5-2016

Oxidative Stress Response in Archaea: Elucidation of Oxidant Sensing and Tolerance Mechanisms in Methanosarcina acetivorans

Matthew Edward Jennings
University of Arkansas, Fayetteville

Oxidative Stress Response in Archaea: Elucidation of Oxidant Sensing and Tolerance Mechanisms in *Methanosarcina acetivorans*

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Biology

by

Matthew Edward Jennings
University of Scranton
Bachelor of Science in Biology, 2005
Villanova University
Master of Science in Biology, 2010

May 2016
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Dr. Daniel Lessner
Dissertation Director

_____
Dr. Ralph Henry
Committee Member

_____
Dr. Suresh Thallapuranam
Committee Member

_____
Dr. Inés Pinto
Committee Member

_____
Dr. David Ivey
Committee Member

ABSTRACT

Methanogens are archaea possessing a conserved metabolic pathway which produces methane. Many of the enzymes in the methanogenesis pathway are Fe-S proteins, meaning methanogens are sensitive to conditions which disrupt Fe-S clusters. Molecular oxygen is capable of disrupting Fe-S clusters through oxidation of the iron atoms. Furthermore, reduced iron can facilitate the production of reactive oxygen species (ROS), meaning methanogens must possess antioxidant mechanisms. Detection and eradication of ROS is important for all cells, due to the potentially fatal consequences of unchecked oxidation. This dissertation presents two separate projects investigating mechanisms the model methanogen *Methanosarcina acetivorans* possess for dealing with ROS. One project investigated the roles two [4Fe-4S] clusters present in RNA polymerase (RNAP) subunit D play in assembly and activity of RNAP; to determine if a mechanism exists for linking sensitivity of the clusters to oxygen to RNAP function. My data shows that both clusters and the cluster binding domain play an important role in assembly of RNAP downstream of D-L heterodimer formation, preventing optimal assembly of at least subunits B' and A'' when the clusters are absent. Cluster one plays a more critical role in this process compared to cluster two. Coupled with experimental evidence that the clusters are oxygen sensitive, this provides support for the hypothesis that the clusters regulate RNAP assembly in response to redox state of the cell. The second project investigated two putative catalase genes present in the *M. acetivorans* genome. Experimental evidence showed neither catalase was functional. Engineering of a *M. acetivorans* strain to express functional catalase from *Escherichia coli* increased the tolerance of *M. acetivorans* to H_2O_2, but not oxygen during growth in standard conditions. Catalase does not appear to be an important component in the oxidative stress response of *M. acetivorans.*

ACKNOWLEDGEMENTS

All science is collaboration, and this dissertation was finished with help from many different people. William Metcalf and his lab at the University of Illinois Urbana-Champaign developed the genetic system for *Methanosarcina acetivorans*, which I used liberally in my experiments. Liz Karr and her lab at the University of Oklahoma provided the protocol for the non-specific transcription assays.

Jim Wilson was my advisor at Villanova University where I pursued my M.S. degree. He encouraged me to continue my education, and if the experience doing research in his lab wasn't pleasant, I doubt I would have continued to get my doctorate. I learned many good lab practices from him, and I was able to hit the ground running here in Arkansas because of his guidance.

My fellow graduate students provided a sounding board for scientific questions, technical advice for experiments, and emotional support for the times when I questioned my sanity in continuing to pursue a doctorate. I want to thank Tom Deere, Addison McCarver, and Ryan Sheehan, in particular, for their help throughout the years. In retrospect, having other grad students join the lab was the best thing that could have happened, even with the extra noise and smells.

Faith Lessner deserves my eternal gratitude for providing additional hands for experiments when I was overwhelmed, technical advice for when I was stymied, delicious food for eating, and harsh criticism for when I did dumb things both in and out of the lab. This dissertation would not have come together without her help and support, and I appreciate everything that she did for me.

Finally, this dissertation owes much to the enthusiasm of my advisor, Dan Lessner. It was at his suggestion that I work on both projects, which turned out to be very fruitful from a publication standpoint. It was his excitement which drove me on, even which I was sick of all

my experiments failing. Observing other graduate students who have come and gone in our department has shown me that Dan was able to a walk a fine line between letting me figure things out on my own without making me feel like I was abandoned, and I always thought he cared deeply about the projects. I will be the first Ph.D. student that graduates from his lab, and I know I won't be the last.

DEDICATION

Committing to a Ph.D. is not something you should do lightly; it's a huge undertaking, and, in retrospect, I can definitely say it is not for everyone. However, my parents, Ken and Rosemary Jennings, always told me that I was a smart boy and that I could do anything if I put my mind to it. They encouraged me to go for it, comforted me when things were tough and it looked like I would never finish, and constantly offered financial assistance even though I kept telling them I was fine (I still accepted the money in holiday greeting cards though). They supported me even though they don't really understand what I do, but they know I'll have some extra letters after my name when I'm done. So, Mom and Dad, six years in Arkansas and I finally have the degree. Thanks in no small part to you.

My sister, Lauren Jennings, was able to commiserate with me over my graduate student woes in ways my parents couldn't. I was able to vent to her about my parents when they started getting on my case about when I was planning on finishing, what my plans were following graduation, and how much money I'd be making. Thanks for being a sympathetic ear sis, and I seem to have one upped you in the degree department.

TABLE OF CONTENTS

LIST OF PUBLISHED PAPERS

Chapter I - Lessner FH, Jennings ME, Hirata A, Duin EC, Lessner DJ. Subunit D of RNA Polymerase from Methanosarcina acetivorans Contains Two Oxygen-labile [4Fe-4S] Clusters: Implications for Oxidant-Dependent Regulation of Transcription. *Journal of Biological Chemistry.* 2012 287(22):18510-18523. (http://www.jbc.org/content/287/22/18510.full)

Chapter II – Jennings ME, Lessner FH, Karr EA, Lessner DJ. The [4Fe-4S] clusters of subunit D are key determinants in the post D-L heterodimer assembly of RNA polymerase in *Methanosarcina acetivorans*. *Molecular Microbiology.* Submitted 2016.

Chapter III - Jennings ME, Schaff CW, Horne AJ, Lessner FH, Lessner DJ. Expression of a bacterial catalase in a strictly anaerobic methanogen significantly increases tolerance to hydrogen peroxide but not oxygen. *Microbiology.* 2014, 160(Pt 2):270-278. (http://mic.sgmjournals.org/content/160/Pt_2/270.long)

INTRODUCTION

When investigation of natural phenomena was codified into science, people began to understand that the sheer diversity of life on Earth is staggering. Early attempts to categorize living things did so by morphology, but microorganisms were harder to categorize because differences in morphology at a cellular level are hard to distinguish, especially without specialized instruments. Carl Woese and his collaborators brought about a revolution in how scientists categorize living things by measuring differences in genetic sequences as a way to estimate how closely organisms are related (Fox *et al.*, 1977). Due largely to their work, current biologists classify all known species into three distinct evolutionary lineages; the domains Bacteria, Eukarya, and Archaea (Woese *et al.*, 1990).

Archaea occupy a unique niche in microbiology that has, historically, not received the same level of investigation as bacteria or eukaryotes. Archaea were originally classified as bacteria due to their similarities; both are single-celled organisms which lack membrane bound organelles, including nuclei, and typically possess a single circular chromosome. When Woese et al determined that archaea, as a group, were entirely distinct from bacteria, an increase in research interest occurred but still lagged behind bacteria and eukaryotes. A likely reason for this disparity is that there are no known pathogens, of humans or any other organisms, among the archaea; understanding and treating the agents of major diseases was a huge driver of early microbiology research. Among archaea are many species which specialize in surviving in extreme environments, including halophiles, acidophiles, and thermophiles (Woese *et al.*). The high number of extremophiles has led scientists to speculate that archaea arose early in Earth's history, since the environment of ancient Earth was much more extreme compared to modern Earth (Gribaldo & Brochier-Armanet, 2006). However, environmental sequencing projects have

revealed the presence of archaea in a wide variety of environments as well, indicating that modern archaea have not been relegated exclusively to extreme environments. One widely distributed group organisms within the domain exhibit the unique capability of synthesizing methane as part of their energy conserving metabolic pathway; the methanogens.

Methanogens and methanogenesis

Methanogens are obligate anaerobes belonging to the Phylum Euryarchaeota in the Domain Archaea. Although methanogens are strict anaerobes, they can be found in a variety of environments which experience only periodic episodes without oxygen. Many are aquatic organisms living in the deep sediment of both fresh water, including lakes, rivers, and rice paddies, and salt water including the oceans. Others are part of the intestinal microbiome of humans and other animals, especially cows and other ruminants. Research has shown that methanogens play an important role in gut microbiota homeostasis in humans (Gaci *et al.*, 2014). Still others are found in lower layers of soil in environments ranging from deserts to frozen tundra. These organisms comprise an important component of the global carbon cycle because they are capable of utilizing the terminal components generated by the metabolism of other organisms (**Fig. 1**). Methanogens have been the subject of considerable research because they are the only organisms known to produce methane. Estimates suggest methanogens are responsible for the emission of one gigaton of methane annually (Thauer *et al.*, 2008). Methane is a more potent greenhouse gas compared to CO_2; absorbing 20% more heat per molecule, meaning methanogens are an important component to consider when addressing global climate change (Ipcc, 2014). Methane is also a potential fuel source for human civilization, and methanogens have been harnessed for generation of methane from organic precursors using bioreactors filled with microbial consortia.

Methanogens utilize a unique central metabolic pathway called methanogenesis, which reduces simple carbon compounds to methane and couples that reduction with the generation of an ion gradient, which is then used to synthesize ATP (Lessner, 2001). Three major pathways of methanogenesis exist, which differ in the starting substrate and electron source but share the same terminal steps (**Fig. 2**). Most methanogens grow using only CO_2 as the carbon source and H_2 as the electron donor. Certain members of the Order Methanosarcinales can utilize other organic compounds as the carbon source such as acetate or methylated compounds (e.g. methanol and methyl amines). Carbon dioxide is reduced using an electron donor (usually H_2) in a stepwise fashion to a methyl group bound to the co-factor tetrahydromethanopterin (THMPT). When acetate is the growth substrate, it is first converted to acetyl-CoA which is then split through the actions of the enzyme complex dehydrogenase/acetyl-CoA synthase (Cdh/Acs). The carboxyl group is oxidized to CO_2 while the methyl group is transferred to THMPT. In both pathways, the enzyme THMPT:coenzyme M methyltransferase (Mtr) then catalyzes the transfer of the methyl group from THMPT to coenzyme M (CoM). During growth on methylated compounds, specialized transfer enzymes move the methyl group directly to CoM. Once a methyl group is bound to CoM, all three pathways of methanogenesis converge. The enzyme methyl-coenzyme M reductase (Mcr) reduces the methyl group to methane which is then released, and couples this to the oxidation of coenzyme B to form a CoM-CoB heterodisulfide. The heterodisulfide is reduced to regenerate the coenzymes by the enzyme heterodisulfide reductase (Hdr), which uses electrons from various sources depending on the growth substrate. During growth on CO_2 the donor is typically hydrogen, during growth on acetate the carboxyl group of acetate is oxidized to CO_2 to provide reducing power, and during growth on methylated compounds, one molar equivalent of the compound is oxidized to CO_2 by reversing the CO_2

reduction pathway. Pumping of ions for generation of a gradient occurs at different steps depending on the organism; occurring at Mtr in most organisms, and Mtr and Hdr in some members of the Order Methanosarcinales, but the details are not well understood in all organisms (Thauer *et al.*, 2008). Methanogenesis constitutes the only mechanism of energy conservation by these organisms, and many of the reactions in the pathway occur near their thermodynamic limit, meaning that methanogens are extremely sensitive to environmental factors which disrupt these reactions (Deppenmeier & Müller, 2008).

The importance of Fe-S clusters to methanogens

Some of the essential enzymes in the methanogenesis pathway; including Cdh/Acs, Hdr, Mtr, formylmethanofuran dehydrogenase, and ferredoxin (**Fig. 2**) contain iron-sulfur (Fe-S) prosthetic groups (Major *et al.*, 2004, Thauer *et al.*, 2008). Fe-S clusters are comprised of iron and sulfur atoms bound to the functional group of an amino acid; typically cysteine, but histidine and aspartic acid can also act as ligands. They can occur in multiple different configurations: including [2Fe-2S], [3Fe-4S], [8Fe-8S], or the most common arrangement [4Fe-4S] (Johnson *et al.*, 2005). Fe-S clusters are found in a wide range of proteins where they play a variety of roles. In the enzyme MutY, a DNA mismatch repair enzyme, a [4Fe-4S] cluster is essential for the protein to assume a functional conformation (Porello *et al.*, 1998). Iron-sulfur clusters can also act as reaction centers, such as in biotin synthase where the [2Fe-2S] cluster donates the sulfur during biotin synthesis (Ugulava *et al.*, 2001). The [2Fe-2S] cluster of SoxR acts as switch which allows for DNA binding by the protein, only when the cluster is present (Kiley & Beinert, 2003). Fe-S clusters can play a role in redox reactions because the iron atoms of the cluster can exist in multiple oxidation states, allowing them to transfer electrons from one compound to another (Johnson *et al.*, 2005). Fe-S proteins are thought to have evolved early on in the diversification

of life, and are found in all three domains of life. However, methanogens have a higher prevalence of Fe-S proteins compared to other organisms, including many which play important roles in key metabolic processes (Major *et al.*, 2004). Because of the large number of Fe-S proteins, and the fact that many play roles in essential metabolic processes, methanogens are extremely sensitive to conditions which would disrupt existing clusters or prevent cluster assembly, such as exposure to oxygen or reactive oxygen species (ROS).

Protection of Fe-S clusters from ROS in methanogens

Molecular oxygen (O_2) is capable of oxidizing the iron of an Fe-S cluster, causing the iron to transition from Fe^{2+} to Fe^{3+} and dissociate from the cluster (Imlay, 2006). Furthermore, exposure to O_2 can prevent the assembly of new Fe-S clusters by reacting with free Fe atoms, resulting in the formation of insoluble iron oxides. In addition to the destruction of Fe-S clusters, O_2 is dangerous to cells due to the generation of ROS. ROS are generated through the uncontrolled oxidation of cellular components by O_2 (**Fig. 3**) (Imlay, 2003). The resulting ROS are much more reactive than O_2 and can damage important components of the cell such as DNA and protein, which can lead to cell death (**Fig. 3**) (Imlay, 2003). The Fe atoms of Fe-S clusters can act as a source of ROS because they are readily oxidized by hydrogen peroxide (H_2O_2) to form even more reactive hydroxide (OH^-) and hydroxyl radical (•OH) through the Fenton reaction (**Fig. 2**) (Imlay, 2003). Organisms must be proactive in dealing with ROS to prevent widespread damage or even death.

Organisms in all three domains have evolved mechanisms to sense and respond to O_2 and ROS because of the dangers they pose. For example, some cells control expression of ROS detoxification enzymes and repair pathways by ensuring that relevant genes are only expressed when the danger from ROS is high. In some of these cases, Fe-S clusters have been adapted to

serve as O_2 sensors in transcription factors, which then regulate expression of genes in accordance with levels of intracellular O_2. This process has been extensively studied in the transcriptional regulators SoxR and Fnr, both of which use Fe-S clusters (two [2Fe-2S] clusters in SoxR, one [4Fe-4S] in Fnr) (Crack *et al.*, 2004, Imlay, 2013). The oxidation of the [2Fe-2S] clusters in SoxR results in the upregulation of an operon encoding genes for dealing with superoxide (O_2^-) (Imlay, 2013). Fnr regulates the expression of genes necessary for nitrate reduction and anaerobic respiration, processes that can only occur in anoxic environments. FNR is able to turn on gene expression when the [4Fe-4S] cluster is present. However, in the presence of O_2, the [4Fe-4S] cluster is oxidized to a [2Fe-2S] cluster, which induces a conformational change and prevents DNA binding (Crack *et al.*, 2004). Since methanogens possess a large numbers of Fe-S proteins, linking the transcription of genes involved in oxidative stress response to stability of Fe-S clusters would be a useful mechanism. Additionally, methanogens must have mechanisms to respond to ROS in order to mitigate potential damage and disruption of methanogenesis and other key pathways. Despite the importance of ROS management, little research has been conducted into the oxidative stress pathways of methanogens.

Once organisms detect the presence of O_2 and ROS, detoxification enzymes are deployed to remove the damaging molecules. Detoxification enzymes take ROS and convert them to less active molecules, such as water. Suites of detoxification enzymes vary between anaerobic and aerobic organisms, due to the terminal products which are generated. Enzymes such as catalase ($2H_2O_2 \rightarrow 2H_2O + O_2$) and superoxide dismutase ($2O_2^- + 2H+ \rightarrow H_2O_2 + O_2$) both produce O_2 as a terminal product (Jenney *et al.*, 1999). This is not a large issue in aerobic organisms, since the O_2 can be consumed during respiration. However, in anaerobic organisms, the action of these enzymes would not clear the cell of ROS, merely trade one species for another. Canonical ROS

detoxification enzymes from anaerobic organisms include NADPH–rubredoxin oxidoreductase, rubrerythrin, and superoxide reductase, which are capable of reducing ROS without generating oxygen (Jenney *et al.*, 1999). However, all three require exogenous addition of electrons for the reactions to proceed (Thorgersen *et al.*, 2012). Enzymes which catalyze the detoxification of ROS are common across all three domains of life and have been characterized from some methanogen species.

It has been documented that methanogens can survive transient exposure to O_2 and subsequently resume normal growth once anaerobic conditions are reestablished, and even producing methane in the presence of low concentrations of O_2 (Angel *et al.*, 2011, Tholen *et al.*, 2007). The ability of methanogens to survive exposure to aerobic environments suggests methanogens must possess mechanisms to mitigate O_2 and ROS exposure. A number of mechanisms from various methanogen species have been discovered; afunctional catalase has been characterized from *Methanosarcina barkeri*, catalase and $F_{420}H_2$ oxidase have been characterized from *Methanobrevibacter arboriphilus*, and active transcription of a catalase gene has been detected in desert soils dominated by methanogen species from the *Methanosarcina* and *Methanocella* genera (Shima *et al.*, Shima *et al.*, 2001, Seedorf *et al.*, 2004, Angel *et al.*, 2011). Clearly there exist mechanisms for detoxifying ROS in methanogens, however, there has not yet been sufficient research to determine if particular strategies are favored by all methanogens or if differences exist between certain methanogen lineages or if pathways are distributed based on differences in habitat.

The overarching goal of this dissertation is to investigate two aspects of the methanogen response to oxidative stress. As described previously, numerous transcription factors sense changes in the levels of O_2 within a cell through the oxidation and reduction of a bound Fe-S

cluster. Recent evidence has shown the presence of Fe-S clusters bound to the RNA polymerase complex (RNAP) of certain archaea and eukaryotes, including many methanogens. The high prevalence of Fe-S clusters bound to RNAP among anaerobic archaea species suggests the clusters may be acting as an oxidative stress sensor. The first and second chapters present a project which investigated the function of the clusters in the assembly and activity of RNAP, and whether the clusters they may be acting as a sensor in the core transcription machinery to regulate transcription in response to certain environmental factors (e.g. O_2 or Fe-availability). The third chapter presents a project examining the importance of catalase, an ROS detoxification enzyme not common in strict anaerobes, to the tolerance of methanogens to O_2 and H_2O_2. Both projects utilized the methanogen *Methanosarcina acetivorans* as a model system for *in vivo* experiments.

The methanogen *M. acetivorans,* a member of the Order Methanosarcinales, was selected as a model organism because it is relatively easy to grow and manipulate for *in vivo* experimentation (Sowers *et al.*, 1984). Members of the Methanosarcinales are typically capable of growing with methanol or acetate, as well as with CO_2/H_2, unlike other methanogens which are restricted to using CO_2/H_2. *M. acetivorans* is a well-studied methanogen; the genome has been completely sequenced, and a large amount of physiological information has been collected from experiments. The combination of genetic and physiological information can help guide studies of oxidative stress responses in this organism (Benedict *et al.*, 2012, Galagan *et al.*, 2002). A robust genetic system exists for *M. acetivorans,* allowing for genetic manipulation of the organism, including site-specific insertion of DNA into the chromosome (Metcalf *et al.*, 1997, Guss *et al.*, 2008). This allows for results from experiments utilizing *in vitro* systems to be

confirmed with experiments using the *in vivo* system, leading to a clearer understanding of the oxidative stress response system in *M. acetivorans.*

Studies have shown that members of the genus *Methanosarcina* are resilient in the face of oxidative stress. Members of this genus are capable of surviving prolonged exposure to aerobic conditions both naturally and in laboratory settings, even producing methane in desert soil during aerobic conditions (Angel *et al.*, 2011). The sequenced genome of *M. acetivorans* revealed the presence of a number of genes shown to be important in the oxidative stress response of other methanogens. These include a number of putative genes coding for ROS detoxification enzymes, such as $F_{420}H_2$ oxidase, rubrerythrin, superoxide reductase, catalase, superoxide dismutase, and peroxidase. The presence of catalase and superoxide dismutase is surprising, since both produce O_2 as a terminal product and are typically found in aerobic organisms. However, there is evidence for active catalase and superoxide dismutase in other anaerobic organisms (Brioukhanov & Netrusov, 2004, Hewitt & Morris, 1975). Thus, the importance of these enzymes in the oxidative stress response of methanogens is unclear. Additionally, the genome contains a gene encoding MsvR, a recently characterized transcriptional regulator from *Methanothermobacter thermautotrophicus (Karr, 2010).* In *M. thermautotrophicus,* MsvR modulates expression of an operon composed of putative ROS detoxification enzymes through the redox state of cysteine residues, another typical target of ROS (Karr, 2010). The genome sequence shows *M. acetivorans* possesses an extensive suite of genes encoding ROS detoxification enzymes, including enzymes typically found in both anaerobes and aerobes, and at least one regulatory element known for modulating expression of an operon important for responding to oxidative stress. This suggests the presence of a complex system for responding to

aerobic conditions, making it an ideal model system for oxidative stress responses in methanogens.

The presence of a [4Fe-4S] binding ferredoxin-like domain in *M. acetivorans* RNAP

RNAP is the multi-subunit enzyme complex responsible for synthesizing RNA from a DNA template. The enzyme is found in all cellular life from each of the three domains; *Archaea*, *Bacteria*, and *Eukarya*. The RNAP from each of the three domains is highly conserved likely due to the essential role it plays; however, some differences exist between RNAP from each domain. In bacteria, a single RNAP is responsible for synthesizing all RNA. Structurally it is the simplest polymerase, being comprised of only four subunits, $\alpha_2\beta\beta'\omega$ (Decker & Hinton, 2013). Eukaryotes possess at least three separate RNAPs, which transcribe different subsets of RNA and are composed of twelve subunits, including subunits not present in the bacterial RNAP (Decker & Hinton, 2013). Archaea utilize a single RNAP like bacteria, however, structurally the RNAP is most similar to eukaryotic RNAPII (**Fig. 4**) (Jun *et al.*, 2011). Although the RNAPs from each domain differ in subunit composition, the catalytic region of RNAPs share a high degree of similarity, indicating the process of RNA synthesis is highly conserved among all three domains.

Differences also exist in the initiation of transcription between RNAP from the three domains. Bacteria possess multiple initiation subunits, called sigma factors, which bind to the RNAP to form the RNAP holoenzyme (Helmann & Chamberlin, 1988) (**Fig. 5**). Each individual sigma factor recognizes different promoter sequences throughout the bacterial genome, and thus allows for binding of RNAP to and subsequent transcription of certain genes. Sigma factors are modular and the organism can change which subset of genes the RNAP is binding and transcribing by switching out specific sigma factors. RNA synthesis in eukaryotes is more complex and requires five general transcription factors to bind to DNA and initiate transcription

(Krishnamurthy & Hampsey, 2009). Archaeal transcription initiation is a simpler version of the initiation system present in Eukaryotes, requiring only TBP (TFIID homolog) and TFB (TFIIB homolog) (**Fig. 5**) (Geiduschek & Ouhammouch, 2005, Bell & Jackson, 1998). Archaea and eukaryotes are thought to share a more recent common ancestor compared to bacteria, which is likely the reason the transcription machinery and initiation process from both domains are more similar to each other compared to bacteria (Bell & Jackson, 1998).

The presence of an iron-sulfur binding domain (domain three aka D3) in the D subunit of RNAP was identified from the sequenced archaeon *Sulfolobus acidocaldarius* by a team of investigators in 1998 (Rodriguez-Monge *et al.*, 1998). The authors examined other genomes, publicly available at the time, and discovered D3 occurred in the subunit D from other archaea and in the homologous subunits (Rpb3/AC40) of eukaryotes as well, but not in the homologous bacterial subunit (α). No additional findings concerning D3 were published until the first X-ray crystal structure of an archaeal RNAP, from *Sulfolobus solfataricus*, was solved in 2008 (Hirata *et al.*, 2008). The crystal structure revealed a [3Fe-4S] cluster bound to D3 of subunit D in fully assembled RNAP. The presence of four cysteine residues in D3 in a canonical [4Fe-4S] cluster binding motif suggested to the authors that the cluster exists as a [4Fe-4S] cluster *in vivo* and that the observed [3Fe-4S] cluster was an artifact of aerobic preparation (Hirata *et al.*, 2008). This was the first evidence of RNAP as a Fe-S protein, and a survey of sequenced archaeal and eukaryotic genomes revealed that a canonical [4Fe-4S] cluster binding motif occurred in D3 of subunit D/Rpb3/AC40 in numerous organisms, but was not universal in either domain. The authors determined the [4Fe-4S] cluster is located 45 Å from the RNAP reaction center, which suggests the cluster is not involved in the synthesis of RNA, since the clusters are too far from the reaction center to allow for an interaction between the clusters and the DNA template or

RNA product. Domain three has only been found in archaea and eukaryotes, suggesting the feature was acquired early during the evolution of archaea and eukaryotes. This suggests domain three and associated [4Fe-4S] cluster may have evolved to perform a unique role in the assembly and/or activity of complete RNAP in both archaea and eukaryotes, which are more similar in the subunit composition and transcription initiation process compared to the RNAP found in bacteria.

