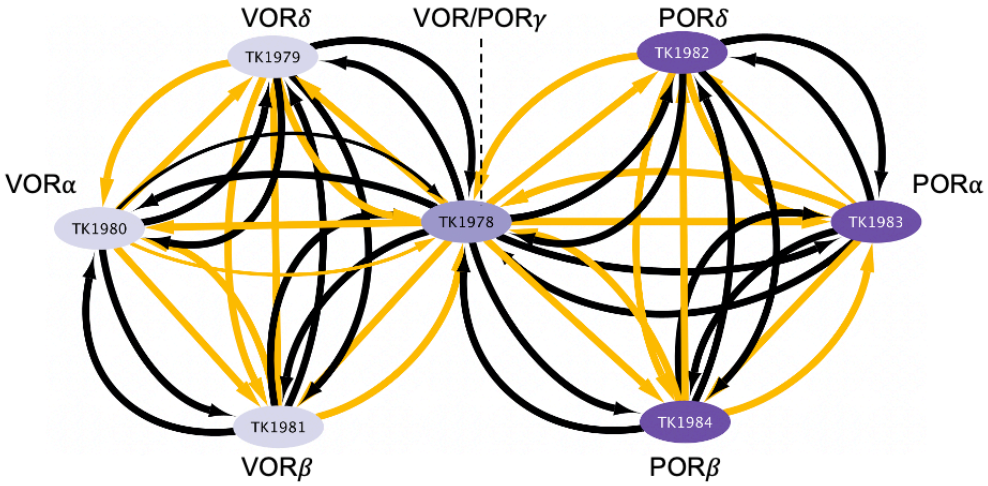


Figure 3.17: Every tagged subunit of VOR and POR identifies every other subunit of the respective complex. VOR (TK1978 - 1981) shown in light purple and POR (TK1978 and TK1982 - 1984) shown in dark purple were tagged, purified, and analyzed by LC-MSMS. A tagged protein was found interacting with the proteins shown here as indicated by a line originating from the tagged protein with an arrow pointing to the identified protein. Proteins were considered significant as compared to the parental control strain (TS559) in +S˚ (orange) or -S˚ (black) using a Fisher’s Exact Test with a Benjamini Hochberg multiple testing correction if the p-value < 0.05 and the $\log_2$ fold change ≥ 2. As visualized using Cytoscape (v3.10.1), every subunit of VOR (TK1978 - 1981) was found by every other tagged-subunit of VOR in +S˚ and -S˚ conditions. Every subunit of POR (TK1978 and TK1982 - TK1984) was found by every other tagged-subunit of POR in +S˚ and -S˚ conditions. Every protein identified was ranked 5 (identified in all three replicates, p-value < 0.01 and the $\log_2$ fold change ≥ 4) except for three association. When TK1980 was tagged, TK1978 was found at a $\log_2$ fold change of 2 in +S˚ and 2.9 in -S˚ resulting in a rank score of 4 (indicated by a thin black or orange line from TK1980 to TK1978). When TK1983 was tagged in +S˚, TK1982 was found in 2 out of 3 replicates, resulting in a rank score of 4 (indicated by a thin orange line from TK1983 to TK1982).

Indolepyruvate oxidoreductase (IOR)

Previous studies suggest that *Pf*IOR functions as an $\alpha 2\beta 2$ tetramer.[3,58] *Tk*IORα (TK0136) and *Tk*IORβ (TK0135) were identified with their respective partners in both +S˚ and -S˚ conditions (Figure 3.18). Data from the tagged-IOR complex supports that all method parameters discussed above were sufficient and appropriate.

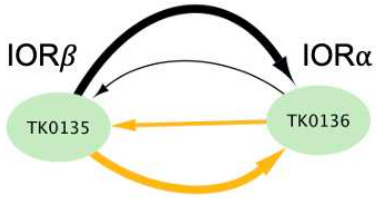


Figure 3.18: Tagged-IOR subunits identify one another. IOR subunits α and β (TK0136 and TK0135, respectively) identify one another in AP-MS data supporting that these operonic proteins associate within the cell as previously described (Schut 2001). Depicted using Cytoscape (v3.10.1), associations seen in +S˚ are shown in orange, while associations seen in -S˚ are shown in black. The thickness of the lines corresponds to the rank observed for each association, with a thicker line representing a higher rank. TK0135 identified TK0136 with a rank of 5 in both +S˚ and -S˚ conditions, while TK0136 identified TK0135 with less confidence (rank 4 in +S˚ and rank 3 in -S˚).

Ferredoxin NAD(P)H oxidoreductase 1 and 2 (FNOR1 and FNOR2)

T. kodakarensis maintains two heterodimeric FNORs which have been studied extensively.[59,61–63] FNOR1 is encoded by TK1325 (FNOR1α) and TK1326 (FNOR1β), while FNOR2 is encoded by TK1684 (FNOR2α) and TK1685 (FNOR2β). When either α subunit was tagged, the β subunit was identified in both +S˚ and -S˚ conditions. However, when the β subunit was tagged, neither α subunit was identified in +S˚ or -S˚ conditions. Interestingly, this was the case for both FNOR1 and FNOR2 (Figure 3.19). Instead, FNOR1α identified FNOR2β and vice versa. These findings will be discussed extensively in Chapter 4, but for the purposes of method validation, it can be observed that reciprocal interactions were not present for previously described heterodimers.

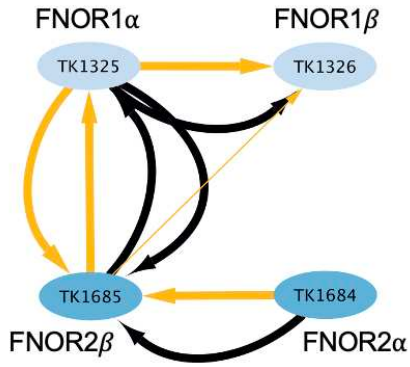


Figure 3.19: Interactome of the two FNOR complexes partially identify one another. All subunits of FNOR1 and FNOR2 were tagged and provided valid AP-MS data. AP-MS data for FNOR1 (TK1325 and TK1326) and FNOR2 (TK1684 and TK1685) shows that both FNORβ subunits were identified from tagged FNORα subunits in both +S˚ (orange) and -S˚ (black) conditions. Neither FNORβ identified the respective FNORα, however both FNOR1α and FNOR2β were observed to interact with reciprocal data. Line thickness corresponds to the rank of each interaction with a thicker line representing a higher rank score. All data was observed with a rank of 5 except for the interaction between tagged-TK1685 and TK1326 in +S˚ which was observed with a rank of 3.

While all subunits of both FNOR1 and FNOR2 were tagged, previously determined relationships were identified in only one direction and not reciprocally. Respective partners that were not identified were identified in the LC-MSMS data but were not significant in the tagged-samples. All FNOR comparisons included three control samples with the exception of TK1684C +S˚ which only included two control samples. All FNOR subunits are >30 kDa and should maintain multiple detectable peptides. It is unknown why respective subunits were not identified.

2-oxoacid:ferredoxin oxidoreductase (OGOR)

OGOR is predicted to be an octamer ($2\alpha2\beta2\delta2\gamma$) encoded by seven subunits: α, TK1125 and TK1130; β, TK1124 and TK1129; γ, TK1123 and TK1126; and δ, TK1131.[64] Homologs have 56, 53, and 45% sequence similarity scores, respectively (Clustal W, v1.83). We tagged OGORα (TK1130), OGORβ (TK1129), and OGORγ (TK1123) (Figure 3.20, ovals). OGORγ (TK1123) was found in association with all six other OGOR subunits in -S˚ and four OGOR subunits in +S˚. OGORβ (TK129) was found in association with three subunits in -S˚ and three subunits in +S˚, while OGORα (TK1130) was found in association with three subunits in -

S˚ and four in +S˚. While the specific composition of OGOR is unknown, the three subunits tagged identified minimally three and maximally all seven corresponding complex subunits. As the structure of the OGOR octamer could be composed of hetero- or homo-dimeric subunits, it is not possible to conclude if the data here fully encapsulates the functional composition of OGOR, however the three tagged subunits of OGOR maintain associations with other subunits of the OGOR complex and such findings can be identified under these method parameters.

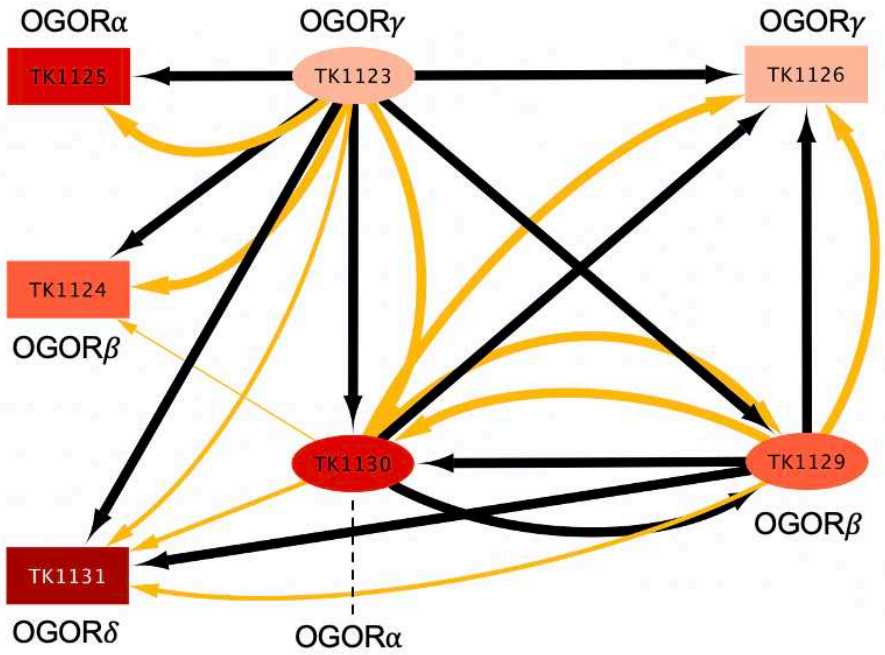


Figure 3.20: Tagged-OGOR subunits α, β, and γ identify all OGOR subunits. AP-MS data from three tagged subunits of OGOR identify all seven subunits of the proposed $2\alpha2\beta2\delta2\gamma$ hetero-octamer. The α subunits (TK1125 and TK1130) are shown in red. The β subunits (TK1124 and TK1129) are shown in orange. The γ subunits (TK1123 and TK1126) are shown in salmon, and the sole δ subunit (TK1131) is shown in dark red. Tagged subunits are shown in ovals while non-tagged but identified subunits are shown in rectangles. Interactions observed in the presence of sulfur are shown in orange, while interactions observed in the absence of sulfur are shown in black. One α (TK1130), one β (TK1129), and one γ (TK1123) subunit have been tagged (ovals). TK1130 and TK1129 identified one another reciprocally in both +S˚ (orange) and -S˚ (black) conditions. While TK1123 was not identified by TK1130 or TK1129, TK1123 identified all other subunits of the OGOR complex (7/7 subunits in -S˚ and 4/7 subunits in +S˚). Line width represents the rank score with a thicker line representing a higher rank. All associations were observed with a rank of 5 except for all three associations between the tagged subunits and TK1131 in +S˚ were observed with a rank of 4, and one association, TK1130 with TK1124, was observed with a rank of 3 in +S˚ conditions.

Formate dehydrogenase (FDH)

FDH has not been thoroughly characterized in *T. kodakarensis*, but FDH has been studied in *T. onnurineus,* a closely related hyperthermophilic archaeon.[8] *T. onnurineus* encodes multiple FDH complexes of which *Ton*FDH3 is the closest to *Tk*FDH. *Ton*FDH3 is a cytoplasmic FDH composed of four subunits: FDH-3α (the tungsten-containing catalytic center), FDHγ_1, and FDHγ_2 which contain Fe-S centers, and FDHβ which is an NAD(P) binding subunit (76.5, 18.3, 18.1, and 39.1kDa, respectively). *T. kodakarensis* encodes one α subunit (TK2076, 80.1 kDa) and three Fe-S cluster subunits (TK2075, TK2077, and TK2078) of 18.3, 18.6, and 17.8 kDa, respectively. The *Ton*FDHβ is proposed to be homologous to TK2074 (39 kDa, Yang et al. 2022), however TK2074 is annotated as a glutamate synthase gene in the Kyoto Encyclopedia of Genes and Genomes (KEGG, genome.jp/kegg).

This study tagged FDHα (TK2076) and three Fe-S binding subunits of FDH (TK2075, TK2077, and TK2078) (Figure 3.21, ovals). TK2074 was not tagged in this study (Figure 3.21, rectangle). Tagged-TK2075 and TK2077 identified all other tagged-FDH subunits and TK2078 identified TK2076 and TK2077 (lacking TK2075) in +S˚ conditions. TK2076 did not identify any other tagged FDH-related protein in +S˚ or -S˚. Tagged TK2077 identified all tagged FDH-related proteins in -S˚, but it was the only one to do so. Tagged TK2075 identified TK2077, but not TK2076 or TK2078 in -S˚. There is strong support for these tagged-proteins to function together in the cell, but without more thorough evidence for the exact functional composition of *Tk*FDH, no assumptions of method validation can be made from these data.

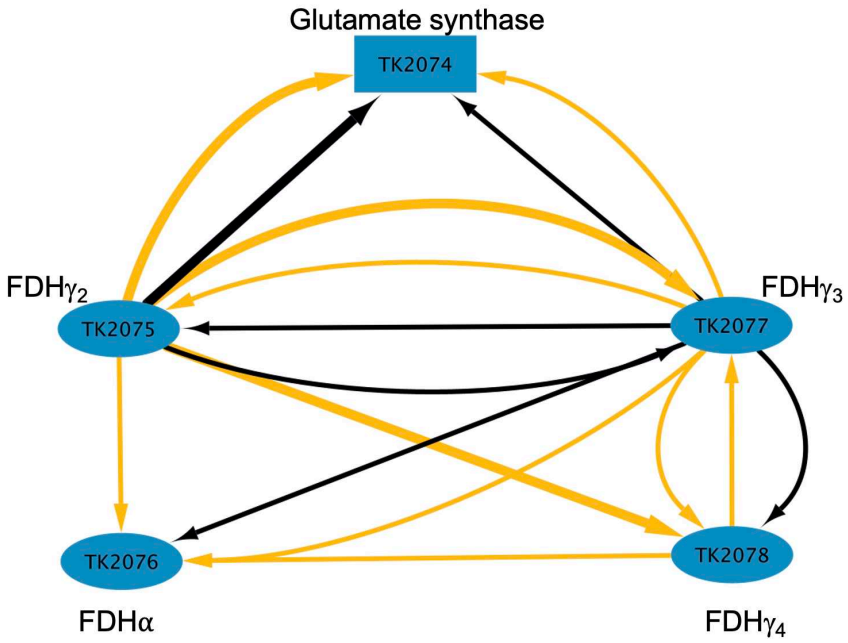


Figure 3.21: Tagged-FDH subunits partially identify one another reciprocally. Four FDH-related subunits (TK2075 – TK2078) were tagged in this study. TK2075, TK2077, and TK2078 are annotated as 4Fe-4S cluster binding proteins associated with FDH while TK2076 is annotated as FDHα. TK2075 and TK2077 identified one another reciprocally in both +S˚ (orange) and -S˚ (black) conditions. TK2075 was found to associate with both TK2076 and TK2078 in +S˚ and TK2077 was found to associate with both TK2076 and TK2078 in both +S˚ and -S˚, but these associations were not reciprocal. TK2078 identified TK2077 in +S˚, but no other FDH-related subunits. TK2076 did not identify any FDH-related subunits in either condition. TK2074 is not tagged (rectangle) but included, because while annotated as a glutamate synthase in T. kodakarensis, it is homologous to FDH3B in *T. onnurineus* and was identified with TK2075 and TK2077 in both +S˚ and -S˚. As data defining FDH in T. kodakarensis is lacking, the presence or absence of reciprocal interactions cannot be used to validate the method but is included here for thoroughness. All interactions displayed here were observed with a rank score of 4 (thinner black or orange line) except for tagged-TK2075 interactions with TK2074 in +S˚ and -S˚ and TK2077 and TK2078 in +S˚ which were observed with a rank of 5.

In sum, three defined protein complexes, VOR, POR, and IOR, reciprocally identified all respective subunits with extreme confidence in both +S˚ and -S˚ conditions. Such clear and strong data supports that the method employed here is identifying associations that have been previously determined. Data for FHD and OGOR supports that the method is identifying probable complex subunits, but these data cannot support or refute the validity of the method. Previous work on the FNOR complexes is thorough. The AP-MS data provided here does not fully support previous findings for the activity of the FNOR1 and FNOR2 heterodimers.

3.4.3 Reciprocal protein interactions aid in method validation

It may not come as a surprise to find that a number of proteins involved in metabolism were found to interact across complexes (discussed in detail in Chapter 4). Many complexes were reciprocally identified, more so in +S˚ conditions than in -S˚ conditions, though those found in -S˚ were also found in +S˚. Reciprocal identification of complexes cannot be assumed but such data supports a robust method.

Reciprocal associations in sulfur and non-sulfur conditions

Four complexes were found to identify reciprocal partners independent of the presence or absence of S˚. POR and VOR, FDH and IOR, FDH and OGOR, and FNOR1 and FNOR2 identified one another in both conditions.

Reciprocal associations only observed in sulfur conditions

OGOR and MBS, IOR and FNOR2, and VOR and FDH were found to associate reciprocally in +S˚ conditions but these associations were not seen in -S˚.

Reciprocal interactions with previously identified ferredoxin proteins

As many of these proteins were chosen because they were found to interact with the three Fds encoded in the *T. kodakarensis* genome, Fd-1 (TK1694), Fd-2 (TK1087), and Fd-3 (TK2012), the Fds provide a means to further identify reciprocal protein identifications.[49] As discussed in depth in Section 3.2.3, Fd-1 and Fd-2 are small proteins (~7 kDa) and their trypsin digested peptides are difficult to impossible to detect using LC-MSMS. Fd-1 maintains no detectable peptides, while Fd-2 maintains one (Table 3.4). When they were identified by the same method in Burkhart *et al*. (2019), it is assumed that the tag provided sufficient protein abundance for these proteins to be identified. It should be noted that in the data published in Burkhart *et al*., the large number of normalized total spectra for Fd-1 (Tk1694) observed in

Table 2 is indicative of a very low number of actual spectral counts for this protein. In Burkhart *et al*., 2019, Fd-1 was observed with many proteins that were included in this study including VOR, FNOR1, RBR1, FNOR2, POR, and IOR. Reciprocal identification of these subunits to Fd-1 was not observed but this is likely due to the indetectable nature of Fd-1. In Burkhart et al. 2019, Fd-2 was observed with FDH, GGR, FNOR1, POR, OGOR, RBR2, and MBH. In this study, Fd-2 was identified with GGR.

Fd-3 is slightly larger (~15 kDa) and digests into ~8 detectable peptides. In Burkhart *et al*., 2019, Fd-3 was found with FDH, MBH, POR, and VOR in -S˚ conditions while no associations are reported in +S˚. In this study, Fd-3 was found with RBR2, GGR, and OGOR in -S˚ and with RBR1, OGOR, MBS, FNOR2, and FDH in +S˚ conditions.

While the AP-MS method will not capture the entirety of interactions, copious identifications can provide novel insights which enhance previous data.

3.5 Discussion

In this chapter the combinatorial nature of genetic, protein, and LC-MSMS techniques and the optimization of these techniques to produced AP-MS data was described in depth. Here a summary of optimized methods, alternative approaches, and method validation is discussed.

Overview of optimized methods for AP-MS

A genetically accessible microbe of interest, *T. kodakarensis*, was modified to express a tagged protein at native expression levels. Twenty-nine strains encoding epitope and affinity-tagged proteins were confirmed by whole genome sequencing and used to generate a metabolic protein network in using AP-MS. These strains were grown in biological replicate and harvested wherein only ~330 mL was necessary to produce AP-MS data. The parental strain was also grown in biological replicate and used as the near-isogenic control. A clarified cell lysate (CCL) was achieved through brisk sonication and the cell contents were purified using

affinity chromatography. Extensive troubleshooting of the purification process provided a consistent procedure to purify all tagged proteins and their co-purifying partners.

Many steps were optimized. A shorter sonication time of five minutes was preferred over a longer sonication. Elution using a linear gradient from 10 - 200 mM imidazole was preferred over multiple isocratic steps or higher imidazole concentrations (500 mM imidazole). Not all tagged-proteins could be purified using a Ni^{2+}-charged column, in which case Cu^{2+} was employed as a divalent cation which offered stronger affinity. Fractions from the purified CCL were analyzed using SDS-PAGE and Western blot against the epitope encoded on the tagged-protein. The tagged-protein most often eluted in 3 - 12 fractions just past the elution of the bulk of proteins ("peak 1") which maintain a natural affinity for the charged-column. The ~three fractions which were identified to contain the majority of tagged-protein were pooled, quantified, TCA precipitated, and sent to our collaborative facility at The OSU's CCIC. It was thought that TCA precipitation was selecting for larger proteins so a raw sample was sent for analysis. The TCA precipitated sample was found to maintain small proteins, so this method of protein handling was chosen.

Protein content of each sample was analyzed using a proteomic tandem mass spectrometry. While the MS facility typically ionizes samples at multiple voltages using FAIMS, it was found that a single voltage was appropriate to identify proteins in our samples. Spectral counts were analyzed using Mascot Daemon via ProteomeDiscoverer and mapped to the *T. kodakarensis* proteome before being returned. Scaffold 5 software was used to compare biological replicates of experimental (tagged-protein) samples to biological replicates of control (parental) samples. Various methods of statistical analysis were compared. Based on the limitations of the software and desiring a consistent analysis for every sample included in the trial, it was determined that a Fisher's Exact Test using a p-value of 0.05 and a Benjamini-Hochberg multiple testing correction was appropriate. Statistical thresholds were then determined and a $\log_2$ fold change of 2 between experimental and control samples and a p-

value cut-off of < 0.05 was deemed appropriate. All samples were required to identify the tagged-protein as significant in the experimental samples as compared to control. No MBH-tagged samples passed this threshold and they were omitted from the analysis, however the remaining twenty-five tagged proteins were identified as highly significant in both +S° and -S° conditions and in all biological replicates.

Alternative approaches to determine MBS and MBH protein networks

T. kodakarensis maintains two membrane-bound respiratory complexes, MBS and MBH, that are of interest as the balance of electrons directed away from MBS and towards MBH will impact H_2 production. While three subunits of MBS and four subunits of MBH were originally included in this study, only one subunit of MBS provided AP-MS data. Alternative methods for gathering information about MBS and MBH protein partnerships will be discussed here.

MBS is a thirteen-subunit complex of which four subunits (MBS-J, K, L, and N) reside in the cytoplasm.[47] Three of the four subunits (MBS-E, H', and L; TK1222, TK1218, and TK1215) were included in the original study group, however only one subunit (MBS-L, TK1215) could be genetically manipulated to encode an affinity and epitope tag. A strength of the AP-MS method as applied in this study is that all protein interactions are seen at native expression levels, however if a protein cannot be genetically modified, data are limited. As a means of fleshing out the MBS network, tagged-MBS subunits could be expressed ectopically using a parental strain which can selectively maintain foreign plasmids such as TS543.[34]

MBH is a fourteen-subunit complex of which four subunits (MBH-J, K, L, and N) reside in the cytoplasm.[46] Five subunits (MBH-I, J, K, L, and N; TK2088 - 2091 and TK2093) were tagged in this study. All tagged MBH-subunits were purified using Cu^{2+} as they were not purified sufficiently when Ni^{2+} was employed as evidenced by Western blot, while MBH-I was not sufficiently purified using Ni^{2+} or Cu^{2+} and was not sent for LC-MSMS analysis. Of the four tagged-subunits that were sent for LC-MSMS analysis, data did not identify any tagged-MBH

subunit as significant compared to the control. Such results indicate that the purification procedure was not sufficient for separating a MBH-tagged protein or that MBH when expressed normally in the parental strain occurs in the same purified fractions when not tagged. The cytoplasmic arm of the MBH complex maintains three Fe-S clusters and one Ni-Fe cluster,[46] thus this protein complex has a natural affinity for metal and may bind to the divalent-charged chelating column.

An alternative tagging and purification method, which does not employ charged metals, could be beneficial to purify MBH and/or MBS. The Strep II tag is a synthetic peptide sequence of 8 amino acids (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) which binds to Strep-Tactin resin and can be eluted using desthiobiotin.[65,66] This alternative method of affinity purification may prove useful for exploring *in vivo* interactions of the MBH complex.

An additional alternative which could be applied to the entirety of this data set involves chemical crosslinking and identification of protein-protein interactions using LC-MSMS, often termed XL-MS. Chemical crosslinking with an agent such as formaldehyde (reversable by heat treatment) or disuccinimidyl sulfoxide (DSSO, cleaved in the gas phase of collision induced disassociation [CID] during MS) have successfully resolve protein interactions.[20,24,67] Such techniques do not rely on genetic manipulation and could be used in combination with the methods presented in this study to define protein-protein interaction networks.

Overview of AP-MS validation

Three methods were employed to validate the data: a ranking system, reciprocal identification of known heterocomplexes, or reciprocal identification of novel complex interactions. First, a ranking system to identify highly significant data from data that minimally passed statistical thresholds was created. Proteins which scored a rank of 5 (most significant) were identified in all biological replicates, at a $\log_2$ fold change ≥ 4 and a p-value ≤ 0.01. A rank of 4 meant that the protein did not meet the more stringent criteria in one area, and a rank of 3

meant that the protein did not meet the more stringent criteria in two areas, etc. 78% of proteins ranked 4 - 5 and 97% of proteins ranked 3 - 5. In this way, significance could further be measured without a loss of data.

Second, the twenty-five proteins included in this study make up seven heterocomplexes for which ~ five (POR, VOR, IOR, FNOR1, and FNOR2) maintain significant *in vitro* data to reliably determine if the respective subunits were found by their tagged counterpart. Three complexes were identified overwhelmingly by all respective subunits in both +S˚ and -S˚ conditions. POR is composed of four subunits and all four subunits reciprocally identified one another. VOR is likewise composed of four subunits and, likewise, all four subunits reciprocally identified one another. All but two reciprocal interactions of the POR and VOR complexes ranked 5 (the other two interactions ranked 4). IOR is composed of two subunits, and both subunits reciprocally identified one another (two interactions were at a rank of 5 and the other two were at a rank of 4). The FNOR complexes did not identify previously determined partners as expected. While the FNORα subunit for both FNOR1 and FNOR2 identified the respective β subunit, the respective β subunit did not reciprocally identify the α subunit. Instead, FNOR1α and FNOR2β reciprocally identified one another. It should be noted that while these associations do not fully support previous data, the FNOR complexes did reciprocally identify one another with high confidence (all associations discussed here were observed with a rank score of 5).

The two complexes for which specific heteromeric data are lacking in *T. kodakarensis* are the OGOR and FDH complexes. OGOR is a supposed $2\alpha2\beta2\delta2\gamma$ hetero-octamer of which three subunits were tagged in this study. Two of the three subunits reciprocally identified one another, while all three subunits identified minimally three and maximally all six of the subunits in the OGOR complex with a majority of the interactions ranking a score of 5. Such data can confirm that the OGOR complex was identified with confidence by the three-tagged subunits.

The composition of *Tk*-FDH is likewise undefined. Four subunits of the proposed complex were tagged in this study, and while all subunits were identified, only three out of the four displayed reciprocal identifications. Many of the interactions observed for FDH occurred at a rank score of 4.

Lastly, the reciprocal identification of complexes can provide confidence in the method and analysis parameters. Four complex pairs identified one another in both +S˚ and -S˚ conditions and an additional three pairs displayed reciprocal identification in +S˚ only. The data generated in this study can be compared to data generated by Burkhart *et al*., 2019 where the three Fds were tagged. Fd-2 and Fd-3 were identified in this study and while data for Fd-3 does not coincide with results published for Fd-3 in Burkhart *et al*., 2019, Fd-2 was identified with GGR in both studies and in both +S˚ and -S˚ conditions.[49]

Conclusion

It was the purpose of this chapter to describe the methodological steps and process for choosing these steps to generate reliable and verifiable AP-MS data to be used to map a novel metabolic protein interactome. When multiple proteins from the same heterologous complex have been analyzed, overlapping, complex networks resolve the dynamic nature of intracellular protein interactions reproducibly. A lack of reciprocal protein interactions suggest that previous data may be incomplete and that proteins may function differently *in vivo* than previously determined by limited *in vitro* environments. Such novel findings of the protein metabolic interactome will be discussed in Chapter 4.

REFERENCES

1. Blarney, J. M. & Adams, M. W. W. *Purification and Characterization of Pyruvate Ferredoxin Oxidoreductase from the Hyperthermophilic Archaeon Pyrococcus Furiosus*. *Biochimica et Biophysica Acta* (1993).
2. Smith, E. T., Blarney, J. M. & Adams, M. W. W. Pyruvate Ferredoxin Oxidoreductases of the Hyperthermophilic Archaeon, Pyrococcus furiosus, and the Hyperthermophilic Bacterium, Thermotoga marítima, Have Different Catalytic Mechanisms'* 1".
3. Schut~, G. J., Menon, A. L. & Adams, M. W. W. *2-Keto Acid Oxidoreductases from Pyrococcus and Thermococcus Litoralis*. (2001).
4. Schut~, G. J., Menon, A. L. & Adams, M. W. W. *ENZYMES OF CENTRAL METABOLISM [12]*. (2001).
5. Schut, G. J., Boyd, E. S., Peters, J. W. & Adams, M. W. W. The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiology Reviews* vol. 37 182–203 Preprint at https://doi.org/10.1111/j.1574-6976.2012.00346.x (2013).
6. Lipscomb, G. L., Schut, G. J., Scott, R. A. & Adams, M. W. W. SurR is a master regulator of the primary electron flow pathways in the order Thermococcales. *Mol Microbiol* **104**, 869–881 (2017).
7. Wu, C. H., Schut, G. J., Poole, F. L., Haja, D. K. & Adams, M. W. W. Characterization of membrane-bound sulfane reductase: A missing link in the evolution of modern day respiratory complexes. *Journal of Biological Chemistry* **293**, 16687–16696 (2019).
8. Yang, J. I., Lee, S. H., Ryu, J. Y., Lee, H. S. & Kang, S. G. A Novel NADP-Dependent Formate Dehydrogenase From the Hyperthermophilic Archaeon Thermococcus onnurineus NA1. *Front Microbiol* **13**, (2022).
9. Allegretti, M. *et al.* In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586**, 796–800 (2020).
10. Russell, R. B. *et al.* A structural perspective on protein-protein interactions. *Curr Opin Struct Biol* **14**, 313–324 (2004).
11. Apic, G., Gough, J. & Teichmann, S. A. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J Mol Biol* **310**, 311–325 (2001).
12. Babu, M., Kagan, O., Guo, H., Greenblatt, J. & Emili, A. Identification of protein complexes in Escherichia coli using sequential peptide affinity purification in combination with tandem mass spectrometry. *Journal of Visualized Experiments* (2012) doi:10.3791/4057.
13. Baspinar, A., Cukuroglu, E., Nussinov, R., Keskin, O. & Gursoy, A. PRISM: A web server and repository for prediction of protein-protein interactions and modeling their 3D complexes. *Nucleic Acids Res* **42**, (2014).
14. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **36**, 880–898 (2018).
15. Clegg, C. & Hayes, D. Identification of Neighbouring Proteins in the Ribosomes of Escherichia coli A Topographical Study with the Cross‐Linking Reagent Dimethyl Suberimidate. *Eur J Biochem* **42**, 21–28 (1974).
16. Cong, Q., Anishchenko, I., Ovchinnikov, S. & Baker, D. *Protein Interaction Networks Revealed by Proteome Coevolution*. *Science* vol. 365 https://www.science.org (2019).
17. Fields, S. & Song, O.-K. *A Novel Genetic System to Detect Protein-Protein Interactions*. (1989).
18. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* **12**, 51–54 (2014).
19. Licata, L. *et al.* MINT, the molecular interaction database: 2012 Update. *Nucleic Acids Res* **40**, (2012).

20. Liu, F., Lössl, P., Rabbitts, B. M., Balaban, R. S. & Heck, A. J. R. The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. *Molecular and Cellular Proteomics* **17**, 216–232 (2018).
21. Robinson, C. V., Sali, A. & Baumeister, W. The molecular sociology of the cell. *Nature* vol. 450 973–982 Preprint at https://doi.org/10.1038/nature06523 (2007).
22. Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).
23. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology* **196**, 801–810 (2012).
24. Schweppe, D. K. *et al.* Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. *Proc Natl Acad Sci U S A* **114**, 1732–1737 (2017).
25. Szklarczyk, D. *et al.* The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605–D612 (2021).
26. Weisbrod, C. R. *et al.* In vivo protein interaction network identified with a novel real-time cross-linked peptide identification strategy. *J Proteome Res* **12**, 1569–1579 (2013).
27. Zhang, Y., Gao, P. & Yuan, J. S. *Plant Protein-Protein Interaction Network and Interactome*. *Current Genomics* vol. 11 (2010).
28. Elhabashy, H., Merino, F., Alva, V., Kohlbacher, O. & Lupas, A. N. Exploring protein-protein interactions at the proteome level. *Structure* vol. 30 462–475 Preprint at https://doi.org/10.1016/j.str.2022.02.004 (2022).
29. Meyer, K. & Selbach, M. Quantitative affinity purification mass spectrometry: A versatile technology to study protein-protein interactions. *Frontiers in Genetics* vol. 6 Preprint at https://doi.org/10.3389/fgene.2015.00237 (2015).
30. Lee, K., Thorneycroft, D., Achuthan, P., Hermjakob, H. & Ideker, T. Mapping plant interactomes using literature curated and predicted protein-protein interaction data sets. *Plant Cell* **22**, 997–1005 (2010).
31. Michaelis, A. C. *et al.* The social and structural architecture of the yeast protein interactome. *Nature* (2023) doi:10.1038/s41586-023-06739-5.
32. Schicho, R. N., Ma, K., Adams, M. W. W. & Kelly3, R. M. *Bioenergetics of Sulfur Reduction in the Hyperthermophilic Archaeon Pyrococcus Furiosus*. *JOURNAL OF BACTERIOLOGY* vol. 175 https://journals.asm.org/journal/jb (1993).
33. Morikawa,' Yoshifumi Izawa,', M., Rashid,', N., Hoaki, T. & Imanakal, T. *Purification and Characterization of a Thermostable Thiol Protease from a Newly Isolated Hyperthermophilic Pyrococcus Sp*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* vol. 60 (1994).
34. Santangelo, T. J., Čuboňová, L. & Reeve, J. N. Thermococcus kodakarensis Genetics: Tk1827-encoded β-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl Environ Microbiol* **76**, 1044–1052 (2010).
35. Bornhorst, J. A. & Falke, J. J. *POLYHISTIDINE AFFINITY TAGS 245 [1 61 Purification of Proteins Using Polyhistidine Affinity Tags*. (2000).
36. Hileman, T. H. & Santangelo, T. J. Genetics techniques for Thermococcus kodakarensis. *Front Microbiol* **3**, 1–12 (2012).
37. Scott, K. A., Williams, S. A. & Santangelo, T. J. Thermococcus kodakarensis provides a versatile hyperthermophilic archaeal platform for protein expression. in *Methods in Enzymology* vol. 659 243–273 (Academic Press Inc., 2021).
38. Lee, C. M., Adamchek, C., Feke, A., Nusinow, D. A. & Gendron, J. M. Mapping protein–protein interactions using affinity purification and mass spectrometry. in *Methods in Molecular Biology* vol. 1610 231–249 (Humana Press Inc., 2017).

39. Gehring, A., Sanders, T. & Santangelo, T. J. Markerless Gene Editing in the Hyperthermophilic Archaeon Thermococcus kodakarensis. *Bio Protoc* **7**, (2017).
40. Atomi, H., Fukui, T., Kanai, T., Morikawa, M. & Imanaka, T. Description of Thermococcus kodakaraensis sp. nov., a well studied hyperthermophilic archaeon previously reported as Pyrococcus sp. KOD1. *Archaea* **1**, 263–267 (2004).
41. Arnold, F. H. *MnAL-AFFINm SEPARATIONS: A NEW DIMENSION IN PR0111N PROCESSING*. http://www.nature.com/naturebiotechnology (1991).
42. Wolters, D. A., Washburn, M. P. & Yates, J. R. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73**, 5683–5690 (2001).
43. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **5**, 699–711 (2004).
44. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376–386 (2002).
45. Graumann, J. *et al.* Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. in *Molecular and Cellular Proteomics* vol. 7 672–683 (2008).
46. Yu, H. *et al.* Structure of an Ancient Respiratory System. *Cell* **173**, 1636-1649.e16 (2018).
47. Yu, H. *et al.* Structure of the respiratory MBS complex reveals iron-sulfur cluster catalyzed sulfane sulfur reduction in ancient life. *Nat Commun* **11**, (2020).
48. Lenski, R. E. *et al.* Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with Escherichia coli. *Proceedings of the Royal Society B: Biological Sciences* **282**, (2015).
49. Burkhart, B. W., Febvre, H. P. & Santangelo, T. J. Distinct physiological roles of the three ferredoxins encoded in the hyperthermophilic archaeon thermococcus kodakarensis. *mBio* **10**, (2019).
50. Senko, M. W. *et al.* Novel parallelized quadrupole/linear ion trap/orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Anal Chem* **85**, 11710–11714 (2013).
51. Geromanos, S. J. *et al.* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683–1695 (2009).
52. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **4**, 923–925 (2007).
53. Searle, B. C. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **10**, 1265–1269 (2010).
54. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods in Molecular Biology* **1151**, 165–188 (2014).
55. Coulter, E. D., Shenvi, N. V & Kurtz, D. M. *NADH Peroxidase Activity of Rubrerythrin*. *Biochemical and Biophysical Research Communications* vol. 255 http://www.idealibrary.comon (1999).
56. Mukund, S. & Adams, M. W. W. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon Pyrococcus furiosus. *Journal of Biological Chemistry* **270**, 8389–8392 (1995).
57. Heider, J., Mai, X. & Adams, M. W. W. *Characterization of 2-Ketoisovalerate Ferredoxin Oxidoreductase, a New and Reversible Coenzyme A-Dependent Enzyme Involved in Peptide Fermentation by Hyperthermophilic Archaea*. *JOURNAL OF BACTERIOLOGY* vol. 178 (1996).

58. Mai, X. & Adams, M. W. W. Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon Pyrococcus furiosus. A new enzyme involved in peptide fermentation. *Journal of Biological Chemistry* **269**, 16726–16732 (1994).
59. Ma, K. & Adams, M. W. W. *Sulfide Dehydrogenase from the Hyperthermophilic Archaeon Pyrococcus Furiosus: A New Multifunctional Enzyme Involved in the Reduction of Elemental Sulfur*. *JOURNAL OF BACrERIOLOGY* vol. 176 https://journals.asm.org/journal/jb (1994).
60. Mai, X. & Adams, M. W. W. *Characterization of a Fourth Type of 2-Keto Acid-Oxidizing Enzyme from a Hyperthermophilic Archaeon: 2-Ketoglutarate Ferredoxin Oxidoreductase from Thermococcus Litoralis*. *JOURNAL OF BACTERIOLOGY* vol. 178 https://journals.asm.org/journal/jb (1996).
61. Adams, M. W. W. & Kletzin', A. *OXIDOREDUCTASE-TYPE ENZYMES AND REDOX PROTEINS INVOLVED IN FERMENTATIVE METABOLISMS OF HYPERTHERMOPHILIC ARCHAEA*. (1996).
62. Nguyen, D. M. N. *et al.* Two functionally distinct NADP+-dependent ferredoxin oxidoreductases maintain the primary redox balance of Pyrococcus furiosus. *Journal of Biological Chemistry* **292**, 14603–14616 (2017).
63. Hidese, R. *et al.* Gene regulation of two ferredoxin:NADP+ oxidoreductases by the redox-responsive regulator SurR in Thermococcus kodakarensis. *Extremophiles* **21**, 903–917 (2017).
64. Yokooji, Y., Sato, T., Fujiwar, S., Imanak, T. & Atomi, H. Genetic examination of initial amino acid oxidation and glutamate catabolism in the hyperthermophilic archaeon Thermococcus kodakarensis. *J Bacteriol* **195**, 1940–1948 (2013).
65. Schmidt, T. G. M. & Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* **2**, 1528–1535 (2007).
66. Skerra, A. & Schmidt, T. O. M. *StFep-TAG/STREPTAVIDIN 271 [18] Use of the Strep-Tag and Streptavidin for Detection and Purification of Recombinant Proteins*. (2000).
67. Liu, F., Rijkers, D. T. S., Post, H. & Heck, A. J. R. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat Methods* **12**, 1179–1184 (2015).

CHAPTER 4
A HIGHLY CONNECTED METABOLIC PROTEIN INTERACTOME DRIVES ENERGY PRODUCTION IN *THERMOCOCCUS KODAKARENSIS* [3]

Summary

Proper shuttling of electrons is necessary both for cellular redox balance and energetic gains. Identification of *bona fide in vivo* protein assemblies -- and how such assemblies dictate the totality of cellular metabolism -- is thus an important milestone to understand the regulation imposed on metabolism, energy-production, and energy-conservation. Proteinaceous electron carriers allow for additional layers of regulation in metabolism.[1,2] While *in vitro* investigations of enzyme activities reveal key parameters under defined conditions, *in vitro* studies alone often fail to reveal dynamic, regulatory, and key protein-protein interactions.[3,4] While efforts to define large networks of protein-protein partnerships have been completed for some model species,[5–8] efforts to establish protein-protein partnerships in Archaea, and particularly hyperthermophilic Archaea with biotechnological and biofuel potential, are rare.[2,9] Given the potential for rational, genetic manipulations of biofuel- and biotech-promising archaea to yield transformative results for major markets, it is a priority to define how the major metabolisms of such species are controlled, at least in part, by *in vivo* protein assemblies, and from such, define routes of energy flux that can be most efficiently altered towards biofuel or biotechnological gains.[10–13] Affinity purification followed by proteomic mass spectrometry (AP-MS) is one such combinatorial method which has the potential to identify intracellular partnerships *en masse* quickly and accurately aiding in the ability to understand protein partnerships in a manner more closely resembling the complexity found in biological systems.[7,14] Purification of tagged-protein complexes and retention of protein interactions established *in vivo* in the absence and presence

[3] Data presented in this chapter in part represents the efforts of talented students and/or colleagues under the tutelage of Seré Williams including: Danielle Riley, Teagan Rockwood, David Crosby, Kate Call, Jared LeCuyer, Lucy Latta, and Isabella Bendorf.

of S° permits identification of redox-associated binding partnerships.[3] AP-MS of twenty proteins involved in central metabolism redefined long-standing assumptions about protein assembly of multimeric complexes *in vivo.* We demonstrated that metabolic networks exhibit flux depending on the availability of terminal electron carriers and identified novel metabolic partners, such as rubrerythrin proteins which likely play a key role in maintaining local redox balance, possibly in lab evolved strains which exhibit fast growth from ideal nutrient conditions and likely exposure to oxic stress. Repeated protein associations were observed which linked amino acid fermentation, acid oxidation, and respiratory systems suggesting that proteins which perform disparate metabolic activities may form distinct metabolic hubs within the cell. Such *in vivo* resolution of metabolic protein networks will allow for rational tuning of electron flux.

4.1 Introduction

Thermococcus kodakarensis is an anaerobic, hyperthermophilic archaeon that efficiently ferments amino acids and sugars resulting in sufficient energy gains to allow culture doubling every ~ forty minutes under optimal conditions.[15,16] *T. kodakarensis* utilizes a modified Embden-Meyerhof-Parnas (EMP) glycolytic pathway that yields a net gain of only two ATP per unit glucose directly from central metabolic processes.[17,18] Catabolic efforts also liberate many high-energy electrons that can be partitioned to i) mature lipids, ii) generate reduced small molecules such as NADH, or iii) utilized to generate a chemiosmotic gradient to yield higher levels of ATP production via a charged membrane and a A_0A_1-ATP synthase.[19–21]

The selective partitioning of electrons in *T. kodakarensis* is facilitated by three ferredoxin (Fd) isoforms that direct and modulate electron flux.[17] *T. kodakarensis* can utilize two distinct terminal electron acceptors (H^+ or S°), and recent work has begun to define the metabolic routes of electron flux *in vivo*.[1,2,22,23]

In this chapter, we will begin with an overview of the unique metabolism of *T. kodakarensis* and the value of understanding protein networks in the context of metabolism. We

will then describe affinity purification mass spectrometry (AP-MS), a novel method to identify protein partnerships *in vivo* and how it has been employed to first identify native protein interactions but then can be used to ultimately direct electron flux. Results for the native metabolic interactome are presented here, while the engineering of novel proteins towards directed electron flux is presented in Chapter 5.

4.1.1 Overview of metabolism in *T. kodakarensis*

Mitochondrial oxidation phosphorylation, reliant on reduced co-factors generated via glycolysis and the Krebs cycle, provides the bulk of energy (ATP and ATP-equivalents) in all eukaryotes.[21] The energy gain per unit glucose in eukaryotes (~36-38 ATP) dwarfs the energetic gain per glucose in prokaryotes, particularly Archaea that lack a conventional TCA cycle.[18,24,25] Most non-methanogenic archaeal metabolisms utilize a modified glycolytic pathway, largely conforming to the Embden-Meyerhof Parnas (EMP) pathway but reliant on unique co-factors, that yields no net ATP-gain in the conversion of glucose to pyruvate and yields just two ATP per unit glucose through the complete deconstruction of glucose to acetate and CO_2 (Figure 4.1).[20] Such minimal energetic yields are typically deemed too low to support substantial growth, and thus fast-growing archaea are dependent on auxiliary mechanisms to extract additional ATP from low-energy inputs.[26] Although direct ATP-yields from archaeal glycolysis are modest, additional energy gains can be extracted from the valuable reduced co-factors resultant from glycolysis. Most prokaryotes use electron transport mediated phosphorylation, dependent on extracellular ion gradients for these additional energy gains. However, the mechanisms employed to generate extracellular ion gradients and the terminal electron acceptors vary dramatically in the biosphere and mechanisms to direct electrons to these diverse pathways are equally complex and poorly understood.[24]

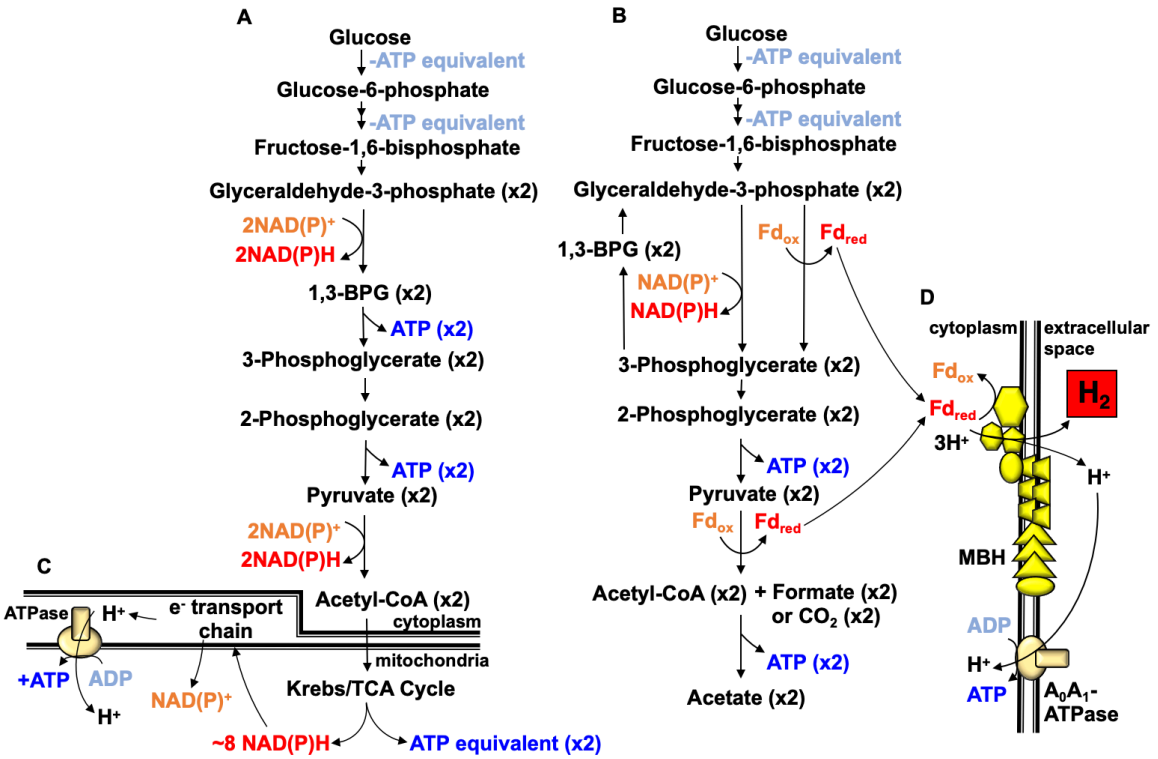


Figure 4.1: Eukaryotic and archaeal glycolysis compared. In eukaryotes (A, C) and archaea (B, D), two ATP-equivalents are consumed converting glucose into glyceraldehyde-3-phosphate (x2). When glyceraldehyde-3-phosphate is converted into 3-phosphoglycerate, 2ATP and 2NAD(P)H are produced in eukaryotes, but archaea only gain reduced products. 2-phosphoglycerate conversion into pyruvate produces 2ATP in both domains. In eukaryotes, pyruvate is converted into acetyl coenzyme A (CoA) and enters the Krebs/TCA Cycle, where many reduced cofactors (including NAD(P)H among others) are produced along with 2ATP equivalents. Reduced cofactors feed (C) the electron transport chain, generating a proton gradient in the inner membrane space and driving ATP production. Depending on terminal electron acceptors, about 1.5-2.5 ATP are produced per reduced cofactor, NAD(P)H shown here. In archaea, (B) pyruvate is converted into acetyl-CoA, and with the addition of formate or CO_2, an additional 2ATP are gained in the final breakdown to acetate. Reduced products feed a membrane-bound hydrogenase (D) where a proton gradient is generated in the production of extracellular H_2, driving ATP production through an ATPase.

Given the importance of the reduced co-factors resultant from glycolysis, proper handling of these co-factors is often tightly regulated. Proper shuttling of electrons is necessary both for cellular redox balance and energetic gains and often these processes are in direct competition.[23] Reducing equivalents are required for the maturation of lipids, as enzymatic co-factors, and for the synthesis of many small- and macro-molecules, but the flux of electrons to each biosynthetic pathway necessarily removes electrons from reactions coupled to total

cellular energy gains.[19] Cells must therefore control the intracellular electron flux based on current and changing needs, including the availability of different terminal electron acceptors.[20,27–30]

High-energy electrons are too valuable and too reactive to indiscriminately distribute. Electron flux is thus typically controlled by i) the production of electron carriers and ii) selective interactions of electron carriers with specific electron donors and electron acceptors.[1,23,31] Many species rely on the small molecule electron carriers FAD^+ or $NAD(P)^+$ (reduced by electron capture to FADH and NAD(P)H, respectively) for redox balance and transfer of electrons to acceptors linked to energy-producing ion gradients, but the thermo-intolerance of FAD^+ and $NAD(P)^+$ precludes their use in hyperthermophiles.[32,33] These common small molecules simply break down too quickly to provide a meaningful electron transport system at high temperature.

4.1.2 Proteinaceous electron carriers (PECs)

Hyperthermophilic species encode proteinaceous electron carriers (PECs), classically termed ferredoxins (Fds), but PECs also include flavodoxins, thioredoxins, rubredoxins and many cytochrome C proteins (Figure 4.2).[1]

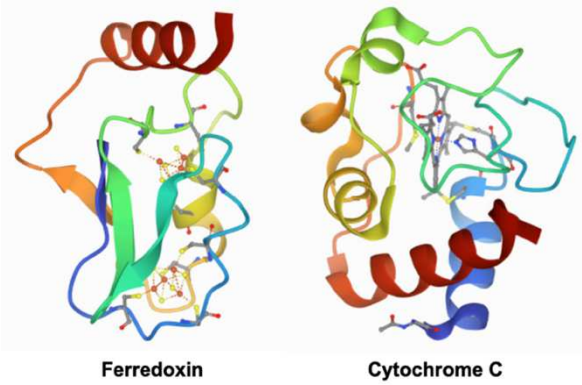


Figure 4.2: Common proteinaceous electron carriers. *Three dimensional structures of 4Fe/4S-ferredoxin (left) and cytochrome C (right) with coordinating central iron atoms. PDB: 1BLU and 1HRC, respectively.*

The use of proteinaceous electron carriers is likely due to three factors; nicotinamide dinucleotide-based electron carriers are known to break down into toxic byproducts at high temperatures,[32] proteinaceous electron carriers allow for additional layers of regulation in metabolism,[2,3,30] and certain proteinaceous electron carriers can reduce protons to hydrogen while NADH cannot (Figure 4.3).[21]

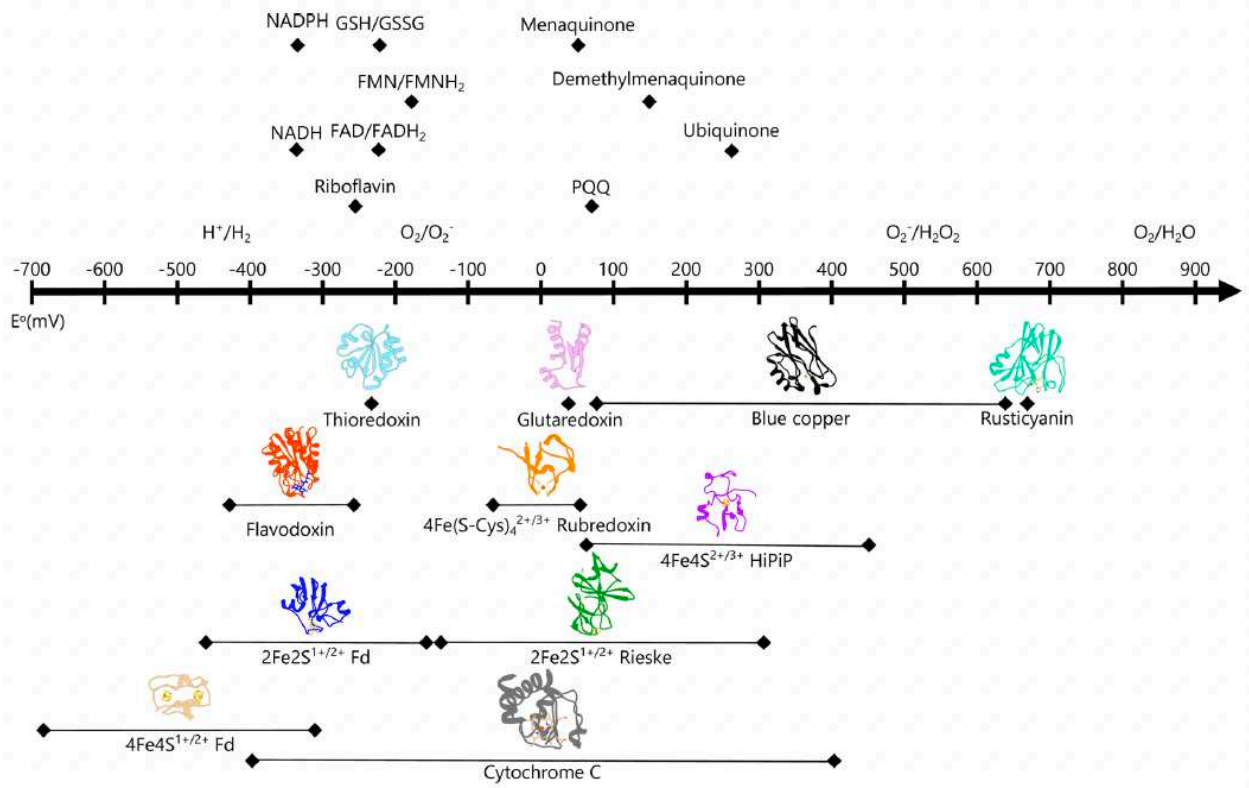


Figure 4.3: Reduction potentials of small molecules and proteinaceous electron carriers. Reduction potential ranges in E˚ (mV) for small molecules (above scale bar) and common proteinaceous electron carriers (below scale bar). A range of midpoint reduction potentials is observed for proteins as demonstrated by the range between diamonds at either end. Fd with varying Fe/S centers display different reductions potentials. Fd which maintain 2Fe2S metal centers cannot reach lower reduction potentials observed from Fd with 4Fe4S centers. Note that the reduction potential of H^+/H_2 sits ~ -410 E˚ (Atkinson *et al*., 2016).

Protein-based electron carriers can be more thermotolerant than $NAD(P)^+$ and classically coordinate metal centers that permit electron shuttling. Additional advantages have been long predicted, as coordinated production of a specific Fd could, in principle, permit selective transport of electrons from specific electron donors to specific electron acceptors.

Selective electron transport could provide routes for electron flux that favor individual reactions without demanding cellular-level shifts in redox balancing.

The *in vivo* redox landscape of protein associations and the transfer of electrons from one protein to another is of great interest for regulating reductant flux. The dynamic nature of protein surface composition, charge distribution, and structure influences protein-protein associations thus,[34] such associations between redox proteins as it relates to the transfer of electrons can define the metabolic outcomes of these systems, yet little is known about the full network of metabolic protein interactions *in vivo*. In theory, the nature of catabolic inputs could direct electron flux through a specific pathway generating specific products (Figure 4.4). For instance, catabolic inputs such as carbohydrates will necessarily require different catalytic protein responses than those involved in amino acid fermentation. As depicted in Figure 4.4, if Donor #1 is reduced upon the breakdown of glucose and preferentially transfers this electron to proteinaceous electron carrier (PEC) #1, PEC #1 may preferentially reduce Acceptor #1. The reduction of Acceptor #1 will generate a specific product that is distinct from that which would result from the reduction Acceptor #2 or #3. Similarly, these networks may be responsive to the presence or absence of a terminal electron acceptor, such as elemental sulfur (S°).[29,31] PEC #3 may most often associate with Acceptor #3 when S° is present, however, in the absence of this terminal electron acceptor, PEC #3 may associate with Acceptor #2, there by generating H_2 gas in -S° conditions when H_2S would otherwise be produced. Such defined and adaptable regulation is not possible with reduced small molecules such as NADH.

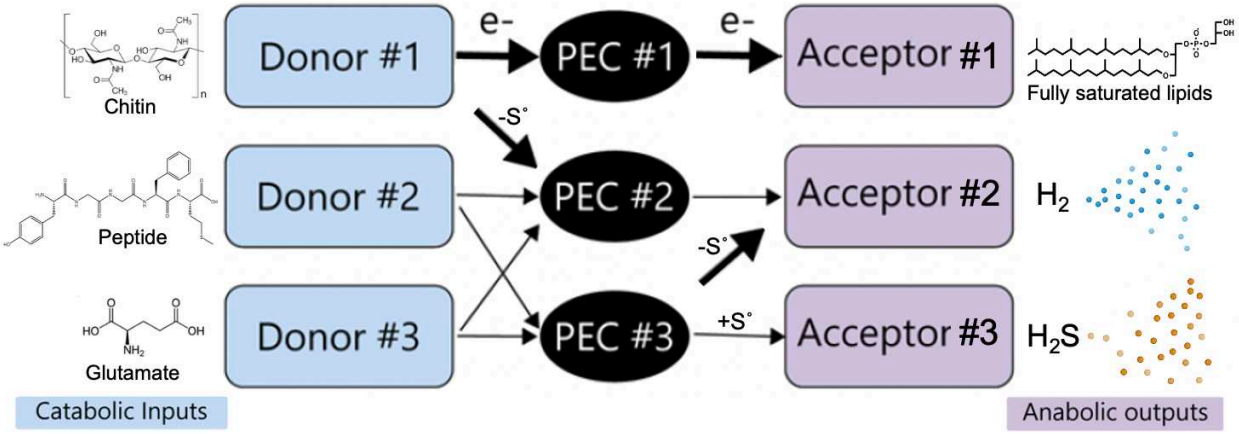


Figure 4.4: Electron transfer pathways between PECs with preferred electron donors and electron acceptors. Proteinaceous electron carriers (PECs, black ellipses at center) will associate with specific electron donors (light blue rectangles) and electron acceptors (light purple rectangles) such that electron flux may be regulated towards one anabolic outcome over another. In the case of chitin (a catabolic input at left), Donor #1 will associate with PEC #1, wherein PEC#1 will become reduced and associate with Acceptor #1. Acceptor #1 will transfer the electron to an unsaturated lipid moiety, further saturating the lipid chain. In the absence of sulfur, the electrons which would otherwise be directed to PEC #1 are directed to PEC #2. When PEC #2 is reduced, PEC #2 preferentially associates with Acceptor #2 leading to the generation of hydrogen gas. A similar regulatory pathway is depicted in the bottom row, but instead of the donor and PEC responding to the presence or absence of sulfur, the association between the PEC and the acceptor is responsive to the presence or absence of sulfur. In this way, PECs provide regulated and adaptable electron transfer. Mapping specific metabolic Donor:PEC:Acceptor relationships is of interest for directing electrons to valuable bioproducts such as isoprene lipids which can be extracted and used for the manufacturing of perfumes, for example, or towards H_2 which is a valuable biofuel. (Adapted from Atkinson et al., 2016.)

4.1.3 Affinity purification tandem mass spectrometry (AP-MS) identifies *in vivo* protein partnerships and will resolve metabolic networks

Affinity purification followed by tandem mass spectrometry (AP-MS) is a method to determine *in vivo* protein associations,[14,35,36] thus this method can be used to define proteinaceous electron donors and acceptors and the impact of different terminal electron acceptors (sulfur or protons) on these associations (Figure 4.5). The methodological optimization of AP-MS for realizing a metabolic interactome in *T. kodakarensis* is described in detail in Chapter 3. The potential for funneling of electrons via Fds between coupled electron donors and acceptors is the focus of Chapter 5.

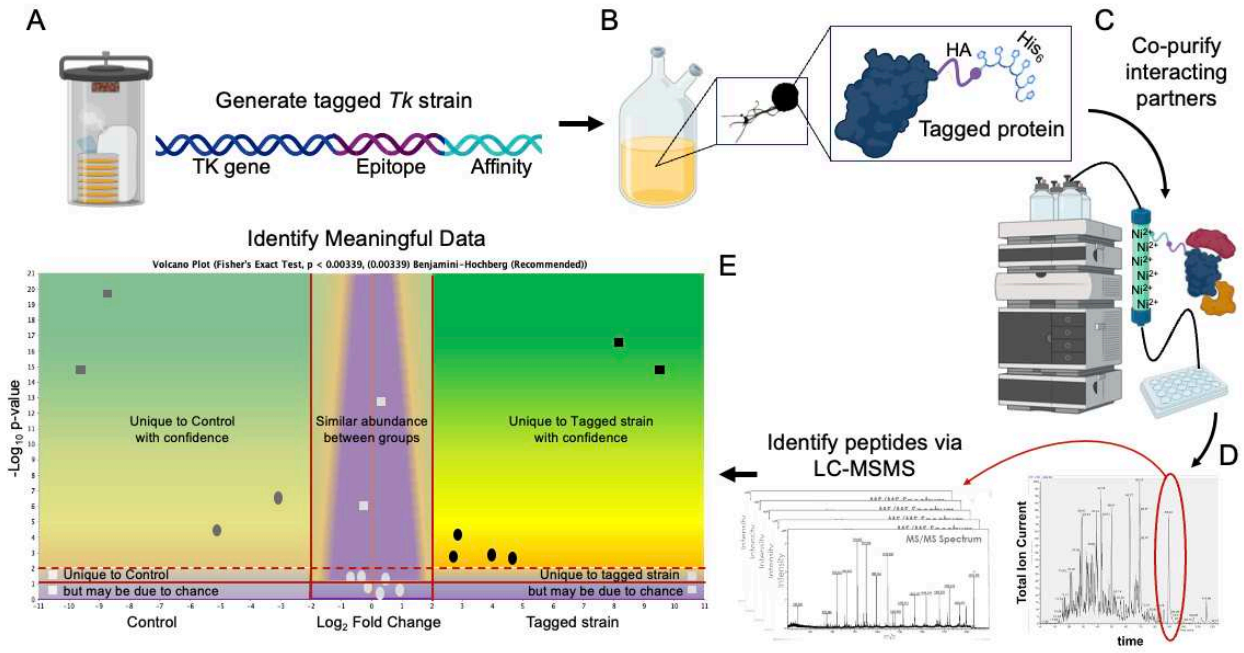


Figure 4.5: Overview of the AP-MS method identifies *in vivo* partnerships from tagged proteins at native expression levels. *T. kodakarensis* strains are modified to encode a genomic epitope and affinity tag at the end of the gene sequence (A) using targeted mutagenesis (Sections 2.5.2 and 2.5.3) and selection strategies (Section 2.2.2). Sequence confirmed strains are cultured to produce biomass in biological replicates (B) (Section 3.2.1). Actively dividing cells are pelleted before being lysed and clarified. The clarified cell lysate (CCL) is purified over a prepared Ni^{2+}-charged affinity column (C) (Section 3.2.2) and fractions containing the tagged protein of interest and co-purifying partners are identified by Western blotting (Figure 3.8D). Fractions are pooled and proteins are TCA precipitated in preparation for shipment to the MS facility. Peptides from trypsin-digested proteins are separated by liquid chromatography followed by tandem mass spectrometry (LC-MSMS) (D) (Section 3.2.3). Identified peptides are mapped to the *T. kodakarensis* proteome. Proteins identified in tagged samples are compared to those identified in the parent strain (a near isogenic control, TS559) (Chapter 2.2) to determine statistically significant, meaningful differences in the tagged strain as compared to the parental strain (E) (Section 3.3). Data from three biological replicates of each tagged protein of interest are used to map a metabolic protein network.

Using the AP-MS method, Burkhart *et al*. showed that the three Fds exhibited specific and distinct interacting partnerships (Figure 4.6).[2]

Fd1 was found to interact with multiple catabolic protein complexes involved in amino acid fermentation including ketoisovalerate (or 2-oxo-isovalerate oxidoreductase, VOR), indolepyruvate oxidoreductase (IOR), and 2-oxoacid:ferredoxin oxidoreductase (OGOR) as well as pyruvate oxidoreductase (POR) which converts pyruvate into acetyl-CoA and CO_2. Fd2 interacted with IOR, OGOR, and POR, but was not found in association with VOR, while Fd3

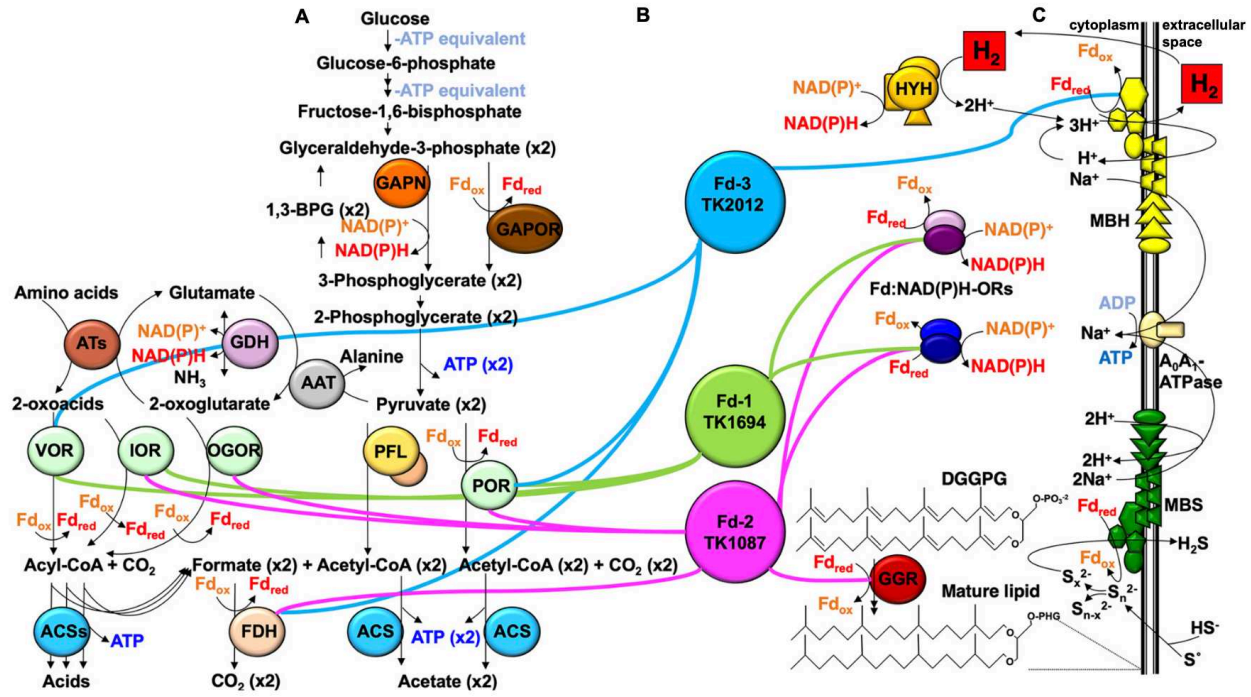


Figure 4.6: Ferredoxins link catabolic and anabolic processes through distinct partnerships. Glucose and amino acids are fermented to generate minimal ATP, pyruvate, and acyl-CoA (A). In addition, cofactors such as $NAD(P)^+$ and PECs such as Fd_{ox} are reduced to NAD(P)H and Fd_{red}, respectively. The three Fds encoded in the T. kodakarensis genome (Fd-1, TK1694; Fd-2, TK1087; and Fd-3, TK2012) displayed distinct associations with electron donors and electron acceptors (B, colored lines linking Fds with respective interacting partners). Fd-1 appears to primarily be involved in the generation of reduced cofactors, while Fd-2 interacts tightly with GGR resulting in the maturation of isoprenoid lipids. T. kodakarensis maintains two membrane bound respiratory complexes (C), membrane bound hydrogenase (MBH, yellow) or membrane bound sulfane reductase (MBS, green) which generate H_2 and H_2S, respectively, creating a proton gradient which drives ATP synthase at the membrane (light yellow). Fd-3 was the only Fd found in association with the MBH, suggesting Fd-3 is linked to H_2 production. (Adapted from Burkhart *et al.*, 2019.)

was only found with VOR and POR. Energetic gains from the resulting acyl-CoA moieties produced in fermentation reactions are further realized in two ways: i) from the conversion of acyl-CoA to acids whereby an ATP is generated or ii) by formate dehydrogenase (FDH) which can further reduce a ferredoxin and in the oxidation of formate to CO_2. Both Fd2 and Fd3 were found in association with FDH.