Subunit D/Rpb3/AC40 plays an essential role during assembly of RNAP. It forms a heterodimer with subunit L (Rpb11/AC19 in eukaryotes), which is the first step in the assembly of RNAP; all other subunits assemble onto the heterodimer (**Fig. 6**) (Goede *et al.*, 2006, Eloranta *et al.,* 1998). In archaea, formation of the D/L heterodimer is followed by assembly of RNAP subunits BNP to form a BDLNP subcomplex (Goede *et al.,* 2006). In members of the Euryarchaeota, including *M. acetivorans,* the B subunit is comprised of two separate subunits, B' and B'' (Goede *et al.,* 2006). Following formation of the subcomplex, subunits A'A''EFHK assemble to form the complete RNAP complex. Regions homologous to the D/L heterodimer are present in the bacterial subunit α and play a similar role during assembly of bacterial RNAP (Grohmann *et al.*, 2009). Recombinant experiments using the *S. solfataricus* subunit D revealed that without the [4Fe-4S] cluster, subunit D does not readily assemble into a heterodimer with subunit L, and instead aggregates (Hirata *et al.,* 2008). Interference with the assembly of subunit D with subunit L would impact downstream assembly of other RNAP subunits as well. This observation suggests a potential mechanism by which a cell could regulate assembly of RNAP, and therefore global transcription, through the state of the [4Fe-4S] cluster (**Fig. 6**).

The RNAP from *S. solfataricus* was crystalized in aerobic conditions, indicating the single [4Fe-4S] cluster is not sensitive to oxygen (Hirata *et al.*, 2008). However, *S. solfataricus* is

an aerobic organism, as are many of the other sequenced archaea and most sequenced eukaryotes which possess D3 with putative cluster binding motifs. Domain three is present in subunit D belonging to a number of methanogens and other anaerobes, where is possess a high degree of similarity to ferredoxin (also known as a ferredoxin-like domain or FLD) (Rodriguez-Monge *et al.*, 1998). Ferredoxins are small proteins containing Fe-S clusters, which function as electron carriers in a wide variety of reactions. Ferredoxins are especially abundant in anaerobic organisms, including acting as an electron carrier during methanogenesis (**Fig. 2**) (D C Yoch & Valentine, 1972). This is significant because ferredoxins play more important roles in the central metabolism of anaerobes compared to aerobes, likely due to the sensitivity of the Fe-S clusters to O_2, and because ferredoxins possess much lower reducing potentials compared to other electron carriers (Buckel & Thauer, 2013). The similarity of the FLD to ferredoxin suggests a mechanism, whereby an anaerobic cell could link transcription with the metabolic state of the cell through the oxidation of the [4Fe-4S] cluster of subunit D. Since subunit D is part of the first step of RNAP assembly, loss of the cluster due to oxidative stress would prevent assembly of subunit D with subunit L and therefore prevent downstream assembly with other RNAP subunits (**Fig 6**). If the sensitivity of the [4Fe-4S] cluster in subunit D is similar to the sensitivity of the clusters in ferredoxin, RNAP assembly would turn off when conditions prevent ferredoxin from functioning as an electron carrier. Synthesis of RNA and protein synthesis require ATP, so it would be beneficial for these processes to be down-regulated in conditions where the cell is unable to synthesize ATP. The [4Fe-4S] cluster in *S. solfataricus* is not sensitive to O_2, but the cluster may be sensitive to more reactive ROS, which are a danger to both aerobes and anaerobes. Alternatively, the cluster may be acting as a sensor of intracellular Fe or S levels; as an aerobe *S. solfataricus* would have a harder time finding soluble iron than an anaerobe would,

since the most oxidized form of iron is insoluble in water. In either case, maintenance of the Fe-S cluster in the FLD may be a previously unrecognized regulatory mechanism that links cluster stability with transcription in response to intracellular conditions. In methanogens, coupling an oxygen sensitive cluster to RNAP assembly or activity could be a potentially useful adaptation, as it would allow cells to link transcription with the integrity of the central metabolic pathway, preventing synthesis of RNA if energy cannot be conserved within the cell (**Fig. 2**).

The sequence of *rpoD,* the gene which encodes subunit D, from *M. acetivorans* revealed the presence of a FLD within the subunit (Hirata *et al.*, 2008). A novel discovery was that the FLD possesses two complete [4Fe-4S] cluster binding motifs; one homologous to the motif initially identified in *S. solfataricus* and a second motif located closer to the N-terminus. The sequence of the FLD shares a high degree of homology with the ferredoxin found in *M. acetivorans,* including the position of the two [4Fe-4S] cluster binding motifs (**Fig. 7**). The cluster homologous to the single cluster found in *S. solfataricus,* which is also the more highly conserved cluster among both archaea and eukaryotes was designated cluster one, while the additional motif was designated cluster two. The presence of two [4Fe-4S] clusters raised some interesting questions about the nature of the FLD and its distribution among archaea. Additional searching of sequenced archaeal genomes revealed organisms possessing a FLD with two [4Fe-4S] cluster binding motifs were exclusively obligate anaerobic archaea, including many methanogens. This observation provided additional support for the hypothesis that the clusters may be playing an important regulatory role in the RNAP of anaerobic archaea, one that links transcription with oxidative stress. The similarity of the *M. acetivorans* FLD with its own ferredoxin suggests the clusters share similar sensitivity to O_2; meaning clusters from the FLD and ferredoxin would be lost in similar conditions (**Fig. 7**). The genetic system of *M. acetivorans*

offered a unique opportunity to probe the role of the FLD and each individual [4Fe-4S] cluster in RNAP assembly and activity. Modified versions of subunit D, which lack cluster one, cluster two, or the entire FLD could be inserted into the chromosome, and the effects that loss of each cluster have on RNAP assembly and activity, as well as other physiological effects, could be observed. The ability to probe each cluster individually allows us to determine if the clusters play non-overlapping roles in assembly of RNAP, and could help explain why certain species possess one cluster while certain anaerobes possess two.

The role of catalase enzymes in the oxidative stress response of *M. acetivorans*

Although the genome sequence of *M. acetivorans* revealed the presence of a large number of putative genes encoding ROS detoxification enzymes, there is little experimental data regarding expression of the genes or activity of the gene products. The presence of superoxide dismutase (MA1574) and rubrerythrin (MA0639) was detected in proteomic studies of *M. acetivorans* grown on acetate, but the enzymatic activity of either enzyme was not determined (Li *et al.*, 2005). There is, however, evidence of a functional catalase enzyme in *M. barkeri*, an organism closely related to *M. acetivorans* (Shima *et al.*). Catalase was originally thought to only be a component of the oxidative stress response of aerobes, since one of the products is molecular oxygen (Imlay, 2002). However, evidence now exists which shows the presence of catalase in anaerobes as well (Brioukhanov & Netrusov, 2004). Moreover, catalase expression was detected from desert soil dominated by *Methanosarcina* species (Angel *et al.*, 2011). The genome of *M. acetivorans* revealed the presence of two putative catalase genes, one a mono-functional catalase and one a bifunctional catalase, based on sequence homology to catalase proteins from *Escherichia coli*. A bifunctional catalase possesses both catalase and peroxidase ($H_2O_2 + e^-\ donor_{red} \rightarrow 2H_2O + e^-\ donor_{ox}$) activity. Additionally, preliminary data from the

Lessner lab showed addition of exogenous catalase to *M. acetivorans* cultures increased survival during exposure to H_2O_2 (Horne & Lessner, 2013). These data suggest that catalase may be an important component of the oxidative stress response of *M. acetivorans* and other methanogens as well. A recombinant approach was selected to measure activity of each catalase through expression and purification from *E. coli*, to determine if either gene product possess enzymatic activity. Additionally, the catalase activity of *M. acetivorans* cell lysate would also be measured in both aerobic and anaerobic conditions to determine if the enzymes are active *in vivo* and if activity is induced upon exposure to oxidative stress.

Experimental sections of this dissertation

The experiments detailed in this dissertation examined two aspects of the oxidative stress response of the model methanogen *Methanosarcina acetivorans*. The data are divided into three chapters, each coinciding with a paper submitted for publication in a scientific journal, or already published.

Chapter I – This chapter details the initial study of *M. acetivorans* subunit D using mainly recombinant methods. The results include a more detailed survey of sequenced archaeal genomes and classifying D3 into six distinct types based on the number of complete [4Fe-4S] binding motifs. Experimental evidence proves that subunit D from *M. acetivorans* binds two [4Fe-4S] clusters as predicted from the protein sequence, and that those clusters are redox active and oxygen sensitive. Generation of a variant of subunit D which lacked the FLD (DΔD3) is also detailed. Finally the DΔD3 variant is used to show the FLD is not required for the assembly of subunit D with subunit L, using both *in vitro* and *in vivo* experiments. The experiments in this chapter have been published and are available online. The data generated from recombinant studies presented in this chapter were generated primarily by Dan Lessner and Faith Lessner.

However, since later experiments included in this dissertation built on the results of these initial experiments, it made sense to include them in this dissertation.

Chapter II - The second chapter details the generation of additional subunit D variants which lack each individual [4Fe-4S] cluster. Variants of subunit D (including the DΔD3 detailed in the previous chapter) were purified from *M. acetivorans* strains, and co-purification of other RNAP subunits was detected using Western blots. Deletion of the FLD and individual clusters results in a negative impact on assembly of subunit D with two catalytic subunits, B' and A''. Inability to bind cluster one has a more severe impact versus inability to bind cluster two or loss of the FLD. Strains of *M. acetivorans*, where native subunit D was replaced with a subunit D variant, were generated and a negative effect on lag phase and generation time was observed when the FLD was deleted or when clusters were unable to bind. Again inability to bind cluster one had the most severe impact. Transcription activity of RNAP incorporating the subunit D variants was measured, and again a negative impact on activity was observed when the FLD or a [4Fe-4S] cluster was missing with inability to bind cluster one having the most severe impact. This chapter has been submitted for publication.

Chapter III - The third chapter details experiments which grew out of a project conducted by undergraduate student Cody Schaff which involved the characterization of two catalase genes from *M. acetivorans.* One gene contained an in frame stop codon mutation which truncated the encoded protein and was not characterized further. The remaining catalase enzyme did not possess significant activity when expressed recombinantly in *E. coli*, nor was activity detected in *M. acetivorans* cell lysate. A *M. acetivorans* strain was generated which expressed recombinant *E. coli* catalase, and expression of this catalase affords the cell protection from subsequent challenge with H_2O_2 but not O_2 exposure. This chapter was published and is available online.

The dissertation concludes with a short conclusions section which summarizes the results of the three data chapters. The conclusions illustrate the advances in our understanding of the oxidative stress responses of *M. acetivorans* made possible by the experiments presented in this dissertation. This section also suggests possible future directions for research into the oxidative stress response of *M. acetivorans,* based on the data presented in this dissertation.

References

Angel, R., D. Matthies & R. Conrad, (2011) Activation of Methanogenesis in Arid Biological Soil Crusts Despite the Presence of Oxygen. *PLoS One* 6: e20453.

Bell, S.D. & S.P. Jackson, (1998) Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends in microbiology* 6: 222-228.

Benedict, M.N., M.C. Gonnerman, W.W. Metcalf & N.D. Price, (2012) Genome-Scale Metabolic Reconstruction and Hypothesis Testing in the Methanogenic Archaeon Methanosarcina acetivorans C2A. *J Bacteriol* 194: 855-865.

Brioukhanov, A.L. & A.I. Netrusov, (2004) Catalase and Superoxide Dismutase: Distribution, Properties, and Physiological Role in Cells of Strict Anaerobes. *Biochemistry (Moscow)* 69: 949-962.

Buckel, W. & R.K. Thauer, (2013) Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1827: 94-113.

Crack, J., J. Green & A.J. Thomson, (2004) Mechanism of Oxygen Sensing by the Bacterial Transcription Factor Fumarate-Nitrate Reduction (FNR). *Journal of Biological Chemistry* 279: 9278-9286.

D C Yoch, a. & R.C. Valentine, (1972) Ferredoxins and Flavodoxins of Bacteria. *Annual Review of Microbiology* 26: 139-162.

Decker, K.B. & D.M. Hinton, (2013) Transcription Regulation at the Core: Similarities Among Bacterial, Archaeal, and Eukaryotic RNA Polymerases. *Annual Review of Microbiology* 67: 113-139.

Deppenmeier, U. & V. Müller, (2008) Life Close to the Thermodynamic Limit: How Methanogenic Archaea Conserve Energy. In: Bioenergetics. G. Schäfer & H. Penefsky (eds). Springer Berlin Heidelberg, pp. 123-152.

Eloranta, J.J., A. Kato, M.S. Teng & R.O.J. Weinzierl, (1998) In vitro assembly of an archaeal D-L-N RNA polymerase subunit complex reveals a eukaryote-like structural arrangement. *Nucleic Acids Res* 26: 5562-5567.

Fox, G.E., L.J. Magrum, W.E. Balch, R.S. Wolfe & C.R. Woese, (1977) Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *P Natl Acad Sci USA* 74: 4537-4541.

Gaci, N., G. Borrel, W. Tottey, P.W. O'Toole & J.-F. Brugère, (2014) Archaea and the human gut: New beginning of an old story. *World Journal of Gastroenterology : WJG* 20: 16062-16078.

Galagan, J.E., C. Nusbaum, A. Roy, M.G. Endrizzi, P. Macdonald, W. FitzHugh, S. Calvo, R. Engels, S. Smirnov, D. Atnoor, A. Brown, N. Allen, J. Naylor, N. Stange-Thomann, K. DeArellano, R. Johnson, L. Linton, P. McEwan, K. McKernan, J. Talamas, A. Tirrell, W. Ye, A. Zimmer, R.D. Barber, I. Cann, D.E. Graham, D.A. Grahame, A.M. Guss, R. Hedderich, C. Ingram-Smith, H.C. Kuettner, J.A. Krzycki, J.A. Leigh, W. Li, J. Liu, B. Mukhopadhyay, J.N. Reeve, K. Smith, T.A. Springer, L.A. Umayam, O. White, R.H. White, E.C. de Macario, J.G. Ferry, K.F. Jarrell, H. Jing, A.J.L. Macario, I. Paulsen, M. Pritchett, K.R. Sowers, R.V. Swanson, S.H. Zinder, E. Lander, W.W. Metcalf & B. Birren, (2002) The Genome of M. acetivorans Reveals Extensive Metabolic and Physiological Diversity. *Genome Research* 12: 532-542.

Geiduschek, E.P. & M. Ouhammouch, (2005) Archaeal transcription and its regulators. *Mol Microbiol* 56: 1397-1407.

Goede, B., S. Naji, O. von Kampen, K. Ilg & M. Thomm, (2006) Protein-Protein Interactions in the Archaeal Transcriptional Machinery: BINDING STUDIES OF ISOLATED RNA POLYMERASE SUBUNITS AND TRANSCRIPTION FACTORS. *Journal of Biological Chemistry* 281: 30581-30592.

Gribaldo, S. & C. Brochier-Armanet, (2006) The origin and evolution of Archaea: a state of the art. *Philosophical Transactions of the Royal Society B: Biological Sciences* 361: 1007-1022.

Grohmann, D., A. Hirtreiter & F. Werner, (2009) Molecular mechanisms of archaeal RNA polymerase. *Biochem Soc T* 37: 12.

Guss, A.M., M. Rother, J.K. Zhang, G. Kulkkarni & W.W. Metcalf, (2008) New methods for tightly regulated gene expression and highly efficient chromosomal integration of cloned genes for Methanosarcina species. *Archaea* 2: 193-203.

Helmann, J.D. & M.J. Chamberlin, (1988) Structure and function of bacterial sigma factors. *Annual Review of Biochemistry* 57: 839-872.

Hewitt, J. & J.G. Morris, (1975) Superoxide dismutase in some obligately anaerobic bacteria. *FEBS Letters* 50: 315-318.

Hirata, A., B.J. Klein & K.S. Murakami, (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451: 851-854.

Horne, A.J. & D.J. Lessner, (2013) Assessment of the oxidant tolerance of Methanosarcina acetivorans. *FEMS microbiology letters* 343: 13-19.

Imlay, J.A., (2002) How oxygen damages microbes: Oxygen tolerance and obligate anaerobiosis. In: Advances in Microbial Physiology. Academic Press, pp. 111-153.

Imlay, J.A., (2003) Pathways of Oxidative Damage. *Annual Review of Microbiology* 57: 395-418.

Imlay, J.A., (2006) Iron-sulphur clusters and the problem with oxygen. *Mol Microbiol* 59: 1073-1082.

Imlay, J.A., (2013) The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat Rev Micro* 11: 443-454.

Ipcc, (2014) *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]*, p. 1132. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jenney, F.E., M.F.J.M. Verhagen, X. Cui & M.W.W. Adams, (1999) Anaerobic Microbes: Oxygen Detoxification Without Superoxide Dismutase. *Science* 286: 306-309.

Johnson, D.C., D.R. Dean, A.D. Smith & M.K. Johnson, (2005) Structure, Function, and Formation of Biological Iron-Sulfur Clusters *Annual Review of Biochemistry* 74: 247-281.

Jun, S.-H., M.J. Reichlen, M. Tajiri & K.S. Murakami, (2011) Archaeal RNA polymerase and transcription regulation. *Crit Rev Biochem Mol* 46: 27-40.

Karr, E.A., (2010) The Methanogen-Specific Transcription Factor MsvR Regulates the fpaA-rlp-rub Oxidative Stress Operon Adjacent to msvR in Methanothermobacter thermautotrophicus. *J Bacteriol* 192: 5914-5922.

Kiley, P.J. & H. Beinert, (2003) The role of Fe–S proteins in sensing and regulation in bacteria. *Curr Opin Microbiol* 6: 181-185.

Krishnamurthy, S. & M. Hampsey, (2009) Eukaryotic transcription initiation. *Current Biology* 19: R153-R156.

Lessner, D.J., (2001) Methanogenesis Biochemistry. In: eLS. John Wiley & Sons, Ltd, pp.

Li, Q., L. Li, T. Rejtar, B.L. Karger & J.G. Ferry, (2005) Proteome of Methanosarcina acetivorans Part I: An Expanded View of the Biology of the Cell. *Journal of Proteome Research* 4: 112-128.

Major, T.A., H. Burd & W.B. Whitman, (2004) Abundance of 4Fe–4S motifs in the genomes of methanogens and other prokaryotes. *FEMS microbiology letters* 239: 117-123.

Metcalf, W.W., J.K. Zhang, E. Apolinario, K.R. Sowers & R.S. Wolfe, (1997) A genetic system for Archaea of the genus Methanosarcina: Liposome-mediated transformation and construction of shuttle vectors. *Proceedings of the National Academy of Sciences* 94: 2626-2631.

Porello, S.L., M.J. Cannon & S.S. David, (1998) A Substrate Recognition Role for the [4Fe-4S]2+ Cluster of the DNA Repair Glycosylase MutY. *Biochemistry* 37: 6465-6475.

Rodriguez-Monge, L., C.A. Ouzounis & N.C. Kyrpides, (1998) A ferredoxin-like domain in RNA polymerase 30/40-kDa subunits. *Trends in Biochemical Sciences* 23: 169-170.

Seedorf, H., A. Dreisbach, R. Hedderich, S. Shima & R.K. Thauer, (2004) F420H2 oxidase (FprA) from Methanobrevibacter arboriphilus, a coenzyme F420-dependent enzyme involved in O2 detoxification. *Archives of Microbiology* 182: 126-137.

Shima, S., A. Netrusov, M. Sordel, M. Wicke, C.G. Hartmann & K.R. Thauer, Purification, characterization, and primary structure of a monofunctional catalase from Methanosarcina barkeri. *Archives of Microbiology* 171: 317-323.

Shima, S., M. Sordel-Klippert, A. Brioukhanov, A. Netrusov, D. Linder & R.K. Thauer, (2001) Characterization of a Heme-Dependent Catalase fromMethanobrevibacter arboriphilus. *Applied and Environmental Microbiology* 67: 3041-3045.

Sowers, K.R., S.F. Baron & J.G. Ferry, (1984) Methanosarcina acetivorans sp. nov., an Acetotrophic Methane-Producing Bacterium Isolated from Marine Sediments. *Applied and Environmental Microbiology* 47: 971-978.

Thauer, R.K., A.-K. Kaster, H. Seedorf, W. Buckel & R. Hedderich, (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Micro* 6: 579-591.

Tholen, A., M. Pester & A. Brune, (2007) Simultaneous methanogenesis and oxygen reduction by Methanobrevibacter cuticularis at low oxygen fluxes. *FEMS Microbiology Ecology* 62: 303-312.

Thorgersen, M.P., K. Stirrett, R.A. Scott & M.W.W. Adams, (2012) Mechanism of oxygen detoxification by the surprisingly oxygen-tolerant hyperthermophilic archaeon, Pyrococcus furiosus. *Proceedings of the National Academy of Sciences* 109: 18547-18552.

Ugulava, N.B., B.R. Gibney & J.T. Jarrett, (2001) Biotin Synthase Contains Two Distinct Iron-Sulfur Cluster Binding Sites: Chemical and Spectroelectrochemical Analysis of Iron-Sulfur Cluster Interconversions(). *Biochemistry* 40: 8343-8351.

Woese, C.R., O. Kandler & M.L. Wheelis, (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* 87: 4576-4579.

Woese, C.R., L.J. Magrum & G.E. Fox, Archaebacteria. *J Mol Evol* 11: 245-252.

Figures

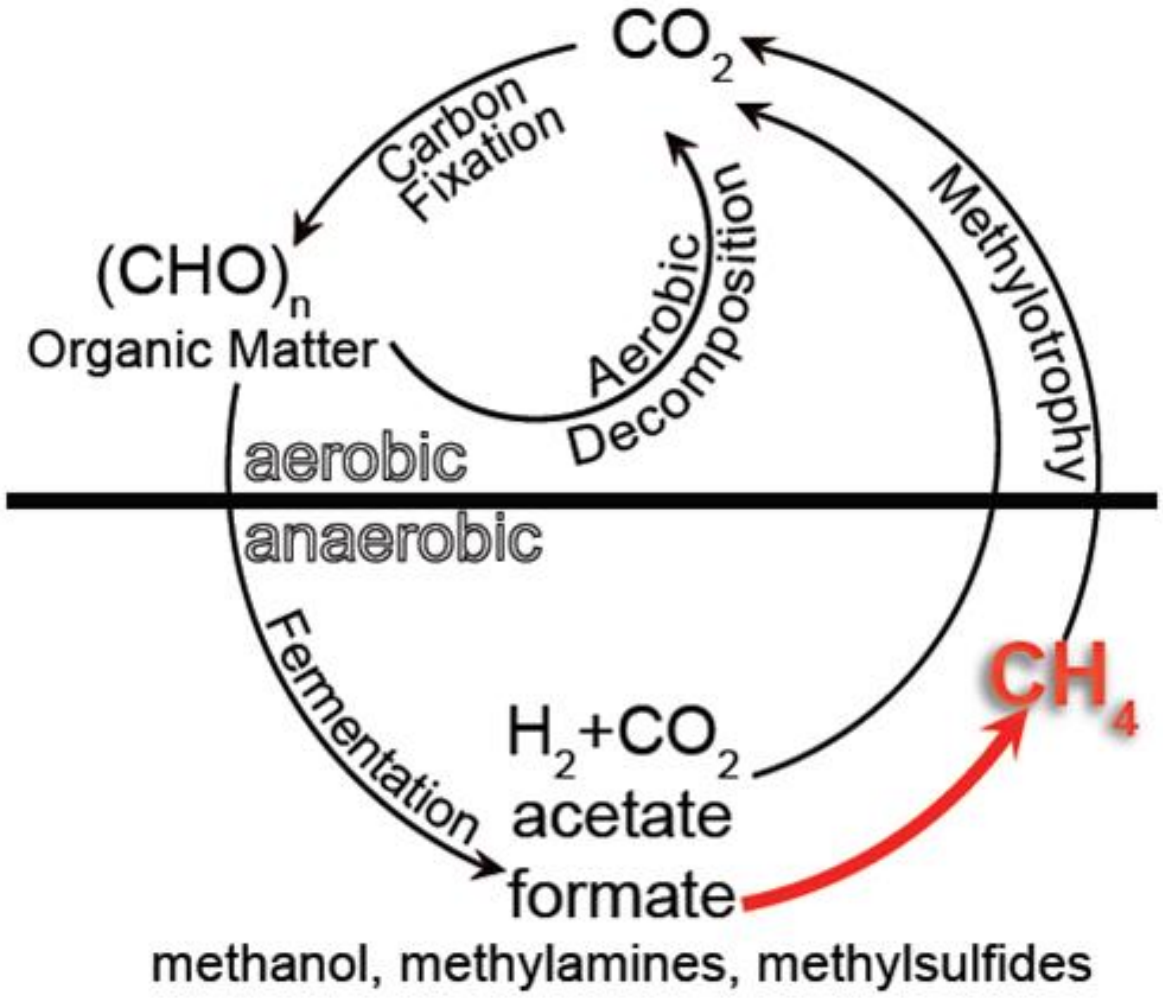

Figure 1. The importance of methanogenesis to the carbon cycle. CO_2 is fixed by certain organisms into biomass, which is then consumed by other organisms through respiration or fermentation. Methanogens utilize the terminal products of fermentation and respiration and reduce them to CH_4 (delineated by red line). The CH_4 is then emitted from the sediment, where it re-enters the carbon cycle. Figure adapted from Nicole Buan's laboratory website (http://unlcms.unl.edu/biochemistry/buanlab/research-overview).