Anabolic activities of the Fd interactome were likewise distinct (Figure 4.6B). Fd-1 interacted with the two heterodimeric ferredoxin NAD(P)H oxidoreductases (FNOR1 and

FNOR2) in the generation of reduced small molecules. While Fd-2 showed association with the FNORs, the primary anabolic interacting partner was the monomeric geranylgeranylreductase (GGR), responsible for lipid maturation. Fd3 was found in association with four subunits of the membrane bound hydrogenase (MBH-L, K, J, and I) and four subunits of the ATP synthase (subunits H, A, B, and D) suggesting the primary role of Fd3 is related to H_2 generation.

4.2 Proteins included in this study

We have generated a large collection of strains wherein key members of both catabolic and anabolic pathways are tagged for AP-MS to develop a more wholistic understanding of the metabolic network in *T. kodakarensis* (Table 4.1). Purification of tagged-protein complexes and retention of protein interactions established *in vivo* in the absence and presence of S° permits identification of redox-associated binding partnerships. The high-confidence peptide identification and low-input requirements offered by LC-MSMS provide a simple and economical route to establishing protein interactions in complex mixtures. Many of the proteins included in this analysis were shown to interact with the three Fds by the same method.[2] While optimization of the AP-MS method is discussed in detail in Chapter 3, results of the collective analysis of the data are presented here.

In summary, twenty-five proteins were tagged and successfully analyzed using the AP-MS method (described in detail in Chapter 3) in both sulfur and non-sulfur conditions. Three biological replicates of each protein were compared to biological replicates of a control sample to identify significant proteins in the tagged sample representing *in vivo* protein associations of the tagged protein.

Table 4.1: *T. kodakarensis* proteins included in generating the metabolic protein interactome.

Gene[a]	Protein	kDa	Acronym and Function	Composition, No. Tagged Subunits	Source	Identified with[b]		
						Fd1	Fd2	Fd3
TK0135	Indolepyruvate-Fd oxidoreductase, β subunit	21.9	IOR; Breakdown of aromatic amino acids.	$\alpha_2\beta_2$, 2/2	Mai and Adams, 1994	X		
TK0136	Indolepyruvate-Fd oxidoreductase, α subunit	70.9				X	X	X
TK0650	Rubrerythrin Related Protein	22.5	RBR1; Fe-containing protein involved in oxidative protection; peroxidase function.	Likely homodimer, 1/1	Coulter *et al.*, 1999	X		
TK0826	Rubrerythrin Mn-Catalase	22.5	RBR2; Fe-containing protein involved in oxidative protection; peroxidase function.	Likely homodimer, 1/1	Coulter *et al.*, 1999			
TK1088	Geranylgeranyl reductase	44.2	GGR; Reduces isoprenoid chains producing fully saturated lipids.	Monomer, 1/1	Masuchi *et al.* 1998		X	
TK1123	2-Oxoacid:ferredoxin oxidoreductase, γ subunit	20.3	OGOR; Breakdown of glutamate into acyl-CoA and CO_2. Oxidative decarboxylation of deaminated glutamate.	Possible heterooctamer: $\alpha_2\beta_2\delta_2\gamma_2$, 3/7[c]	Mai and Adams, 1996		X	X
TK1129	2-Oxoacid:ferredoxin oxidoreductases, β subunit	31.5					X	
TK1130	2-Oxoacid:ferredoxin oxidoreductase, α subunit	44.0					X	X
TK1215	Membrane-bound sulfane reductase, NiFe-hydrogenase large subunit 2, Mbs-L	45.3	MBS; Membrane-bound respiratory complex generates a proton/Na^+ gradient and H_2S in the presence of S˚.	13-subunit membrane-bound complex, 1/13	Yu *et al.,* 2020			
TK1218	Membrane-bound sulfane reductase, Mbs-M	32.6						
TK1222	Membrane-bound sulfane reductase, Mbs-F	26.1						
TK1325	Ferredoxin:NADH oxidoreductase, α subunit	53.3	FNOR1; Oxidize Fd to generate reduced NAD(P)H; more abundant in S˚ conditions.	$\alpha\beta$, 2/2	Ma and Adams, 1994;	X	X	
TK1326	Ferredoxin:NADH oxidoreductase, β subunit	32.5				X	X	
TK1684	Ferredoxin:NADP oxidoreductase, α subunit	52.7	FNOR2; Oxidize Fd to generate reduced NAD(P)H; more abundant in -S˚ conditions.	$\alpha\beta$, 2/2	Ma and Adams, 1994;	X	X	
TK1685	Ferredoxin:NADP oxidoreductase, β subunit	32.0				X	X	
TK1978	2-Oxoisovalarate-ferredoxin oxidoreductase, γ subunit	19.7	VOR & POR	$\alpha\beta\delta\gamma$, 4/4	Heider *et al.*, 1996	X		X
TK1979	2-Oxoisovalarate-ferredoxin oxidoreductase, δ subunit	11.8	VOR; Breakdown of branched chain amino acids into acyl-CoA and CO_2.			X		X
TK1980	2-Oxoisovalarate-ferredoxin oxidoreductase, α subunit	44.4				X		

TK1981	2-Oxoisovalarate-ferredoxin oxidoreductase, β subunit	32.6				X		X
TK1982	Pyruvate:ferredoxin oxidoreductase, δ subunit	12.0	POR; Breakdown of pyruvate into acetyl-CoA and CO_2.	$\alpha\beta\delta\gamma$, 4/4	Blamey and Adams, 1993; Schut *et al.*, 2001	X		X
TK1983	Pyruvate:ferredoxin oxidoreductase, α subunit	43.9				X	X	X
TK1984	Pyruvate:ferredoxin oxidoreductase, β subunit	36.1				X	X	X
TK2075	4Fe-4S cluster-binding protein associated with FDH	18.3	FDH; Reversable conversion of formate into CO_2.	Unknown, 4/~6	Yang *et al.*, 2022		X	X
TK2076	Formate dehydrogenase, α subunit	80.1					X	X
TK2077	Formate dehydrogenase, 4Fe-4S cluster-binding protein	18.7					X	X
TK2078	4Fe-4S cluster-binding protein associated with FDH	17.8					X	
TK2088	Membrane-bound hydrogenase, MbhI	13.5	MBH; Membrane-bound respiratory complex that generates a proton/Na^+ gradient and H_2 in the absence of S°.	14-subunit membrane-bound complex, 5/14	Yu *et al.*, 2018			X
TK2089	Membrane-bound hydrogenase, MbhJ	19.1						X
TK2090	Membrane-bound hydrogenase, MbhK	21.4						X
TK2091	Membrane-bound hydrogenase, MbhL	48.2					X	X
TK2093	Membrane-bound hydrogenase, MbhN	18.1						
TK2163	Glyceraldehyde-3-phosphate:ferredoxin oxidoreductase	74.2	GAPOR; Converts glyceraldehyde-3-phosphate to 3-phosphoclycerate in glycolysis.	Monomer, 1/1	Makund and Adams, 1995			

a: AP-MS data was produced for bolded genes. Two Mbs genes could not be genetically modified to encode a tag sequence (Chapter 2.5.4). No tagged-MBH subunits were identified as significant compared to the control (Chapter 3.3.3.)
b: Burkhart *et al*., 2019.
c: While the OGOR functional protein complex is predicted to be formed from eight subunits ($\alpha_2\beta_2\delta_2\gamma_2$), only one δ subunit is encoded, thus a total of seven genes are annotated as OGOR-encoding genes, three of which were tagged in this study.

4.3 Results

AP-MS data identifies *in vivo* protein partnerships of genetically tagged proteins expressed at native levels. Tagged protein samples containing co-purified interacting partners are compared to samples from a near-isogenic control (the parental strain) and proteins which appear significant in the tagged sample are identified (Figure 4.7). To aid in identifying protein-protein interactions that occur with high confidence, a ranking system has been developed to identify proteins which appear as significant with high confidence (rank score of 5) and proteins which pass all statistical thresholds but with less confidence (rank score of 1). The application of a rank score system can allow for a confidence matric to be applied to the data without losing information (protein associations) by imposing more stringent thresholds.

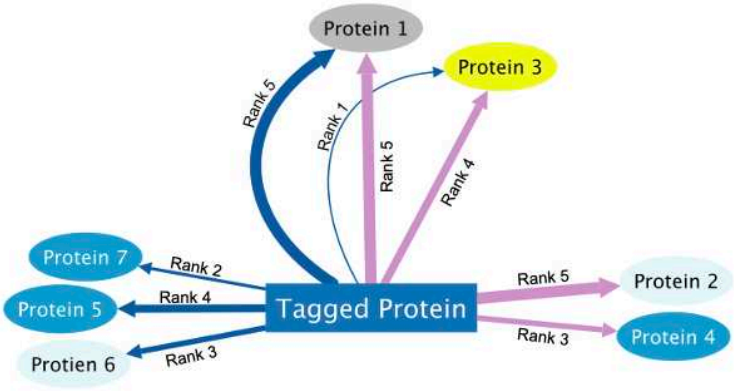
Every tagged-protein was identified as significant in the data of the respective tagged protein. For instance, when GGR was tagged, GGR was identified as significant compared to the control. On average, ~ thirty proteins were identified as interacting partners with each tagged protein, however associations ranged from as few as three (seen when TK1325 was tagged in -S° conditions) and as many as 95 (seen when TK1129 was tagged in +S° conditions) (Figure 4.8). Over 95% of interactions were identified with a rank score of 3 or greater, and almost 80% were identified with a rank score of 4 or greater.

While a global metabolic network is of interest, such an interconnected interactome is difficult to interpret (Figure 4.9). Instead, these data will be broken down and analyzed first by complex, then by reciprocal interactions between all tagged proteins, and finally select, novel associations will be discussed.

A

	Gene No.	P-Value	Fold Change	Spectral Counts: Tagged1	Tagged2	Tagged3	Control1	Control2	Control3	Rank
Sulfur	Protein 1	0.00001	49	379	390	410	12	10	2	5
	Protein 2	0.0045	4.4	125	132	128	27	27	23	5
	Protein 3	0.038	INF	5	5	5	0	0	0	4
	Protein 4	0.046	3.4	5	10	9	6	0	1	3
Non-Sulfur	Protein 1	0.00001	96	411	778	637	3	0	16	5
	Protein 5	0.0045	2.1	24	23	20	13	5	6	4
	Protein 6	0.012	13	0	8	5	1	0	0	3
	Protein 7	0.02	3	4	0	2	1	0	1	2
	Protein 3	0.04	3	0	3	0	1	0	0	1

B



C

Protein No.	Annotation
Protein 1	Rubrerythrin domain-containing protein
Protein 2	V-type ATP synthase subunit I
Protein 3	Ferredoxin 2
Protein 4	ABC-type iron(III) transport system, periplasmic component
Protein 5	Peroxiredoxin, AhpC/TSA family
Protein 6	V-type ATP synthase subunit C
Protein 7	ABC-type multidrug transport system, ATPase component

Figure 4.7: Example AP-MS data from three biological replicates is analyzed and ranked to display interacting partners of the tagged protein. Genetically encoded, tagged proteins are grown in three biological replicates until mid-exponential phase. The tagged protein is purified from the clarified cell lysate using affinity purification. Fractions containing the tagged protein and co-purifying partners are analyzed using proteomic mass spectrometry (LC-MSMS, Chapter 3.2.3). Spectral masses of intact and fractured peptides (hence, tandem mass spec.) are used to sequence peptides which are then mapped to the T. kodakarensis proteome to identify proteins in each sample. Proteins identified in the tagged samples are compared to proteins identified in the control samples (A) using a Fisher's Exact Test (p-value < 0.05) with a Benjamini-Hochberg multiple testing correction (Chapter 3.3). Proteins which are identified with confidence (p-value < 0.05) at a $\log_2$ fold change of ≥ 2 in the tagged sample (compared to the control) are identified as significant proteins in the tagged sample (A). Proteins identified as significant are assigned a rank score representing the confidence of the identified protein (Chapter 3.4). Proteins earn a point if they are identified with a p-value < 0.01, with a $\log_2$ fold change > 4, and for every tagged biological replicate in which they are identified. Here, Protein 1 was identified with a rank score of 5 in both sulfur (+S˚) and non-sulfur (-S˚) conditions. Such a confident association is depicted in (B) by connecting the tagged protein (dark blue rectangle, center) to the light grey ellipse representing Protein 1 by both a thick blue line (representing the

association observed in -S˚) and a thick lilac line (representing the association observed in +S˚). Line thickness decreases with a decreased rank score (B). Protein 3 was identified in +S˚ conditions with a rank score of 4; Protein 3 was observed in all three biological replicates at a fold change of 7 in the tagged strain compared to control, but only with a p-value of 0.039 which still passes the < 0.05 cut-off but is greater than 0.01 (A). When Protein 3 was identified in -S˚ conditions, it was only observed in one biological replicate, at a fold change of 3, and with a p-value of 0.04, thus while the identification of this protein passes all statistical thresholds, it is identified with less confidence in -S˚ conditions than in +S˚ conditions which is demonstrated by the rank score and line thickness (B). Proteins which are identified as significant are observed in colored ellipses radiating from the tagged protein (colored rectangle). Proteins identified as significant are colored bright blue (default color). Select proteins which are of interest have been colored to allow for easy and consistent reference throughout the analysis (B). Here, Protein 2 and Protein 6 are annotated as V-type ATP synthase subunits I and C, respectively (gene numbers and annotations are listed in (C)), thus the color of the ellipse has been changed from bright blue to light blue to indicate that these proteins are of interest and they have the same light blue color to indicate that they are proteins from the same complex (the V-type ATP synthase). For simplicity, the protein number is displayed in the ellipse, but in later figures the T. kodakarensis gene number (e.g., TK0570) is displayed in the ellipse and can be referenced to the protein annotation (C) provided in tables following each network diagram. Proteins have been annotated using the most recent Uniprot database available for T. kodakarensis, although the Kyoto Encyclopedia of Genes and Genomes (KEGG, genome.jp/kegg) and the Archaeal Genome Browser (archaea.ucsc.edu) maintain databases which use these gene numbers and are also recommended resources.

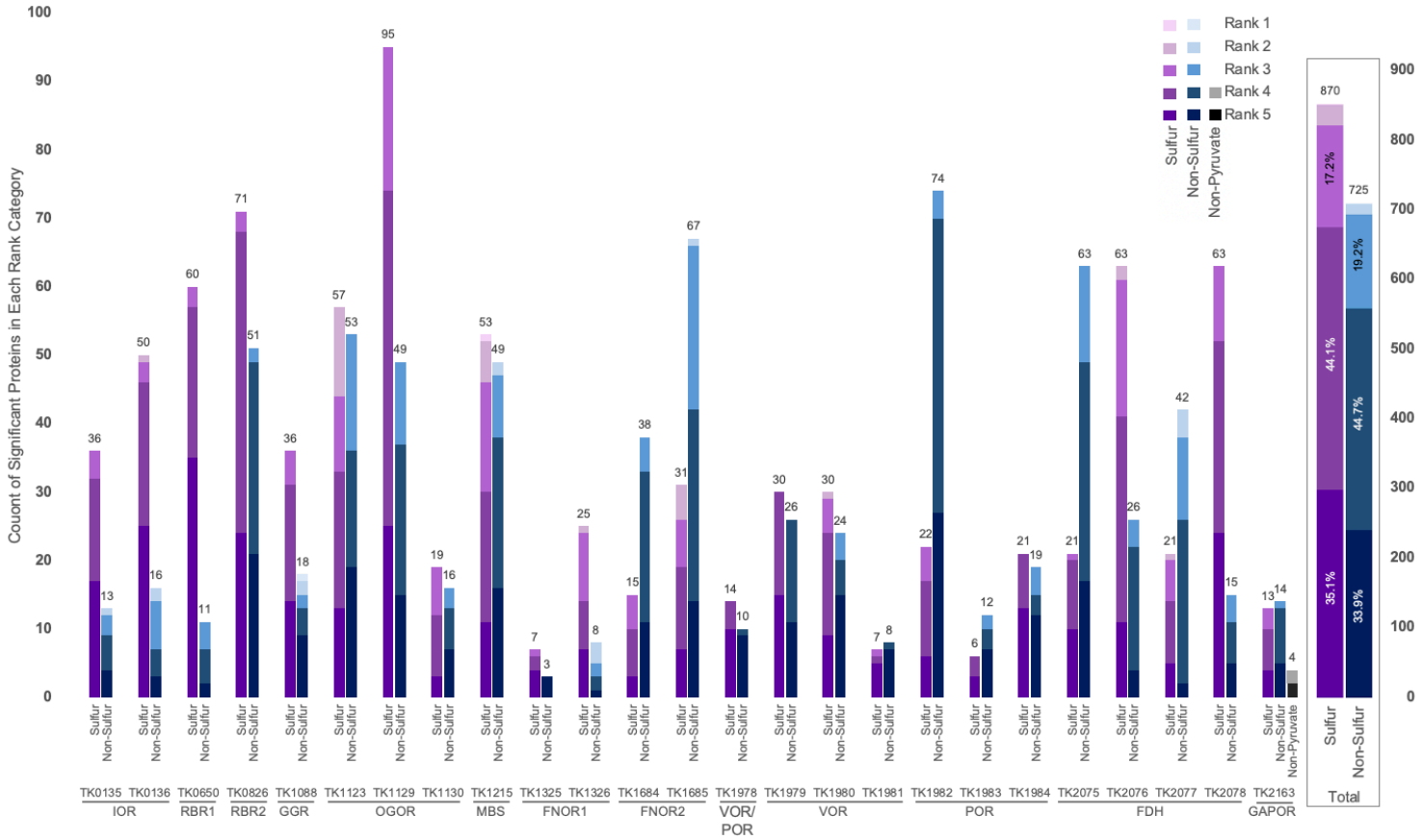


Figure 4.8: Number of proteins identified from each tagged protein in sulfur and non-sulfur conditions. Tagged proteins were observed to interact with a varying number of protein partnerships in sulfur (purple bars) and non-sulfur (blue bars) conditions. Significant proteins are tallied for each tagged-protein which are identified by gene number (e.g., TK0135) and acronym below (refer to Table 4.1 for tagged protein descriptions). Each tagged protein was found to associate with an average of ~ thirty proteins ranging from 6 - 95 in sulfur conditions and 3 - 74 in non-sulfur conditions. Total proteins identified per tagged protein in each respective condition (sulfur or non-sulfur) are noted at the top of each bar, and rank score is indicated by the darkness of the color where dark purple or dark blue indicates a rank score of 5, with lighter colors indicating lower rank score values. A total of 870 and 725 proteins were identified as significant in sulfur and non-sulfur conditions, respectively (bars at far right with modified y-axis scale). The majority of proteins (~ 44%) were identified with a rank score of 4, while ~34% were identified with a rank score of 5 followed by ~18% identified with a rank score of 3. Associations observed for GAPOR in sulfur conditions without pyruvate ("Non-Pyruvate") are depicted using a grey scale.

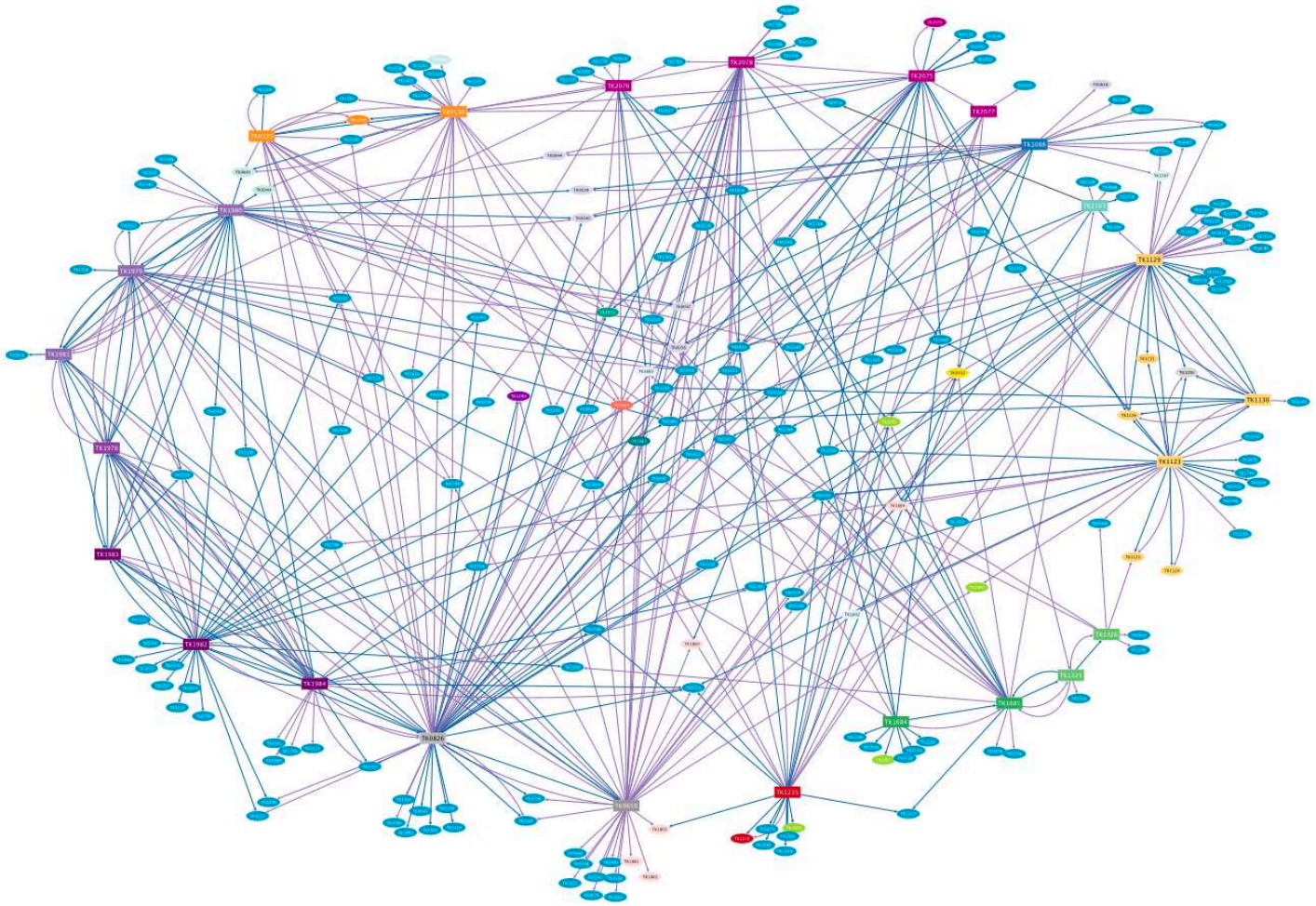


Figure 4.9: Global metabolic network of rank 5 proteins in sulfur and non-sulfur conditions. Tagged proteins in colored rectangles identified proteins in colored ellipses in sulfur (purple lines) or non-sulfur (blue lines) conditions. All proteins identified here were observed with a rank score of 5. Tagged proteins that exist as subunits of known complexes are colored to match and arranged on the periphery: POR is dark purple, VOR is light purple, OGOR is orange, FDH is fuchsia, and FNORs are green. Tagged proteins which function as monomers or homodimers are displayed as colored rectangles and again placed on the periphery: GAPOR is sea foam green, GGR is dark blue, MBS is red, and the RBRs are grey. Proteins which were identified by only a single tagged-protein are placed radiating outwards from the tagged protein, while proteins that were identified by multiple proteins are placed in the interior.

The amino acid fermentation complexes (IOR, OGOR, and VOR)[37–41] will be discussed, followed by proteins POR[42,43] and the rubrerythrins (RBR1 and RBR2).[3,44] The only protein involved in glycolysis, monomeric GAPOR,[18,45,46] will be discussed, followed by FDH and three 4Fe-4S cluster binding proteins (FDH$\gamma_{2\text{-}4}$) which are involved in the conversion of formate to CO_2.[47–49] Then we will turn our attention to complexes involved in anabolic reactions and focus on the generation of reduced small molecules by FNOR1 and FNOR2[50,51] and the maturation of lipids catalyzed by GGR.[19,52] Finally, we will conclude the discussion with the membrane-bound respiratory systems. While no tagged-subunits of MBH generated significant AP-MS data and only one subunit of MBS could be genetically modified, multiple subunits of both MBH and MBS appeared in the data and are worthy of discussion as the balance of electron transfer to these complexes is the realization of the regulated electron flux.[53,54]

The known function of each protein or protein complex will be described as well as an overview of the number of proteins identified by each subunit (if relevant) in each condition (+S˚ or -S˚). A network map depicting protein associations from each tagged protein will show shared and independently identified proteins with the line thickness representing the rank score for each association. Tables which list significant co-purifying partners for each tagged-protein, the p-value, fold change, and rank score in both sulfur and non-sulfur conditions will follow the network map.

Overview of amino acid fermentation

Hyperthermophilic archaea commonly utilize amino acid fermentation as a main energy source.[55,56] In the catabolic breakdown of amino acids, amino transferases (AT) remove the amino group generating 2-oxoacids (or 2-ketoacids) (Figure 4.10).[38–40] Three multimeric oxidoreductases are responsible for the further conversion of 2-oxoacids into acyl-CoA in *T. kodakarensis*: IOR, OGOR, and VOR, each exhibiting a preference for certain classes of amino acids, depending on the composition of the R-group.[41] All amino acid catabolizing complexes

were found to interact with Fd1, however while OGOR and IOR (and not VOR) were found with Fd2, only VOR (and not IOR or OGOR) were found with Fd3.[2]

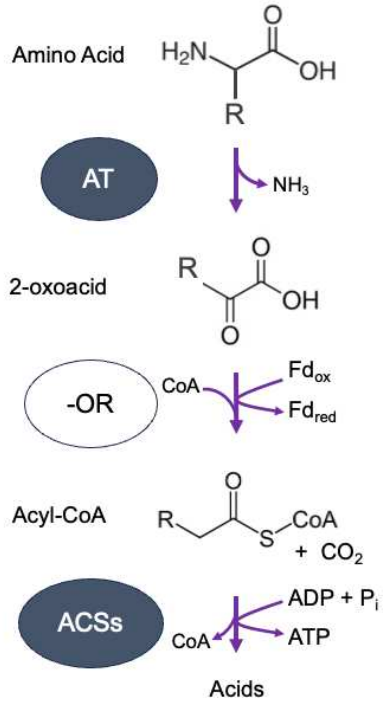


Figure 4.10: Classic amino acid fermentation pathway. Amino acids undergo deamination by amino transferase enzymes (AT) to generate 2-oxoacids. Depending on the R group, various oxidoreductase enzymes (-OR) catalyze the conversion of a 2-oxoacid into an acyl-CoA, thereby reducing a Fd (Fd_{red}). VOR preferentially catabolizes short chain 2-oxoacids, while IOR acts on 2-oxoacids from aryl amino acids, and OGOR displays the greatest substrate preference by specifically catabolizing α-ketoglutarate from deaminated glutamate.

Indolepyruvate oxidoreductase (IOR)

IOR catabolizes the oxidative decarboxylation of aryl pyruvates following transamination of aromatic amino acids (Figure 4.11A).[37] Previous studies suggest that *Pf*-IOR functions as a tetramer ($\alpha_2\beta_2$),[40] and *Tk*-IORα is 70.9 kDa while β is 21.9 kDa. Both IOR subunits identified more than double the number of associations in +S˚ (~42) than in -S˚ (~14) (Figure 4.11B).

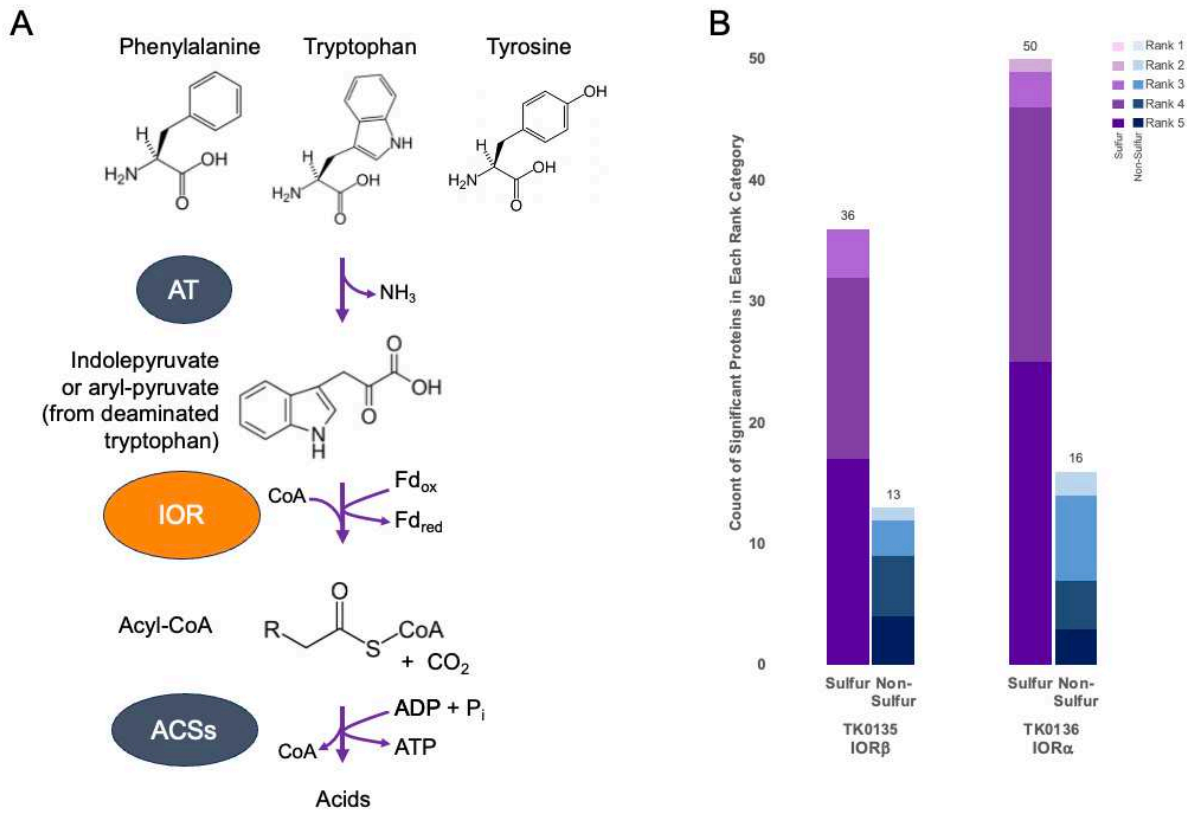


Figure 4.11: IOR function and number of interacting partners observed in sulfur and non-sulfur conditions. Amino acids which contain cyclic R-groups are deaminated by an amino transferase (AT) to produce aryl pyruvates (here, an indole pyruvate from a deaminated tryptophan is shown) (A). IOR (orange) preferentially catalyzes the oxidation of aryl-pyruvates into acyl-CoA while generating a reduced Fd. The acyl-CoA is further oxidized with the addition of CO_2 to generate ATP and resulting acids by acyl-CoA synthase (ACS). In (B), the number of interacting partners for each tagged subunit of IOR is shown by rank score. Overall, more proteins were co-purified with both IOR subunits in sulfur conditions (thirty-six interacting partners for TK0135 and 50 for TK1036) than in non-sulfur conditions (thirteen for TK0135 and sixteen for TK0136). The number of interactions that were observed with a rank score of 5 is shown in dark purple or dark blue, rank 4 is shown in purple or blue, rank 3 is light purple or light blue, and rank 2 is faint blue. No interactions with a rank score of 1 were observed in either analysis for the tagged-IOR proteins.

IORα and IORβ subunits identified one another in both +S˚ and -S˚ conditions (Figure 4.12). Interestingly, there is a second, non-operonic IORβ (TK2244) encoded which produces a 21.7 kDa protein. TK2244 was found to associate with both TK0136 and TK0135 in +S˚ and -S˚, suggesting that the complex may consist of heteroβ subunits. Conversely, a second, non-operonic IORα, TK1643, is also encoded in the genome, however this protein was not identified as significant in the data.

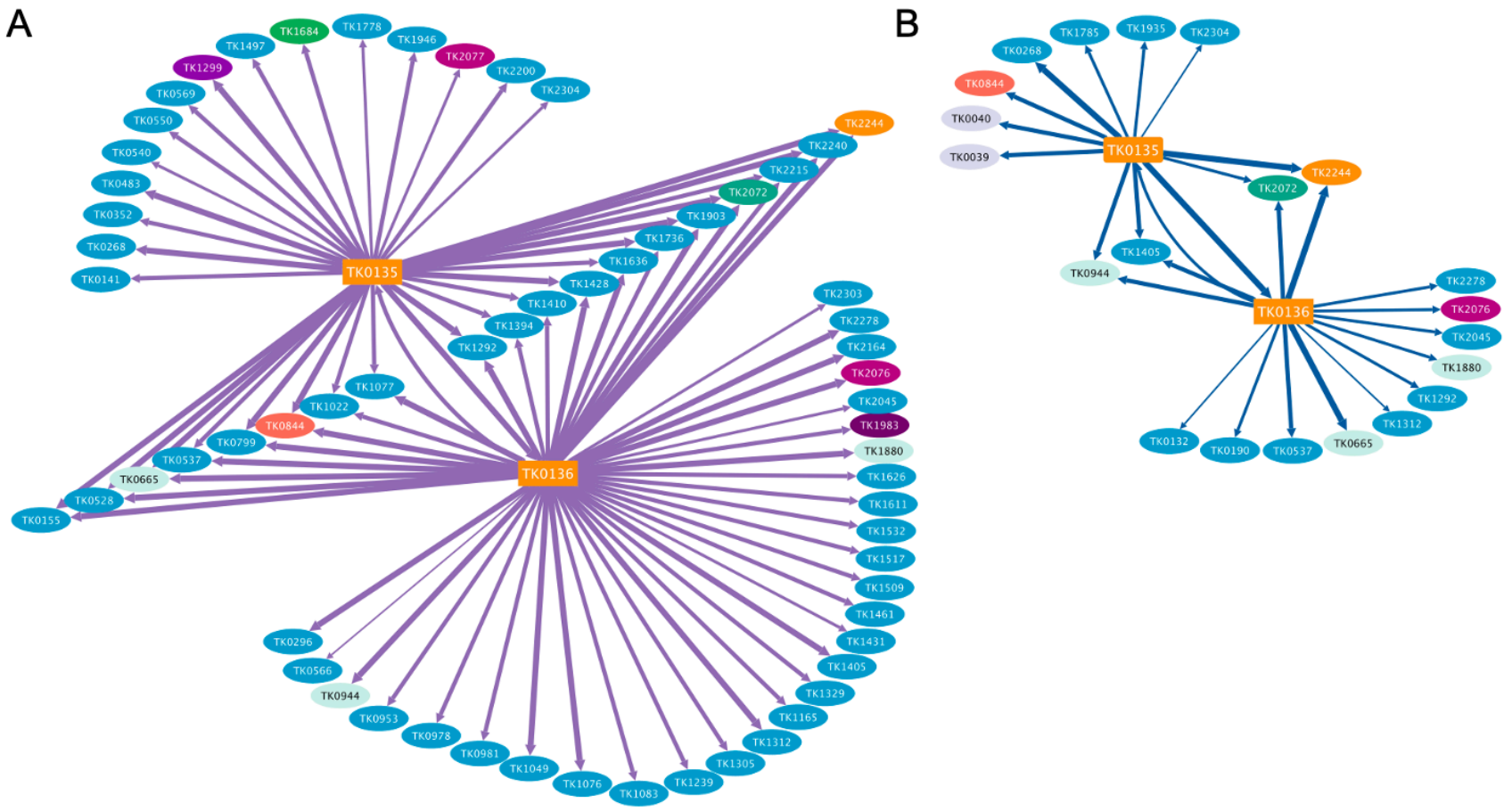


Figure 4.12: Interactome of IOR in sulfur and non-sulfur conditions. AP-MS data for tagged-IORα (TK0136) and IORβ (TK0135) (orange rectangles) were seen to interact with > ~2X proteins in sulfur (A) compared to non-sulfur (B) conditions. Significant proteins are shown in colored ellipses, identified by gene number and generally arranged from lower gene number on the right to higher gene numbers moving left. Line thickness refers to the rank score observed in each association with the thickest line correlating to a rank sore of 5 and thinner lines identifying lower rank scores. Both tagged-IOR subunits identified one another in both conditions. Additionally, both subunits identified a second IORβ subunit (TK2244) (orange ellipse). Both subunits identified the β subunit of a cytosolic hydrogenase (Hydβ, TK2072, teal-green ellipse). Three acetate Co-A ligases were identified: TK0665, TK0944, and TK1880 (light aqua ellipse). A Co-A dependent NAD(P)H sulfur oxidoreductase (TK1299, bright purple ellipse) was identified by IORβ

in +S˚. Two subunits of FDH (TK2076 and TK2077, fuchsia ellipses) were identified by each IOR in +S˚ conditions while only TK2076 was identified with IORα in -S˚. FNOR2α (bright green ellipse) was identified by IORβ and PORα (purple ellipse) was identified by IORα in +S˚ conditions. Two flagellin proteins (TK0039, FlaB2 and TK0040, FlaB3; grey ellipses) were identified in association with IORβ in -S˚.

Table 4.2: Co-purified interacting partners of IORβ (TK0135) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Indolepyruvate oxidoreductase subunit IorB	TK0135	0.00001	72	5	0.00001	29	5
Indolepyruvate oxidoreductase subunit IorA	TK0136	0.00001	190	5	0.00001	96	5
2-dehydro-3-deoxyphosphoheptonate aldolase	TK0268	0.00001	INF	5	0.00001	INF	5
Aldehyde ferredoxin oxidoreductase	TK0844	0.0048	31	5	0.049	5	4
Cytosolic NiFe-hydrogenase, beta subunit	TK2072	0.0061	INF	5	0.012	13	3
Indolepyruvate: ferredoxin oxidoreductase, beta subunit	TK2244	0.00001	INF	5	0.00001	INF	5
tRNA/rRNA cytosine-C5-methylase, NOL1/NOP2/Sun family, fused to N-terminal NusB regulator domain	TK2304	0.00048	INF	3	0.035	5.3	2
UPF0173 metal-dependent hydrolase TK0141	TK0141	0.026	9.1	4			
RecJ-like exonuclease, containing OB-fold nucleic acid-binding domains	TK0155	0.0024	21	5			
AMP phosphorylase	TK0352	0.00001	2.2	4			
Uncharacterized protein	TK0483	0.00065	40	5			
Serine hydroxymethyltransferase	TK0528	0.00087	39	5			
Peroxiredoxin	TK0537	0.022	INF	4			
Metal-dependent phosphohydrolase, HD superfamily	TK0540	0.041	INF	3			
Proline--tRNA ligase	TK0550	0.033	21	4			
Uncharacterized protein	TK0569	0.032	9.7	4			
Acetate--CoA ligase (ADP-forming)	TK0665	0.00066	INF	5			
Type 2 DNA topoisomerase 6 subunit B	TK0799	0.0015	12	5			
D-aminopeptidase	TK1022	0.016	INF	4			
30S ribosomal protein S7	TK1077	0.023	15	4			
Phosphoenolpyruvate synthase	TK1292	0.0012	INF	5			
NAD(P)H sulfur oxidoreductase (CoA-dependent)	TK1299	0.009	12	5			
Uncharacterized protein	TK1394	0.03	INF	4			
DNA primase DnaG	TK1410	0.0045	2.1	4			
Cleavage and polyadenylation specificity factor subunit homolog	TK1428	0.00089	18	5			
Phosphopyruvate hydratase	TK1497	0.038	13	4			
Ribosome maturation protein SDO1 homolog	TK1636	0.03	14	4			
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1684	0.041	INF	4			

Vitamin B12-dependent ribonucleotide reductase	TK1736	0.0061	6	5			
Predicted hydrolase, metallo-beta-lactamase superfamily	TK1778	0.03	INF	3			
DNA polymerase II large subunit	TK1903	0.0066	8.8	5			
Translation initiation factor 2 subunit gamma	TK1946	0.049	6.3	4			
4Fe-4S cluster-binding protein	TK2077	0.022	2	3			
cobW domain-containing protein	TK2200	0.028	6.1	4			
tRNA-splicing endonuclease	TK2215	0.03	14	4			
Lysine--tRNA ligase	TK2240	0.0029	13	5			
Flagellin B2	TK0039				0.0035	16	4
Flagellin B3	TK0040				0.022	INF	4
Acetate--CoA ligase (ADP-forming)	TK0944				0.014	INF	4
Phosphoenolpyruvate carboxykinase [GTP]	TK1405				0.028	7	4
Probable tRNA/rRNA methyltransferase	TK1785				0.0054	INF	3
tRNA/rRNA cytosine-C5-methylase, NOL1/NOP2/Sun family	TK1935				0.0033	INF	3

Table 4.3: Co-purified interacting partners of IORα (TK1036) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Indolepyruvate oxidoreductase subunit IorB	TK0135	0.013	8.7	4	0.028	3.2	3
Indolepyruvate oxidoreductase subunit IorA	TK0136	0.00001	140	5	0.00001	66	5
Peroxiredoxin	TK0537	0.00001	INF	5	0.00071	6.2	4
Acetate--CoA ligase (ADP-forming)	TK0665	0.00001	INF	5	0.00034	9	5
Acetate--CoA ligase (ADP-forming)	TK0944	0.00001	INF	5	0.00001	INF	4
Phosphoenolpyruvate synthase	TK1292	0.0032	INF	5	0.048	4.7	3
V4R domain-containing protein	TK1312	0.0023	INF	5	0.048	INF	2
Phosphoenolpyruvate carboxykinase [GTP]	TK1405	0.00001	INF	5	0.00046	11	4
Acetate--CoA ligase (ADP-forming)	TK1880	0.00001	INF	5	0.0023	INF	3
SAM-dependent methyltransferase, UPF0020 family	TK2045	0.00001	INF	3	0.0023	INF	3
Cytosolic NiFe-hydrogenase, beta subunit	TK2072	0.0032	INF	5	0.0022	16	4
Probable formate dehydrogenase, alpha subunit	TK2076	0.0045	4.4	5	0.021	1.3	3
Indolepyruvate: ferredoxin oxidoreductase, beta subunit	TK2244	0.00001	INF	5	0.00001	INF	5
Myo-inositol-1-phosphate synthase	TK2278	0.00001	INF	5	0.00001	INF	3
RecJ-like exonuclease, containing OB-fold nucleic acid-binding domains	TK0155	0.0019	20	5			
Quinolinate synthase A	TK0296	0.0089	INF	5			
Serine hydroxymethyltransferase	TK0528	0.0011	21	5			
DEAD/DEAH box RNA helicase	TK0566	0.012	INF	2			

Type 2 DNA topoisomerase 6 subunit B	TK0799	0.00047	11	5			
Aldehyde ferredoxin oxidoreductase	TK0844	0.0043	29	5			
Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains	TK0953	0.012	INF	4			
Glycine--tRNA ligase	TK0978	0.017	INF	4			
N2, N2-dimethylguanosine tRNA methyltransferase	TK0981	0.039	7	4			
D-aminopeptidase	TK1022	0.034	INF	4			
Methionine--tRNA ligase	TK1049	0.0089	INF	5			
DNA-directed RNA polymerase subunit	TK1076	0.00026	25	5			
30S ribosomal protein S7	TK1077	0.0057	17	5			
DNA-directed RNA polymerase subunit beta	TK1083	0.035	6.8	4			
Predicted AP endonuclease	TK1165	0.029	5.9	4			
Peptide chain release factor subunit 1	TK1239	0.017	INF	4			
Probable translation initiation factor IF-2	TK1305	0.019	23	4			
Uncharacterized protein	TK1329	0.034	20	4			
Uncharacterized protein	TK1394	0.034	INF	4			
DNA primase DnaG	TK1410	0.0068	2	4			
Cleavage and polyadenylation specficity factor subunit homolog	TK1428	0.00081	17	5			
Glutamate dehydrogenase	TK1431	0.024	2.3	3			
Leucine--tRNA ligase	TK1461	0.024	INF	4			
Probable tRNA pseudouridine synthase B	TK1509	0.022	14	4			
Adenylate kinase	TK1517	0.034	INF	4			
30S ribosomal protein S17	TK1532	0.036	10	4			
Predicted metal-dependent hydrolase	TK1611	0.034	20	4			
Uncharacterized protein	TK1626	0.034	9	4			
Ribosome maturation protein SDO1 homolog	TK1636	0.0019	20	5			
Vitamin B12-dependent ribonucleotide reductase	TK1736	0.015	5.2	4			
DNA polymerase II large subunit	TK1903	0.00001	16	5			
Pyruvate:ferredoxin oxidoreductase, alpha subunit	TK1983	0.0034	2	4			
Fructose-1,6-bisphosphate aldolase/phosphatase	TK2164	0.0023	INF	5			
tRNA-splicing endonuclease	TK2215	0.036	13	4			
Lysine--tRNA ligase	TK2240	0.0088	11	5			
Thermosome subunit beta	TK2303	0.032	2	3			
TATA-box-binding protein	TK0132				0.048	INF	2
GMP synthase [glutamine-hydrolyzing] subunit A	TK0190				0.046	3.3	3

One protein was found to be significant in both conditions and by both subunits: TK2072 encodes the β subunit of a cytosolic hydrogenase (Hydβ). Two acetate-CoA ligases, TK0665 and TK0944, were identified in both +S˚ and -S˚ conditions, however TK0665 was identified by both IOR subunits in +S˚ and TK0944 was identified by both IOR subunits in -S˚, while only IORα identified TK0665 in -S˚ and TK0944 in +S˚. An aldehyde ferredoxin oxidoreductase (AOR), TK0844, was found to associate with both subunits in +S˚ and IORβ in -S˚. AOR uses tungsten in the enzymatic oxidation of aldehydes, thereby reducing a Fd in the process.[57]

2-oxoacid ferredoxin oxidoreductase (OGOR)

Of the known 2-ketoacid catabolizing complexes, 2-oxoacid-ferredoxin oxidoreductase (or 2-ketoglutarate-ferredoxin oxidoreductase, OGOR) displays the greatest substrate specificity by preferentially acting on deaminated glutamate (Figure 4.13A).[38,40,56] The composition of OGOR is not confirmed but seven OGOR subunits (TK1123 – TK1126 and TK1129 – TK1131) encode two α (TK1125 and TK1130), two β (TK1124 and TK1129), two γ (TK1123 and TK1126), and one δ (TK1131) subunit. OGOR has been reported to function as a heterotetramer[40] and a heterooctamer.[56]

Three distinct subunits of OGOR were tagged in this study: OGORγ_1 (TK1123), OGORβ_2 (TK1129), and OGORα_2 (TK1130). Like IOR, OGOR displayed more interactions in +S˚ (average 57) than -S˚ (average 39) (Figure 4.13B). OGORα_2 displayed considerably fewer interactions (~17) than the other two subunits. All seven OGOR subunits were identified by the three tagged OGORs (Figure 4.14). Interestingly, while OGORγ_1 was identified will all subunits, OGORα_2 and OGORβ_2 were only identified with OGOR$_2$ subunits, thus there may be a slight preference for this multimer to be composed of homodimeric subunits in some instances.

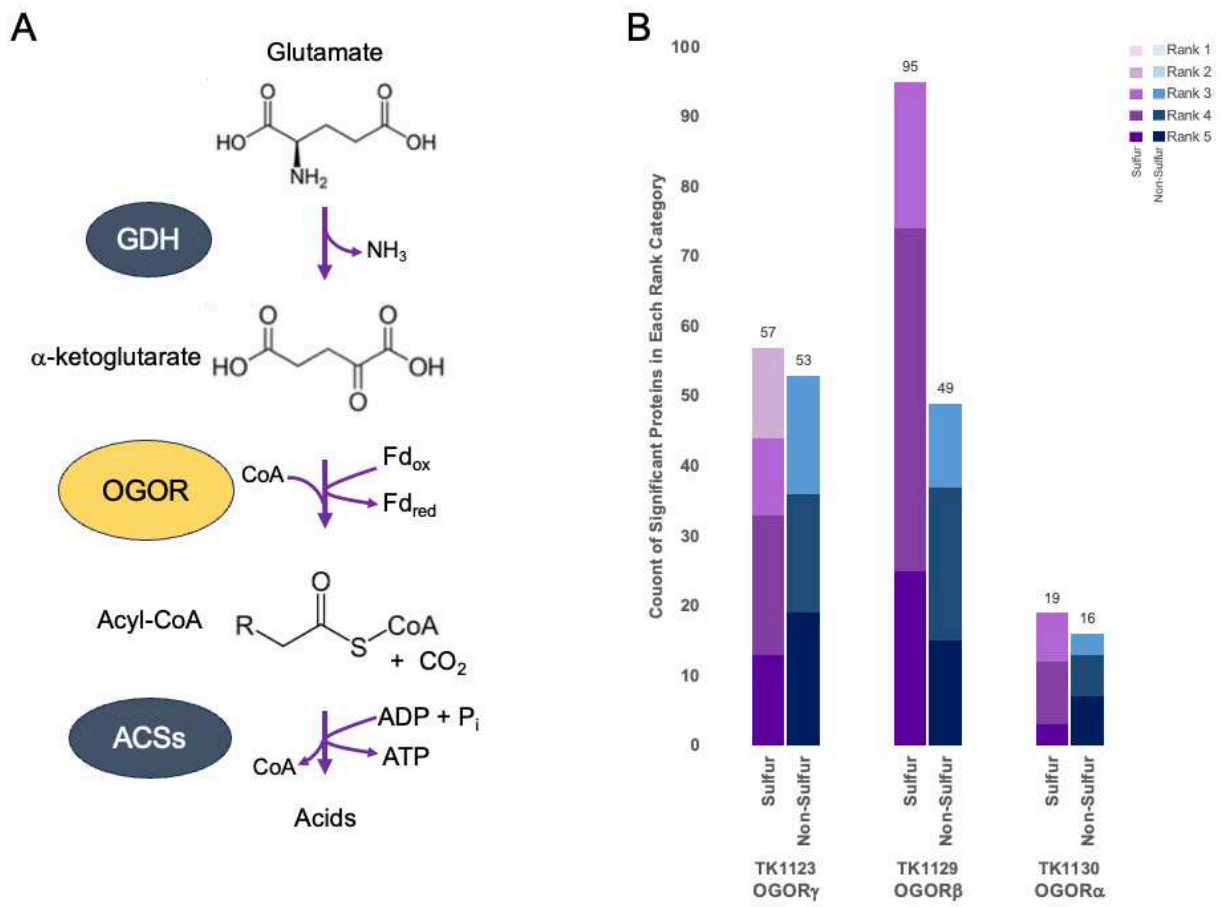


Figure 4.13: OGOR function and number of interacting partners observed in sulfur and non-sulfur conditions. Glutamate is deaminated by glutamate dehydrogenase (GDH) (A). OGOR (light yellow) preferentially catalyzes the oxidation of deaminated glutamate into acyl-CoA while generating a reduced Fd. The acyl-CoA is further oxidized with the addition of CO_2 to generate ATP and resulting acids by acyl-CoA synthase (ACS). In (B), the number of interacting partners for each tagged subunit of OGOR is shown by rank score. Similar numbers of proteins were identified in sulfur and non-sulfur conditions for TK1123 and TK1130, however almost double the number of interactions were observed for TK1129 in sulfur conditions compared to non-sulfur conditions. The number of interactions that were observed with a rank score of 5 is shown in dark purple or dark blue, rank 4 is shown in purple or blue, rank 3 is light purple or light blue, and rank 2 is faint blue. Only TK1123 in +S˚ identified interactions with a rank score of 2. No interactions with a rank score of 1 were observed in either analysis for the tagged-OGOR proteins.

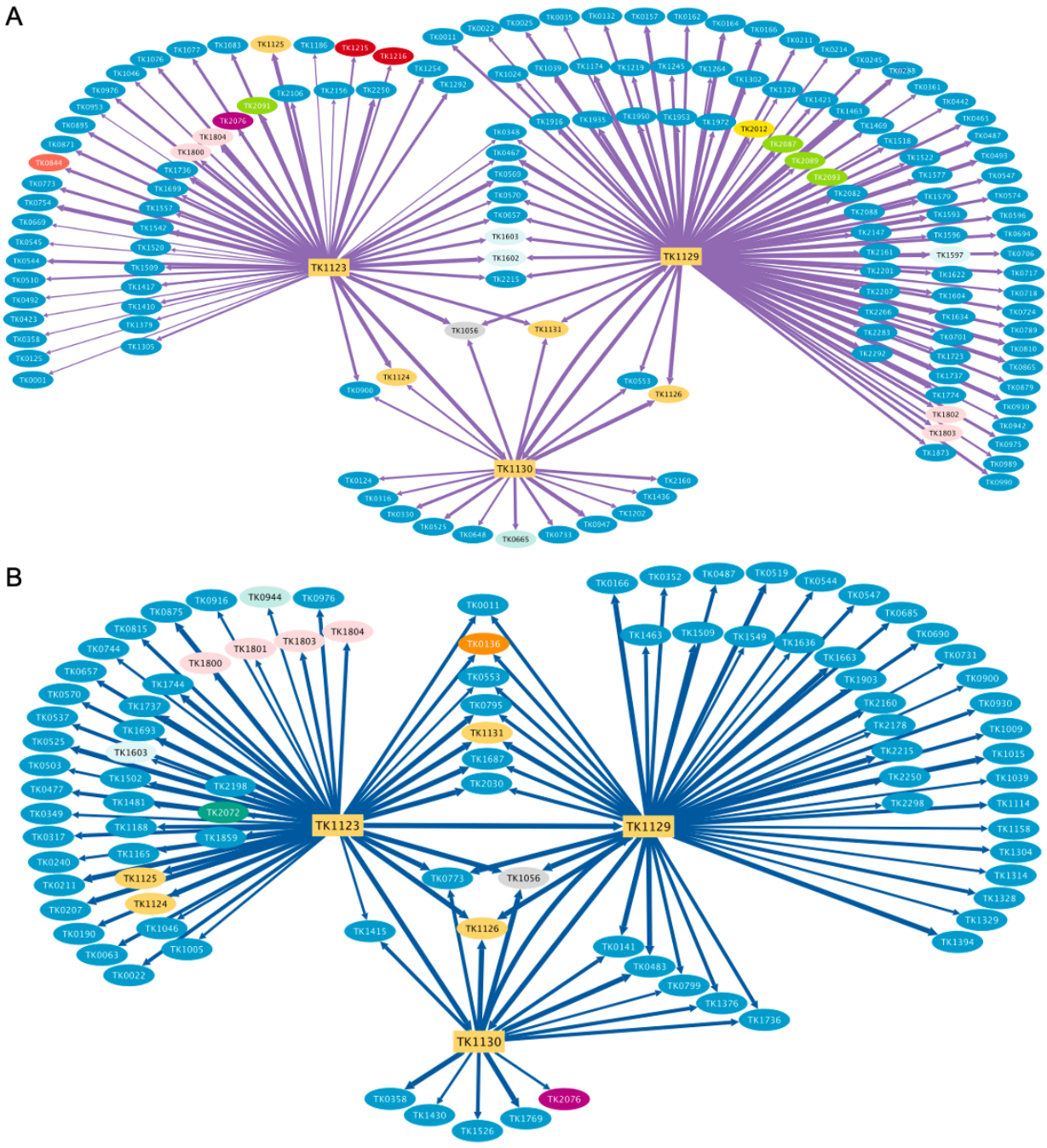


Figure 4.14: Interactome of OGOR in sulfur and non-sulfur conditions. AP-MS data from three tagged-OGOR subunits (gold rectangles) in +S˚ (A) and -S˚ (B) conditions identifies significant proteins as co-purifying partners of the respective tagged-OGOR subunit. All seven OGOR subunits are identified in each interactome (gold ellipses), though not by all tagged subunits. OGORγ_1 (TK1123) was found with OGORβ_1 (TK1124), both subunits of OGORα (TK1125 and TK1130), and OGOR δ (TK1131) in both conditions and OGORγ_2 (TK1126) and OGORβ_2 (TK1129) in -S˚. Tagged-OGORβ_2 (TK1129) copurified OGORγ_2 (TK1126), OGORα_2 (TK1130), and OGORδ (TK1131), but not OGORγ_1, OGORβ_1, or OGORα_1 (TK1123 – TK1125) in either condition. Tagged-OGORα_2 (TK1130) was found with OGORγ_2 (TK1126), OGORβ_2 (TK1129), and OGORδ (TK1131) in +S˚, and almost the same subunits in -S˚ with the exception

of OGORγ_2 (TK1126). Tagged-OGORα_2 (TK1130) was not observed to interact with OGORα_1, OGORβ_1, or OGORγ_1 (TK1123 – TK1125, respectively) in either condition. All tagged OGORs identified RBR3 (TK1056) as an interacting partner in both conditions. OGORγ_1 and OGORβ_2 identified IORβ (TK0136) in -S˚ conditions. OGORβ_2 identified Fd3 (TK2012, yellow ellipse), three subunits of the MBH complex (lime green ellipses, TK2087, TK2089, and TK2093) while OGORγ_1 identified another MBH subunit (lime green ellipse, TK2091) and two MBS subunits (red, TK1215 and TK1216) in +S˚. Four subunits of the ABC-type dipeptide/oligopeptide transport system (light pink, TK1800 – TK1804) were identified in both +S˚ and -S˚. Association with two ATP synthase subunits (light blue, TK1597 and TK1602) were seen in +S˚.

Table 4.4: Co-purified interacting partners of OGORγ (TK1123) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
ABC-type iron(III) transport system, periplasmic component	TK0570	0.025	INF	3	0.017	5	3
ABC-type transport system, probable periplasmic component	TK0657	0.04	24	4	0.0027	3	7
Predicted ATP-dependent endonuclease, OLD family	TK0773	0.021	INF	4	0.00001	6.2	11
Putative snRNP Sm-like protein	TK0976	0.00018	2.1	4	0.00001	2.1	33
Uncharacterized protein	TK1046	0.047	INF	3	0.025	3.7	2
Rubrerythrin-related protein	TK1056	0.004	INF	5	0.00001	24	45
2-oxoacid:ferredoxin oxidoreductases, gamma subunit	TK1123	0.00028	INF	5	0.00001	23	40
2-oxoacid:ferredoxin oxidoreductases, beta subunit	TK1124	0.00051	INF	5	0.00001	50	68
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1125	0.00001	INF	5	0.00001	45	172
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1130	0.0049	INF	5	0.00001	5.7	17
2-oxoacid: ferredoxin oxidoreductase, delta subunit	TK1131	0.038	INF	4	0.00001	INF	20
V-type ATP synthase beta chain	TK1603	0.02	28	4	0.027	2.8	4
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1800	0.024	27	4	0.041	3.3	3
ABC-type dipeptide/oligopeptide transport system, probable periplasmic component	TK1804	0.007	17	5	0.00001	3.5	20
DNA polymerase	TK0001	0.017	INF	2			
N-acetyltransferase, GNAT family	TK0125	0.047	INF	2			
Predicted membrane protease subunit, stomatin/prohibitin homolog	TK0348	0.047	INF	2			
RNA-splicing ligase RtcB	TK0358	0.021	INF	2			
Prolyl endopeptidase	TK0423	0.038	INF	2			
Uncharacterized protein	TK0467	0.047	INF	3			
Aspartate—tRNA(Asp) ligase	TK0492	0.031	INF	2			
Diadenylate cyclase	TK0510	0.038	2	3			
Molybdenum cofactor biosynthesis protein B	TK0544	0.00001	2.6	4			
S-adenosylmethionine synthase	TK0545	0.021	INF	2			
Uncharacterized protein	TK0569	0.0069	34	5			
CDC48/VCP homolog, AAA superfamily	TK0669	0.047	INF	2			
tRNA(Met) cytidine acetyltransferase TmcA	TK0754	0.00001	INF	5			
Aldehyde ferredoxin oxidoreductase	TK0844	0.00034	INF	4			
Ornithine carbamoyltransferase	TK0871	0.025	INF	4			
S-layer protein	TK0895	0.029	7.9	4			
Uncharacterized protein	TK0900	0.038	INF	4			
Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains	TK0953	0.038	INF	2			
DNA-directed RNA polymerase subunit	TK1076	0.0026	INF	4			

30S ribosomal protein S7	TK1077	0.048	23	4			
DNA-directed RNA polymerase subunit beta	TK1083	0.0012	26	4			
Uncharacterized protein	TK1186	0.047	INF	2			
Membrane bound hydrogenase, NiFe-hydrogenase large subunit 2	TK1215	0.00001	INF	3			
Membrane bound hydrogenase, NiFe-hydrogenase large subunit 1	TK1216	0.025	INF	3			
30S ribosomal protein S3Ae	TK1254	0.031	INF	4			
Phosphoenolpyruvate synthase	TK1292	0.0018	INF	4			
Probable translation initiation factor IF-2	TK1305	0.038	INF	3			
Probable glycine dehydrogenase (decarboxylating) subunit 2	TK1379	0.031	INF	2			
DNA primase DnaG	TK1410	0.03	2.8	3			
50S ribosomal protein L1	TK1417	0.031	INF	3			
Probable tRNA pseudouridine synthase B	TK1509	0.034	25	4			
50S ribosomal protein L30	TK1520	0.038	INF	2			
50S ribosomal protein L3	TK1542	0.0012	INF	5			
Predicted dehydrogenase	TK1557	0.019	2.6	3			
V-type ATP synthase alpha chain	TK1602	0.0049	INF	5			
DNA-directed RNA polymerase subunit A''	TK1699	0.017	INF	3			
Vitamin B12-dependent ribonucleotide reductase	TK1736	0.0011	15	5			
Probable formate dehydrogenase, alpha subunit	TK2076	0.00034	2.5	4			
Membrane bound hydrogenase, NiFe-hydrogenase large subunit 2	TK2091	0.0091	INF	5			
Enolase	TK2106	0.017	INF	4			
Archaeosine synthase subunit alpha	TK2156	0.047	INF	2			
tRNA-splicing endonuclease	TK2215	0.042	15	4			
Non-specific serine/threonine protein kinase	TK2250	0.0038	23	5			
Uncharacterized protein	TK0011				0.00001	2	4
Uncharacterized protein	TK0022				0.032	7	4
Nucleotidyltransferase, fused to N-terminal DNA-binding domain	TK0063				0.0029	2.7	4
Indolepyruvate oxidoreductase subunit IorA	TK0136				0.00001	2.4	4
GMP synthase [glutamine-hydrolyzing] subunit A	TK0190				0.0004	2.7	4
Formate-dependent phosphoribosylglycinamide formyltransferase	TK0207				0.00001	21	5
Amidophosphoribosyltransferase	TK0211				0.00001	4.2	5
Arginase	TK0240				0.022	2.7	3
TatD-related deoxyribonuclease	TK0317				0.007	INF	4
ATPase involved in chromosome partitioning, ParA/MinD family	TK0349				0.029	INF	3
Glyceraldehyde 3-phosphate phosphatase	TK0477				0.00001	2.3	4
Cupin_2 domain-containing protein	TK0503				0.032	7	3
Superoxide reductase	TK0525				0.007	INF	5
Peroxiredoxin	TK0537				0.0051	2.3	4

tRNA(Ile2) 2-agmatinylcytidine synthetase TiaS	TK0553				0.014	2	3
Uncharacterized protein	TK0744				0.018	2.6	3
Putative 5-methylcytosine restriction system, GTPase subunit	TK0795				0.001	5.7	5
Peroxiredoxin, AhpC/TSA family	TK0815				0.007	INF	4
Saccharopine reductase	TK0875				0.00001	INF	5
L-threonine 3-dehydrogenase	TK0916				0.026	2	3
Acetate—CoA ligase (ADP-forming)	TK0944				0.032	7	3
Prefoldin subunit alpha 1	TK1005				0.029	INF	3
2-oxoacid:ferredoxin oxidoreductases, gamma subunit	TK1126				0.00001	30	5
2-oxoacid:ferredoxin oxidoreductases, beta subunit	TK1129				0.00001	7	5
Predicted AP endonuclease	TK1165				0.00001	2	4
Glucosamine-1-phosphate N-acetyltransferase	TK1188				0.015	4	4
50S ribosomal protein L12	TK1415				0.014	2.8	3
NADH:polysulfide oxidoreductase	TK1481				0.02	2.2	3
50S ribosomal protein L18e	TK1502				0.032	7	4
Cysteine synthase	TK1687				0.0051	10	5
FeS_assembly_P domain-containing protein	TK1693				0.0034	INF	5
Xanthine/guanine phosphoribosyltransferase	TK1737				0.0055	6	5
Uncharacterized protein	TK1744				0.00001	INF	5
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1801				0.042	2.6	3
ABC-type dipeptide/oligopeptide transport system, permease component	TK1803				0.015	4	4
Probable lipoprotein releasing system, ATP-binding protein	TK1859				0.011	2.2	3
Uncharacterized protein	TK2030				0.00083	INF	5
Cytosolic NiFe-hydrogenase, beta subunit	TK2072				0.032	2.2	3
Uncharacterized protein	TK2198				0.022	2.7	3

Table 4.5: Co-purified interacting partners of OGORβ (TK1129) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Uncharacterized protein	TK0011	0.0033	2.1	4	0.041	4.1	4
Uncharacterized protein	TK0166	0.00001	4.3	5	0.0022	INF	5
Uncharacterized protein	TK0487	0.00001	INF	5	0.00021	INF	4
Metallophosphoesterase, calcineurin superfamily	TK0547	0.00049	2.7	4	0.021	5.7	4
tRNA(Ile2) 2-agmatinylcytidine synthetase TiaS	TK0553	0.00001	2.2	4	0.0034	3.6	4
Uncharacterized protein	TK0930	0.00001	5.5	5	0.028	6.2	4
Cyclic 2,3-diphosphoglycerate synthetase	TK1039	0.00001	2	4	0.018	3.5	3

Rubrerythrin-related protein	TK1056	0.001	2.1	4	0.0021	8.9	5
2-oxoacid:ferredoxin oxidoreductases, gamma subunit	TK1126	0.00001	18	5	0.00001	INF	5
2-oxoacid:ferredoxin oxidoreductases, beta subunit	TK1129	0.00001	7.6	5	0.00001	41	5
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1130	0.00001	5.2	5	0.00001	39	5
2-oxoacid: ferredoxin oxidoreductase, delta subunit	TK1131	0.00001	3.4	4	0.00017	11	5
tRNA (1-methyladenosine) methyltransferase	TK1328	0.00001	INF	3	0.023	13	4
Uncharacterized protein	TK1463	0.00001	8.3	5	0.048	11	4
tRNA-splicing endonuclease	TK2215	0.00001	2.3	4	0.0044	4.2	5
Uncharacterized protein	TK0022	0.021	2.5	3			
Hypothetical membrane protein, conserved	TK0025	0.021	4.5	4			
Hypothetical membrane protein, conserved	TK0035	0.023	INF	3			
TATA-box-binding protein	TK0132	0.037	4	4			
Xanthine/uracilpermease	TK0157	0.0065	9	5			
Hypothetical membrane protein	TK0162	0.023	INF	4			
S-layer-like array protein	TK0164	0.00001	INF	5			
Amidophosphoribosyltransferase	TK0211	0.00037	2.2	4			
Probable formate dehydrogenase, alpha subunit	TK0214	0.049	INF	3			
Imidazoleglycerol-phosphate dehydratase	TK0245	0.023	INF	4			
Phosphonate 202etabolism protein PhnP homolog, metallo-beta-lactamase superfamily	TK0288	0.0098	4	5			
Predicted membrane protease subunit, stomatin/prohibitin homolog	TK0348	0.021	2.1	3			
Protein archease	TK0361	0.011	INF	3			
Uncharacterized protein	TK0442	0.029	3.3	3			
Acetyl-CoA synthetase I (NDP forming), beta subunit	TK0465	0.023	INF	4			
Uncharacterized protein	TK0467	0.00001	6.9	5			
Hypothetical membrane protein, conserved, containing DUF11 domain	TK0493	0.00029	6	5			
Uncharacterized protein	TK0569	0.00001	2.5	4			
ABC-type iron(III) transport system, periplasmic component	TK0570	0.003	2.5	4			
Metallophosphoesterase, calcineurin superfamily	TK0574	0.00001	3.4	4			
Hypothetical membrane protein	TK0596	0.049	INF	4			
ABC-type transport system, probable periplasmic component	TK0657	0.00001	3.2	4			
ABC-type multidrug transport system, ATPase component	TK0694	0.049	INF	4			
ATPase, ParA/MinD family, containing ferredoxin domains	TK0701	0.049	INF	3			
ABC-type iron(III)-siderophore transport system, periplasmic component fused to N-terminal uncharacterized domain	TK0706	0.00001	3.8	4			
ABC-type molybdate transport system, periplasmic component	TK0717	0.011	INF	4			
ABC-type molybdate transport system, permease component	TK0718	0.022	3	3			
Iron-molybdenum cofactor-binding protein	TK0724	0.045	6	4			

Glycerol-1-phosphate dehydrogenase [NAD(P)+]	TK0789	0.00062	2.7	4
Oligosaccharyl transferase	TK0810	0.00043	13	5
Probable vitamin B12 transport protein	TK0865	0.00098	7	5
Uncharacterized protein	TK0879	0.012	5	4
ABC-type multidrug transport system, ATPase component	TK0942	0.023	INF	3
50S ribosomal protein L37e	TK0975	0.049	INF	4
Fructose-bisphosphate aldolase class 1	TK0989	0.018	2	3
Oxaloacetate decarboxylase, alpha subunit	TK0990	0.012	5	4
Hypothetical membrane protein	TK1024	0.049	INF	4
Predicted acetyltransferase, isoleucine patch superfamily	TK1174	0.0063	5.5	5
Membrane bound hydrogenase, MbxH' subunit	TK1219	0.012	5	4
Uncharacterized protein	TK1245	0.0083	2.8	4
Archaeal Lon protease	TK1264	0.049	INF	3
Hypothetical membrane protein, conserved	TK1302	0.023	INF	4
Cell division protein FtsZ 1	TK1421	0.0051	3	4
Ribonuclease J	TK1469	0.00001	28	3
Protein translocase subunit SecY	TK1518	0.0023	3.6	4
50S ribosomal protein L18	TK1522	0.021	2.5	3
Hypothetical membrane protein, conserved	TK1577	0.0017	11	5
ABC-type multidrug transport system, ATPase component	TK1579	0.00001	5.4	5
Protein-export membrane protein SecF	TK1593	0.0033	10	5
Archaeal/vacuolar-type H+-ATPase, subunit H	TK1596	0.013	3.2	3
V-type ATP synthase subunit I	TK1597	0.0018	6.5	5
V-type ATP synthase alpha chain	TK1602	0.00075	3.1	4
V-type ATP synthase beta chain	TK1603	0.0018	2.4	4
V-type ATP synthase subunit D	TK1604	0.021	4.5	4
Methylmalonyl-CoA decarboxylase, alpha subunit	TK1622	0.00001	3.6	4
Exosome complex component Rrp41	TK1634	0.0033	10	5
Glycosyltransferase, family 4	TK1723	0.049	INF	3
Xanthine/guanine phosphoribosyltransferase	TK1737	0.00015	5.2	5
Amylopullulanase, GH57 family	TK1774	0.0063	5.5	5
ABC-type dipeptide/oligopeptide transport system, permease component	TK1802	0.00038	2.6	4
ABC-type dipeptide/oligopeptide transport system, permease component	TK1803	0.00045	3.2	4
Peptidase_M1 domain-containing protein	TK1873	0.049	INF	4
Carbon-nitrogen hydrolase	TK1916	0.0051	INF	3
tRNA/rRNA cytosine-C5-methylase, NOL1/NOP2/Sun family	TK1935	0.00001	6.2	5
Uncharacterized protein	TK1950	0.0017	11	5

Uncharacterized protein	TK1953	0.045	6	4			
Hypothetical membrane protein, conserved, containing TPR-repeat domains	TK1972	0.00043	13	4			
Ferredoxin 3	TK2012	0.00015	2.8	4			
Membrane bound hydrogenase, MbhC subunit	TK2082	0.037	4	4			
Membrane bound hydrogenase, MbhH subunit	TK2087	0.017	3.7	3			
Membrane bound hydrogenase, MbhI subunit	TK2088	0.0051	INF	4			
Membrane bound hydrogenase, NiFe-hydrogenase small subunit	TK2089	0.00015	2.8	4			
Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	TK2093	0.00053	2	4			
Methyl-accepting chemotaxis protein	TK2147	0.011	INF	4			
ABC-type multidrug transport system, ATPase component	TK2161	0.049	INF	3			
Hypothetical membrane protein	TK2201	0.049	INF	3			
Proteasome subunit beta 2	TK2207	0.049	INF	4			
Predicted permease, major facilitator superfamily	TK2266	0.049	INF	4			
Uncharacterized protein	TK2283	0.0033	10	5			
30S ribosomal protein S17e	TK2292	0.048	3	3			
Indolepyruvate oxidoreductase subunit IorA	TK0136				0.022	4.2	4
UPF0173 metal-dependent hydrolase TK0141	TK0141				0.0062	6	5
AMP phosphorylase	TK0352				0.00011	3	4
Uncharacterized protein	TK0483				0.00001	8.1	5
Uncharacterized protein	TK0519				0.0077	9.7	5
Molybdenum cofactor biosynthesis protein B	TK0544				0.00001	3.8	4
NAD-dependent protein deacylase	TK0685				0.037	7.3	4
Molybdopterin oxidoreductase, molybdopterin-binding subunit	TK0690				0.046	7	4
ABC-type transport system involved in Fe-S cluster assembly, ATPase component	TK0731				0.017	3.9	3
Predicted ATP-dependent endonuclease, OLD family	TK0773				0.00022	4.8	5
Putative 5-methylcytosine restriction system, GTPase subunit	TK0795				0.028	6.2	4
Type 2 DNA topoisomerase 6 subunit B	TK0799				0.0026	3	4
Uncharacterized protein	TK0900				0.027	3.4	3
Putative 5-methylcytosine restriction system, GTPase subunit	TK1009				0.002	3.1	4
Large helicase-related protein	TK1015				0.00001	5.9	5
Ribonuclease Z	TK1114				0.012	5.1	4
Site-specific DNA-methyltransferase (adenine-specific)	TK1158				0.047	3.7	3
Uncharacterized protein	TK1304				0.03	INF	4
Predicted ATPase, AAA superfamily	TK1314				0.026	3.9	3
Uncharacterized protein	TK1329				0.031	3.4	3
Uncharacterized protein	TK1376				0.023	INF	4
Uncharacterized protein	TK1394				0.0076	5.9	5

Probable tRNA pseudouridine synthase B	TK1509		0.0003	4.2	5
Predicetd ATPase	TK1549		0.012	3	3
Ribosome maturation protein SDO1 homolog	TK1636		0.048	2.6	3
Uncharacterized protein	TK1663		0.00067	6.8	5
Cysteine synthase	TK1687		0.041	INF	4
Vitamin B12-dependent ribonucleotide reductase	TK1736		0.00001	2.6	4
DNA polymerase II large subunit	TK1903		0.025	2.8	3
Uncharacterized protein	TK2030		0.03	INF	4
Radical_SAM domain-containing protein	TK2160		0.019	6.7	4
LTD domain-containing protein	TK2178		0.041	INF	3
Non-specific serine/threonine protein kinase	TK2250		0.027	3	3
Anaerobic ribonucleoside-triphosphate reductase	TK2298		0.026	3.9	3

Table 4.6: Co-purified interacting partners of OGORα (TK1130) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Rubrerythrin-related protein	TK1056	0.0088	2.2	4	0.0077	7.6	5
2-oxoacid:ferredoxin oxidoreductases, gamma subunit	TK1126	0.00001	11	5	0.00001	INF	5
2-oxoacid:ferredoxin oxidoreductases, beta subunit	TK1129	0.00001	5.6	5	0.00001	22	5
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1130	0.00001	3.4	4	0.00001	22	5
Uncharacterized protein	TK0124	0.043	2.8	3			
Rubrerythrin-related protein, fused to C-terminal DUF835 domain	TK0316	0.019	2.6	3			
Methylmalonyl-CoA epimerase	TK0330	0.041	4.5	4			
Superoxide reductase	TK0525	0.043	7	4			
tRNA(Ile2) 2-agmatinylcytidine synthetase TiaS	TK0553	0.007	1.8	4			
Uncharacterized protein	TK0648	0.027	2.2	3			
Acetate—CoA ligase (ADP-forming)	TK0665	0.00018	2.8	4			
Uncharacterized protein	TK0733	0.037	INF	4			
Uncharacterized protein	TK0900	0.036	1.8	3			
TPR_REGION domain-containing protein	TK0947	0.009	6	5			
2-oxoacid:ferredoxin oxidoreductases, beta subunit	TK1124	0.048	1.8	3			
2-oxoacid: ferredoxin oxidoreductase, delta subunit	TK1131	0.0046	2.8	4			
Uncharacterized protein	TK1202	0.048	1.8	3			
1,4-alpha-glucan branching enzyme TK1436	TK1436	0.033	3.2	3			
Radical_SAM domain-containing protein	TK2160	0.00036	2.1	4			
UPF0173 metal-dependent hydrolase TK0141	TK0141				0.013	5.5	4

RNA-splicing ligase RtcB	TK0358		0.0072	4.2	5
Uncharacterized protein	TK0483		0.0017	4.3	5
Predicted ATP-dependent endonuclease, OLD family	TK0773		0.0085	3.8	4
Type 2 DNA topoisomerase 6 subunit B	TK0799		0.036	2.6	3
Uncharacterized protein	TK1376		0.04	INF	4
50S ribosomal protein L12	TK1415		0.046	11	4
Uncharacterized protein	TK1430		0.025	3.2	3
30S ribosomal protein S8	TK1526		0.014	6	4
Vitamin B12-dependent ribonucleotide reductase	TK1736		0.0056	2.3	4
Predicted transcription regulator, DUF118 helix-turn-helix family	TK1769		0.0047	4.2	5
Probable formate dehydrogenase, alpha subunit	TK2076		0.04	2.3	3

OGOR was identified with multiple subunits of the ATP synthase, MBS, and MBH in +S˚, while multiple subunits of the ABC-type dipeptide/oligopeptide transport system in both conditions, suggesting OGOR may associate with proteins at the membrane involved in both the transport of small peptides into the cell and anabolic reactions related to the two respiratory systems.

2-oxoisovalerate ferredoxin oxidoreductase (VOR)

Ketoisovalerate oxidoreductase (or 2-oxo-isovalerate oxidoreductase, VOR) displays the least specificity of the amino acid fermenting complexes included in this study in that VOR catalyzes the conversion of 2-ketoacids with branched side into acyl-CoA and CO_2 (Figure 4.15A).[38–40,42] *In vitro* studies of VOR from *Thermococcus litoralis* show that *Tl*-VOR displays substrate preference for 2-ketoacid forms of valine, leucine, isoleucine, and methionine while displaying 1/5th of the activity on pyruvate or aryl-pyruvates.[39] The reaction is dependent on Co-A oxidation and generates a reduced Fd in the process.[40,41]

VOR is an $\alpha\beta\delta\gamma$ heterotetramer encoded by the operon TK1978 – TK1981,[38,40] and each subunit was tagged to generate AP-MS data which revealed a moderate to low number of associations for each VOR subunit (range 7 - 30; Figure 4.15B, Figure 4.16). Slightly more associations were observed in +S˚ conditions for most subunits.

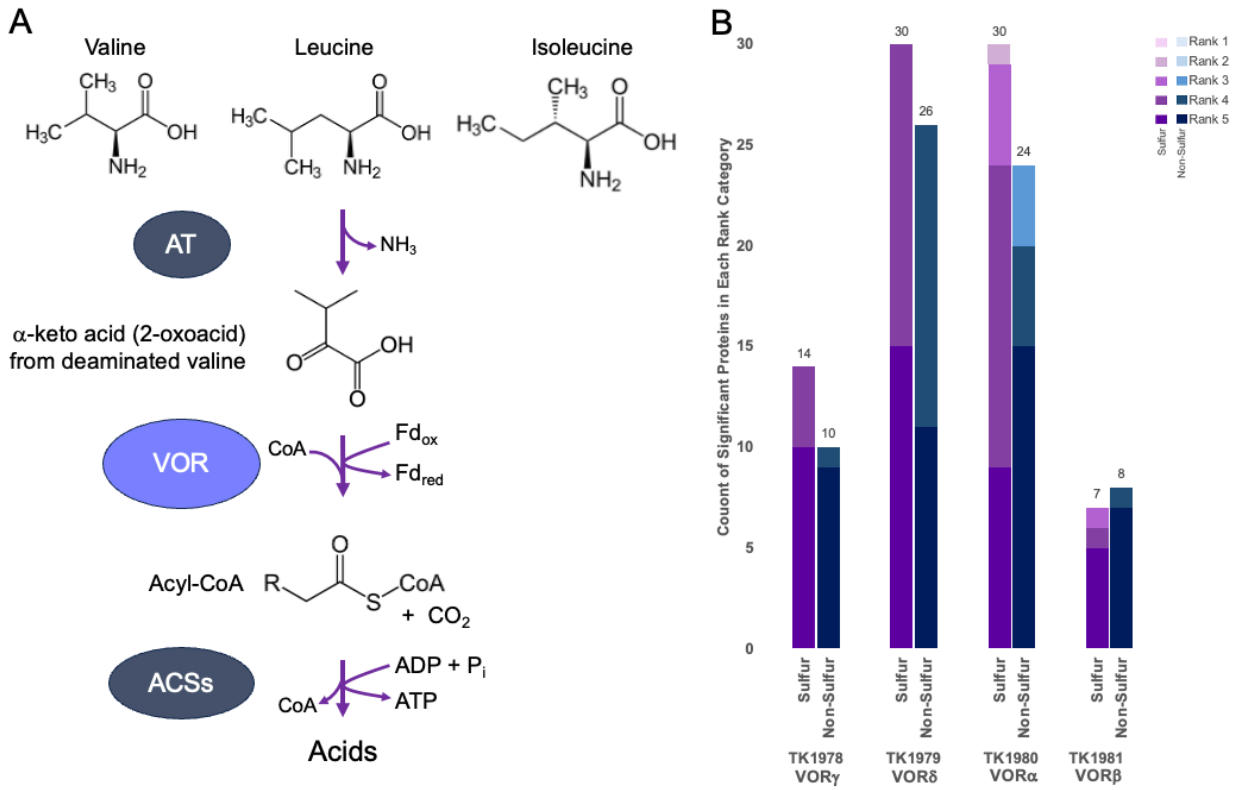


Figure 4.15: VOR function and number of interacting partners observed in sulfur and non-sulfur conditions. Branched chain amino acids are deaminated by an amino transferase (AT) (A) to produce α-keto acids (or 2-oxoacids). VOR (light purple ellipse) preferentially catalyzes the oxidation of α-keto acids into acyl-CoA while generating a reduced Fd. The acyl-CoA is further oxidized with the addition of CO_2 to generate ATP and resulting acids by acyl-CoA synthase (ACS). In (B), the number of interacting partners for each tagged subunit of VOR is shown by rank score. More protein interactions were observed for TK1979 and TK1980 (> 20) and fewer were observed for TK1978 and TK1981 (< 15) in both conditions. The number of interactions that were observed with a rank score of 5 is shown in black, rank 4 is shown in grey, rank 3 is light blue, and rank 2 is dark blue. All proteins identified with TK1978 and TK1979 showed a rank score of 4 or 5 (dark purple or dark blue, suggesting high confidence for these interactions. Only TK1980 in +S˚ identified interactions with a rank score of 2. No interactions with a rank score of 1 were observed for tagged-VOR.

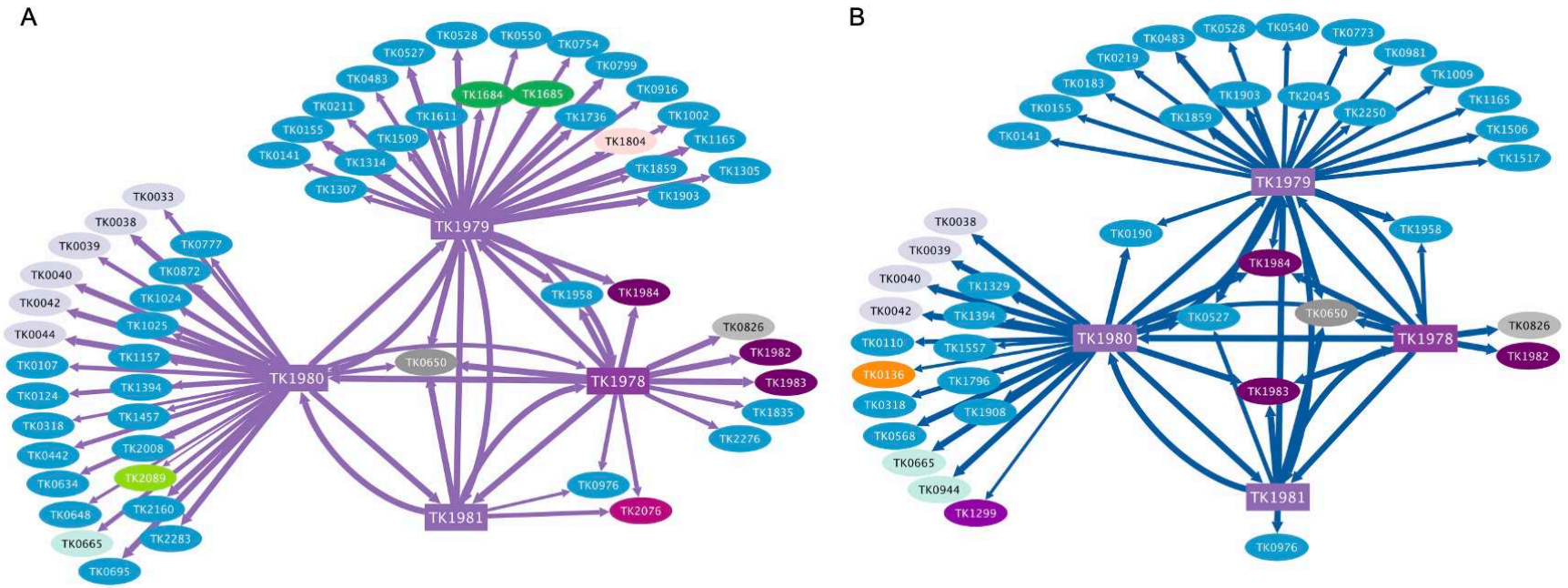


Figure 4.16: Interactome of VOR in sulfur and non-sulfur conditions. The four subunits of the $\alpha\beta\delta\gamma$ heterotetrameric VOR were tagged to generate AP-MS data. All tagged-VOR subunits identified all other VOR subunits (purple) in both sulfur (A) and non-sulfur (B) conditions. Additionally, all VOR subunits identified RBR1 (dark grey ellipse, TK0650) in both conditions and an FeS assembly P-domain containing protein (TK0527) in -S˚. A second rubrerythrin (grey ellipse, RBR2, TK0826) was only identified by VORγ. VORδ (TK1979) identified both subunits of the FNOR2 complex (green ellipses, TK1684 and TK1685) in +S˚ while VORα (TK1980) identified six flagellin proteins in +S˚ and four in -S˚ (light grey, TK0033 - TK0044). VORγ is a shared subunit of the POR heterotetramer (dark purple, TK1982 – TK1984) encoded just upstream of VOR.[3,40,41] VORγ identified all other POR subunits in both conditions as expected, while three out of four subunits identified two additional POR subunits (TK1983 and TK1984) in -S˚ conditions.

Table 4.7: Co-purified interacting partners of VORγ/PORγ (TK1978) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Rubrerythrin-related protein	TK0650	0.00001	INF	5	0.00001	INF	5
Rubrerythrin domain-containing protein	TK0826	0.00001	INF	5	0.00001	INF	5
Uncharacterized protein	TK1958	0.0044	INF	5	0.018	INF	4
Pyruvate/ketoisovalerate oxidoreductases common subunit gamma	TK1978	0.00001	49	5	0.00001	62	5
2-oxoisovalerate:ferredoxin oxidoreductase, delta subunit	TK1979	0.00001	INF	5	0.00001	200	5
2-oxoisovalerate:ferredoxin oxidoreductase, alpha subunit	TK1980	0.00001	83	5	0.00001	270	5
2-oxoisovalerate:ferredoxin oxidoreductase, beta subunit	TK1981	0.00001	160	5	0.00001	INF	5
Pyruvate:ferredoxin oxidoreductase, delta subunit	TK1982	0.00001	INF	5	0.00001	140	5
Pyruvate:ferredoxin oxidoreductase, alpha subunit	TK1983	0.00001	93	5	0.00001	60	5
Pyruvate:ferredoxin oxidoreductase, beta subunit	TK1984	0.00001	420	5	0.00001	340	5
Putative snRNP Sm-like protein	TK0976	0.00001	1.9	4			
Pyrrolidone-carboxylate peptidase	TK1835	0.038	INF	4			
Probable formate dehydrogenase, alpha subunit	TK2076	0.00001	1.7	4			
Orotidine 5'-phosphate decarboxylase	TK2276	0.02	INF	4			

Table 4.8: Co-purified interacting partners of VORδ (TK1979) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
UPF0173 metal-dependent hydrolase TK0141	TK0141	0.026	7.6	4	0.032	6.8	4
RecJ-like exonuclease, containing OB-fold nucleic acid-binding domains	TK0155	0.0034	INF	5	0.013	18	4
Uncharacterized protein	TK0483	0.049	6.2	4	0.00001	18	5
FeS_assembly_P domain-containing protein	TK0527	0.0049	INF	5	0.00041	INF	5
Serine hydroxymethyltransferase	TK0528	0.0024	INF	5	0.028	7	4
Rubrerythrin-related protein	TK0650	0.00001	INF	5	0.00001	INF	5
Predicted AP endonuclease	TK1165	0.00029	8.2	5	0.03	6.3	4
Probable lipoprotein releasing system, ATP-binding protein	TK1859	0.035	12	4	0.004	22	5
DNA polymerase II large subunit	TK1903	0.00043	11	5	0.008	5.5	5
Uncharacterized protein	TK1958	0.0085	INF	5	0.0081	INF	5
Pyruvate/ketoisovalerate oxidoreductases common subunit gamma	TK1978	0.00001	24	5	0.00001	32	5
2-oxoisovalerate:ferredoxin oxidoreductase, delta subunit	TK1979	0.00001	150	5	0.00001	88	5
2-oxoisovalerate:ferredoxin oxidoreductase, alpha subunit	TK1980	0.00001	41	5	0.00001	48	5

2-oxoisovalerate:ferredoxin oxidoreductase, beta subunit	TK1981	0.00001	90	5	0.00001	64	5
Pyruvate:ferredoxin oxidoreductase, beta subunit	TK1984	0.003	18	5	0.03	6.3	4
Amidophosphoribosyltransferase	TK0211	0.03	12	4			
Proline--tRNA ligase	TK0550	0.039	18	4			
tRNA(Met) cytidine acetyltransferase TmcA	TK0754	0.018	21	4			
Type 2 DNA topoisomerase 6 subunit B	TK0799	0.00001	12	5			
L-threonine 3-dehydrogenase	TK0916	0.046	11	4			
Adenylosuccinate synthetase	TK1002	0.021	INF	4			
Probable translation initiation factor IF-2	TK1305	0.018	8.5	4			
Nucleoside diphosphate kinase	TK1307	0.046	11	4			
Predicted ATPase, AAA superfamily	TK1314	0.034	19	4			
Probable tRNA pseudouridine synthase B	TK1509	0.015	9.7	4			
Predicted metal-dependent hydrolase	TK1611	0.037	INF	4			
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1684	0.0085	INF	5			
Ferredoxin:NADP oxidoreductase, beta subunit	TK1685	0.031	INF	4			
Vitamin B12-dependent ribonucleotide reductase	TK1736	0.0048	5.1	5			
ABC-type dipeptide/oligopeptide transport system, probable periplasmic component	TK1804	0.02	11	4			
Fibrillarin-like rRNA/tRNA 2'-O-methyltransferase	TK0183				0.035	15	4
GMP synthase [glutamine-hydrolyzing] subunit A	TK0190				0.02	INF	4
Sugar-phosphate nucleotidyltransferase	TK0219				0.013	18	4
Metal-dependent phosphohydrolase, HD superfamily	TK0540				0.019	11	4
Predicted ATP-dependent endonuclease, OLD family	TK0773				0.02	INF	4
N2, N2-dimethylguanosine tRNA methyltransferase	TK0981				0.044	5	4
Putative 5-methylcytosine restriction system, GTPase subunit	TK1009				0.032	6.8	4
30S ribosomal protein S13	TK1506				0.0081	INF	5
Adenylate kinase	TK1517				0.043	14	4
SAM-dependent methyltransferase, UPF0020 family	TK2045				0.00013	INF	4
Non-specific serine/threonine protein kinase	TK2250				0.014	5.6	4

Table 4.9: Co-purified interacting partners of VORα (TK1980) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Flagellin B1	TK0038	0.00001	4.5	5	0.00001	31	5
Flagellin B2	TK0039	0.00001	3.8	4	0.00001	40	5
Flagellin B3	TK0040	0.00001	8.2	5	0.00001	INF	5
Flagellin B5	TK0042	0.00001	4.1	5	0.00013	23	5

DUF3216 domain-containing protein	TK0318	0.017	2.7	3	0.0011	INF	4
Acetate--CoA ligase (ADP-forming)	TK0665	0.004	2.2	4	0.00051	9	5
Uncharacterized protein	TK1394	0.00001	2.6	4	0.012	INF	4
Pyruvate/ketoisovalerate oxidoreductases common subunit gamma	TK1978	0.00061	2	4	0.0024	2.9	4
2-oxoisovalerate:ferredoxin oxidoreductase, delta subunit	TK1979	0.00001	9.1	5	0.0019	17	5
2-oxoisovalerate:ferredoxin oxidoreductase, alpha subunit	TK1980	0.00001	3.5	4	0.00001	53	5
2-oxoisovalerate:ferredoxin oxidoreductase, beta subunit	TK1981	0.00001	4.4	5	0.00001	12	5
Uncharacterized protein	TK0033	0.00081	2.7	4			
Archaeal flagella-related protein D, internal insertion	TK0044	0.00001	11	5			
Manganese-dependent transcription regulator	TK0107	0.046	6	4			
Uncharacterized protein	TK0124	0.03	6.7	4			
Uncharacterized protein	TK0442	0.03	6.7	4			
Chemotaxis histidine kinase, flame shift	TK0634	0.00013	3.9	4			
Uncharacterized protein	TK0648	0.023	2.6	3			
Rubrerythrin-related protein	TK0650	0.0067	3.7	4			
RHH_1 domain-containing protein	TK0695	0.01	INF	5			
Tyrosine recombinase XerA	TK0777	0.04	4	4			
tRNA/rRNA cytosine-C5-methylase, NOL1/NOP2/Sun family	TK0872	0.03	6.7	4			
Hypothetical membrane protein	TK1024	0.017	INF	3			
Uncharacterized protein	TK1025	0.00001	3.6	4			
CDC48/VCP homolog, AAA superfamily	TK1157	0.028	INF	3			
Probable phosphate transport system regulator, fused to TrkA-C domain	TK1457	0.029	2.7	3			
Hydrogenase maturation factor HypA	TK2008	0.011	4	4			
Membrane bound hydrogenase, NiFe-hydrogenase small subunit	TK2089	0.047	INF	2			
Radical_SAM domain-containing protein	TK2160	0.00001	5.9	5			
Uncharacterized protein	TK2283	0.0061	INF	5			
Hydrolase, HAD superfamily	TK0110				0.0075	INF	4
Indolepyruvate oxidoreductase subunit IorA	TK0136				0.036	2.6	3
GMP synthase [glutamine-hydrolyzing] subunit A	TK0190				0.00001	6.3	5
FeS_assembly_P domain-containing protein	TK0527				0.0004	INF	5
Tyrosine--tRNA ligase	TK0568				0.00001	15	5
Acetate--CoA ligase (ADP-forming)	TK0944				0.0004	INF	5
NAD(P)H sulfur oxidoreductase (CoA-dependent)	TK1299				0.046	3.4	3
Uncharacterized protein	TK1329				0.02	INF	4
Predicted dehydrogenase	TK1557				0.046	2.1	3
Glutamine synthetase	TK1796				0.013	3.9	3

Lipoate-protein ligase A, N-terminal section	TK1908				0.00025	INF	5
Pyruvate:ferredoxin oxidoreductase, alpha subunit	TK1983				0.00001	12	5
Pyruvate:ferredoxin oxidoreductase, beta subunit	TK1984				0.00001	44	5

Table 4.10: Co-purified interacting partners of VORβ (TK1981) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Rubrerythrin-related protein	TK0650	0.00001	INF	5	0.00001	INF	5
Putative snRNP Sm-like protein	TK0976	0.032	2.3	3	0.00001	6.9	5
Pyruvate/ketoisovalerate oxidoreductases common subunit gamma	TK1978	0.00001	67	5	0.00001	47	5
2-oxoisovalerate:ferredoxin oxidoreductase, delta subunit	TK1979	0.00001	INF	5	0.00001	INF	5
2-oxoisovalerate:ferredoxin oxidoreductase, alpha subunit	TK1980	0.00001	200	5	0.00001	180	5
2-oxoisovalerate:ferredoxin oxidoreductase, beta subunit	TK1981	0.00001	INF	5	0.00001	350	5
Probable formate dehydrogenase, alpha subunit	TK2076	0.04	5.5	4			
FeS_assembly_P domain-containing protein	TK0527				0.015	INF	4
Pyruvate:ferredoxin oxidoreductase, alpha subunit	TK1983				0.00001	6.3	5

VOR maintains ~ eleven iron and ~ twelve sulfur atoms per complex,[40] thus it can be assumed that the association of the Fe-S assembly protein (TK0527) in -S° conditions is aiding the acquisition of sulfur atoms when such are scarce. Rubrerythrins are non-heme, iron-containing proteins which act as peroxidases suggesting that they aid in oxidative stress responses.[44,58] The clear associations of VOR with rubrerythrins, primarily RBR1 (TK0650), suggests that VOR benefits from associating with proteins that assist in maintaining appropriate local redox conditions. The association of VORα with the flagellin proteins in both +S° and -S° is not reconciled.

A discussion of the results from tagged-POR ensues, followed by summative remarks concerning the activity and function of POR and VOR *in vivo.*

Pyruvate ferredoxin oxidoreductase (POR)

Pyruvate generated in glycolysis is converted to acetyl-CoA and CO_2 via the Fd-dependent pyruvate oxidoreductase (POR) (Figure 4.17).[43,45] POR was thought to be encoded by three genes in the POR operon (TK1982 - TK1984)[42] until it was determined that a fourth subunit (PORγ, TK1978) is shared between POR and VOR (TK1978 - TK1981) which is encoded directly upstream of POR (Figure 4.17A).[3,39,40] While algorithms which predict microbial operons consider the entire genomic region encoding VOR and POR (TK1978 - TK1984) series within a single operon,[59,60] analysis of the *T. kodakarensis* transcriptome report three primary transcription start sites each preceding the following genes: TK1978, TK1979, or TK1982.[31] Thus, it seems that expression of the γ subunit (TK1978) may be uncoupled from expression of the remaining three genes encoding either VOR (TK1979 – TK1981) or POR (TK1982 – TK1984) (Figure 4.17A). POR is thought to be composed of four homodimeric subunits ($\alpha_2\beta_2\gamma_2\delta_2$, TK1978 and TK1982 - TK1984).[41]

POR was seen to interact with all three Fds,[42] and resolved ~ 6 - 22 interactions except for one subunit, PORδ which identified 74 interacting partners in -S˚ conditions (Figure 4.17B, Figure 4.18).

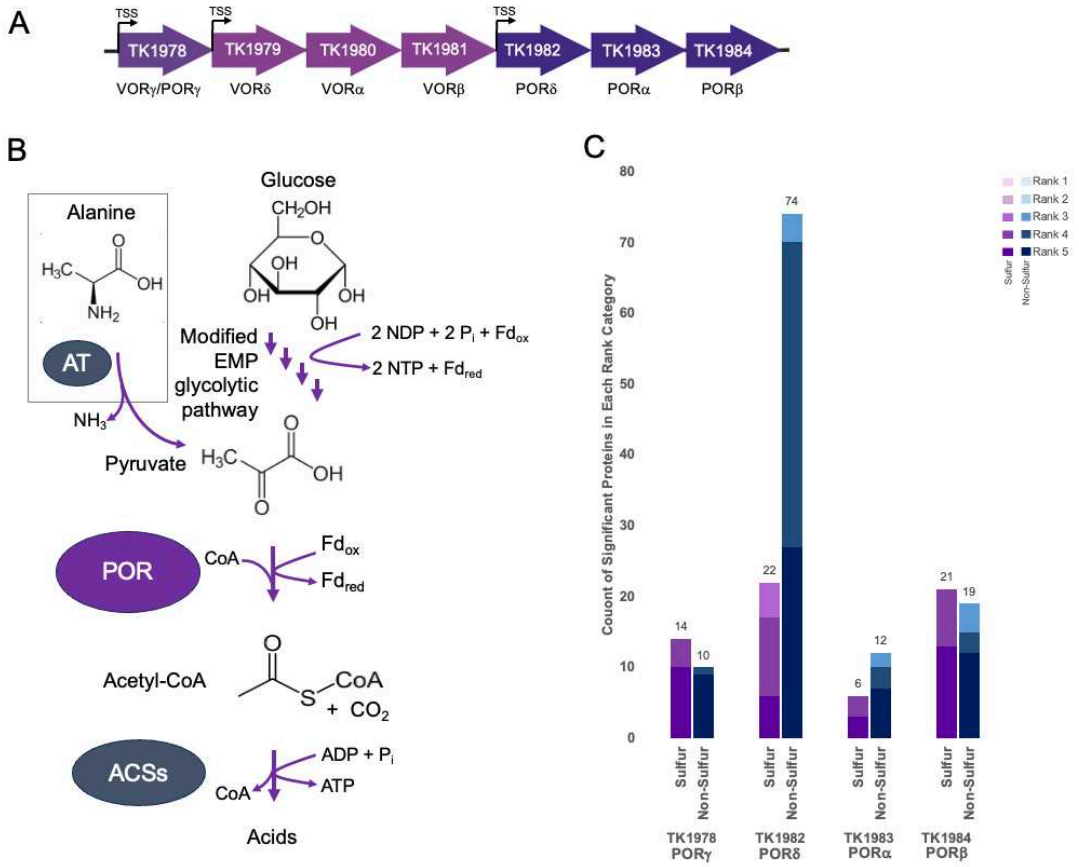


Figure 4.17: POR genomic organization, function, and number of interacting partners observed in sulfur and non-sulfur conditions. POR and VOR share a γ subunit encoded by the first gene (purple, TK1978) in the VOR/POR operon (A). The remaining three subunits which make up the POR heterotetramer (dark purple, TK1982 - TK1984) are encoded directly downstream of VOR (lighter purple, TK1978 - TK1981). Three primary transcription start sites (TSS) appear upstream of TK1978, TK1979, and TK1984 (A). In (B), T. kodakarensis utilizes a modified EMP pathway in glycolysis generating net 2 ATP and 2 reduced Fds for the conversion of glucose into pyruvate. POR catalyzes the conversion of pyruvate into acetyl-CoA. POR (dark purple) preferentially catalyzes the oxidation of pyruvate (from glycolysis) into acetyl-CoA while generating a reduced Fd. Note that pyruvate is homologous to an α-ketoacid from a deaminated alanine (grey box), thus POR can act on substrates from amino acid fermentation pathways. The acetyl-CoA is further oxidized with the addition of CO_2 to generate ATP and resulting acids by acyl-CoA synthase (ACS). In (C), the number of interacting partners for each tagged subunit of POR is shown by rank score. Most subunits identified ~10 - 20 interacting partners with the exception of TK1982 which identified 74 interacting partners in -S˚ conditions. The number of interactions that were observed with a rank score of 5 is shown in dark purple or dark blue, rank 4 is shown in purple or blue, and rank 3 is light purple or light blue. No interactions with a rank score of 2 or 1 were observed for tagged-POR suggesting interactions were identified with high confidence.

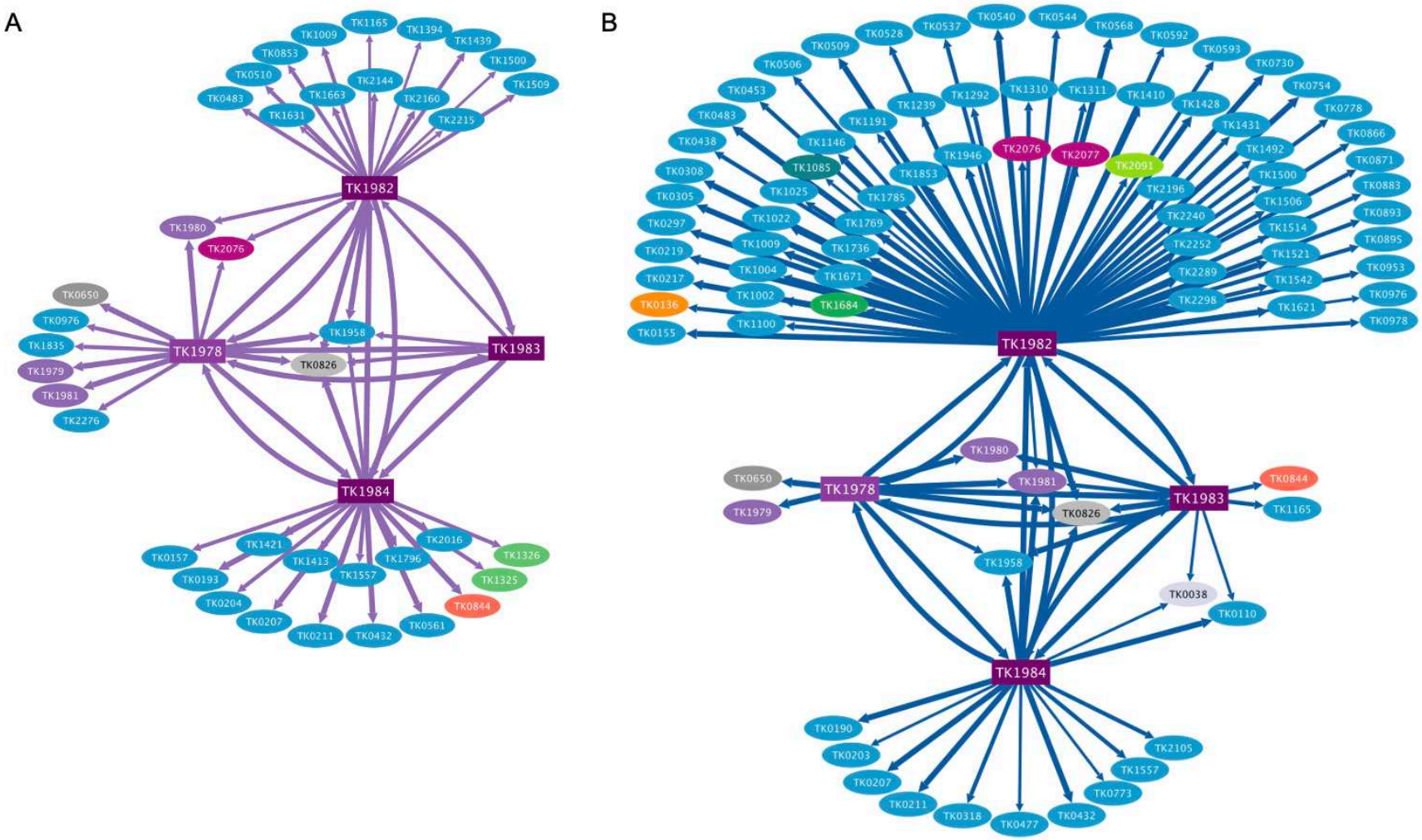


Figure 4.18: Interactome of POR in sulfur and non-sulfur conditions. All four subunits of POR were tagged for AP-MS analysis and the sulfur (A) and non-sulfur (B) interactomes are shown here. Proteins which were identified as significant are shown in colored ellipses and line thickness indicates the rank of each association (rank 5 is the thickest, with a rank of 1 the thinnest). The tagged proteins are shown in purple rectangles. Most subunits of POR displayed an average number of associations (range 6 - 22) with the exception of PORδ (TK1982) in -S˚ where 74 associations were observed. All tagged subunits identified all other POR subunits in both conditions. RBR2 (grey ellipse) was identified by all subunits in +S˚ and ¾ subunits in -S˚, suggesting that association with this

RBR is important for POR activity. RBR1 (grey ellipse) was found in association with PORγ in both conditions. An uncharacterized protein, TK1958, was also identified by all subunits in +S˚ and ¾ subunits in -S˚, however the importance of this association is unknown. PORβ associated with both subunits of FNOR1 (TK1325 and TK1326) in +S˚, while PORδ was found with one FNOR2 subunit (green ellipse, TK1684) in -S˚. One subunit of FDH (TK2076, fuchsia ellipse) was identified by two subunits in +S˚, while two subunits of FDH (TK2076 and TK2077, fuchsia ellipses) were identified by PORδ in -S˚. Only one flagellin protein (light grey ellipse, TK0038) was identified by two subunits of POR in -S˚. AOR (salmon ellipse, TK0844) was identified by PORβ in +S˚ and PORδ in -S˚ conditions. PORγ identified all subunits of VOR (light purple) in +S˚. The three remaining POR subunits identified VORβ, and PORα identified VORδ in -S˚, suggesting a strong association between POR and VOR in -S˚ conditions.

Table 4.11: Co-purified interacting partners of PORδ (TK1982) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Uncharacterized protein	TK0483	0.00001	3.2	4	0.00001	56	5
Rubrerythrin domain-containing protein	TK0826	0.00001	87	5	0.00001	INF	5
Putative 5-methylcytosine restriction system, GTPase subunit	TK1009	0.00062	2.3	4	0.00001	140	5
30S ribosomal protein S9	TK1500	0.038	2.3	3	0.017	50	4
Pyruvate/ketoisovalerate oxidoreductases common subunit gamma	TK1978	0.00001	19	5	0.00001	46	5
Pyruvate:ferredoxin oxidoreductase, delta subunit	TK1982	0.00001	66	5	0.00001	INF	5
Pyruvate:ferredoxin oxidoreductase, alpha subunit	TK1983	0.00001	29	5	0.00001	110	5
Pyruvate:ferredoxin oxidoreductase, beta subunit	TK1984	0.00001	56	5	0.00001	270	5
Probable formate dehydrogenase, alpha subunit	TK2076	0.00001	2	4	0.00001	3.6	4
Diadenylate cyclase	TK0510	0.0077	2.3	4			
UPF0284 protein TK0853	TK0853	0.02	6.5	4			
Predicted AP endonuclease	TK1165	0.039	2	3			
Uncharacterized protein	TK1394	0.022	3.4	3			
Putative nickel-responsive regulator	TK1439	0.038	9	4			
Probable tRNA pseudouridine synthase B	TK1509	0.00001	3.6	4			
Uncharacterized protein	TK1631	0.02	6.5	3			
Uncharacterized protein	TK1663	0.031	6	4			
Uncharacterized protein	TK1958	0.00001	INF	5			
2-oxoisovalerate:ferredoxin oxidoreductase, alpha subunit	TK1980	0.0071	2.3	4			
Uncharacterized protein	TK2144	0.046	5.5	4			
Radical_SAM domain-containing protein	TK2160	0.0052	3.5	4			
tRNA-splicing endonuclease	TK2215	0.032	2.4	3			
Indolepyruvate oxidoreductase subunit IorA	TK0136				0.00036	2	4
RecJ-like exonuclease, containing OB-fold nucleic acid-binding domains	TK0155				0.00059	82	5

Pyridoxal 5'-phosphate synthase subunit PdxS	TK0217		0.00001	100	5
Sugar-phosphate nucleotidyltransferase	TK0219		0.00001	38	5
L-aspartate oxidase	TK0297		0.0033	INF	5
Uridylate kinase	TK0305		0.0034	41	5
Elongation factor 1-alpha	TK0308		0.00001	4	5
Uncharacterized protein	TK0438		0.022	INF	4
Uncharacterized protein	TK0453		0.011	INF	4
Predicted GTPase, containing TGS domain	TK0506		0.02	INF	4
Flavin prenyltransferase UbiX	TK0509		0.0086	INF	5
Serine hydroxymethyltransferase	TK0528		0.012	54	4
Peroxiredoxin	TK0537		0.032	INF	4
Metal-dependent phosphohydrolase, HD superfamily	TK0540		0.00001	99	5
Molybdenum cofactor biosynthesis protein B	TK0544		0.00001	2.5	4
Tyrosine--tRNA ligase	TK0568		0.00082	79	5
Uncharacterized protein	TK0592		0.046	INF	4
Uncharacterized protein	TK0593		0.018	INF	4
ABC-type transport system involved in Fe-S cluster assembly, permease component	TK0730		0.003	67	5
tRNA(Met) cytidine acetyltransferase TmcA	TK0754		0.006	INF	5
Toprim domain-containing protein	TK0778		0.0011	2.6	4
2,3-bisphosphoglycerate-independent phosphoglycerate mutase	TK0866		0.0023	2.5	4
Ornithine carbamoyltransferase	TK0871		0.04	INF	4
Uncharacterized protein	TK0883		0.036	INF	4
Pyruvate fromate-lyase activating enzyme-related protein, radical SAM superfamily	TK0893		0.021	48	4
S-layer protein	TK0895		0.005	2.6	4
Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains	TK0953		0.032	44	4
Putative snRNP Sm-like protein	TK0976		0.024	2.5	3
Glycine--tRNA ligase	TK0978		0.02	INF	4
Adenylosuccinate synthetase	TK1002		0.00001	50	5
UDP-glucose 4-epimerase	TK1004		0.029	45	4
D-aminopeptidase	TK1022		0.035	43	4
Uncharacterized protein	TK1025		0.028	INF	4
Protein disulfide oxidoreductase	TK1085		0.027	2.2	3
Translation initiation factor 2 subunit alpha	TK1100		0.024	47	4
Phosphoglycerate kinase	TK1146		0.0054	INF	5
30S ribosomal protein S8e	TK1191		0.00028	2.9	4
Peptide chain release factor subunit 1	TK1239		0.013	53	4

Phosphoenolpyruvate synthase	TK1292		0.032	INF	4
30S ribosomal protein S28e	TK1310		0.001	2.3	4
50S ribosomal protein L7Ae	TK1311		0.0097	2.5	4
DNA primase DnaG	TK1410		0.00001	3.3	4
Cleavage and polyadenylation specficity factor subunit homolog	TK1428		0.00001	40	5
Glutamate dehydrogenase	TK1431		0.00001	2.8	4
Uncharacterized protein	TK1492		0.043	18	4
30S ribosomal protein S13	TK1506		0.012	INF	4
Cytidylate kinase	TK1514		0.016	INF	4
30S ribosomal protein S5	TK1521		0.0066	4.6	5
50S ribosomal protein L3	TK1542		0.018	INF	4
Translation initiation factor 2 subunit beta	TK1621		0.00001	INF	5
S-adenosyl-L-methionine-dependent tRNA 4-demethylwyosine synthase	TK1671		0.022	INF	4
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1684		0.036	INF	4
Vitamin B12-dependent ribonucleotide reductase	TK1736		0.00001	16	5
Predicted transcription regulator, DUF118 helix-turn-helix family	TK1769		0.003	INF	5
Probable tRNA/rRNA methyltransferase	TK1785		0.032	INF	4
Type II/IV secretion system ATPase	TK1853		0.00062	2	4
Translation initiation factor 2 subunit gamma	TK1946		0.00045	23	5
2-oxoisovalerate:ferredoxin oxidoreductase, beta subunit	TK1981		0.021	2.7	3
4Fe-4S cluster-binding protein	TK2077		0.00062	2.8	4
Membrane bound hydrogenase, NiFe-hydrogenase large subunit 2	TK2091		0.029	45	4
Aspartate carbamoyltransferase	TK2196		0.032	29	4
Lysine--tRNA ligase	TK2240		0.021	3.4	3
Proteasome-activating nucleotidase	TK2252		0.018	INF	4
Archaeal histone B	TK2289		0.0081	2.1	4
Anaerobic ribonucleoside-triphosphate reductase	TK2298		0.0018	INF	5

Table 4.12: Co-purified interacting partners of PORα (TK1983) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Rubrerythrin domain-containing protein	TK0826	0.00001	53	4	0.00001	INF	5
Uncharacterized protein	TK1958	0.00001	INF	4	0.0013	INF	5
Pyruvate/ketoisovalerate oxidoreductases common subunit gamma	TK1978	0.00001	13	5	0.00001	41	5
Pyruvate:ferredoxin oxidoreductase, delta subunit	TK1982	0.00001	41	4	0.00001	INF	5
Pyruvate:ferredoxin oxidoreductase, alpha subunit	TK1983	0.00001	16	5	0.00001	89	5

Pyruvate:ferredoxin oxidoreductase, beta subunit	TK1984	0.00001	29	5	0.00001	170	5
Flagellin B1	TK0038				0.023	INF	3
Hydrolase, HAD superfamily	TK0110				0.031	INF	3
Aldehyde ferredoxin oxidoreductase	TK0844				0.012	12	4
Predicted AP endonuclease	TK1165				0.028	6.8	4
2-oxoisovalerate:ferredoxin oxidoreductase, alpha subunit	TK1980				0.00001	7.7	5
2-oxoisovalerate:ferredoxin oxidoreductase, beta subunit	TK1981				0.049	4.9	4

Table 4.13: Co-purified interacting partners of PORβ (TK1984) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Formate-dependent phosphoribosylglycinamide formyltransferase	TK0207	0.0021	9.3	5	0.00001	72	5
Amidophosphoribosyltransferase	TK0211	0.00001	24	5	0.00001	21	5
Phosphoribosylaminoimidazole-succinocarboxamide synthase	TK0432	0.00001	INF	5	0.00001	INF	5
Rubrerythrin domain-containing protein	TK0826	0.00001	9.9	5	0.00001	INF	5
Predicted dehydrogenase	TK1557	0.00001	3.9	4	0.0067	2	4
Uncharacterized protein	TK1958	0.011	INF	4	0.00001	INF	5
Pyruvate/ketoisovalerate oxidoreductases common subunit gamma	TK1978	0.00001	16	5	0.00001	48	5
Pyruvate:ferredoxin oxidoreductase, delta subunit	TK1982	0.00001	51	5	0.00001	36	5
Pyruvate:ferredoxin oxidoreductase, alpha subunit	TK1983	0.00001	8.9	5	0.00001	37	5
Pyruvate:ferredoxin oxidoreductase, beta subunit	TK1984	0.00001	29	5	0.00001	78	5
Xanthine/uracilpermease	TK0157	0.037	INF	4			
GMP synthase [glutamine-hydrolyzing] subunit B	TK0193	0.00001	8.8	5			
Phosphoribosylamine--glycine ligase	TK0204	0.016	INF	4			
Adenylosuccinate lyase	TK0561	0.009	5.1	5			
Aldehyde ferredoxin oxidoreductase	TK0844	0.0033	5.4	5			
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1325	0.041	7.5	4			
Ferredoxin:NADP oxidoreductase, beta subunit	TK1326	0.025	INF	4			
Archaeal histone A	TK1413	0.0021	INF	4			
Cell division protein FtsZ 1	TK1421	0.0014	21	5			
Glutamine synthetase	TK1796	0.00001	8	5			
Iron-molybdenum cofactor-binding protein	TK2016	0.011	INF	4			
Flagellin B1	TK0038				0.017	12	3
Hydrolase, HAD superfamily	TK0110				0.0005	INF	5

GMP synthase [glutamine-hydrolyzing] subunit A	TK0190		0.00001	5.2	5
ATP-grasp domain-containing protein	TK0203		0.022	INF	3
DUF3216 domain-containing protein	TK0318		0.042	INF	4
Glyceraldehyde 3-phosphate phosphatase	TK0477		0.026	2.3	3
Predicted ATP-dependent endonuclease, OLD family	TK0773		0.035	3.8	3
2-oxoisovalerate:ferredoxin oxidoreductase, beta subunit	TK1981		0.00001	8	5
Uncharacterized protein	TK2105		0.022	INF	4

POR and VOR parallels and joint associations

VOR functions as a heterotetramer encoded by TK1978 - TK1981. It should be noted that while POR is fully described as acting in the conversion of pyruvate to acetyl-CoA following glycolysis (Figure 4.17A), pyruvate is the conjugate base of a 2-oxo-acid and synonymous with a deaminated alanine, thus POR and VOR can function on similar substrates, albeit from different progenitors. It is hypothesized that while VOR and POR clearly display substrate specificity *in vitro*,[3,39–43] their *in vivo* activity may be more dynamic. Further, while they can clearly function as independent multimeric complexes, the data suggests that multiple subunits of each respective complex identify the other complex, especially in -S˚ conditions; there may be a tendency for heteromeric subunit cooperation in -S˚ conditions (Figure 4.19).

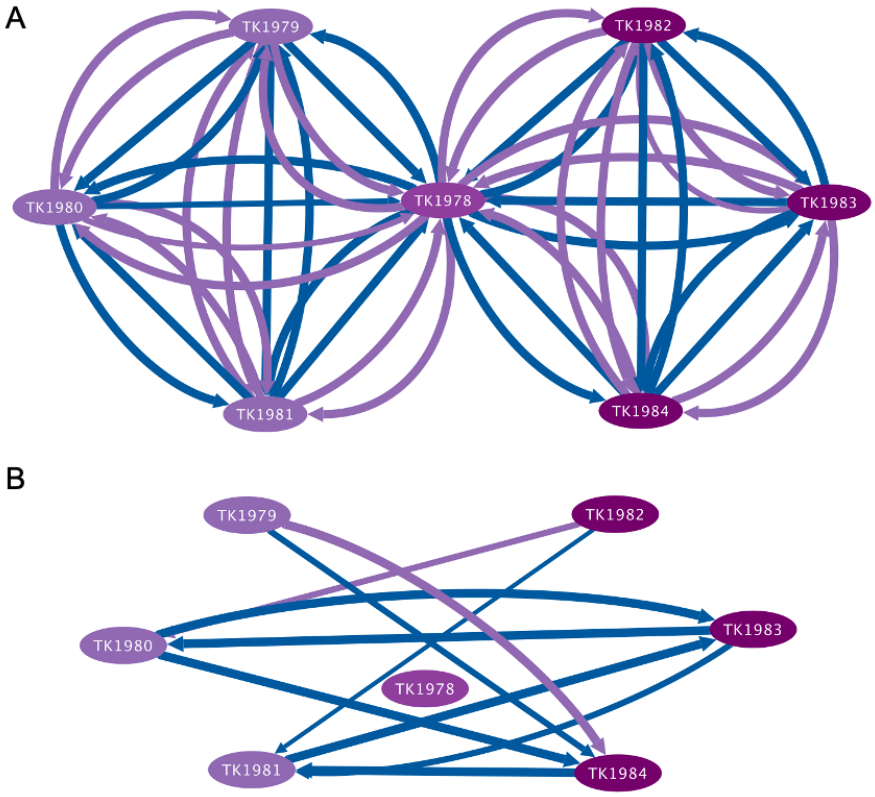


Figure 4.19: VOR and POR complexes maintain overlapping associations connecting the two complexes. POR and VOR are both heterotetramers and share a γ subunit, TK1978. All seven POR/VOR subunits were tagged and most tagged subunits identified all three of the other subunits of their respective complexes (A) in both sulfur (purple lines) and non-sulfur (blue lines) conditions. In (B), intra-complex associations have been removed and instead, associations between complexes (excluding VORg, TK1978) are shown. Eight inter-complex associations are observed in -S˚ while two inter-complex associations are observed in +S˚.

Additional comparisons can be made between the *in vivo* associations of POR and VOR. While VOR associated with both subunits of FNOR2 in +S˚, POR was seen to associate with both subunits of FNOR1 in +S˚. While all four subunits of POR identified the uncharacterized protein, TK1958, this protein was only identified by two VOR subunits. Both complexes showed some associations with FDH, specifically FDHα, TK2076. While VOR associated with RBR1 (TK0650), POR seems to associate with RBR2 (TK0826). It is plausible that these oxidoreductases benefit from the local peroxidase activity of rubrerythrins. As RBR1 and RBR2 were identified with such repetition and confidence, it was determined that analyzing the network of these two proteins would be of interest. These are the only two proteins which were added to the study after initial results were gathered.

Rubrerythrins

Rubrerythrins are iron-containing proteins which assist in oxygen and peroxide detoxification (Figure 4.20A).[44,58] *T. kodakarensis* encodes three rubrerythrins (RBR1-3: TK0650, TK0826, and TK1056), two of which were identified as interacting partners with the Fds: TK0650 was found in association with Fd1 and TK1056 was found in association with Fd2.[2] While these proteins were not originally included in this study, the consistent association between RBR1 and VOR and RBR2 and POR suggested that including them in the AP-MS metabolic network was of interest (Figure 4.16 and Figure 4.18). RBR3 (TK1056) was not included in this study. Both RBR1 and RBR2 displayed a large number of interactions except in -S˚, RBR1 only identified eleven interacting partners (Figure 4.20B, Figure 4.21).

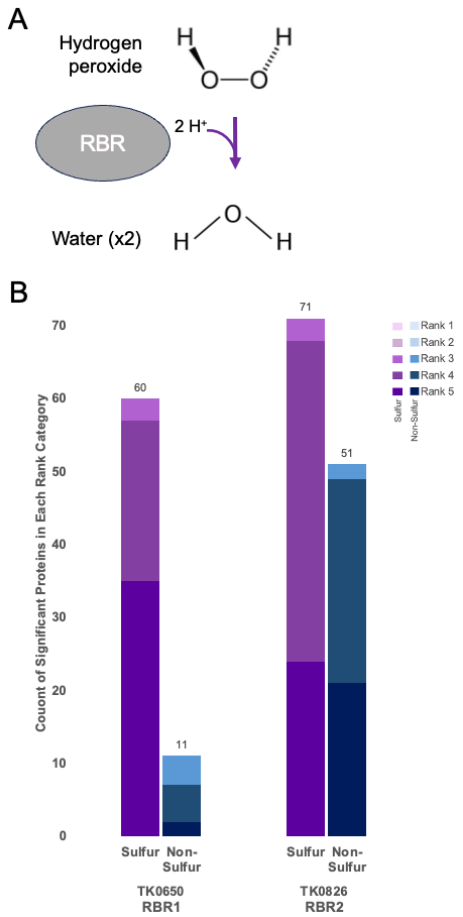


Figure 4.20 RBR function and number of interacting partners observed in sulfur and non-sulfur conditions. Rubrerythrins (RBR) contain non-heme di-iron centers and function as peroxidases. Here, RBR is shown converting hydrogen peroxide into water (A). Two RBRs were tagged and analyzed via AP-MS and were found to have abundant interacting partners (B) with the exception of TK0650 in -S˚ conditions where only eleven copurified proteins were identified as significant. The number of interactions that were observed with a rank score of 5 is shown in dark purple or dark blue, rank 4 is shown in purple or blue, and rank 3 is light purple or light blue. No interactions with a rank score of 2 or 1 were observed for either tagged-RBR.

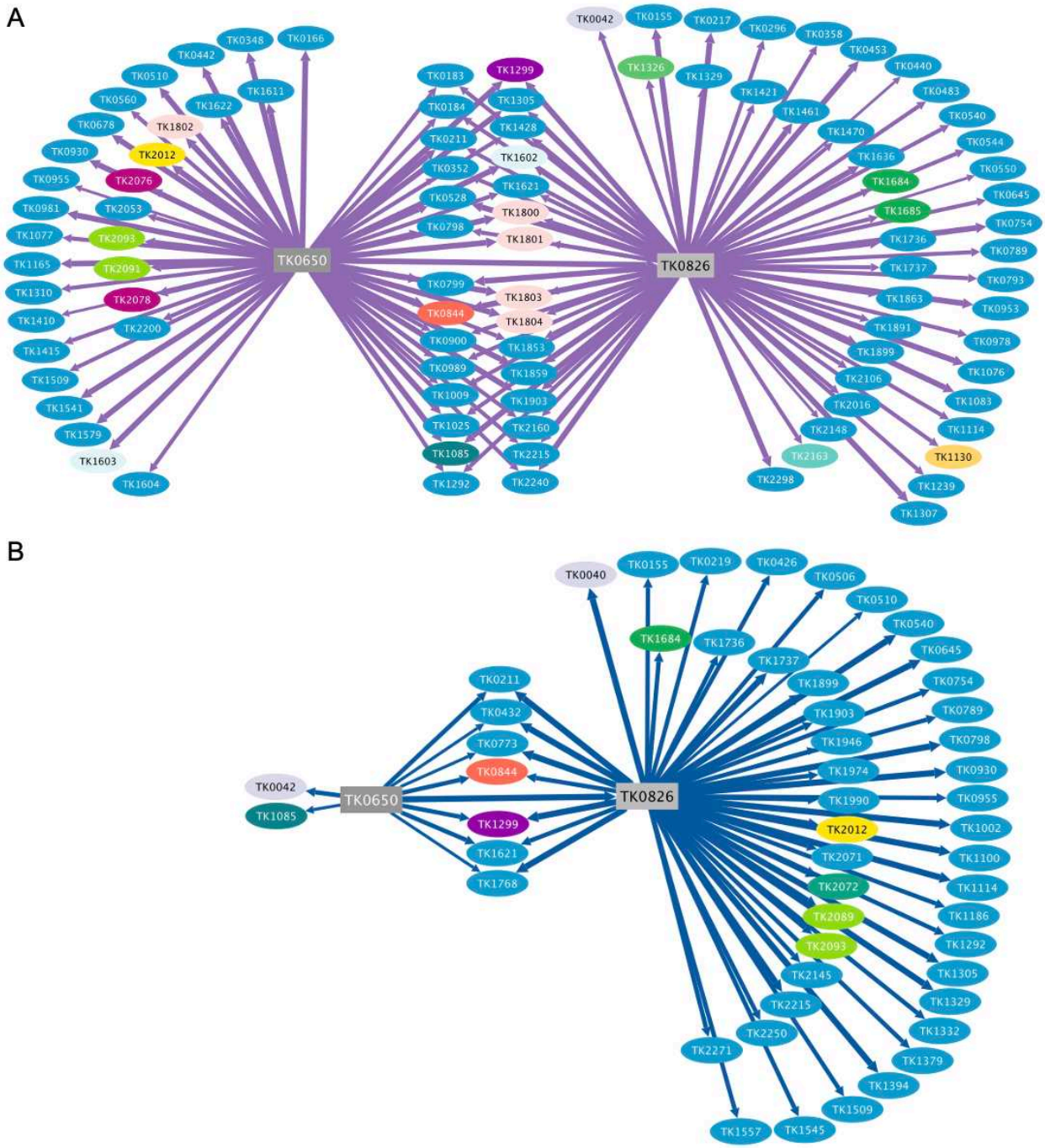


Figure 4.21: Interactome of Rubrerythrins in sulfur and non-sulfur conditions. Two rubrerythrins were tagged for AP-MS analysis: RBR1 (TK0650) and RBR2 (TK0826). Resulting proteins identified as interacting partners are shown here in sulfur conditions (A) and non-sulfur conditions (B) with line thickness relating to the rank score of the protein identified (thicker lines represent a greater rank score). Both RBRs identified four proteins in both +S˚ and -S˚ conditions: amidophosphoribosyltransferase (blue ellipse, TK0211), AOR (salmon ellipse, TK0844), NAD(P)H sulfur oxidoreductase (bright purple ellipse, TK1299), and a translation initiation factor 2β (blue ellipse, TK1621). In +S˚, both proteins identified protein disulfide oxidoreductase (dark teal ellipse, TK1085), an ATP synthase (light blue ellipse, TK1602), and four subunits of an ABC-type dipeptide/oligopeptide transport system (light peach ellipses,

TK1800 - TK1804), none of which were identified in -S˚ with the exception of RBR1 associating with the protein disulfide oxidoreductase (dark teal ellipse, TK1085). Interestingly, RBR1 associated with Fd3 (yellow ellipse, TK2012) and two subunits of the MBH complex (lime ellipses, TK2091 and TK2093) in +S˚, while associations with Fd3 and MBH were seen with RBR2 in -S˚. RBR1 associated with two subunits of FDH (fuchsia ellipses, TK2076 and TK2078) in +S˚, and RBR1 associated with both subunits of FNOR2 (green ellipses, TK1684 and TK1685) in +S˚ while only FNOR2α (TK1684) was seen in -S˚. RBR2 was found with FNOR1β (light green ellipse, TK1326), OGORα2 (light yellow ellipse, TK1130), and GAPOR (sea foam green ellipse, TK2163) in +S˚, but not in -S˚.