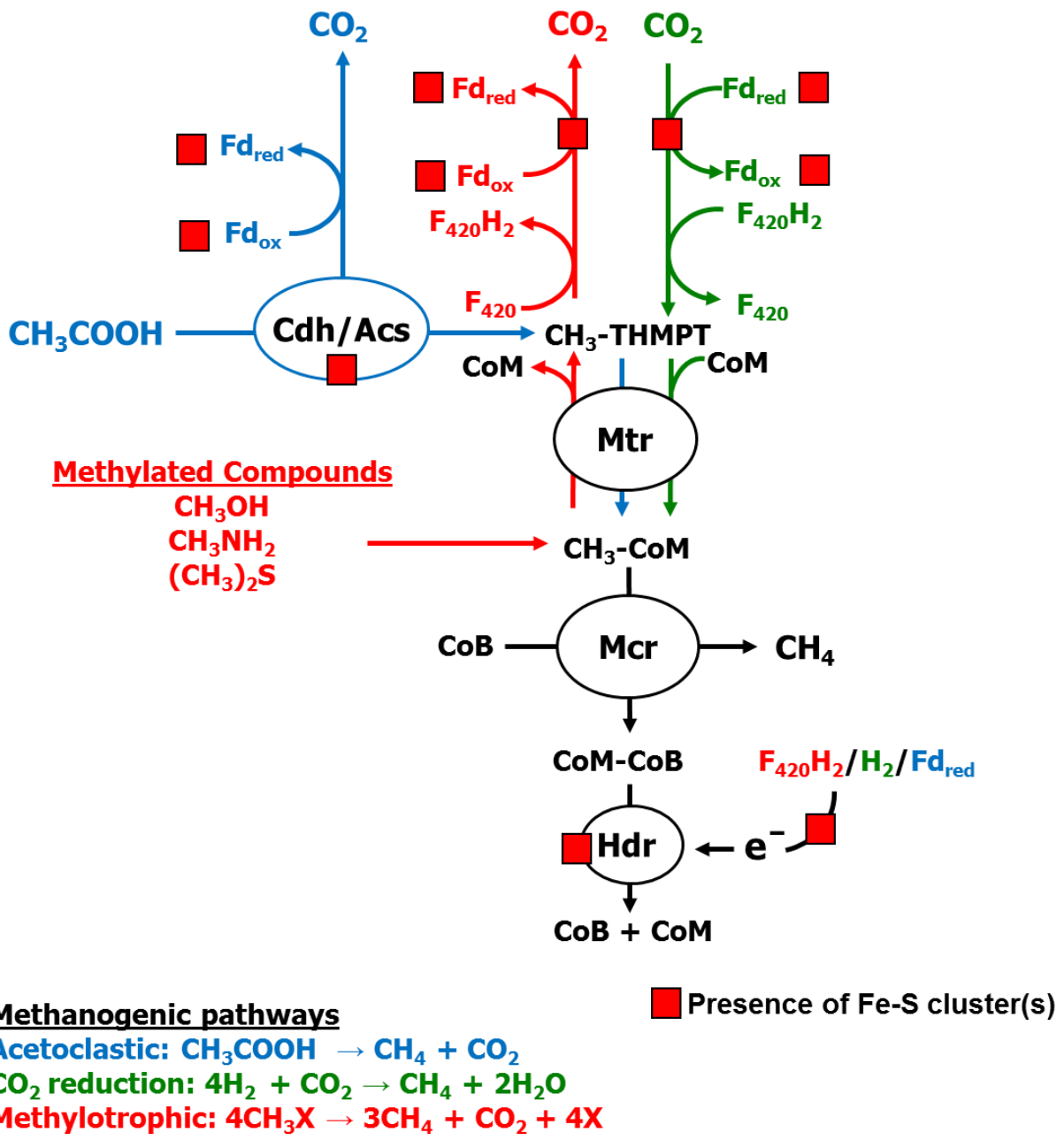

Figure 2. The three major methanogenesis pathways. The colored lines represent the three major pathways of methanogenesis which differ in the starting substrate and source of reducing power. Terminal steps are shared between pathways and denoted by black lines. The presence of iron-sulfur clusters in various enzymes and other proteins is denoted by a red box. Abbreviations: coenzyme F_{420} (F_{420}); ferredoxin (Fd); carbon monoxide dehydrogenase/acetyl-CoA synthase (Cdh/Acs); tetrahydromethanopterin (THMPT); coenzyme M (CoM); THMPT:coenzyme M methyltransferase (Mtr); coenzyme B (CoB); methyl-coenzyme M reductase (Mcr); heterodisulfide reductase (Hdr).

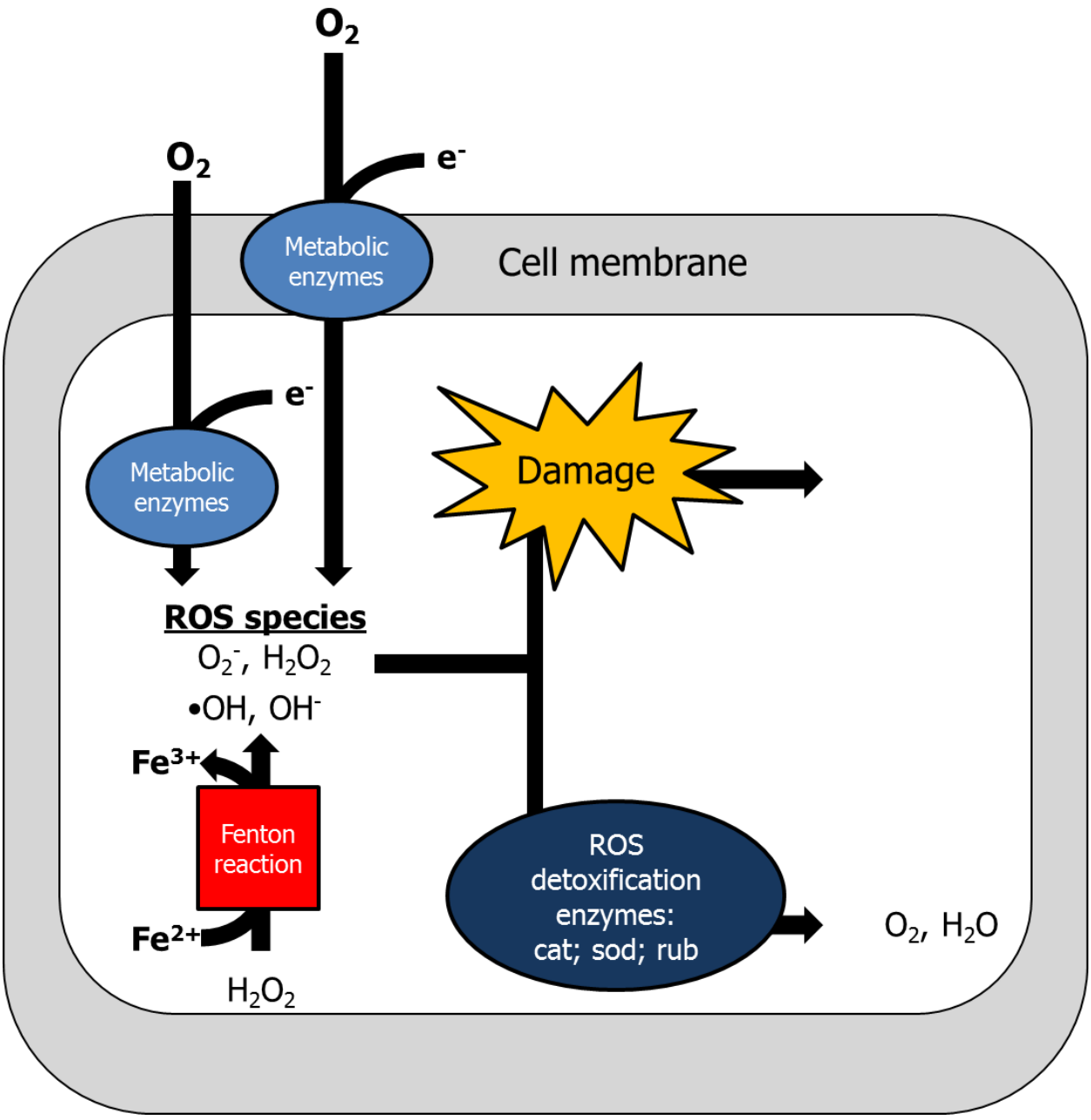

Figure 3. Generation of intracellular ROS. ROS are generated when O_2 becomes an electron acceptor for metabolic enzymes. Soluble iron exacerbates the problem by catalyzing the generation of even more ROS from H_2O_2 via the Fenton reaction. ROS can cause cell damage and ultimately death through the introduction of mutations to DNA, the oxidation of proteins and lipid membranes, and disruption of Fe-S clusters. Cells synthesize ROS detoxification enzymes to convert the ROS into water or less harmful molecular oxygen. Abbreviations: catalase (cat); super oxide dismutase (sod), rubrerythrin (rub).

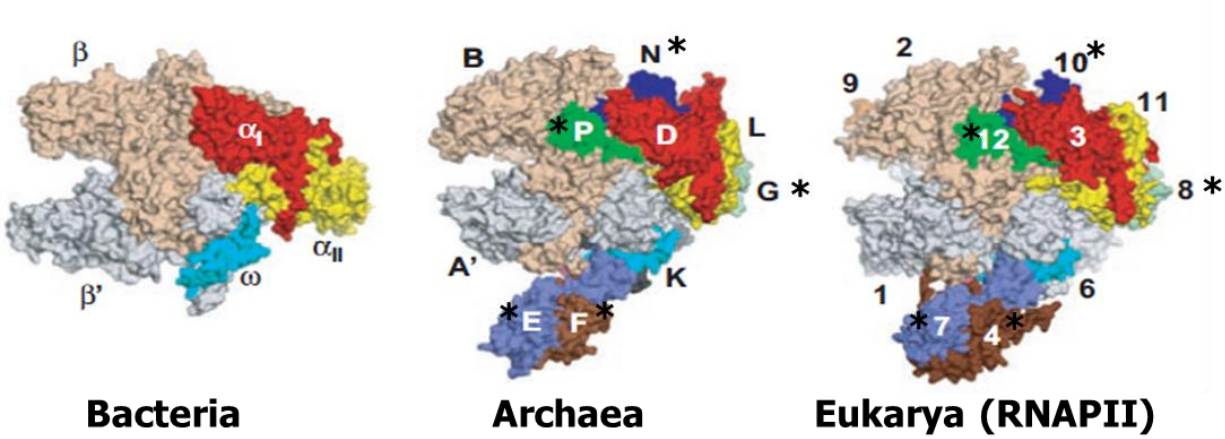

Figure 4. Comparison of RNAP complexes between the three domains of life. The crystal structure of RNAP from Bacteria (*Thermus aquaticus*), Archaea (*Sulfolobus solfataricus*), and Eukarya (*Saccharomyces cerevisiae* RNAPII) are shown. The catalytic region (tan and grey colored subunits) is highly conserved between all three domains. Subunits designated by asterisks are present in both eukaryotes and archaea but absent in bacteria. Figure adapted from reference (Jun *et al.*, 2011).

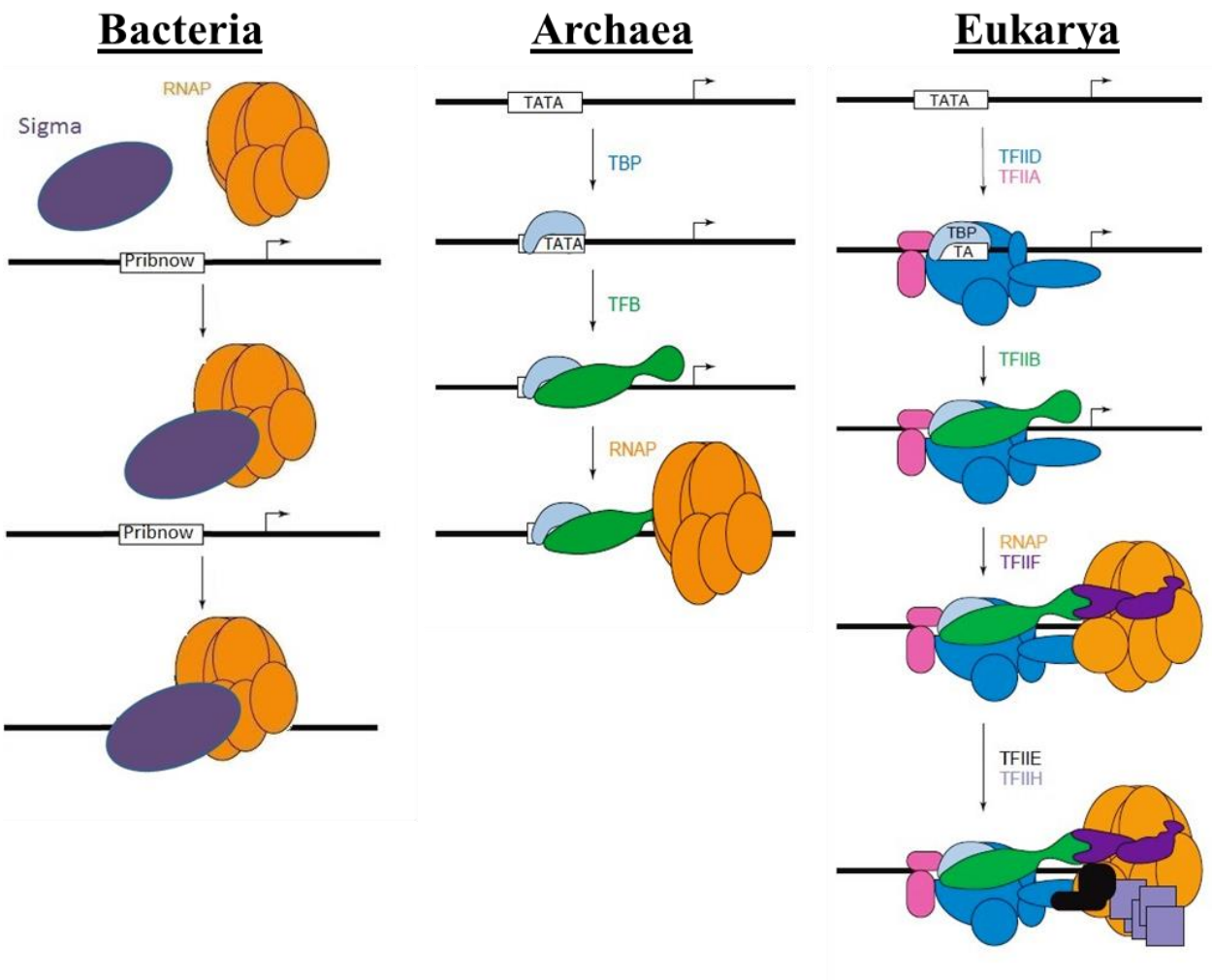

Figure 5. Comparison of transcription initiation between the three domains of life. Bacteria modulate RNAP activity using modular sigma factors which bind unique sets of genes. Eukarya require a large number of transcription factors to recruit the RNAP to initiation sites. Archaea possess a scaled down eukaryotic system requiring only TBP and TFB for recruitment of RNAP. Figure adapted from reference (Bell & Jackson, 1998).

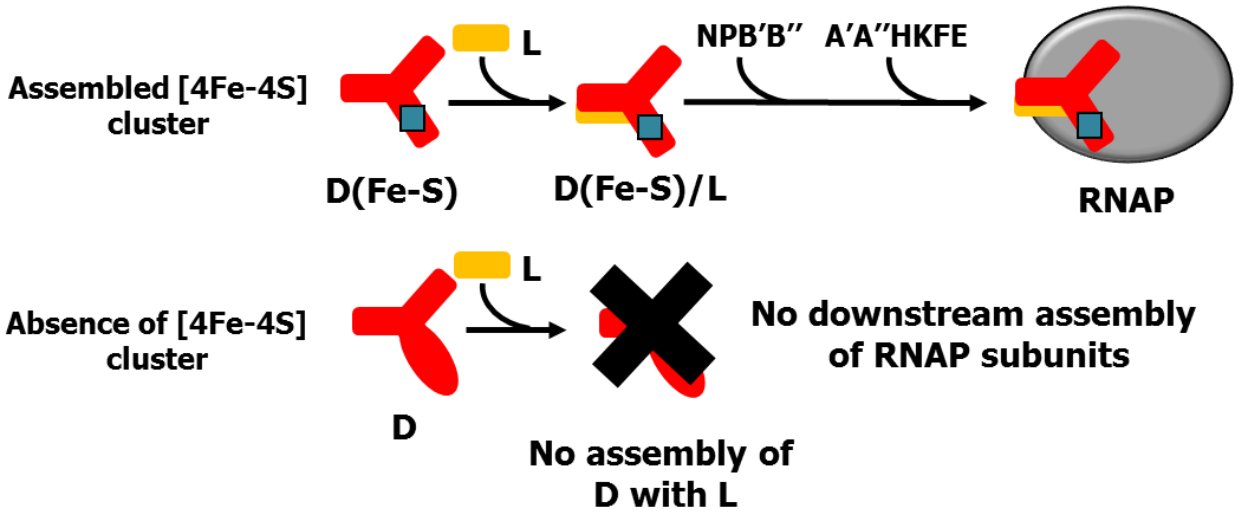

Figure 6. Model of RNAP assembly regulation. Formation of the DL heterodimer is the first step of RNAP assembly. When the [4Fe-4S] cluster is present, the DL heterodimer forms a platform, upon which the other RNAP subunits assemble, to generate complete RNAP. However, in the absence of cluster assembly, the DL heterodimer is prevented.

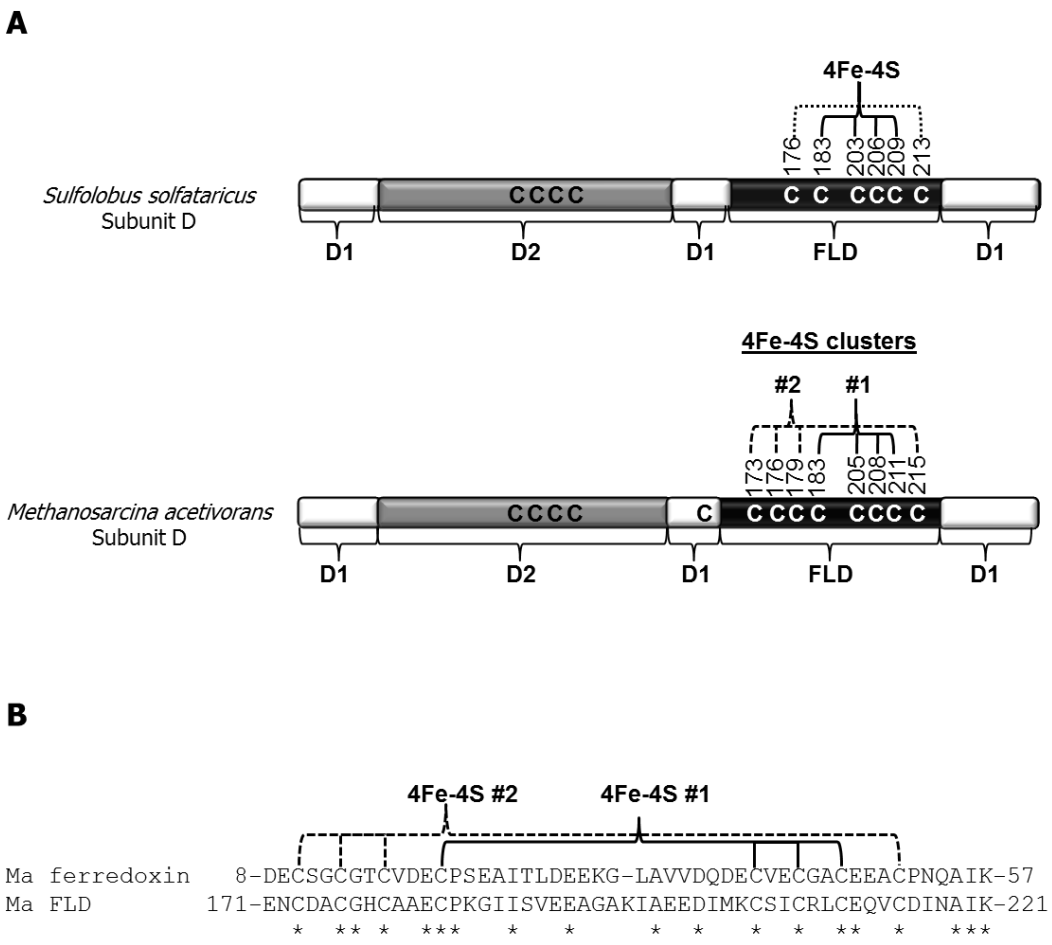

Figure 7. Comparison of the D subunits from *M. acetivorans* and *S. solfataricus.* A. Representations of subunit D from each species are shown with domains (D1 = domain 1; D2 = domain 2; FLD = ferredoxin-like domain) delineated below the illustrations. The FLD of *M. acetivorans* contains two [4Fe-4S] cluster binding motifs labeled #1 (homologous to the motif present in the *S. solfataricus* subunit D) and #2. Numbers indicate the amino acid residue participating in the binding motif(s). B. Alignment of *M. acetivorans* ferredoxin and FLD, showing the high degree of homology including the identical binding motifs of the two [4Fe-4S] clusters.

Chapter I

Subunit D of RNA Polymerase from *Methanosarcina acetivorans* contains two oxygen-labile [4Fe-4S] clusters: implications for oxidant-dependent regulation of transcription

Matthew E. Jennings

Department of Biological Sciences, University of Arkansas,

Fayetteville, AR 72701 USA

Preface

Some of the results presented in this chapter were derived from data collected by Dan Lessner, Faith Lessner, Akira Hirata, and Evert Duin; specifically the data generated from the study of recombinant subunit D, subunit DΔD3, and subunit L expressed in *E. coli*. These experiments were being conducted immediately before and during my first semester in the Lessner lab. However, since later experiments included in this dissertation built on the results of these initial experiments, it made sense to include them in this dissertation. Akira Hirata provided the homology model of *M. acetivorans* D-L heterodimer, and Evert Duin conducted the EPR measurements of D-L heterodimers. My contributions to this chapter include analysis of the D subunit from sequenced Archaea, and the generation of *M. acetivorans* strains capable of the expression of His-tagged subunit D and DΔD3.Followed by the purification and analysis of expressed proteins.

Abstract

Subunit D of multi-subunit RNA polymerase (RNAP) from many species of Archaea is predicted to bind one to two iron-sulfur (Fe-S) clusters, the function of which is unknown. A survey of encoded subunit D in the genomes of sequenced Archaea revealed six distinct groups based on the number of complete or partial [4Fe-4S] cluster motifs within domain 3. Only the subunits D from strictly anaerobic Archaea, including all members of the *Methanosarcinales*, are predicted to bind two [4Fe-4S] clusters. We report herein the purification and characterization of *Methanosarcina acetivorans* subunit D in complex with subunit L. Expression of subunit D and subunit L in *Escherichia coli* resulted in the purification of a D/L heterodimer with only partial [4Fe-4S] cluster content. Reconstitution *in vitro* with iron and sulfide revealed that the *M. acetivorans* D/L heterodimer is capable of binding two redox-active [4Fe-4S] clusters. *M. acetivorans* subunit D deleted of domain 3 (DΔD3) was still capable of co-purifying with subunit L, but was devoid of [4Fe-4S] clusters. Affinity-purification of subunit D or subunit DΔD3 from *M. acetivorans* resulted in the co-purification of endogenous subunit L with each tagged subunit D. Overall, these results suggest that domain 3 of subunit D is required for [4Fe-4S] cluster binding, but the [4Fe-4S] clusters and domain 3 are not required for the formation of the D/L heterodimer. However, exposure of two [4Fe-4S] cluster containing D/L heterodimer to oxygen resulted in loss of the [4Fe-4S] clusters and subsequent protein aggregation, indicating that the [4Fe-4S] clusters influence the stability of the D/L heterodimer and therefore have the potential to regulate the assembly and/or activity of RNAP in an oxidant-dependent manner.

Introduction

DNA-dependent RNA polymerase (RNAP) is a multi-subunit enzyme that is essential in all cellular organisms, where it acts in concert with a vast array of gene-specific activators, repressors, and general transcription factors to modulate the synthesis of RNA. The core catalytic component of RNAP is conserved in all three domains of life (*Bacteria*, *Archaea*, and *Eukarya*). The diversity in RNAPs lies in the number of RNAP complexes and their subunit composition. Eukaryotes have three separate nuclear RNAPs (Pol I, II, and III) that transcribe non-overlapping subsets of genes (Ebright, 2000, Werner, 2008). Each eukaryotic RNAP contains at least 10 subunits. In addition to the three RNAPs, higher plants possess Pol IV (NRPD1) and Pol V (NRPE1) that transcribe non-coding RNAs involved in RNA silencing (Haag & Pikaard, 2011). *Archaea* and *Bacteria* contain a single RNAP that transcribes all genes. Bacterial RNAP is the least complex, containing only the core component comprised of five subunits ($\alpha_2\beta\beta'\omega$). Archaeal RNAP consists of 12-13 subunits (RpoA-P) depending on phyla, similar to the subunit composition of eukaryotic RNAPs. Archaeal RNAP is most similar to eukaryotic Pol II, which consists of 12 subunits (Rpb1-12). The similarities between archaeal and eukaryotic RNAP were confirmed following the recent elucidation of the structures of archaeal RNAP (Hirata *et al.*, 2008, Korkhin *et al.*, 2009, Kusser *et al.*, 2008). One of the more striking features of the *Sulfolobus solfataricus* RNAP structure was the presence of an iron-sulfur (Fe-S) cluster in subunit D. Fe-S clusters are cofactors typically found in metabolic enzymes where they function in the transfer of electrons, but also have catalytic, sensing and structural roles (Johnson *et al.*, 2005). The function of Fe-S cluster(s) in RNAP is unknown.

Subunit D of archaeal RNAP contains three domains. Domains 1 and 2 are conserved in the homologous Rpb3 subunit of Pol II, AC40 subunit of Pol I/III, and bacterial α subunit (Hirata & Murakami, 2009). Domain 1 is divided into three regions and is involved in dimerization with

subunit L (RpoL), the initial step in the assembly of RNAP (Eloranta *et al.*, 1998, Goede *et al.*, 2006). Domain 2 is involved in the interaction with additional RNAP subunits and general transcription factors. Domain 2 of *S. solfataricus* subunit D contains four cysteines that form two intramolecular disulfides (Hirata *et al.*, 2008). However, these cysteines are not conserved in all RpoD/Rpb3/AC40/α subunits, and the four cysteines of domain 2 in yeast Rpb3 chelate a zinc ion (Spahr *et al.*, 2009). Domain 3 of *S. solfataricus* subunit D contains six cysteine residues, four of which are found in a typical [4Fe-4S] cluster binding motif. The structure of the D/L heterodimer from *S. solfataricus* revealed a [3Fe-4S] cluster present in domain 3 of subunit D, ligated by C183, C203, and C209 (Hirata *et al.*, 2008). The cluster is likely a [4Fe-4S] cluster *in vivo*, with C206 serving as the fourth ligand. The two remaining cysteines of domain 3 form an intramolecular disulfide. The [4Fe-4S] cluster binding motif is not restricted to *S. solfataricus* subunit D, but is found in D and Rpb3/AC40 subunits in various species of *Archaea* and *Eukarya*. However, all bacterial α subunits lack the residues of domain 3 and therefore are not predicted to bind Fe-S clusters (Hirata & Murakami, 2009).

Previous phylogenetic analysis indicated that there are four groups of archaeal D subunits based on the number of complete [4Fe-4S] cluster binding motifs in domain 3 (Hirata & Murakami, 2009). A re-examination here of the cysteine content in subunit D domain 3 encoded in the genomes of sequenced Archaea in the NCBI database (NCBI) indicates the classification should be expanded to six distinct groups based on complete or partial [4Fe-4S] cluster motifs (see results). Conservation of the [4Fe-4S] cluster binding motif(s) in RNAP in multiple archaeal lineages and in some eukaryotes suggests that the clusters serve an important role(s). The Fe-S clusters have been postulated to function in regulation of the *de novo* assembly of RNAP, in

which the presence of the cluster(s) is required for the formation of the D/L heterodimer (Hirata & Murakami, 2009), the first step in the assembly of RNAP in archaea.

Only strictly anaerobic archaea contain RNAP subunit D with two [4Fe-4S] cluster motifs. However, not every anaerobic archaeon possesses a RNAP subunit D with [4Fe-4S] cluster motifs. For example, all methane-producing archaea (methanogens) are strict anaerobes, yet there is extensive diversity in the properties of domain 3 of subunit D. Moreover, the ability of subunit D to bind two [4Fe-4S] clusters has not been documented. We report herein that subunit D from the methanogen *Methanosarcina acetivorans* binds two [4Fe-4S] clusters, consistent with the presence of two [4Fe-4S] cluster binding motifs, and that the clusters are similar to the [4Fe-4S] clusters in ferredoxin. Expression studies combined with mutational analysis also demonstrate that although domain 3 of subunit D is required for [4Fe-4S] cluster binding, the [4Fe-4S] clusters and amino acid residues of domain 3 are not required for the initial association of subunit D with subunit L. However, the [4Fe-4S] clusters are oxygen-labile, and oxygen-dependent loss of the [4Fe-4S] clusters impacts the structural stability of the D/L heterodimer. Implications for the regulation of RNAP assembly and activity by the [4Fe-4S] clusters in methanogens are presented.

Materials and Methods

Homology model of the M. acetivorans D/L heterodimer. A structural model of the D/L heterodimer from *M. acetivorans* was generated by SWISS-MODEL (Arnold *et al.*, 2006). The PDB coordinates of the *Sulfolobus solfataricus* D/L heterodimer (PDB ID: 2PA8) and ferredoxin-like domain of *Archaeoglobus fulgidus* adenylylsulfate reductase (PDB ID: 1JNR) were used as templates for the generation of the structural model. The model generated for

domain 3 including the two [4Fe-4S] clusters was manually traced onto domain 3 of the *Sulfolobus solfataricus* structure by COOT (Emsley & Cowtan, 2004).

Plasmid construction and mutagenesis. For co-expression of *M. acetivorans* subunits D and L in *E. coli*, the plasmid pMaRpoDL was constructed. Primers were used to amplify *rpoL* from *M. acetivorans* C2A genomic DNA resulting in a product with *Nde*I and *Xho*I sites at the 5' and 3' ends, respectively. The PCR product was digested with *Nde*I and *Xho*I and ligated to pET21a that had been similarly digested. The resulting plasmid (pRpoL) contained *rpoL* fused with a C-terminal six-histidine tag (His tag) under the control of the T7 promoter. The gene encoding *rpoD* was amplified from *M. acetivorans* C2A genomic DNA using a forward primer with a *Nde*I site and a reverse primer containing a *Bam*HI site. The PCR product was digested with *Nde*I and *Bam*HI and ligated with pET21a that was digested with the same enzymes. The resultant plasmid, pRpoD, contained an untagged *rpoD* gene under the control of the T7 promoter. pRpoD was digested with *Bgl*II and *Bam*HI. The roughly 1 kb fragment containing the T7 promoter and *rpoD* was gel-extracted, purified, and ligated with *Bgl*II-digested pRpoL, creating the pRpoDL plasmid. pRpoDL contains both *rpoD* and *rpoL*, each regulated by a separate T7 promoter. Importantly, only subunit L contains a His tag, designated subunit L(his). *E. coli* DH5α was used as the parent strain for all manipulations and was grown in Luria broth or agar containing appropriate antibiotics at 37 °C.

PCR amplification was used to remove the domain 3 amino acid residues 171 to 221 from *M. acetivorans* subunit D, the region which encompasses both [4Fe-4S] cluster motifs. Primers RpoDFeS2KORev (5'- AAT GGT AAT TAC AGG CAT GTT TTT G -3') and RpoDFeS1KOFor (5'- GTG GAC TTC TAT GAA AAC TCT TTT G -3') were 5'-

phosphorylated and used to amplify pMaRpoDL minus domain 3 of subunit D. The ends of the amplified product were blunt-ligated forming pDL408. In-frame deletion of the codons encoding amino acid residues 171-221 of subunit D was confirmed by DNA sequencing. Subunit D harboring the deletion was designated subunit DΔD3. Plasmid pDL408 was used for co-expression of subunit DΔD3 with subunit L(his).

Expression and purification of recombinant M. acetivorans D/L heterodimer. All recombinant proteins were expressed in *E. coli* Rosetta (DE3) (pLacI) cells containing pRKISC, which increases the level of iron-sulfur cluster biogenesis proteins (Takahashi & Nakamura, 1999). Unless stated otherwise, cells were grown in Terrific broth containing 50 µg/ml ampicillin, 17 µg/ml chloramphenicol, and 5 µg/ml tetracycline at 37 °C with shaking at 250 rpm. Once an optical density at 600 nm of 0.5 to 0.7 was reached, the growth temperature was adjusted to 25 °C, 0.2 mM ferrous ammonium sulfate was added to the medium, and the culture was induced with the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were harvested by centrifugation 16 h after induction.

All subsequent purification procedures were performed anaerobically within an anaerobic chamber (Coy Laboratory Products) containing an atmosphere of 95% N_2 and 5% H_2. Approximately 10 g of cells was suspended in 35 ml of 20 mM Tris-HCl (pH 8) containing 500 mM NaCl, 2 mM benzamidine, and a few crystals of DNaseI. The cells were lysed by two passages through a French pressure cell at over 110 Mpa. The lysate was centrifuged at 70,000 x *g* for 35 min at 4 °C. The supernatant containing the expressed protein(s) was filtered (pore size, 0.45 µm) and applied at a flow rate of 0.5 ml min^{-1} to a column containing 5 ml of Ni^{2+}- agarose resin (Genscript). The column was then washed with 25 ml of 20 mM Tris-HCl (pH 8), 500 mM

NaCl, followed by a 25 ml wash with the same buffer containing 10 mM imidazole. Bound protein(s) were eluted from the column by the addition of 15 ml of 20 mM Tris-HCl (pH 8), 500 mM NaCl containing 250 mM imidazole. The eluate was concentrated using a Vivacell concentrator (Sartorius) with a 10,000-molecular-weight cutoff under nitrogen flow inside the anaerobic chamber. The concentrated protein was desalted into 50 mM HEPES (pH 7.5) containing 150 mM NaCl using a PD-10 column (GE Healthcare). The desalted protein was either frozen under N_2 at -80 °C or immediately processed by size-exclusion chromatography.

Size-exclusion chromatography was performed using a Sephacryl HiPrep S-200 gel filtration fast protein liquid chromatography column (GE Healthcare) with a Biologic LP system (Biorad) within an anaerobic chamber. Samples were analyzed with a running buffer of 50 mM HEPES (pH 7.5) containing 150 mM NaCl and 2 mM DTT at a flow rate of 0.5 ml min^{-1}. The column was calibrated with the following proteins having known molecular masses: β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), and cytochrome *c* (12.4 kDa). Samples (2 ml) containing 50-75 mg of total protein were loaded onto the column per run.

D/L(his) proteins were reconstituted with iron and sulfide similar to as described (Cruz & Ferry, 2006). All additions and incubations were done under anaerobic conditions at 4 °C. Briefly, the eluate containing subunit D and subunit L(his) from the Ni^{2+}-agarose resin was concentrated and desalted into 50 mM Tris-HCl (pH 8.0). Approximately 50 mg of total protein was diluted in 50 ml of 50 mM Tris-HCl pH 8.0 containing 2 mM β-mercaptoethanol in a 125 ml serum bottle. After 30 minutes, ferrous ammonium sulfate was added drop-wise to a final concentration of 120 mM. After another incubation of 30 minutes, sodium sulfide was added drop-wise to a final concentration of 120 mM. The protein/iron/sulfide mixture was incubated for

16 h at 4 °C. The protein was then concentrated using a stirred cell concentrator (Millipore) with a 10,000-molecular weight cutoff and subsequently desalted into 50 mM HEPES pH 7.5 containing 150 mM NaCl and analyzed by size-exclusion chromatography.

Analytical methods. Protein concentrations were determined by the method of Bradford using bovine serum albumin as a standard (Bradford, 1976). The concentration of the D/L(his) heterodimer was also determined using the calculated extinction coefficient at 280 nm of 21,890 mM^{-1} cm^{-1}. The results from both methods were in good agreement (< 5% deviation). The iron and acid-labile sulfide content of D/L(his) heterodimers was determined as previously described (Lessner & Ferry, 2007). The number of thiols in D/L(his) heterodimers was determined using Ellman's reagent (dithionitrobenzoate, DTNB) (Ellman, 1959). In brief, thiol measurements were performed in 100 mM phosphate buffer (pH 8) with 0.2 mM DTNB in both the presence and absence of 6 M guanidine hydrochloride. The absorbance at 412 nm was measured after room temperature incubation for 15 minutes. An extinction coefficient at 412 nm of 14,150 mM^{-1} cm^{-1} and 13,700 mM^{-1} cm^{-1} was used for sulfhydryl content determination in buffer or buffer containing guanidine hydrochloride, respectively. Mass spectrometry of purified proteins was performed at the University of Arkansas Statewide Mass Spectrometry facility (Fayetteville, AR). SDS-PAGE and Western blotting were performed by standard procedures. The anti-RpoDL antibody was raised in rabbit against recombinantly-purified D/L(his) heterodimer.

CW electron paramagnetic resonance (EPR) spectra were measure at X-band (9 GHz) frequency on a Bruker EMX spectrometer, fitted with the ER-4119-HS high sensitivity perpendicular-mode cavity. General EPR conditions were: microwave frequency, 9.385 GHz;

field modulation frequency, 100 kHz; field modulation amplitude, 0.6 mT. Sample specific conditions are indicated in the figure legends.

Cloning, expression, and purification of subunit D and DΔD3 with a N-terminal His-tag in M. acetivorans. PCR was used to amplify *rpoD* from *M. acetivorans* C2A genomic DNA. The forward primer for the amplification contained the sequence for an *Nde*I restriction site and an N-terminal His-tag (5'- ATT AAG GCA TAT GCA TCA TCA TCA TCA TCA TAC GAT GGA AGT AGA CAT TCT -3'), while the reverse primer contained a *Hind*III restriction site (5'- GGT GGT AAG CTT TCA GAG CTG GTC CAG AAT TGC -3'). The PCR product was digested with *Nde*I and *Hind*III and ligated with similarly digested pJK027A (Guss *et al.*, 2008). The resulting plasmid was named pDL516. To construct a similar plasmid carrying the RpoDΔD3 deletion, the same procedure listed here was performed, only using pDL408 as a PCR template, with the resulting plasmid named pDL409.

M. acetivorans strain WWM73 was transformed with pDL516 and pDL409 as previously described (Metcalf *et al.*, 1997). Successful integration of the plasmid was determined as described (Guss *et al.*, 2008), and the resulting strains named DJL30 (His-D) and DJL31 (His-DΔD3). These strains allow for the tetracycline-regulated chromosomal expression of His-D or His-DΔD3. Cultures of DJL30 or DJL31 were grown anaerobically in HS medium (Sowers *et al.*, 1993) supplemented with methanol and sodium sulfide in the presence and absence of 100 µg/ml tetracycline at 37 °C. The cells were harvested by centrifugation at an optical density at 600 nm of 0.6 under anaerobic conditions.

All purification steps were performed in an anaerobic environment using either, lysis buffer (10 mM Tris-HCl (pH 8), 1 M KCl, 10 µM $ZnCl_2$, 10 mM imidazole, 15% glycerol), wash

buffer (10 mM Tris-HCl (pH 8), 0.1 M KCl, 10 µM $ZnCl_2$, 10 mM imidazole, 15% glycerol) or elution buffer (10 mM Tris-HCl (pH 8), 0.1 M KCl, 10 µM $ZnCl_2$, 200 mM imidazole, 15% glycerol). Cell pellets were thawed and suspended on ice in lysis buffer supplemented with a few crystals of DNaseI and benzamidine-HCl. Cells were lysed by sonication, and clarified lysates were obtained after a 35 minute centrifugation step at 70,000 x *g* at 4 °C. The soluble fraction was passed through a 0.45 µm filter and slowly loaded onto a Ni^{2+}-agarose column pre-equilibrated with lysis buffer. The column was washed with 10 column volumes (CV) of lysis buffer followed by 20 CV of wash buffer. Protein was eluted from the column with the addition of 3 CV of elution buffer. Eluates were stored anaerobically in sealed vials at -80 °C.

Results

Archaeal RNAP subunit D classification and diversity of methanogen subunit D. A survey of RNAP D subunits from ninety-nine sequenced archaeal genomes present in the NCBI database circa 2011 revealed six distinct groups based on the absence of the ferredoxin-like domain, or the presence of complete or partial [4Fe-4S] cluster motifs within the ferredoxin-like domain. The domain organization and cysteine residue content of a representative subunit D from each group is shown in Figure 1, with *S. solfataricus* subunit D shown as a reference. A complete list of subunit D from individual species and their grouping is listed in Table 1. Briefly, Group 1, complete [4Fe-4S] cluster #1 and #2 motifs (restricted to strict anaerobes); Group 2, a complete [4Fe-4S] cluster #1 motif; Group 3, a complete [4Fe-4S] cluster #2 motif; Group 4, partial [4Fe-4S] cluster #1 and/or #2 motifs; Group 5, absence of [4Fe-4S] cluster motif cysteines; Group 6, lack the amino acid residues comprising the ferredoxin-like domain (only *Methanococcales*). The corresponding subunit Rpb3 in Pol II or AC40 Pol I/III in certain eukaryotes contain only [4Fe-4S] cluster binding motif #1 and therefore fall in archaeal group 2.

All members of an order typically encode a subunit D which belongs to the same group, indicating the presence or absence of the clusters is correlated with particular taxonomic lineages. For example, all Halobacteriales encode a group 5 subunit D, one that is devoid of any cysteine residues. Group 1 subunit D is found only in strictly anaerobic Archaea, but is not a universal feature of strict anaerobes. For example, all methanogenic archaea are strict anaerobes and are only capable of growth by methanogenesis using a limited number of substrates, including acetate, H_2/CO_2, and methylated compounds (Zinder, 1993). However, despite the uniform mode of growth and nutritional requirements, there is extensive diversity in the cysteine residue content of subunit D, in particular the number of [4Fe-4S] cluster motifs (Table 1). Subunit D from all sequenced methanogens contains four cysteines in domain 2, except for members of the *Methanococcales*. Members of *Methanosarcinales*, *Methanomicrobiales*, and *Methanobacteriales* encode subunit D that contains one cysteine residue in domain 1. The majority of methanogens encode subunit D with a ferredoxin-like domain predicted to bind at least one [4Fe-4S] cluster. All members of the *Methanosarcinales* and *Methanomicrobiales* encode a group 1 subunit D, while members of *Methanobacteriales* encode a group 1 or 2 subunit D. All species of the *Methanococcales* encode subunit D which lacks the amino acid residues of domain 3 and are the only members which comprise group 6. The sole species of *Methanopyrales* encodes a group 4 subunit D. Thus, methanogens provide a unique opportunity to examine the effect of the presence or absence of the [4Fe-4S] clusters within a single group of organisms and correlate results obtained with the metabolism and environment of particular species.

Homology model of the group 1 subunit D from *Methanosarcina acetivorans* in complex with subunit L. The genome of *M. acetivorans* contains a single *rpoD* gene encoding a 266 amino acid protein with 32% identity to the 265 amino acid subunit D from *S. solfataricus*. However, in the original annotation of the *M. acetivorans* genome, *rpoD* was incorrectly identified as two separate genes (NTO2MA1356-7) (Galagan *et al.*, 2002). Based on the X-ray crystal structure of the *S. solfataricus* D/L heterodimer (Hirata *et al.*, 2008), a homology model of the *M. acetivorans* subunit D in complex with subunit L was generated (Fig. 2).

Similar to *S. solfataricus* subunit D, *M. acetivorans* subunit D contains three domains (Fig. 2A). There are however differences in the cysteine content of each domain between *S. solfataricus* and *M. acetivorans* D subunits. Domain 1 makes up the dimerization interface with subunit L. *M. acetivorans* subunit D contains a single cysteine residue (C160) in domain 1, unlike that of *S. solfataricus*. Domain 2 of *M. acetivorans* subunit D contains four cysteine residues similar to *S. solfataricus*; however, there is a difference in the spacing of the cysteines. The ferredoxin-like domain of *M. acetivorans* subunit D contains eight cysteine residues that comprise the two [4Fe-4S] cluster motifs (Group 1 subunit D), whereas the ferredoxin-like domain of *S. solfataricus* subunit D contains six cysteines (Group 2 subunit D). Moreover the 50-amino acid region of *M. acetivorans* subunit D ferredoxin-like domain (residues 171-221) has 39% identity to the predicted 60-amino acid 2[4Fe-4S] ferredoxin encoded by *ma0431* in the *M. acetivorans* genome (Fig. 2A). Based on the homology model of the *M. acetivorans* D/L heterodimer, neither domain 2 nor the ferredoxin-like domain is in contact with subunit L indicating they are likely unnecessary for dimerization with subunit L. In addition, all members of the *Methanoccocales* completely lack the amino acid residues of the ferredoxin-like domain suggesting it is likely not necessary for dimerization with subunit L in methanogens.

Expression and purification of recombinant *M. acetivorans* D/L heterodimer. To determine the capacity of *M. acetivorans* subunit D to bind Fe-S clusters and whether the clusters are required for the interaction and stable association of subunit D with subunit L, each subunit was separately expressed with a C-terminal six-histidine (his) tag in *E. coli*. Subunit L-His expressed in aerobically-grown *E. coli* was found in the soluble fraction of cell lysates and was purified to homogeneity using Ni^{2+}-affinity and size-exclusion chromatography (data not shown). However, subunit D-His expressed in aerobically-grown *E. coli* was found only in inclusion bodies in the insoluble fraction of cell lysates. Expression of untagged subunit D or subunit D with a glutathione-S-transferase (GST) tag also resulted in the formation of inclusion bodies (data not shown). The formation of inclusion bodies indicates that subunit D is improperly folded, potentially as a result of the absence of subunit L or incomplete Fe-S cluster incorporation. The lack of Fe-S clusters incorporated into subunit D may be a result of limiting cellular levels of iron, sulfur, or iron-sulfur biogenesis machinery during protein expression in *E. coli*. However, subunit D expressed in anaerobically-grown *E. coli* containing pRKISC, which encodes iron-sulfur cluster biogenesis proteins and has been shown to increase the cluster-content in recombinant Fe-S proteins (Nakamura *et al.*, 1999), combined with the addition of supplemental iron, still resulted in the protein being found in insoluble inclusion bodies (data not shown).

Subunit D was subsequently co-expressed with subunit L-His in aerobically-grown *E. coli*. Size-exclusion chromatography and SDS-PAGE analyses of imidazole-eluted protein from the Ni^{2+}-agarose loaded with cell lysate from *E. coli* co-expressing subunit D and subunit L-His revealed that subunit D co-purifies with subunit L-His (Fig. 3A). Three major peaks were detected, all of which contained subunit D and/or subunit L-His. Based on the elution profile of

known molecular weight standards, the elution volume of Peak 1 is consistent with D/L-His complexes larger than a heterodimer (M_r > 45 kDa). The elution volume of Peak 3 is consistent with the molecular weight of subunit L-His (M_r = 11.1 kDa). The elution volume of the major Peak (#2) is consistent with a heterodimer of D/L-His (M_r = 40.1 kDa). The presence of higher molecular weight D/L-His complexes, as well as subunit L-His monomers, may be a result of differences in expression levels of subunit D and subunit L-His. Co-purification of untagged subunit D with subunit L-His, combined with the insolubility of subunit D in the absence of subunit L under the same conditions, indicates that subunit D is likely unstable or improperly folded when not associated with subunit L.

Purified D/L-His heterodimer was red-brown in color, indicative of the presence of a chromophore. The UV-visible spectrum of the D/L-His heterodimer contained absorbance maxima centered at 320 nm and 390 nm, in addition to 280 nm (Fig. 4), consistent with the presence of Fe-S clusters (Sweeney & Rabinowitz, 1980). Although *M. acetivorans* subunit D is predicted to contain two [4Fe-4S] clusters, the A_{390}/A_{280} ratio is low, and the experimentally-determined iron and acid-labile sulfide content of as-purified D/L-His heterodimer is substantially less than predicted (Table 2). The ability to purify D/L-His heterodimer, which lacks two [4Fe-4S] clusters indicates that the presence of both clusters is not required for the interaction and dimerization of *M. acetivorans* subunit D with subunit L. Attempts to increase the cluster content by co-expression of subunit D with subunit L-His in anaerobically-grown *E. coli* containing pRKISC and supplementation of the medium with iron did not significantly increase the cluster content of as-purified D/L-His heterodimer (data not shown).

Reconstitution of [4Fe-4S] clusters in recombinant *M. acetivorans* D/L-His heterodimer. To determine the full capacity and type of Fe-S clusters within *M. acetivorans* subunit D, *in vitro* reconstitution of the eluate from the Ni^{2+}-agarose column containing a mixture of subunit D and subunit L-His was performed by incubation with iron and sulfide as previously described (Cruz & Ferry, 2006). The elution profile from size-exclusion chromatography of the Fe/S-reconstituted sample was identical to that of the non-reconstituted sample (Fig. 3). However, the major peak (#2) corresponding to the D/L-His heterodimer was of greater intensity than that of non-reconstituted D/L-His heterodimer, despite loading a similar amount of protein onto the column. The fractions containing peak #2 from the Fe/S-reconstituted sample were dark-brown in color, indicative of increased Fe-S cluster content. These fractions were collected and concentrated, and the sample was designated as the D/L-FeS heterodimer. The D/L-FeS heterodimer had a substantial increase in the A_{390}/A_{280} ratio and extinction coefficient at 390 nm compared to the D/L-His heterodimer (Table 2), consistent with an increase in Fe-S cluster content. Additionally, compared to the non-reconstituted D/L-His heterodimer, the D/L-FeS heterodimer contained eight-times the amount of iron and acid-labile sulfide (Table 2), consistent with the incorporation of two [4Fe-4S] clusters into subunit D. These data support that *M. acetivorans* subunit D coordinates two [4Fe-4S] clusters.

The D/L-FeS heterodimer showed a broad absorption band centered around 390 nm and a second peak centered around 300 nm, similar to the non-reconstituted heterodimer (Fig. 4). The similarity of the ferredoxin-like domain of *M. acetivorans* subunit D to 2[4Fe-4S] ferredoxin, which functions in electron transfer (Terlesky & Ferry, 1988b, Terlesky & Ferry, 1988a), indicates that the [4Fe-4S] clusters in subunit D may have the capacity to participate in oxidation-reduction reactions. The addition of the reductant dithionite to the D/L-FeS

heterodimer under anaerobic conditions resulted in a decrease of the absorption band at 390 nm, consistent with reduction of the clusters. EPR spectroscopic analyses of the D/L-FeS heterodimer in the absence of reductant did not reveal any significant signals (Fig. 5). The lack of a g = 2.01 EPR signal under these conditions indicates that the D/L-FeS heterodimer does not contain [3Fe-4S]$^{+}$ clusters, which is consistent with the elemental analyses (Table 2). Addition of dithionite to the sample of D/L-FeS resulted in the appearance of EPR signals with g values at 2.075, 1.974, 1.938, and 1.897 (Fig. 5). The complex g_{av} = 1.94 is consistent with the presence of two [4Fe-4S]$^{+}$ clusters, as is seen with ferredoxins harboring two spin-coupled [4Fe-4S] clusters (Clements *et al.*, 1994, Prince & Adams, 1987). The complexity is caused by spin-spin coupling of the two clusters in close vicinity of each other. Taken together these results support that the D/L-FeS heterodimer contains two redox-coupled [4Fe-4S]$^{2+/+}$ that can possibly participate in one-electron transfers. EPR spectroscopic analyses of the non-reconstituted D/L-His heterodimer did not produce an EPR signal in the as-purified state or after the addition of dithionite, which may be attributed to insufficient cluster in the samples analyzed. Attempts to obtain more concentrated samples of non-reconstituted D/L-His for EPR analysis resulted in the precipitation of the protein (data not shown).

Domain 3 of subunit D is required for [4Fe-4S] cluster binding, but is not required for interaction with subunit L. To determine if domain 3 of subunit D is required for [4Fe-4S] cluster binding and dimerization with subunit L, a variant of subunit D deleted of amino acid residues 171 to 221 (domain 3) was generated. This variant (subunit DΔD3) was co-expressed with subunit L-His and subjected to the same purification procedure as was used for wild-type subunit D. The elution profile from the size-exclusion column for the sample containing subunit

DΔD3 and subunit L-His) was similar to that observed for purification of the wild-type D/L-His heterodimer (Fig. 6). However, the DΔD3/L-His) heterodimer eluted at a greater elution volume, consistent with the expected smaller size of a the DΔD3/L-His heterodimer (35 kDa) compared to wild-type heterodimer (40 kDa). The sizes of D subunits and the presence of subunit L-His in each heterodimer was confirmed by SDS-PAGE (Fig. 6). The ability of subunit DΔD3 to co-purify with subunit L-His and form a stable heterodimer reveals that domain 3 is not required for the interaction of subunit D with subunit L. Moreover, deletion of domain 3 from subunit D resulted in a heterodimer devoid of [4Fe-4S] clusters, as revealed by UV-visible spectroscopy and elemental analyses (Fig. 4 and Table 2). These results support that the eight cysteine residues in domain 3 are necessary for the binding of the two [4Fe-4S] clusters in *M. acetivorans* subunit D.

The formation of the *M. acetivorans* D/L heterodimer does not require disulfides in subunit D. *M. acetivorans* subunit D contains thirteen cysteine residues (Fig. 2). *M. acetivorans* subunit L has one cysteine residue. Each cysteine may be in the free reduced state (thiol), participate in disulfide bonding or serve to coordinate a cofactor (e.g. [4Fe-4S] cluster). The structure of the *S. solfataricus* D/L heterodimer revealed that each cysteine that is not a ligand to the [4Fe-4S] cluster participates in disulfide bonding, suggesting disulfides may be required for the formation and stability of the heterodimer (Hirata *et al.*, 2008). Thus, the number of thiols in purified *M. acetivorans* D/L(his) heterodimers was quantified using the thiol-specific reagent DTNB to infer the presence or absence of disulfides (Ellman, 1959).