Table 4.14: Co-purified interacting partners of RBR1 (TK0650) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Amidophosphoribosyltransferase	TK0211	0.0012	19	5	0.00032	2.5	4
Rubrerythrin-related protein	TK0650	0.00001	INF	5	0.00001	350	5
Rubrerythrin domain-containing protein	TK0826	0.00094	INF	5	0.00088	INF	5
Aldehyde ferredoxin oxidoreductase	TK0844	0.0041	16	5	0.0037	3.2	4
Protein disulfide oxidoreductase	TK1085	0.006	INF	5	0.044	2	3
NAD(P)H sulfur oxidoreductase (CoA-dependent)	TK1299	0.0062	15	5	0.0021	1.9	4
Translation initiation factor 2 subunit beta	TK1621	0.039	INF	4	0.023	6	4
Uncharacterized protein	TK0166	0.00023	INF	5			
Fibrillarin-like rRNA/tRNA 2'-O-methyltransferase	TK0183	0.031	11	4			
snoRNP component, Nop56p/58p homolog	TK0184	0.015	INF	4			
Predicted membrane protease subunit, stomatin/prohibitin homolog	TK0348	0.0096	INF	5			
AMP phosphorylase	TK0352	0.018	2	3			
Uncharacterized protein	TK0442	0.0096	INF	5			
Diadenylate cyclase	TK0510	0.00001	29	5			
Serine hydroxymethyltransferase	TK0528	0.024	INF	4			
DNA/RNA-binding protein Alba	TK0560	0.039	INF	4			
Thermosome subunit alpha	TK0678	0.0032	4.3	5			
Type 2 DNA topoisomerase 6 subunit A	TK0798	0.00023	INF	5			
Type 2 DNA topoisomerase 6 subunit B	TK0799	0.00033	13	5			
Uncharacterized protein	TK0900	0.039	INF	4			
Uncharacterized protein	TK0930	0.00001	INF	5			
Sugar-phosphate nucleotidyltransferase	TK0955	0.048	2.8	3			
N2, N2-dimethylguanosine tRNA methyltransferase	TK0981	0.0054	6.2	5			
Fructose-bisphosphate aldolase class 1	TK0989	0.015	INF	4			
Putative 5-methylcytosine restriction system, GTPase subunit	TK1009	0.031	11	4			
Uncharacterized protein	TK1025	0.00051	21	5			
30S ribosomal protein S7	TK1077	0.021	12	4			
Predicted AP endonuclease	TK1165	0.0037	5.8	5			
Phosphoenolpyruvate synthase	TK1292	0.039	INF	4			
Probable translation initiation factor IF-2	TK1305	0.00094	INF	5			
30S ribosomal protein S28e	TK1310	0.045	10	4			
DNA primase DnaG	TK1410	0.032	3	3			
50S ribosomal protein L12	TK1415	0.039	INF	4			
Cleavage and polyadenylation specificity factor subunit homolog	TK1428	0.031	11	4			
Probable tRNA pseudouridine synthase B	TK1509	0.014	13	4			

50S ribosomal protein L4	TK1541	0.006	INF	5			
ABC-type multidrug transport system, ATPase component	TK1579	0.0038	INF	5			
V-type ATP synthase alpha chain	TK1602	0.00001	INF	5			
V-type ATP synthase beta chain	TK1603	0.006	INF	5			
V-type ATP synthase subunit D	TK1604	0.024	INF	4			
Predicted metal-dependent hydrolase	TK1611	0.0096	INF	5			
Methylmalonyl-CoA decarboxylase, alpha subunit	TK1622	0.00001	INF	5			
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1800	0.006	INF	5			
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1801	0.00015	INF	5			
ABC-type dipeptide/oligopeptide transport system, permease component	TK1802	0.0096	INF	5			
ABC-type dipeptide/oligopeptide transport system, permease component	TK1803	0.006	INF	5			
ABC-type dipeptide/oligopeptide transport system, probable periplasmic component	TK1804	0.0011	11	5			
Type II/IV secretion system ATPase	TK1853	0.0056	3.5	4			
Probable lipoprotein releasing system, ATP-binding protein	TK1859	0.0093	14	5			
DNA polymerase II large subunit	TK1903	0.00001	INF	5			
Ferredoxin 3	TK2012	0.0038	INF	5			
ABC-type multidrug transport system, ATPase component	TK2053	0.039	INF	4			
Probable formate dehydrogenase, alpha subunit	TK2076	0.00017	2.8	4			
4Fe-4S cluster-binding protein	TK2078	0.0024	INF	5			
Membrane bound hydrogenase, NiFe-hydrogenase large subunit 2	TK2091	0.0062	15	5			
Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	TK2093	0.0038	INF	5			
Radical_SAM domain-containing protein	TK2160	0.039	INF	4			
cobW domain-containing protein	TK2200	0.018	4.1	4			
tRNA-splicing endonuclease	TK2215	0.024	INF	4			
Lysine--tRNA ligase	TK2240	0.015	INF	4			
Flagellin B5	TK0042				0.03	INF	4
Phosphoribosylaminoimidazole-succinocarboxamide synthase	TK0432				0.03	INF	3
Predicted ATP-dependent endonuclease, OLD family	TK0773				0.031	3.5	3
Glycogen synthase	TK1768				0.012	INF	3

Table 4.15: Co-purified interacting partners of RBR2 (TK0826) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
RecJ-like exonuclease, containing OB-fold nucleic acid-binding domains	TK0155	0.0049	28	5	0.014	4.6	4
Amidophosphoribosyltransferase	TK0211	0.014	9.1	4	0.00001	7.9	5
Metal-dependent phosphohydrolase, HD superfamily	TK0540	0.011	INF	4	0.0017	6.4	5
Predicted phosphohydrolase, DHH family	TK0645	0.044	INF	4	0.0081	INF	5
tRNA(Met) cytidine acetyltransferase TmcA	TK0754	0.0078	INF	5	0.035	4.6	4
Glycerol-1-phosphate dehydrogenase [NAD(P)+]	TK0789	0.031	INF	4	0.022	9.5	4
Type 2 DNA topoisomerase 6 subunit A	TK0798	0.015	10	4	0.0052	5.1	5
Rubrerythrin domain-containing protein	TK0826	0.00001	INF	5	0.00001	INF	5
Aldehyde ferredoxin oxidoreductase	TK0844	0.00001	45	5	0.016	8	4
Ribonuclease Z	TK1114	0.044	INF	4	0.00082	25	5
Phosphoenolpyruvate synthase	TK1292	0.022	INF	4	0.02	7.7	4
NAD(P)H sulfur oxidoreductase (CoA-dependent)	TK1299	0.00035	23	5	0.00019	8	5
Probable translation initiation factor IF-2	TK1305	0.00001	INF	5	0.0044	5	5
Uncharacterized protein	TK1329	0.016	INF	4	0.0027	INF	5
Translation initiation factor 2 subunit beta	TK1621	0.031	INF	4	0.011	17	4
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1684	0.031	INF	4	0.023	6.5	4
Vitamin B12-dependent ribonucleotide reductase	TK1736	0.00021	7.7	5	0.0014	3.4	4
Xanthine/guanine phosphoribosyltransferase	TK1737	0.044	INF	4	0.004	9.7	5
DNA repair and recombination protein RadA	TK1899	0.016	INF	4	0.0081	INF	5
DNA polymerase II large subunit	TK1903	0.00001	19	5	0.0024	3.8	4
tRNA-splicing endonuclease	TK2215	0.00017	INF	5	0.00018	7.1	5
Flagellin B5	TK0042	0.022	INF	4			
Fibrillarin-like rRNA/tRNA 2'-O-methyltransferase	TK0183	0.031	10	4			
snoRNP component, Nop56p/58p homolog	TK0184	0.041	12	4			
Pyridoxal 5'-phosphate synthase subunit PdxS	TK0217	0.0089	12	5			
Quinolinate synthase A	TK0296	0.031	INF	4			
AMP phosphorylase	TK0352	0.0011	2.6	4			
RNA-splicing ligase RtcB	TK0358	0.031	INF	4			
Uncharacterized protein	TK0440	0.03	20	3			
Uncharacterized protein	TK0453	0.0055	INF	5			
Uncharacterized protein	TK0483	0.031	INF	4			
Serine hydroxymethyltransferase	TK0528	0.00077	36	5			
Molybdenum cofactor biosynthesis protein B	TK0544	0.0033	2.4	4			
Proline--tRNA ligase	TK0550	0.022	INF	3			
GTP cyclohydrolase MptA	TK0793	0.031	INF	4			

Type 2 DNA topoisomerase 6 subunit B	TK0799	0.00083	10	5
Uncharacterized protein	TK0900	0.04	19	4
Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains	TK0953	0.0078	INF	5
Glycine--tRNA ligase	TK0978	0.04	19	4
Fructose-bisphosphate aldolase class 1	TK0989	0.032	13	4
Putative 5-methylcytosine restriction system, GTPase subunit	TK1009	0.039	9.8	4
Uncharacterized protein	TK1025	0.04	19	4
DNA-directed RNA polymerase subunit	TK1076	0.0027	31	5
DNA-directed RNA polymerase subunit beta	TK1083	0.0014	16	5
Protein disulfide oxidoreductase	TK1085	0.00098	INF	5
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1130	0.022	INF	4
Peptide chain release factor subunit 1	TK1239	0.022	INF	4
Nucleoside diphosphate kinase	TK1307	0.0039	INF	5
Ferredoxin:NADP oxidoreductase, beta subunit	TK1326	0.016	INF	4
Cell division protein FtsZ 1	TK1421	0.0039	INF	4
Cleavage and polyadenylation specficity factor subunit homolog	TK1428	0.0011	16	5
Leucine--tRNA ligase	TK1461	0.0039	INF	5
Isopentenyl-diphosphate delta-isomerase	TK1470	0.022	INF	4
V-type ATP synthase alpha chain	TK1602	0.012	24	4
Ribosome maturation protein SDO1 homolog	TK1636	0.0027	19	5
Ferredoxin:NADP oxidoreductase, beta subunit	TK1685	0.022	INF	4
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1800	0.04	19	4
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1801	0.04	19	4
ABC-type dipeptide/oligopeptide transport system, permease component	TK1803	0.022	INF	4
ABC-type dipeptide/oligopeptide transport system, probable periplasmic component	TK1804	0.04	7.7	4
Type II/IV secretion system ATPase	TK1853	0.024	5.3	4
Probable lipoprotein releasing system, ATP-binding protein	TK1859	0.0014	33	5
Predicted N6-adenine-specific DNA methylase	TK1863	0.044	INF	4
5-methylthioadenosine/S-adenosylhomocysteine deaminase	TK1891	0.011	INF	4
Iron-molybdenum cofactor-binding protein	TK2016	0.044	INF	4
Enolase	TK2106	0.022	21	4
Uncharacterized protein	TK2148	0.022	INF	3
Radical_SAM domain-containing protein	TK2160	0.011	INF	4
Tungsten-containing glyceraldehyde-3-phosphate:ferredoxin oxidoreductase	TK2163	0.022	INF	4

Lysine--tRNA ligase	TK2240	0.0024	15	5			
Anaerobic ribonucleoside-triphosphate reductase	TK2298	0.00024	INF	5			
Flagellin B3	TK0040				0.0056	INF	5
Sugar-phosphate nucleotidyltransferase	TK0219				0.029	4.8	4
Uncharacterized protein	TK0426				0.036	INF	4
Phosphoribosylaminoimidazole-succinocarboxamide synthase	TK0432				0.00001	INF	5
Predicted GTPase, containing TGS domain	TK0506				0.036	INF	4
Diadenylate cyclase	TK0510				0.035	3.4	3
Predicted ATP-dependent endonuclease, OLD family	TK0773				0.0038	12	5
Uncharacterized protein	TK0930				0.00001	32	5
Sugar-phosphate nucleotidyltransferase	TK0955				0.0033	3.5	4
Adenylosuccinate synthetase	TK1002				0.0091	8.7	5
Translation initiation factor 2 subunit alpha	TK1100				0.0068	11	5
Uncharacterized protein	TK1186				0.018	6.8	4
ATP-dependent DNA helicase Hel308	TK1332				0.025	INF	4
Probable glycine dehydrogenase (decarboxylating) subunit 2	TK1379				0.029	9	4
Uncharacterized protein	TK1394				0.0056	INF	5
Probable tRNA pseudouridine synthase B	TK1509				0.017	4.5	4
Uncharacterized protein	TK1545				0.036	INF	4
Predicted dehydrogenase	TK1557				0.00049	3.7	4
Glycogen synthase	TK1768				0.00001	INF	5
Translation initiation factor 2 subunit gamma	TK1946				0.041	3.6	3
Carboxymuconolactone decarboxylase-related protein	TK1974				0.017	INF	4
Cysteine desulfurase	TK1990				0.017	4.4	4
Ferredoxin 3	TK2012				0.014	4.5	4
Cytosolic NiFe-hydrogenase, gamma subunit	TK2071				0.022	15	4
Cytosolic NiFe-hydrogenase, beta subunit	TK2072				0.0083	18	5
Membrane bound hydrogenase, NiFe-hydrogenase small subunit	TK2089				0.014	4.6	4
Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	TK2093				0.014	4.5	4
Radical_SAM domain-containing protein	TK2145				0.036	INF	4
Non-specific serine/threonine protein kinase	TK2250				0.02	7.7	4
Cell division protein FtsZ 2	TK2271				0.034	7	4

While *in vitro* data describing the rubrerythrins in *T. kodakarensis* is lacking, studies from closely related archaeal species suggest these proteins function as homodimers.[44,58] Rubrerythrins are non-heme iron-containing proteins that aid in oxidative stress response in anaerobic organisms. Their association with all subunits of tagged-POR and tagged-VOR (Figure 4.16 and Figure 4.18) suggests that they may provide such protection by helping maintain a local environment that supports appropriate redox conditions.

Many associations (average ~65) were seen from both RBRs in +S˚ conditions, and while RBR2 (TK0826) displayed a similar abundance of associations in -S˚ (51 significant proteins), RBR1 displayed considerably less associations in -S˚ with only eleven proteins being identified as significant. The tagged-RBRs interact with many proteins of metabolic interest also tagged in this study. RBR1 was seen to interact with Fd3 and MBH in +S˚ while RBR2 was seen to interact with these same proteins in -S˚.

While the RBRs were initially identified in the data of multiple subunits of tagged-VOR and POR, these relationships were not identified reciprocally; no subunits of POR or VOR were identified when either RBR was tagged. Such data suggests that while there is clear evidence for the interaction between POR/VOR and RBRs, there may be additional proteins which aid in these associations which are unknown at this time.

While RBR3 was not tagged in this study, it was found in association with all three tagged subunits of OGOR. As RBR1 was identified by all tagged-subunits of VOR and RBR2 was identified with all tagged subunit of POR, each RBR may preferentially associate with distinct 2-ketoacid oxidoreductase complexes. Such a hypothesis warrants further investigations but is beyond the scope of this study.

Glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR)

In the classic EMP pathway, glyceraldehyde-3-phosphate (GAP) is reversibly converted to 3-phosphoglycerate (3PG) by phosphorylating GAP dehydrogenase (GAPDH) and

phosphoglycerate kinase (PGK).[45] In *T. kodakarensis*, a modified EMP pathway involves GAP conversion to 3PG by non-phosphorylating GAPDH (also known as GAPN) or GAP oxidoreductase (GAPOR).[46] While GAPN reduces NAD(P)$^+$ to NADPH, GAPOR reduces a Fd.[61] Expression of GAPOR was observed to be ~ five-fold higher in glycolytic conditions,[62] however cultures in this study were grown in media supplemented with pyruvate, thereby inducing gluconeogenic conditions. Considering the expression activity of GAPOR, additional trials of tagged-GAPOR and the parental control strain were evaluated in media without pyruvate supplementation ("non-pyruvate") in the presence of sulfur. While no Fds identified GAPOR as an interacting partner,[2] GAPOR is a monomeric 1-tungsten, 6-iron protein which demonstrated minimal activity in the presence of NAD(P)+ but clear activity in the presence of Fd.[61]

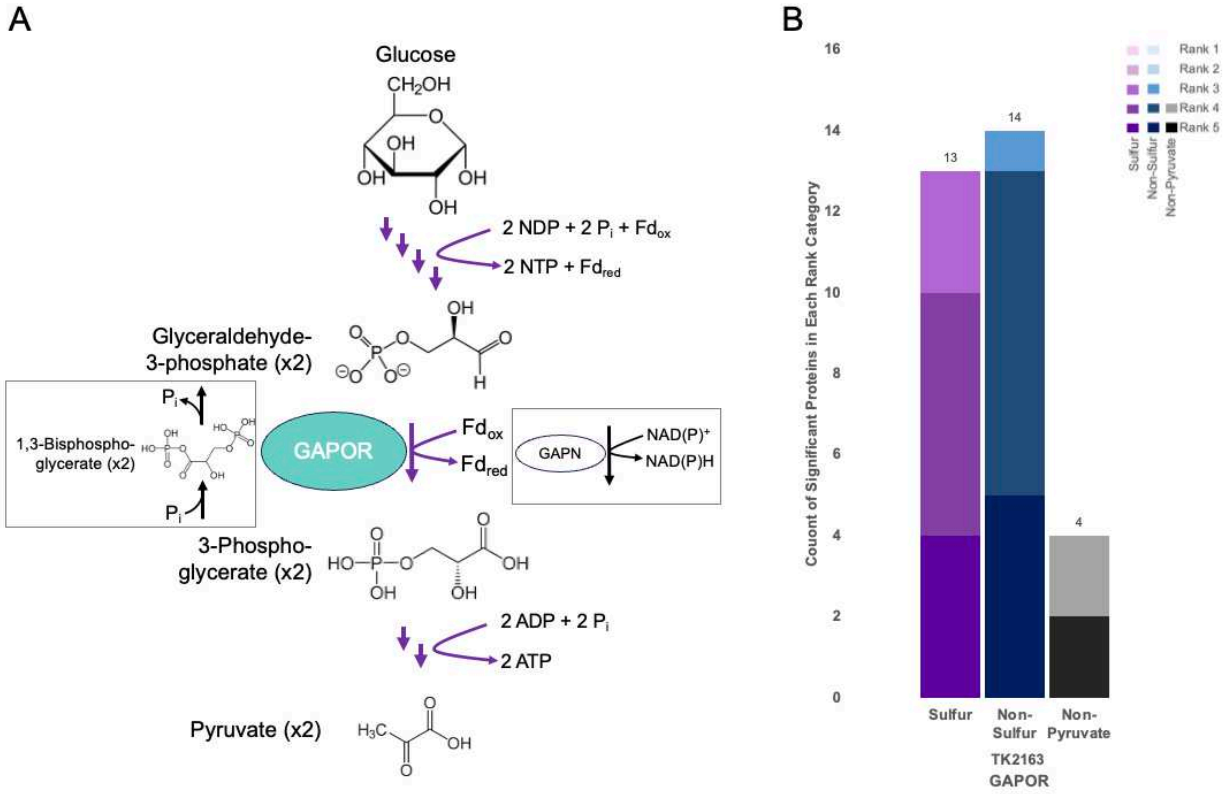


Figure 4.22: GAPOR function and number of interacting partners observed in sulfur and non-sulfur conditions. The modified EMP glycolytic pathway (A) in *T. kodakarensis* employs two enzymes which irreversibly convert glyceraldehyde-3-phosphate (GAP) to 3-phosphoglycerate (3PG) by the reduction of either a Fd by GAPOR (sea foam green) or NAD(P)$^+$ by GAPN (box at right). 3PG undergoes a two-step conversion to GAP via 1,3-bisphosphoglycerate (box at left). GAPOR is the only protein included in this study which functions in glycolysis. 3PG is converted to pyruvate and ATP is generated. GAPOR is only

expressed at low levels in gluconeogenic conditions which are induce with the supplementation of pyruvate. As the focus of this study was to recognize sulfur and non-sulfur metabolic networks, pyruvate was added to all media. (*T. kodakarensis* must be supplemented with sulfur pyruvate to encourage doubling times near ~1 hour.) However, assessing interactions of GAPOR in glycolysis was of interest, thus tagged-GAPOR was assessed in three conditions: sulfur and pyruvate ("sulfur"), pyruvate only ("non-sulfur"), and sulfur only ("non-pyruvate"). In (B), the number of interacting partners for tagged-GAPOR in each condition are shown. Overall, GAPOR showed less than average (~25) interactions in all conditions with the lowest number of interactions observed in glycolytic conditions (non-pyruvate) out of all conditions tested. The number of interactions that were observed with a rank score of 5 is shown in dark purple or dark blue, rank 4 is shown in purple or blue, and rank 3 is light purple or light blue. No interactions with a rank score of 2 or 1 were observed. GAPOR was the only protein included in this study that was affinity purified using Cu^{2+} instead of Ni^{2+} which is explained in detail in Chapter 3.2.2.

GAPOR is a ~74 kDa monomer and the only protein included in this study involved in glycolysis. *T. kodakarensis* metabolizes complex and simple sugars, but monosaccharides do not support growth.[61] Optimal growth (doubling time of ~40 minutes) is seen with the addition of pyruvate and inorganic sulfur (as the preferred terminal electron acceptor), and moderate growth rates (doubling times of ~60 minutes) are observed with the supplementation of only pyruvate or inorganic sulfur.[63] Pyruvate supplementation supports gluconeogenesis, but GAPOR is lowly expressed in gluconeogenesis as it catalyzes the directional conversion of GAP to 3PG. As routing electron flux in the presence and absence of sulfur was a primary focus of this study, and *T. kodakarensis* must be supplemented with pyruvate or sulfur, non-sulfur growth trials required the addition of pyruvate. In a desire to compare sulfur and non-sulfur conditions consistently, pyruvate was added to all growth media with the exception of the trial for GAPOR. Because GAPOR is lowly expressed in gluconeogenesis, it was considered important to identify interacting partners of GAPOR in glycolytic conditions, thus GAPOR interacting partners were tested in the presence of sulfur and pyruvate (+S˚+P), pyruvate only (-S˚, non-sulfur), and sulfur only (+S˚-P, non-pyruvate). The control strain was similarly grown in the appropriate, respective, media.

Tagged-GAPOR identified fewer partners on average than the majority of tagged proteins: thirteen partners in +S˚+P, fourteen in -S˚+P, and only four partners in +S˚-P (Figure

4.22B, Figure 4.23). While low expression in gluconeogenic conditions could be a reason for low numbers of interacting partners, even fewer interacting partners (four) were observed in glycolytic conditions when GAPOR is expressed at higher levels.

The lack of interacting partners identified in glycolytic conditions is surprising. One could presume that GAPOR may require space to function, such that interacting with Additionally, while this Fd-dependent GAPOR was not identified when each of the Fds were tagged,[2] consistently Fds were not identified in this study when GAPOR was tagged.

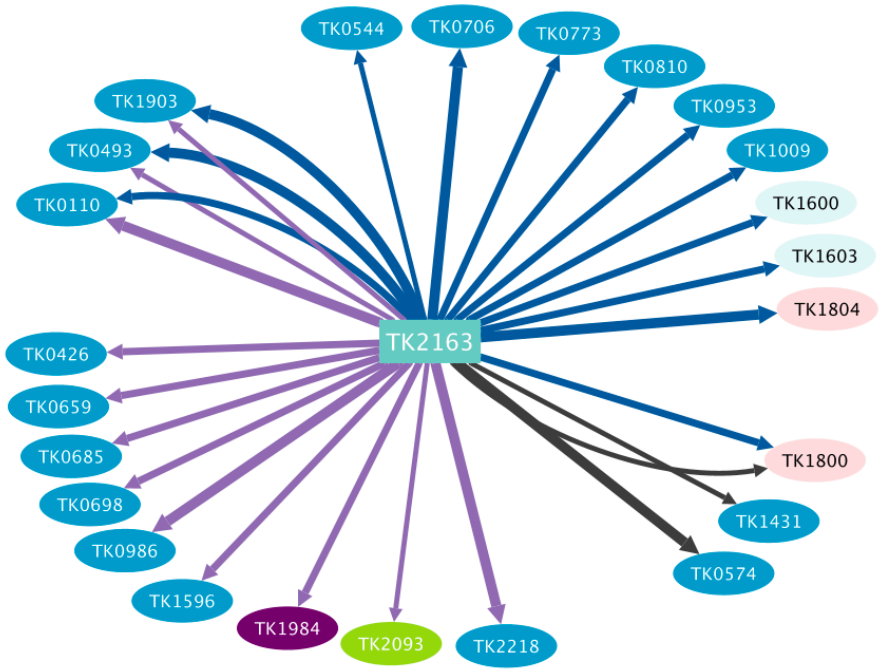


Figure 4.23: Interactome of GAPOR in sulfur, non-sulfur, and non-pyruvate conditions. AP-MS data from GAPOR identified interacting partners in sulfur (purple lines), non-sulfur (blue lines), and non-pyruvate (dark grey lines) conditions. Line thickness corresponds to the rank score of the interaction with thicker lines representing a higher rank score. GAPOR identified three proteins in both +S˚ and -S˚ conditions: a HAD family hydrolase (TK0110), a hypothetical membrane protein (TK0493), and the DNA polymerase II large subunit (TK1903). GAPOR identified only one similar protein in -S˚ and -P conditions: an ABC-type dipeptide/oligopeptide transport system component (light pink ellipse, TK1800) of which another (TK1804) was identified in only -S˚ conditions. Two components of the ATP synthase (light blue ellipses, TK1600 and TK1603) were identified in -S˚ conditions, along with a molybdenum cofactor biosynthesis protein (TK0544) and an oligosaccharyl transferase (TK0810). In glycolytic conditions (-P), a glutamate dehydrogenase (TK1431) and a metallophosphoesterase (TK0574) were also identified. GAPOR was found to associate with PORβ (purple ellipse, TK1984) and MBH-N (lime green ellipse, TK2093) in +S˚ conditions as well as a proton/glutamate symporter (blue ellipse, TK0986).

Table 4.16: Co-purified interacting partners of GAPOR (TK2163) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur			Non-Pyruvate		
Identified Proteins	**Gene No.**	**P-Value**	**FC**	**Rank**	**P-Value**	**FC**	**Rank**	**P-Value**	**FC**	**Rank**
Tungsten-containing glyceraldehyde-3-phosphate:ferredoxin oxidoreductase	TK2163	0.00001	INF	5	0.00001	INF	5	0.00001	INF	5
Hydrolase, HAD superfamily	TK0110	0.0027	INF	5	0.032	INF	4			
Hypothetical membrane protein, conserved, containing DUF11 domain	TK0493	0.014	3.4	3	0.0029	INF	5			
DNA polymerase II large subunit	TK1903	0.023	3.7	3	0.00053	INF	5			
Uncharacterized protein	TK0426	0.021	INF	4						
ABC-type transport system, permease component	TK0659	0.032	6.7	4						
NAD-dependent protein deacylase	TK0685	0.021	7.3	4						
Zinc-dependent protease, TldD/PmbA family	TK0698	0.016	INF	4						
Proton/glutamate symporter, SDF family	TK0986	0.0097	INF	5						
Archaeal/vacuolar-type H+-ATPase, subunit H	TK1596	0.045	INF	4						
Pyruvate:ferredoxin oxidoreductase, beta subunit	TK1984	0.026	7	4						
Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	TK2093	0.047	3.2	3						
Replication factor C small subunit	TK2218	0.0093	5.3	5						
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1800				0.032	7	4	0.017	2	3
Molybdenum cofactor biosynthesis protein B	TK0544				0.044	2.5	3			
ABC-type iron(III)-siderophore transport system, periplasmic component fused to N-terminal uncharacterized domain	TK0706				0.0016	28	5			
Predicted ATP-dependent endonuclease, OLD family	TK0773				0.012	INF	4			
Oligosaccharyl transferase	TK0810				0.048	INF	4			
Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains	TK0953				0.022	18	4			
Putative 5-methylcytosine restriction system, GTPase subunit	TK1009				0.017	19	4			
V-type ATP synthase subunit C	TK1600				0.014	INF	4			
V-type ATP synthase beta chain	TK1603				0.032	7	4			
ABC-type dipeptide/oligopeptide transport system, probable periplasmic component	TK1804				0.0044	7.7	5			
Metallophosphoesterase, calcineurin superfamily	TK0574							0.0056	INF	5
Glutamate dehydrogenase	TK1431							0.032	2.3	3

Formate dehydrogenase (FDH)

The breakdown of amino acids or pyruvate generates an acyl-CoA (or acetyl-CoA, in the case of pyruvate catabolism).[41] With the addition of CO_2, such compounds can be converted to acids (such as acetate) by Acyl-CoA Synthase (ACS) resulting in the generation of 1 ATP.[64] Alternatively, formate dehydrogenase (FDH) converts acyl-CoA and formate into CO_2, generating a reduced cofactor (NAD(P)H or Fd) in the process (Figure 4.24).[47,65] When formate oxidation is coupled with a hydrogenase and H_2 generation, members of the *Thermococcales* have been observed to exhibit the rare feat of slow growth on formate alone.[13,47] While FDH is of great interest and the diversity of metal-dependent FDH complexes is wide ranging,[65] the specific structure of the tungsten-containing enzymatic complex in *T. kodakarensis* has yet to be determined.

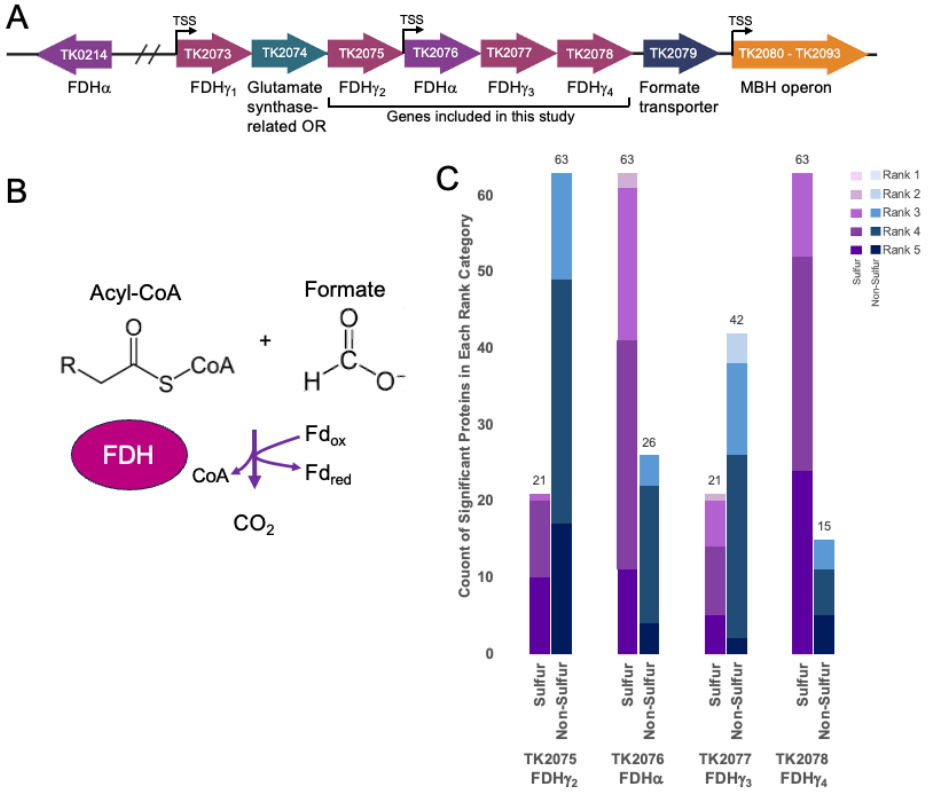


Figure 4.24: FDH genomic organization, function, and number of interacting partners observed in sulfur and non-sulfur conditions. FDH is not well characterized in *T. kodakarensis*. Two FDHα subunits are encoded in the genome at TK0214 and TK2076 (purple) (A), while only FDHα encoded by TK2076 is surrounded by FDH-related subunits (maroon and dark blue) and preceding the MBH operon (orange) as seen in other closely related

Thermococcales.[49] In *T. kodakarensis*, two 4Fe-4S cluster binding proteins of ~ 18 kDa (maroon, FDH$\gamma_{1\text{-}4}$) are encoded both upstream and downstream of FDHα, followed by a formate transporter (navy blue, TK2079), and the MBH operon (orange, TK2080 – TK2093). Four FDH-related proteins were tagged in this study: FDHγ_2, FDHα, FDHγ_3, and FDHγ_4 (TK2075 - TK2078). FDH converts formate and acyl-CoA to CO_2 therein reducing a Fd (Fd_{red}) (B). Of the tagged-FDH-related subunits included in this study, TK2076 and TK2078 displayed greater than average interactions in +S° conditions while TK2075 and TK2077 displayed greater than average interactions in -S° conditions (C). The number of interactions that were observed with a rank score of 5 is shown in dark purple or dark blue, rank 4 is shown in purple or blue, rank 3 is light purple or light blue, and rank 2 is faint purple or faint blue. No interactions with a rank score of 1 were observed.

T. kodakarensis encodes two FDHα subunits (TK0214 and TK2076). While the genomic region of TK0214 is bracketed by two hypothetical proteins, the genomic region of TK2076 is bracketed by multiple 4Fe-4S cluster binding proteins that are likely FDH-related genes (TK2073 - TK2079) (Figure 4.24A).[49] Thus, TK2076 and select surrounding genes were included in this study, while TK0214 was not.

FDHα subunits are classically 70-80 kDa, while β subunits are 20-35 kDa, and γ subunits are 12-18 kDa.[49] While TK2073 is ~14 kDa, the remaining 4Fe-4S cluster binding subunits surrounding TK2076 are ~18 kDa, thus they will be considered FDH$\gamma_{1\text{-}4}$ subunits (TK2073, TK2075, TK2077, and TK2078). TK2074 is annotated as a glutamate synthase related oxidoreductase and while some studies report this gene to encode a protein homologous to FDHβ,[49] at ~39 kDa, it is too large to be classified as a β subunit by other reports.[65]

The TK2073 - TK2078 gene cluster sits downstream of a formate transporter (TK2079) and the MBH operon (TK2080 - TK2093) (Figure 4.24A). While microbial operon prediction suggests that a single transcription start site (TSS) may regulate expression of genes TK2073 – TK2079,[60] other sources (including a transcriptomic analysis of *T. kodakarensis*) suggest that a second TSS occurs directly upstream of TK2076 (Figure 4.24A).[31,59] Transcription of FDHγ_1, the glutamate synthase related oxidoreductase, and FDHγ_2 is initiated by a promoter which sits upstream of FDHγ_1.

A closely related yet genetically inaccessible hyperthermophile, *T. onnurineus*, maintains three FDHs.[49] *To*-FDH is thought to be composed of five genes which sit upstream of a formate transporter. Homologs in *T. kodakarensis* include three FDH$\gamma_{1\text{-}4}$ (TK2073, TK2075, TK2077, and TK2078), TK2074 (which is annotated as a glutamate synthase related oxidoreductase but homologous to *To*-FDH3β), and FDHα (TK2076).

FDHα (TK2076) and the surrounding FDH$\gamma_{2\text{-}4}$ (TK2075 – TK2078) were tagged and analyzed via AP-MS. Two subunits (TK2076 and TK2078) displayed greater numbers of interacting partners (average ~sixty) in +S˚ while the other two subunits (TK2075 and TK2077) displayed greater numbers (average ~fifty) in -S˚ (Figure 4.24C). The respective subunits showed low numbers of associations in the inverse condition, averaging ~twenty associations (Figure 4.24C and Figure 4.25). While TK2075, TK2077, and TK2078 identified other FDH subunit components, FDHα (TK2076) did not suggesting that FHDα may only associate with the other 4Fe-4S cluster binding subunits loosely.

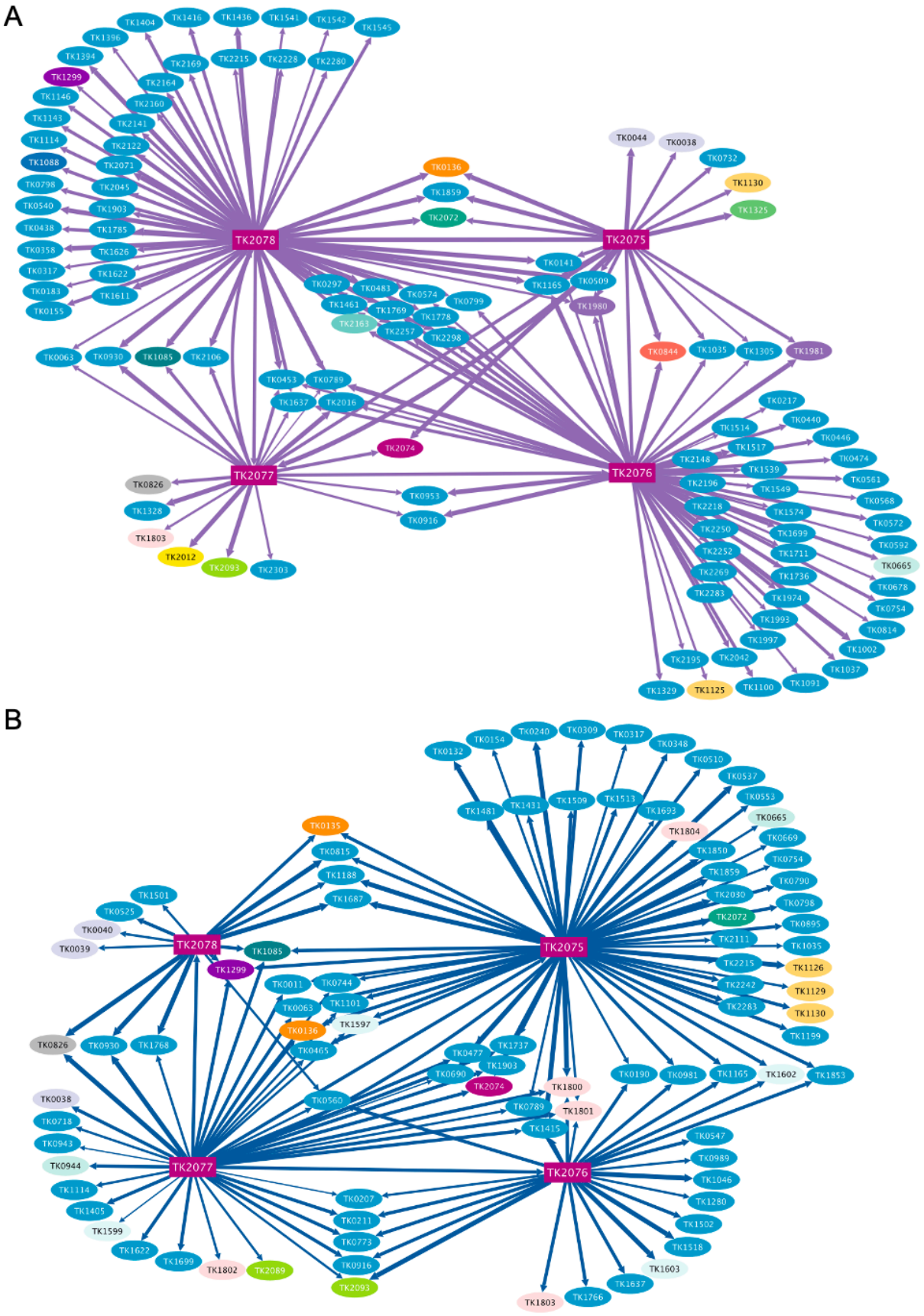

A
TK2078
TK2075
TK2077
TK2076
TK1404
TK1416
TK1436
TK1541
TK1542
TK1545
TK1396
TK1394
TK1299
TK1146
TK1143
TK1114
TK1088
TK0798
TK0540
TK0438
TK0358
TK0317
TK0183
TK0155
TK2169
TK2215
TK2228
TK2280
TK2164
TK2160
TK2141
TK2122
TK2071
TK2045
TK1903
TK1785
TK1626
TK1622
TK1611
TK0136
TK1859
TK2072
TK0044
TK0038
TK0732
TK1130
TK1325
TK0141
TK0509
TK1165
TK1980
TK0297
TK0483
TK0574
TK0799
TK1461
TK1769
TK1778
TK2163
TK2257
TK2298
TK0063
TK0930
TK1085
TK2106
TK0453
TK0789
TK1637
TK2016
TK2074
TK0844
TK1035
TK1305
TK1981
TK0826
TK1328
TK1803
TK2012
TK2093
TK2303
TK0953
TK0916
TK0217
TK0440
TK0446
TK0474
TK0561
TK0568
TK0572
TK0592
TK0665
TK0678
TK0754
TK0814
TK1002
TK1037
TK1091
TK1514
TK1517
TK2148
TK1539
TK2196
TK1549
TK2218
TK1574
TK2250
TK1699
TK2252
TK1711
TK2269
TK1736
TK2283
TK1974
TK1993
TK1997
TK2195
TK2042
TK1329
TK1125
TK1100
B
TK0132
TK0154
TK0240
TK0309
TK0317
TK0348
TK0510
TK0537
TK0553
TK0665
TK0669
TK0754
TK0790
TK0798
TK0895
TK1035
TK1126
TK1129
TK1130
TK1199
TK1481
TK1431
TK1509
TK1513
TK1693
TK1804
TK1850
TK1859
TK2030
TK2072
TK2111
TK2215
TK2242
TK2283
TK1853
TK1602
TK1165
TK0981
TK0190
TK0135
TK0815
TK1188
TK1687
TK1501
TK0525
TK0040
TK0039
TK1085
TK1299
TK0011
TK0744
TK0063
TK1101
TK1597
TK0136
TK0465
TK0826
TK0930
TK1768
TK0477
TK1737
TK1903
TK0690
TK2074
TK1800
TK0789
TK1801
TK1415
TK0560
TK0038
TK0718
TK0943
TK0944
TK1114
TK1405
TK1599
TK1622
TK1699
TK1802
TK2089
TK0207
TK0211
TK0773
TK0916
TK2093
TK0547
TK0989
TK1046
TK1280
TK1502
TK1518
TK1603
TK1637
TK1766
TK1803

Figure 4.25: Interactome of formate dehydrogenase in sulfur and non-sulfur conditions. Four subunits related to formate metabolism were tagged including three 4Fe-4S cluster binding proteins (FDH$\gamma_{2\text{-}4}$: TK2075, TK2077, and TK2078) and FDHα (TK2076). Significant associations identified from AP-MS data were found in sulfur (A) and non-sulfur (B) conditions and are displayed with line thickness relating to a confidence rank score: a thicker line relates to a rank score of 5, representing the highest confidence in the association. No proteins were identified by all four tagged FDH-related proteins in +S˚ or -S˚. TK2077 and TK2078 identified one another in +S˚, as did TK2077 and TK2075. TK2075 identified TK2078 and TK2076, but these associations were not observed reciprocally. Both TK2077 and TK2075 were found to associate with TK2074, a glutamate synthase related oxidoreductase encoded upstream of the genes encoding the FDH-operon. While TK2073 (FDHγ_1, is reported as a 4Fe-4S cluster binding protein related to FDH, it was not identified as significant with any tagged FDH-related protein in either condition. TK2075 and TK2077 were seen to associate with more proteins in -S˚ than in +S˚, while TK2076 and TK2078 showed the opposite. TK2076 and TK2078 identified eighteen similar interacting partners in +S˚ but only one in -S˚. IORα (orange ellipse, TK0136) was identified by TK2075 and TK2078 in +S˚ while both IOR subunits were identified in -S˚. TK2075 identified one subunit of OGOR (light yellow ellipse, TK1130) and TK2076 identified another subunit of OGOR (light yellow ellipse, TK1125) in +S˚ and TK2075 identified three subunits (TK1126, TK1129, and TK1130) in -S˚. The NAD(P)H sulfur oxidoreductase (bright purple ellipse, TK1299) was identified by TK2078 in +S˚ and TK2075, TK2077, and TK2078 in -S˚. TK2077 was seen to associate with one subunit of MBH (TK2093) +S˚ and one subunit (TK2089) in -S˚, while both TK2077 and TK2076 identified TK2093 in -S˚. Two and three flagellin proteins (light grey ellipses, TK0038 - TK0044) were observed in +S˚ and -S˚ conditions, respectively. Five subunits of the ABC-type dipeptide/oligopeptide transport system (light pink ellipses, TK1800 - TK1804) were observed in -S˚, while only one subunit was observed in +S˚. GAPOR (light blue/green ellipse, TK2163) was identified by two FDH-related proteins in +S˚ conditions. Line thickness represents the rank score observed for the interaction with thicker lines representing greater rank scores.

Table 4.17: Co-purified interacting partners of FDHγ_2 (TK2075) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Indolepyruvate oxidoreductase subunit IorA	TK0136	0.00045	14	5	0.00001	3.7	4
Predicted glutamine amidotransferase, class II	TK1035	0.047	11	4	0.024	2.7	3
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1130	0.033	12	4	0.00099	3.9	4
Predicted AP endonuclease	TK1165	0.00001	5.5	5	0.00001	3.3	4
Probable lipoprotein releasing system, ATP-binding protein	TK1859	0.0019	12	5	0.00001	4.2	5
Cytosolic NiFe-hydrogenase, beta subunit	TK2072	0.0098	INF	4	0.0025	4	5
Glutamate synthase beta chain-related oxidoreductase	TK2074	0.00001	68	5	0.00001	20	5
4Fe-4S cluster-binding protein	TK2075	0.00001	45	5	0.00001	18	5
4Fe-4S cluster-binding protein	TK2077	0.00001	4	5	0.00036	2.1	4
Flagellin B1	TK0038	0.00092	13	4			
Archaeal flagella-related protein D, internal insertion	TK0044	0.00001	INF	5			
UPF0173 metal-dependent hydrolase TK0141	TK0141	0.027	4	4			
Flavin prenyltransferase UbiX	TK0509	0.033	12	4			
Iron-molybdenum cofactor-binding protein	TK0732	0.023	INF	4			
Aldehyde ferredoxin oxidoreductase	TK0844	0.0085	7.7	5			
Probable translation initiation factor IF-2	TK1305	0.029	3.3	3			
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1325	0.0028	INF	5			
2-oxoisovalerate:ferredoxin oxidoreductase, alpha subunit	TK1980	0.021	6.7	4			
2-oxoisovalerate:ferredoxin oxidoreductase, beta subunit	TK1981	0.035	INF	4			
Probable formate dehydrogenase, alpha subunit	TK2076	0.00001	3.6	4			
4Fe-4S cluster-binding protein	TK2078	0.0094	5	5			
Uncharacterized protein	TK0011				0.0013	2.1	4
Nucleotidyltransferase, fused to N-terminal DNA-binding domain	TK0063				0.002	3.5	4
TATA-box-binding protein	TK0132				0.00025	INF	5
Indolepyruvate oxidoreductase subunit IorB	TK0135				0.00099	3.9	4
Uncharacterized protein	TK0154				0.049	INF	3
GMP synthase [glutamine-hydrolyzing] subunit A	TK0190				0.014	2.7	3
Arginase	TK0240				0.0011	4.7	5
Elongation factor 2	TK0309				0.0065	3.6	4
TatD-related deoxyribonuclease	TK0317				0.023	INF	3
Predicted membrane protease subunit, stomatin/prohibitin homolog	TK0348				0.049	4	4
Acetyl-CoA synthetase I (NDP forming), beta subunit	TK0465				0.049	INF	4
Glyceraldehyde 3-phosphate phosphatase	TK0477				0.00037	2.5	4
Diadenylate cyclase	TK0510				0.004	2	4

Peroxiredoxin	TK0537		0.00001	6	5
tRNA(Ile2) 2-agmatinylcytidine synthetase TiaS	TK0553		0.00016	3.7	4
Acetate--CoA ligase (ADP-forming)	TK0665		0.014	2.7	3
CDC48/VCP homolog, AAA superfamily	TK0669		0.032	2.7	3
Molybdopterin oxidoreductase, molybdopterin-binding subunit	TK0690		0.00098	7.1	5
Uncharacterized protein	TK0744		0.017	3.7	3
tRNA(Met) cytidine acetyltransferase TmcA	TK0754		0.019	2	3
Glycerol-1-phosphate dehydrogenase [NAD(P)+]	TK0789		0.014	3	3
UPF0179 protein TK0790	TK0790		0.006	2.2	4
Type 2 DNA topoisomerase 6 subunit A	TK0798		0.0024	2.1	4
Peroxiredoxin, AhpC/TSA family	TK0815		0.049	INF	4
S-layer protein	TK0895		0.00001	2.4	4
N2, N2-dimethylguanosine tRNA methyltransferase	TK0981		0.00001	2.9	4
Protein disulfide oxidoreductase	TK1085		0.00001	3.8	4
Ribosome biogenesis protein Nop10	TK1101		0.045	8	4
2-oxoacid:ferredoxin oxidoreductases, gamma subunit	TK1126		0.0011	INF	5
2-oxoacid:ferredoxin oxidoreductases, beta subunit	TK1129		0.033	3.1	3
Glucosamine-1-phosphate N-acetyltransferase	TK1188		0.00098	7.1	5
DNA helicase	TK1199		0.029	4.4	4
NAD(P)H sulfur oxidoreductase (CoA-dependent)	TK1299		0.00001	3.5	4
50S ribosomal protein L12	TK1415		0.00056	4.6	5
Glutamate dehydrogenase	TK1431		0.00001	2.2	4
NADH:polysulfide oxidoreductase	TK1481		0.00001	4.1	5
Probable tRNA pseudouridine synthase B	TK1509		0.00065	2.3	4
50S ribosomal protein L14e	TK1513		0.029	4.4	4
V-type ATP synthase subunit I	TK1597		0.049	INF	4
V-type ATP synthase alpha chain	TK1602		0.0065	3.6	4
Cysteine synthase	TK1687		0.00001	48	5
FeS_assembly_P domain-containing protein	TK1693		0.049	INF	3
Xanthine/guanine phosphoribosyltransferase	TK1737		0.0065	12	5
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1800		0.0032	6.2	5
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1801		0.037	3.7	3
ABC-type dipeptide/oligopeptide transport system, probable periplasmic component	TK1804		0.044	2	3
Peptidyl-prolyl cis-trans isomerase	TK1850		0.00012	INF	5
Type II/IV secretion system ATPase	TK1853		0.00001	3.2	4
DNA polymerase II large subunit	TK1903		0.00001	2.4	4

Uncharacterized protein	TK2030		0.0051	INF	5
dITP/XTP pyrophosphatase	TK2111		0.00001	INF	3
tRNA-splicing endonuclease	TK2215		0.00039	2.3	4
Beta-ribofuranosylaminobenzene 5'-phosphate synthase	TK2242		0.023	INF	4
Uncharacterized protein	TK2283		0.012	6.7	4

Table 4.18: Co-purified interacting partners of FDHα (TK2076) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Glycerol-1-phosphate dehydrogenase [NAD(P)+]	TK0789	0.0083	10	5	0.002	3	4
L-threonine 3-dehydrogenase	TK0916	0.0032	5.4	5	0.00041	2.4	4
Predicted AP endonuclease	TK1165	0.00001	4	5	0.00001	2.1	4
Proteasome subunit alpha	TK1637	0.032	6.3	4	0.0007	3	4
Probable formate dehydrogenase, alpha subunit	TK2076	0.00001	4.5	5	0.00001	2.4	4
UPF0173 metal-dependent hydrolase TK0141	TK0141	0.023	3.9	3			
Pyridoxal 5'-phosphate synthase subunit PdxS	TK0217	0.048	3.2	3			
L-aspartate oxidase	TK0297	0.011	6.5	4			
Uncharacterized protein	TK0440	0.042	5.2	4			
PDDEXK_1 domain-containing protein	TK0446	0.011	INF	4			
Uncharacterized protein	TK0453	0.025	5.7	3			
Arginase	TK0474	0.0029	19	4			
Uncharacterized protein	TK0483	0.029	3.2	3			
Flavin prenyltransferase UbiX	TK0509	0.00066	23	4			
Adenylosuccinate lyase	TK0561	0.029	4.4	4			
Tyrosine--tRNA ligase	TK0568	0.016	9	3			
ABC-type iron(III) transport system, ATPase component	TK0572	0.041	7.5	4			
Metallophosphoesterase, calcineurin superfamily	TK0574	0.024	INF	3			
Uncharacterized protein	TK0592	0.037	INF	3			
Acetate--CoA ligase (ADP-forming)	TK0665	0.018	14	4			
Thermosome subunit alpha	TK0678	0.021	2.8	3			
tRNA(Met) cytidine acetyltransferase TmcA	TK0754	0.014	4.6	4			
Type 2 DNA topoisomerase 6 subunit B	TK0799	0.00096	3.4	4			
Type A flavoprotein	TK0814	0.041	2.8	3			
Aldehyde ferredoxin oxidoreductase	TK0844	0.0049	6.4	5			
Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains	TK0953	0.0028	6.8	5			
Adenylosuccinate synthetase	TK1002	0.00022	6.6	5			

Predicted glutamine amidotransferase, class II	TK1035	0.012	15	4			
Metallophosphoesterase, calcineurin superfamily	TK1037	0.018	14	4			
DNA topoisomerase 1	TK1091	0.025	13	3			
Translation initiation factor 2 subunit alpha	TK1100	0.024	5.2	4			
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1125	0.036	12	3			
Probable translation initiation factor IF-2	TK1305	0.021	3.2	3			
Uncharacterized protein	TK1329	0.013	4.4	4			
Leucine--tRNA ligase	TK1461	0.04	4.8	4			
Cytidylate kinase	TK1514	0.037	INF	2			
Adenylate kinase	TK1517	0.0051	5.8	5			
50S ribosomal protein L2	TK1539	0.032	5.5	4			
Predicetd ATPase	TK1549	0.017	2.8	3			
tRNA uridine(34) acetyltransferase	TK1574	0.036	12	3			
DNA-directed RNA polymerase subunit A"	TK1699	0.022	8.5	4			
Sugar-phosphate nucleotydyltransferase	TK1711	0.011	INF	3			
Vitamin B12-dependent ribonucleotide reductase	TK1736	0.00001	3.3	4			
Predicted transcription regulator, DUF118 helix-turn-helix family	TK1769	0.00059	INF	4			
Predicted hydrolase, metallo-beta-lactamase superfamily	TK1778	0.0011	13	5			
Carboxymuconolactone decarboxylase-related protein	TK1974	0.016	INF	4			
2-oxoisovalerate:ferredoxin oxidoreductase, alpha subunit	TK1980	0.0084	3.3	4			
2-oxoisovalerate:ferredoxin oxidoreductase, beta subunit	TK1981	0.00038	6.4	5			
Hydrogenase expression/formation protein HypE	TK1993	0.037	INF	3			
Carbamoyltransferase	TK1997	0.016	9	3			
Iron-molybdenum cofactor-binding protein	TK2016	0.037	INF	3			
ATPase, RecA superfamily	TK2042	0.011	6.5	4			
Uncharacterized protein	TK2148	0.041	7.5	4			
Tungsten-containing glyceraldehyde-3-phosphate:ferredoxin oxidoreductase	TK2163	0.0031	INF	4			
Aspartate carbamoyltransferase regulatory chain	TK2195	0.024	INF	3			
Aspartate carbamoyltransferase	TK2196	0.018	3.7	3			
Replication factor C small subunit	TK2218	0.0021	7	4			
Non-specific serine/threonine protein kinase	TK2250	0.0014	3.7	4			
Proteasome-activating nucleotidase	TK2252	0.0045	7.2	5			
Deoxycytidylate deaminase	TK2257	0.025	13	4			
Uncharacterized protein	TK2269	0.037	INF	2			
Uncharacterized protein	TK2283	0.036	12	4			
Anaerobic ribonucleoside-triphosphate reductase	TK2298	0.049	4.3	4			
GMP synthase [glutamine-hydrolyzing] subunit A	TK0190				0.0044	2.5	4
Formate-dependent phosphoribosylglycinamide formyltransferase	TK0207				0.0043	8	3

Amidophosphoribosyltransferase	TK0211	0.0037	2	4
Metallophosphoesterase, calcineurin superfamily	TK0547	0.009	3.3	4
DNA/RNA-binding protein Alba	TK0560	0.009	3.3	4
Predicted ATP-dependent endonuclease, OLD family	TK0773	0.00028	3.6	4
N2, N2-dimethylguanosine tRNA methyltransferase	TK0981	0.00001	2	4
Fructose-bisphosphate aldolase class 1	TK0989	0.011	2.8	3
Uncharacterized protein	TK1046	0.0068	4.5	5
Transcription initiation factor IIB 1	TK1280	0.028	3.5	3
50S ribosomal protein L12	TK1415	0.0024	2.7	4
50S ribosomal protein L18e	TK1502	0.044	5	4
Protein translocase subunit SecY	TK1518	0.00001	INF	5
V-type ATP synthase alpha chain	TK1602	0.0046	3.7	4
V-type ATP synthase beta chain	TK1603	0.00038	11	5
Radical_SAM domain-containing protein	TK1766	0.0046	3.7	4
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1800	0.003	3.2	4
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1801	0.012	2.4	3
ABC-type dipeptide/oligopeptide transport system, permease component	TK1803	0.011	INF	4
Type II/IV secretion system ATPase	TK1853	0.00001	2.2	4
Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	TK2093	0.0096	7	5

Table 4.19: Co-purified interacting partners of FDHγ_3 (TK2077) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Nucleotidyltransferase, fused to N-terminal DNA-binding domain	TK0063	0.016	2.3	3	0.0027	1.5	4
Glycerol-1-phosphate dehydrogenase [NAD(P)+]	TK0789	0.042	3.5	2	0.021	1.3	3
Rubrerythrin domain-containing protein	TK0826	0.00001	3.8	4	0.00001	INF	5
L-threonine 3-dehydrogenase	TK0916	0.022	1.8	3	0.002	1.4	4
Uncharacterized protein	TK0930	0.00001	8.6	5	0.00001	10	5
Protein disulfide oxidoreductase	TK1085	0.0093	2.5	4	0.00001	1.7	4
Glutamate synthase beta chain-related oxidoreductase	TK2074	0.00038	2.2	4	0.00001	1.8	4
4Fe-4S cluster-binding protein	TK2075	0.00093	2.8	4	0.00013	1.6	4
Probable formate dehydrogenase, alpha subunit	TK2076	0.00001	1.9	4	0.00001	1.3	4
4Fe-4S cluster-binding protein	TK2077	0.00047	1.5	4	0.00001	1.6	4
4Fe-4S cluster-binding protein	TK2078	0.00001	3.8	4	0.00001	1.6	4
Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	TK2093	0.00001	4.7	5	0.012	1.4	3

Uncharacterized protein	TK0453	0.022	4	4			
Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains	TK0953	0.026	2.2	3			
tRNA (1-methyladenosine) methyltransferase	TK1328	0.0035	9	5			
Proteasome subunit alpha	TK1637	0.016	3.3	3			
ABC-type dipeptide/oligopeptide transport system, permease component	TK1803	0.042	3.5	3			
Ferredoxin 3	TK2012	0.00001	4.7	5			
Iron-molybdenum cofactor-binding protein	TK2016	0.003	INF	5			
Enolase	TK2106	0.015	7	4			
Thermosome subunit beta	TK2303	0.011	1.3	3			
Uncharacterized protein	TK0011				0.00001	1.3	4
Flagellin B1	TK0038				0.00036	9	4
Indolepyruvate oxidoreductase subunit IorA	TK0136				0.00001	1.4	4
Formate-dependent phosphoribosylglycinamide formyltransferase	TK0207				0.02	3	2
Amidophosphoribosyltransferase	TK0211				0.00001	1.9	4
Acetyl-CoA synthetase I (NDP forming), beta subunit	TK0465				0.012	INF	3
Glyceraldehyde 3-phosphate phosphatase	TK0477				0.00001	1.3	4
DNA/RNA-binding protein Alba	TK0560				0.02	2.3	3
Molybdopterin oxidoreductase, molybdopterin-binding subunit	TK0690				0.043	2	2
ABC-type molybdate transport system, permease component	TK0718				0.012	INF	3
Uncharacterized protein	TK0744				0.013	1.5	3
Predicted ATP-dependent endonuclease, OLD family	TK0773				0.00001	3	4
Acetyl-CoA synthetase II (NDP forming), beta subunit	TK0943				0.045	2.5	2
Acetate--CoA ligase (ADP-forming)	TK0944				0.0039	3	4
Ribosome biogenesis protein Nop10	TK1101				0.045	2.5	3
Ribonuclease Z	TK1114				0.019	1.3	3
NAD(P)H sulfur oxidoreductase (CoA-dependent)	TK1299				0.00001	1.4	4
Phosphoenolpyruvate carboxykinase [GTP]	TK1405				0.0089	2.2	4
50S ribosomal protein L12	TK1415				0.00088	2.3	4
V-type ATP synthase subunit I	TK1597				0.012	INF	3
V-type ATP synthase subunit E	TK1599				0.037	INF	2
Methylmalonyl-CoA decarboxylase, alpha subunit	TK1622				0.01	1.4	4
DNA-directed RNA polymerase subunit N	TK1699				0.037	INF	4
Xanthine/guanine phosphoribosyltransferase	TK1737				0.0089	2.7	4
Glycogen synthase	TK1768				0.00001	INF	3
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1800				0.0017	3.3	4
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1801				0.0041	2.2	4

ABC-type dipeptide/oligopeptide transport system, permease component	TK1802				0.039	1.7	3
DNA polymerase II large subunit	TK1903				0.00001	1.4	4
Membrane bound hydrogenase, NiFe-hydrogenase small subunit	TK2089				0.037	INF	3

Table 4.20: Co-purified interacting partners of FDHγ_4 (TK2078) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Uncharacterized protein	TK0930	0.00001	12	5	0.00001	14	5
Protein disulfide oxidoreductase	TK1085	0.00022	6.3	5	0.00001	1.7	4
NAD(P)H sulfur oxidoreductase (CoA-dependent)	TK1299	0.047	2.6	3	0.00001	1.5	4
4Fe-4S cluster-binding protein	TK2078	0.00001	8.2	5	0.00001	1.7	4
Nucleotidyltransferase, fused to N-terminal DNA-binding domain	TK0063	0.029	3.8	3			
Indolepyruvate oxidoreductase subunit IorA	TK0136	0.0021	7.7	5			
UPF0173 metal-dependent hydrolase TK0141	TK0141	0.00001	4.9	5			
RecJ-like exonuclease, containing OB-fold nucleic acid-binding domains	TK0155	0.009	2.9	4			
Fibrillarin-like rRNA/tRNA 2'-O-methyltransferase	TK0183	0.025	2.7	3			
L-aspartate oxidase	TK0297	0.032	4.5	4			
TatD-related deoxyribonuclease	TK0317	0.023	11	3			
RNA-splicing ligase RtcB	TK0358	0.00025	12	5			
Uncharacterized protein	TK0438	0.023	11	4			
Uncharacterized protein	TK0453	0.0068	8.5	5			
Uncharacterized protein	TK0483	0.00001	5.6	5			
Flavin prenyltransferase UbiX	TK0509	0.0063	14	5			
Metal-dependent phosphohydrolase, HD superfamily	TK0540	0.0096	6.3	5			
Metallophosphoesterase, calcineurin superfamily	TK0574	0.0071	INF	5			
Glycerol-1-phosphate dehydrogenase [NAD(P)+]	TK0789	0.01	8	5			
Type 2 DNA topoisomerase 6 subunit A	TK0798	0.0052	2.8	4			
Type 2 DNA topoisomerase 6 subunit B	TK0799	0.00012	3	4			
Digeranylgeranylglycerophospholipid reductase	TK1088	0.039	5	4			
Ribonuclease Z	TK1114	0.044	4.2	4			
Transcription regulator, PadR-like family	TK1143	0.012	INF	4			
Phosphoglycerate kinase	TK1146	0.036	4	4			
Predicted AP endonuclease	TK1165	0.00001	4.4	5			
Uncharacterized protein	TK1394	0.0041	15	5			
Glycerol kinase	TK1396	0.015	12	3			

Phosphomannomutase-related protein	TK1404	0.032	6.5	4			
50S ribosomal protein L10	TK1416	0.015	7.5	4			
1,4-alpha-glucan branching enzyme TK1436	TK1436	0.0026	INF	5			
Leucine--tRNA ligase	TK1461	0.0096	6.3	5			
50S ribosomal protein L4	TK1541	0.029	3.8	3			
50S ribosomal protein L3	TK1542	0.035	3.2	3			
Uncharacterized protein	TK1545	0.046	6	4			
Predicted metal-dependent hydrolase	TK1611	0.00051	5.1	5			
Methylmalonyl-CoA decarboxylase, alpha subunit	TK1622	0.033	3.4	3			
Uncharacterized protein	TK1626	0.032	6.5	4			
Proteasome subunit alpha	TK1637	0.028	5.3	4			
Predicted transcription regulator, DUF118 helix-turn-helix family	TK1769	0.0097	13	5			
Predicted hydrolase, metallo-beta-lactamase superfamily	TK1778	0.022	7	4			
Probable tRNA/rRNA methyltransferase	TK1785	0.00001	6.8	5			
Probable lipoprotein releasing system, ATP-binding protein	TK1859	0.0019	5.8	5			
DNA polymerase II large subunit	TK1903	0.00001	3.8	4			
2-oxoisovalerate:ferredoxin oxidoreductase, alpha subunit	TK1980	0.026	3.5	3			
Iron-molybdenum cofactor-binding protein	TK2016	0.012	INF	4			
SAM-dependent methyltransferase, UPF0020 family	TK2045	0.035	10	3			
Cytosolic NiFe-hydrogenase, gamma subunit	TK2071	0.0071	INF	5			
Cytosolic NiFe-hydrogenase, beta subunit	TK2072	0.0043	INF	5			
Probable formate dehydrogenase, alpha subunit	TK2076	0.00001	3	4			
4Fe-4S cluster-binding protein	TK2077	0.01	2.4	4			
Enolase	TK2106	0.0026	16	5			
tRNA (cytosine(72)-C(5))-methyltransferase	TK2122	0.00001	INF	4			
Pantoate kinase	TK2141	0.032	6.5	4			
Radical_SAM domain-containing protein	TK2160	0.031	INF	4			
Tungsten-containing glyceraldehyde-3-phosphate:ferredoxin oxidoreductase	TK2163	0.019	INF	4			
Fructose-1,6-bisphosphate aldolase/phosphatase	TK2164	0.0041	15	5			
Zinc-dependent protease, TldD/PmbA family	TK2169	0.035	10	4			
tRNA-splicing endonuclease	TK2215	0.041	2.4	3			
Ribonuclease VapC	TK2228	0.035	10	4			
Deoxycytidylate deaminase	TK2257	0.015	12	4			
ATP-binding protein	TK2280	0.031	INF	4			
Anaerobic ribonucleoside-triphosphate reductase	TK2298	0.022	4	4			
Flagellin B2	TK0039				0.017	4	3
Flagellin B3	TK0040				0.017	INF	3

Indolepyruvate oxidoreductase subunit IorB	TK0135		0.002	1.7	4
Superoxide reductase	TK0525		0.017	INF	4
DNA/RNA-binding protein Alba	TK0560		0.011	2	3
Peroxiredoxin, AhpC/TSA family	TK0815		0.0011	INF	5
Rubrerythrin domain-containing protein	TK0826		0.00001	INF	5
Glucosamine-1-phosphate N-acetyltransferase	TK1188		0.00041	2.5	4
50S ribosomal protein L13	TK1501		0.041	2	3
Cysteine synthase	TK1687		0.0051	5	5
Glycogen synthase	TK1768		0.00001	INF	5

While it is possible that *Tk*-FDH is composed of a four to six subunit complex consisting of FDHα and 4Fe-4S cluster binding proteins (FDH$\gamma_{1\text{-}4}$) encoded up- and down-stream of FDHα, the data does not support such associations. Instead, it seems that there are some interactions between FDH-related proteins encoded by the FDH operon, but these associations are not sufficiently stable to be identified as a complex. Functional studies using *Tk*-FDH and associated 4Fe-4S cluster binding proteins would be required to confirm such activity.

As FDH is often seen to function directly with membrane-bound hydrogenase components and encoded directly downstream of the MBH operon (TK2080 - TK2093) it could be assumed that FDH may associate with MBH in -S˚ conditions.[49] TK2089 and TK2093 (lime green ellipses, Figure 4.25B) were identified as significant in the data from TK2077 and TK2076 in -S˚, suggesting that these proteins may be involved in formate oxidation directly coupled to hydrogen generation. TK2075 identified the β subunit of the cytosolic hydrogenase (teal ellipse, Hydβ, TK2072) in -S˚ suggesting that this 4Fe-4S cluster binding subunit may be involved in hydrogen consumption in the cytoplasm.

FDH converts acyl-CoA and formate into CO_2, reducing a Fd in the process.[65] Acetate CoA synthase (ACS) converts acyl-CoA and CO_2 into acids (such as acetate) resolving an ATP in the process, thus FDH and ACS have similar substrates, while one reaction produces energy which is stored as Fd_{red} and the other phosphorylates an ADP to ATP.[64] *T. kodakarensis* encodes two heterodimeric ACS proteins. ACS1$\alpha\beta$ is encoded by TK0665 and TK0465, respectively while ACS2 $\alpha\beta$ is encoded by TK0139 and TK0943, respectively. ACS2α was not found to associate with any subunit of FDH in this study, however in non-sulfur conditions, both FDHγ_2 and FDHγ_3 identified ACS1β while FDHγ_3 identified ACS2β. FDHα identified ACS1α in sulfur conditions, suggesting that acid oxidation from acyl-CoA substrates may occur in a localized area within the cell.

An overview of reductant disposal systems

Reductant disposal pathways in *T. kodakarensis* include the generation of reduced small molecules such as NAD(P)H, the production of fully saturated lipids, or the generation of a proton gradient at the membrane which is used to indirectly drive ATP synthesis (Figure 4.26).[20] Two respiratory complexes reside at the membrane and utilize different terminal electron acceptors; the membrane bound sulfane reductase (MBS) uses inorganic S˚ as a terminal electron acceptor thereby producing H_2S gas, while the membrane bound hydrogenase (MBH) reduces protons thereby generating H_2 gas.[53,54] H_2 gas, a valuable biofuel, thus directing the regulation of electron flux towards MBH is of great interest.[10–13,20,66]

As depicted in Figure 4.4, it is hypothesized that specific PECs interact with distinct electron acceptor proteins which in turn produce unique biproducts, but only recently have such networks begun to be identified (Figure 4.26).[1,2]

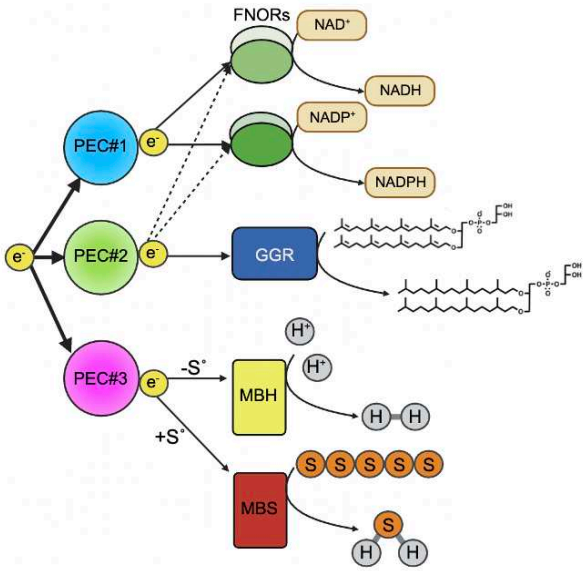


Figure 4.26: Proteinaceous electron carriers (PECs) direct the flux of electrons to specific reductant disposal systems. Electron flux is directed to the generation of reduced small molecules, such as NAD(P)H, via the FNORs. Here, PEC#1 (blue) directs the majority of electron flux towards the generation of reduced cofactors. PEC#2 (pink) directs the flux of electrons to GGR (red) which saturates isoprenoid lipids, thereby providing a second pathway for reductant disposal. At bottom, PEC#3 (pink), provides reductant to the MBH complex whereby protons are reduced to generate H_2 gas. A final route of disposal involves donation of electrons to the MBS complex whereby inorganic S˚ is reduced and H_2S gas is produced. Both MBH and MBS generate a proton gradient at the membrane which drives ATP synthase (not shown).