The purified D/L-His heterodimers were first incubated in non-denaturing buffer containing DTNB under anaerobic conditions. Incubation of the D/L-His heterodimer with

DTNB resulted in the release of TNB that corresponds to fourteen thiols (Table 2), indicating that all fourteen cysteine residues are in the reduced state. The DΔD3/L-His heterodimer reacted with DTNB under non-denaturing conditions to release TNB that corresponds to six thiols, consistent with the deletion of the eight cysteine residues of domain 3. There was no significant difference in the reactivity of each heterodimer with DTNB under denaturing conditions, indicating that each cysteine is readily accessible to DTNB in the folded state of the heterodimer. This result is consistent with the model showing that the domain 2 and 3 cysteine residues are relatively surface exposed within the heterodimer (Fig. 2). The cysteine residues of domain 2 of Rbp3 from yeast have been shown to coordinate a zinc ion, instead of participating in disulfide bonding. However, the reaction of C82, C85, C90, and C93 of domain 2 of *M. acetivorans* subunits D and subunit DΔD3 with DTNB under both non-denaturing and denaturing conditions suggests that these residues do not bind zinc or participate in disulfides under the examined conditions.

The quantitation of thiols in the D/L-FeS heterodimer was more variable. Incubation of D/L-FeS with DTNB under non-denaturing conditions produced TNB corresponding to sixteen thiols (Table 2). The number of thiols detected in the D/L-FeS heterodimer is likely a combination of TNB produced by thiols and by sulfide released from the [4Fe-4S] clusters, as the reaction of one sulfide anion with DTNB yields two TNB anions (Crack *et al.*, 2006). The addition of denaturant resulted in an increase in the number of thiols detected, consistent with an increase in sulfide released from the [4Fe-4S] clusters. These data indicate that the cysteine residues of the purified *M. acetivorans* D/L heterodimers are either in the thiol state or participate in coordination of the [4Fe-4S] clusters, indicating disulfides are not required for dimerization of *M. acetivorans* subunit D with subunit L.

The [4Fe-4S] clusters of *M. acetivorans* subunit D are oxygen-labile and affect the structural stability of the D/L heterodimer. The Fe-S clusters in proteins from methanogens are often oxygen-labile, including the [4Fe-4S] clusters in ferredoxin from *Methanosarcina* (Clements *et al.*, 1994). The exposure of the D/L-FeS heterodimer to air resulted in a time-dependent decrease in the absorbance at 390 nm, whereas no significant decrease at 390 nm was seen when the heterodimer was kept anaerobic (Fig. 7). To determine if destruction of the clusters due to oxygen exposure affects the stability of the D/L-His heterodimer, oxygen-induced structural changes were examined by size-exclusion chromatography. Each purified heterodimer was incubated under nitrogen or air for one hour and then subjected to size-exclusion chromatography under anaerobic conditions (Fig. 8A-C). For each sample, the fractions corresponding to heterodimer were collected, concentrated, and the UV-visible spectrum recorded (Fig. 8D-E). The elution profile of the non-reconstituted D/L-His heterodimer following anaerobic incubation resulted in a major heterodimer peak and a minor peak which was consistent with a larger molecular weight aggregate of D/L-His. Exposure to oxygen resulted in a very slight increase in the population of the minor, non-heterodimer, species. Importantly, oxygen did not cause disassociation of subunit D and subunit L, as indicated by the lack of peaks that corresponded to D or L monomers. Compared to the collected, anaerobically-incubated heterodimer, a decrease in absorbance in the 300-420 nm range of the collected aerobically-incubated heterodimer was observed, consistent with some loss of the small amount of Fe-S cluster present in the non-reconstituted D/L-His heterodimer.

The elution profile of the anaerobically-incubated D/L-FeS heterodimer displayed a single heterodimer peak. While the non-reconstituted D/L-His heterodimer exhibited some

aggregation after anaerobic incubation, the lack of this aggregation by the D/L-FeS suggests that the heterodimer is more stable (i.e. less structural heterogeneity) when both [4Fe-4S] clusters are present. In contrast, a substantial increase in the structural heterogeneity of the D/L-FeS heterodimer was observed following aerobic incubation, as indicated by the appearance of two additional peaks in the elution profile (Fig. 8B). The increased structural heterogeneity is a result of oxygen-induced [4Fe-4S] cluster loss in the D/L-FeS sample as revealed by the UV-visible spectrum of the collected heterodimer sample (Fig. 8E). The most likely explanation for the increased aggregation observed in the D/L(his)-FeS heterodimer, compared to the D/L-His heterodimer, upon oxygen exposure is due to Fe-catalyzed production of reactive oxygen species leading to amino acid oxidation and protein aggregation. The presence of only a heterodimer peak in the elution profile of both anaerobically-incubated and aerobically-incubated DΔD3/L-His heterodimer, demonstrates that domain 3 is involved in the formation of the D/L(his) aggregates. These results suggest that the presence or absence of domain 3, along with the [4Fe-4S] clusters affect the structural stability of the D/L heterodimer. Overall, these data support that the presence of both [4Fe-4S] clusters are important for the integrity of domain 3 and serve to stabilize the D/L heterodimer.

Domain 3 of subunit D is not required for the formation of the D/L heterodimer within *M. acetivorans*. Co-purification of subunit DΔD3 with subunit L(his) after co-expression in *E. coli*, along with the complete lack of domain 3 in subunit D from all members of the *Methanococcales*, indicate that domain 3 of subunit D is not required for the formation of the D/L heterodimer within *M. acetivorans*. To test this hypothesis, two merodiploid strains of *M. acetivorans* were constructed. In addition to expressing native subunit D, these strains contain a

second *rpoD* gene which encodes an N-terminally His-tagged subunit D and is under the control of a tetracycline-dependent promoter. Strain DJL30 expresses His-subunit D (His-D) and strain DJL31 expresses His-subunit DΔD3 (His-DΔD3). Importantly, subunit L was not tagged or over-expressed in these strains. Western blot analysis with anti-D/L antibodies of lysates from induced DJL30 cells did not reveal an increase in the concentration of subunit D, when compared to uninduced cells (data not shown). The inability to detect an increase in the cellular levels of subunit D in the merodiploid His-D expression strain is likely due to the degradation of unassembled subunits by proteases, which has been seen in yeast (Mitobe *et al.*, 2001). This result is consistent with His-D competing with native D for association with endogenous subunit L and assembly into RNAP, with unassembled His-D or D being degraded. Inducible expression was seen in the DJL31 strain where a smaller immunoreactive protein (His-DΔD3) was detected only when DJL31 cells were grown in the presence of tetracycline (data not shown). Strains DJL30 and DJL31 grown in the absence and presence of tetracycline exhibit a similar growth rate and cell yield (data not shown), indicating the expression of subunit D containing an affinity tag or the presence of subunit DΔD3 is not detrimental to *M. acetivorans*.

To verify expression and to determine if His-D and His-DΔD3 can associate with native subunit L, clarified cell lysates from strains DJL30 and DJL31 grown in the presence or absence of tetracycline were used for Ni^{2+}-affinity purification. Imidazole eluates from the columns were analyzed by Western blot using anti-D/L antibodies. A band consistent with subunit D or DΔD3 was only detected in the eluate from lysate of cells grown in the presence of tetracycline (Fig. 9), consistent with inducible expression of His-D and His-DΔD3 in strains DJL30 and DJL31, respectively. Moreover, endogenous subunit L (untagged) co-purified with both subunit His-D and His-DΔD3, revealing that domain 3 of subunit D is not required for the formation of the D/L

heterodimer within *M. acetivorans* (Fig. 9). The additional immunoreactive proteins in the eluates are degradation products of unassembled His-D and His-DΔD3, similar to the appearance of Rpb3 degradation products when affinity-tagged Rpb3 is expressed in yeast (Kimura & Ishihama, 2000, Mitobe *et al.*, 2001, Svetlov *et al.*, 1998).

Discussion

The results presented herein reveal the ability of subunit D of RNAP from *M. acetivorans* to bind two redox-active and oxygen-labile [4Fe-4S] clusters. Given that all archaeal species harboring group 1 subunit D are strictly anaerobic, similar to *M. acetivorans*, it is highly likely that all group 1 subunit D contain two [4Fe-4S] clusters. However, the presence of a group 1 subunit D is not a universal feature of strictly anaerobic Archaea. In fact, there is extensive diversity in the properties of subunit D within the five orders of methanogens, indicating that the [4Fe-4S] clusters are not essential to the function of RNAP in methanogens. The presence of one or two [4Fe-4S] clusters in subunit D of archaeal RNAP likely imparts some advantage to those species that have acquired or retained this feature within their RNAP. Other DNA processing and repair enzymes, including helicase, primase, glycosylases and endonucleases, bind [4Fe-4S] clusters (Genereux *et al.*, 2010), indicating these cofactors are a common, but not universal, feature of DNA processing enzymes. In many of these enzymes the [4Fe-4S] cluster is important to the structural stability of the enzyme and serves to monitor the redox state of the cell and modulate enzyme activity. For example, the DNA damage-inducible protein DinG, a DNA helicase, contains a [4Fe-4S] cluster that regulates DinG helicase activity (Ren *et al.*, 2009). The [4Fe-4S] clusters in RNAP may serve a similar role. Based on previous structure-function studies of bacterial α N-terminal domain (αNTD) (Ebright & Busby, 1995, Ishihama, 1981), eukaryotic

Rpb3 (Kimura *et al.*, 1997, Kimura & Ishihama, 2000), and archaeal D of RNAPs (Werner, 2008), along with results obtained with *M. acetivorans* subunit D, a plausible function for RNAP clusters is to regulate the *de novo* assembly of RNAP or to modulate the post-assembly activity of RNAP. A proposed model for the possible role(s) of the [4Fe-4S] clusters in *M. acetivorans* subunit D is presented in Fig. 10.

Iron is used as a cofactor in numerous metabolic enzymes in virtually all organisms. This is especially true in anaerobes, including methanogens. Methanogens are predicted to contain the largest number of [4Fe-4S] proteins, many of which are oxygen-labile (Major *et al.*, 2004). Many enzymes directly involved in methanogenesis including hydrogenases, carbon monoxide dehydrogenase, and heterodisulfide reductase, contain Fe-S clusters (Ferry, 2003, Lessner, 2009). These enzymes are rapidly inactivated in the presence of oxygen due to the oxidation and/or destruction of the Fe-S clusters, resulting in a decrease in energy available for biosynthesis. Moreover, 2[4Fe-4S] ferredoxin is a key electron transfer protein that interacts with a number of redox partners, including carbon monoxide dehydrogenase (Ferry, 2003). In *M. acetivorans* and other organisms, the [4Fe-4S] clusters in ferredoxin-like domain 3 of subunit D may impart the ability to directly integrate global gene expression with the metabolic state of the cell.

One plausible role for the [4Fe-4S] clusters is to regulate the assembly of RNAP in response to cellular conditions, which is supported by previous studies documenting archaeal subunit D, bacterial α subunit and eukaryotic Rpb3 as key subunits that initiate the assembly of RNAP (Eloranta *et al.*, 1998, Goede *et al.*, 2006, Ishihama, 1981, Kimura *et al.*, 1997). The D/α/Rpb3 subunits are not directly involved in binding of template DNA and synthesis of RNA. Instead, one function of these subunits is to serve as a scaffold for the assembly and proper

interaction of the catalytic subunits. In bacteria, the assembly of RNAP is initiated by α subunit dimerization and is assembled in the following order: $\alpha_I + \alpha_{II} \rightarrow \alpha_{I/II} \rightarrow \alpha_{I/II}\beta \rightarrow \alpha_{I/II}\beta\beta' \rightarrow \alpha_{I/II}\beta\beta'\sigma$ (Ebright & Busby, 1995). The determinants for α dimerization are located within the αNTD. The β and β' subunits comprise the catalytic core. The assembly of eukaryotic Pol II is initiated by Rpb3 dimerization with Rpb11, followed by addition of Rpb2 to form the assembly subcomplex Rpb2-Rpb3-Rpb11, analogous to $\alpha_{I/II}\beta$. The catalytic subunit Rpb1, and the auxiliary subunits Rpb4-10 and Rbp12 then assemble to form complete Pol II. Similarly, the assembly of archaeal RNAP is initiated by the formation of a heterodimer of subunits D and L, followed by the addition of subunits N and P, forming the DLNP assembly platform (Werner, 2007). A stable BDLNP subcomplex that is competent in DNA binding and transcription factor interaction has also been identified, indicating subunit B is likely added prior to the addition of the core A' and A" subunits and auxiliary HKFE subunits (Goede *et al.*, 2006). Subunit B is split into two polypeptides (B' and B") in some Euryarchaeota, including *M. acetivorans* (Werner, 2007).

The ability to purify the *M. acetivorans* D/L heterodimer which lacks the [4Fe-4S] clusters reveals that the clusters are not required for initial D/L heterodimer formation. However, the presence and redox state of the [4Fe-4S] clusters in domain 3 of subunit D may influence the interaction of subunit D with additional assembly or catalytic subunits. Moreover, the instability of recombinant subunit D when expressed in the absence of subunit L indicates that subunit L serves to stabilize subunit D and that dimerization occurs prior to [4Fe-4S] cluster incorporation (Fig. 10). Based on the *S. solfataricus* RNAP structure and previous interaction studies with *Pyrococcus furious* RNAP, the D/L heterodimer directly contacts subunits B, N, and P (Goede *et al.*, 2006, Hirata *et al.*, 2008). In particular, domain 3 of *S. solfataricus* subunit D is in close proximity to subunit B (B' in *M. acetivorans* RNAP), indicating that the conformational state of

domain 3 could impact the interaction between the D/L heterodimer and the B subunit. The lack of cluster incorporation or loss of incorporated clusters likely increases the flexibility of domain 3 which could prevent favorable interaction of the D/L heterodimer with subunits B, N and P (Fig. 10).

Because *M. acetivorans* is a strict anaerobe and exposure to oxygen shuts down metabolism (Brioukhanov *et al.*, 2006, Lessner & Ferry, 2007), one possible mechanism to globally conserve energy in order to maintain critical functions during times of iron-depletion or oxidative stress would be to decrease production of RNA to minimum levels adequate for cellular maintenance. The lack of cluster incorporation, the oxidation of the clusters, or oxidative loss of incorporated clusters within domain 3 of subunit D may prevent the *de novo* assembly of RNAP. Thus, the [4Fe-4S] clusters in RNAP may be used to sense metabolic factors within the cell, such as iron/sulfur levels and intracellular redox state, which are influenced by environmental factors (e.g. nutrients and oxygen). The reduction potential and oxidative/reductive stability of the clusters is likely tuned to the metabolism of each particular species. For example, the single [4Fe-4S] cluster in *S. solfataricus* subunit D, an aerobic archaeon, is quite stable in air compared with the two [4Fe-4S] clusters in subunit D from *M. acetivorans*, a strict anaerobe. One possible explanation for this difference is the presence of the disulfide bond, instead of a second cluster, in domain 3 of *S. solfataricus* subunit D (Hirata *et al.*, 2008), which is absent from *M. acetivorans* subunit D. The disulfide may serve to stabilize the cluster in *S. solfataricus* subunit D, such that the cluster is only lost under more extreme oxidative conditions. In addition to preventing the assembly of RNAP, oxidation or loss of the clusters in assembled RNAP may induce dispersion and subsequent inactivation of RNAP.

An alternative function of the clusters may be to regulate RNAP activity, whereby the clusters serve as a recognition element necessary for optimal protein-protein interactions with general or gene-specific transcription factors. Importantly, subunit D is on the periphery of archaeal RNAP, as is Rpb3 on eukaryotic Pol II, in a position to interact with general or specific transcription factors (Cramer *et al.*, 2001, Hirata *et al.*, 2008). The α subunit of bacterial RNAP is also the primary subunit that interacts with specific transcription factors to regulate promoter-dependent gene expression in bacteria (Ebright & Busby, 1995). Moreover, Rpb3 has been shown to directly interact with a number of transcription factors (De Angelis *et al.*, 2003, Oufattole *et al.*, 2006). An Rpb3 mutagenesis study in yeast identified Rpb3 mutants with a temperature-sensitive defect in activator-dependent transcription, indicating that Rpb3 contains determinants for interaction with transcription factors (Tan *et al.*, 2000). More recently, a direct interaction of Rpb3 with the Med17 subunit of the Mediator complex was demonstrated using an *in vivo* cross-linking approach (Soutourina *et al.*). Mediator is a large multi-subunit complex conserved in eukaryotes that is required for transcription of most Pol II-transcribed genes. Mediator also serves to link specific regulatory proteins with the Pol II transcription complex. Therefore, Rpb3 is important for assembly and the direct interaction with general and specific transcription factors. Proteins that interact with subunit D of archaeal RNAP have not been identified, but given the high degree of similarity to Rpb3 it is highly likely subunit D contains regions required for interaction with transcription factors.

In *M. acetivorans,* the two[4Fe-4S] cluster-containing domain 3 of subunit D may function as a recognition element for the interaction with regulatory proteins needed to recruit RNAP to specific promoters or to control transcription initiation rates. Furthermore, given the rapid loss of the [4Fe-4S] clusters in the purified D/L(FeS) heterodimer upon exposure to air, the

clusters may be used to sense the redox state of the cell to direct changes in gene expression. However, there may be differences in the stability of the [4Fe-4S] clusters in fully assembled RNAP. For example, the [4Fe-4S] cluster in *S. solfataricus* RNAP was more recalcitrant to removal by chelators as compared to the [4Fe-4S] cluster in D/L heterodimer. As seen with other DNA-interacting proteins, changes in the redox state of the [4Fe-4S] clusters in *M. acetivorans* RNAP may be enough to induce structural changes in domain 3 that alter the affinity for interacting partners. Oxidized RNAP may be recruited to the promoters of genes involved in response to stress. To test this hypothesis it will be necessary to purify *M. acetivorans* RNAP and compare the properties of the [4Fe-4S] clusters to those of clusters in the D/L heterodimer. It is possible the clusters affect both the assembly and activity of RNAP. For example, exposure of *M. acetivorans* to oxygen and loss of the clusters from the D/L heterodimer may prevent assembly and oxidation or loss of the clusters from RNAP may direct RNAP to the promoters of genes needed for cellular maintenance. Using the established *M. acetivorans* genetic system, combined with *in vitro* approaches, it will be possible to test both the assembly and activity function of the [4Fe-4S] clusters. For example, the ability to generate the *M. acetivorans* strain capable of expressing subunit DΔD3 without a negative phenotype, indicates it will be possible to use this mutant to determine subunit D interacting partners which are dependent on domain 3.

Conclusions

Our results from a survey of RNAP subunit D from sequenced Archaea reveal extensive diversity in the cysteine content and number of [4Fe-4S] cluster motifs, dividing archaeal RNAP into six distinct groups. Subunit D from species of an individual order typically falls within a single group, indicating that subunit D may be used as an additional marker for phylogenetic

analyses. For example, all members of the Methanosarcinales encode a group 1 subunit D, whereas all Methanococcales encode group 6 subunit D. We have also demonstrated that subunit D from *M. acetivorans* is capable of binding two redox-active and oxygen-labile [4Fe-4S] clusters. Data obtained from recombinant studies reveal that the clusters and domain 3 are not required for the formation of the D/L heterodimer, which was supported by *in vivo* studies revealing that domain 3 is not required for D/L heterodimer formation within *M. acetivorans*. Overall, these results suggest the clusters are not essential to RNAP, consistent with a regulatory role. Given the extreme sensitivity of the clusters to oxygen and the cluster-induced structural changes of the D/L heterodimer, a potential function is in the redox-dependent regulation of RNAP assembly and/or activity.

Acknowledgements

We thank Professor Y. Takahashi for providing pRKISC, Bill Metcalf for *M. acetivorans* strains and plasmids, Mack Ivey for critical reading of the manuscript, and Liz Karr for helpful discussions.

References

Arnold, K., L. Bordoli, J. Kopp & T. Schwede, (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201.

Bradford, M.M., (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-254.

Brioukhanov, A.L., A.I. Netrusov & R.I. Eggen, (2006) The catalase and superoxide dismutase genes are transcriptionally up-regulated upon oxidative stress in the strictly anaerobic archaeon Methanosarcina barkeri. *Microbiology* 152: 1671-1677.

Clements, A.P., L. Kilpatrick, W.P. Lu, S.W. Ragsdale & J.G. Ferry, (1994) Characterization of the iron-sulfur clusters in ferredoxin from acetate-grown Methanosarcina thermophila. *J. Bacteriol.* 176: 2689-2693.

Crack, J.C., J. Green, N.E. Le Brun & A.J. Thomson, (2006) Detection of sulfide release from the oxygen-sensing [4Fe-4S] cluster of FNR. *J Biol Chem* 281: 18909-18913.

Cramer, P., D.A. Bushnell & R.D. Kornberg, (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292: 1863-1876.

Cruz, F. & J.G. Ferry, (2006) Interaction of iron-sulfur flavoprotein with oxygen and hydrogen peroxide. *Biochim. Biophys. Acta.* 1760: 858-864.

De Angelis, R., S. Iezzi, T. Bruno, N. Corbi, M. Di Padova, A. Floridi, M. Fanciulli & C. Passananti, (2003) Functional interaction of the subunit 3 of RNA polymerase II (RPB3) with transcription factor-4 (ATF4). *FEBS Lett.* 547: 15-19.

Ebright, R.H., (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* 304: 687-698.

Ebright, R.H. & S. Busby, (1995) The *Escherichia coli* RNA polymerase alpha subunit: structure and function. *Curr. Opin. Genet. Dev.* 5: 197-203.

Ellman, G.L., (1959) Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* 82: 70-77.

Eloranta, J.J., A. Kato, M.S. Teng & R.O. Weinzierl, (1998) In vitro assembly of an archaeal D-L-N RNA polymerase subunit complex reveals a eukaryote-like structural arrangement. *Nucleic Acids Res.* 26: 5562-5567.

Emsley, P. & K. Cowtan, (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* 60: 2126-2132.

Ferry, J.G., (2003) One-carbon metabolism in methanogenic anaerobes. In: Biochemistry and Physiology of Anaerobic Bacteria. L.G. Ljungdahl, M.W. Adams, L.L. Barton, J.G. Ferry & M.K. Johnson (eds). New York: Springer-Verlag, pp. 143-156.

Galagan, J.E., C. Nusbaum, A. Roy, M.G. Endrizzi, P. Macdonald, W. FitzHugh, S. Calvo, R. Engels, S. Smirnov, D. Atnoor, A. Brown, N. Allen, J. Naylor, N. Stange-Thomann, K. DeArellano, R. Johnson, L. Linton, P. McEwan, K. McKernan, J. Talamas, A. Tirrell, W. Ye, A. Zimmer, R.D. Barber, I. Cann, D.E. Graham, D.A. Grahame, A.M. Guss, R. Hedderich, C. Ingram-Smith, H.C. Kuettner, J.A. Krzycki, J.A. Leigh, W. Li, J. Liu, B. Mukhopadhyay, J.N. Reeve, K. Smith, T.A. Springer, L.A. Umayam, O. White, R.H. White, E. Conway de Macario, J.G. Ferry, K.F. Jarrell, H. Jing, A.J. Macario, I. Paulsen, M. Pritchett, K.R. Sowers, R.V. Swanson, S.H. Zinder, E. Lander, W.W. Metcalf & B. Birren, (2002) The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome. Res.* 12: 532-542.

Genereux, J.C., A.K. Boal & J.K. Barton, (2010) DNA-mediated charge transport in redox sensing and signaling. *J. Am. Chem. Soc.* 132: 891-905.

Goede, B., S. Naji, O. von Kampen, K. Ilg & M. Thomm, (2006) Protein-protein interactions in the archaeal transcriptional machinery: binding studies of isolated RNA polymerase subunits and transcription factors. *J. Biol. Chem.* 281: 30581-30592.

Guss, A.M., M. Rother, J.K. Zhang, G. Kulkarni & W.W. Metcalf, (2008) New methods for tightly regulated gene expression and highly efficient chromosomal integration of cloned genes for Methanosarcina species. *Archaea* 2: 193-203.

Haag, J.R. & C.S. Pikaard, (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell. Biol.* 12: 483-492.

Hirata, A., B.J. Klein & K.S. Murakami, (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451: 851-854.

Hirata, A. & K.S. Murakami, (2009) Archaeal RNA Polymerase. *Curr. Opin. Struct. Biol.* 19: 724-731.

Ishihama, A., (1981) Subunit of assembly of Escherichia coli RNA polymerase. *Adv. Biophys.* 14: 1-35.

Johnson, D.C., D.R. Dean, A.D. Smith & M.K. Johnson, (2005) Structure, function, and formation of biological iron-sulfur clusters. *Annu. Rev. Biochem.* 74: 247-281.

Kimura, M., A. Ishiguro & A. Ishihama, (1997) RNA polymerase II subunits 2, 3, and 11 form a core subassembly with DNA binding activity. *J. Biol. Chem.* 272: 25851-25855.

Kimura, M. & A. Ishihama, (2000) Involvement of multiple subunit-subunit contacts in the assembly of RNA polymerase II. *Nucleic Acids Res.* 28: 952-959.

Korkhin, Y., U.M. Unligil, O. Littlefield, P.J. Nelson, D.I. Stuart, P.B. Sigler, S.D. Bell & N.G. Abrescia, (2009) Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLoS Biol.* 7: e102.

Kusser, A.G., M.G. Bertero, S. Naji, T. Becker, M. Thomm, R. Beckmann & P. Cramer, (2008) Structure of an archaeal RNA polymerase. *J. Mol. Biol.* 376: 303-307.

Lessner, D., (2009) Methanogenesis: Biochemistry. In: Encyclopedia of Life Sciences. Nature Publishing Group, pp.

Lessner, D.J. & J.G. Ferry, (2007) The Archaeon *Methanosarcina acetivorans* Contains a Protein Disulfide Reductase with an Iron-Sulfur Cluster. *J. Bacteriol.* 189: 7475-7484.

Major, T.A., H. Burd & W.B. Whitman, (2004) Abundance of 4Fe-4S motifs in the genomes of methanogens and other prokaryotes. *FEMS Microbiol. Lett.* 239: 117-123.

Metcalf, W.W., J.K. Zhang, E. Apolinario, K.R. Sowers & R.S. Wolfe, (1997) A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. *Proc. Natl. Acad. Sc.i USA* 94: 2626-2631.

Mitobe, J., H. Mitsuzawa & A. Ishihama, (2001) Functional analysis of RNA polymerase II Rpb3 mutants of the fission yeast Schizosaccharomyces pombe. *Curr. Genet.* 39: 210-221.

Nakamura, M., K. Saeki & Y. Takahashi, (1999) Hyperproduction of recombinant ferredoxins in escherichia coli by coexpression of the ORF1-ORF2-iscS-iscU-iscA-hscB-hs cA-fdx-ORF3 gene cluster. *J. Biochem.* 126: 10-18.

Oufattole, M., S.W. Lin, B. Liu, D. Mascarenhas, P. Cohen & B.D. Rodgers, (2006) Ribonucleic acid polymerase II binding subunit 3 (Rpb3), a potential nuclear target of insulin-like growth factor binding protein-3. *Endocrinology* 147: 2138-2146.