While all reductant disposal systems were initially included in the study, not all components were successful in generating AP-MS data (see Chapter 3.1.2 for details). The two heterodimeric FNORs were tagged and generated significant data, as did monomeric GGR. However, between the eight proteins involved in respiration, only one tagged subunit of MBS produced AP-MS data. Both MBS and MBH are multimeric, membrane-bound complexes which include cytoplasmic arms.[53,54] Multiple subunits of the cytoplasmic component of both MBS and MBH were attempted, however only one subunit of MBS (MBS-L, TK1215) could be genetically modified and while five subunits of MBH were tagged (MBH-I – -N, TK2088 – TK2091 and TK2092), none could be purified sufficiently to be identified as significant compared to the control strain via LC-MSMS analysis. Interactomes for the FNORs, GGR, and one subunit of MBS will be presented here.

Ferredoxin NAD(P)H Oxidoreductases (FNOR1 and FNOR2)

FNORs primary function is believed to be the reduction of $NAD(P)^+$ to NAD(P)H using electrons donated by reduced Fds.[50,51] Phylogenetic analyses of FNORs have shown that there are three distinct FNOR homologs present in the *Thermococcales*: NADH-dependent ferredoxin $NADP^+$ oxidoreductase I (NfnI), NfnII, and NfnIII.[30] *T. kodakarensis* encodes for FNORs of the NfnII and NfnIII type, which are believed to have an unknown secondary enzymatic function. The unknown secondary enzymatic function, along with the fact that the two FNORs in *T. kodakarensis* are sensitive to SurR transcriptional control,[51] indicates the possibility of different metabolic functions for each of the FNORs.

FNOR1 is composed of an α and β subunit (~ 53 kDa and ~ 33 kDa) encoded by TK1325 and TK1326, respectively. FNOR2 is likewise composed of an α and β subunit (~53 kDa and ~32 kDa) encoded by TK1684 and TK1685, respectively (Figure 4.27, Figure 4.28, Figure 4.29). Fd-1 was primarily found to interact with both FNORs, indicating a role in

regulating the levels of NAD(P)H in the cell.[2] Due to the thermo-instability of NAD(P)H the cell likely requires electrons to be constantly supplied to the FNORs to regenerate reduced NAD(P)H.[32,33]

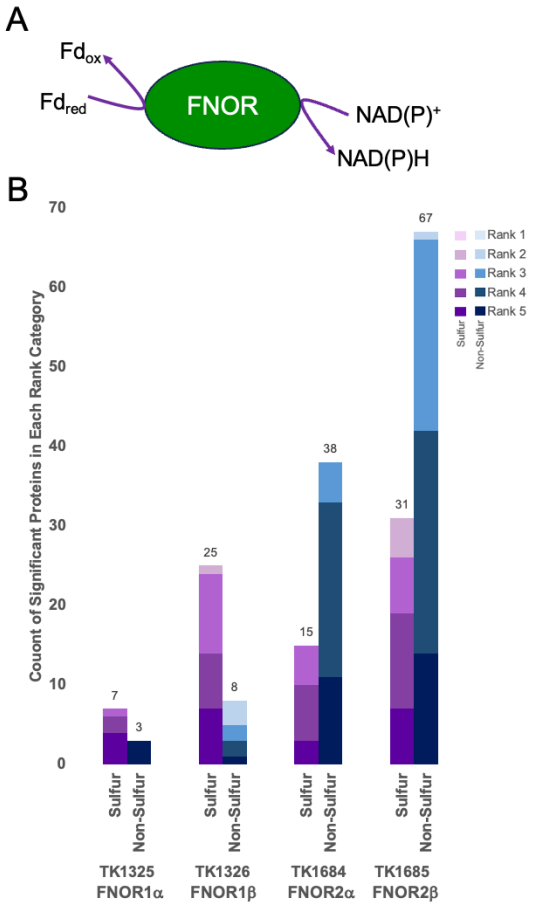


Figure 4.27: FNOR function and number of interacting partners observed in sulfur and non-sulfur conditions. FNORs catalyze the reduction of small molecule $NAD(P)^+$ to NADPH by the donation of an electron from a reduced Fd (Fd_{red}) (A). Two distinct $\alpha\beta$ heterodimeric FNORs are encoded by *T. kodakarensis* (FNOR1: TK1325 - TK1326 and FNOR2: TK1684 - TK1685). FNOR1 has been shown to be more active in +S˚ conditions while FNOR2 was shown to be more active in -S˚ conditions.[51] All four subunits were tagged in this study (B). FNOR1 identified low numbers of proteins on average while FNOR2 displayed a wide range (15 - 67). Consistent with previous findings, FNOR1 (TK1325 - TK1326) identified more interacting partners in +S˚ and FNOR2 (TK1684 - TK1685) identified more interactions in -S˚ conditions. The number of interactions that were observed with a rank score of 5 is shown in dark purple or dark blue, rank 4 is shown in purple or blue, rank 3 is light purple or light blue, and rank 2 is faint purple or faint blue. No interactions with a rank score of 1 were observed for any tagged-FNOR subunit.

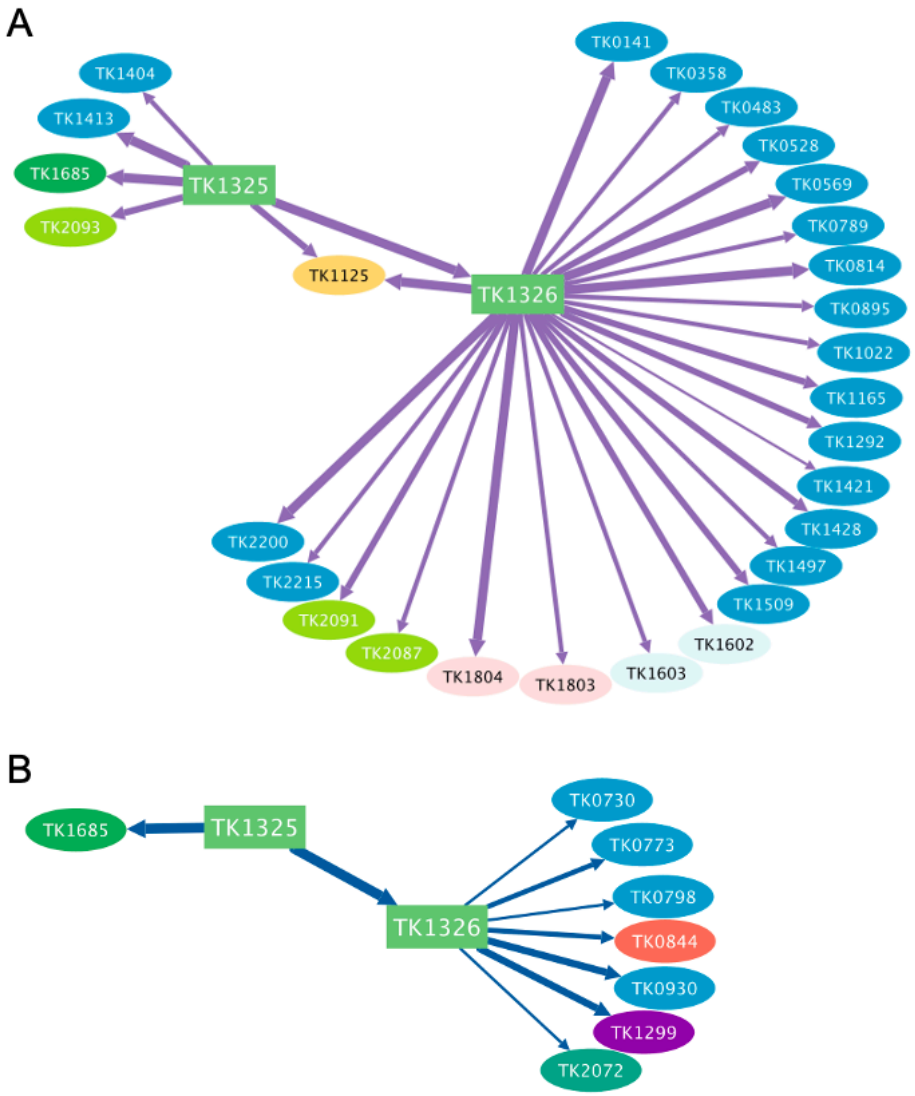


Figure 4.28: Interactome of Ferredoxin NADH oxidoreductase 1 (FNOR1) in sulfur and non-sulfur conditions. FNOR1 is composed of an α (TK1325) and a β subunit (TK1326) (bright green rectangles). AP-MS data shows lower than average numbers of interacting partners for FNOR1 compared to tagged proteins included in this study in both sulfur (A) and non-sulfur (B) conditions, with the lowest number of partners (two) seen for any tagged-protein observed here for FNOR1α in non-sulfur conditions (B). Slightly more interactions were observed in sulfur conditions, supporting previous evidence that FNOR1 is more active in the presence of sulfur.[51] Here, FNOR1α identified FNOR1β, as well as FNOR2α (TK1685) in both conditions. FNOR1β did not identify FNOR1α, thus this partnership was not reciprocally confirmed. Three associations with subunits of the MBH complex (TK2087 - TK2093, lime green ellipses) were observed in sulfur conditions, while FNOR1β was observed to associate with an aldehyde ferredoxin oxidoreductase (AOR, TK0844, salmon ellipse), NAD(P)H sulfur oxidoreductase (NSOR, TK1299, purple ellipse), and the cytosolic hydrogenase (Hydβ, Tk2072, teal ellipse) in non-sulfur conditions. Line thickness represents the rank score observed for the interaction with a thicker line representing a greater rank score.

Table 4.21: Co-purified interacting partners of FNOR1α (TK1325) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1325	0.00001	120	5	0.00001	INF	5
Ferredoxin:NADP oxidoreductase, beta subunit	TK1326	0.00001	INF	5	0.0001	INF	5
Ferredoxin:NADP oxidoreductase, beta subunit	TK1685	0.0003	5	5	0.0017	INF	5
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1125	0.046	5	4			
Phosphomannomutase-related protein	TK1404	0.029	3.5	3			
Archaeal histone A	TK1413	0.0017	5.5	5			
Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	TK2093	0.028	INF	4			

Table 4.22: Co-purified interacting partners of FNOR1β (TK1326) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Ferredoxin:NADP oxidoreductase, beta subunit	TK1326	0.00001	INF	5	0.00001	INF	5
UPF0173 metal-dependent hydrolase TK0141	TK0141	0.001	6.5	5			
RNA-splicing ligase RtcB	TK0358	0.042	INF	3			
Uncharacterized protein	TK0483	0.026	10	3			
Serine hydroxymethyltransferase	TK0528	0.0066	13	4			
Uncharacterized protein	TK0569	0.0002	7.5	5			
Glycerol-1-phosphate dehydrogenase [NAD(P)+]	TK0789	0.042	INF	3			
Type A flavoprotein	TK0814	0.0041	4	5			
S-layer protein	TK0895	0.033	2.3	3			
D-aminopeptidase	TK1022	0.042	INF	3			
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1125	0.001	INF	5			
Predicted AP endonuclease	TK1165	0.00001	4.7	4			
Phosphoenolpyruvate synthase	TK1292	0.0051	INF	4			
Cell division protein FtsZ 1	TK1421	0.025	INF	2			
Cleavage and polyadenylation specficity factor subunit homolog	TK1428	0.0026	7	4			
Phosphopyruvate hydratase	TK1497	0.04	9	3			
Probable tRNA pseudouridine synthase B	TK1509	0.0065	8	4			
V-type ATP synthase alpha chain	TK1602	0.0026	15	4			
V-type ATP synthase beta chain	TK1603	0.026	10	3			
ABC-type dipeptide/oligopeptide transport system, permease component	TK1803	0.042	INF	3			
ABC-type dipeptide/oligopeptide transport system, probable periplasmic component	TK1804	0.00042	6.2	5			

Membrane bound hydrogenase, MbhH subunit	TK2087	0.042	INF	3			
Membrane bound hydrogenase, NiFe-hydrogenase large subunit 2	TK2091	0.0098	7.5	4			
cobW domain-containing protein	TK2200	0.00001	4.2	5			
tRNA-splicing endonuclease	TK2215	0.015	7	3			
ABC-type transport system involved in Fe-S cluster assembly, permease component	TK0730				0.027	4	2
Predicted ATP-dependent endonuclease, OLD family	TK0773				0.047	4.5	3
Type 2 DNA topoisomerase 6 subunit A	TK0798				0.046	2	2
Aldehyde ferredoxin oxidoreductase	TK0844				0.00052	9	3
Uncharacterized protein	TK0930				0.00001	28	4
NAD(P)H sulfur oxidoreductase (CoA-dependent)	TK1299				0.0033	3.4	4
Cytosolic NiFe-hydrogenase, beta subunit	TK2072				0.039	7.5	2

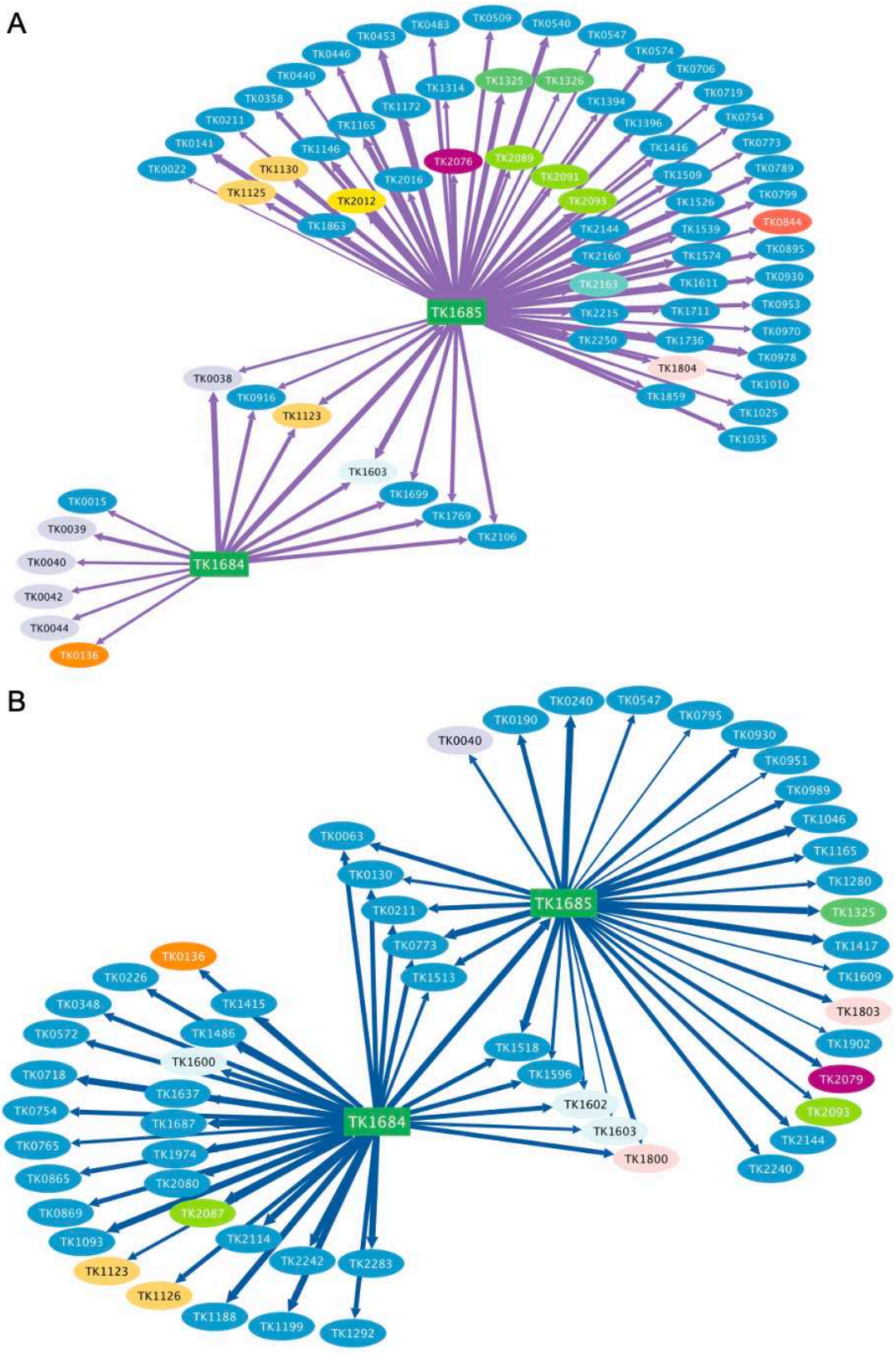

A
TK0509
TK0483
TK0540
TK0453
TK0547
TK0446
TK0574
TK0706
TK0440
TK1325
TK1326
TK1314
TK0719
TK0358
TK1172
TK1394
TK0211
TK1165
TK1396
TK0754
TK0141
TK1146
TK1416
TK0773
TK0022
TK1130
TK2076
TK2089
TK1509
TK0789
TK1125
TK2016
TK2091
TK0799
TK2012
TK2093
TK1526
TK0844
TK1863
TK2144
TK1539
TK0895
TK2160
TK1574
TK0930
TK2163
TK1611
TK0953
TK1685
TK2215
TK1711
TK0970
TK2250
TK1736
TK0978
TK1804
TK1010
TK0038
TK0916
TK1859
TK1025
TK1123
TK1035
TK1603
TK0015
TK1699
TK0039
TK1769
TK2106
TK0040
TK1684
TK0042
TK0044
TK0136
B
TK0240
TK0547
TK0190
TK0795
TK0040
TK0930
TK0951
TK0989
TK1046
TK0063
TK1165
TK0130
TK1280
TK1685
TK0211
TK1325
TK0773
TK1417
TK0136
TK1513
TK1609
TK0226
TK1415
TK1803
TK0348
TK1902
TK0572
TK1486
TK1518
TK1600
TK1596
TK2079
TK0718
TK1637
TK1602
TK2093
TK0754
TK1684
TK1687
TK1603
TK2144
TK0765
TK1974
TK1800
TK2240
TK0865
TK2080
TK0869
TK2087
TK1093
TK2114
TK2242
TK2283
TK1123
TK1126
TK1188
TK1199
TK1292

Figure 4.29: Interactome of Ferredoxin NADPH oxidoreductase 2 (FNOR2) in sulfur and non-sulfur conditions. FNOR2 is composed of an α (TK1684) and a β subunit (TK1685) (green rectangles). AP-MS data shows slightly fewer than average interacting partners in non-sulfur conditions (B) and slightly greater than average interacting partners in sulfur conditions (A). These trends support previous evidence that FNOR2 is more active in the absence of sulfur.[51] As was the case for FNOR1, FNOR2α identified FNOR2β, but this partnership was not reciprocally confirmed in either condition. Instead, FNOR2β identified both subunits of FNOR1 (bright green ellipses) in sulfur and FNOR1α in non-sulfur conditions. Associations with MBH (TK2089 - TK2093, lime green ellipses), OGOR (TK1123 - TK1126, light yellow ellipses), and IOR (TK0136, orange ellipse) were observed in both conditions. Of note, FNOR2β was found in association with Fd3 (TK2012, gold ellipse), GAPOR (TK2163, sea foam green ellipse), and FDHα (TK2076, fuchsia ellipse) in sulfur conditions. Line thickness refers to ranks score with a thicker line representing an association with a greater rank score.

Table 4.23: Co-purified interacting partners of FNOR2α (TK1684) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Indolepyruvate oxidoreductase subunit IorA	TK0136	0.015	2.9	3	0.00001	1.8	4
Indolepyruvate oxidoreductase subunit IorA	TK0136	0.00001	1.8	4	0.015	2.9	3
2-oxoacid:ferredoxin oxidoreductases, gamma subunit	TK1123	0.014	6	4	0.019	1.9	3
2-oxoacid:ferredoxin oxidoreductases, gamma subunit	TK1123	0.019	1.9	3	0.014	6	4
V-type ATP synthase beta chain	TK1603	0.024	5.3	4	0.03	2.2	3
V-type ATP synthase beta chain	TK1603	0.03	2.2	3	0.024	5.3	4
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1684	0.00001	27	5	0.00001	22	5
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1684	0.00001	22	5	0.00001	27	5
Ferredoxin:NADP oxidoreductase, beta subunit	TK1685	0.00001	78	5	0.00001	30	5
Ferredoxin:NADP oxidoreductase, beta subunit	TK1685	0.00001	30	5	0.00001	78	5
Molybdate/tungstate-binding protein WtpA	TK0015	0.025	3.3	3			
Flagellin B1	TK0038	0.00001	9	5			
Flagellin B2	TK0039	0.00019	3.1	4			
Flagellin B3	TK0040	0.015	2.9	3			
Flagellin B5	TK0042	0.015	2.9	3			
Archaeal flagella-related protein D, internal insertion	TK0044	0.019	INF	3			
L-threonine 3-dehydrogenase	TK0916	0.0065	2.1	4			
DNA-directed RNA polymerase subunit A"	TK1699	0.014	6	4			
Predicted transcription regulator, DUF118 helix-turn-helix family	TK1769	0.019	INF	4			
Enolase	TK2106	0.024	5.3	4			
Nucleotidyltransferase, fused to N-terminal DNA-binding domain	TK0063				0.00022	2.6	4
Uncharacterized protein	TK0130				0.039	5	4
Amidophosphoribosyltransferase	TK0211				0.0024	2.1	4
Uncharacterized protein	TK0226				0.025	INF	4
Predicted membrane protease subunit, stomatin/prohibitin homolog	TK0348				0.0036	3.7	4
ABC-type iron(III) transport system, ATPase component	TK0572				0.0053	2.6	4
ABC-type molybdate transport system, permease component	TK0718				0.0098	INF	5
tRNA(Met) cytidine acetyltransferase TmcA	TK0754				0.00001	1.8	4
Glyceraldehyde-3-phosphate dehydrogenase	TK0765				0.048	3	3
Predicted ATP-dependent endonuclease, OLD family	TK0773				0.0059	2.3	4
Probable vitamin B12 transport protein	TK0865				0.018	6	4
Uncharacterized protein	TK0869				0.018	6	4
Sulfur transfer protein involved in thiamine biosynthesis	TK1093				0.0039	INF	5
2-oxoacid:ferredoxin oxidoreductases, gamma subunit	TK1126				0.0039	INF	4
Glucosamine-1-phosphate N-acetyltransferase	TK1188				0.00085	4.3	5

DNA helicase	TK1199				0.00041	4.7	5
Phosphoenolpyruvate synthase	TK1292				0.00019	2.2	4
50S ribosomal protein L12	TK1415				0.0032	2.5	4
Signal recognition particle 54 kDa protein	TK1486				0.018	6	4
50S ribosomal protein L14e	TK1513				0.028	2.7	3
Protein translocase subunit SecY	TK1518				0.012	4	4
Archaeal/vacuolar-type H+-ATPase, subunit H	TK1596				0.039	5	4
V-type ATP synthase subunit C	TK1600				0.018	6	4
V-type ATP synthase alpha chain	TK1602				0.0011	2.6	4
Proteasome subunit alpha	TK1637				0.0036	2.3	4
Cysteine synthase	TK1687				0.0081	7	5
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1800				0.0036	3.7	4
Carboxymuconolactone decarboxylase-related protein	TK1974				0.039	5	4
Membrane bound hydrogenase, MbhA subunit	TK2080				0.025	INF	3
Membrane bound hydrogenase, MbhH subunit	TK2087				0.0098	INF	5
SPASM domain-containing protein	TK2114				0.0039	INF	5
Beta-ribofuranosylaminobenzene 5'-phosphate synthase	TK2242				0.0098	INF	5
Uncharacterized protein	TK2283				0.0012	5.5	5

Table 4.24: Co-purified interacting partners of FNOR2β (TK1685) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Predicted ATP-dependent endonuclease, OLD family	TK0773	0.0052	3.9	3	0.00001	4.6	5
Predicted AP endonuclease	TK1165	0.00001	4.4	5	0.00001	2.1	4
Ferredoxin:NADP oxidoreductase, alpha subunit	TK1325	0.00001	98	5	0.00001	11	5
Ferredoxin:NADP oxidoreductase, beta subunit	TK1685	0.00001	96	5	0.00001	15	5
Amidophosphoribosyltransferase	TK0211	0.026	3.2	3	0.00026	2.1	4
Uncharacterized protein	TK0930	0.012	4.1	4	0.00028	3.2	4
Uncharacterized protein	TK2144	0.049	6	4	0.001	3.2	4
Metallophosphoesterase, calcineurin superfamily	TK0547	0.042	5	3	0.0083	3	3
V-type ATP synthase beta chain	TK1603	0.0012	18	5	0.017	2.7	2
Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	TK2093	0.00001	INF	5	0.018	INF	3
Uncharacterized protein	TK0022	0.012	INF	2			
Flagellin B1	TK0038	0.049	6	3			
UPF0173 metal-dependent hydrolase TK0141	TK0141	0.00029	4.2	5			
RNA-splicing ligase RtcB	TK0358	0.021	5.7	4			
Uncharacterized protein	TK0440	0.033	3.8	3			

PDDEXK_1 domain-containing protein	TK0446	0.02	INF	4
Uncharacterized protein	TK0453	0.0085	5	5
Uncharacterized protein	TK0483	0.00001	3.7	4
Flavin prenyltransferase UbiX	TK0509	0.011	13	4
Metal-dependent phosphohydrolase, HD superfamily	TK0540	0.0054	4.8	5
Metallophosphoesterase, calcineurin superfamily	TK0574	0.0076	INF	4
ABC-type iron(III)-siderophore transport system, periplasmic component fused to N-terminal uncharacterized domain	TK0706	0.049	6	4
ABC-type molybdate transport system, ATPase component	TK0719	0.034	6.5	3
tRNA(Met) cytidine acetyltransferase TmcA	TK0754	0.014	3.7	3
Glycerol-1-phosphate dehydrogenase [NAD(P)+]	TK0789	0.03	5.3	4
Type 2 DNA topoisomerase 6 subunit B	TK0799	0.0014	2.6	4
Aldehyde ferredoxin oxidoreductase	TK0844	0.017	3.8	3
S-layer protein	TK0895	0.01	2.4	4
L-threonine 3-dehydrogenase	TK0916	0.016	3.4	3
Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains	TK0953	0.015	4	4
tRNA (guanine(26)-N(2))-dimethyltransferase	TK0970	0.024	11	3
Glycine--tRNA ligase	TK0978	0.01	4.5	5
Putative 5-methylcytosine restriction system, catalytic subunit	TK1010	0.033	INF	3
Uncharacterized protein	TK1025	0.034	6.5	3
Predicted glutamine amidotransferase, class II	TK1035	0.0045	15	4
2-oxoacid:ferredoxin oxidoreductases, gamma subunit	TK1123	0.016	7.5	4
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1125	0.024	7	4
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1130	0.034	6.5	3
Phosphoglycerate kinase	TK1146	0.0021	5.3	5
TBP-interacting protein	TK1172	0.012	INF	3
Predicted ATPase, AAA superfamily	TK1314	0.029	3.5	3
Ferredoxin:NADP oxidoreductase, beta subunit	TK1326	0.033	INF	3
Uncharacterized protein	TK1394	0.011	13	3
Glycerol kinase	TK1396	0.016	12	3
50S ribosomal protein L10	TK1416	0.049	6	4
Probable tRNA pseudouridine synthase B	TK1509	0.00011	3.8	4
30S ribosomal protein S8	TK1526	0.012	4.8	4
50S ribosomal protein L2	TK1539	0.016	4.6	4
tRNA uridine(34) acetyltransferase	TK1574	0.0012	18	5
Predicted metal-dependent hydrolase	TK1611	0.0035	4.7	5
DNA-directed RNA polymerase subunit A"	TK1699	0.034	6.5	4
Sugar-phosphate nucleotydyltransferase	TK1711	0.012	INF	3

Vitamin B12-dependent ribonucleotide reductase	TK1736	0.0048	2.1	4			
Predicted transcription regulator, DUF118 helix-turn-helix family	TK1769	0.0004	INF	4			
ABC-type dipeptide/oligopeptide transport system, probable periplasmic component	TK1804	0.0028	3.3	4			
Probable lipoprotein releasing system, ATP-binding protein	TK1859	0.016	4.6	4			
Predicted N6-adenine-specific DNA methylase	TK1863	0.035	3.6	3			
Ferredoxin 3	TK2012	0.00001	INF	5			
Iron-molybdenum cofactor-binding protein	TK2016	0.033	INF	4			
Probable formate dehydrogenase, alpha subunit	TK2076	0.0012	2.1	4			
Membrane bound hydrogenase, NiFe-hydrogenase small subunit	TK2089	0.0047	INF	3			
Membrane bound hydrogenase, NiFe-hydrogenase large subunit 2	TK2091	0.039	3.2	3			
Enolase	TK2106	0.013	5.2	4			
Radical_SAM domain-containing protein	TK2160	0.033	INF	3			
Tungsten-containing glyceraldehyde-3-phosphate:ferredoxin oxidoreductase	TK2163	0.0047	INF	5			
tRNA-splicing endonuclease	TK2215	0.012	2.6	3			
Non-specific serine/threonine protein kinase	TK2250	0.0082	2.7	4			
Lysine--tRNA ligase	TK2240				0.00001	2	4
Uncharacterized protein	TK1046				0.0001	6.5	5
Arginase	TK0240				0.00025	6	5
GMP synthase [glutamine-hydrolyzing] subunit A	TK0190				0.00044	2.2	4
Nucleotidyltransferase, fused to N-terminal DNA-binding domain	TK0063				0.0008	2.2	4
Protein translocase subunit SecY	TK1518				0.00092	INF	5
Fructose-bisphosphate aldolase class 1	TK0989				0.0012	2.8	4
ABC-type dipeptide/oligopeptide transport system, permease component	TK1803				0.0025	INF	4
Flagellin B3	TK0040				0.0068	INF	3
50S ribosomal protein L1	TK1417				0.0069	4	5
V-type ATP synthase alpha chain	TK1602				0.0083	3	3
50S ribosomal protein L14e	TK1513				0.0091	2.5	4
ABC-type dipeptide/oligopeptide transport system, ATPase component	TK1800				0.0091	2.5	4
Transcription initiation factor IIB 1	TK1280				0.015	3.5	3
Uncharacterized protein	TK1609				0.017	2.7	2
Uncharacterized protein	TK0130				0.018	INF	3
XPA-binding protein 1 homolog	TK0951				0.018	INF	2
Archaeal/vacuolar-type H+-ATPase, subunit H	TK1596				0.018	INF	3
DNA polymerase II small subunit	TK1902				0.018	2.2	2
Probable formate transporter	TK2079				0.028	5	4
Putative 5-methylcytosine restriction system, GTPase subunit	TK0795				0.035	2.3	2

Lipid maturation: Geranylgeranyl reductase (GGR)

The Archaeal domain is defined by prokaryotes which maintain isoprenoid-based lipids with ether linkages connecting glycerol head groups.[67–69] Lipids are an important reductant sink and geranylgeranyl reductase (GGR) is a monomeric protein responsible for maturation of lipids from unsaturated to saturated isoprenoid chains (Figure 4.30).[19] GGR (TK1088) is encoded in a three-protein operon (TK1086 - TK1088) regulated by the sulfur response regulator (SurR, TK1086), a transcription factor which functions as a sulfur-sensitive redox switch (Figure 4.30A).[28] In the presence of sulfur, the dimeric SurR maintains oxidized cystine residues and cannot initiate transcription, however in the absence of sulfur, these cystine residues become reduced and transcription of genes involved in non-sulfur metabolism (such as the *mbh* operon and *fnor2*) are transcribed.[27–29,31,51]

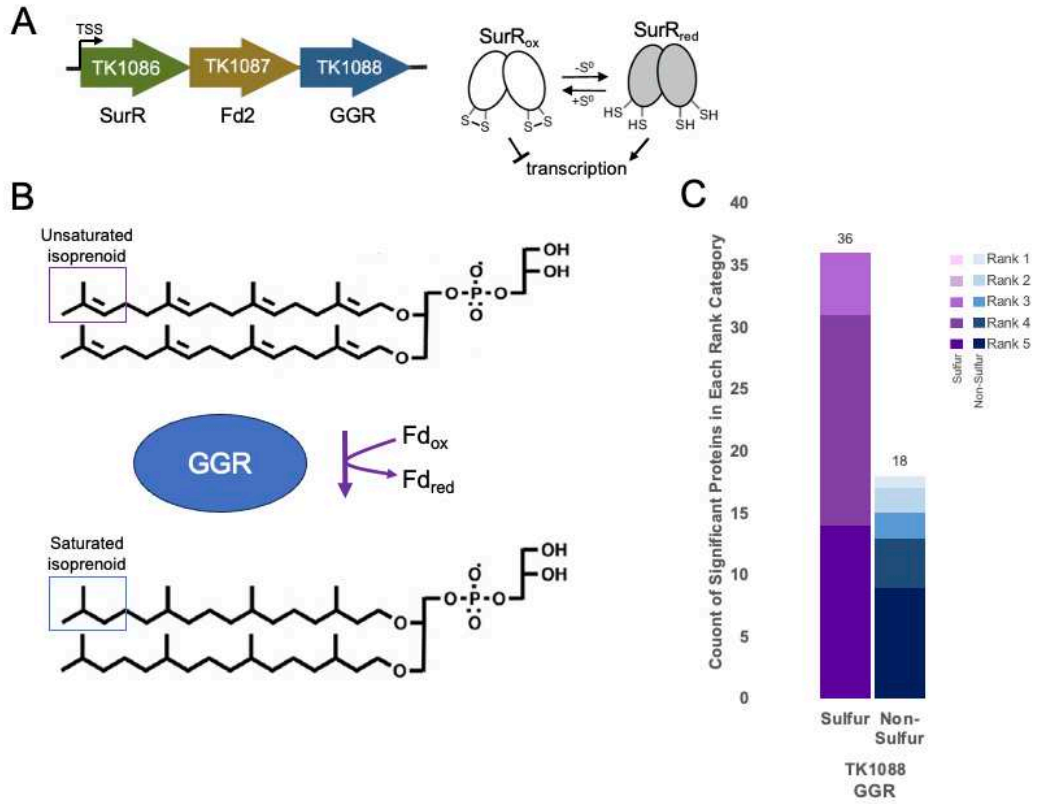


Figure 4.30: GGR genomic organization, transcriptional regulation, function, and number of interacting partners observed in sulfur and non-sulfur conditions. *Ggr* is encoded by the final gene (TK1088) in the SurR operon (TK1086 - TK1088) which includes *surR* and *fd2* encoded upstream of *ggr*. SurR is redox-sensitive transcription factor which is active in the reduced form (grey ovals) in the absence of sulfur (A) (figure adapted from Hidese et al., 2017).

Archaeal membranes are composed of isoprenoid lipids attached to a glycerol moiety via an ether linkage (B). Electrons provided by a reduced Fd are transferred to GGR which is responsible for converting unsaturated isoprene groups (showing in the purple box) to saturated isoprene groups (shown in the blue box) to produce a mature lipid. Tagged-GGR was observed to associate with thirtysix proteins in sulfur and eighteen proteins in non-sulfur conditions in this AP-MS study (C). While the majority of associations were observed with a rank score of 5 - 3 (dark purple or dark blue), some associations were observed with a rank score of 2 or 1 in non-sulfur conditions. (TSS, transcription start site.)

GGR and Fd2 are not only linked in the SurR operon, GGR activity requires Fd2; recombinant expression of GGR from a closely related hyperthermophile, *Metahnosarcina acetivorans*, resulted in fully saturated lipids in *Escherichia coli* only when the archaeal ferredoxin encoded near *Ma*-GGR was expressed.[19] The AP-MS data supported a strong association between Fd2 and GGR in *T. kodakarensis*, as well (Figure 4.31). At ~7 kDa, Fd2 (TK1087) is a small protein and difficult to detect in proteomic mass spectrometry (discussed in depth in Chapter 3.2.3). Even with only a few detectable peptides, Fd2 was found in association with tagged-GGR in both sulfur and non-sulfur conditions (and was not found in any other AP-MS dataset) (Figure 4.31).

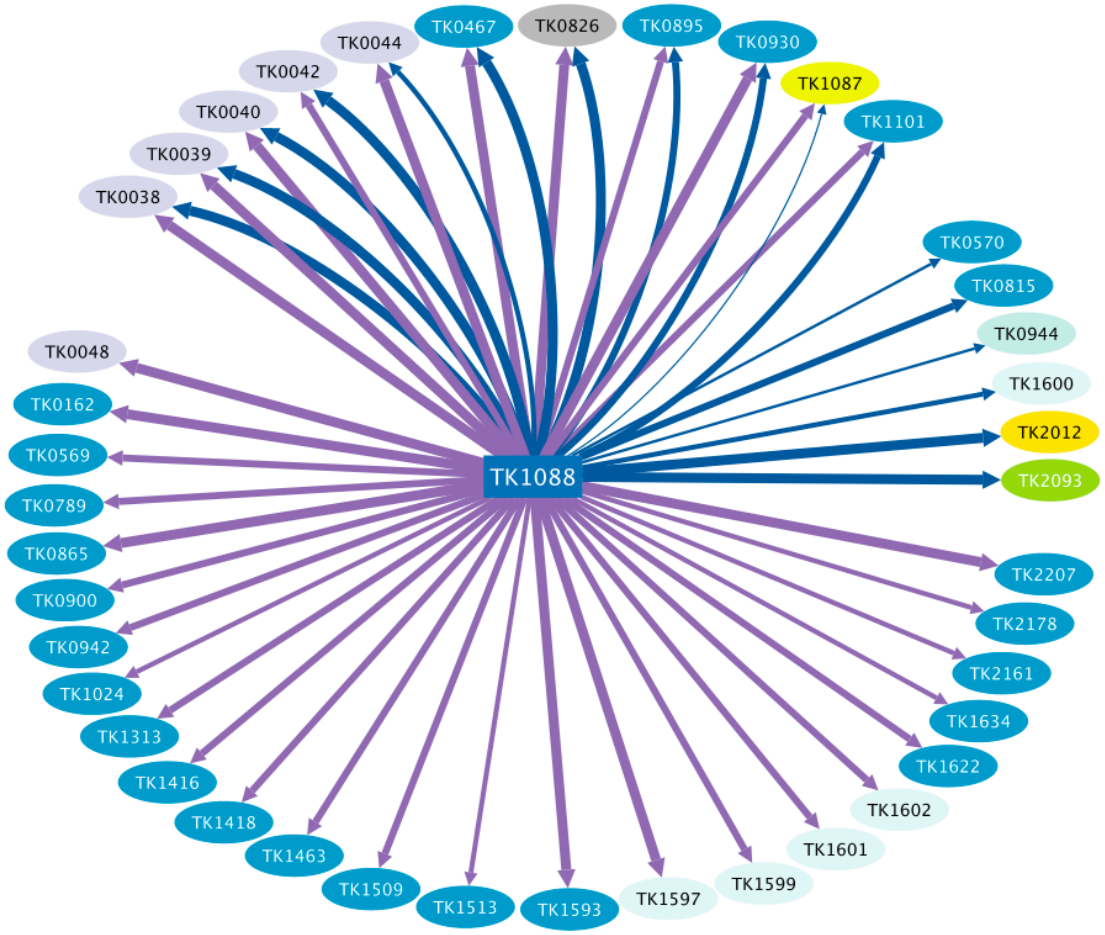


Figure 4.31: Interactome of GGR in sulfur and non-sulfur conditions. AP-MS data identified a greater number of interacting partners in +S˚ (purple lines) than in -S˚ (blue lines), however eleven partners were identified in both conditions including Fd2 (TK1087, lime green ellipse), RBR2 (TK0826, grey ellipse) and five subunits of the flagellin complex (TK0038 - TK0044, light grey ellipses). Another Fd, Fd3 (TK2012, yellow ellipse) and the MBH-N subunit (TK2093, green ellipse) of the MBH complex were identified in non-sulfur conditions. Four subunits of ATP synthase (TK1597 – TK1602, light blue ellipse) were identified in sulfur conditions. Line width represents rank score with thicker lines representing a higher rank score.

Table 4.25: Co-purified interacting partners of GGR (TK1088) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**Fold Change**	**Rank**	**P-Value**	**Fold Change**	**Rank**
Flagellin B1	TK0038	0.00001	6.8	5	0.00001	22	5
Flagellin B2	TK0039	0.00001	8.6	5	0.00001	INF	5
Flagellin B3	TK0040	0.00001	8.8	5	0.00001	INF	5
Flagellin B5	TK0042	0.00001	INF	4	0.00001	INF	5
Archaeal flagella-related protein D, internal insertion	TK0044	0.00001	5.1	5	0.00063	INF	3
Uncharacterized protein	TK0467	0.00001	4	5	0.00001	INF	5
Rubrerythrin domain-containing protein	TK0826	0.00001	5.3	5	0.00001	INF	5
S-layer protein	TK0895	0.00001	3.5	4	0.00001	2.8	4
Uncharacterized protein	TK0930	0.00001	5.5	5	0.00001	5	4
Ferredoxin 2	TK1087	0.00089	9.3	4	0.04	3	1
Digeranylgeranylglycerophospholipid reductase	TK1088	0.00001	7.6	5	0.00001	17	5
Ribosome biogenesis protein Nop10	TK1101	0.0004	3.7	4	0.0087	2.5	4
Archaeal flagella-related protein I, predicted secretion ATPase	TK0048	0.0018	INF	5			
Hypothetical membrane protein	TK0162	0.0065	INF	5			
Uncharacterized protein	TK0569	0.00001	2.3	4			
Glycerol-1-phosphate dehydrogenase [NAD(P)+]	TK0789	0.00043	2.1	4			
Probable vitamin B12 transport protein	TK0865	0.0011	5.3	5			
Uncharacterized protein	TK0900	0.00001	3.2	4			
ABC-type multidrug transport system, ATPase component	TK0942	0.023	INF	4			
Hypothetical membrane protein	TK1024	0.023	INF	3			
Uncharacterized protein	TK1313	0.00001	2.1	4			
50S ribosomal protein L10	TK1416	0.0014	2.2	4			
50S ribosomal protein L11	TK1418	0.00089	9.3	4			
Uncharacterized protein	TK1463	0.0076	3.1	4			
Probable tRNA pseudouridine synthase B	TK1509	0.00001	2.6	4			
50S ribosomal protein L14e	TK1513	0.016	2.3	3			
Protein-export membrane protein SecF	TK1593	0.0028	8	5			
V-type ATP synthase subunit I	TK1597	0.0011	5.3	5			
V-type ATP synthase subunit E	TK1599	0.025	5.3	4			
V-type ATP synthase subunit F	TK1601	0.023	INF	4			
V-type ATP synthase alpha chain	TK1602	0.00001	3	4			
Methylmalonyl-CoA decarboxylase, alpha subunit	TK1622	0.00011	2.7	4			
Exosome complex component Rrp41	TK1634	0.025	5.3	3			

ABC-type multidrug transport system, ATPase component	TK2161	0.023	INF	3			
LTD domain-containing protein	TK2178	0.0028	8	3			
Proteasome subunit beta 2	TK2207	0.0065	INF	5			
ABC-type iron(III) transport system, periplasmic component	TK0570				0.04	3	2
Peroxiredoxin, AhpC/TSA family	TK0815				0.0028	INF	4
Acetate--CoA ligase (ADP-forming)	TK0944				0.04	3	2
V-type ATP synthase subunit C	TK1600				0.0087	2.5	3
Ferredoxin 3	TK2012				0.00001	12	5
Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	TK2093				0.00001	12	5

Membrane Bound Sulfane Reductase

The membrane bound sulfane reductase (MBS) respiratory system generates a chemiosmotic gradient at the membrane upon harnessing the energy of spontaneous electron transfer reactions associated with the conversion of polysulfides into hydrogen sulfide (H_2S) gas (Figure 4.32A).[21,53] The thirteen-subunit complex (TK1214 – TK1226, MBS-A – N) consists of a membrane-bound arm which includes both sodium and proton translocation modules and a cytoplasmic arm (MBS-J, K, L, and N) which maintains three 4Fe-4S centers (Table 4.26).[53] The structure of MBS closely resembles Complex 1, while maintaining unique features such as two proton translocation modules instead of one and including a Na^+/H^+ antiporter at the distal end of the membrane arm.[54] It is proposed that upon reduction of the Fd-S moieties in the cytoplasmic region, two protons are pumped from the cytoplasm to the exterior of the cell and then exchanged for 2 Na^+ ions which are pumped across the membrane and drive a sodium-dependent A_0A_1-ATPase (Figure 4.32A).[21,53] MBS oxidizes a Fd, however no subunits of MBS were identified when the Fds were tagged.[2]

Here, three (TK1215, TK1218, and TK1222; MBS-L, MBS-H', and MBS-E) subunits of MBS were attempted to be tagged, however only one subunit, MBS-L (TK1215) which resides on the cytosolic arm of MBS was successfully genetically modified to include an epitope and affinity tag sequence (Chapter 2.5.4). While MBS is lowly expressed in non-sulfur conditions (Schut et al., 2007),[71] AP-MS data for MBS-L identified ~ fifty associations in both sulfur and non-sulfur conditions (Figure 4.32B, Figure 4.33).

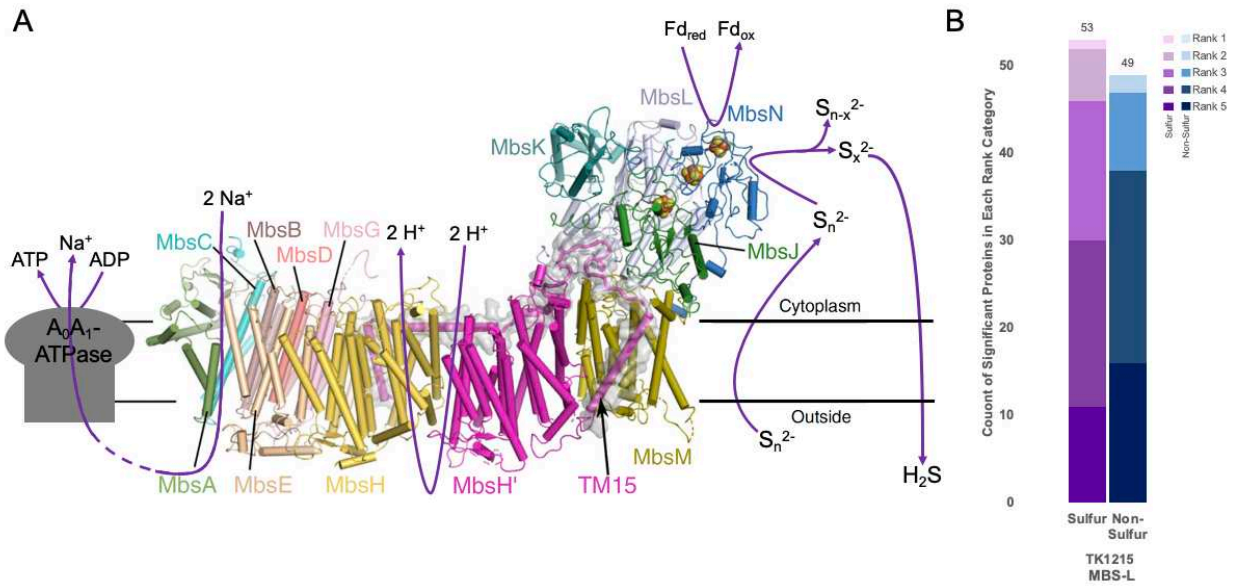


Figure 4.32: The membrane bound sulfane reductase structure and function and AP-MS results by count of rank score in sulfur and non-sulfur conditions. The membrane bound sulfane reductase (MBS) complex resembles Complex 1 and is composed of thirteen-subunits: nine make up a membrane arm and four compose a cytosolic arm (A). The membrane arm is composed of two proton pumps (MBS-H and -H', yellow and bright pink) and a Na^+/H^+ antiporter (MBS-A – C). The cytosolic arm (MBS-J – N, at right, in the cytoplasm) undergoes reduction and the conversion of sulfane (S_n^{2-}) chains into smaller sulfane (S_{n-x}^{2-}) chains.[70] Trisulfides can spontaneously convert into organic sulfur and hydrogen sulfide gas (H_2S). While MBS-L is considered to be the functional catalytic subunit, MBS-N maintains three 4Fe-4S sulfur moieties which are reduced by a Fd. Upon reduction, two protons are pumped across the membrane and exchanged for two sodium ions which drive the A_0A_1-ATPase in the production of ATP. (Figure adapted from Yu et al., 2020.) The MBH-L subunit (TK1215) was tagged in this study, and AP-MS results returned ~50 significant proteins in both sulfur and non-sulfur conditions (B). Rank score in sulfur is indicated by dark purple (rank score of 5) to light purple (rank score of 1) while non-sulfur rank scores are indicated in a blue gradient.

Table 4.26: Membrane bound sulfane reductase gene numbers and annotations in *T. kodakarensis* and *P. furiosis*.

TK Gene No.	PF Gene No.	KEGG Annotation	GenBank Annotation	Published Annotation	Membrane or Cytosolic Arm
TK1226	PF1453	multicomponent Na+:H+ antiporter subunit E	membrane bound hydrogenase, MbxA	MBS-A	Membrane
TK1225	PF1452	multicomponent Na+:H+ antiporter subunit F	membrane bound hydrogenase, MbxB	MBS-B	Membrane
TK1224	PF1451	multicomponent Na+:H+ antiporter subunit G	membrane bound hydrogenase, MbxC	MBS-C	Membrane
TK1223	PF1450	no KO assigned	membrane bound hydrogenase, MbxD subunit	MBS-D	Membrane
TK1222	PF1449	multicomponent Na+:H+ antiporter subunit B	membrane bound hydrogenase, MbxF	MBS-E	Membrane
TK1221	PF1448	NADH-quinone oxidoreductase subunit K	membrane bound hydrogenase, MbxG subunit	MBS-F	Membrane
TK1220	PF1447	NADH-quinone oxidoreductase subunit N	membrane bound hydrogenase, MbxH subunit	MBS-G	Membrane
TK1219	PF1446	NADH-quinone oxidoreductase subunit M	membrane bound hydrogenase, MbxH' subunit	MBS-H	Membrane
TK1218	PF1445	NADH-quinone oxidoreductase subunit H	membrane bound hydrogenase, MbxM subunit	MBS-H'	Membrane
TK1217	PF1444	NADH-quinone oxidoreductase subunit B	membrane bound hydrogenase, NiFe-hydrogenase small subunit	MBS-J	Cytosolic
TK1216	PF1443	NADH-quinone oxidoreductase subunit C	membrane bound hydrogenase, NiFe-hydrogenase large subunit 1	MBS-K	Cytosolic
TK1215	PF1442	NADH-quinone oxidoreductase subunit D	membrane bound hydrogenase, NiFe-hydrogenase large subunit 2	MBS-L	Cytosolic
TK1214	PF1441	NADH-quinone oxidoreductase subunit I	membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	MBS-N	Cytosolic

a: Annotation used in Wu *et al*., 2018 and Yu *et al*., 2020.

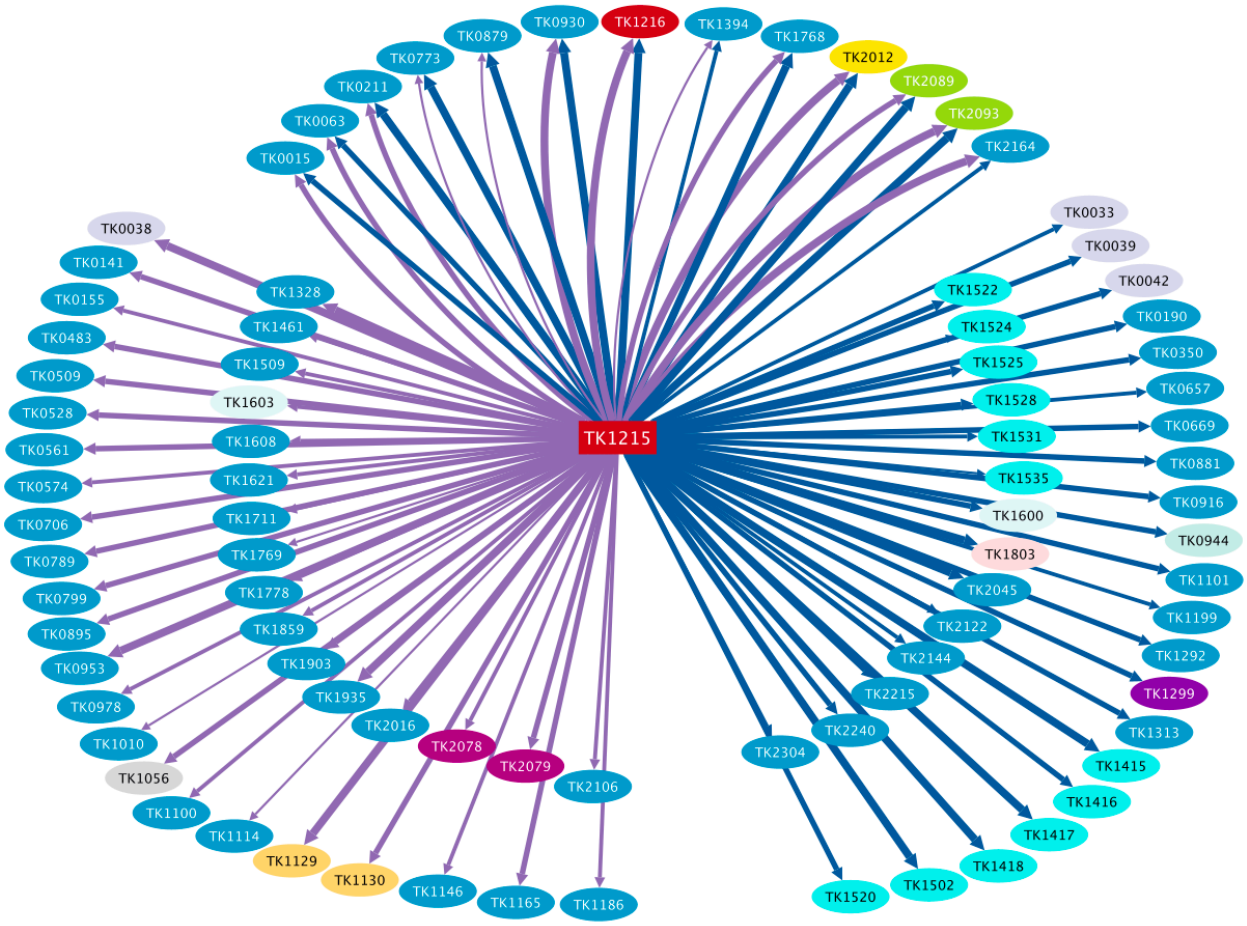


Figure 4.33: Interactome of MBS-L in sulfur and non-sulfur conditions. AP-MS data for the tagged subunit of MBS-L (TK1215, red rectangle) identified ~ fifty interacting partners in both sulfur (purple lines, at left) and non-sulfur (blue lines, at right) conditions, with thirteen of these interacting partners being identified in conditions (purple and blue lines, at top). One other subunit of MBS, MBS-K (TK1216, red ellipse) was identified in both conditions, along with two subunits of MBH (TK2089 and TK2093, lime green ellipses), and Fd3 (TK2012, yellow ellipse). FDHγ_3 and FDHγ_4 (TK2078 and TK2079, fuchsia ellipses) were identified only in sulfur conditions, along with two subunits of OGOR (OGORα, TK1130 and OGORβ, TK1129, light yellow ellipses) and RBR3 (TK1056, grey ellipse). In non-sulfur conditions, twelve ribosomal proteins (TK1415 - TK1535, bright teal ellipses) were identified, along with an acetate-CoA ligase (TK0944, sea foam green ellipse) and NAD(P)H sulfur oxidoreductase (NSOR, TK1299, purple ellipse). The rank score of each association is indicated by line thickness with thicker lines representing greater rank scores.

Table 4.27: Co-purified interacting partners of MBS-L (TK1215) in sulfur and non-sulfur conditions.

		Sulfur			Non-Sulfur		
Identified Proteins	**Gene No.**	**P-Value**	**FC**	**Rank**	**P-Value**	**FC**	**Rank**
Molybdate/tungstate-binding protein WtpA	TK0015	0.022	4.3	4	0.006	3.2	4
Nucleotidyltransferase, fused to N-terminal DNA-binding domain	TK0063	0.007	3	4	0.00001	3.7	4
Amidophosphoribosyltransferase	TK0211	0.0041	2.9	4	0.00001	4	5
Predicted ATP-dependent endonuclease, OLD family	TK0773	0.024	2.5	2	0.00001	8	5
Uncharacterized protein	TK0879	0.044	INF	2	0.0028	INF	5
Uncharacterized protein	TK0930	0.00001	14	5	0.00001	21	5
Membrane bound hydrogenase, NiFe-hydrogenase large subunit 2	TK1215	0.00001	INF	5	0.00001	67	5
Membrane bound hydrogenase, NiFe-hydrogenase large subunit 1	TK1216	0.00001	34	5	0.00001	85	5
Uncharacterized protein	TK1394	0.011	10	2	0.035	2	3
Glycogen synthase	TK1768	0.0037	INF	4	0.00001	INF	5
Ferredoxin 3	TK2012	0.00001	INF	5	0.00001	8.3	5
Membrane bound hydrogenase, NiFe-hydrogenase small subunit	TK2089	0.00001	INF	4	0.00001	7.5	5
Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	TK2093	0.00001	INF	5	0.00001	8.3	5
Fructose-1,6-bisphosphate aldolase/phosphatase	TK2164	0.0021	13	5	0.02	2.1	3
Flagellin B1	TK0038	0.008	6.5	5			
Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains	TK0953	0.0011	4.3	5			
2-oxoacid:ferredoxin oxidoreductases, beta subunit	TK1129	0.001	INF	5			
tRNA (1-methyladenosine) methyltransferase	TK1328	0.0064	11	5			
tRNA/rRNA cytosine-C5-methylase, NOL1/NOP2/Sun family	TK1935	0.00001	19	5			
UPF0173 metal-dependent hydrolase TK0141	TK0141	0.0074	2.7	4			
Uncharacterized protein	TK0483	0.0019	2.4	4			
Flavin prenyltransferase UbiX	TK0509	0.0021	13	4			
Serine hydroxymethyltransferase	TK0528	0.0082	2.5	4			
Adenylosuccinate lyase	TK0561	0.0081	3.3	4			
ABC-type iron(III)-siderophore transport system, periplasmic component fused to N-terminal uncharacterized domain	TK0706	0.034	5	4			
Glycerol-1-phosphate dehydrogenase [NAD(P)+]	TK0789	0.021	5.5	4			
Type 2 DNA topoisomerase 6 subunit B	TK0799	0.0022	2	4			
S-layer protein	TK0895	0.00084	2	4			
Rubrerythrin-related protein	TK1056	0.024	INF	4			
2-oxoacid:ferredoxin oxidoreductases, alpha subunit	TK1130	0.013	6	4			
Predicted AP endonuclease	TK1165	0.00001	2.5	4			
DNA polymerase II large subunit	TK1903	0.001	2.1	4			
Probable formate transporter	TK2079	0.024	INF	4			
RecJ-like exonuclease, containing OB-fold nucleic acid-binding domains	TK0155	0.016	2.2	3			

Metallophosphoesterase, calcineurin superfamily	TK0574	0.002	INF	3			
Glycine--tRNA ligase	TK0978	0.038	2.8	3			
Translation initiation factor 2 subunit alpha	TK1100	0.013	3.6	3			
Phosphoglycerate kinase	TK1146	0.038	2.8	3			
Uncharacterized protein	TK1186	0.043	3	3			
Probable tRNA pseudouridine synthase B	TK1509	0.017	2.1	3			
V-type ATP synthase beta chain	TK1603	0.032	8	3			
Uncharacterized protein	TK1608	0.044	INF	3			
Translation initiation factor 2 subunit beta	TK1621	0.033	2.4	3			
Sugar-phosphate nucleotydyltransferase	TK1711	0.044	INF	3			
Predicted hydrolase, metallo-beta-lactamase superfamily	TK1778	0.021	5.5	3			
Probable lipoprotein releasing system, ATP-binding protein	TK1859	0.013	3.6	3			
Iron-molybdenum cofactor-binding protein	TK2016	0.024	INF	3			
4Fe-4S cluster-binding protein	TK2078	0.013	3.3	3			
Enolase	TK2106	0.032	3.5	3			
Putative 5-methylcytosine restriction system, catalytic subunit	TK1010	0.044	INF	2			
Ribonuclease Z	TK1114	0.032	3.5	2			
Predicted transcription regulator, DUF118 helix-turn-helix family	TK1769	0.024	INF	2			
Leucine--tRNA ligase	TK1461	0.043	3	1			
50S ribosomal protein L12	TK1415				0.00001	4.3	5
50S ribosomal protein L1	TK1417				0.0097	4	5
50S ribosomal protein L11	TK1418				0.00041	13	5
50S ribosomal protein L18e	TK1502				0.0057	9.3	5
ABC-type dipeptide/oligopeptide transport system, permease component	TK1803				0.008	5.3	5
SAM-dependent methyltransferase, UPF0020 family	TK2045				0.00001	INF	5
Flagellin B2	TK0039				0.0075	INF	4
Flagellin B5	TK0042				0.0075	INF	4
GMP synthase [glutamine-hydrolyzing] subunit A	TK0190				0.0063	2.1	4
Uncharacterized protein	TK0350				0.02	INF	4
CDC48/VCP homolog, AAA superfamily	TK0669				0.0021	2.7	4
Universal stress protein	TK0881				0.02	INF	4
L-threonine 3-dehydrogenase	TK0916				0.00069	2.6	4
Acetate--CoA ligase (ADP-forming)	TK0944				0.013	8	4
Ribosome biogenesis protein Nop10	TK1101				0.031	6.7	4
Phosphoenolpyruvate synthase	TK1292				0.0004	2.6	4
NAD(P)H sulfur oxidoreductase (CoA-dependent)	TK1299				0.00001	2.3	4
Uncharacterized protein	TK1313				0.00013	2.5	4
50S ribosomal protein L10	TK1416				0.00099	3.2	4

50S ribosomal protein L30	TK1520		0.0031	3.5	4
50S ribosomal protein L18	TK1522		0.0016	6.7	4
50S ribosomal protein L6	TK1525		0.00088	3.6	4
50S ribosomal protein L5	TK1528		0.037	4	4
V-type ATP synthase subunit C	TK1600		0.017	4.7	4
tRNA-splicing endonuclease	TK2215		0.00001	2.1	4
Lysine--tRNA ligase	TK2240		0.00001	2.3	4
Uncharacterized protein	TK0033		0.021	2.3	3
ABC-type transport system, probable periplasmic component	TK0657		0.035	2	3
DNA helicase	TK1199		0.039	3.1	3
50S ribosomal protein L14	TK1531		0.012	2.7	3
tRNA (cytosine(72)-C(5))-methyltransferase	TK2122		0.0075	INF	3
Uncharacterized protein	TK2144		0.038	2.2	3
tRNA/rRNA cytosine-C5-methylase, NOL1/NOP2/Sun family, fused to N-terminal NusB regulator domain	TK2304		0.00001	3.7	3
50S ribosomal protein L32e	TK1524		0.039	3.1	2
50S ribosomal protein L29	TK1535		0.02	INF	2

While MBS is not thought to generate H_2S directly, di- and tri-sulfides can spontaneously convert to H_2S and S˚ at ~ neutral pH.[71] Additionally, *T. kodakarensis* encodes two H_2S producing proteins: NAD(P)H sulfur reductase (TK1299) and NAD(P)H polysulfide oxidoreductase (TK1481), thus MBS may be involved in the generation of reduced small molecules which in turn can be used to generate H_2S via sulfur reductases.[20] Tagged-MBS-L was found to associate with TK1299 in -S˚ conditions, but TK1481 was only found when OGOR and FDHγ_2 were tagged, again, in -S˚ conditions (Figure 4.33).

Tagged-MBS and tagged-OGOR reciprocally identified multiple subunits of one another in +S˚ conditions suggesting that these two complexes interact. OGOR may oxidize α-ketoglutarate and transfer these electrons either directly to MBS or via an electron carrier. Of the three tagged subunits of OGOR and the one tagged subunit of MBS, nineteen proteins were identified in +S˚ by both complexes, including Fd3. Fd3 may be reduced by OGOR and subsequently reduce MBS, linking amino acid fermentation and the MBS respiratory system at the membrane.

Likewise, tagged-MBS and tagged-FDH identified multiple subunits of one another in +S˚ conditions suggesting that additional electrons garnered from the conversion of acyl-CoA and formate to CO_2 by FDH may be donated directly to MBS or via an electron carrier. Considering that FDH and MBS also both associated with Fd3, it is theorized that Fd3 not only aids in transferring electrons from OGOR, but from FDH to MBS as well. Twenty-nine additional proteins were identified by at least one subunit of both complexes, suggesting considerable metabolic network connectivity.

Such overlap between MBS, OGOR, and FDH suggests these complexes are linked *in vivo*. In total, eight proteins were identified by at least one subunit of MBS, OGOR, and FDH (Table 4.28).

Table 4.28: Proteins identified in MBS, OGOR, and FDH overlap in sulfur conditions.

Bait Complex	Bait Gene No.	Rank Score	Prey Gene No.	Annotation, acronym, and function
OGOR	TK1129	4	TK0789	Glycerol-1-phosphate dehydrogenase [NAD(P)+] (egsA); forms sn-glycerol 1-phosphate which composes the backbone of Archaeal membrane phospholipids.
MBS	TK1215	4		
FDH	TK2076	5		
FDH	TK2077	2		
FDH	TK2078	5		
OGOR	TK1129	4	TK0574	Metallophosphoesterase, calcineurin superfamily; metal-dependent hydrolysis of phospho-esters.
MBS	TK1215	3		
FDH	TK2076	3		
OGOR	TK1129	5	TK0930	Uncharacterized protein.
MBS	TK1215	5		
FDH	TK2077	5		
FDH	TK2078	5		
OGOR	TK1123	2	TK0953	Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains.
MBS	TK1215	5		
FDH	TK2076	5		
FDH	TK2077	3		
OGOR	TK1129	3	TK1328	tRNA (1-methyladenosine) methyltransferase.
MBS	TK1215	5		
FDH	TK2077	5		
OGOR	TK1129	4	TK2012	Ferredoxin 3, Fd3; proteinaceous electron carrier.
MBS	TK1215	5		
FDH	TK2077	5		
OGOR	TK1129	4	TK2093	Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit, MBH-N; cytosolic component of the membrane bound hydrogenase.
MBS	TK1215	5		
FDH	TK2077	5		
OGOR	TK1123	4	TK2106	Enolase, eno; reversibly converts 2-phosphoglycerate (2PG) to phosphoenolpyruvate (PEP) in glycolysis.
MBS	TK1215	3		
FDH	TK2077	4		

While MBS is lowly expressed in -S°, expression changes can be observed within ten minutes after the addition of sulfur.[71] Tagged-MBS-L was identified with twelve ribosomal-related proteins in -S° conditions (TK1415 - TK 1535, bright teal ellipses, Figure 4.33). Such associations cannot be due to an upregulation of translation of MBS-L because MSB-L is tagged at the C-terminus, thus all tagged-protein subunits are fully translated. Instead, ribosomal proteins are known to co-localize to the membrane in prokaryotes[72] and such activity could be redox dependent in *T. kodakarensis.*

4.4 Discussion

AP-MS data of an expansive number of proteins collectively involved in metabolism provides insights about the complexity of the *in vivo* protein landscape. Here, twenty-five genes encoding metabolic enzymes composing at least ten functional complexes were modified at their native genomic loci to generate tagged-proteins at native expression levels. These data identified novel partnerships, refined subunit associations for previously determined protein complexes, and revealed a more interconnected landscape than could be deciphered previously.

Protein complexes are defined and refined

Most of our understanding of protein function comes form *in vitro* studies where protein complexes were identified based on size composition followed by functional studies from native or recombinantly purified protein components, the addition of substrates, and the measurement of resulting products.[3,41,42] For instance, POR was originally classified as a heterotrimer and could catabolize pyruvate[42] without the inclusion of the fourth, non-operonic, subunit which was only identified later.[3,39,41] Subunit composition can be redefined using AP-MS data to provide a more realistic view of *in vivo* associations.

AP-MS data for IOR demonstrated unique associations for this supposed tetramer of homodimers. IORα_1 and IORβ_1 (encoded by TK0135 and TK0136) were tagged in this study, however non-operonic α_2 and β_2 subunits are encoded by TK1643 and TK2244, respectively. While IORα_2 was not observed in this data, IORβ_2 was found to associate with both tagged subunits of IORα_1 and β_1 in both sulfur and non-sulfur conditions, suggesting that IOR may be composed of a $2\alpha_1\beta_1\beta_2$ tetramer and such distinctions can further be confirmed with *in vitro* studies.

While IOR was observed with novel associations, previous and well-studied associations of the FNORs were not supported by the AP-MS data in this study. Previous work with the FNOR complexes identified each consisting of an $\alpha\beta$ dimer with FNOR1 being more abundant in +S˚ conditions while FNOR2 is more abundant in -S˚ conditions.[30,50,51] In this study, both FNOR1α and FNOR2β identified one another in both conditions. FNOR1α and FNOR2α identified the respective β subunits in both conditions, but neither β subunit reciprocally identified the α subunit in either sulfur or non-sulfur condition. *In vitro* functional studies with these newly observed partnerships are of interest to confirm if these associations provide additional efficiencies.

Some protein associations are dependent on the presence or absence of sulfur

Previously determined POR and VOR subunit associations were confirmed in sulfur conditions; POR and VOR share a common γ subunit which was found to associate with all other respective subunits in sulfur conditions. In non-sulfur conditions, however, POR and VOR function with considerably more overlap as both α and β subunits of VOR or POR of the alternative complex were identified, while the δ subunit (TK1979 for VORδ and TK1982 for PORδ) was the only subunit which didn't cross-associate. In -S˚, POR and VOR may interchange subunits more readily or function as a heterooctamer.

FDH is not well described in *T. kodakarensis*, and AP-MS data in this study suggests that the proteins included in this study (TK2075 – TK2078) do not function as a stable multimeric complex. While a number of associations between the tagged subunits were observed in both sulfur and non-sulfur conditions, there was not consistent enough overlap nor reciprocal interactions to suggest these proteins function as a stable complex. Instead, like POR and VOR, two FDH-related proteins, FDHα (TK2076) and FDHγ_4 (TK2078), were seen with

more interacting partners in +S˚ conditions, while FDHγ_2 (TK2075) and FDHγ_3 (TK2077) were observed with more interacting partners in -S˚ conditions.

MBS is the respiratory complex involved in the generation of a chemiosmotic gradient upon reduction in the presence of sulfur. The *mbs* thirteen-gene operon is lowly expressed in the absence of sulfur, however expression changes can be observed at the protein level within ten minutes following sulfur supplementation.[71] AP-MS data in this study observed MBS-L bound to twelve ribosomal related proteins in -S˚ conditions, none of which were observed in +S˚ conditions. Such an abundance of ribosomal related proteins in the data suggests that ribosomal proteins may associate with MBS at the membrane in -S˚.

In the presence of sulfur, data suggests that MBS associates with FDH and OGOR as significant overlapping associations were observed for these tagged proteins; nine proteins, including Fd3 and MBH-N, were identified when at least one subunit of MBS, FDH, and OGOR was tagged suggesting significant overlap of these networks in +S˚ conditions. OGOR decarboxylates α-ketoglutarate to generate acyl-CoA and Fd_{red}.[38] FDH converts acyl-CoA and formate to CO_2, again generating Fd_{red}.[65] MBS oxidizes Fd to generate an ion gradient at the membrane, driving ATP synthase and producing H_2S gas in the process. The overlap of OGOR, FDH, and MBS networks brings together catabolic and anabolic processes, which may be indicative of functionally distinct *in vivo* processes.

Ferredoxins identified in this study

Many of the proteins included in this study were identified as interacting partners when the three *Tk*-Fds were tagged previously.[2] Additionally, *in vitro* studies with many oxidoreductases, such as GAPOR, IOR, OGOR, VOR, POR, and FNORs, identified efficient catalysis when paired with Fds.[37,39–43,50,61] Unfortunately, LC-MSMS does not equally identify all peptides, especially peptides which are too large, too small, or highly charged (see Chapter

3.2.3). Fd1 is beyond detection limits and was not identified in this study, while Fd2 lies at the detection limit and was only identified in partnership with GGR, confirming an established and well documented relationship.[19,52] Fortunately, Fd3 is detectable and was identified by OGOR, MBS, GGR, FNOR2, FDH, and both RBRs (Figure 4.34).

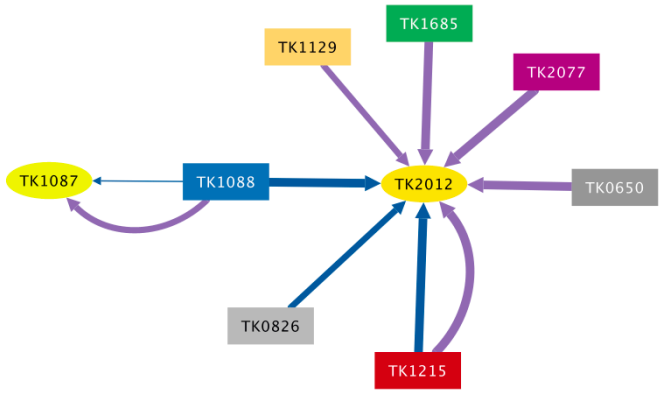


Figure 4.34: Ferredoxin 2 and Ferredoxin 3 were identified in this study. Tagged proteins (rectangles) identified Fd2 (TK1087, lime green ellipse) and Fd3 (TK2012, yellow ellipse). Fd2 (TK1087) was identified by GGR (TK1088, blue rectangle) in both sulfur (purple line) and non-sulfur (blue line) conditions, with greater confidence in the association in sulfur conditions (as indicated by the thicker line representing rank score). GGR also identified Fd3 in non-sulfur conditions. Both RBR1 (TK0650, grey rectangle) and RBR2 (TK0826, light grey rectangle) identified Fd3, though the associations were observed in different metabolic conditions: RBR1 was observed in sulfur and RBR2 was observed in non-sulfur. OGORβ (TK1129, yellow rectangle), FNOR2β (TK1685, green rectangle), and FDHγ_3 (TK2077, fuchsia rectangle) identified Fd3 in sulfur conditions, while MBH-L (TK1215) was the only protein to identify Fd3 in both sulfur and non-sulfur conditions.

The association between GGR and Fd2 is well established; considering that Fd2 is at the edge of the detection limit of LC-MSMS, this association appears strong enough to be detected despite detection limitations.

Tagged-FDHγ_3 (TK2077) identified Fd3 which supports previous findings that Fd3 is linked to FDH.[2] The associations of Fd3 with OGOR, FNOR2, GGR, and MBS are novel and suggest that Fd3 may link glutamate catabolism, acid oxidation, and the generation of a proton gradient at the membrane. RBRs appear to be important peroxidases and their specific association with Fd3 in different conditions suggests that they play a role in maintaining the local redox environment surrounding Fd3.

Rubrerythrins maintain specific associations and are abundant in metabolic networks

Rubrerythrins (RBRs) are non-heme, iron containing proteins known to alleviate oxidative stress through peroxidase activity.[44,58,73] *T. kodakarensis* encodes three annotated RBRs, TK0650, TK0826, and TK1056 (RBR1, RBR2, and RBR3), all of which were observed when at least one Fd was tagged.[2] RBR1 and RBR2 were tagged in this study after each was identified with multiple subunits of the VOR and POR complexes, respectively. It was later shown that RBR3 was identified by every tagged-OGOR subunit, thus it appears that specific oxidoreductase complexes preferentially associate with specific RBRs (Figure 4.35).

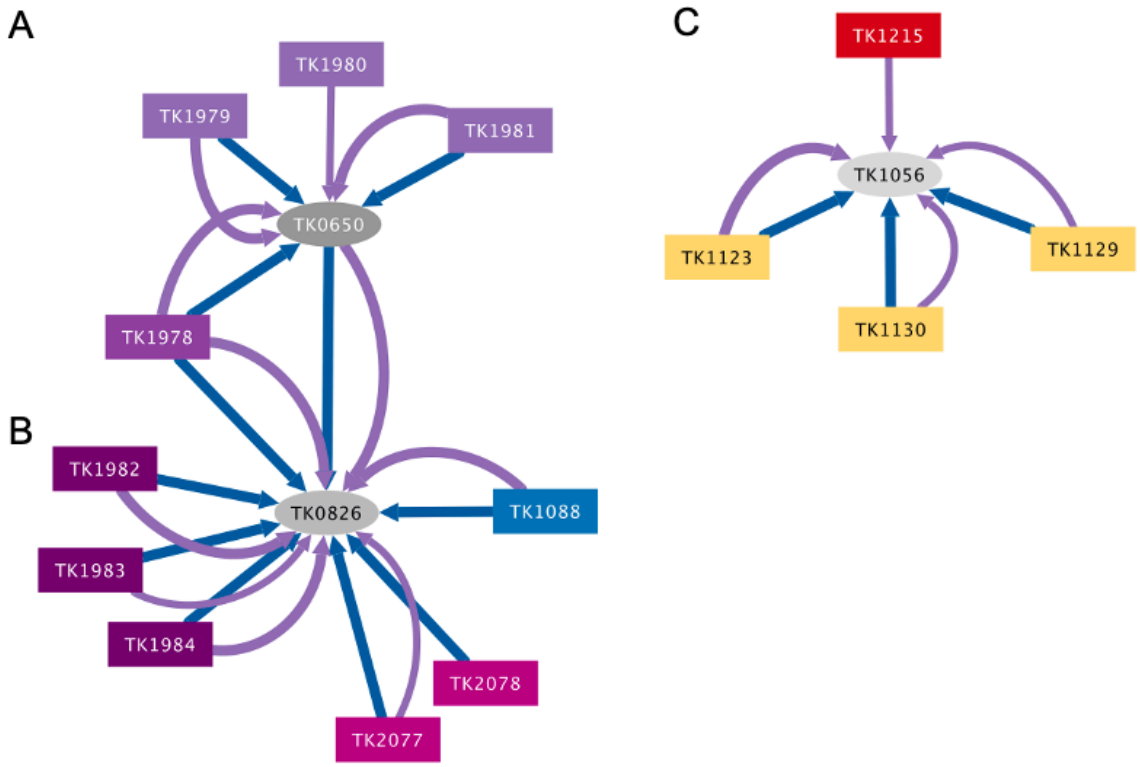


Figure 4.35: RBRs were identified by more than half of tagged proteins included in this study. *T. kodakarensis* encodes three rubrerythrins, RBR1 (TK0650, dark grey ellipse at center in A), RBR2 (TK0826, grey ellipse at center in B), and RBR3 (TK1056, light grey ellipse at center in C), which were identified in the AP-MS data by fourteen different tagged proteins (colored rectangles). These associations were often seen in both sulfur (purple lines) and non-sulfur conditions (blue lines) at a rank score of 4 (thinner lines) or 5 (thicker lines). All four subunits of VOR (TK1978 - TK1981, purple rectangles in A) identified RBR1. All four subunits of POR (TK1978, TK1982 -TK1984, dark purple rectangles in B) identified RBR2. Additionally, FDH γ_3 (TK2077, fuchsia rectangle), FDHγ_4 (TK2078, fuchsia rectangle), and GGR (TK1088, blue rectangle) identified RBR2 (B). Three subunits of OGOR (TK1123 – TK1129, yellow rectangles) and MBS-L (TK1215, red rectangle) identified RBR3 (C).

RBR1 and RBR2 were tagged in this study, but the tagged RBRs identified different interacting partners. For instance, tagged-FDHγ_3 and tagged-FDHγ_4 proteins identified RBR2, however tagged-RBR1 (and not tagged-RBR2) identified FHDγ_4 and FDHα. Additionally, no subunits of POR or VOR were identified when either RBR1 or RBR2 was tagged.

We propose that the RBRs are employed as facilitators of metabolic protection, providing support for oxidative stress observed especially in lab cultivation where nutrient availability is high and contamination from oxygen stress is nearly impossible to avoid completely. It would be ideal to test the activity and expression of RBRs in native environments, albeit unlikely, as access to hyperthermophilic vents and cultivation of sufficient biomass of pure cultures in poor nutrient conditions is limiting. However, considering that hyperthermophiles with unique metabolisms are of interest for bioproducts such as H_2 gas or isoprene lipids, knowledge of the expression and activity of RBRs may benefit industrial processes where strains are grown at great rates and efficiency is paramount for profit.

False positive interactions can sometimes be rationally identified but should not be discarded

Large data sets, such as those produced by AP-MS, struggle to remove all false positives. Replicates and stringent statistical thresholds aid in limiting false positive results, however rational evaluation of the data is appropriate and required. While all data are presented, there are certain associations which make little sense, such as the abundance of associations observed with the flagellin proteins. *T. kodakarensis* encodes fourteen flagellin or flagella-related proteins (TK0038 – TK0049, TK1852 – TK1853) which make up the motor and extracellular structure of the archaellum, a polar structure resembling flagella in many bacteria.[74] While the motor proteins associate in the membrane, the majority of the flagellin proteins are maintained extracellularly, thus association with metabolic components within the cell is unlikely. Out of the fifty-one AP-MS trials performed (twenty-five in each sulfur or non-sulfur conditions

and one in non-pyruvate conditions), the flagellin proteins were observed twenty-three times in sulfur and twenty-four times in non-sulfur with FDH, FNOR2, GGR, IOR, MBS, POR, VOR, RBR1, or RBR2. If these proteins are simply "sticky," they would presumably appear at equal rates in control samples and be discarded from the data (by not passing the fold change threshold of being 4X more abundant in the tagged sample – see Chapter 3.3.3), however this was not the case. It is possible that several flagellin proteins associated with the amino acids of the tag itself; the six histidines or nine amino acids which compose the HA sequence (YPYDVPDYA) may confer sufficient attractive forces to maintain association with the flagellin proteins. While little sense can be made of the abundance of flagellin proteins identified, they can be considered false positives until a better understanding is determined.

Generalized stats of AP-MS data and overrepresented proteins may form central hubs

T. kodakarensis encodes ~ 2,300 proteins, of which 436 were identified as significant in this study (235 in both +S˚ and -S˚, 120 in +S˚ only, and 57 in -S˚ only). On average, each tagged protein identified ~30 interacting partners, just slightly over 1% of all proteins. 151 proteins were only observed once, while many proteins were observed multiple times.

Some proteins were overrepresented in the data which could suggest that they form hubs or central connectivity for metabolic protein networks. Proteins which were observed at least seven times in sulfur or non-sulfur conditions are noted in Table 4.26. Interestingly, three endonucleases (TK2215, TK1165, and TK0773) were observed fourteen, sixteen, and seventeen times in the data, respectively, with TK0773 present more so in sulfur conditions and TK1165 and TK2215 present more so in non-sulfur conditions. No other proteins were observed more than seventeen times.

Table 4.29: Highly identified proteins by gene number in sulfur and non-sulfur conditions.

Prey Gene No.	Annotation	Count in +S°	Count in -S°	Total
TK0773	Predicted ATP-dependent endonuclease, OLD family	3	14	17
TK1165	Predicted AP endonuclease	10	6	16
TK0211	Amidophosphoribosyltransferase	7	9	16
TK0826	Rubrerythrin domain-containing protein	7	8	15
TK0930	Uncharacterized protein	7	8	15
TK2076	Probable formate dehydrogenase, alpha subunit	10	4	14
TK2215	tRNA-splicing endonuclease	10	4	14
TK0844	Aldehyde ferredoxin oxidoreductase	9	5	14
TK1903	DNA polymerase II large subunit	8	6	14
TK0789	Glycerol-1-phosphate dehydrogenase [NAD(P)+]	9	4	13
TK2093	Membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	7	6	13
TK0483	Uncharacterized protein	9	3	12
TK1509	Probable tRNA pseudouridine synthase B	9	3	12
TK1603	V-type ATP synthase beta chain	7	5	12
TK1978	Pyruvate/ketoisovalerate oxidoreductases common subunit gamma	6	6	12
TK0136	Indolepyruvate oxidoreductase subunit IorA	4	8	12
TK1299	NAD(P)H sulfur oxidoreductase (CoA-dependent)	4	8	12
TK0799	Type 2 DNA topoisomerase 6 subunit B	9	2	11
TK0141	UPF0173 metal-dependent hydrolase TK0141	8	3	11
TK1394	Uncharacterized protein	7	4	11
TK1736	Vitamin B12-dependent ribonucleotide reductase	7	4	11
TK1980	2-oxoisovalerate:ferredoxin oxidoreductase, alpha subunit	7	4	11
TK1800	ABC-type dipeptide/oligopeptide transport system, ATPase component	4	7	11
TK1859	Probable lipoprotein releasing system, ATP-binding protein	7	3	10
TK0528	Serine hydroxymethyltransferase	7	2	9
TK0953	Predicted ATPase, AAA superfamily, containing PIN and KH nucleic acid-binding domains	7	2	9
TK0190	GMP synthase [glutamine-hydrolyzing] subunit A	0	9	9
TK1305	Probable translation initiation factor IF-2	7	1	8
TK2160	Radical_SAM domain-containing protein	7	1	8
TK0944	Acetate--CoA ligase (ADP-forming)	1	7	8
TK1415	50S ribosomal protein L12	1	7	8
TK2016	Iron-molybdenum cofactor-binding protein	7	0	7
TK2106	Enolase	7	0	7

Of the proteins which were observed frequently, only a handful were observed ~evenly between sulfur and non-sulfur conditions, while the majority were observed with more frequency in one condition over the other. Such shifts suggest that extensive re-programming occurs in the two metabolic conditions studied, supporting previous findings.[27,28,31]

REFERENCES

1. Atkinson, J. T., Campbell, I., Bennett, G. N. & Silberg, J. J. Cellular Assays for Ferredoxins: A Strategy for Understanding Electron Flow through Protein Carriers That Link Metabolic Pathways. *Biochemistry* **55**, 7047–7064 (2016).
2. Burkhart, B. W., Febvre, H. P. & Santangelo, T. J. Distinct physiological roles of the three ferredoxins encoded in the hyperthermophilic archaeon thermococcus kodakarensis. *mBio* **10**, (2019).
3. Menon, A. L. *et al.* Novel multiprotein complexes identified in the hyperthermophilic archaeon pyrococcus furiosus by non-denaturing fractionation of the native proteome. *Molecular and Cellular Proteomics* **8**, 735–751 (2009).
4. Rohila, J. S. *et al.* Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. *Plant J* **46**, 1–13 (2006).
5. Zhang, Y., Gao, P. & Yuan, J. S. *Plant Protein-Protein Interaction Network and Interactome*. *Current Genomics* vol. 11 (2010).
6. Babu, M., Kagan, O., Guo, H., Greenblatt, J. & Emili, A. Identification of protein complexes in Escherichia coli using sequential peptide affinity purification in combination with tandem mass spectrometry. *Journal of Visualized Experiments* (2012) doi:10.3791/4057.
7. Michaelis, A. C. *et al.* The social and structural architecture of the yeast protein interactome. *Nature* (2023) doi:10.1038/s41586-023-06739-5.
8. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623–627 (2000).
9. Li, Z., Santangelo, T. J., Čuboňová, L., Reeve, J. N. & Kelman, Z. Affinity purification of an archaeal DNA replication protein network. *mBio* **1**, (2010).
10. Lee, S. H., Kim, M. S., Kang, S. G. & Lee, H. S. Biohydrogen production of obligate anaerobic archaeon Thermococcus onnurineus NA1 under oxic conditions via overexpression of frhAGB-encoding hydrogenase genes. *Biotechnol Biofuels* **12**, (2019).
11. Kanai, T. *et al.* Continuous hydrogen production by the hyperthermophilic archaeon, Thermococcus kodakaraensis KOD1. *J Biotechnol* **116**, 271–282 (2005).
12. Aslam, M. *et al. Engineering of the Hyperthermophilic Archaeon Thermococcus Kodakarensis for Chitin-Dependent Hydrogen Production*. https://doi.org/10 (2017).
13. Verhaart, M. R. A., Bielen, A. A. M., Van Der Oost, J., Stams, A. J. M. & Kengen, S. W. M. Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: Mechanisms for reductant disposal. *Environmental Technology* vol. 31 993–1003 Preprint at https://doi.org/10.1080/09593331003710244 (2010).
14. Lee, C. M., Adamchek, C., Feke, A., Nusinow, D. A. & Gendron, J. M. Mapping protein–protein interactions using affinity purification and mass spectrometry. in *Methods in Molecular Biology* vol. 1610 231–249 (Humana Press Inc., 2017).
15. Atomi, H., Fukui, T., Kanai, T., Morikawa, M. & Imanaka, T. Description of Thermococcus kodakaraensis sp. nov., a well studied hyperthermophilic archaeon previously reported as Pyrococcus sp. KOD1. *Archaea* **1**, 263–267 (2004).
16. Santangelo, T. J., Čuboňová, L. & Reeve, J. N. Thermococcus kodakarensis Genetics: Tk1827-encoded β-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl Environ Microbiol* **76**, 1044–1052 (2010).
17. Fukui, T. *et al.* Complete genome sequence of the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1 and comparison with Pyrococcus genomes. *Genome Res* **15**, 352–363 (2005).
18. Schut, G. J., Brehm, S. D., Datta, S. & Adams, M. W. W. Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: Pyrococcus furiosus grown on carbohydrates or peptides. in *Journal of Bacteriology* vol. 185 3935–3947 (2003).

19. Isobe, K. *et al.* Geranylgeranyl reductase and ferredoxin from methanosarcina acetivorans are required for the synthesis of fully reduced archaeal membrane lipid in escherichia coli cells. *J Bacteriol* **196**, 417–423 (2014).
20. Santangelo, T. J., Čuboňová, L. & Reeve, J. N. Deletion of alternative pathways for reductant recycling in Thermococcus kodakarensis increases hydrogen production. *Mol Microbiol* **81**, 897–911 (2011).
21. Mayer, F. & Müller, V. Adaptations of anaerobic archaea to life under extreme energy limitation. *FEMS Microbiol Rev* **38**, 449–472 (2014).
22. Truong, A., Myerscough, D., Campbell, I., Atkinson, J. & Silberg, J. J. A cellular selection identifies elongated flavodoxins that support electron transfer to sulfite reductase. *Protein Science* **32**, (2023).
23. Lubner, C. E. *et al.* Mechanistic insights into energy conservation by flavin-based electron bifurcation. *Nat Chem Biol* **13**, 655–659 (2017).
24. Schäfer, G., Engelhard, M. & Müller, V. Bioenergetics of the Archaea. *Microbiology and Molecular Biology Reviews* **63**, 570–620 (1999).
25. Sapra, R., Bagramyan, K. & Adams, M. W. W. A simple energy-conserving system: Proton reduction coupled to proton translocation. *Proc Natl Acad Sci U S A* **100**, 7545–7550 (2003).
26. Mayer, F. & Müller, V. Adaptations of anaerobic archaea to life under extreme energy limitation. *FEMS Microbiol Rev* **38**, 449–472 (2014).
27. Lipscomb, G. L., Schut, G. J., Scott, R. A. & Adams, M. W. W. SurR is a master regulator of the primary electron flow pathways in the order Thermococcales. *Mol Microbiol* **104**, 869–881 (2017).
28. Lipscomb, G. L. *et al.* SurR: A transcriptional activator and repressor controlling hydrogen and elemental sulphur metabolism in Pyrococcus furiosus. *Mol Microbiol* **71**, 332–349 (2009).
29. Yang, H. *et al.* SurR regulates hydrogen production in Pyrococcus furiosus by a sulfur-dependent redox switch. *Mol Microbiol* **77**, 1111–1122 (2010).
30. Nguyen, D. M. N. *et al.* Two functionally distinct NADP+-dependent ferredoxin oxidoreductases maintain the primary redox balance of Pyrococcus furiosus. *Journal of Biological Chemistry* **292**, 14603–14616 (2017).
31. Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J. & Reeve, J. N. Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis. *BMC Genomics* **15**, (2014).
32. Hachisuka, S. I., Sato, T. & Atomi, H. Metabolism dealing with thermal degradation of NAD+ in the hyperthermophilic archaeon Thermococcus kodakarensis. *J Bacteriol* **199**, 1–14 (2017).
33. Hachisuka, S., Sato, T. & Atomi, H. Hyperthermophilic Archaeon Thermococcus kodakarensis Utilizes a Four-Step Pathway for NAD+ Salvage through Nicotinamide Deamination. *J Bacteriol* **200**, 1–12 (2018).
34. Russell, R. B. *et al.* A structural perspective on protein-protein interactions. *Curr Opin Struct Biol* **14**, 313–324 (2004).
35. Van Leene, J. *et al.* A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana. *Mol. Cell Proteomics* **6**, 1226–1238 (2007).
36. Meyer, K. & Selbach, M. Quantitative affinity purification mass spectrometry: A versatile technology to study protein-protein interactions. *Frontiers in Genetics* vol. 6 Preprint at https://doi.org/10.3389/fgene.2015.00237 (2015).
37. Mai, X. & Adams, M. W. W. Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon Pyrococcus furiosus. A new enzyme involved in peptide fermentation. *Journal of Biological Chemistry* **269**, 16726–16732 (1994).

38. Mai, X. & Adams, M. W. W. *Characterization of a Fourth Type of 2-Keto Acid-Oxidizing Enzyme from a Hyperthermophilic Archaeon: 2-Ketoglutarate Ferredoxin Oxidoreductase from Thermococcus Litoralis*. *JOURNAL OF BACTERIOLOGY* vol. 178 https://journals.asm.org/journal/jb (1996).
39. Heider, J., Mai, X. & Adams, M. W. W. *Characterization of 2-Ketoisovalerate Ferredoxin Oxidoreductase, a New and Reversible Coenzyme A-Dependent Enzyme Involved in Peptide Fermentation by Hyperthermophilic Archaea*. *JOURNAL OF BACTERIOLOGY* vol. 178 (1996).
40. Schut~, G. J., Menon, A. L. & Adams, M. W. W. *2-Keto Acid Oxidoreductases from Pyrococcus and Thermococcus Litoralis*. (2001).
41. Schut~, G. J., Menon, A. L. & Adams, M. W. W. *ENZYMES OF CENTRAL METABOLISM [12]*. (2001).
42. Blarney, J. M. & Adams, M. W. W. *Purification and Characterization of Pyruvate Ferredoxin Oxidoreductase from the Hyperthermophilic Archaeon Pyrococcus Furiosus*. *Biochimica et Biophysica Acta* (1993).
43. Smith, E. T., Blarney, J. M. & Adams, M. W. W. Pyruvate Ferredoxin Oxidoreductases of the Hyperthermophilic Archaeon, Pyrococcus furiosus, and the Hyperthermophilic Bacterium, Thermotoga marítima, Have Different Catalytic Mechanisms'* 1".
44. Coulter, E. D., Shenvi, N. V & Kurtz, D. M. *NADH Peroxidase Activity of Rubrerythrin*. *Biochemical and Biophysical Research Communications* vol. 255 http://www.idealibrary.comon (1999).
45. Bräsen, C., Esser, D., Rauch, B. & Siebers, B. Carbohydrate metabolism in archaea: Current insights into unusual enzymes and pathways and their regulation. *JAMA Ophthalmol* **132**, 326–331 (2014).
46. Matsubara, K., Yokooji, Y., Atomi, H. & Imanaka, T. Biochemical and genetic characterization of the three metabolic routes in Thermococcus kodakarensis linking glyceraldehyde 3-phosphate and 3-phosphoglycerate. *Mol Microbiol* **81**, 1300–1312 (2011).
47. Kim, Y. J. *et al.* Formate-driven growth coupled with H2 production. *Nature* **467**, 352–355 (2010).
48. Kobayashi, A. *et al.* Structure and function relationship of formate dehydrogenases: an overview of recent progress. *IUCrJ* **10**, (2023).
49. Yang, J. I., Lee, S. H., Ryu, J. Y., Lee, H. S. & Kang, S. G. A Novel NADP-Dependent Formate Dehydrogenase From the Hyperthermophilic Archaeon Thermococcus onnurineus NA1. *Front Microbiol* **13**, (2022).
50. Ma, K. & Adams, M. W. W. *Sulfide Dehydrogenase from the Hyperthermophilic Archaeon Pyrococcus Furiosus: A New Multifunctional Enzyme Involved in the Reduction of Elemental Sulfur*. *JOURNAL OF BACrERIOLOGY* vol. 176 https://journals.asm.org/journal/jb (1994).
51. Hidese, R. *et al.* Gene regulation of two ferredoxin:NADP+ oxidoreductases by the redox-responsive regulator SurR in Thermococcus kodakarensis. *Extremophiles* **21**, 903–917 (2017).
52. Murakami, M. *et al.* Geranylgeranyl reductase involved in the biosynthesis of archaeal membrane lipids in the hyperthermophilic archaeon Archaeoglobus fulgidus. *FEBS Journal* **274**, 805–814 (2007).
53. Yu, H. *et al.* Structure of the respiratory MBS complex reveals iron-sulfur cluster catalyzed sulfane sulfur reduction in ancient life. *Nat Commun* **11**, (2020).
54. Yu, H. *et al.* Structure of an Ancient Respiratory System. *Cell* **173**, 1636-1649.e16 (2018).
55. Sato, T. & Atomi, H. Novel metabolic pathways in Archaea. *Current Opinion in Microbiology* vol. 14 307–314 Preprint at https://doi.org/10.1016/j.mib.2011.04.014 (2011).

56. Yokooji, Y., Sato, T., Fujiwar, S., Imanak, T. & Atomi, H. Genetic examination of initial amino acid oxidation and glutamate catabolism in the hyperthermophilic archaeon Thermococcus kodakarensis. *J Bacteriol* **195**, 1940–1948 (2013).
57. Adams, M. W. W. & Kletzin', A. *OXIDOREDUCTASE-TYPE ENZYMES AND REDOX PROTEINS INVOLVED IN FERMENTATIVE METABOLISMS OF HYPERTHERMOPHILIC ARCHAEA*. (1996).
58. Weinberg, M. V., Jenney, F. E., Cui, X. & Adams, M. W. W. Rubrerythrin from the hyperthermophilic archaeon Pyrococcus furiosus is a rubredoxin-dependent, iron-containing peroxidase. *J Bacteriol* **186**, 7888–7895 (2004).
59. Price, M. N., Huang, K. H., Alm, E. J. & Arkin, A. P. A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res* **33**, 880–892 (2005).
60. Ermolaeva, M. D., White, O. & Salzberg, S. L. *Prediction of Operons in Microbial Genomes*. *Nucleic Acids Research* vol. 29 www.tigr.org/tigr-scripts/operons/operons.cgi. (2001).
61. Mukund, S. & Adams, M. W. W. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon Pyrococcus furiosus. *Journal of Biological Chemistry* **270**, 8389–8392 (1995).
62. Kanai, T. *et al.* A global transcriptional regulator in Thermococcus kodakaraensis controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes. *Journal of Biological Chemistry* **282**, 33659–33670 (2007).
63. Scott, K. A., Williams, S. A. & Santangelo, T. J. Thermococcus kodakarensis provides a versatile hyperthermophilic archaeal platform for protein expression. in *Methods in Enzymology* vol. 659 243–273 (Academic Press Inc., 2021).
64. Hutchins, A. M., Mai, X. & Adams, M. W. W. *Acetyl-CoA Synthetases I and II from Pyrococcus Furiosus*. (2001).
65. Nielsen, C. F., Lange, L. & Meyer, A. S. Classification and enzyme kinetics of formate dehydrogenases for biomanufacturing via CO2 utilization. *Biotechnology Advances* vol. 37 Preprint at https://doi.org/10.1016/j.biotechadv.2019.06.007 (2019).
66. Bae, S. S. *et al.* Enhancing bio-hydrogen production from sodium formate by hyperthermophilic archaeon, Thermococcus onnurineus NA1. *Bioprocess Biosyst Eng* **38**, 989–993 (2015).
67. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc Natl Acad Sci U S A* **74**, 5088–5090 (1977).
68. Woese, C. R., Kandlert, O. & Wheelis, M. L. *Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya (Euryarchaeota/Crenarchaeota/Kingdom/Evolution)*. *Proc. Nati. Acad. Sci. USA* vol. 87 (1990).
69. Jain, S., Caforio, A. & Driessen, A. J. M. Biosynthesis of archaeal membrane ether lipids. *Frontiers in Microbiology* vol. 5 Preprint at https://doi.org/10.3389/fmicb.2014.00641 (2014).
70. Schicho, R. N., Ma, K., Adams, M. W. W. & Kelly3, R. M. *Bioenergetics of Sulfur Reduction in the Hyperthermophilic Archaeon Pyrococcus Furiosus*. *JOURNAL OF BACTERIOLOGY* vol. 175 https://journals.asm.org/journal/jb (1993).
71. Schut, G. J., Bridger, S. L. & Adams, M. W. W. Insights into the metabolism of elemental sulfur by the hyperthermophilic archaeon Pyrococcus furiosus: Characterization of a coenzyme A-dependent NAD(P)H sulfur oxidoreductase. *J Bacteriol* **189**, 4431–4441 (2007).
72. Roggiani, M. & Goulian, M. Chromosome-Membrane Interactions in Bacteria. *Annu Rev Genet* **49**, 115–129 (2015).

73. Dillard, B. D., Demick, J. M., Adams, M. W. W. & Lanzilotta, W. N. A cryo-crystallographic time course for peroxide reduction by rubrerythrin from Pyrococcus furiosus. *Journal of Biological Inorganic Chemistry* **16**, 949–959 (2011).
74. Jarrell, K. F., Albers, S. V. & de Sousa Machado, J. N. A comprehensive history of motility and Archaellation in Archaea. *FEMS Microbes* vol. 2 Preprint at https://doi.org/10.1093/femsmc/xtab002 (2021).

CHAPTER 5
REROUTING ELECTRON FLUX *IN VIVO* [4]

Summary

Iron-sulfur containing proteins, such as ferredoxins (Fds), are known to transport electrons[1–4] and offer the potential to regulate electron flux by manipulating the expression of each ferredoxin isoform or manipulating the interactions of specific Fds to allow selective or preferred redox reactions *in vivo*.[5,6] The hyperthermophilic, anaerobic archaeon *Thermococcus kodakarensis* encodes three distinct Fds which have been shown to make specific and distinct protein-partnerships with unique metabolic enzymes.[5] Selective Fd-protein interactions can be used to bridge reductive flow from catabolic reactions to the membrane, providing a "proteinaceous electron highway" for efficient electron shuttling. Specific redox protein partnerships have been shown to adapt to changing physiological conditions,[7–11] suggesting that such are tunable and can provide a level of selectivity not possible with small molecule electron transport.

T. kodakarensis energy production strategies result in four common reductive sinks: the generation of fully saturated lipids, the reduction of NAD(P), the reduction of protons to H_2 gas, and the reduction of polysulfides to H_2S gas.[12] H_2 gas is a valuable biofuel and a number of products are synthesized from isoprene lipids, thus the regulation of electrons towards these reductive sinks is of particular interest.[13–19] Further, lipids are an abundant store of reductant thus directing electrons away from these abundant hydrocarbon chains and towards other pathways could redirect electron flux towards other bioproducts.[20,21] We propose that regulated reductive relay networks can be rationally manipulated to boost the production of mature lipids and/or H_2 gas.

[4] Data presented in this chapter in part represents the efforts of talented students and/or colleagues under the tutelage of Seré Williams including: Danielle Riley, Teagan Rockwood, David Crosby, Kate Call, and Savina Miller.

By placing a Fd in direct contact with the protein responsible for each reductant sink of interest we propose that the exchange of electrons through these facilitated partnerships will modify reductive flux. We have generated novel strains in which one of three ferredoxins are tethered (physically linked) to either geranylgeranyl reductase (GGR) which is responsible for lipid maturation[22] or to one of two cytosolic components of the membrane bound hydrogenase,[23] the respiratory system responsible for H_2 production. By measuring fitness outcomes via growth assessments, lipid profiles, and hydrogen generation, we can determine if the constant presence of a Fd and a proposed electron acceptor can modulate reductive flux.

We generated thirteen fusion strains wherein a Fd is fused to MBH or GGR. H_2 gas generation increased when Fd was tethered to MBH, with a ~45% increase in H_2 production per cell when Fd3 was tethered to MBH-J with a short, flexible linker sequence. It was determined that tethering Fd2 to GGR was viable and resulted in a ~40% reduction in lipid saturation.

5.1 Introduction

Electron carriers are necessary to shuttle the electrons produced from glucose and amino acid fermentation reactions to their corresponding electron acceptors. At the optimal growing temperatures of *T. kodakarensis* (85º C), small molecule electron carriers such as $NAD(H)^+$ degrade into ADP-ribose which can spontaneously bind to proteins at elevated temperatures resulting in a toxic byproduct.[24,25] *T. kodakarensis* encodes an ADP-ribose pyrophosphatase (TK2284) which aids in managing ADP-ribose toxicity, but not without cost to the cell.[25,26] Thus, proteinaceous electron carriers provide an alternative means of electron transport which may be especially useful in hyperthermophiles. Many organisms, and essentially all hyperthermophilic species encode proteinaceous electron carriers, classically termed ferredoxins (Fds), but proteinaceous electron carriers also include flavodoxins, thioredoxins, rubredoxins and many cytochrome C proteins (Figure 5.1).[2,6] *T. kodakarensis* utilizes three ferredoxins (Fds) which are not only able to better withstand high temperatures but

are one of a few electron carriers which are able to reduce protons due to maintaining exceptionally low reductions potentials (Figure 5.1).[6,27]

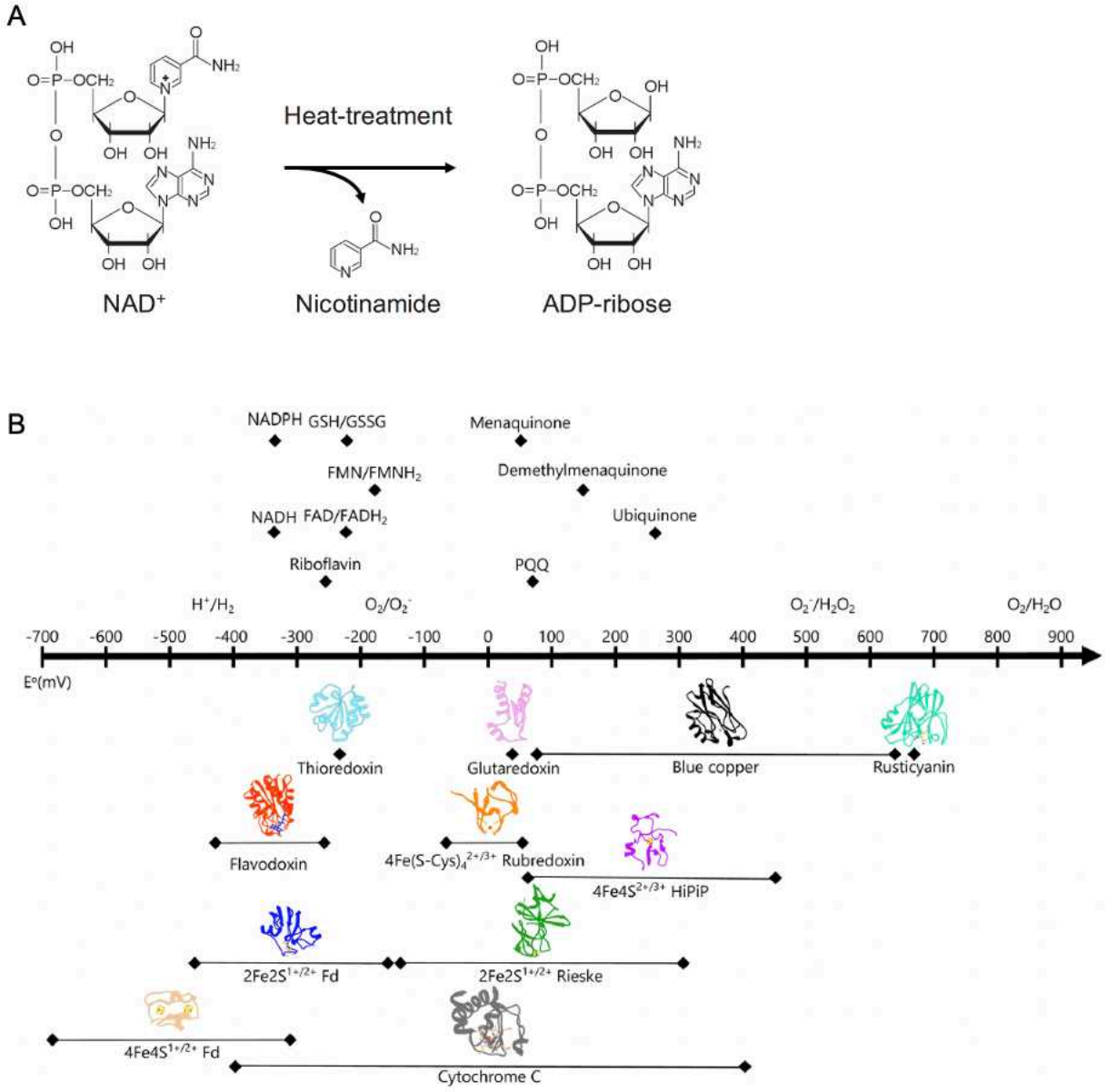
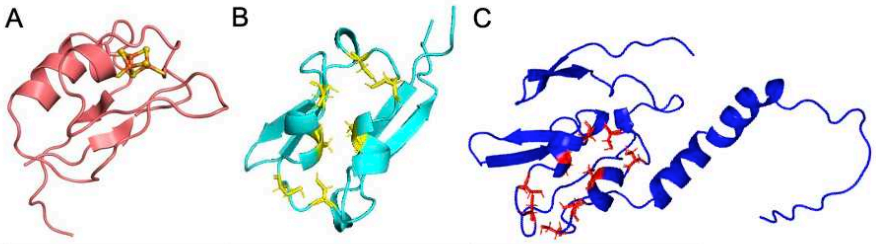


Figure 5.1: NAD+ denaturation and a comparison of reductive potentials of small molecule and protein electron carriers (PECs). NAD^+ spontaneously degrades into nicotinamide and ADP-ribose upon heat-treatment (A), conferring toxicity in hyperthermophiles. Midpoint reduction potentials of small molecule and PECs display a range of potentials (E°(mV)). Note that neither NADH nor NADHP maintain a low enough reduction potential to reduce a proton to H_2, which requires ~ -400 mV. Flavodoxin, Fds, and Cytochrome C display low enough reduction potentials to reduce H^+ and generate H_2. (Adapted from Hachisuka *et al.*, 2017 and Atkinson *et al.*, 2016).

5.1.1 The ferredoxins as valuable proteinaceous electron carriers in *T. kodakarensis*

Fds are PECs of particular interest in the generation of H_2 gas, as they maintain a variety of combinations of metal centers which confer reduction potentials low enough to reduce H^+ to H_2.[14,28,29] *T. kodakarensis* encodes three Fds (Fd1, TK1694; Fd2, TK1087; and Fd3 TK2012) which maintain unique features (Figure 5.2).[2,3,5]



Protein	Fd1	Fd2	Fd3
Mass (kDa)	6.5	7.2	15.6
pI	3.7	5.1	8.4
E_m potential	-407 mV		-409 mV and -363 mV
Essential?	Yes	Yes	No
Abundance, +S°	High	Moderate	Low
Abundance, -S°	High	Moderate	High

Figure 5.2: Ferredoxins in *T. kodakarensis*. Crystal structure of Fd1 to 1.1Å shown in (A) with a 4Fe/3S center (Fe in orange, S in yellow; PDB code: 1SIZ). AlphaFold2 models of Fd2 (B) and Fd3 (C) with cystine residues highlighted in yellow (B) or red (C).[30] A summary of physical properties, essentiality, reduction potential, and relative abundance in +/- S° identifies unique characteristics of the three Fd proteins. At ~7 kDa, Fd1 and Fd2 are small proteins. While Fd1 maintains a lower pI and is expressed at a higher abundance, both are essential. Fd3 is slightly larger and can be deleted from the genome. Reduction potentials for Fd1 using cyclic voltammetry measured a one electron transfer at -407 mV. Two, one electron transfers were measured for Fd3 at -409 mV and -363 mV. (Adapted from Burkhart *et al*., 2019 and M. Stettler's thesis, 2022).

Fd1 and Fd2 are relatively similar in size (6.5 kDa and 7.2 kDa respectively), both have acidic pIs (3.7 and 5.1, respectively) and are predicted to share somewhat similar spatial geometry. The main differences between Fd1 and Fd2 can be seen in their minimal primary sequence beyond important metal-binding residues as well as their distinct electrostatic surface potentials. Fd3 is by far the most distinctive of the three ferredoxins, exhibiting a significantly larger size (15.6 kDa) and a basic pI (8.4). Fd3 is also the only non-essential Fd, while Fd1 and Fd2 cannot be deleted and are thus essential for cell viability.[31]

The diversity of Fds within the biosphere has required that a specific schema be developed which provides useful classification of these diverse Fe-S containing proteins. Such a system has landed on classifying Fds by cluster type (e.g. 2Fe-2S, 2Fe-3S, 4Fe-4S, etc.) which have varying ranges of reduction potentials (Figure 5.3).[2,6] Fd1 has been crystalized and while the model predicted a 4Fe/4S center, Fd1 was observed to maintain a 4Fe/3S center (Figure 5.3A). Theoretical protein structures determined for Fd2 and Fd3 (Figure 5.3B and C), such *in silico* predictions cannot confirm cluster composition.[30] Crystallization trials for *Tk*-Fd2 and *Tk*-Fd3 and voltammetry trials for *Tk*-Fd2 are underway.

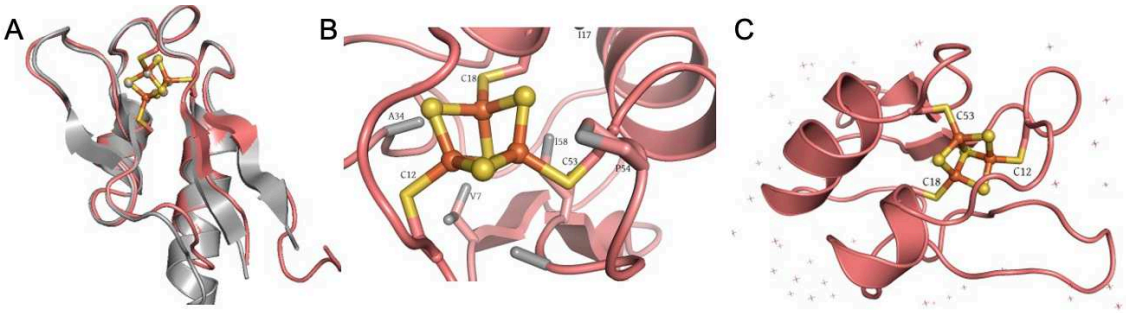


Figure 5.3: The crystal structure of *Tk*-Fd1. Crystalized Fd1 from *T. kodakarensis* (salmon) and *Pyrococcus furiosis* (grey) are aligned showing extreme homology at the 3Fe/4S fold and less homology at the C- and N-termini (A). The 3Fe (orange) and 4S (yellow) center and coordinating residues (C12, C18, and C53) of Fd1 is shown in (B). The full protein structure is shown in (C) with coordinating water molecules surrounding Fd1. (Adapted from M. Stettler's thesis, 2022).

Protein-based electron carriers are more thermotolerant than $NAD(P)^+$ and universally coordinate metal centers that permit electron shuttling. Additional advantages have been long predicted, as coordinated production of a specific Fd could, in principle, permit selective transport of electrons from particular electron donors to specific electron acceptors. Selective electron transport could provide routes for electron flux that favor individual reactions without demanding cellular-level shifts in redox balancing (Figure 5.4).

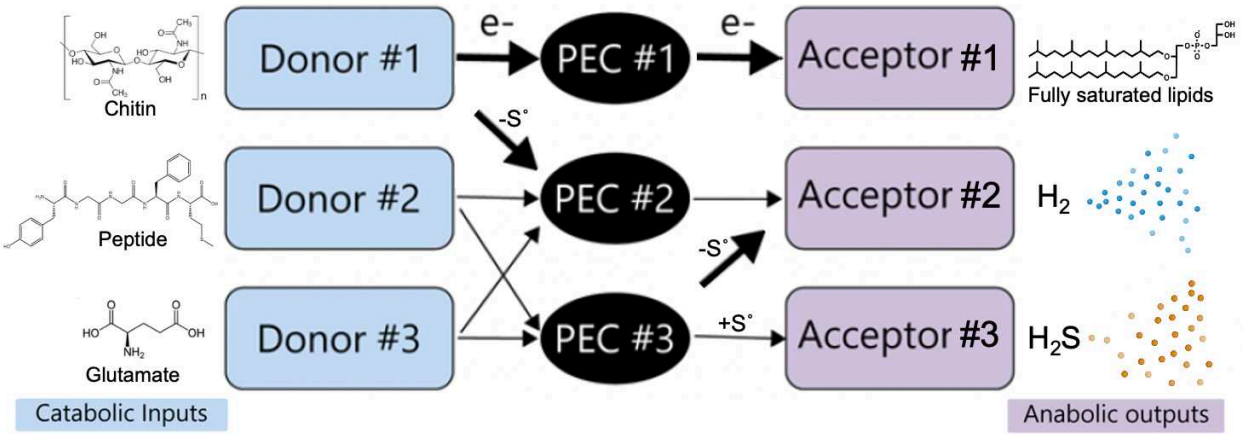


Figure 5.4: Electron transfer pathways between PECs with preferred electron donors and electron acceptors. Proteinaceous electron carriers (PECs, black ellipses at center) will associate with specific electron donors (light blue rectangles at left) and electron acceptors (light purple rectangles at right) such that electron flux may be regulated towards one anabolic outcome over another. In the case of chitin (a catabolic input at left), Donor #1 will associate with PEC #1, wherein PEC#1 will become reduced and associate with Acceptor #1. Acceptor #1 will transfer the electron to an unsaturated lipid moiety, further saturating the lipid chain. In the absence of sulfur, the electrons which would otherwise be directed to PEC #1 are directed to PEC #2. When PEC #2 is reduced, PEC #2 preferentially associates with Acceptor #2 leading to the generation of hydrogen gas. A similar regulatory pathway is depicted in the bottom row, but instead of the donor and PEC responding to the presence or absence of sulfur, the association between the PEC and the acceptor is responsive to the presence or absence of sulfur. In this way, PECs provide regulated and adaptable electron transfer. Mapping specific metabolic Donor:PEC:Acceptor relationships is of interest for directing electrons to valuable bioproducts such as isoprene lipids which can be extracted and used for the manufacturing of perfumes, for example, or towards H_2 which is a valuable biofuel. (Adapted from Atkinson *et al*., 2016.)

5.1.2 Ferredoxins may provide a means to regulate electron flux to distinct reduced products

Burkhart *et al*. (2019) showed that the three Fds have unique interacting partners (Figure 5.5).[5] Fd1 was found to interact with multiple catabolic protein complexes involved in amino acid fermentation including ketoisovalerate or 2-oxo-isovalerate oxidoreductase (VOR), indolepyruvate oxidoreductase (IOR), and 2-oxoacid:ferredoxin oxidoreductase (OGOR) as well as pyruvate oxidoreductase (POR) which converts the product of glycolysis, pyruvate, into acetyl-CoA and CO_2.[32–38] Fd2 interacted with IOR, OGOR, and POR, but was not found in association with VOR, while Fd3 was only found with VOR and POR. Energetic gains can further be realized from the conversion of Acyl-CoA to formate by the reversable action of

acetyl-CoA synthase (ACS).[39,40] Formate dehydrogenase (FDH) can further reduce a ferredoxin and in the oxidation of formate to CO_2, and both Fd2 and Fd3 were found in association with FDH.

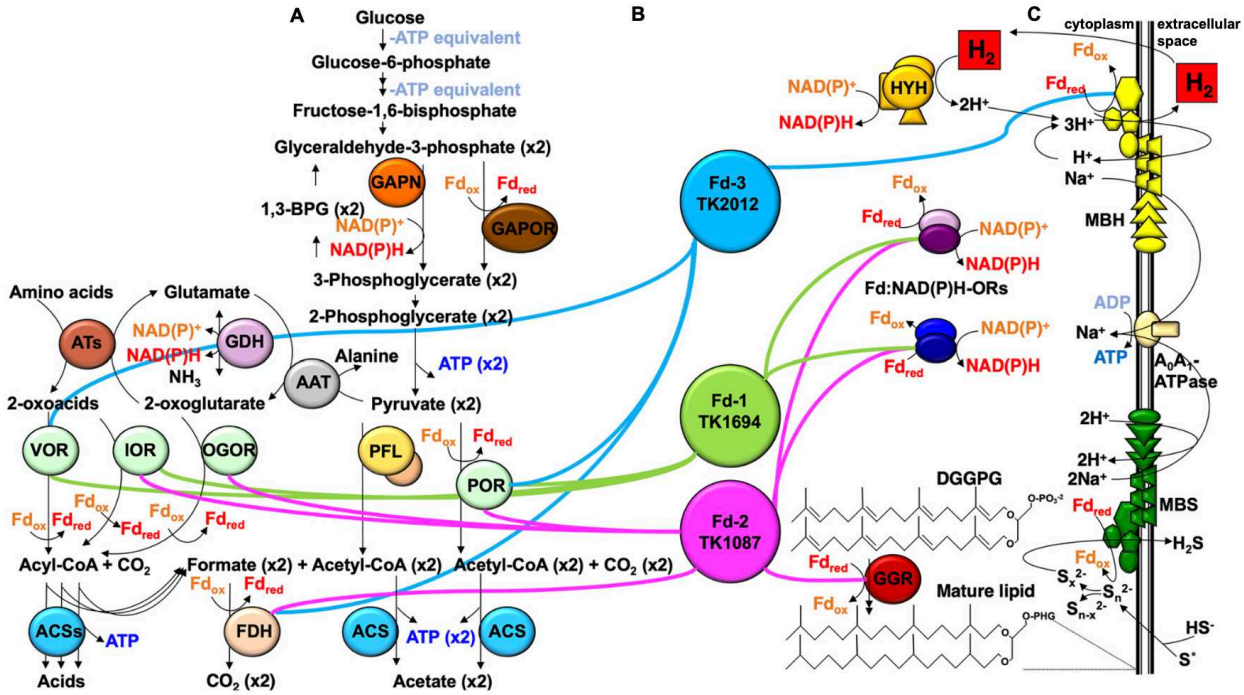


Figure 5.5: Ferredoxins link catabolic and anabolic processes in *T. kodakarensis* through distinct partnerships. Glucose and amino acids are fermented to generate minimal ATP, pyruvate, and acyl-CoA (A).[4,41] In addition, cofactors such as $NAD(P)^+$ and PECs such as Fd_{ox} are reduced to NAD(P)H and Fd_{red}, respectively. The three Fds encoded in the *T. kodakarensis* genome (Fd-1, TK1694; Fd-2, TK1087; and Fd-3, TK2012) displayed distinct associations with electron donors and electron acceptors (B, colored lines linking Fds with respective interacting partners). Fd1 appears to primarily be involved in the generation of reduced cofactors, while Fd2 interacts tightly with GGR resulting in the maturation of isoprenoid lipids. *T. kodakarensis* maintains two membrane bound respiratory complexes (C), membrane bound hydrogenase (MBH, yellow) or membrane bound sulfane reductase (MBS, green) which generate H_2 and H_2S, respectively, creating a proton gradient which drives ATP synthase at the membrane (light yellow). Fd3 was the only Fd found in association with the MBH, suggesting Fd-3 is linked to H_2 production. (Adapted from Burkhart et al., 2019.)

Anabolic activities of the Fd interactome were distinct. The main role of Fd1 appears be in interacting with the two heterodimeric ferredoxin:NAD(P)H oxidoreductases (FNOR1 and FNOR2) in the generation of reduced small molecules.[10] In order to meet NAD(P)H demands, the two FNORs likely require a constant supply of electrons via Fd1, explaining the high expression levels and abundance of Fd1 in all conditions and growth phases.

While Fd2 showed some association with the FNORs, the primary anabolic interacting partner was the monomeric geranylgeranyl reductase (GGR, TK1088), responsible for lipid maturation.[5,42] These findings were further supported as GGR also identified Fd2 in this work (Chapter 4). Because lipid maturation is a known electron sink[12] and because GGR was only recovered with Fd2, Fd2 is likely the sole electron donor to the complex.

Fd3 is primarily involved in H_2 evolution and electrochemical gradient generation by MBH for ATP production, returning high confidence partnerships with the at least four of the MBH complex subunits (TK2088, TK2089, TK2090, and TK2091) as well as at least four subunits of the Na^+-dependent ATPase subunits (TK1596, TK1602, TK1603, and TK1604).[5]

We hypothesize that PECs can direct electrons to specific pathways, thereby increasing or decreasing the generation of certain products. When considering the various reductive products, a metabolic map can include PECs (Figure 5.6).

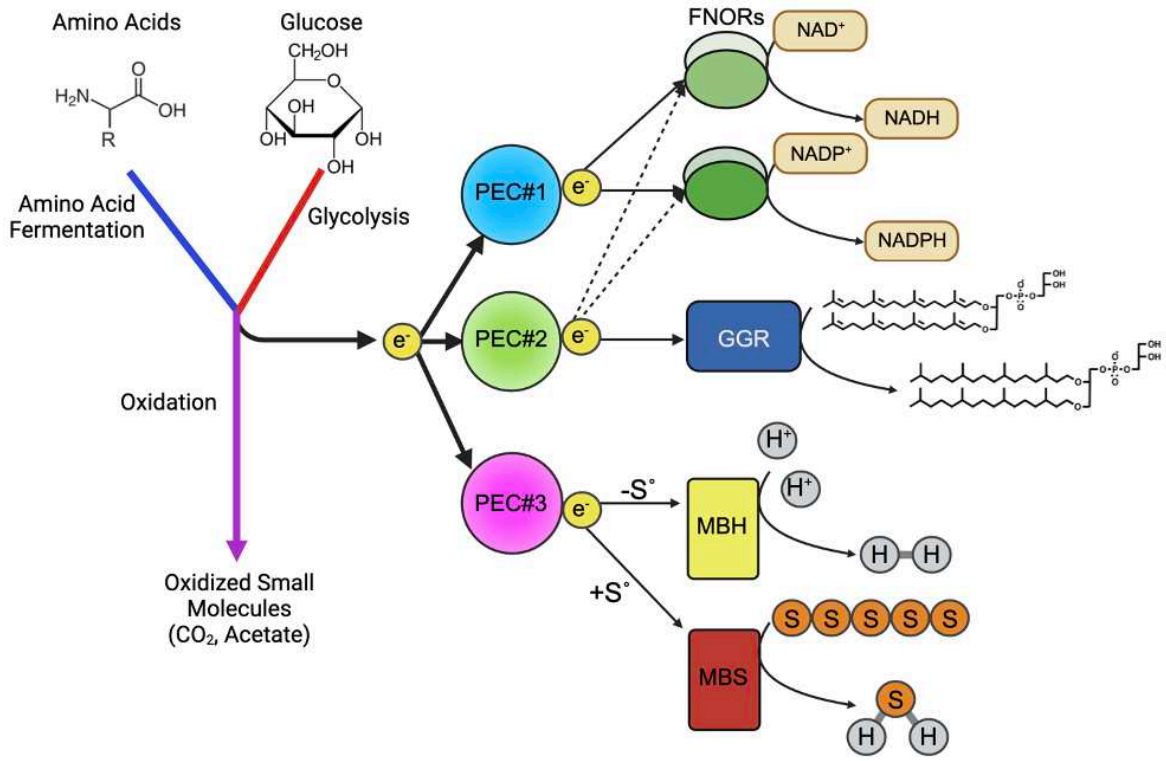


Figure 5.6: An overview of metabolic processes wherein PECs direct electron flux to specific reductant products. Amino acids and glucose are fermented to release electrons which are directed to distinct metabolic pathways via proteinaceous electron carriers (PECs). Here, PEC#1 (blue circle) directs electrons to the FNORs which generate reduced small

molecules such as NAD(P)H. PEC#2 (bright green circle) may also donate electrons to the FNORs, however, the primary electron acceptor is GGR (dark blue rectangle) which saturates isoprene lipids. PEC#3 (pink circle) is slightly more dynamic than PEC#1 or PEC#2 in that the electron acceptor complex varies depending on the presence or absence of sulfur, a terminal electron acceptor in *T. kodakarensis*. In the presence of sulfur (+S˚), PEC#3 interacts with the membrane bound sulfane reductase complex, thereby producing H_2S gas. In the absence of sulfur (-S˚), PEC#3 interacts with the MBH complex, reducing protons in the generation of H_2 gas. While association for the Fds are supported by previous findings (Burkhart et al., 2019), this model is hypothetical in nature and displayed here to illustrate the role PECs can play in directing electron flux to specific reductive pathways.

5.2 Tethering a ferredoxin to an electron acceptor protein to direct reductive flux

The distinct features and physiological role(s) of each of the Fds raises the question of whether they form specific and distinct interactions with specific electron donors and acceptors. The relationship of the Fds and MBH is of particular interest considering the value of H_2 as a biofuel,[14–17,43] thus assessing this relationship will be the focus of this chapter. Strong evidence suggests an important role between Fd2 and GGR in the production of mature lipids.[22,42] The role of Fd2 and GGR will be addressed in this chapter, however considerably more attention to this relationship is warranted than will be provided here. Lastly, Fd1 is abundantly expressed and essential, though data suggests its primary function is in the generation of reduced small molecules via the FNORs. While reduced small molecules serve an important role in metabolism and the specificity of interaction between the FNORs and Fds is of interest, it will not be discussed here except to direct an interested reader to T. Rockwood's thesis (2023) wherein Fd2 was tethered to the FNORs and ΔFd1 was attempted, to no avail. These findings suggesting that Fd2 cannot replace Fd1 in this role or that Fd1 is required in other cellular processes.

5.2.1 Tethering ferredoxins to the membrane bound hydrogenase complex

Due to its role in hydrogen gas generation, the partnership between a Fd and the MBH complex is of particular interest. Upon Fd oxidation, MBH evolves H_2 gas via a proton pump.[23] Protons are exchanged for sodium ions which drive the A_0A_1-ATPase. The cryo-EM structure of

fourteen-subunit MBH complex from the closely related archaeal hyperthermophile *Pyrococcus furiosos*[44,45] has been determined and maintains a membrane arm which contains both the Na^+ and proton pumps and a cytosolic arm which contains three, 4Fe4S centers and a NiFe cluster at the active site (Figure 5. 7 and Table 5.1).[23]

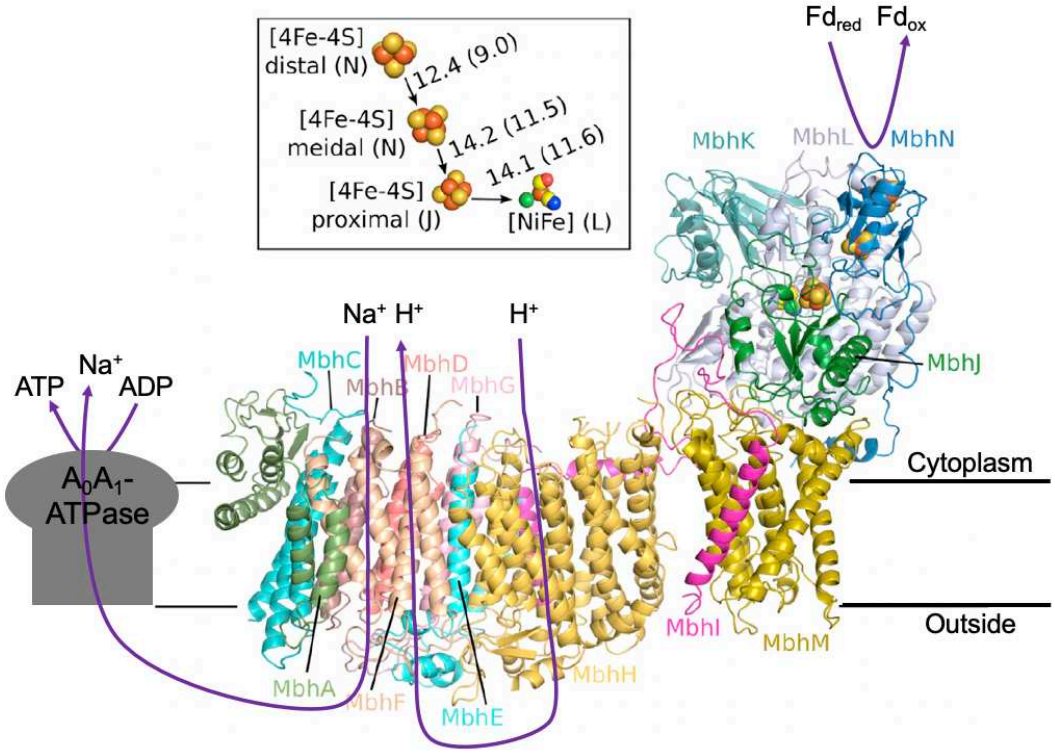


Figure 5.7: Structure of *Pf*-MBH complex. The structure of the fourteen-subunit MBH complex from *Pyrococcus furiosis* was determined by cryo-EM.[23] Ten subunits compose the membrane arm (MbhA-I and MbhM) while four subunits compose the cytosolic arm (MbhJ-L and Mbh-N). The membrane arm contains a proton pump and an proton/Na^+ antiporter which results in the generation of ATP via the Na^+-active A_0A_1-ATPase. The cytosolic arm contains three 4Fe/4S centers and one NiFe center which compose the redox sensitive active site. Upon reduction by Fd_{red}, a proton is translocated across the membrane. A proton is exchanged for a Na^+ ion, thereby generating a chemiosmotic gradient which drives ATP production. (Adapted from Yu *et al*., 2018.)

Table 5.1: Membrane bound hydrogenase gene numbers and annotations

TK Gene No.	PF Homolog	KEGG Annotation	GenBank Annotation	Published Annotation[a]	Membrane or Cytosolic Arm
TK2080	PF1423	multicomponent Na+:H+ antiporter subunit E	membrane bound hydrogenase, MbhA subunit	MBS-A	Membrane
TK2081	PF1424	multicomponent Na+:H+ antiporter subunit F	membrane bound hydrogenase, MbhB subunit	MBS-B	Membrane
TK2082	PF1425	multicomponent Na+:H+ antiporter subunit G	membrane bound hydrogenase, MbhC subunit	MBS-C	Membrane
TK2083	PF1426	no KO assigned	membrane bound hydrogenase, MbhD subunit	MBS-D	Membrane
TK2084	PF1427	no KO assigned	membrane bound hydrogenase, MbhE subunit	MBS-E	Membrane
TK2085	PF1428	multicomponent Na+:H+ antiporter subunit B	membrane bound hydrogenase, MbhF subunit	MBS-F	Membrane
TK2086	PF1429	multicomponent Na+:H+ antiporter subunit C	membrane bound hydrogenase, MbhG subunit	MBS-G	Membrane
TK2087	PF1430	multicomponent Na+:H+ antiporter subunit D	membrane bound hydrogenase, MbhH subunit	MBS-H	Membrane
TK2088	PF1431	no KO assigned	membrane bound hydrogenase, MbhI subunit	MBS-I	Membrane
TK2089	PF1432	membrane-bound hydrogenase subunit mbhJ	membrane bound hydrogenase, NiFe-hydrogenase small subunit	MBS-J	Cytosolic
TK2090	PF1433	membrane-bound hydrogenase subunit beta	membrane bound hydrogenase, NiFe-hydrogenase large subunit 1	MBS-K	Cytosolic
TK2091	PF1434	membrane-bound hydrogenase subunit alpha	membrane bound hydrogenase, NiFe-hydrogenase large subunit 2	MBS-L	Cytosolic
TK2092	PF1435	no KO assigned	membrane bound hydrogenase, MbhM subunit	MBS-M	Membrane
TK2093	PF1436	no KO assigned	membrane bound hydrogenase, 4Fe-4S cluster-binding subunit	MBS-N	Cytosolic

a: Annotation used in Yu et al., 2018.