Prince, R.C. & M.W. Adams, (1987) Oxidation-reduction properties of the two Fe4S4 clusters in Clostridium pasteurianum ferredoxin. *J. Biol. Chem.* 262: 5125-5128.

Ren, B., X. Duan & H. Ding, (2009) Redox control of the DNA damage-inducible protein DinG helicase activity via its iron-sulfur cluster. *J. Biol. Chem.* 284: 4829-4835.

Soutourina, J., S. Wydau, Y. Ambroise, C. Boschiero & M. Werner, Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* 331: 1451-1454.

Sowers, K.R., J.E. Boone & R.P. Gunsalus, (1993) Disaggregation of Methanosarcina spp. and Growth as Single Cells at Elevated Osmolarity. *Appl. Environ. Microbiol.* 59: 3832-3839.

Spahr, H., G. Calero, D.A. Bushnell & R.D. Kornberg, (2009) Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution. *Proc. Natl. Acad. Sc.i USA* 106: 9185-9190.

Svetlov, V., K. Nolan & R.R. Burgess, (1998) Rpb3, stoichiometry and sequence determinants of the assembly into yeast RNA polymerase II in vivo. *J. Biol. Chem.* 273: 10827-10830.

Sweeney, W.V. & J.C. Rabinowitz, (1980) Proteins containing 4Fe-4S clusters: an overview. *Annu. Rev. Biochem.* 49: 139-161.

Takahashi, Y. & M. Nakamura, (1999) Functional assignment of the ORF2-iscS-iscU-iscA-hscB-hscA-fdx-ORF3 gene cluster involved in the assembly of Fe-S clusters in Escherichia coli. *J. Biochem.* 126: 917-926.

Tan, Q., K.L. Linask, R.H. Ebright & N.A. Woychik, (2000) Activation mutants in yeast RNA polymerase II subunit RPB3 provide evidence for a structurally conserved surface required for activation in eukaryotes and bacteria. *Genes Dev.* 14: 339-348.

Terlesky, K.C. & J.G. Ferry, (1988a) Ferredoxin requirement for electron transport from the carbon monoxide dehydrogenase complex to a membrane-bound hydrogenase in acetate-grown Methanosarcina thermophila. *J. Biol. Chem.* 263: 4075-4079.

Terlesky, K.C. & J.G. Ferry, (1988b) Purification and characterization of a ferredoxin from acetate-grown Methanosarcina thermophila. *J. Biol. Chem.* 263: 4080-4082.

Werner, F., (2007) Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* 65: 1395-1404.

Werner, F., (2008) Structural evolution of multisubunit RNA polymerases. *Trends Microbiol* 16: 247-250.

Zinder, S., (1993) Physiological ecology of methanogens. In: Methanogenesis. J.G. Ferry (ed). New York, NY: Chapman and Hall, pp. 128-206.

Figures and Tables

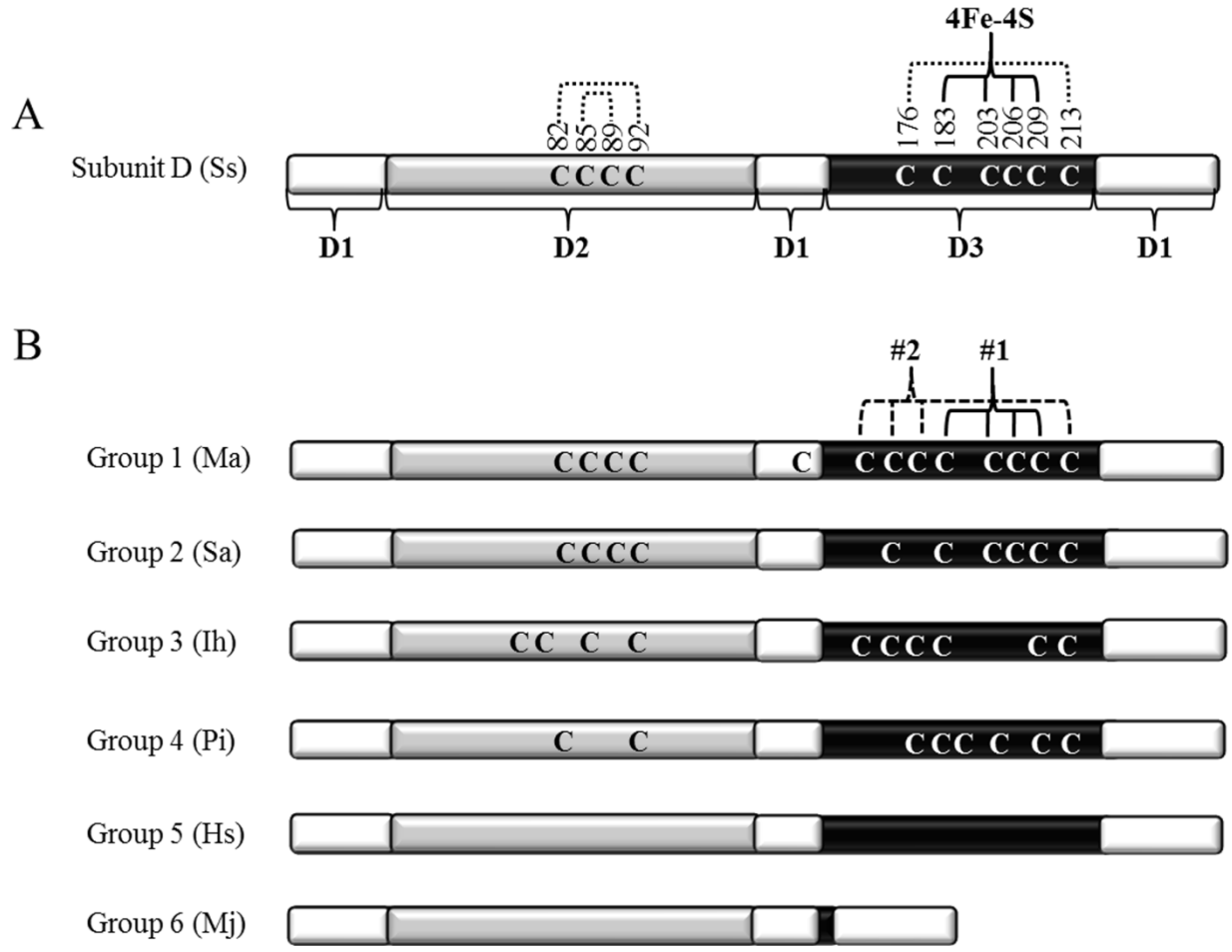

Figure 1. Domain architecture and diversity of cysteine residue content of subunit D of archaeal RNAP. A) Domain architecture and cysteine content of subunit D from *Sulfolobus solfataricus* (Ss) (4). The regions comprising domains 1, 2 and3 are indicated and shaded differently. The cysteine residues that function as ligands to a [4Fe-4S] cluster are indicated by 4Fe-4S. Cysteine residues documented to participate in disulfide bonds are indicated by dotted lines. B) Diagram depicting the six distinct groups of archaeal subunit D based on the number of complete or partial [4Fe-4S] cluster motifs within domain 3. The domain architecture is the same as depicted in panel A. The two [4Fe-4S] cluster motifs in group 1 subunit D are labeled #1 and #2, with #1 corresponding to the single [4Fe-4S] cluster motif in *S. solfataricus* Subunit D. Representative members of each group are shown: Ma, *Methanosarcina acetivorans;* Sa = *Sulfolobus acidocaldarius;* Ih = *Ignicoccus hospitalis;* Pi = *Pyrobaculum islandicum*; Hs = *Halobacterium salinarum*; Mr = *Methanocaldococcus jannaschii*.

Table 1. Archaeal RNAP subunit D classification based on the number of cysteine residues comprising partial or complete [4Fe-S] cluster binding motifs in domain 3.

			Domain Cysteine residue content					
Phylum or Order	Organism	*rpoD* gene ID	D2	D1	D3 cluster 1	D3 cluster 2	[4Fe-4S] clusters	Subunit D Group
Desulfurococcales	*Aeropyrum pernix* K1	1446211	2	0	2	2	none	4
	Hyperthermus butylicus DSM 5456	4782414	2	0	2	2	none	4
	Pyrolobus fumarii 1A	11138404	2	0	2	2	none	4
	Desulfurococcus mucosus DSM 2162	10154080	2	0	4	4	#1, #2	1
	Desulfurococcus kamchatkensis 1221n	7171405	2	0	4	4	#1, #2	1
	Ignisphaera aggregans DSM 17230	9715509	4	0	4	4	#1, #2	1
	Staphylothermus hellenicus DSM 12710	9234945	2	0	4	4	#1, #2	1
	Staphylothermus marinus F1	4907289	2	0	4	4	#1, #2	1
	Thermosphaera aggregans DSM 11486	9166443	2	0	4	4	#1, #2	1
	Ignicoccus hospitalis KIN4/I	5562263	2	0	2	4	#2	3
Sulfolobales	*Acidianus hospitalis* W1	10600112	4	0	4	2	#1	2
	Metallosphaera cuprina Ar-4	10492339	4	0	4	2	#1	2
	Metallosphaera sedula DSM 5348	5104884	4	0	4	2	#1	2
	Sulfolobus acidocaldarius DSM 639	3473859	4	0	4	2	#1	2
	Sulfolobus islandicus L.S.2.15	7798192	4	0	4	2	#1	2
	Sulfolobus solfataricus	1455324	4	0	4	2	#1	2
	Sulfolobus tokodaii str. 7	1460132	4	0	4	2	#1	2
Thermoproteales	*Caldivirga maquilingensis* IC-167	5709429	2	0	3	3	none	4
	Pyrobaculum aerophilum str. IM2	1465165	2	0	3	3	none	4
	Pyrobaculum arsenaticum DSM 13514	5054812	2	0	3	3	none	4
	Pyrobaculum calidifontis JCM 11548	4910199	2	0	3	3	none	4
	Pyrobaculum islandicum DSM 4184	4616570	2	0	3	3	none	4
	Thermoproteus neutrophilus V24Sta	6164675	2	0	3	3	none	4
	Thermoproteus uzoniensis 768-20	10361352	2	0	3	3	none	4
	Vulcanisaeta distributa DSM 14429	9751000	2	0	3	3	none	4
	Vulcanisaeta moutnovskia 768-28	10288618	2	0	3	3	none	4
	Thermofilum pendens Hrk 5	4602090	0	0	0	4	#1	2
Archaeoglobales	*Archaeoglobus fulgidus* DSM 4304	1485514	4	1	4	4	#1,#2	1
	Archaeoglobus profundus DSM 5631	8739703	4	1	4	4	#1,#2	1
	Archaeoglobus veneficus SNP6	10393458	4	1	4	4	#1,#2	1
	Ferroglobus placidus DSM 10642	8778351	4	1	4	4	#1,#2	1
Halobacteriales	*Haladaptatus paucihalophilus* DX253		0	0	0	0	none	5
	Halalkalicoccus jeotgali B3	9419552	0	0	0	0	none	5
	Haloarcula hispanica ATCC 33960	11049700	0	0	0	0	none	5
	Haloarcula marismortui ATCC 43049	3129579	0	0	0	0	none	5
	Halobacterium salinarum R1	5954259	0	0	0	0	none	5
	Haloferax volcanii DS2	8926874	0	0	0	0	none	5
	Halogeometricum borinquense DSM 11551	9993073	0	0	0	0	none	5
	Halomicrobium mukohataei DSM 12286	8412139	0	0	0	0	none	5
	Haloquadratum walsbyi DSM 16790	4194635	0	0	0	0	none	5
	Halopiger xanaduensis SH-6	10799252	0	0	0	0	none	5
	Halorhabdus utahensis DSM 12940	8384824	0	0	0	0	none	5
	Halorubrum lacusprofundi ATCC 49239	7399695	0	0	0	0	none	5
	Haloterrigena turkmenica DSM 5511	8743163	0	0	0	0	none	5
	Natrialba magadii ATCC 43099	8823246	0	0	0	0	none	5
	Natronomonas pharaonis DSM 2160	3703154	0	0	0	0	none	5
Methanobacteriales	*Methanobacterium* sp. AL-21	10277257	4	1	4	4	#1,#2	1
	Methanobrevibacter ruminantium M1	8770560	3	1	4	0	#1	2
	Methanobrevibacter smithii ATCC 35061	5217217	0	1	4	1	#1	2
	Methanosphaera stadtmanae DSM 3091	3854742	4	1	4	0	#1	2
	Methanothermobacter marburgensis str. Marburg	9704233	4	1	4	4	#1,#2	1
	Methanothermobacter thermautotrophicus str. Delta H	1469999	4	1	4	4	#1,#2	1
	Methanothermus fervidus DSM 2088	9962622	4	1	4	4	#1,#2	1

Table 1. Cont.

			Domain Cysteine residue content					
Phylum or Order	Organism	*rpoD* gene ID	D2	D1	D3 cluster 1	D3 cluster 2	[4Fe-4S] clusters	Subunit D Group
Methanococcales	*Methanocaldococcus fervens* AG86	8365160	1	0	0	0	no domain 3	6
	Methanocaldococcus infernus ME	9131706	1	0	0	0	no domain 3	6
	Methanocaldococcus jannaschii DSM 2661	1451040	1	0	0	0	no domain 3	6
	Methanocaldococcus sp. FS406-22	8803920	1	0	0	0	no domain 3	6
	Methanocaldococcus vulcanius M7	8513993	1	0	0	0	no domain 3	6
	Methanococcus aeolicus Nankai-3	5326605	1	0	0	0	no domain 3	6
	Methanococcus maripaludis S2	2762136	1	0	0	0	no domain 3	6
	Methanococcus vannielii SB	5324690	1	0	0	0	no domain 3	6
	Methanococcus voltae A3	9275437	1	0	0	0	no domain 3	6
	Methanothermococcus okinawensis IH1	10773159	1	0	0	0	no domain 3	6
	Methanotorris igneus Kol 5	10643705	1	0	0	0	no domain 3	6
Methanomicrobiales	*Methanocorpusculum labreanum* Z	4794628	4	1	4	4	#1,#2	1
	Methanoculleus marisnigri JR1	4846177	4	1	4	4	#1,#2	1
	Methanoplanus petrolearius DSM 11571	9743205	4	1	4	4	#1,#2	1
	Methanoregula boonei 6A8	5410157	4	1	4	4	#1,#2	1
	Methanosphaerula palustris E1-9c	7272107	4	1	4	4	#1,#2	1
	Methanospirillum hungatei JF-1	3924460	4	1	4	4	#1,#2	1
Methanopyrales	*Methanopyrus kandleri* AV19	1478069	4	0	2	0	none	4
Methanosarcinales	*Methanococcoides burtonii* DSM 6242	3996903	4	1	4	4	#1,#2	1
	Methanohalobium evestigatum Z-7303	9346454	4	1	4	4	#1,#2	1
	Methanohalophilus mahii DSM 5219	8983411	4	1	4	4	#1,#2	1
	Methanosaeta concilii GP-6	10460370	4	1	4	4	#1,#2	1
	Methanosaeta thermophila PT	4462198	4	1	4	4	#1,#2	1
	Methanosarcina acetivorans C2A	1473000	4	1	4	4	#1,#2	1
	Methanosarcina barkeri str. Fusaro	3627569	4	1	4	4	#1,#2	1
	Methanosarcina mazei Go1	1480500	4	1	4	4	#1,#2	1
Thermococcales	*Pyrococcus abyssi* GE5	1495432	0	0	0	0	none	5
	Pyrococcus furiosus DSM 3638	1469524	0	0	0	0	none	5
	Pyrococcus horikoshii OT3	1442487	0	0	0	0	none	5
	Pyrococcus sp. NA2	10553683	0	0	0	0	none	5
	Pyrococcus yayanosii CH1	10836830	0	0	0	0	none	5
	Thermococcus barophilus MP	10040449	0	0	0	0	none	5
	Thermococcus gammatolerans EJ3	7987168	0	0	0	0	none	5
	Thermococcus kodakarensis KOD1	3234626	0	0	0	0	none	5
	Thermococcus onnurineus NA1	7017753	0	0	0	0	none	5
	Thermococcus sibiricus MM 739	8095316	0	0	0	0	none	5
	Thermococcus sp. AM4	7419843	0	0	0	0	none	5
Thermoplasmales	*Ferroplasma acidarmanus* fer1	ZP_05571044	4	0	2	0	none	4
	Picrophilus torridus DSM 9790	2844373	4	0	2	0	none	4
	Thermoplasma acidophilum DSM 1728	1456551	4	0	2	0	none	4
	Thermoplasma volcanium GSS1	1441081	4	0	2	0	none	4
Nanoarchaeota	*Nanoarchaeum equitans* Kin4-M	2654414	0	0	0	0	none	5
Korarchaeota	Candidatus *Korarchaeum cryptofilum* OPF8	Kcr_1582	0	0	2	4	#2	3

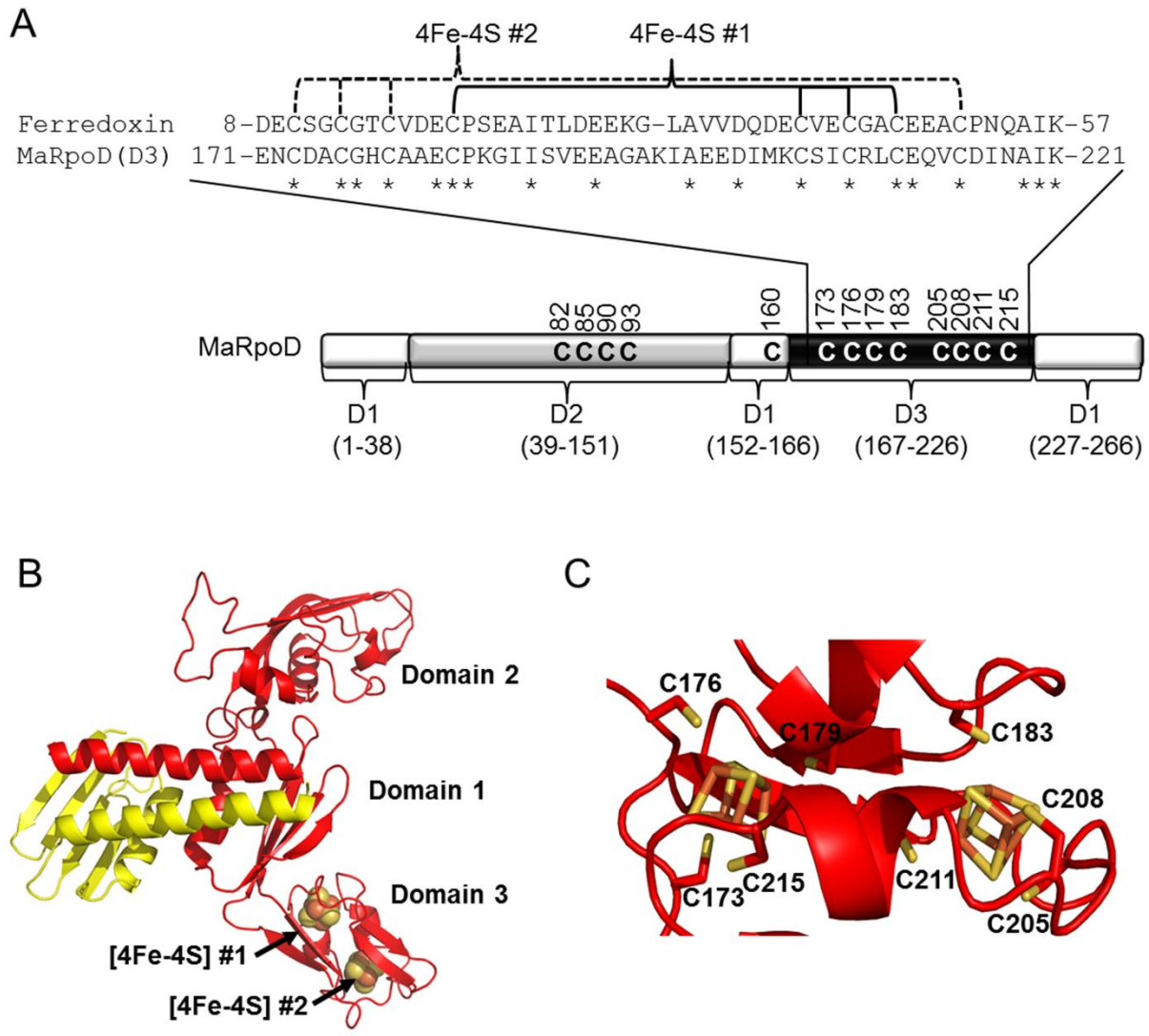

Figure 2. Domain 3 of subunit D of RNAP from *M. acetivorans* is predicted to bind two [4Fe-4S] clusters similar to 2[4Fe-4S] cluster ferredoxin. A) Schematic of the domain architecture of *M. acetivorans* subunit D including an amino acid alignment of Domain 3 to 2[4Fe-4S] cluster ferredoxin (MA0431) from *M. acetivorans*. Conserved residues are indicated by an asterisk including the cysteine residues postulated to bind the two [4Fe-4S] clusters. B) Homology model of the heterodimer of *M. acetivorans* subunit D (red) and subunit L (yellow) D/L heterodimer. The two [4Fe-4S] clusters in domain 3 are represented by sphere model. C) Close-up view of domain 3 of subunit D showing putative cysteine ligands to each [4Fe-4S] cluster. The two [4Fe-4S] clusters in domain 3 are represented in stick model.

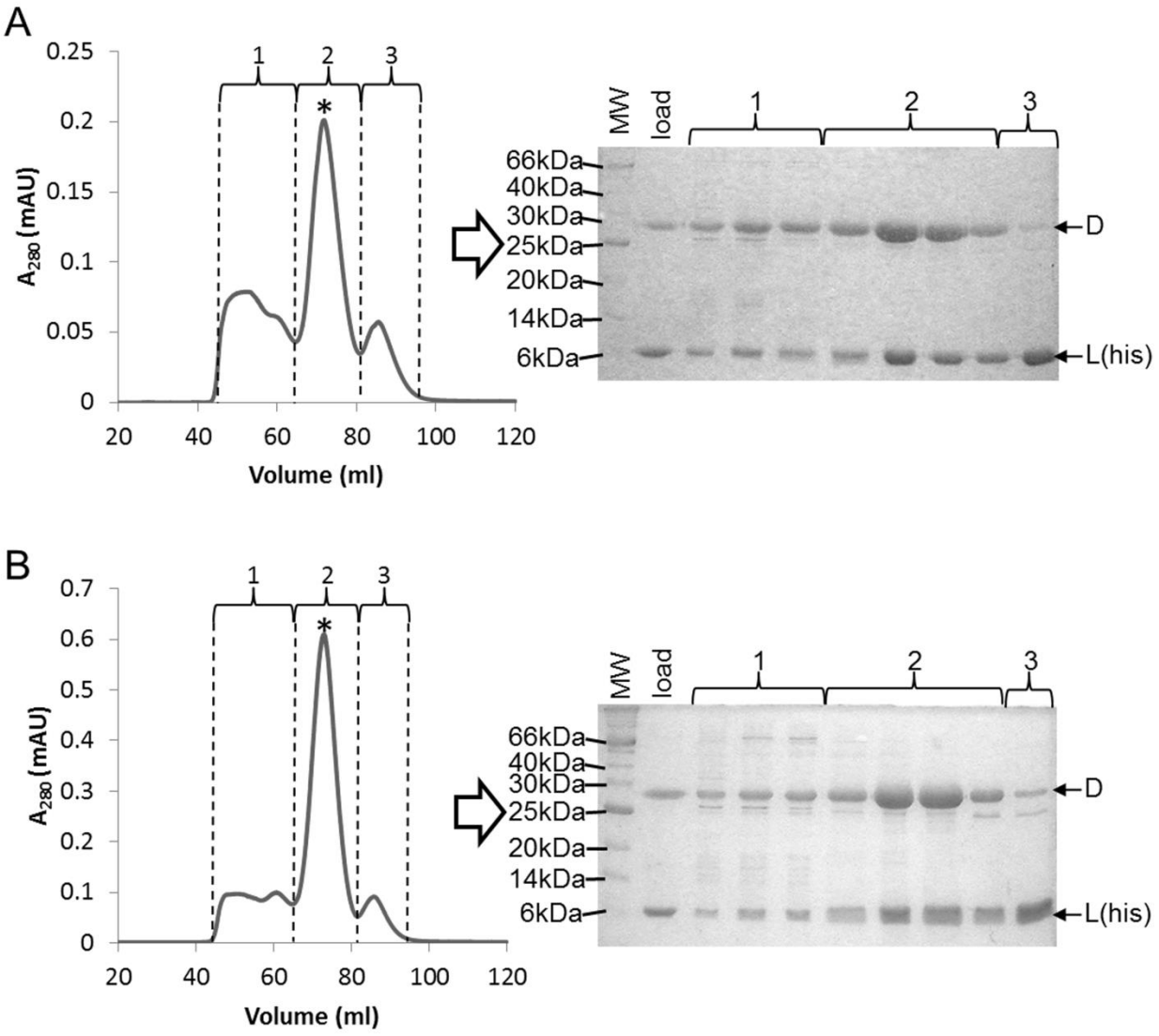

Figure 3. Purification of recombinant *M. acetivorans* D/L(his) and D/L(his)-FeS heterodimers. A) Anaerobic size-exclusion chromatography of eluate (~50 mg of total protein) from Ni^{2+}-affinity chromatography of *E. coli* cell lysate containing *M. acetivorans* subunit L(his) and subunit D. B) Same as panel A, except that the eluate was reconstituted with iron and sulfide prior to loading onto the size-exclusion column (see Materials and Methods for details). For each purification, samples from fractions containing peaks 1, 2, and 3 were analyzed by SDS-PAGE as indicated. The asterisk indicates the peak that is consistent with a the molecular weight of a D/L(his) heterodimer (40.2 kDa).