To evaluate the activity of the Fds as electron donors to the MBH complex, the *T. kodakarensis* genome will be modified such that the *fd* sequence will be encoded within the *mbh* operon. Prior to this project, two proteins had never been fused (or tethered) in *T. kodakarensis*, thus what follows is a description of the first viable tethered strains in *T. kodakarensis*. MBH is non-essential and can be deleted from the genome, thus it seems an appropriate candidate for genetic modification. Additionally, the cryo-EM structure of *Pf*-MBH was accomplished by inserting an affinity tag sequence at the 5' end of *Pf-mbhJ*, thus the genomic region of *mbh* can be modified (Santangelo et al., 2011; Yu et al., 2018).[23,46]

To generate tethered strains, the gene sequence for each *fd* was inserted upstream or downstream of one *mbh* gene. The stop codon between the two genes was removed, and a linker sequence which serves as the connector between the two tethered proteins was added. In total, all three Fd were tethered to two different subunits of MBH using two different linker sequences, thus twelve strains were generated for this project to fully evaluate the ability of each Fd to reduce MBH in the generation of H_2 gas. All tethered proteins also encoded an hemagglutinin (HA: YPYDVPDYA, tacccatacgacgttccggactacgca) epitope sequence and an affinity (His_6: HHHHHH, catcaccatcaccatcac) sequence preceding the Fd on MBH-J tethers and following the Fd on MBH-N tethers. These affinity and epitope tags allow for purification and identification via Western blot for future endeavors. A description of the MBH-subunits chosen for tethering will be described followed by how the *fd* sequences were added, and finally the design of the two linker sequences. Markerless genomic modification (or markerless allelic exchange, MAE) of *T. kodakarensis* is described in detail in Chapter 2.2.[46–49]

Tethering to two MBH subunits: MBH-J and MBH-N

MBH maintains four proteins in the cytosolic arm: MBH-J, MBH-K, MBH-L, and MBH-N. *Mbh-M* is encoded between *mbh-L* and *mbh-N* but the MBH-M protein is situated directly "below" the cytosolic arm in the membrane, forming the elbow of the L-shaped complex.[23]

Ideally, a terminus of a cytosolic subunit to be accessible in the cytoplasm and facing away from the membrane to reduce potential steric conflicts of the addition of a protein and to assure that each protein folds into a native, functional confirmation (Figure 5.8).

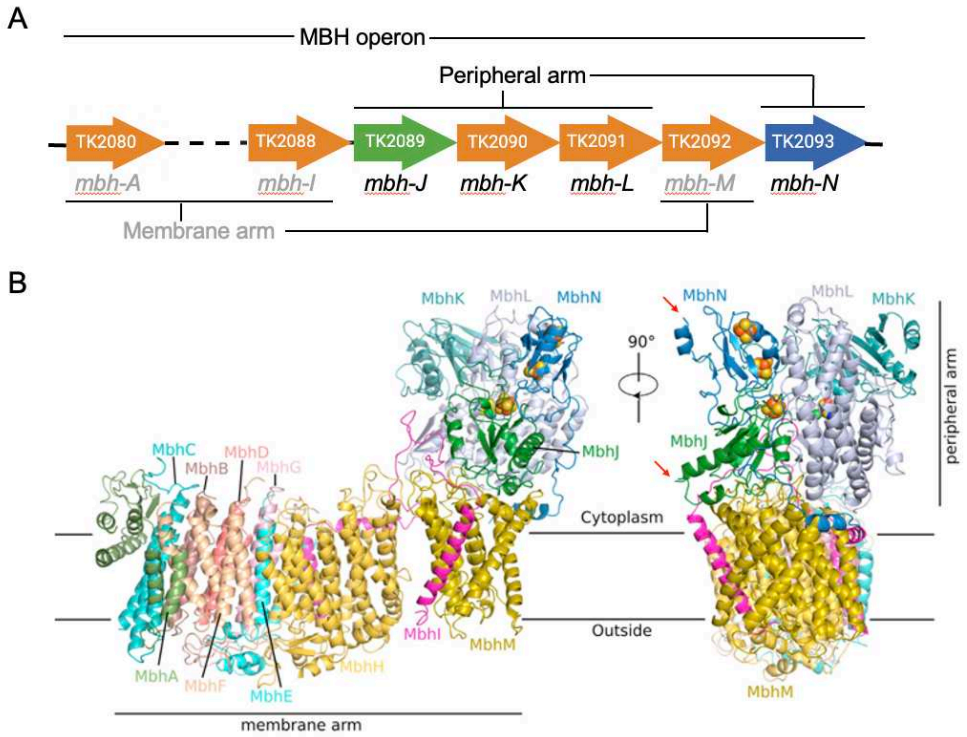


Figure 5.8: The MBH operon and protein structure offer two MBH subunits as plausible targets for tethering. The fourteen-subunit MBH complex is encoded in an operon (TK2080 – TK2093; orange, green, and blue) (A). Mbh-J, mbh-K, mbh-L, and mbh-N compose the peripheral arm which reaches into the cytoplasm as depicted in (B). The 5' end of mbh-J (green) was tagged for purification and cryo-EM, thus this region can be genetically modified and is spatially accessible; both termini of MBH-J are at the exterior of the protein complex (green protein, red arrow red arrow, B). Mbh-N (blue) is encoded at the end of the operon, thus the 3' end of the gene is likely available for modification. The C-terminus of MBH-N (blue protein, red arrow) is exposed and points away from the membrane, thus mbh-N is a plausible target for generating a tethered protein. (Adapted from Yu *et al.*, 2018)

Tethering at the genomic level (as opposed to ectopic expression of a tether sequence encoded on a vector) allows for native expression levels and requires that the genomic region can be modified. *Pf-mbh-J* was modified at the 5' end to encode an affinity tag sequence for purification for cryo-EM trials, thus evidence suggests that this region can be modified.[23] *Mbh-N* is the last gene encoded in the operon thus it follows that tethering at the 3' end of *Mbh-N* is permitted.

Due to the genetic and structural accessibility, the N-terminus of MBH-J and the C-terminus of MBH-N were chosen for tethering (Figure 5.9). MBH-J was chosen as there was confidence in modifying this region of the genome, although considering that the N-terminus points towards the membrane, there could be steric hinderance with the addition of a tethered protein. As MBH-N maintains two of the 4Fe-4S metal centers involved in reduction, it was considered somewhat risky to attempt a modification to the terminus of MBH-N, however the accessibility of the C-terminus pointing away from the membrane and the genetic position at the end of the operon suggested it was worth attempting.

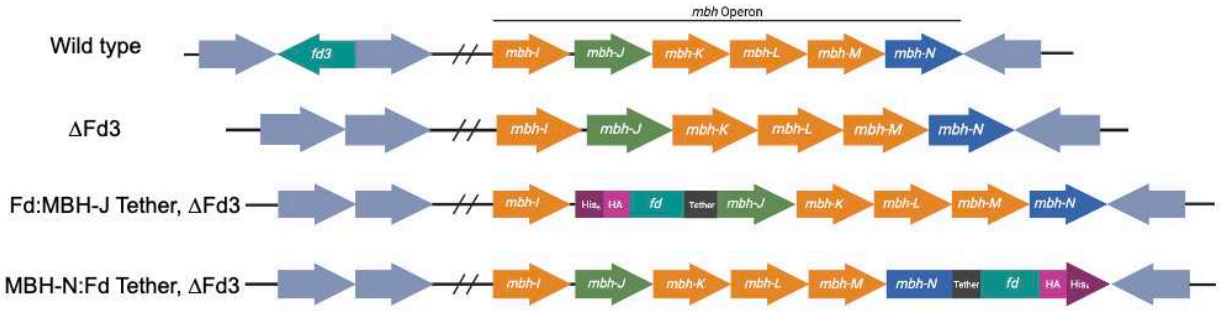


Figure 5.9: Wildtype, ΔFd3, and Fd:MBH tether genomes. The wild type genome encodes *fd3* (TK2012, teal) and the fourteen-gene *mbh* operon (TK2080 – TK2093, orange). Surrounding genes are shown in grey. *Fd3* can be deleted from the genome (ΔFd3). Mbh-J (TK2089, green) and mbh-N (TK2093, navy blue) were chosen as the desired MBH-subunits for generating fused proteins by tethering a Fd to either subunit in the ΔFd3 background. Fd:MBH-J tethers encode a His_6 affinity sequence (purple) and HA epitope sequence (pink) preceding the *fd* sequence (teal) which is linked to the 5' end of *mbh-J*. MBH-N:Fd tethers encode the *mbh-N* sequence tethered at the 3' end to *fd* with the HA and His_6 sequences following.

Encoding ferredoxins in the mbh-operon

Previous findings suggest that Fd3 (TK2012) is likely the primary electron donor to the MBH complex.[5] Fd3 is more highly expressed in non-sulfur conditions, and it was found to associate with four subunits of MBH when analyzed using AP-MS. Additionally, Fd3 is the only Fd which can be deleted from the genome. ΔFd3 displayed an extended lag phase when grown in non-sulfur conditions,[5] further suggesting that Fd3 is required for hydrogen production and likely coupled to the MBH complex.

The goal of this study was to determine which Fd would be the optimal donor to the MBH complex (as measured by H_2 gas production per cell), thus all three Fds were tethered to both MBH-J and MBH-N subunits. Ideally, only a single copy of each Fd would be present in the genome of a tethered strain, however both Fd1 (TK1694) and Fd2 (TK1087) are essential and cannot be deleted. ΔFd3 is viable, thus to limit Fds when possible, all Fd:MBH tether strains were generated in the ΔFd3 parent strain.

Fd1 and Fd2 tethered strains were generated using synonymous codon sequences to assure that neither sequence could integrate into the genome at the natively encoded *fd1* or *fd2* gene (Table 5.2). Synonymous *fd* sequences were designed with codons optimized for *T. kodakarensis* with alternative codons appearing at a frequency greater than 23/1000 to assure for ease during translation (kazusa.or.jp). To allow for double Fd:MBH tether constructs in the future, two synonymous Fd1 sequences were designed such that Fd1z was used with MBH-J tethers and Fd1y was used with MBH-N tethers. Likewise, two synonymous Fd2 sequences were designed: Fd2z sequences were used with MBH-J tethers and Fd2y sequences were used with MBH-N tethers.

Table 5.2: Synonymous *fd1* and *fd2* sequences.

fd1	ATGGCTTGGAAGGTAAGTGTTGATGTCGACACCTGCATTGGAGATGCCATCTGTGCTAGC
fd1y	ATGGCCTGGAAAGTAAGCGTTGACGTCGATACCTGCATTGGCGATGCCATCTGCGCTAGC
fd1z	ATGGCTTGGAAGGTTAGTGTCGATGTTGACACGTGCATAGGAGACGCCATATGTGCCAGC
	***** ***** ** ** ** ** ** ** ** ***** ** ** ***** ** ** ***
fd1	CTCTGCCCGGACGTCTTTGAGATGGGCGACGACGGCAAGGCCCACCCGGTAGTTGAGACC
fd1y	CTCTGCCCGGATGTCTTCGAGATGGGCGATGACGGAAAGGCGCACCCCGTAGTCGAGACG
fd1z	CTCTGCCCCGACGTTTTTGAAATGGGAGACGATGGCAAAGCCCACCCGGTCGTTGAAACC
	******** ** ** ** ** ***** ** ** ** ** ** ***** ** ** ** **
fd1	ACCGACCTTGACTGCGCCCAGGAGGCCGCCGAGGCCTGCCCGGTCGGCGCTATAACCCTC
fd1y	ACCGATCTTGATTGCGCCCAGGAAGCCGCGGAGGCCTGCCCCGTCGGAGCTATCACCCTC
fd1z	ACGGATCTCGATTGCGCCCAGGAAGCCGCCGAAGCGTGCCCGGTTGGAGCCATAACGCTC
	** ** ** ** ********** ***** ** ** ***** ** ** ** ** ** ***
fd1	GAAGAGGCCTGA
fd1y	GAAGAAGCC---
fd1z	GAGGAGGCC---
	** ** ***

```
fd2   ATGCCGGAGAAGATTAAAGTGGTCGTGAACGAAGATAGATGCTATCTCTGCGGCGGTTGT
fd2y  ATGCCGGAAAAGATAAAAGTTGTCGTTAACGAGGATAGGTGCTACCTCTGTGGCGGCTGT
fd2z  ATGCCCGAGAAAATTAAGGTGGTTGTCAACGAAGACAGATGTTATCTGTGCGGAGGTTGC
      ***** ** ** ** ** ** ** ** ***** ** ** ** ** ** ** ** ** **

fd2   GCCGGTGTCTGCCCGACACTCGCGATAGAGGTGCACTCAACAGGCTGGGAGTTCCTTCAA
fd2y  GCGGGTGTTTGCCCCACACTGGCGATCGAGGTTCACAGCACAGGATGGGAATTCCTCCAA
fd2z  GCCGGCGTCTGTCCGACCCTCGCCATAGAAGTCCACTCAACCGGGTGGGAGTTTCTTCAG
      ** ** ** ** ** ** ** ** ** ** ** ***   ** ** ***** ** ** **

fd2   GACAAGTGCATAAGCTGCAGGATATGCATCAACGCCTGCCCCGTTGGAGCCCTGAGCGCT
fd2y  GATAAGTGTATATCCTGCAGAATATGTATCAACGCGTGCCCGGTTGGCGCCCTCAGCGCC
fd2z  GACAAATGCATCAGCTGTAGGATCTGCATAAACGCCTGTCCCGTCGGAGCGCTGTCCGCT
      ** ** ** **   *** ** ** ** ** ***** ** ** ** ** ** **   ***

fd2   AAACCCCTGGAGGTGAGCGAATGA
fd2y  AAACCGCTGGAAGTGTCCGAA---
fd2z  AAGCCCCTCGAGGTTAGCGAG---
      ** ** ** ** **   ***
```

Designing two ideal tether sequences

The specific sequence of amino acids confers protein structure and function, which must not be disrupted when two proteins are linked to one another.[50–52] The goal of the linker sequence is to position each protein in local proximity without disrupting function, and even possibly enhancing function. The specific case of linking an electron donor protein to an electron acceptor protein necessitates that the two surfaces which interact will do so ideally such that an electron can pass from the Fd to the first 4Fe-4S moiety of the MBH complex with ease. A survey of proteins involved in electron transfer suggests that electrons can travel up to ~14 Å but optimally travel ~6 Å through a protein medium.[53] A linker sequence must not only allow for electron transfer but for an effective protein-protein interface to form between the donor and acceptor proteins.

Redox protein fusions using flexible glycine-rich and rigid proline-rich linkers have demonstrated that linker length as well as a proper balance of structure and flexibility confers the greatest activity in the case of cytochrome C reduction by fused flavodoxin and flavodoxin reductase in *E. coli*.[54] Shorter flexible linkers (~ten glycine residues) displayed similar increases in redox function compared to longer, more rigid linkers (~twenty proline residues), while even longer linkers (> twenty-five amino acid residues) displayed much lower activity and greater rates of protein turnover.[54]

Owing to the small size of the Fds (Figure 5.2) and the inflexible nature of the large, membrane-bound complex, two linkers were chosen (Table 5.3). The first consisted of five glycine (G) residues and was chosen to provide ultimate flexibility and a reduced distance. Each glycine residue is ~ 3.8 Å, suggesting that a G_5 linker sequence could reach ~ 19 Å. The second linker sequence consisted of a combination of G, serine (S), and proline (P) residues: GGSPPPSGG. This nine amino acid series contains G_2 on either end which provide flexibility and P_3 in the middle which provides rigidity, connected by a small, polar S residue between the G and P residues to assure that the linker sequence remains soluble. The two likers will further be referred to as G_5 and 9aa. Linker sequences were designed with codons optimized for *T. kodakarensis* with synonymous codons appearing at a frequency greater than 23/1000 to assure ease during translation (kazusa.or.jp).

Table 5.3: Nucleotide and amino acid sequences of the two tether sequences.

5 glycine linker	Nucleotides:	GGCGGAGGCGGAGGC
	Amino acids:	G G G G G

9 amino acid linker	Nucleotides:	GGCGGAAGCCCGCCCCCGAGCGGCGGA
	Amino acids:	G G S P P P S G G

Final Fd:MBH tethered strains

Each combination of Fd, MBH, and linker sequence was generated in the ΔFd3 parent, resulting in twelve viable, tethered strains (Figure 5.10). Tethered strains were confirmed by whole genome sequencing and PCR (Figure 5.11).

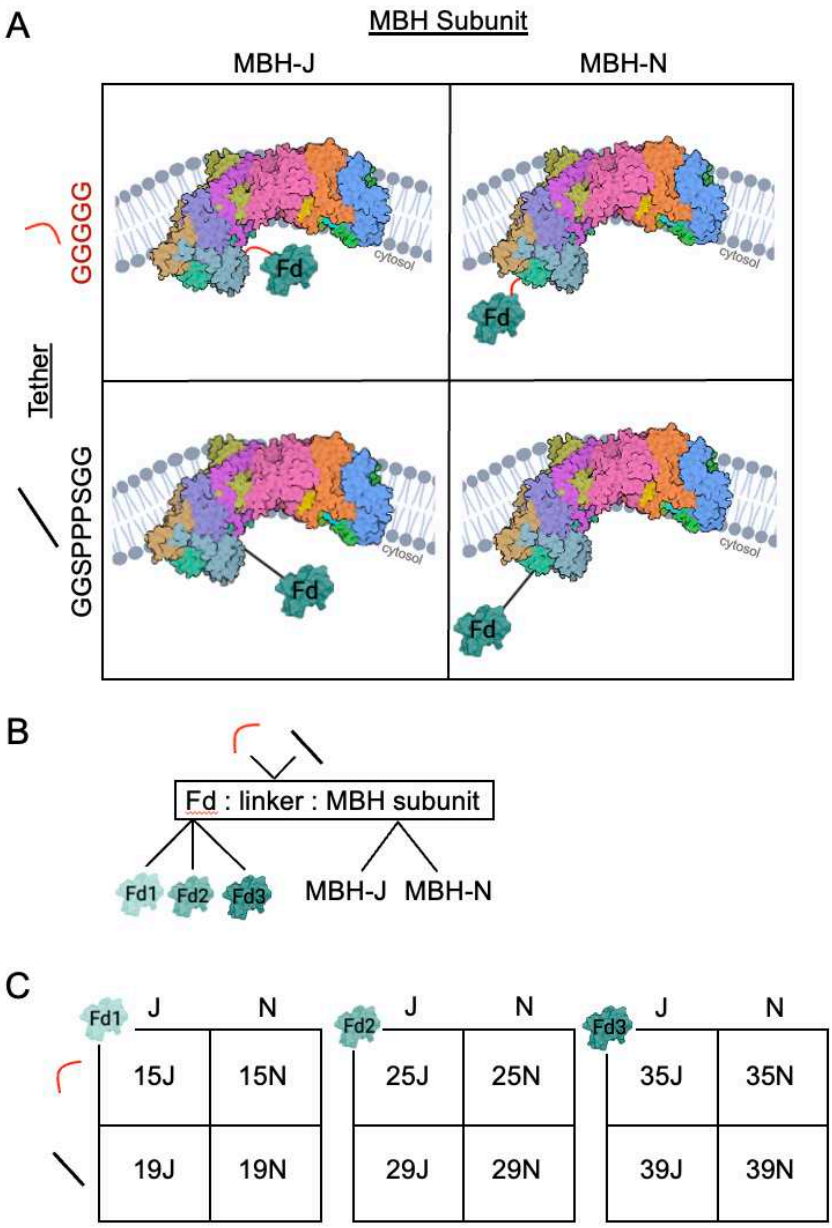


Figure 5.10: Fds are tethered to the MBH-J or MBH-N subunit via a short, flexible (G_5) or longer, more rigid (9aa) linker. The MBH complex resides in the membrane with four subunits reaching into the cytosol.[23] Each Fd protein was linked to either the N-terminus of the MBH-J subunit (A, left) which points into the cytosol towards the membrane arm or the C-terminus of the MBH-N subunit (A, right) which points into the cytosol away from the membrane arm. Two amino acid linker sequences were used to fuse a Fd to either of the MBH subunits: a short, flexible linker consisting of five glycine residues (G_5, red, top) and a longer, more rigid sequence consisting of nine amino acids (9aa, black, bottom). Tethered strains are named for the Fd:tether:MBHsubunit (B) such that each strain is identified by a three-character code (C). For example, Fd1 tethered with a G_5 linker to the MBH-J subunit is coded by "15J." Each of the Fds (Fd1, Fd2, or Fd3) was fused with each combination of linker and MBH-subunit, thus twelve strains in total were generated.

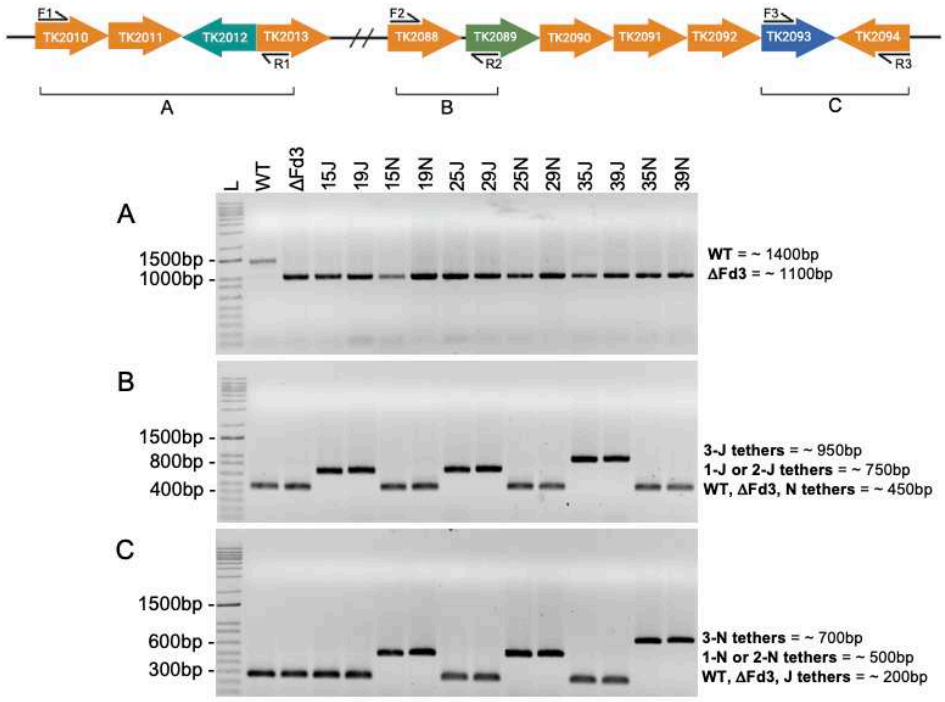


Figure 5.11: PCR amplification of wild type, ΔFd3, and tethered strains confirms the expected mutant genotypes for all strains. A genomic map (top) of the genes surrounding *fd3* (TK2012), *mbh-J* (TK2089), and *mbh-N* (TK2093) shows forward (F1, F2, or F3) and reverse (R1, R2, or R3) primer binding sites to identify Δ*fd3* (A), MBH-J tethers (B), or MBH-N tethers (C). PCR amplicons are separated by agarose gel electrophoresis and compared to a 1kb+ ladder (Invitrogen, 10787026). The gel in (A) identifies the presence of *fd3* in the WT strain (amplicon size ~1400bp) while all tethers show amplicon sizes which align to Δ*fd3* (amplicon size ~300 bp), confirming that all tethered strains were generated in the Δ*fd3* background and maintain this genotype. The gel in (B) identifies the presence of a *fd* gene preceding the *mbh-J* gene with an amplicon size of ~ 750bp for *fd1-J* and *fd2-J* tethered strains and a slightly larger amplicon (~950bp) for *fd3-J* tethered strains. *Mbh-N* tethered strains show amplicon sizes which match the parental (Δ*fd3*) strains ~450bp. In (C), the presences of the *fd* gene following the *mbh-N* gene with an amplicon size of ~ 500bp for *fd1-N* and *fd2-N* tethered strains and a slightly larger amplicon (~700bp) for *fd3-N* tethered strains. *Mbh-J* tethered strains show amplicon sizes which match the parental (Δ*fd3*) strains ~200bp. Size differences of the nucleotides encoding the G_5 (15 nucleotides) or 9aa (27 nucleotides) linker are not distinguished at this resolution. F1 = 001-2011 and R1 = 002-2011; F2 = 001-2090 and R2 = 002-2087; F3 = 001-2095 and R3 = 002-2092.

5.2.2 Tethering Fd2 to GGR

The Archaeal domain is defined by prokaryotes which maintain isoprenoid-based lipids with ether linkages connecting polar head groups.[22] Lipids are an important reductant sink and GGR is a monomeric protein responsible for maturation of lipids from unsaturated to saturated isoprenoid chains (Figure 5.12).[12,42]

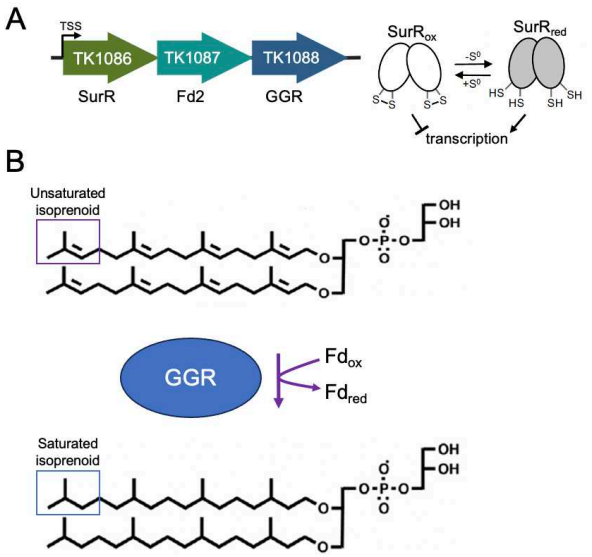


Figure 5.12: GGR genomic organization, transcriptional regulation, and function. *Ggr* is encoded by the final gene (TK1088) in the *surR* operon (TK1086 – TK1088) which includes *surR* and f*d2* encoded upstream of *ggr*. SurR is redox-sensitive transcription factor which is active in the reduced form (grey ovals) in the absence of sulfur (A) (figure adapted from Hidese *et al*., 2017). Archaeal membranes are composed of isoprenoid lipids attached to a glycerol moiety via an ether linkage (B). Electrons provided by a reduced Fd are transferred to GGR which is responsible for converting unsaturated isoprene groups (shown in the purple box, B) to saturated isoprene groups (shown in the blue box, B) to produce a fully saturated lipid.

While GGR from some species (e.g. *Thermoplasma acidophilum*) can utilize NADPH as a direct electron source,[18] other species require a Fd_{red} to provide electrons to GGR.[42] Recombinant expression of GGR from a closely related hyperthermophile, *Methanosarcina acetivorans*, resulted in fully saturated lipids in *Escherichia coli* only when the archaeal *fd* encoded near *Ma*-GGR was expressed.[42] AP-MS data from *T. kodakarensis* supported a strong association between Fd2 and GGR when Fd2 was tagged[5] or when GGR was tagged (Chapter 4.3). At ~7 kDa, Fd2 (TK1087) is a small protein and difficult to detect in proteomic mass spectrometry (discussed in depth in Chapter 3.2.3). Even with only a few detectable peptides, Fd2 was found in association with tagged-GGR in both sulfur and non-sulfur conditions (and was not found in any other AP-MS dataset).

Ggr and *fd2* are also linked genomically in the *surR* operon. *T. kodakarensis* undergoes a transcriptional shift in response to sulfur.[9,55] The sulfur response regulator (SurR) is a redox

sensitive transcription factor which upregulates the expression of genes which are useful in non-sulfur metabolism, such as the *mbh* operon, and downregulates the expression of genes which are not needed in sulfur metabolism, such as the *mbs* operon.[8,56] *SurR* is encoded directly upstream of *fd2* and *ggr* in the *surR* operon (TK1086 – TK1088) (Figure 5.12A). Electrons provided by $Fd2_{red}$ are transferred to GGR which is responsible for converting unsaturated isoprene groups (showing in the purple box) to saturated isoprene groups (shown in the blue box) to produce a mature lipid (Figure 5.12B).

We hypothesize that if Fd2 is tethered to GGR, electron flux could be directed to the lipid reductant pool. Considering that Fd2 is essential yet *fd2* is encoded directly upstream of *ggr*, such a protein fusion would result in every instance of Fd2 and GGR being linked in the cell, which is not possible in the Fd1-MBH or Fd1-MBH tethers unless *fd1* or *fd2* were deleted from the tethered strain, which has not been attempted and is considered highly unlikely from previous Δ*fd1* and Δ*fd2* attempts. Here, we have designed three tether constructs wherein Fd2 is linked to GGR via five, nine, and twenty-four amino acids (Figure 5.13). As was done with the Fd:MBH tethers, these constructs also include an affinity (His_6) and epitope (HA) tag sequence to allow for purification and identification via Western blot for future assays. Constructs will further be referred to as 25G, 29G, and 224G to follow suit with the Fd:MBH tether naming scheme.

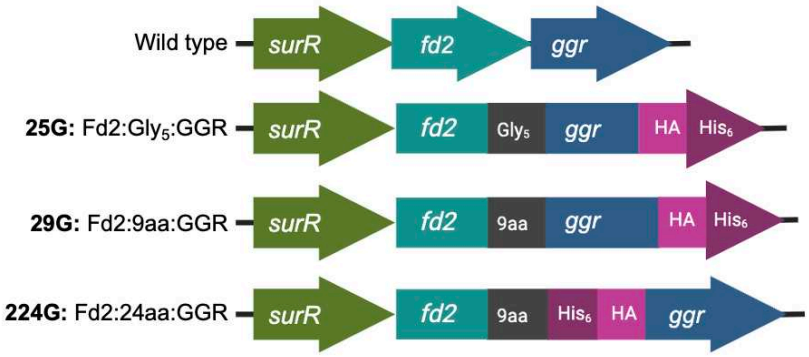


Figure 5.13: Wild type *surR* operon and Fd2:GGR tether constructs. Wild type genomic sequence of the *surR* operon (top) encodes surR (green), fd2 (teal), and ggr (blue). Three strains in which Fd2 is physically linked to GGR are constructed by removing the stop codon of *fd2* and inserting sequences which encode three different linker lengths. Nucleotide sequences

encoding a short flexible linker (G_5; 25G), a longer, more rigid linker (9aa, GGSPPPSGG; 29G), or the 9aa linker with a His_6 and HA encoded sequence resulting in a 24aa linker connecting Fd2 and GGR (224G) have been designed. All three tethers contain an affinity (His_6) and epitope (HA) sequences. In the 25G and 29G tethers, the His_6 and HA encoding sequences follow *ggr* at the 3' end, however in the longest tether (224G), the His_6 and HA sequences precede *ggr* at the 3' end thus form part of the linker between Fd2 and GGR.

GGR is a ~44 kDa protein, thus to generate a fusion wherein Fd2 is tethered to GGR with a G_5 or 9aa linker and includes epitope and affinity tags (His6 and HA, respectively) at the C-terminus of GGR, the genome must undergo two mutational events. We have previously generated a strain wherein *ggr* encodes the epitope and affinity tag sequence at the 3' end, resulting in a C-tagged protein (referred to as TK1088C in Chapter 4 or GGR-tag). 25G and 29G will be constructed in the GGR-tag background. To date, only the 224G tether strain has been successfully constructed (Figure 5.14). Efforts to generate the shorter linker strains, 25G and 29G, continue. As the longest linker has been constructed, it is supposed that the other two linker combinations are viable, but after five transformation attempts to date, a viable 25G or 29G strain has not yet been confirmed.

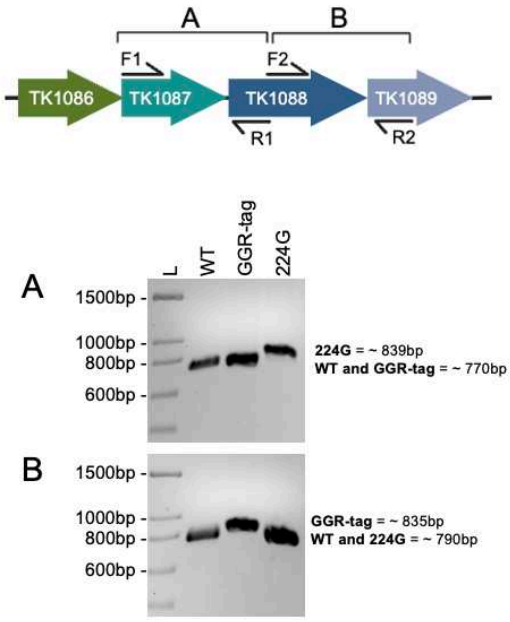


Figure 5.14: Confirmation of genomic modification of GGR-tag, and 224G tether strains. Genomic amplification via PCR of the region upstream and downstream of the 3' end of TK1087 (*fd2*) and 5' end of TK1088 (*ggr*) identifies the presence of the 224G modification with ~839 bp (A, primers F1 and R1) when compared to WT and GGR-tag strain (~770 bp) which do not

contain a Fd2:GGR fusion. In 224G, the stop codon of TK1087 has been removed, nucleotides encoding the 9aa linker followed by the affinity tag and HA tag sequence have been inserted preceding the start codon of *ggr*. A 69bp shift is evident in the 224G strain compared to WT and GGR-tag strains (A) as observed by 2% agarose gel electrophoresis. Bands were separated at 80V for ~ 2 hours. A 1kb+ ladder (Invitrogen, 10787026) is included to identify band size. In (B), amplification of the 3' end of TK1088 (*ggr*) identifies the presence of the sequence encoding the epitope and affinity tag sequence ~ 835 bp in GGR-tag strain compared to WT and 224G (~790 bp) which do not contain any modification at the 3' end of TK1088. F1 = 001-1088 and R1 = 008-1088; F2 = 001-1089 and R2 = 008-1088).

Where all three Fds were tethered to MBH, only one Fd (Fd2) has been tethered to GGR. Considering the strong evidence for the association between Fd2 and GGR and the gene sequence positioning in the *surR* operon, tethering Fd1 or Fd3 to GGR is of less interest at this time but could be valuable for directing reductive flux in future endeavors.

5.2.3 Measuring a change in reductive flux

A shift in reductive outcomes in a tethered strain would indicate that the fusion construct impacts the flow of electrons. To successfully measure this change in flux, we will use three metrics: growth curves, hydrogen gas production, and lipid profiles.[5,12,57,58]

Growth curves as a measure of microbial strain fitness

Microbial cultures which demonstrate exponential doubling follow predictable growth patterns.[59–61] Cultures are typically inoculated into fresh media at a 1:100 ratio with freshly grown starter cultures. Initial doublings do not produce great enough cell densities to be read via a standard optical density reading using a spectrophotometer at 600nm (OD_{600}), however after ~3-4 hours of growth, cell densities can be observed at ~0.1 OD_{600} (Figure 5.15).[44] Growth from inoculation to the first readings are considered the lag phase. Once cell density can be detected, there are standardly sufficient cells that doubling every ~40-60 minutes can be observed by spectrophotometric readings increasing from 0.1 OD_{600} to 0.8 OD_{600} in a matter of hours which is referred to as exponential growth stage. Cells grown in batch culture are nutrient

limited and subjected in increasing amounts of toxic biproducts, thus growth plateaus at 0.8 – 1.2 OD_{600} until reaching death phase where the only nutrients remaining are scavenged from dead cells.[16] Measuring the OD_{600} each hour provides a metric for fitness in which multiple strains in replicate can be compared.

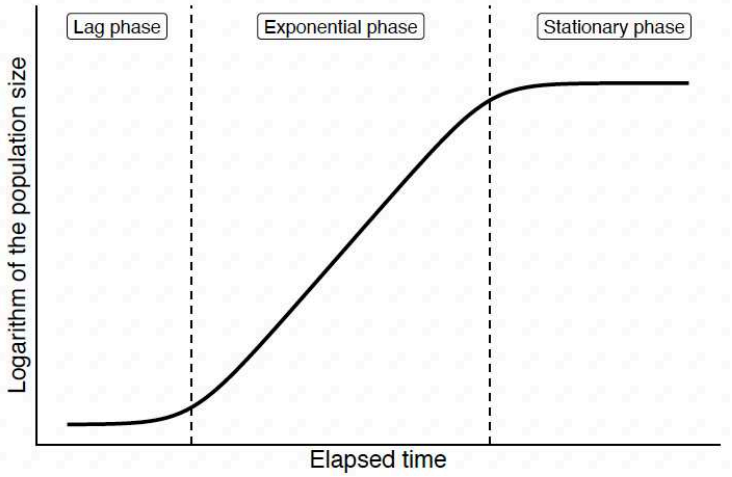
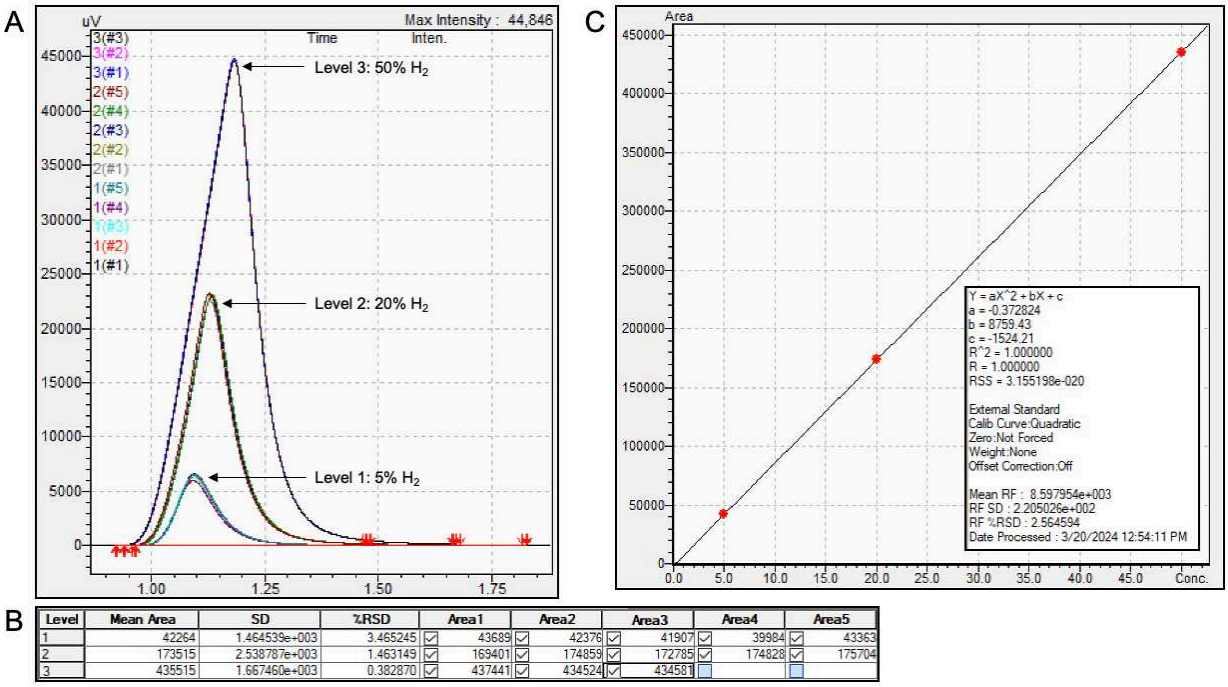


Figure 5.15: Model growth curve of a batch microbial culture. Microbial cultures grow with predictable stages as observed here. Cells divide but do not reach sufficient numbers to indicate growth as determined by optical density using a spectrophotometer at 600 nm (OD_{600}) which is referred to as "Lag phase." Once cells reach sufficient density, their doubling can be measured in frequent intervals indicating "exponential phase" whereby the slope of the line can be used as a measure of fitness. Cells with a continuous influx of nutrients could hypothetically continue to divide at the rate observed in exponential phase. Cells in batch culture (as opposed to continuous culture) reach a maximal growth OD_{600}, referred to as "stationary phase," as nutrients are limited. The maximum OD_{600} reached in stationary phase can be used as another measure of microbial fitness. Death phase (not shown here) occurs following stationary phase and is often indicated by a reduction in OD_{600}, especially if dead cells flocculate. (Adapted form Garre *et al*., 2023.)

Hydrogen gas generation as a measure of reductive flux

Hydrogen gas can be detected using a gas chromatograph (GC) equipped with a thermal coupled detector (TCD) (Shimadzu, GC-2014).[12] Nitrogen gas flows through two packed columns in a GC-TCD and voltage is measured across each column by the TCD. A change in voltage indicates the presence of a gas other than the N_2 carrier gas and the application of pure gas mixtures (*e.g*. 5% H_2, 95% N_2) at specified concentrations allows for a quantitative measurement of H_2 gas (Figure 5.16). For the purposes of measuring H_2 production in *T.*

kodakarensis, strains are grown sealed tubes where OD_{600} measurements indicate growth and gas can be extracted from the headspace of the sealed culture tube and % H_2 can be measured against a standard curve on the gas chromatograph.



Level	Mean Area	SD	%RSD	Area1	Area2	Area3	Area4	Area5
1	42264	1.464539e+003	3.465245	43689	42376	41907	39984	43363
2	173515	2.538787e+003	1.463149	169401	174859	172785	174828	175704
3	435515	1.667460e+003	0.382870	437441	434524	434581		

Figure 5.16: A standard curve is used to provide quantitative H_2 measurements on a GC-TCD. Three H_2 standards at 5% H_2 (Level 1), 20% H_2 (Level 2), and 50% H_2 (Level 3) are measured in 3-5 replicates on a GC-TCD (Shimadzu, GC-2014; Lab Solutions software) and a chromatogram showing a signal in μV over time in minutes is displayed. The retention time for each peak is averaged and the area under each curve (B) for each standard replicate is used to generate a standard curve. A quadratic curve (C) is generated for $Y = aX^2 + bX + c$, where Y is the area under the curve and X is solved for the % H_2 in a sample. Here, a = -0.372824, b = 8759.43, and c = -1524.21. The R^2 value = 1.0. A standard curve should be generated on the day the samples are tested.

Lipid profiles can identify saturation patterns

Unique features of archaeal lipids distinguish the domain from Bacteria and Eukarya which maintain fatty acids bound via ester linkages.[62,63] Archaeal lipids are uniquely composed of isoprenoid lipid chains attached to glycerol head groups with ether bonds.[58] Additionally, membranes are composed of both bilayer and monolayer moieties (diether and tetraether lipids) with various numbers of cyclopentane rings (Figure 5.17).[57]

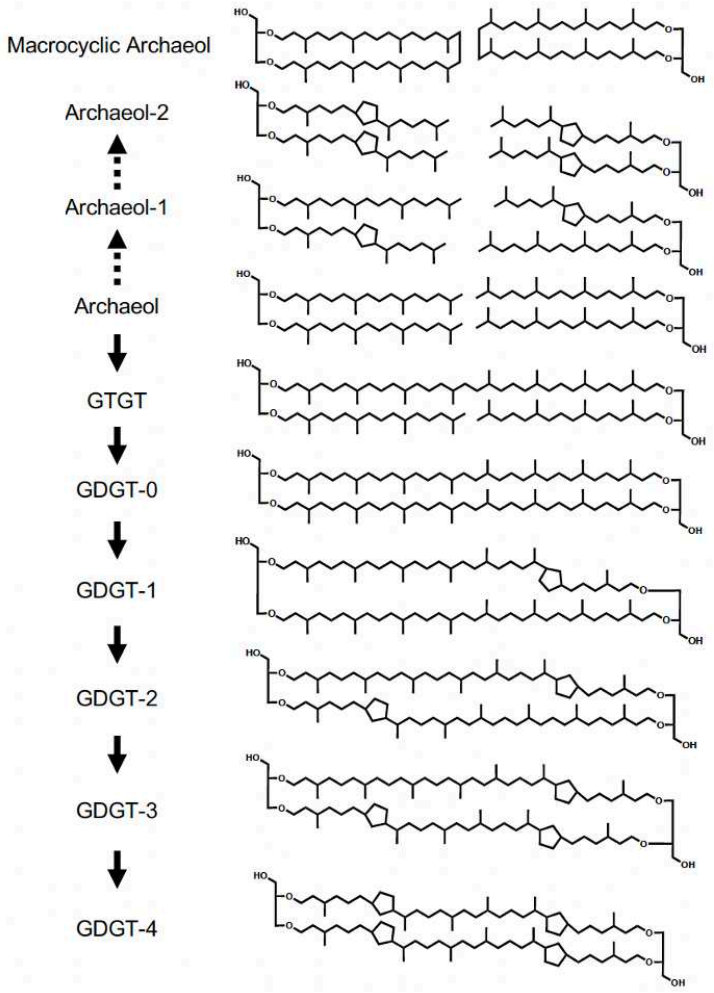


Figure 5.17: Fully saturated archaeal membrane lipids maintain a variety of diether, tetraether, and cyclopentane ring structures. Archaea maintain unique lipid moieties that consist of isoprenoid chains bound to polar head groups via ether linkages. Archaeol lipids can be modified to contain 1 – 4 cyclopentane rings (Archaeol-1 or Archaeol-2) as diether lipids (above) or as tetraether lipids (below). Diether lipids can be converted to macrocyclic archaeols (top). Monolayer lipid structures with tetraether headgroups can form with a single bond (GTGT) or by generating a cyclicmonolayer from glycerol dibiphytanyl glycerol tetraethers (GDGT). Both diether and tetraether lipids can contain a variety of cyclopentane rings (archaeol-1, archaeol-2 at top or GDGT-1, GDGT-2, GDGT-3, or GDGT-4 at bottom) can also contain 1 – 4 cyclopentane rings (GDGT-1 – GDGT-4). Lipids depicted here consist of fully saturated isoprenoid chains. (Adapted from Liman *et al*., 2024.)

The Fd2:GGR tether may impact lipid saturation of any of the many lipid moieties observed in *T. kodakarensis* (Figure 5.17). Acid hydrolyzed lipids can be analyzed using liquid chromatography mass spectrometry (LC-MS) wherein samples are separated via reverse phase chromatography, charged via electrospray ionization, and the resulting masses are analyzed in

a quadrupole mass spectrometer.[57,64] Diether lipids can further be analyzed using gas chromatography-MS (GC-MS). The combination of these methods can provide a thorough analysis of lipid moieties in tethered and non-tethered strains which can be used to infer the impact and function of tethering a likely electron donor (Fd2 in this case) to a likely electron acceptor (GGR in this case), thus furthering our understanding of reductive flux in *T. kodakarensis*.

Evaluating the Fd:MBH tethers for growth and hydrogen gas production

Confirmed Fd:MBH tethered strains were analyzed using a growth curve with concurrent H_2 measurements, and all tether and control strains were tested in biological quadruplicate. Tether strains were compared to the parental (ΔFd3) strain and WT (TS559)[49] strain which maintains a functional *fd3* (a detailed description of TS559 can be found in Chapter 2).

While *T. kodakarensis* is viable in anaerobic, sealed jars at room temperature for extended periods of time (*i.e*. years), the number of viable cells in an original culture bottle cannot be known. To synchronize cultures for growth and H_2 measurements, it was determined that sequence-confirmed cultures would be used to inoculate pre-starter cultures and pre-starters would be used to inoculate starter cultures which would be used to inoculate test strains. Additionally, all long-term, sequence confirmed cultures are grown with sulfur, so passaging cultures twice before growth in test vials should sufficiently limit the chance of sulfur contamination in non-sulfur experimentation. Sequence confirmed cultures were used to inoculate pre-starters at 1:100 dilution. When inoculating the starters and the experimental cultures, the volume for inoculation was adjusted based on the OD_{600} of the freshly grown culture. For instance, a 1:100 inoculation ratio is performed by inoculating a 100 mL bottle with 1 mL of freshly grown culture at an OD_{600} of 1.0. If the OD_{600} of the freshly grown culture is only 0.5 at the time of inoculation, the volume of the inoculum is doubled, thus 2 mL of a culture at 0.5 OD_{600} would be used to inoculate a 100 mL culture.

Cultures were grown in artificial sea water with yeast, tryptone, and pyruvate (ASWYTP) with the addition of 1:1000 KOD Vitamins and 1M agmatine sulfate.[5,46–49] A pea-sized scoop of sulfur was supplemented to each culture bottle (or tube) in sulfur experimentation, which is ample and should not result in sulfur limitation in batch culture conditions. Tubes were inoculated with the respective culture, sealed with new, autoclaved stoppers, and then gasses were purged from the headspace of the culture tube and replaced with pure N_2 over three cycles and pressurized to ~ 10 psi with N_2. Cultures were grown at 85˚C and OD_{600} measurements were taken each hour. Three H_2 measurements were taken for each tube during the growth curve. H_2 was measured when the culture reached 0.4 OD_{600}, 0.8 OD_{600}, and an end point measurement after the OD_{600} remained stable for ~2 hours. Each H_2 measurement returned the pressure in the headspace of the culture tube to ambient conditions.

Pure N_2 is used as the carrier gas and flows through two packed columns, i. Molecular Sieve 5A 60/80 mesh 1/8th SS 2.0 m (Shimadzu, 220-94714-20) and ii. Hayesep Q 80/100 mesh 1/8 stainless 2.0 m (Shimadzu, 220-94715-20) at 15 mL/minute. The GC-TCD injector and detector are maintained at 110˚C while the columns are kept at 40˚C. A 100 μL loop attached to an automated valve is opened at the start of a run and filled with gas from the headspace of the culture tube for ~ 6 seconds. Excess gas flows through an outlet and into a bottle of ~80% ethanol where bubbles resulting from the excess gas are observed. The 100 μL gas sample from the culture tube flows through the MS5 column and across the detector where a change in voltage between the gas sample in one column and the carrier gas in the other column is measured. A chromatogram of the voltage over time is displayed (Lab Solutions software). In this system, a peak from H_2 gas is observed at a retention time of ~1 minute. The area under the curve is compared to a standard curve constructed with three H_2 standards and the %H_2 is recorded.

Evaluating the Fd2:GGR tether for changes in lipid composition

The parental strain (TS559) and Fd2:GGR tether were grown at 85˚C in ASWYTP inoculated from freshly growing starter culture at a 1:100 ratio and supplemented with 1M agmatine sulfate and KOD vitamins at 1:1000.[46–49,65] Cells were grown to stationary phase OD_{600} ~ 0.9 before being put on ice and harvested via centrifugation at 4˚C. Cell pellets were frozen at -80˚C and shipped on ice to Dr. Paula Welander's lab at Stanford University for analysis.[57]

5.3 Results

Tethering a Fd to an electron acceptor protein could direct electrons to specific reductant sinks. Metabolic flux can be measured using a variety of assays which evaluate growth at different stages (lag, exponential, or stationary) via a growth curve, along with H_2 gas production, or in the case of lipids, a lipid panel to distinguish the presence various lipid moieties.[5,12,46,57–59,64] In this section, results for growth and H_2 production will be discussed for the Fd:MBH tether strains and then growth and a lipid analysis will be discussed for the Fd2:GGR tether strain.

5.3.1 Phenotypic outcomes of the Fd:MBH tethered strains

Growth curves and hydrogen measurements of the Fd:MBH strains are discussed here. Lipid profiles of the Fd2:GGR tethered strain will be discussed in section 5.3.2.

Growth curve results of Fd:MBH tethered strains in non-sulfur conditions

All tethered strains were viable in sulfur and non-sulfur conditions (Figures 5.18 - 5.22). Growth curves in non-sulfur conditions revealed that ΔFd3 and 15N reached exponential stage ~1 hour before the other strains, however WT reached the greatest OD_{600} and 35J the lowest

OD_{600} in stationary phase. Maximum standard error for OD_{600} measurements for the four biological replicates was less than 0.07 for all strains.

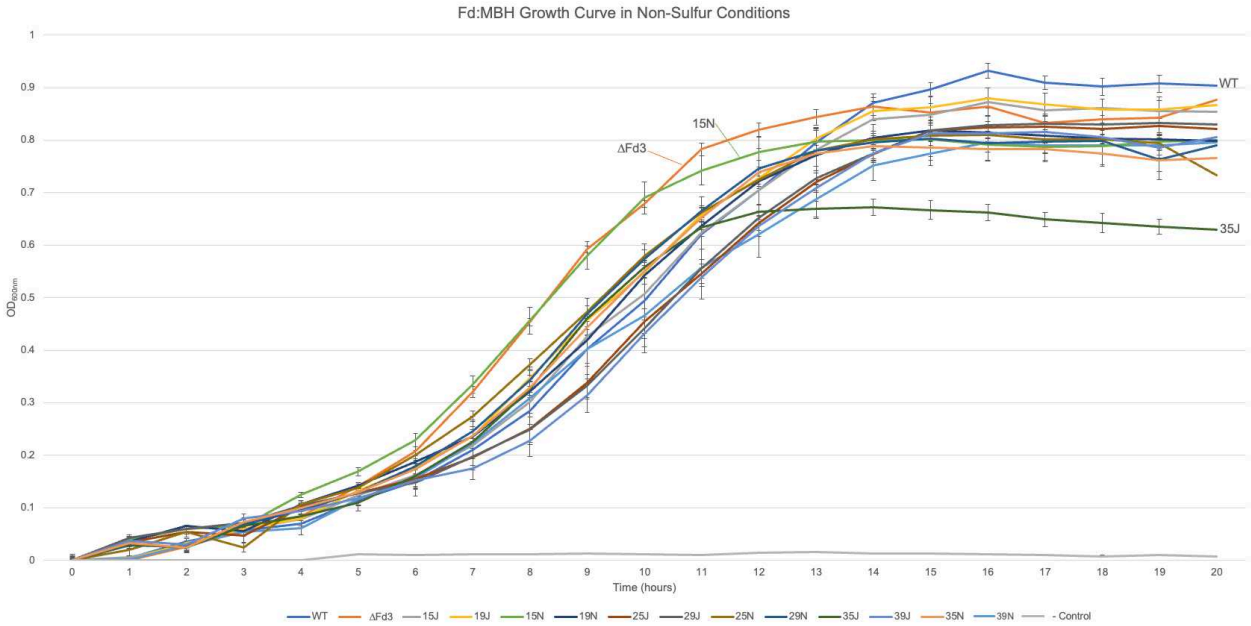
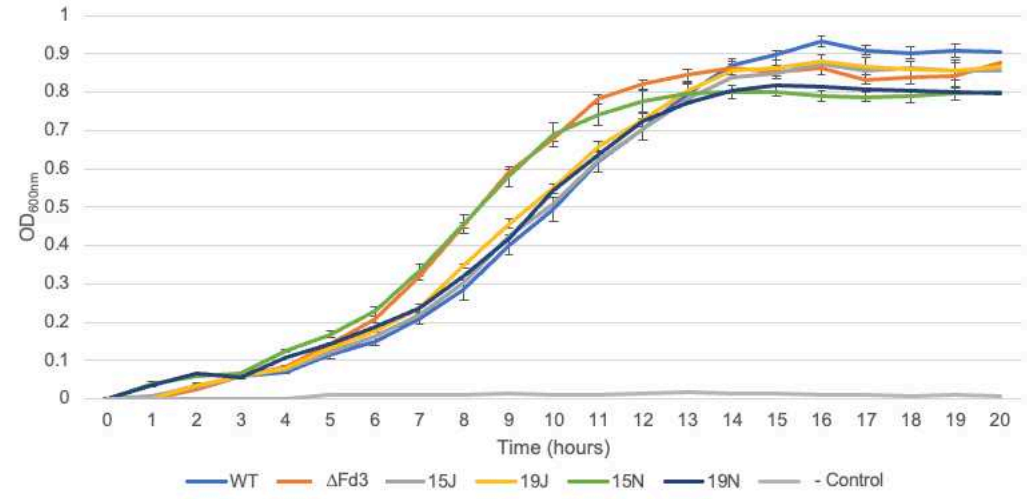


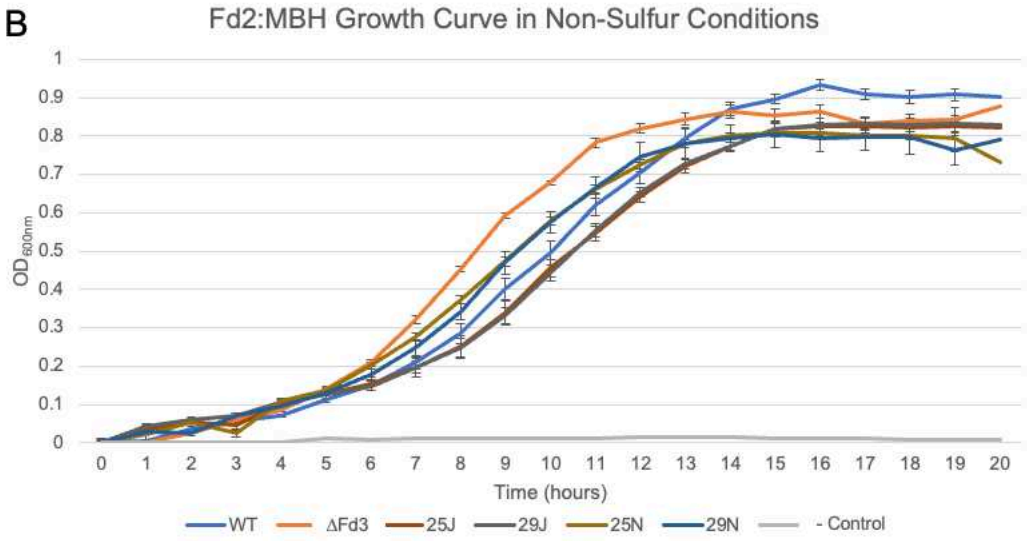
Figure 5.18: Growth curve of all strains in non-sulfur conditions. Cultures inoculated at 1:100 with freshly grown inoculum were evaluated for growth phenotypes at 85˚C in ASWYTP media lacking sulfur for 20 hours. Pre-starter and starter cultures also lacked the addition of sulfur. Optical density at 600 nm (OD_{600}) of the culture was taken every hour. One replicate in ΔFd3 (tube #6) and one replicate in 39J (tube #59) were identified as outliers and removed from the data, thus these groups consist of three biological replicates while all other groups consist of four biological replicates. Standard error of three or four biological replicates are shown with error bars. No strains reached an OD_{600} before 3 hours. On average, strains reached exponential growth phase ~ 6 – 8 hours and stationary phase ~ 10 – 12 hours. Two strains, ΔFd3 (orange) and 15N (lime green) reached exponential phase before all other strains, while WT (blue) reached the greatest OD_{600} at ~ 0.9. 35J showed a clear reduction in maximum OD_{600} by plateauing at ~ 0.67. The remaining strains are difficult to interpret as depicted, thus results for growth phenotypes for tethers will be grouped and compared within or between groups.

As the data set with all strains is difficult to interpret, growth curves have been separated such that all Fd1 tethers are grouped (Figure 5.19A), all Fd2 tethers are grouped (Figure 5.19B), or all Fd3 tethers are grouped (Figure 5.19C).

A


Fd1:MBH Growth Curve in Non-Sulfur Conditions
OD600nm
Time (hours)
WT
ΔFd3
15J
19J
15N
19N
- Control

B


Fd2:MBH Growth Curve in Non-Sulfur Conditions
OD600nm
Time (hours)
WT
ΔFd3
25J
29J
25N
29N
- Control

C

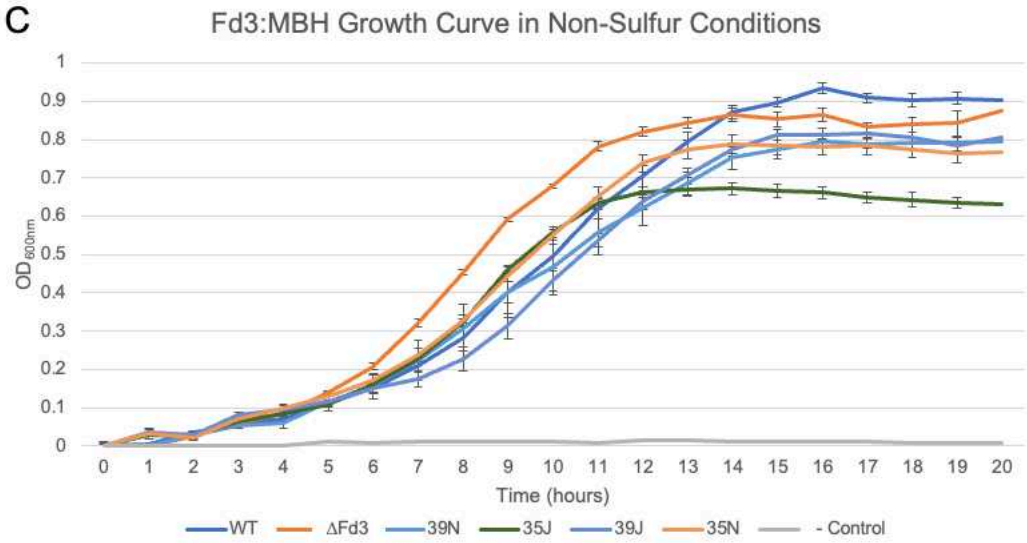

Fd3:MBH Growth Curve in Non-Sulfur Conditions
OD600nm
Time (hours)
WT
ΔFd3
39N
35J
39J
35N
- Control

Figure 5.19: Growth curves of strains grouped by Fd in non-sulfur conditions. Growth curve data is grouped to show all curves from strains with Fd1 (A), Fd2 (B), or Fd3 (C) tethers. WT (blue) and ΔFd3 (orange) are shown in each graph for comparison. Uninoculated, negative controls are included and behaved as expected (light grey, at bottom). Fd1 tethers (A) grew similar to WT during exponential phase with the exception of 15N (green, A) which reached exponential phase ~1 hour earlier showing a growth phenotype more similar to ΔFd3 (orange, A). WT reached the greatest OD_{600} (~0.9, blue, A), while 15N (green, A) and 19N (navy blue, A) displayed slightly reduced stationary OD_{600} (~0.8) compared to 15J, 19J, and ΔFd3 (~0.85, grey, yellow, and orange, respectively, A). Fd2 tethers (B) showed 25N (gold, B) and 29N (dark green, B) reaching exponential after ΔFd3 but before WT, while 25J (brown, B) and 29J (dark gray, B) reached exponential after WT. Fd2 tethers displayed similar OD_{600} at stationary phase, with all reaching a lower maximum OD_{600} compared to WT. Fd3 tethers (C) reached exponential growth phase later than ΔFd3, while 35J and 35N reached exponential growth phase before WT, 39J mirrored WT, and 39N reached exponential after WT. There was a wider difference in final OD measurements during stationary phase for Fd3 tethers than with the other Fds. 35J displayed the lowest maximum OD_{600} (~0.68, green, C), while 35N, 39J, and 39N showed greater maximum OD_{600} than 35J, all were less than ΔFd3 or WT. Error bars depict standard error.

When comparing Fds, 15N reached exponential earlier than other tethered strains, mirroring exponential phase for ΔFd3. Such data suggests that Fd1 may function in place of Fd3 when Fd3 is deleted, however physical linkage of Fd1 using a 9aa linker or when fused to the MBH-J subunit, did not demonstrate a similar growth phenotype to ΔFd3. Both 35J and 35N tethers reached exponential phase before 39J or 39N, suggesting that the G_5 linker is preferred in exponential phase. 35J demonstrated a clear reduction in maximum OD_{600} during stationary phase suggesting that this strain posed some restriction on continued cell division, compared to the other tethers.

Further, growth curves have been separated to group all MBH-J tethers (Figure 5.20A), all MBH-N tethers (Figure 5.20B), all G_5 linkers (Figure 5.20C), and all 9aa linkers (Figure 5.20D). WT and ΔFd3 are included in every dataset for comparison.

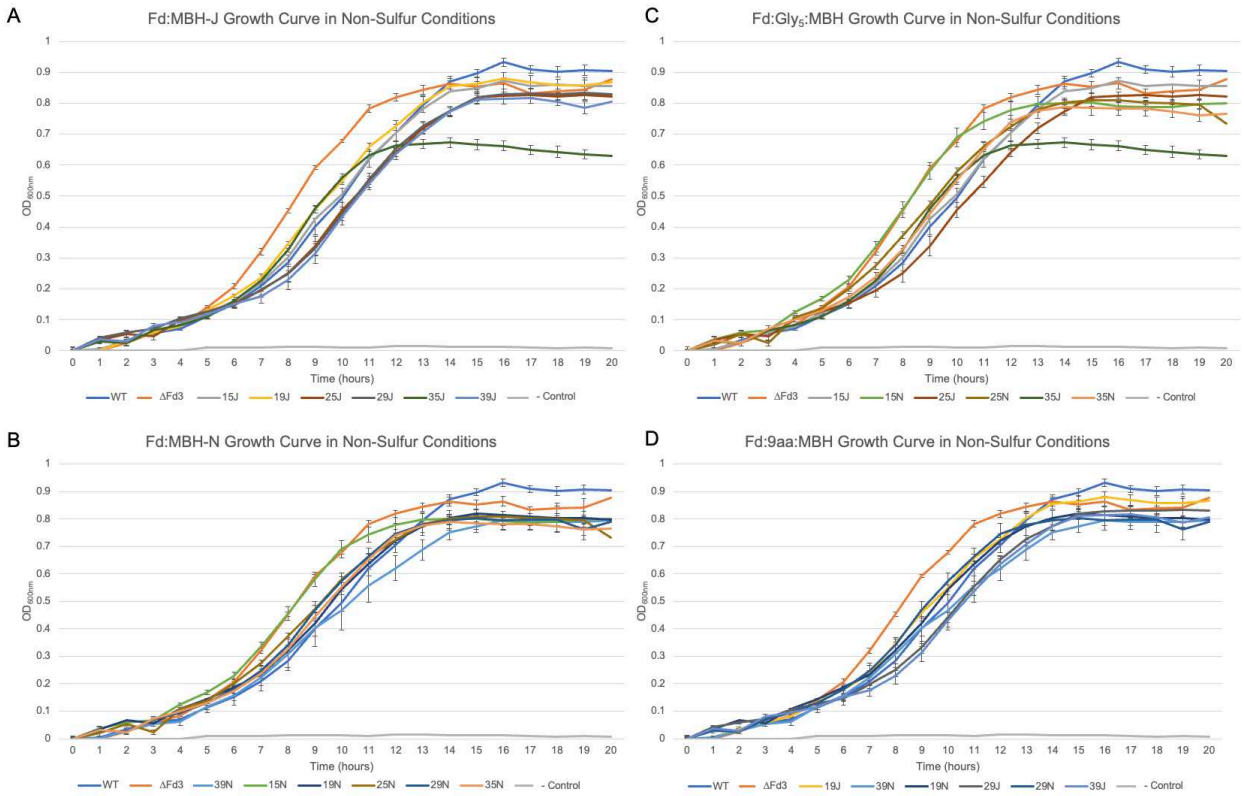


Figure 5.20: Growth curves of strains grouped by MBH-subunit or linker sequence in non-sulfur conditions. Tethers are grouped by MBH-J (A) or MBH-N (B) subunit or by linker sequence where all G_5 tethers are shown in (C) and all 9aa tethers are shown in (D). All growth curves include WT (blue), ΔFd3 (orange), and the uninoculated negative control (light grey, bottom). No MBH-J tethers (A) reached exponential before ΔFd3 (orange, A), however 19J (yellow, A) and 35J (green, A) reached exponential before WT (blue, A), 15J (light grey, A) mirrored WT, and 25J, 29J, and 39J (brown, dark grey, and green, respectively, A) reached exponential after WT. All MBH-J tethers reached lower maximum OD_{600} than WT, while 15J (light grey, A) and 19J (yellow, A) mirrored ΔFd3 (orange, A). 25J (brown, A), 29J (dark grey, A), and 39J (light blue, A) showed similar maximum OD_{600} (~0.82), 35J (green, A) clearly showed a reduced OD_{600} in stationary phase (~0.65). MBH-N tethers (B) show 15N (green, B) mirroring ΔFd3 (orange, B) in exponential phase, while the majority of N-tethers mirror WT in exponential phase with the exception of 39N (light blue, B) which showed a slower rate of growth (indicated by a flatter slope between 10-14 hours of growth). All MBH-N tethers showed a similar reduced maximum OD_{600} compared to WT and ΔFd3 at stationary phase (~0.8). Tethers containing a G_5 linker (C) displayed exponential phase growth similar to WT with the exception of 15N (green, C) which grew quickly like ΔFd3 and 25J (brown, C) which reached exponential phase slightly later than the other G_5 tethers. No G_5 tethers reached a greater maximum OD_{600} than WT, with 35J (dark green, C) showing the lowest maximum OD_{600} (~0.68). The 9aa linker (D) tethers took longer to reach exponential phase than ΔFd3, with the majority mirroring WT. 29J (dark grey, D) and 39J (grey-blue, D) took the longest to reach exponential phase. All 9aa tethers reached similar maximum OD_{600} which were less than WT and ΔFd3 except for 19J (yellow, D) which mirrored ΔFd3. Error bars depict standard error.

When comparing MBH subunits and linkers, 35J reached exponential slightly earlier than all other MBH-J tethers but returned an overall lower OD_{600} at stationary phase. All 9aa tethers reached exponential slightly later and reached a lower maximum OD_{600} compared to ΔFd3 and WT, suggesting that the 9aa linker does not provide a fitness advantage as measured by cell division.