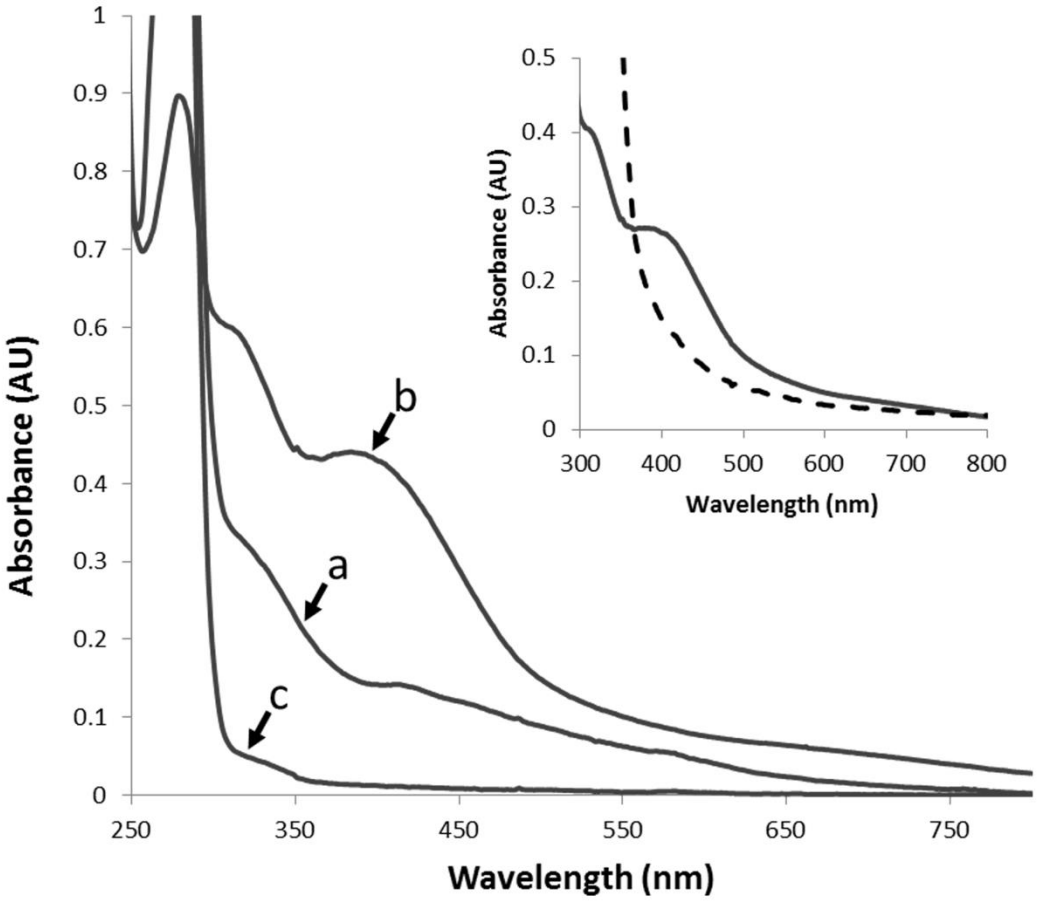

Figure 4. UV-visible spectra of purified D/L(his), D/L(his)-FeS, and DΔD3/L(his) heterodimers. Line a, D/L(his) (67.5 µM); line b, D/L(his)-FeS (13.5 µM); line c, DΔD3/L(his) (67.5 µM). The inset shows the D/L(his)-FeS heterodimer before (solid line) and after (dashed line) the addition of sodium dithionite (10 mM).

Table 2. Comparison of the properties of the *M. acetivorans* D/L(His), D/L(His)-FeS, and DΔD3/L(His) heterodimers.

Heterodimer	A_{390}/A_{280}	ε_{390}[a]	Iron[b]	Sulfide[c]	Thiol[d]	Thiol[e], denatured
D/L-His	0.06	2.2 ± 0.08	1.1 ± 0.1	1.1 ± 0.05	13.8 ± 0.4	14.5 ± 0.7
D/L-His-FeS	0.50	33.2 ± 0.6	8.3 ± 0.2	7.7 ± 0.8	16.2 ± 0.3	21.5 ± 1.6
DΔD3/ L-His	0.01	0.2 ± 0.01	ND	ND	5.9 ± 0.7	5.4 ± 1.5

[a] mM^{-1} cm^{-1}
[b] nmol iron/nmol of DL heterodimer
[c] nmol acid-labile sulfide/nmol of DL heterodimer
[d] nmol thiols/nmol of DL heterodimer
[e] measured in buffer containing 6M guanidine-HCl
ND: not determined

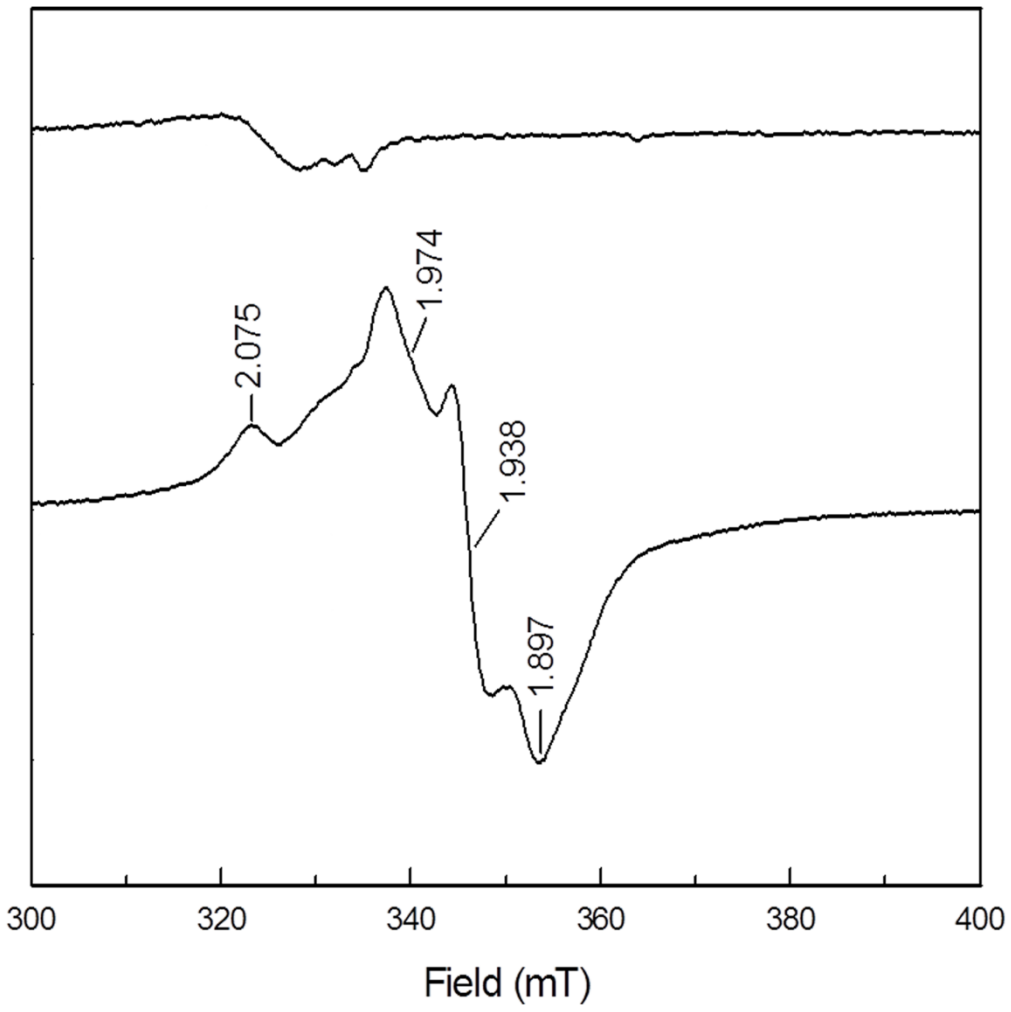

Figure 5. EPR spectra of the D/L(his)-FeS heterodimer before (a) and after the addition of dithionite (b). The sample contained 150 µM D/L(his)-FeS in 50 mM HEPES, 50 mM NaCl, pH 7.5. EPR conditions: microwave frequency, 9.385 GHz; microwave power incident to the cavity, 2.0 mW; field modulation frequency, 100 kHz; microwave amplitude, 0.6 mT; temperature 8 K.

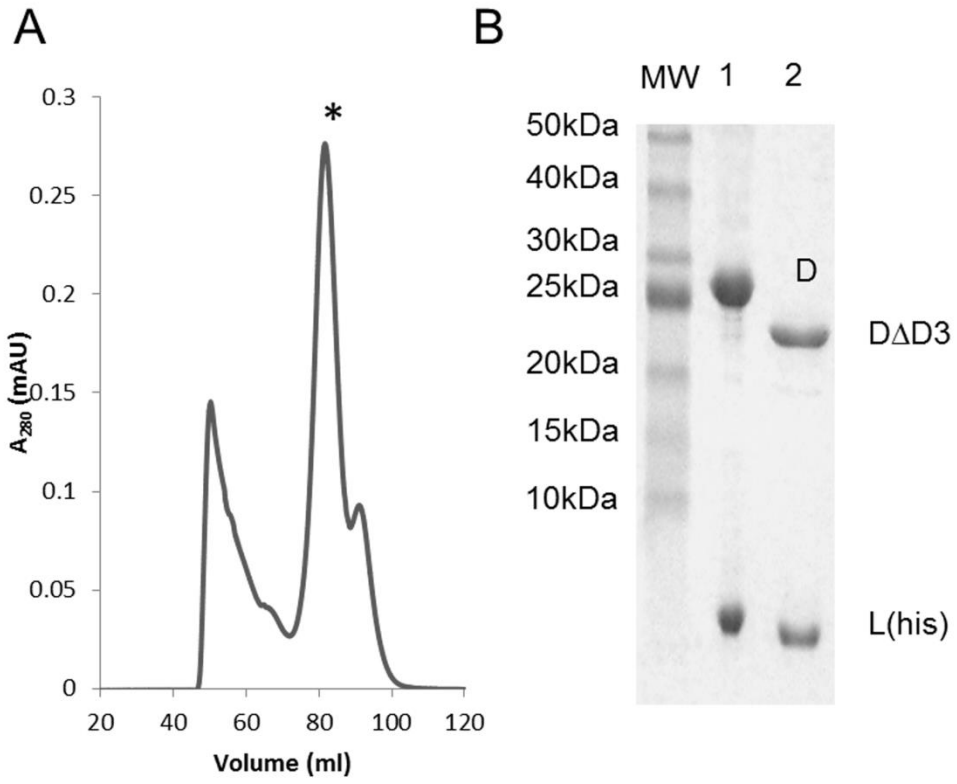

Figure 6. Co-purification of *M. acetivorans* subunit DΔD3 with subunit L(his) after co-expression in *E. coli*. A) Anaerobic size-exclusion chromatography of eluate (~50 mg of total protein) from Ni^{2+}-affinity chromatography of *E. coli* cell lysate containing *M. acetivorans* subunits L(his) and DΔD3. The asterisk indicates the peak which is consistent with a the molecular weight of a heterodimer of DΔD3/L(his) (34.8 kDa). B) SDS-PAGE analysis of the purified D/L(his) (lane 1) and DΔD3/L(his) (lane 2) heterodimers, confirming the smaller molecular weight of DΔD3 (23.3 kDa) compared to D (28.7 kDa).

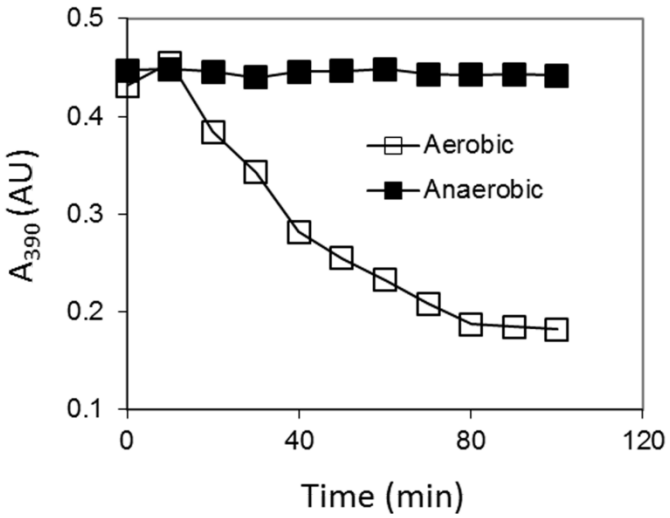

Figure 7. Effect of oxygen on the stability of the [4Fe-4S] clusters in purified D/L(his)-FeS heterodimer. The loss of the [4Fe-4S] clusters in purified D/L(his)-FeS (20 µM) incubated under N_2 or in air in 50 mM HEPES, 50 mM NaCl, pH 7.5 was monitored by measuring A_{390} over time.

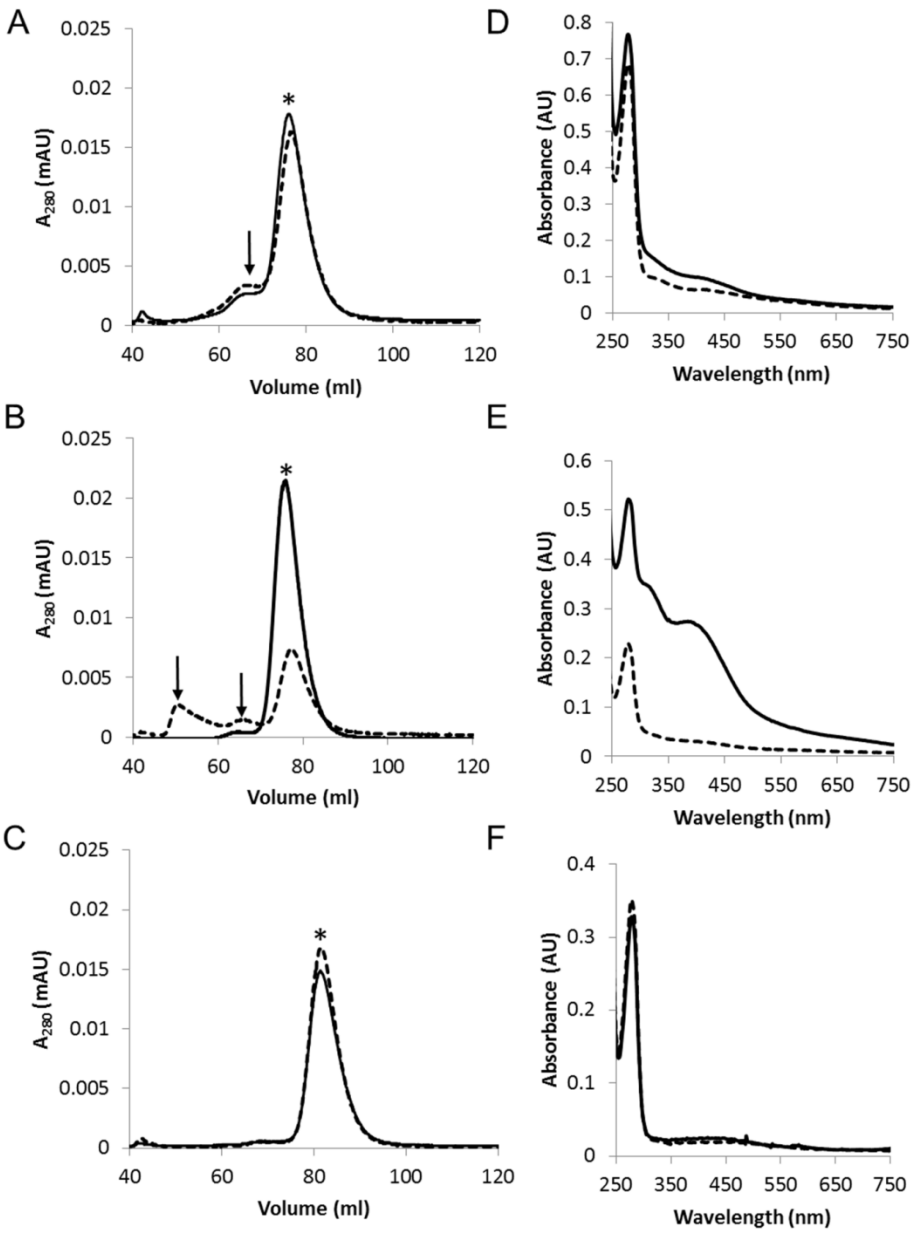

Figure 8. Effect of oxygen on the stability of the D/L(his), D/L(his)-FeS, and DΔD3/L(his) heterodimers. Each heterodimer was incubated anaerobically or aerobically in 50 mM HEPES, 50 mM NaCl, pH 7.5 and analyzed by size-exclusion chromatography under anaerobic conditions using a running buffer of 50 mM HEPES, pH 7.5, 150 mM NaCl. Elution profiles of each heterodimer after anaerobic (solid line) or aerobic incubation (dashed line) are shown in panels A-C. Panel A: D/L(his), 1.1 mg each; panel B: D/L(his)-FeS, 0.8 mg each; panel C: DΔD3/L(his), 1.5 mg each. For each elution profile an asterisk indicates the heterodimer peak and arrows indicate D/L(his) aggregates. The UV-visible spectra of concentrated fractions containing heterodimer from anaerobic samples (solid line) and aerobic samples (dashed line) are shown in panels D-F. Panel D: D/L(his), 26 µM each; panel E: D/L(his)-FeS, 10 µM each; panel F: DΔD3/L(his),18 µM each.

Table 3. Iron and sulfide content of purified recombinant D/His-L heterodimers.

D/L heterodimer	Iron[a]	Sulfide[b]
DΔFeS1/L-His	4.6 ± 0.1	3.3 ± 0.2
DmFeS1/L-His	4.7 ± 0.2	3.5 ± 0.7
DΔFeS2/L-His	3.8 ± 0.1	3.0 ± 0.1
DmFeS2/L-His	4.9 ± 0.2	3.3 ± 0.2

[a]nmol of iron/nmol of D-L heterodimer
[b]nmol of sulfide/nmol of D-L heterodimer

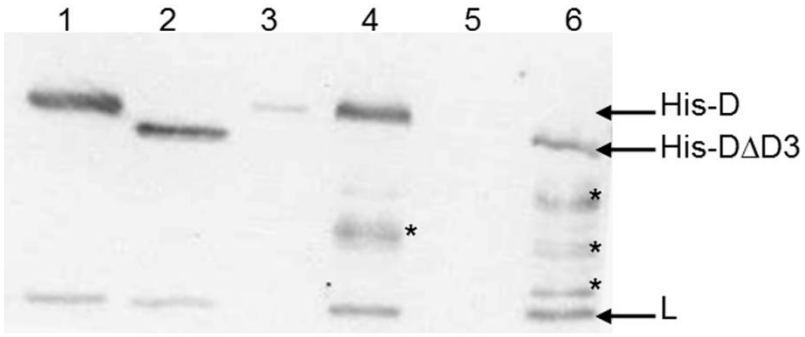

Figure 9. Co-purification of endogenous subunit L with His-tagged subunit D or subunit DΔD3 expressed within *M. acetivorans*. Western blot of a 15% SDS-PAGE gel with anti-D/L antibodies. Lane 1, recombinant D/L(his) heterodimer (50 ng); lane 2, recombinant DΔD3/L(his) heterodimer (150 ng); lane 3, imidazole eulate from a Ni^{2+}-agarose column loaded with cell lysate of DJL30 grown in the absence of tetracycline; lane 4, imidazole eluate from a Ni^{2+}-agarose column loaded with cell lysate of DJL30 grown in the presence of tetracycline; lane 5, imidazole eulate from a Ni^{2+}-agarose column loaded with cell lysate of DJL31 grown in the absence of tetracycline; lane 6, imidazole eulate from a Ni^{2+}-agarose column loaded with cell lysate of DJL31 grown in the presence of tetracycline. The asterisks indicate subunit His-D degradation products.

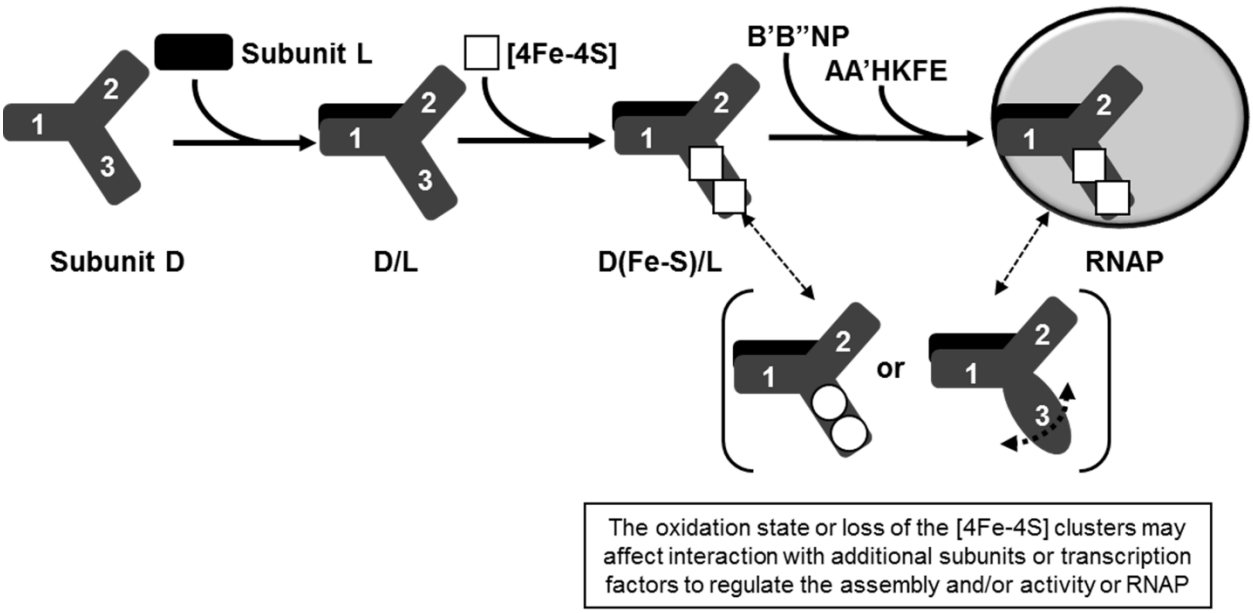

Figure 10. Model depicting the potential roles of the subunit D [4Fe-4S] clusters in modulating the assembly and/or activity of *M. acetivorans* RNAP. The three domains of subunit D are represented by 1, 2, and 3. A change in the reduction state of each [4Fe-4S] cluster is indicated by a circle versus a square. The dashed semi-circle arrow depicts the increased flexibility of domain 3 upon loss of the [4Fe-4S] clusters.

Chapter II

The [4Fe-4S] clusters of subunit D are key determinants in the post D-L heterodimer assembly of RNA polymerase in *Methanosarcina acetivorans*

Matthew E. Jennings

Department of Biological Sciences, University of Arkansas,

Fayetteville, AR 72701 USA

Abstract

Subunits D and Rpb3/AC40 of RNA polymerase (RNAP) from many archaea and some eukaryotes, respectively, contain a ferredoxin-like domain (FLD) predicted to bind one or two [4Fe-4S] clusters postulated to play a role in regulating the assembly of RNAP. To test this hypothesis, the two [4Fe-4S] cluster subunit D from *Methanosarcina acetivorans* was modified to generate variants that lack the FLD or each [4Fe-4S] cluster. Viability of gene replacement mutants revealed neither the FLD, nor either [4Fe-4S] cluster is essential. Nevertheless, each mutant demonstrated impaired growth due to significantly lower RNAP activity when compared to wild type. Affinity purification of tagged subunit D variants from *M. acetivorans* strains revealed that neither the FLD, nor each [4Fe-4S] cluster is required for the formation of a D-L heterodimer, the first step in the assembly of RNAP. However, the association of the D-L heterodimer with catalytic subunits B' and A" was diminished by removal of the FLD and each cluster, with the loss of cluster 1 having a more substantial effect than the loss of cluster 2. These results reveal that the FLD and [4Fe-4S] clusters, particularly cluster 1, are key determinants in the post D-L heterodimer assembly of RNAP in *M. acetivorans*.

Introduction

Iron-sulfur (Fe-S) clusters are protein cofactors that serve diverse functions, including catalytic, sensing, structural, and electron transfer. The most common function of Fe-S clusters is to serve as electron carriers during oxidation-reduction reactions catalyzed by metabolic enzymes, including those found in electron transport systems (Johnson *et al.*, 2005). Fe-S clusters are an ancient prosthetic group that likely served as the primary electron carriers in the anaerobes that were the dominant life form on Earth for approximately 2.5 billion years, when the atmosphere was largely devoid of oxygen (Imlay, 2006). The reliance of strict anaerobes on Fe-S clusters is supported by genomic evidence, which has revealed that the genomes of extant strict anaerobes encode significantly more Fe-S proteins, containing primarily [4Fe-4S] clusters, than the genomes of extant aerobes (Major *et al.*, 2004, Sousa *et al.*, 2013). This disparity is likely due to the fact that Fe-S clusters are typically oxygen-labile. Nonetheless, Fe-S clusters serve critical roles in the vast majority of anaerobes and aerobes.

Fe-S clusters are also found in proteins that function in replication, transcription, and translation, processes where an electron carrier cofactor would not likely be necessary. The incorporation of Fe-S clusters into proteins involved in the central dogma may provide mechanisms to correlate information processing systems with energy conserving processes requiring Fe-S clusters within cells. For example, many of the enzymes that are involved in DNA replication in eukaryotes and archaea, including replicative polymerases, helicase, and primase, contain [4Fe-4S] clusters (Fuss *et al.*, 2015). Often the [4Fe-4S] clusters serve as a structural determinant required for the activity of enzymes, such as in primase (Klinge *et al.*, 2007). The [4Fe-4S] cluster is essential in the B-family of DNA polymerases in yeast, where it is required for subunit interaction and polymerase complex stabilization (Netz *et al.*, 2012). The crystal structure of RNA polymerase (RNAP) from the archaeon *Sulfolobus solfataricus* revealed a

[4Fe-4S] cluster coordinated by cysteine residues within domain 3 (D3) of subunit D (Hirata *et al.*, 2008). Many strictly anaerobic archaea possess a subunit D with a D3 containing two [4Fe-4S] cluster binding motifs, which is homologous to 2[4Fe-4S] cluster ferredoxin, a ubiquitous electron carrier protein (Rodriguez-Monge *et al.*, 1998). Thus, D3 can also be referred to as the ferredoxin-like domain (FLD). The D subunit of RNAP from numerous archaea with diverse physiology, contain a D3/FLD with one or two [4Fe-4S] cluster binding motifs (Lessner *et al.*, 2012). Moreover, the corresponding subunit in eukaryotic RNAPs (Rpb3/AC40) also contains D3 and several eukaryotes possess Rpb3/AC40 subunits with a D3/FLD predicted to bind a [4Fe-4S] cluster (Hirata *et al.*, 2008, Hirata & Murakami, 2009). The conservation of [4Fe-4S] clusters in RNAPs from two domains of life across multiple taxa indicates that the clusters serve an important, but likely not essential, role in RNAP. The precise role the [4Fe-4S] cluster(s) serve in RNAP is unknown.