To compare all possible groups, data for each Fd, each MBH-subunit, and each linker has been averaged and plotted (Figure 5.21).

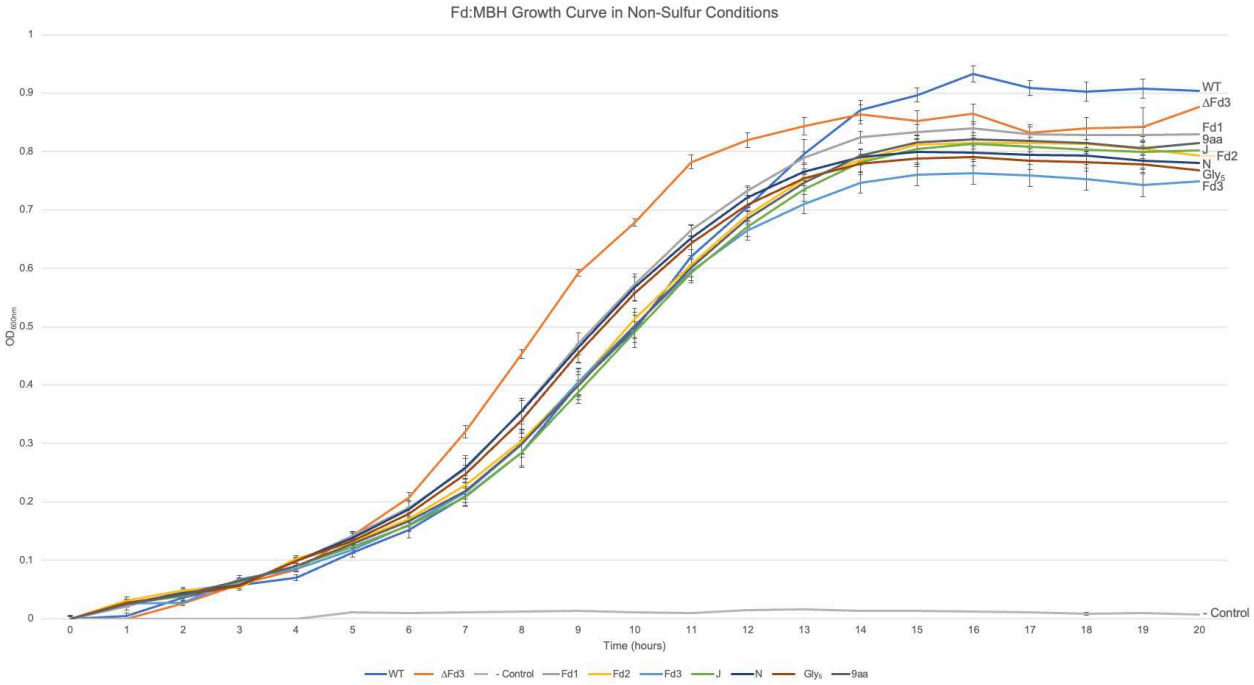


Figure 5.21: Growth curve of strains grouped by Fd, linker, or MBH subunit in non-sulfur conditions. To observe growth phenotypes between combinations of strains, growth curve data has been grouped by Fd, the linker, or the MBH subunit. All strains which include Fd1 (15J, 19J, 15N, and 19N) have been averaged to return results depicted by the grey line. All strains which include Fd2 (25J, 29J, 25N, and 29N) have been averaged to return results depicted by the yellow line. All strains which include Fd3 (35J, 39J, 35N, and 39N) have been averaged to return results depicted by the light blue line. All strains which include MBH-J (15J, 19J, 25J, 29J, 35J, and 39J) have been averaged to return results depicted by the green line. All strains which include MBH-N (15N, 19N, 25N, 29N, 35N, and 39N) have been averaged to return results depicted by the dark blue line. All strains which include a G_5 linker (15J, 25J, 35J, 15N, 25N, and 35N) have been averaged to return results depicted by the brown line. All strains which include a 9aa linker (19J, 29J, 39J, 19N, 29N, and 39N) have been averaged to return results depicted by the dark grey line. WT is shown in blue and ΔFd3 is shown in orange. ΔFd3 (orange) showed the fastest growth in exponential phase, with G_5, N, and Fd1 (brown, dark blue, grey) tethers showing the next fastest growth in exponential. While WT (blue) showed the greatest OD_{600} at stationary phase, Fd3 tethers showed the lowest OD_{600} at stationary phase.

When data are grouped and averaged, some interesting trends emerge. Tethers which include Fd1, MBH-N, or a G_5 linker reached exponential phase before Fd2, Fd3, MBH-J, or 9aa tethers (Figure 5.21). Tethers which contain Fd3, MBH-N, or the G_5 linker tend to result in lower maximum OD_{600} during stationary phase.

These data are the first documentation of a viable fusion protein in *T. kodakarensis.* Some tethered strains reached exponential faster than WT and some tethered strains grew to greater maximum OD_{600} during stationary phase than the parental (ΔFd3) strain, however in general, fusion constructs showed slightly reduced fitness in growth trials in non-sulfur conditions. Such a strain must bring metabolic machinery to the membrane to reduce a Fd while attached to MBH, thus posing some constraints on natural processes. The purpose of generating these tethered constructs was to determine if greater hydrogen gas could be produced by physically linking a Fd (electron donor) to a cytosolic subunit of the MBH complex, thus while growth data is of interest, a discussion of hydrogen production by the Fd:MBH tethered strains is ultimately the point in generating such strains, but first a comparison of growth phenotypes in sulfur vs. non-sulfur conditions is discussed.

Growth curve results of Fd:MBH tethered strains in sulfur and non-sulfur conditions

T. kodakarensis preferentially uses sulfur as the terminal electron acceptor when elemental sulfur (S°) is present. Extensive gene expression changes are observed within 10 minutes of sulfur addition to growing cultures which give rise to an upregulation of sulfur-metabolism genes and down-regulation of non-sulfur metabolism genes.[56] The thirteen-subunit membrane-bound sulfane reductase (*mbs*) is upregulated in the presence of sulfur while the fourteen-subunit *mbh* complex is downregulated.[9,55,56] A thorough analysis of growth and hydrogen production in the presence of sulfur provides for a comparison of the growth phenotype when sulfur is present (Figure 5.22).

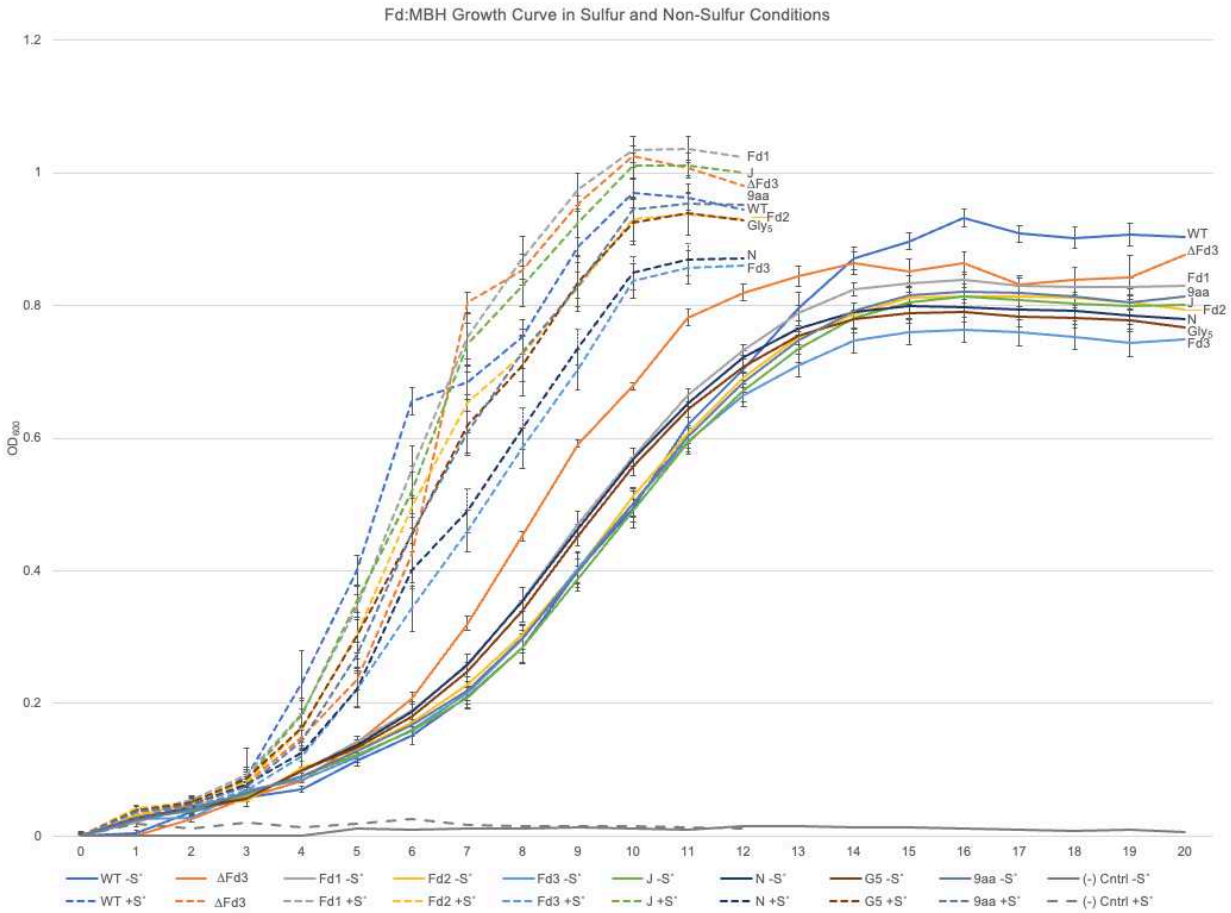


Figure 5.22: Growth curve of strains grouped by Fd, linker, or MBH subunit in sulfur and non-sulfur conditions. To compare growth outcomes in the presence (+S˚, dashed lines) and absence (-S˚, solid lines) of sulfur, growth curves of the Fd:MBH tethered strains were performed. For simplicity, growth curve data has been grouped by Fd, the linker, or the MBH subunit are shown. All strains which include Fd1 (15J, 19J, 15N, and 19N) have been averaged to return results depicted by the grey dashed or solid line. All strains which include Fd2 (25J, 29J, 25N, and 29N) have been averaged to return results depicted by the dashed or solid yellow line. All strains which include Fd3 (35J, 39J, 35N, and 39N) have been averaged to return results depicted by the dashed or solid light blue line. All strains which include MBH-J (15J, 19J, 25J, 29J, 35J, and 39J) have been averaged to return results depicted by dashed or solid the green line. All strains which include MBH-N (15N, 19N, 25N, 29N, 35N, and 39N) have been averaged to return results depicted by the dashed or solid dark blue line. All strains which include a G_5 linker (15J, 25J, 35J, 15N, 25N, and 35N) have been averaged to return results depicted by the dashed or solid brown line. All strains which include a 9aa linker (19J, 29J, 39J, 19N, 29N, and 39N) have been averaged to return results depicted by the dashed or solid dark grey line. WT is shown in blue and ΔFd3 is shown in orange. All strains reached exponential growth phase faster and reached a higher OD_{600} in +S˚. While Fd1:MBH tethers (grey dashed line) showed the greatest OD_{600} in +S˚, WT (solid blue line) showed the greatest OD_{600} in -S˚. Fd3:MBH tethers displayed the most reduced OD_{600} at stationary phase in both +S˚ and -S˚ conditions. Standard error is shown with error bars.

As expected, strains grown in the presence of sulfur reached exponential phase in less time (~ 4 hours less) and reached a higher maximum OD_{600} (~ 0.9 – 1.1) than the same strains grown in the absence of sulfur. While WT and ΔFd3 grew to the greatest OD_{600} in -S˚, tethers which included Fd1 or MBH-J grew to a greater maximum OD_{600} than WT or the parental strain. Fd3 and MBH-N tethers showed the most reduced growth in by exponential and stationary phase in +S˚ conditions, which was mirrored by Fd3 tethers in -S˚.

Hydrogen gas generation in Fd:MBH tethered strains in non-sulfur conditions

WT, ΔFd3, and Fd:MBH tethered strains were analyze for H_2 gas production during the growth curve. The culture headspace of freshly inoculated strains was purged and replaced with pure (99.99%) N_2 gas (AirGas, Ni UHP 300) at ~10 psi. Cultures were incubated at 85˚C and OD_{600} measurements were taken each hour. The pressurized headspace of the closed culture was used to inject 100 μL of gas at ambient pressure onto a Molecular Sieve 5A 60/80 mesh $1/8^{th}$ SS 2.0 m (Shimadzu, 220-94714-20). Pure N_2 is used as the carrier gas and flows over both the MS5 column and the reference column (Hayesep Q 80/100 mesh 1/8 stainless 2.0 m, Shimadzu, 220-94715-20) at 15 mL/min. A thermal coupled detector (TCD) at 110˚C measures the change in conductivity across the two columns as the gas from the headspace flows across the detector and a chromatogram readout displays a spike in voltage indicating the presence of hydrogen gas. Applying known standards of H_2 gas provides for a standard curve to be constructed. Integral values from samples are compared to the standard curve applied to the method and $\%H_2$ in the culture headspace can be assessed.

Headspace measurements were taken when cultures reached an OD_{600} of ~0.4, ~0.8, and at stationary phase (when OD_{600} had not increased for ~ >2 hours), thus three H_2 measurements were observed over the course of the growth curve. Each H_2 measurement

relieves the headspace of excess pressure, thus the pressure of the headspace returns to ambient pressure after each H_2 measurement.

While the percentage of H_2 in the headspace is of interest, this metric does not account for cell growth contributing to the production of hydrogen which is a more accurate representation of the capacity of a strain to produce hydrogen gas. Hydrogen production per cell, commonly referred to as the Ω value, accounts for the number of cells in a culture and is calculated by dividing the % H_2 by the OD_{600}.[12]

Hydrogen and OD_{600} measurements were averaged across four biological replicates. Hydrogen as a percentage of gas in the headspace when cultures had grown to ~0.4 OD_{600} ranged from ~5-10%, while % H_2 production increased to ~24-29% at 0.8 OD_{600}, and ~ 29-38% at stationary phase (Table 5.4 and Figure 5.23).

Table 5.4: Average hydrogen and OD_{600} measurements for WT, ΔFd3, and Fd:MBH tethers in non-sulfur conditions.

	H_2 at ~0.4 OD_{600}				H_2 at ~0.8 OD_{600}				H_2 at Stationary Phase					
Strain	**OD_{600}**	**Signal[a]**	**% H_2**	**Ω[b]**	**OD_{600}**	**Signal[a]**	**% H_2**	**Ω[b]**	**OD_{600}**	**Signal[a]**	**% H_2**	**Ω[b]**	**Total Ω[c]**	**% Δ[d]**
WT	0.408	45309.8	5.35	13.13	0.817	210845.3	24.27	29.73	0.906	295321.5	33.94	37.51	70.17	-1.22
ΔFd3	0.430	44543.8	5.26	12.26	0.835	243530.8	28.01	33.60	0.998	327192.3	37.59	38.50	71.03	0.00
15J	0.417	54109.0	6.35	15.23	0.815	213615.3	24.59	30.16	0.861	306340.0	35.20	40.91	76.79	8.11
19J	0.398	50284.0	5.92	14.90	0.811	222729.8	25.63	31.60	0.888	312565.3	35.91	40.63	75.99	6.97
39N	0.470	62418.8	7.30	15.56	0.785	221784.0	25.52	32.54	0.793	282702.3	32.49	40.96	82.34	15.92
15N	0.426	43725.5	5.17	12.08	0.784	221686.8	25.51	32.61	0.795	284470.0	32.70	41.09	79.69	12.18
19N	0.426	48302.8	5.69	13.38	0.797	228997.5	26.35	33.10	0.805	281323.0	32.34	40.23	79.99	12.61
25J	0.477	65004.5	7.60	15.87	0.821	215114.8	24.76	30.20	0.827	280082.8	32.19	38.93	78.05	9.88
29J	0.493	63648.8	7.44	15.11	0.831	226102.5	26.02	31.36	0.831	282184.5	32.43	39.13	79.34	11.69
25N	0.414	48949.8	5.76	13.93	0.808	227793.5	26.21	32.45	0.799	293288.0	33.70	42.19	82.23	15.75
29N	0.476	59588.3	6.98	14.72	0.788	238075.0	27.39	34.92	0.812	272028.8	31.27	38.67	80.86	13.83
35J	0.537	69994.8	8.17	15.22	0.656	247925.8	28.51	43.53	0.639	254381.0	29.25	45.81	103.14	45.19
39J	0.476	65634.8	7.67	16.05	0.825	214295.0	24.66	29.97	0.969	284592.5	32.71	34.49	67.14	-5.48
35N	0.554	78561.5	9.15	16.53	0.794	217386.0	25.02	31.73	0.776	266322.0	30.62	39.54	83.54	17.60
MIN			5.17	12.08			24.27	29.73			29.25	34.49	67.14	-5.48
MAX			9.15	16.53			28.51	43.53			37.59	45.81	103.14	45.19
Average			6.70	14.57			25.89	32.68			33.02	39.90	79.31	13.69

a: Signal refers to the area under the curve as integrated by Lab Solutions software (Shimadzu, 2010-2016).

b: Ω value is calculated from % H_2/OD.

c: Total Ω value is calculated by summing all %H_2 and dividing by OD at stationary phase.

d: Compared to ΔFd3 (parental) strain.

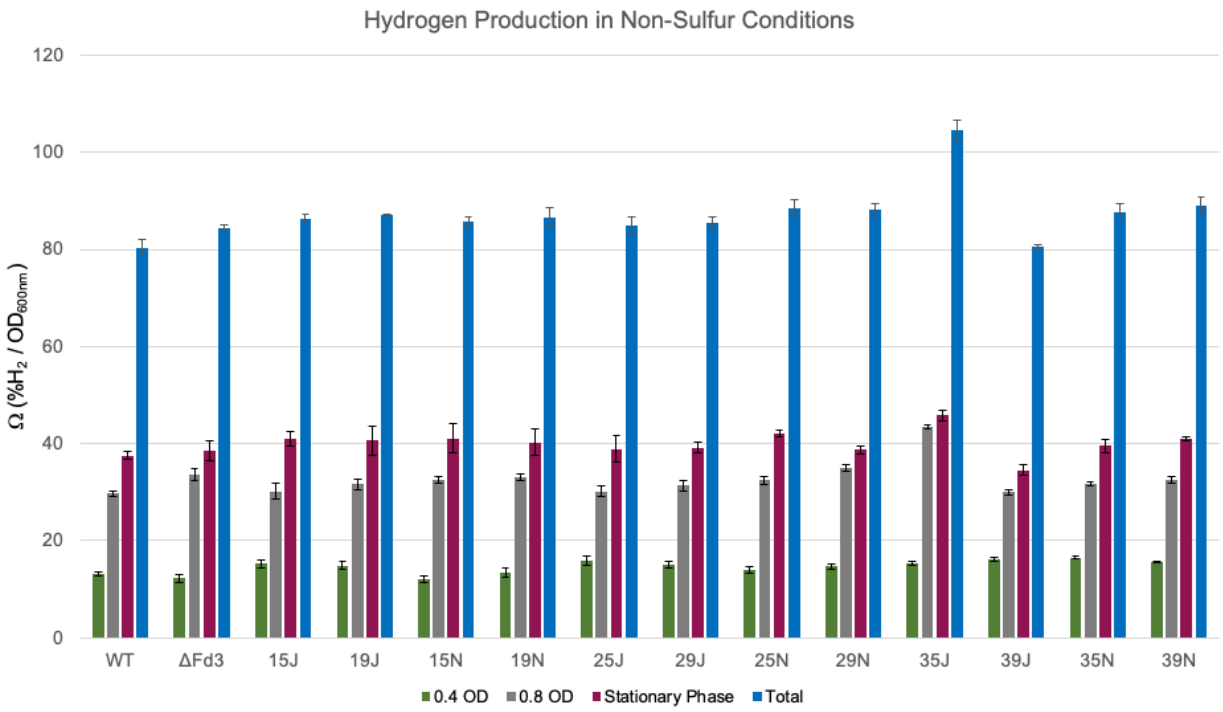


Figure 5.23: Hydrogen production at early exponential, entry into stationary, and stationary phase of Fd:MBH tether strains. Percent hydrogen production of all Fd:MBH tether strains, WT, and ΔFd3 was measured using a GC-TCD from the headspace of four biological replicates of actively growing cultures when they reached ~0.4 OD_{600} (green), ~0.8 OD_{600} (grey), and stationary phase (maroon). The total $\%H_2$ (a sum of the $\%H_2$ at 0.4, 0.8, and stationary phase) is divided by the end point OD_{600} measurement (blue). Ω values are calculated by dividing the percent H_2 by OD_{600} to provide a value which indicates hydrogen production per cell. Standard error from four biological replicates is depicted with error bars.

The Ω value shows that hydrogen production at the start of exponential phase (~0.4 OD_{600}, Figure 5.23 green bars) was relatively consistent among WT, ΔFd3, and tethered strains. Differences in $\%H_2$ begin to be apparent in late exponential/early stationary phase (0.8 OD_{600}, Figure 5.23 grey bars). By late exponential, 35J clearly produced more H_2 gas as indicated by the Ω value ~43.5, while all other strains produced < 35 Ω. $\%H_2$ measurements in stationary were on average higher than at 0.4 and 0.8, and again 35J was observed with the greatest $\%H_2$ in the headspace per cell density at ~45.8 Ω, while all other cultures produced < 43 Ω at stationary phase. While there are slight variations in H_2 production between all other strains, the clear increase in the Ω value for 35J stands out. Total Ω was higher for all tethered strains

compared to WT, with the exception of 39J, which demonstrated similar amounts of H_2 production per cell density compared to WT.

Interestingly, while 35J showed the greatest amount of total H_2 production, 39J showed the least, suggesting that the linker matters and that the shorter flexible linker is preferred when Fd3 is fused to MBH-J. Such a preference in linker composition was not observed for any other Fd:MBH-subunit combination; 35N and 39N performed similarly to one another. While both H_2 measurements at 0.8 and stationary phase were higher for 35J, the greatest difference between this tether combination and the others is observed ~0.8 OD_{600} where the Ω value is 8.5 units greater than the next highest producing strain (25N, 38.7 Ω at 0.8 OD_{600}). It should also be noted that 35J displayed the most reduced phenotype in growth (Figure 5.18) and never reached cell densities of 0.7 OD_{600}. The reduced growth phenotype does not result in a decrease in hydrogen yield, suggesting that 35J may be putting energy into hydrogen production as opposed to cell division.

A more detailed analysis reveals that all tethered strains produced more $\%H_2$ per cell density than WT at 0.4 OD_{600} with the exception of 15N. While some strains showed similar amount of H_2 at 0.8 OD_{600}, most tethered strains (15N, 19N, 25N, 29N, and 35J) produced notably more $\%H_2$. Interestingly the strains which produced more H_2 at 0.8 OD_{600} were tethered to MBH-N, with the exception of 35J, which produced considerably more H_2 than the other strains noted. It seems that tethering Fd1 or Fd2 to MBH-N may be of benefit for H_2 production in late exponential/early stationary phase.

Hydrogen gas generation of Fd:MBH tethered strains in sulfur conditions

Hydrogen gas output was measured in sulfur conditions at 0.4 OD_{600}, 0.8 OD_{600}, and at stationary phase (Table 5.5, Figure 5.24) as was performed in non-sulfur conditions. The Ω value is calculated by dividing the $\%H_2$ by the OD_{600} to account for cell density. Cultures in

+S˚ showed similar hydrogen production at 0.4 OD_{600}, but Ω values did not increase as significantly during exponential or stationary phase as was observed in -S˚ conditions.

Table 5.5: Average hydrogen and OD_{600} measurements for WT, ΔFd3, and Fd:MBH tethers in sulfur conditions.

	H_2 at ~0.4 OD_{600}				H_2 at ~0.8 OD_{600}				H_2 at Stationary Phase				Total	
Strain	**OD_{600}**	**Signal[a]**	**% H_2**	**Ω[b]**	**OD_{600}**	**Signal[a]**	**% H_2**	**Ω[b]**	**OD_{600}**	**Signal[a]**	**% H_2**	**Ω[b]**	**Ω[c]**	**% Δ[d]**
WT	0.402	28746.3	3.46	8.58	0.858	180016.0	20.74	24.19	0.911	228190.0	26.25	28.86	55.41	26.83
ΔFd3	0.409	36186.5	4.31	10.55	0.781	154275.5	17.80	22.81	1.009	215601.3	24.81	24.60	46.50	6.43
15J	0.413	27192.0	3.28	7.94	0.780	133755.0	15.45	19.81	1.056	238168.8	27.40	25.97	43.69	0.00
19J	0.399	28398.0	3.42	8.59	0.801	156734.8	18.08	22.55	1.046	256566.0	29.50	28.22	48.77	11.62
39N	0.492	47201.3	5.56	11.32	0.921	137321.0	15.86	17.41	1.154	225365.5	25.93	22.48	41.05	-6.06
15N	0.396	40015.3	4.74	12.00	0.840	164095.5	18.92	22.53	0.930	201489.3	17.44	24.78	44.22	1.20
19N	0.459	30963.3	3.71	8.08	0.785	124813.5	14.43	18.36	1.040	245604.0	28.25	27.17	44.63	2.13
25J	0.424	37254.3	4.43	10.46	0.901	163149.0	18.81	20.96	0.979	221249.3	19.14	26.31	43.29	-0.92
29J	0.390	38752.3	4.60	11.84	0.765	163464.5	18.85	24.77	0.757	179829.3	20.72	27.44	58.37	33.59
25N	0.461	45826.0	5.41	11.76	0.866	166255.0	19.17	22.23	0.902	207005.0	23.83	26.44	53.67	22.83
29N	0.419	47344.5	5.58	13.31	0.846	160502.3	18.51	22.02	0.916	206733.0	23.80	26.00	52.27	19.63
35J	0.443	34212.3	4.08	9.12	0.910	145801.5	16.83	18.49	1.047	182903.8	21.07	20.10	40.10	-8.22
39J	0.397	43818.3	5.18	13.07	0.743	159890.0	18.44	24.83	0.805	166979.0	14.48	23.68	47.33	8.33
35N	0.501	53944.3	6.33	12.74	0.849	140262.3	16.20	19.10	0.928	179413.3	20.68	21.89	46.59	6.62
MIN			3.28	7.94			14.43	17.41			14.48	20.10	40.10	-8.22
MAX			6.33	13.31			20.74	24.83			29.50	28.86	58.37	33.59
Average			4.58	10.67			17.72	21.43			23.09	25.28	47.56	8.86

a: Signal refers to the area under the curve as integrated by Lab Solutions software (Shimadzu, 2010-2016).

b: Ω value is calculated from % H_2/OD.

c: Total Ω value is calculated by summing all %H_2 and dividing by OD at stationary phase.

d: Compared to ΔFd3 (parental) strain.

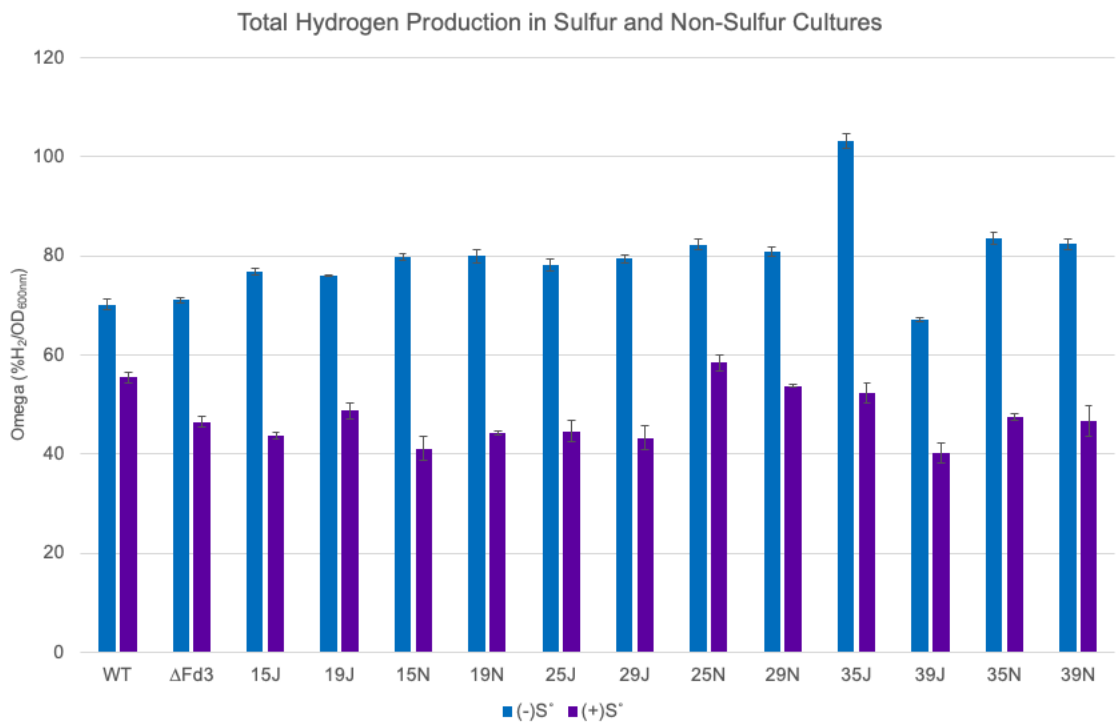


Figure 5.24: Total hydrogen production of Fd:MBH tether strains in sulfur and non-sulfur conditions. Percent hydrogen at 0.4 OD, 0.8 OD, and stationary phase has been summed and divided by the final OD measurement to provide the Ω value indicating hydrogen production per cell in non-sulfur (-S˚, blue) and sulfur (+S˚, purple) conditions. Ω values in +S˚ are lower than in -S˚, with the maximum Ω in +S˚ reaching ~58 Ω for 25N. Most tethered strains produced less hydrogen per cell than WT in +S˚. Interestingly, 39J produced the least amount of hydrogen in both +S˚ and -S˚ conditions compared to all other strains. Standard error is shown with error bars.

5.3.2 Phenotypic outcomes of the Fd2:GGR tethered strain

GGR which is the monomeric protein responsible for lipid saturation in *T. kodakarensis*,[22,45] and significant evidence suggests that Fd2 is the likely electron donor to GGR.[5,42] While three strains wherein Fd2 is fused to GGR have been attempted with linkers of various lengths (Figure 5.13), only the longest fusion strain, 224G, has been confirmed (Figure 5.14).

T. kodakarensis maintains a diverse array of ether linked isoprenoid lipid moieties (Figure 5.17),[57] and an analysis of lipid composition in the WT strain and 224G has revealed significant changes of both diether and tetraether lipids in the tethered strain (Figure 5.25). While the WT strain maintains 0.1 double bonds (DB) in tetraether lipids and 0.6 DB in diether

lipids, the 224G tethered strain maintained 1.2 DB and 2.5 DB in tetraether and diether lipids, respectively (Figures 5.26 and 5.27).

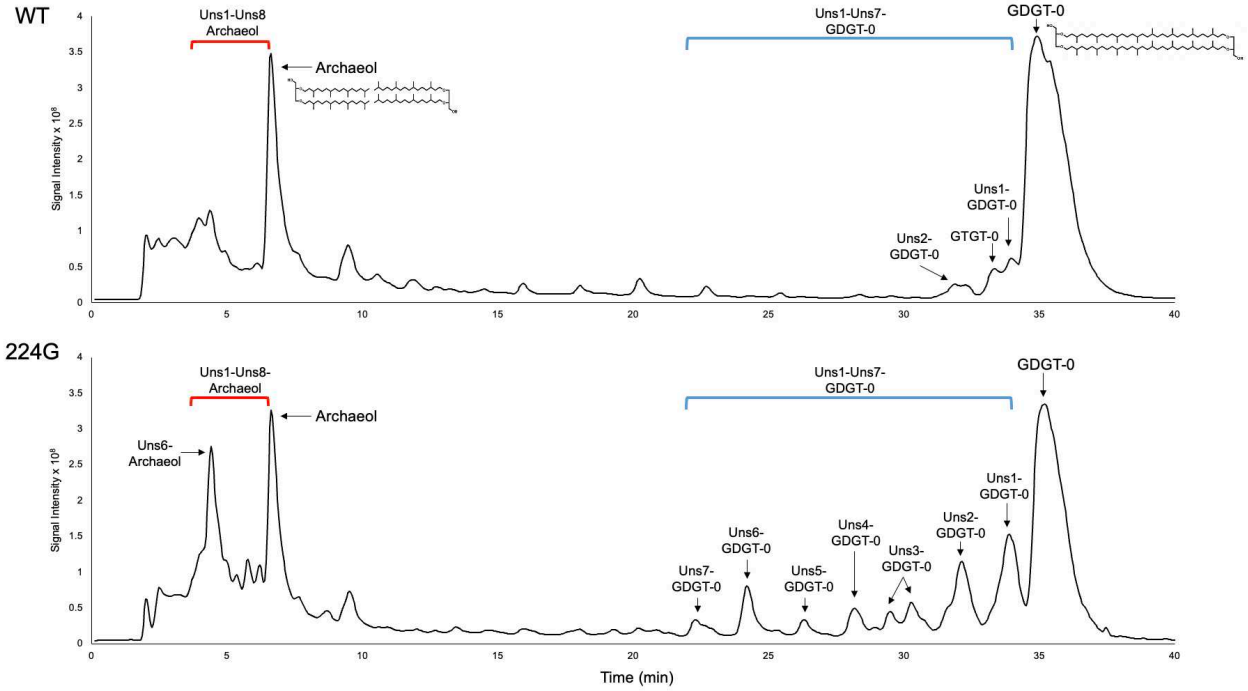


Figure 5.25: Lipid composition of WT and 224G tether. Total ion chromatographs of acid hydrolyzed lipids show lipid composition of WT (top) and 224G (bottom) strains grown to stationary phase. The number of unsaturated bonds (double bonds) is indicated by "Uns-#". Peaks from diether lipid appear at left (~ 2-10 minutes) while peaks from tetraether lipids appear at right (~ 22-38 minutes). Fully saturated archaeol lipids generate a narrow peak (~7.5 minutes, signal intensity of ~3.5 x 10^8) while fully saturated tetraether lipids generate a wider and taller peak (~36 minutes, signal intensity of ~3.7 x 10^8). Unsaturated archaeol lipids appear in peaks preceding saturated archaeols (~ 2-6 minutes, red bracket), and unsaturated tetraether lipids appear preceding the saturated GDGT moiety (~ 22-35 minutes, blue bracket). The tethered strain produces more unsaturated archaeols and unsaturated GDGT moieties as observe by the multiple peaks from ~2-6 minutes and ~22-33 minutes, respectively, in the lower chromatogram.

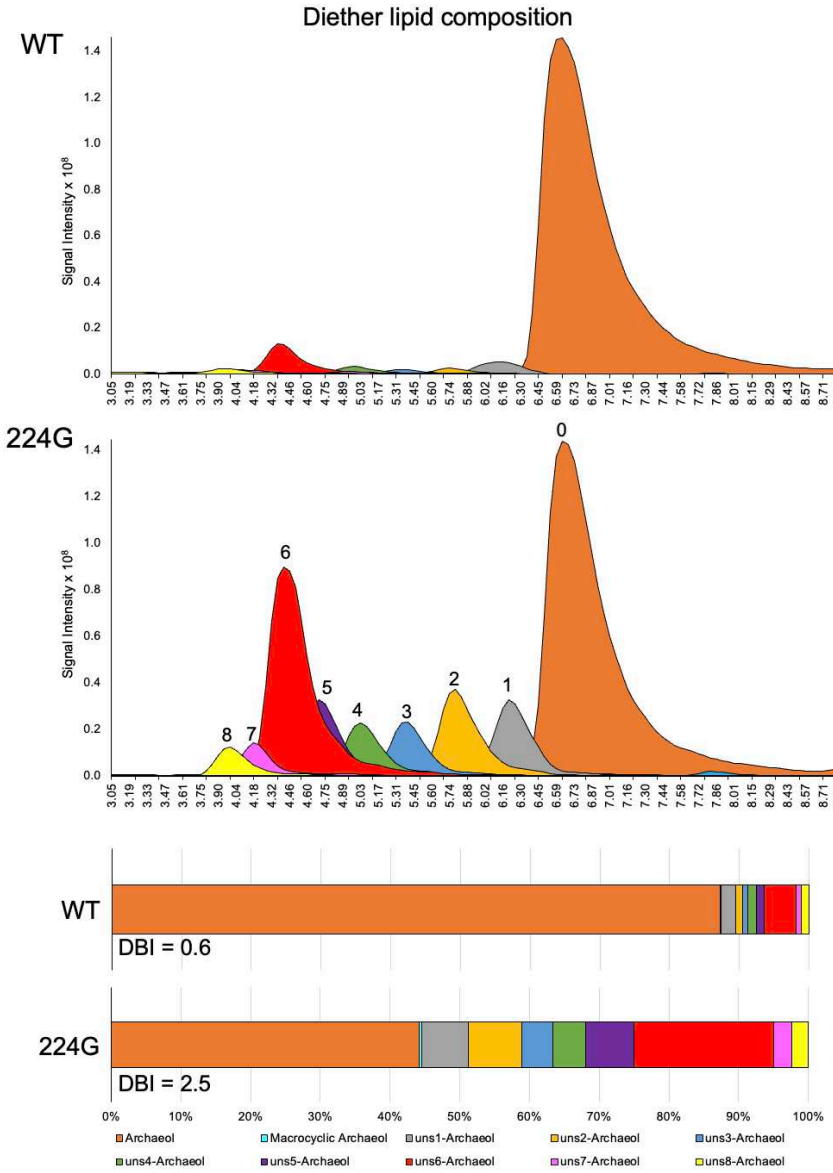


Figure 5.26: Composition of diether lipids in WT and 224G tethered strains. Total ion chromatograph of acid hydrolyzed lipids from WT (A) and 224G (B) strains grown to stationary phase show an abundance of fully saturated archaeol lipids (orange) in both WT and 224G strains. The number of unsaturated bonds (double bonds) is indicated by "Uns-#". Very few unsaturated lipids appear in the WT strain, with the exception of unsaturated-6-archaeols (red) which compose ~ 5% of WT diether lipids (C, red). The average number of double bonds (DBI) in the WT strain is 0.6. The 224G tether displays an abundance of unsaturated archaeols (B) as indicated by the numerous peaks with the number of unsaturated bonds displayed above each peak (B). The percentage of each lipid moiety for 224G is shown in D. While archaeols make up over 85% of diether lipids in WT, they compose < 45% of diether lipids in the 224G tether.

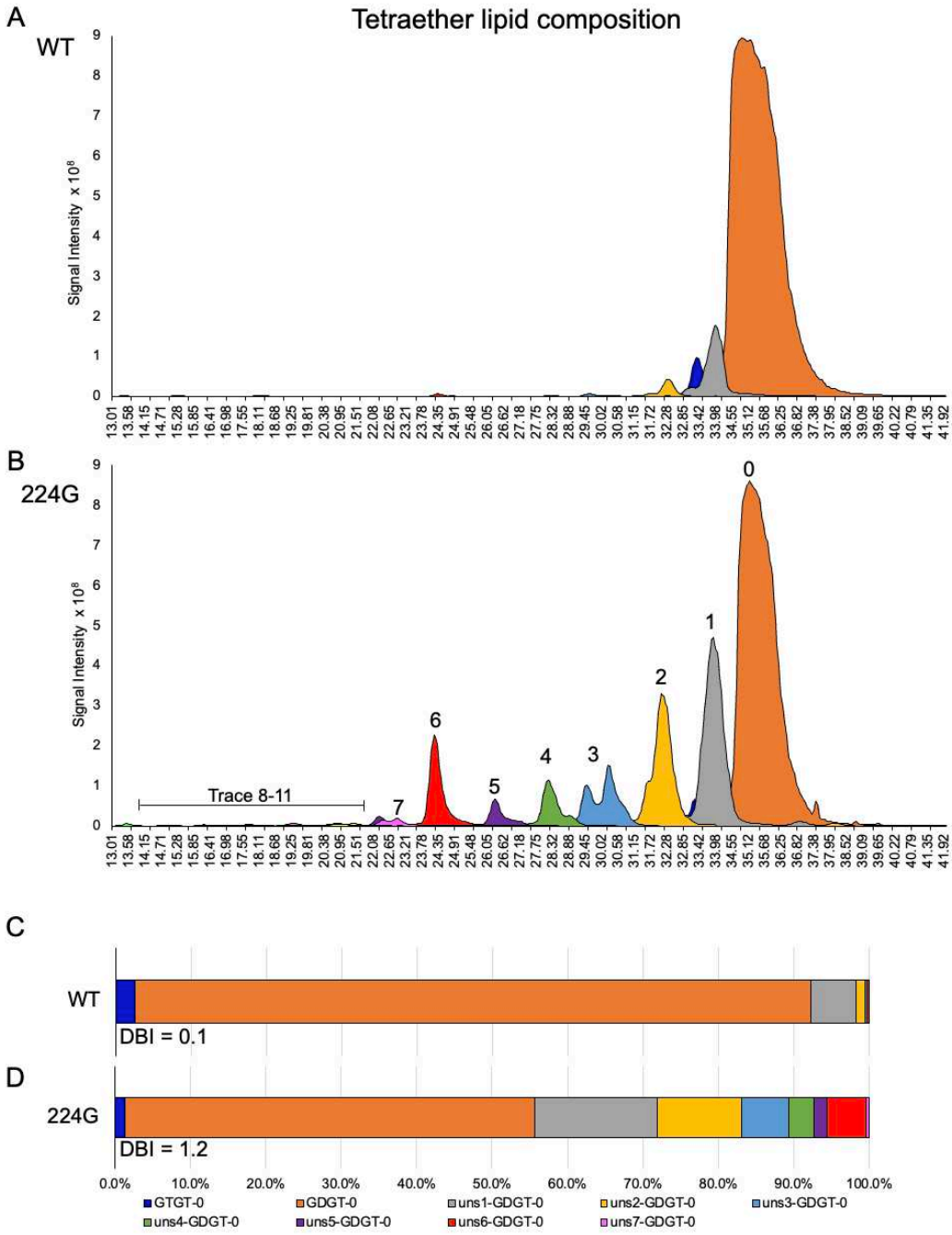


Figure 5.27: Composition of tetraether lipids in WT and 224G tethered strains. Total ion chromatograph of acid hydrolyzed lipids from WT (A) and 224G (B) strains grown to stationary phase show an abundance of fully saturated GDGT lipids (orange) in both WT and 224G strains. The number of unsaturated bonds (double bonds) is indicated by "Uns-#". Very few unsaturated lipids appear in the WT strain, with the exception of unsaturated-1-GDGT (grey) which compose ~ 6% of WT tetraether lipids (C, grey). The average number of double bonds (DBI) in tetraether lipids in the WT strain is 0.1. The 224G tether displays an abundance of unsaturated GDGTs (B) as indicated by the numerous peaks with the number of unsaturated bonds displayed above each peak (B). The percentage of each lipid moiety for 224G is shown in D. While GDGT-0 make up ~90% of tetraether lipids in WT, they compose ~55% of tetraether lipids in the 224G tether.

A clear abundance of double bonds in the 224G tethered strain is not seen in the WT strain, suggesting that fusing Fd2 to GGR via a 24aa linker sequence impacts the ability of GGR to fully saturate lipids. Such a dramatic change in lipid composition indicates that while Fd2 may donate electrons to GGR, the fusion impacts the native function of GGR in saturating both diether and tetraether lipids.

A disruption in GGR activity also identifies interesting findings as to the specific activity of GGR in the order in which double bonds are reduced. Previous findings suggest that double bonds are saturated proximal to the polar headgroup,[18] however these data come from *in vitro* studies of GGR reducing geranylgeranyl pyrophosphate (GGPP) which is not the natural substrate of GGR. Our findings suggest that unsaturated bonds distal to the headgroup are saturated first or that the double bond nearest the headgroup is saturated last (Figure 5.28).

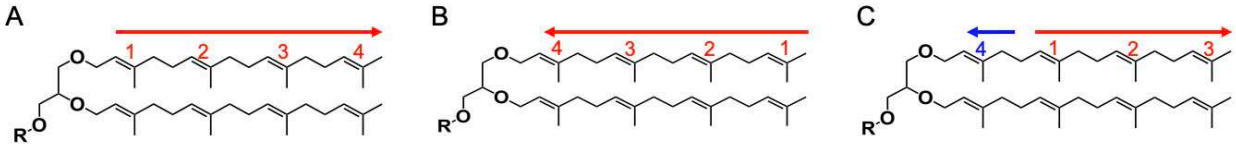


Figure 5.28: Saturation order of isoprene lipids by GGR reveals novel findings. Previous findings suggest that GGR saturates double bonds of isoprenoid lipids by reducing the bond proximal to the head group (A), however disruption to GGR activity imposed by the 224G tether results in lipid moieties which suggest otherwise. As shown in B and C, unsaturated lipids observed in these data suggest that bonds distal to the head group are saturated first while the bond most proximal to the head group is saturated last. Data from the lipid analysis in this study cannot distinguish form B or C, however it is observed that the final double bond remaining is the bond most proximal to the head group.

5.4 Discussion

Novel fusion proteins were generated to redirect electron flux. The physical linkage of two proteins encoded at the genomic level is a first in *T. kodakarensis*. All confirmed tethered strains are viable and grow with only moderate fitness impacts. A thorough collection of Fd:MBH fusion strains (twelve in total) were generated and assessed for growth and hydrogen production, while a single Fd2:GGR fusion strain was generated and impacted the saturation

profile of diether and tetraether lipids. A summary of the findings described in this chapter will be discussed here and plans for future experiments will be proposed.

Tethers are viable in T. kodakarensis

Thirteen novel tethered strains were generated in *T. kodakarensis.* Twelve strains were constructed by tethering Fd1, Fd2, or Fd3 to either the MBH-J or MBH-N subunit in the ΔFd3 parent strain such that if Fd3 was on present if tethered to an MBH-subunit. Neither *fd1* nor *fd2* can be deleted from the genome, thus all Fd1 and Fd2 tethers encoded two copies of *fd1* or *fd2;* the native copy at the original locus and a second copy at the fusion site within the *mbh* operon. One strain wherein Fd2 was fused to GGR was generated in the WT background. *Fd2* is encoded directly before *ggr* in the *surR* operon, thus the *fd2:ggr* fusion occurred at the native locus which allowed for every instance of the Fd2 protein to be tethered to GGR. While growth curves were not performed on the Fd2:GGR tether, cultivation of this strain for genetic confirmation and providing biomass for lipid analysis did not indicate a severe growth defect.

Fd:MBH tethers increase hydrogen production

Growth phenotypes and hydrogen production measurements with the Fd:MBH tethers identified mild growth impacts. While slight growth phenotypes were observed between tethered strains, no tethered strains reached exponential phase earlier than ΔFd3 and no tethered strains grew to a greater final OD_{600} than WT. 35J displayed a notable reduction in growth as this strain only reached a maximum OD_{600} of ~0.65 while most tethered strains grew ~ > 0.75 OD_{600} and WT grew to ~0.9 OD_{600}. Tethering Fd to MBH was predicted to modify H_2 output, which was observed. As all three Fds are predicted to maintain specific and optimal protein-partnerships for reductive flux, it was hypothesized that some Fds would be more natural donors to the MBH complex than others. All tethered strains generated more H_2 relative to cell density (Ω value)

compared to either ΔFd3 (the parental strain) or WT. Tethered strains generated ~13% more H_2 per cell on average than the parent (ΔFd3). 35J, which demonstrated the lowest OD_{600} at stationary phase, demonstrated the greatest production of H_2 gas relative to cell density, generating 45% more H_2 than the parent strain. There were minimal trends concerning H_2 production and Fds as Fd1-tethers generated similar amounts of H_2 as Fd2- and Fd3-tethers. Likewise, there were minimal trends concerning the MBH-subunit involved, as tethering a Fd to either subunit produced similar H_2 per cell. Likewise, there were minimal trends for the linker composition as a short, flexible linker (G_5) often resulted in similar H_2 production as the slightly longer yet more rigid 9aa linker (GGSPPPSG). The most notable deviation from average was apparent when Fd3 was tethered to the MBH-J subunit wherein the short, flexible (G_5) linker resulted in greater amounts of H_2 per cell than any other combination of Fd, MBH-subunit, or linker and when Fd3 was tethered to the MBH-J subunit with the longer, more rigid linker, 39J produced the lowest amount of total hydrogen per cell of any strain, suggesting that the linker sequence matters for the combination of Fd3 and MBH-J fusion.

Fd2:GGR tether impacted lipid composition

Fd2 was tethered to *ggr* via a sequence which encoded the longer, more rigid linker (9aa: GGSPPPSGG), the His_6 affinity tag, and the HA epitope tag resulting in a twenty-four amino acid tether. The Fd2:GGR fusion was evaluated for lipid composition at stationary phase. When compared to WT, the Fd2:GGR fusion resulted in significant changes to the composition of both diether and tetraether membrane lipids. The average number of double bonds in the diethers increased from 0.6 in WT to 2.5 in the tethered strain, and for the tetraether lipids, an increase from 0.1 to 1.2 double bonds was seen. Additionally, the order in which double bonds are saturated has previously been thought to be from more distal to more proximal (relative to

the head group),[18] however our data suggests that GGR saturates double bonds proximal to the head group last, which is a novel finding.

Further exploration of the activity of GGR and the relationship of Fd2 and GGR is warranted. Two additional Fd2:GGR strains are in process (25G and 29G), and once confirmed, the impact of the linker composition can provide further information as to the relationship between Fd2 and GGR. A shorter linker may provide more optimal electron transfer from Fd2 to GGR, thus a more similar saturation pattern to WT could be expected. Conversely, a shorter linker may restrict the activity of either Fd2 or GGR, rendering such a fusion incompatible. Evaluation of the Fd2:GGR tethers at multiple growth stages (*i.e*. early and late exponential) along with growth curves and protein expression analysis will provide valuable insights into the Fd2 and GGR partnership as well as refine our understanding of lipids as a reductant sink.

Redirecting electron flux from lipid sinks to hydrogen gas

The overarching goal of this project is to first recognize protein relationships which are vital to the flow of electrons and then to rationally direct this flux to specific products. Hydrogen gas is an important biofuel and lipids are a vast reductant sink, thus the redirection of electrons from lipids to hydrogen gas production is an enticing frontier in metabolic engineering.[15–17,66] The physical tethering of Fd3 to the MBH complex via a G_5 linker increased H_2 production by 45%. Tethering Fd2 to GGR via a 24aa linker reduced lipid saturation by ~40%. From the work presented here, we propose that next steps towards refining metabolic flux could include a strain which contains multiple fusion proteins, wherein both Fd2:GGR (specifically 224G) and Fd3:MBH-J (specifically 35J) are expressed. Such a strain could potentially inhibit the full saturation of lipids while funneling electrons to the generation of H_2 gas (Figure 5.29).

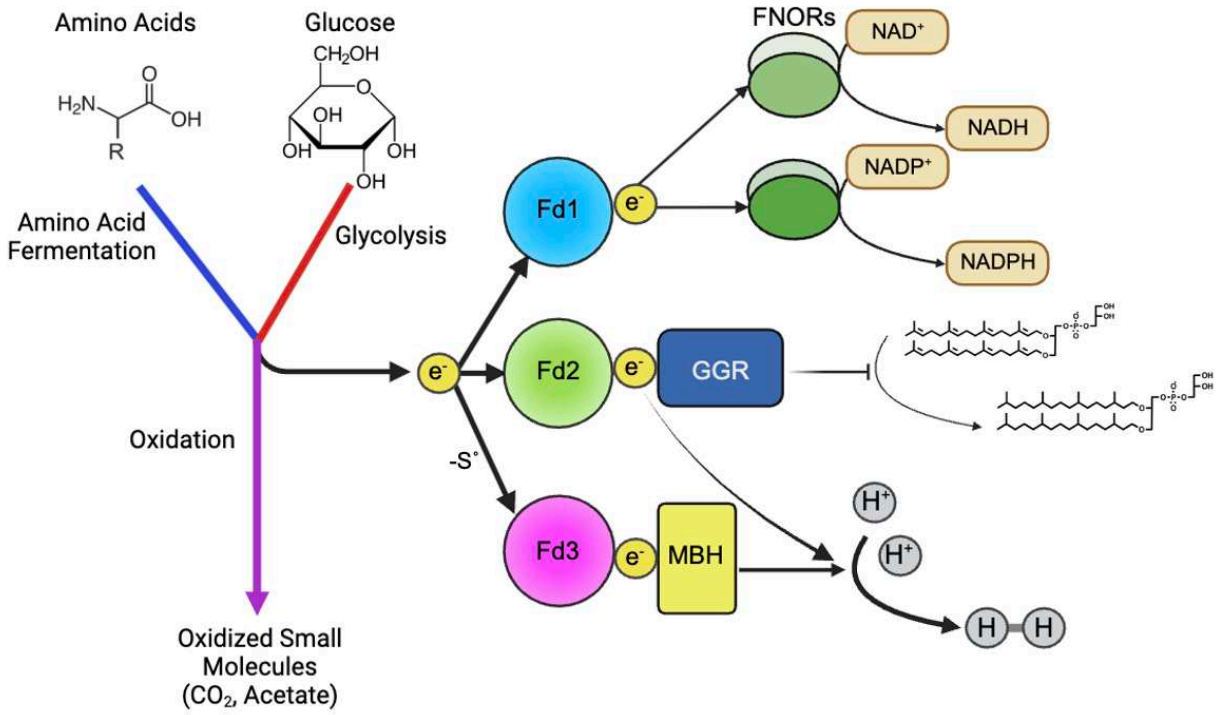


Figure 5.29: Rerouting electron flux from lipid saturation to the generation of H_2 gas. Amino acids and glucose are fermented to release electrons which are directed to distinct metabolic pathways via proteinaceous electron carriers such as Fds. Here, Fd1 (blue circle) directs electrons to the FNORs which generate reduced small molecules such as NAD(P)H. Fd2 (bright green circle) is physically tethered to GGR (dark blue rectangle) which saturates isoprene lipids, however this physical linkage restricts the activity of GGR such that 40% fewer double bonds are reduced. Fd3 (pink circle) is fused to the MBH complex at MBH-J via a short, flexible linker (G_5). The Fd3:MBH-J fusion increases the production of hydrogen gas by ~45% compared to WT. In a strain wherein both Fd2 is fused to GGR and Fd3 is fused to MBH, even greater H_2 yields could be possible.

REFERENCES

1. Hase, T., Schürmann, P. & Knaff, D. B. The interaction of ferredoxin with ferredoxin-dependent enzymes. in *Photosystem I: The Light-Driven Plastocyanin-Ferredoxin Oxidoreductase* (ed. Golbeck, J. H.) 477–498. (Springer, Dordrecht, The Netherlands, 2006).
2. Nzuza, N. *et al.* Diversification of ferredoxins across living organisms. *Curr Issues Mol Biol* **43**, 1374–1390 (2021).
3. Nielsen, M. S., Harris, P., Ooi, B. L. & Christensen, H. E. M. The 1.5 Å Resolution Crystal Structure of [Fe3S 4]-Ferredoxin from the Hyperthermophilic Archaeon Pyrococcus furiosus. *Biochemistry* **43**, 5188–5194 (2004).
4. Mayer, F. & Müller, V. Adaptations of anaerobic archaea to life under extreme energy limitation. *FEMS Microbiol Rev* **38**, 449–472 (2014).
5. Burkhart, B. W., Febvre, H. P. & Santangelo, T. J. Distinct physiological roles of the three ferredoxins encoded in the hyperthermophilic archaeon thermococcus kodakarensis. *mBio* **10**, (2019).
6. Atkinson, J. T., Campbell, I., Bennett, G. N. & Silberg, J. J. Cellular Assays for Ferredoxins: A Strategy for Understanding Electron Flow through Protein Carriers That Link Metabolic Pathways. *Biochemistry* **55**, 7047–7064 (2016).
7. Menon, A. L. *et al.* Novel multiprotein complexes identified in the hyperthermophilic archaeon pyrococcus furiosus by non-denaturing fractionation of the native proteome. *Molecular and Cellular Proteomics* **8**, 735–751 (2009).
8. Lipscomb, G. L., Schut, G. J., Scott, R. A. & Adams, M. W. W. SurR is a master regulator of the primary electron flow pathways in the order Thermococcales. *Mol Microbiol* **104**, 869–881 (2017).
9. Yang, H. *et al.* SurR regulates hydrogen production in Pyrococcus furiosus by a sulfur-dependent redox switch. *Mol Microbiol* **77**, 1111–1122 (2010).
10. Nguyen, D. M. N. *et al.* Two functionally distinct NADP+-dependent ferredoxin oxidoreductases maintain the primary redox balance of Pyrococcus furiosus. *Journal of Biological Chemistry* **292**, 14603–14616 (2017).
11. Hidese, R. *et al.* Gene regulation of two ferredoxin:NADP+ oxidoreductases by the redox-responsive regulator SurR in Thermococcus kodakarensis. *Extremophiles* **21**, 903–917 (2017).
12. Santangelo, T. J., Čuboňová, L. & Reeve, J. N. Deletion of alternative pathways for reductant recycling in Thermococcus kodakarensis increases hydrogen production. *Mol Microbiol* **81**, 897–911 (2011).
13. Andrews, S. C. *et al. A 12-Cistron Escherichia Coli Operon (Huf) Encoding a Putative Proton-Translocating Formate Hydrogenlyase System*. *Microbiology* vol. 143.
14. Maier, R. J. *RESPIRATORY METABOLISM IN HYPERTHERMOPHILIC ORGANISMS: HYDROGENASES. SULFUR REDUCTASES. AND ELECTRON TRANSPORT FACTORS THAT FUNCTION AT TEMPERATURES EXCEEDING 100°C*. (1996).
15. Lee, S. H., Kim, M. S., Kang, S. G. & Lee, H. S. Biohydrogen production of obligate anaerobic archaeon Thermococcus onnurineus NA1 under oxic conditions via overexpression of frhAGB-encoding hydrogenase genes. *Biotechnol Biofuels* **12**, (2019).
16. Kanai, T. *et al.* Continuous hydrogen production by the hyperthermophilic archaeon, Thermococcus kodakaraensis KOD1. *J Biotechnol* **116**, 271–282 (2005).
17. Aslam, M. *et al. Engineering of the Hyperthermophilic Archaeon Thermococcus Kodakarensis for Chitin-Dependent Hydrogen Production*. https://doi.org/10 (2017).
18. Kung, Y. *et al.* Constructing tailored isoprenoid products by structure-guided modification of geranylgeranyl reductase. *Structure* **22**, 1028–1036 (2014).

19. Jacquemet, A., Barbeau, J., Lemiègre, L. & Benvegnu, T. Archaeal tetraether bipolar lipids: Structures, functions and applications. *Biochimie* **91**, 711–717 (2009).
20. Anjum, N. A. *et al.* Lipids and proteins—major targets of oxidative modifications in abiotic stressed plants. *Environmental Science and Pollution Research* **22**, 4099–4121 (2015).
21. Murakami, M. *et al.* Geranylgeranyl reductase involved in the biosynthesis of archaeal membrane lipids in the hyperthermophilic archaeon Archaeoglobus fulgidus. *FEBS Journal* **274**, 805–814 (2007).
22. Jain, S., Caforio, A. & Driessen, A. J. M. Biosynthesis of archaeal membrane ether lipids. *Frontiers in Microbiology* vol. 5 Preprint at https://doi.org/10.3389/fmicb.2014.00641 (2014).
23. Yu, H. *et al.* Structure of an Ancient Respiratory System. *Cell* **173**, 1636-1649.e16 (2018).
24. Morikawa,' Yoshifumi Izawa,', M., Rashid,', N., Hoaki, T. & Imanakal, T. *Purification and Characterization of a Thermostable Thiol Protease from a Newly Isolated Hyperthermophilic Pyrococcus Sp*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* vol. 60 (1994).
25. Hachisuka, S. I., Sato, T. & Atomi, H. Metabolism dealing with thermal degradation of NAD+ in the hyperthermophilic archaeon Thermococcus kodakarensis. *J Bacteriol* **199**, 1–14 (2017).
26. Hachisuka, S., Sato, T. & Atomi, H. Hyperthermophilic Archaeon Thermococcus kodakarensis Utilizes a Four-Step Pathway for NAD+ Salvage through Nicotinamide Deamination. *J Bacteriol* **200**, 1–12 (2018).
27. Amend, J. P. & Shock, E. L. Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol Rev* **25**, 175–243 (2001).
28. Schäfer, G., Engelhard, M. & Müller, V. Bioenergetics of the Archaea. *Microbiology and Molecular Biology Reviews* **63**, 570–620 (1999).
29. Sapra, R., Bagramyan, K. & Adams, M. W. W. A simple energy-conserving system: Proton reduction coupled to proton translocation. *Proc Natl Acad Sci U S A* **100**, 7545–7550 (2003).
30. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
31. Zhang, C., Phillips, A. P. R., Wipfler, R. L., Olsen, G. J. & Whitaker, R. J. The essential genome of the crenarchaeal model Sulfolobus islandicus. *Nat Commun* **9**, (2018).
32. Mai, X. & Adams, M. W. W. Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon Pyrococcus furiosus. A new enzyme involved in peptide fermentation. *Journal of Biological Chemistry* **269**, 16726–16732 (1994).
33. Mai, X. & Adams, M. W. W. *Characterization of a Fourth Type of 2-Keto Acid-Oxidizing Enzyme from a Hyperthermophilic Archaeon: 2-Ketoglutarate Ferredoxin Oxidoreductase from Thermococcus Litoralis*. *JOURNAL OF BACTERIOLOGY* vol. 178 https://journals.asm.org/journal/jb (1996).
34. Schut~, G. J., Menon, A. L. & Adams, M. W. W. *2-Keto Acid Oxidoreductases from Pyrococcus and Thermococcus Litoralis*. (2001).
35. Heider, J., Mai, X. & Adams, M. W. W. *Characterization of 2-Ketoisovalerate Ferredoxin Oxidoreductase, a New and Reversible Coenzyme A-Dependent Enzyme Involved in Peptide Fermentation by Hyperthermophilic Archaea*. *JOURNAL OF BACTERIOLOGY* vol. 178 (1996).
36. Blarney, J. M. & Adams, M. W. W. *Purification and Characterization of Pyruvate Ferredoxin Oxidoreductase from the Hyperthermophilic Archaeon Pyrococcus Furiosus*. *Biochimica et Biophysica Acta* (1993).
37. Smith, E. T., Blarney, J. M. & Adams, M. W. W. Pyruvate Ferredoxin Oxidoreductases of the Hyperthermophilic Archaeon, Pyrococcus furiosus, and the Hyperthermophilic Bacterium, Thermotoga marítima, Have Different Catalytic Mechanisms'* 1".

38. Schut~, G. J., Menon, A. L. & Adams, M. W. W. *ENZYMES OF CENTRAL METABOLISM [12]*. (2001).
39. Nielsen, C. F., Lange, L. & Meyer, A. S. Classification and enzyme kinetics of formate dehydrogenases for biomanufacturing via CO2 utilization. *Biotechnology Advances* vol. 37 Preprint at https://doi.org/10.1016/j.biotechadv.2019.06.007 (2019).
40. Hutchins, A. M., Mai, X. & Adams, M. W. W. *Acetyl-CoA Synthetases I and II from Pyrococcus Furiosus*. (2001).
41. Bräsen, C., Esser, D., Rauch, B. & Siebers, B. Carbohydrate metabolism in archaea: Current insights into unusual enzymes and pathways and their regulation. *JAMA Ophthalmol* **132**, 326–331 (2014).
42. Isobe, K. *et al.* Geranylgeranyl reductase and ferredoxin from methanosarcina acetivorans are required for the synthesis of fully reduced archaeal membrane lipid in escherichia coli cells. *J Bacteriol* **196**, 417–423 (2014).
43. Peters, J. W. *et al.* [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochim Biophys Acta Mol Cell Res* **1853**, 1350–1369 (2015).
44. Atomi, H., Fukui, T., Kanai, T., Morikawa, M. & Imanaka, T. Description of Thermococcus kodakaraensis sp. nov., a well studied hyperthermophilic archaeon previously reported as Pyrococcus sp. KOD1. *Archaea* **1**, 263–267 (2004).
45. Fukui, T. *et al.* Complete genome sequence of the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1 and comparison with Pyrococcus genomes. *Genome Res* **15**, 352–363 (2005).
46. Hileman, T. H. & Santangelo, T. J. Genetics techniques for Thermococcus kodakarensis. *Front Microbiol* **3**, 1–12 (2012).
47. Gehring, A., Sanders, T. & Santangelo, T. J. Markerless Gene Editing in the Hyperthermophilic Archaeon Thermococcus kodakarensis. *Bio Protoc* **7**, (2017).
48. Scott, K. A., Williams, S. A. & Santangelo, T. J. Thermococcus kodakarensis provides a versatile hyperthermophilic archaeal platform for protein expression. in *Methods in Enzymology* vol. 659 243–273 (Academic Press Inc., 2021).
49. Santangelo, T. J., Čuboňová, L. & Reeve, J. N. Thermococcus kodakarensis Genetics: Tk1827-encoded β-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl Environ Microbiol* **76**, 1044–1052 (2010).
50. Reddy Chichili, V. P., Kumar, V. & Sivaraman, J. Linkers in the structural biology of protein-protein interactions. *Protein Science* vol. 22 153–167 Preprint at https://doi.org/10.1002/pro.2206 (2013).
51. Chen, X., Zaro, J. L. & Shen, W. C. Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews* vol. 65 1357–1369 Preprint at https://doi.org/10.1016/j.addr.2012.09.039 (2013).
52. Köhler, V. *Protein Design Methods and Applications Second Edition Methods in Molecular Biology 1216*. http://www.springer.com/series/7651.
53. Page, C. C., Moser, C. C., Chen, X. & Dutton, P. L. *Natural Engineering Principles of Electron Tunnelling in Biological Oxidation±reduction*. *NATURE* vol. 402 www.nature.com (1999).
54. Bakkes, P. J. *et al.* Design and improvement of artificial redox modules by molecular fusion of flavodoxin and flavodoxin reductase from Escherichia coli. *Sci Rep* **5**, (2015).
55. Lipscomb, G. L. *et al.* SurR: A transcriptional activator and repressor controlling hydrogen and elemental sulphur metabolism in Pyrococcus furiosus. *Mol Microbiol* **71**, 332–349 (2009).
56. Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J. & Reeve, J. N. Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis. *BMC Genomics* **15**, (2014).

57. Liman, G. L. S. *et al.* Tetraether archaeal lipids promote long-term survival in extreme conditions. *Mol Microbiol* (2024) doi:10.1111/mmi.15240.
58. Summons, R. E., Welander, P. V. & Gold, D. A. Lipid biomarkers: molecular tools for illuminating the history of microbial life. *Nature Reviews Microbiology* vol. 20 174–185 Preprint at https://doi.org/10.1038/s41579-021-00636-2 (2022).
59. Mcclure, P. J., Cole, M. B. & Davies, K. W. *An Example of the Stages in the Development of a Predictive Mathematical Model for Microbial Growth: The Effects of NaC1, PH and Temperature on the Growth of Aeromonas Hydrophila*. *International Journal of Food Microbiology* vol. 23 (1994).
60. Atomi, H. & Reeve, J. Microbe profile: Thermococcus kodakarensis: The modehyperthermophilic archaeon. *Microbiology (United Kingdom)* **165**, 1166–1168 (2019).
61. Garre, A., Koomen, J., Besten, H. M. W. D. & Zwietering, M. H. Modeling Population Growth in R with the biogrowth Package. *J Stat Softw* **107**, (2023).
62. Woese, C. R. & Fox, G. E. *Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms (Archaebacteria/Eubacteria/Urkaryote/16S Ribosomal RNA/Molecular Phylogeny)*. vol. 74 (1977).
63. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**, 4576–4579 (1990).
64. Rattray, J. E. & Smittenberg, R. H. Separation of Branched and Isoprenoid Glycerol Dialkyl Glycerol Tetraether (GDGT) Isomers in Peat Soils and Marine Sediments Using Reverse Phase Chromatography. *Front Mar Sci* **7**, (2020).
65. Farkas, J. A., Picking, J. W. & Santangelo, T. J. Genetic techniques for the Archaea. *Annual Review of Genetics* vol. 47 539–561 Preprint at https://doi.org/10.1146/annurev-genet-111212-133225 (2013).
66. Dietrich, H. M. *et al.* Membrane-anchored HDCR nanowires drive hydrogen-powered CO2 fixation. *Nature* **607**, 823–830 (2022).