RNAP is a conserved multi-subunit complex found in all three domains of life. Archaea and bacteria possess a single RNAP responsible for synthesizing all RNA, whereas eukaryotes possess between three to five types of RNAP that synthesize different subsets of RNA (Werner & Grohmann, 2011). Structurally, archaeal RNAP is most similar to eukaryotic RNAPII; each comprised of 12-13 subunits, including several subunits not present in bacterial RNAP, which is comprised of 5 subunits (Decker & Hinton, 2013). The α subunit of bacterial RNAP is homologous to the D and Rpb3/AC40 subunits of archaeal and eukaryotic RNAP, respectively. However, all α subunits from sequenced bacteria lack D3, comprising the FLD found in numerous D and Rpb3/AC40 subunits. The structural similarity of archaeal and eukaryotic RNAP, along with the shared presence of D3/FLD, support a shared ancestry, one that incorporated the use of [4Fe-4S] clusters in RNAP in certain lineages.

Subunit D is located a substantial distance from the active site of archaeal RNAP, indicating the [4Fe-4S] cluster(s) likely do not directly participate in catalysis (Hirata *et al.*, 2008). The [4Fe-4S] cluster(s) more likely serve a structural or regulatory function. The homologous D/Rpb3/AC40/α subunits are each involved in the first step in the assembly of RNAP. In archaea, subunit D forms a heterodimer with subunit L, which initiates the assembly of RNAP (Eloranta *et al.*, 1998, Goede *et al.*, 2006). All other RNAP subunits sequentially assemble on the D-L heterodimer. It was initially postulated that the [4Fe-4S] cluster(s) are required for the ability of subunit D to from a heterodimer with subunit L. This hypothesis was supported by recombinant studies with *S. solfataricus* RNAP. *S. solfataricus* subunit D contains an oxygen-stable [4Fe-4S] cluster, and the removal of [4Fe-4S] cluster binding impaired the ability to form a recombinant heterodimer with subunit L (Hirata *et al.*, 2008). Unlike *S. solfataricus*, which is an aerobe, many strict anaerobes, such as methanogens, contain a D3/FLD predicted to bind two [4Fe-4S] clusters. Previous studies with the methanogen *Methanosarcina acetivorans* demonstrated that subunit D is capable of binding two oxygen-labile [4Fe-4S] clusters that are not required for D-L heterodimer formation, but the clusters affect the stability of the heterodimer, and thus may influence assembly of RNAP after the formation of the D-L heterodimer (Lessner *et al.*, 2012). In either case, the [4Fe-4S] cluster(s) may serve as a structural determinant that is required for the optimal assembly of RNAP.

Herein, we have used the *M. acetivorans* genetic system to investigate the importance of the [4Fe-4S] clusters to specific steps in the *in vivo* assembly of RNAP, and to delineate the significance of each cluster to RNAP assembly and activity. A combination of genetic and biochemical experiments revealed that neither the FLD, nor the ability of the FLD to bind either [4Fe-4S] cluster was required for subunit D to form a heterodimer with subunit L, indicating the

first step of RNAP assembly in *M. acetivorans* is not influenced by the FLD or the presence of the [4Fe-4S] clusters. Moreover, the FLD and the ability to bind either [4Fe-4S] cluster 1 or 2 are not essential to RNAP in *M. acetivorans*. However, our results demonstrate that the FLD and the ability to bind the [4Fe-S] clusters, are important for the interaction of the D-L heterodimer with catalytic subunits of RNAP, and therefore, are important for the *in vivo* assembly of RNAP post D-L heterodimer formation. In particular, the inability of subunit D to bind [4Fe-4S] cluster 1 had a more substantial impact, compared to the loss of [4Fe-4S] cluster 2, on the assembly of RNAP in *M. acetivorans*. The greater importance of [4Fe-4S] cluster 1 in subunit D from *M. acetivorans* provides support for the conservation of the analogous cluster in D/Rpb3/AC40 subunits from numerous archaea and eukaryotes, including *S. solfataricus* subunit D.

Materials and Methods

Classification of archaeal *rpoD* genes. The *M. acetivorans* subunit D protein sequence (P0CG28.1) was used in a BLASTP search of archaeal sequences in the non-redundant NCBI database. The returned sequences were screened and duplicates, or those not annotated as an RNAP subunit, were removed. The remaining sequences were aligned using the MEGA program (v 6.06) (Tamura *et al.*, 2013) to identify those sequences that contain a domain similar to domain 3, defined as aligning with residues 171-221 of the *M. acetivorans* sequence. Species were classified into six groups based on the presence of D3/FLD and arrangement of the cysteine residues comprising the predicted [4Fe-4S] cluster binding motifs within the domain as previously described (Lessner *et al.*, 2012).

Generation of subunit D variants and recombinant protein analyses. A complete list of the plasmids and primers used in this study is included in **Table S1** and **Table S2**, respectively. The

construction of *M. acetivorans rpoD* genes harboring mutations in domain 3 was done in pRpoDL, which is used to for the co-expression of subunit D and C-terminally His-tagged subunit L in *E. coli* (Lessner *et al.*, 2012). Briefly, following the manufacturer's instructions, the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies) was used with pRpoDL as the template and primers were designed using the QuikChange Primer Design Program to generate variants of subunit D (**Fig. 1**) encoded in pRpoDL. Specifically, the *rpoD*ΔFeS1 and *rpoD*ΔFeS2 mutations were generated by deleting nucleotides 594-636 (cluster 1; primers QCRpoDΔFeS1For and QCRpoDΔFeS1Rev) and nucleotides 509-539 (cluster 2; primers QCRpoDΔFeS2For and QCRpoDΔFeS2Rev), respectively. Similarly, the *rpoD*mFeS1 and *rpoD*mFeS2 were generated by changing nucleotides 614 and 623 from G to C (cluster 1; primers QCRpoDmFeS1For and QCRpoDmFeS1Rev) and nucleotides 518, 527, and 536 from G to C (cluster 2; primers QCRpoDmFeS2For and QCRpoDmFeS2Rev), respectively. To generate the *rpoD*ΔCterm mutation, primers RpoDΔCtermFor and RpoDΔCtermRev were 5'-phosphorylated and used to amplify the entire pRpoDL plasmid minus nucleotides 679 – 783 of the *rpoD* sequence. The PCR product was then blunt-ligated to generate pDL414. All of the constructs were confirmed by sequencing. Generation of pDL408 containing *rpoD* with the entire domain 3 deleted (*rpoD*ΔD3) has been previously described (Lessner *et al.*, 2012).

Each subunit D variant was co-expressed with subunit L(His) in *E. coli* Rosetta (DE3) (pLacI) and D-L(His) heterodimers were purified anaerobically by Ni^{2+}-affinity chromatography and size-exclusion chromatography as previously described (Lessner *et al.*, 2012). Fe-S clusters were reconstituted into purified D-L(His) heterodimers as described (Cruz & Ferry, 2006).

Generation of merodiploid strains of *M. acetivorans* capable of inducible expression of (His)D. Primers were designed to amplify wild-type and mutated *rpoD* genes from each respective derivative of pRpoDL (**Table S2**). The forward primer (HisRpoDNdeFor) contained an *Nde*I restriction site and an N-terminal His-tag, while the reverse primer (HisRpoDHindRev) contained a *Hind*III restriction site. PCR products were digested with *Nde*I and *Hind*III and ligated with similarly-digested pJK027A (Guss *et al.*, 2008). For the *rpoD*ΔCterm construct, an alternative reverse primer was required (RpoDΔCtermHindRev). All of the derivatives of pJK027A were confirmed by DNA sequencing. Each derivative of pJK027A was used to transform *M. acetivorans* strain WWM73 as described previously (Metcalf *et al.*, 1997). The successful integration of the plasmid into the chromosome of the parent strain was confirmed by PCR (Guss *et al.*, 2008). Each strain is capable of the tetracycline-dependent expression of a separate (His)D variant.

Generation of *M. acetivorans* mutants with *rpoD* replaced with *rpoD* encoding a (His)D variant. The generation of mutant strains of *M. acetivorans*, where native *rpoD* is replaced with *rpoD* encoding a (His)D variant, was attempted using homologous recombination with plasmids derived from pJK301 (Guss *et al.*, 2008). Given the conflicting *Nde*I restriction enzyme site in the pJK301 sequence and the sequence surrounding *rpoD* in the chromosome, the QuikChange mutagenesis kit was used to remove the *Nde*I site from pJK301 (CATATG to CACATG) to generate pDL517. Similarly, 4 kb of *M. acetivorans* genomic DNA, which included the *rpoD* gene and the upstream 3 kb sequence, was cloned into pUC19 to create pDL518. The upstream region contains two natural restriction sites which would interfere with downstream cloning, both of which occur in genes encoding 30S ribosomal proteins (MA1109 and MA110). The

QuikChange mutagenesis kit was used to introduce silent mutations into pDL518 which removed the restriction site from both genes but did not alter the encoded amino acid; resulting in the generation of pDL520. Primers QCRpoDUSXhoF and QCRpoDUSXhoR were used to remove the *Xho*I site from MA1109 (902 bp upstream of *rpoD* start site; CTCGAG to CTAGAG silent mutation for leucine codon) and primers QCRpoDUSApaF and QCRpoDUSApaR were used to remove the *Apa*I site from MA1110 (192 bp upstream of the *rpoD* start site; GGGCCC to GGACCC silent mutation for glycine codon).

Primers (RpoDUSApaFor and RpoDUSXhoRev) were used to amplify 2.8 kb of DNA encompassing *rpoD* at the 3' end and approximately 2 kb of upstream sequence at the 5' end, from pDL520. The primers also added *Apa*I and *Xho*I at the 5' and 3' ends, respectively, to the PCR product. The PCR product was digested with *Apa*I and *Xho*I and ligated with similarly-digested pDL517 to create pDL521. Similarly, the 2.5 kb sequence downstream and adjacent to *rpoD* was amplified from genomic DNA with primers (RpoDDSBamFor and RpoDDSNotRev) that add a *Bam*HI and *Not*I restriction enzyme site at the 5' and 3' ends, respectively. The PCR product was digested with *Bam*HI and *Not*I and ligated with similarly-digested pDL521 to generate pDL522. Plasmid pDL522 contains *rpoD* along with 2 kb upstream sequence and 2.5 kb sequence downstream of *rpoD.* Between the upstream and downstream regions is 2.1 kb of plasmid DNA including *pac*, which encodes resistance to puromycin. Thus, pDL522 and derivatives can be used to replace native *rpoD* with *rpoD* encoded in pDL522 by homologous recombination and selection with puromycin.

Derivatives of pDL522 encoding (His)D variants were generated by amplifying each mutant *rpoD* from each pJK027A-based plasmid (described above) using the primers (HisRpoDNdeFor and RpoDUSXhoRev), which added *Nde*I and *Xho*I restriction enzyme sites at

the 5’ and 3’ end of the PCR product, respectively. The PCR products were digested with *Nde*I and *Xho*I and ligated with similarly-digested pDL522 to create the derivatives of pDL522, each containing *rpoD* encoding a separate (His)D variant. All plasmids were verified by DNA sequencing.

M. acetivorans strain WWM73 was transformed with pDL522 or a derivative as previously described (Metcalf *et al.*, 1997). Multiple transformations were attempted with plasmids encoding (His)*rpoD* variants, and in each case a transformation utilizing pJK027A and pDL522 were included as positive controls for transformation efficiency and homologous recombination, to confirm that the lack of any observed transformants was not due to poor transformation and/or homologous recombination. Mutants were initially screened and identified by PCR, and each mutant confirmed by PCR amplification and sequencing of the entire *rpoD* gene.

Expression and Purification of (His)D and (His)D domain 3 variants in *M. acetivorans*. All of the following procedures were completed under anaerobic conditions in an anaerobic chamber (Coy Laboratories) containing 95% N_2 and 5% H_2 at 25 °C unless otherwise noted. *M. acetivorans* cells were grown in a stoppered flask at 37 °C in 1 L of high-salt (HS) medium supplemented with methanol (125 mM) and 0.025% sodium sulfide (Sowers *et al.*, 1993). For merodiploid strains, cultures were induced with 100 µg/mL tetracycline immediately upon inoculation. Cells were harvested at an optical density at 600 nm of 0.6 - 0.9, as described previously (Sowers *et al.*, 1984). Pelleted cells were harvested and resuspended in lysis buffer (10 mM Tris, 15% glycerol, 10 µM $ZnCl_2$, 2 M KCl, 30 mM $MgCl_2$). Phenylmethylsulfonyl

fluoride and benzamadine were added to the resuspended cells to a final concentration of 1 mM. Cell pellets were stored under nitrogen in sealed vials at -80 °C until use.

For the purification of (His)D, frozen cells were thawed on ice and lysed by sonication or by repetitive freeze/thaw cycles. DNase I was added to a final concentration of 4 μg/mL, and the lysate was incubated for one hour, followed by centrifugation for 15 minutes at 16,000 x *g*. The soluble fraction was filtered (0.45 μm pore size) and loaded onto a pre-equilibrated (lysis buffer plus 10 mM imidazole) Ni^{2+}-agarose resin column (600 μl). The flow-through from the column was collected and re-applied to the column a total of four times. The column was then washed with 12 ml buffer (10 mM Tris, 15% glycerol, 10 μM $ZnCl_2$, 0.1 M KCl, 10 mM imidazole), and bound protein was eluted from the column with the addition of 3 ml of elution buffer (20 mM Tris, 15% glycerol, 10 μM $ZnCl_2$, 30 mM $MgCl_2$, 250 mM imidazole). The eluates were aliquoted into sealed vials and stored under nitrogen at -80 °C or used immediately.

Non-specific transcription assays. Cell lysates and eluates were assayed for non-specific transcription activity using a radiolabeled assay similar to that previously described (Darcy *et al.*, 1999). Due to the observed loss of activity after freezing, all lysates and eluates were assayed prior to freezing. For lysate samples, *M. acetivorans* cells were grown and processed to obtain the soluble fraction as described above, except the cells were resuspended in elution buffer. A transcription assay mixture was freshly prepared by adding 2 μL ^{32}P-UTP (20 μCi) and 2 μL 0.5 M DTT (1 mM) for every ml of transcription buffer (20 mM Tris, 20 mM $MgCl_2$ 1 mM ATP, 1 mM GTP, 1 mM CTP, 100 μg/mL 90 bp oligonucleotide, pH 8.0) as needed. The oligonucleotide sequence was randomly generated using a random oligonucleotide generator with the GC content set to 22%. Two complimentary oligonucleotides were synthesized (IDT) and dissolved to a

concentration of 10 μg/μL in buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 7.5), and annealed by combining equal amounts of each and incubating at 95°C for 5 minutes and then cooled to 25 °C. Ten microliters of sample were mixed with the transcription assay mixture to a final volume of 100 μl and incubated at 35 °C for 30 min. The reaction mixture was then spotted onto cellulose filter discs (23 mm) and washed in succession with 0.5 M Na_2HPO_4 (7X) and water (3X). Filters were rinsed with 95% ethanol and air-dried for ten minutes before being placed in scintillation vials containing 10 ml scintillation fluid (National Diagnostic Ecoscint H). The vials were vortexed for 10 seconds, and radioactivity counts were measured for one minute (CPM) on a Packard 1600 TR Liquid Scintillation Analyzer.

Analytical Methods. Growth studies were performed with *M. acetivorans* strains in HS medium supplemented with 125 mM methanol and 0.025% sodium sulfide as described (Sowers *et al.*, 1993). Generation times were calculated from at least three replicate cultures. For recombinant proteins, concentrations were determined using the Bradford method (Bradford, 1976) and bovine serum albumin as a standard. The concentration of protein in imidazole eluates from (His)D purifications was determined using the Invitrogen Qubit 2.0 Fluorometer and the Qubit Protein Assay Kit as directed by the manufacturer. The iron and acid-labile sulfide content of D-L(His) heterodimers were determined as described previously (Cruz & Ferry, 2006). SDS-PAGE and Western blotting were performed by standard procedures, using anti-RpoDL antibody previously described (Lessner *et al.*, 2012). Antibodies used to specifically detect subunit D, subunit B', or subunit A" were supplied by Genscript and produced using synthesized peptides specific for each protein; subunit D (CISSDPKIQPADPNV), subunit B' (CGKTSPPRFLEEPSD) and subunit A" (CDGEVKQIGRHGISG). The intensity of bands in

Western blots was calculated using ImageJ (Schneider *et al.*, 2012). A standard curve to calculate the subunit D concentration in samples using Western blot band intensity was generated using purified recombinant subunit D.

Results

Conservation of [4Fe-4S] cluster binding motifs in subunit D of archaeal RNAP. Previously, RNAP D subunits (99) from the available sequenced archaeal genomes at the time were analyzed for the presence of D3/FLD and [4Fe-4S] cluster motifs, which revealed six distinct groups (Lessner *et al.*, 2012). Since this initial analysis there are substantially more archaeal genome sequences available in the NCBI database, including sequences from recently identified phyla and orders. Thus, we have updated the subunit D classification and diversity table to include an additional 180 archaeal genome sequences (**Table S3**). A summary of the presence of D3/FLD and [4Fe-4S] cluster motifs in subunit D among phyla/orders is shown in **Table 1**. Importantly, D3/FLD is present in the D subunit from the majority of sequenced archaea, with the exceptions being all orders of the phylum Thaumarchaeota and all species in the order Methanococcales (**Table 1**). Of those archaea that have a D3/FLD-containing D subunit, the majority contain at least one [4Fe-4S] cluster motif (12 of the 20 phyla/orders, **Table 1**). In particular, the cluster 1 motif is conserved, found in D subunits from 11 of the 12 phyla/orders. The cluster 2 motif is less conserved, typically only found in D subunits that also contain the cluster 1 motif. The cluster 1 motif, but not the cluster 2 motif, is found in the D3/FLD of the Rpb3/AC40 subunit from several eukaryotes (Hirata *et al.*, 2008). D subunits with a FLD containing two complete [4Fe-4S] binding motifs are found only in strictly anaerobic members of the Crenarchaeota and Euryarchaeota, as well as the recently deposited sequence from the founding member of the Lokiarchaeota (Spang *et al.*, 2015). Since the D3/FLD and [4Fe-4S] clusters are not universal

among archaea, it is unlikely they are essential for the function of RNAP, but instead may play an accessory role. The higher prevalence of the cluster 1 motif among archaea, and its presence in the Rpb3/AC40 subunit of RNAP from several eukaryotes, indicates cluster 1 plays a more prominent role than cluster 2 in RNAP.

Generation of subunit D variants deficient in binding [4Fe-4S] cluster(s). Previously, expression of subunit D deleted of D3/FLD, named DΔD3, in *E. coli* and within *M. acetivorans*, revealed that the FLD is not required for subunit D to form a heterodimer with subunit L (Lessner *et al.*, 2012). To ascertain the importance of each [4Fe-4S] cluster to D-L heterodimer formation and the interaction of the D-L heterodimer with other RNAP subunits during assembly of RNAP, D variants were generated defective in binding cluster 1 or cluster 2. To specifically assess the effect of the absence of the cluster, but retention of the cluster binding region on subunit D interactions, cysteine residues predicted to coordinate the cluster were changed to serine residues, in addition to deletion mutants (**Fig. 1**). For example, DmFeS1 and DΔFeS1 are each predicted to be incapable of binding cluster 1, but DmFeS1 still contains the cluster 1 binding region. As a control for subsequent *in vivo* studies, a subunit D variant deleted of C-terminal residues 227-262 (DΔCterm) was generated. These residues form an α-helix required for the interaction of subunit D with subunit L (Hirata *et al.*, 2008); thus, DΔCterm should be unable to form a heterodimer with subunit L.

First, to determine if each variant subunit D was capable of forming a heterodimer with subunit L, each variant was co-expressed along with histidine-tagged subunit L [L(His)] in *E. coli* followed by purification of subunit L(His) using Ni^{2+}-affinity and size-exclusion chromatography. Similar to previous results obtained for the purification of recombinant D-

L(His) and DΔD3-L(His) heterodimers (Lessner *et al.*, 2012), each single cluster-binding variant D was capable of forming a recombinant heterodimer with subunit L(His) that was devoid of Fe-S clusters (data not shown). However, co-expression of DΔCterm with subunit L(His) did not generate a heterodimer, but instead produced inclusion bodies containing DΔCterm (data not shown), similar to results obtained when subunit D was expressed in the absence of L(His) (Lessner *et al.*, 2012). These results are consistent with the C-terminal helix of subunit D being required for association with subunit L and formation of the D-L heterodimer, whereas mutation or deletion of the cluster 1 or cluster 2 binding regions does not impact formation of the D-L heterodimer.

Next, to ascertain whether each variant D-L(His) heterodimer is competent in binding the predicted number of [4Fe-4S] clusters, each purified variant D-L(His) heterodimer was reconstituted with iron and sulfide as previously described for D-L(His) and DΔD3-L(His) (Lessner *et al.*, 2012). Each D-L(His) heterodimer comprised of a single cluster-binding D variant contained approximately four molecules of iron and sulfur after reconstitution (**Table 2**) and generated UV-visible spectra (data not shown) consistent with the presence of a single [4Fe-4S] cluster. The ability to purify recombinant D-L(His) heterodimers competent in binding a single [4Fe-4S] cluster indicates that the directed changes to the cluster 1 or 2 binding-regions of subunit D specifically impact [4Fe-4S] cluster incorporation. These results support the use of the single cluster-binding variant D subunits, along with DΔD3, to investigate the effect of the loss of cluster 1, 2, or the entire FLD on the *in vivo* assembly and activity of RNAP.

Loss of [4Fe-4S] cluster binding in subunit D affects the *in vivo* assembly of RNAP.

Previously, two merodiploid strains (DJL30 and DJL31) of *M. acetivorans* were generated

(**Table 3**), each capable of tetracycline-inducible expression of a second subunit D harboring an N-terminal histidine-tag to facilitate purification by Ni^{2+}-affinity chromatography. Purification of (His)DΔD3 from strain DJL31 showed that the FLD of subunit D is not required for the *in vivo* interaction with subunit L and subsequent formation of the D-L heterodimer (Lessner *et al.*, 2012). To specifically test the effect of the absence of cluster 1 or 2 within subunit D on the *in vivo* assembly of RNAP, four additional merodiploid strains of *M. acetivorans* were generated, each capable of expressing a (His)D deficient in [4Fe-4FS] cluster binding (**Table 3**). A fifth strain (DJL40) capable of expressing (His)DΔCterm was also generated to use as a negative control, since DΔCterm cannot form a heterodimer with subunit L. Each strain exhibited wild-type growth rates and cell yields when grown in the presence or absence of tetracycline (data not shown), indicating expression of each (His)D variant is not detrimental to *M. acetivorans*.

To test the ability of each (His)D variant to compete with native D for assembly into RNAP, each strain was grown under inducing conditions and each (His)D variant partially purified from cell lysate by anaerobic Ni^{2+}-affinity chromatography. The imidazole eluates obtained from the purification columns contained similar amounts of total protein, ranging from 50-180 ng μl^{-1}. Western blot analysis using anti-D antibodies revealed the presence of (His)D in all eluates, except the eluate generated from strain DJL40 expressing (His)DΔCterm (**Fig. 2**). The lack of detection of (His)DΔCterm was expected based on the inability of recombinant DΔCterm to form a heterodimer with L(His) within *E. coli*. Thus, (His)DΔCterm within *M. acetivorans* is likely subjected to proteolysis. Since the eluates contained contaminating protein as determined by Coomassie staining (data not shown), the concentration of (His)D in each imidazole eluate was calculated using a calibration curve of band intensity with Western blots containing known amounts of recombinant subunit D. Multiple purifications revealed that the

imidazole eluate containing (His)D typically contained the highest amount of subunit D, whereas the eluates containing (His)D FLD variants contained 2-20X less subunit D. The concentration of subunit D in imidazole eluates from a representative experiment is shown in **Table 4**. These results indicate each [4Fe-4S] cluster (His)D variant is expressed in a stable form within *M. acetivorans*.

First, to determine if the inability to bind cluster 1 or 2 affects the *in vivo* formation of the D-L heterodimer, SDS-PAGE gels were loaded with eluate samples containing equal amounts of each (His)D and examined for the presence of native subunit L by Western blot. Native subunit L was detected at similar levels in all of the normalized (His)D samples (**Fig. 2**), revealing neither the presence of the FLD nor the ability to bind either cluster 1 or 2 is required for subunit D to interact with subunit L to form a stable heterodimer. Next, to ascertain whether the inability to bind cluster 1 or 2, or the loss of the entire FLD, affects the assembly of RNAP post D-L heterodimer formation, samples normalized to (His)D were examined by Western Blot using antibodies specific for subunit B' or A" (**Fig. 2**). Subunits B' and A" comprise part of the catalytic core of RNAP (De Carlo *et al.*, 2010). Subunits B' and A" were detected in the eluates containing (His)D, (His)DΔD3, (His)DΔFeS2 and (His)DmFeS2, but differed significantly in the observed band intensity. Subunits B' and A" were barely detectable in the (His)DΔD3 eluate, and (His)DΔFeS2 and (His)DmFeS2 eluates contained approximately 50% less B' and A" than the (His)D eluate, based on band intensity (**Fig. 2**). Subunits B' and A" were not detected in the eluates containing (His)DΔFeS1 or (His)DmFeS1. These results reveal that the FLD domain is critical to the interaction of the D-L heterodimer with at least subunits B' and A" in *M. acetivorans*. In particular, the inability of (His)DΔFeS1-L or (His)DmFeS1-L heterodimers to compete with native D-L heterodimer for assembly with subunits B' and A' reveals that the

presence of cluster 1 is a key determinant for the assembly of RNAP after D-L heterodimer formation in *M. acetivorans*. Finally, only the eluate containing (His)D exhibited non-specific transcription activity (**Table 4**), consistent with (His)D being assembled into holo-RNAP. The lack of transcription activity with the eluates containing (His)DΔFeS1 and (His)DmFeS1 was expected since these eluates lack RNAP catalytic subunits B' and A" (**Fig. 2**). The lack of activity with imidazole eluates containing (His)DΔD3, (His)DΔFeS2, and (His)DmFeS2 could be due to the lack of holo-RNAP, assembled RNAP is largely inactive, and/or the level of active holo-RNAP is insufficient to detect activity. Nonetheless, these results demonstrate that (His)D is assembled into functional holo-RNAP and reveal the importance of the FLD, in particular the cluster 1 region, of subunit D to the assembly of RNAP after the formation of D-L heterodimer.

Neither the FLD nor the ability of the FLD to bind [4Fe-4S] cluster 1 or 2 are essential to RNAP in *M. acetivorans.* The results from the purification of each (His)D variant from the merodiploid strains of *M. acetivorans* suggests that the entire FLD region and the ability to bind cluster 1 are critical to the assembly of RNAP. To determine whether the FLD region and binding of the [4Fe-4S] clusters are essential for the *in vivo* function of RNAP, we attempted to replace *rpoD* in the chromosome of *M. acetivorans* with mutated *rpoD* encoding each (His)D variant. After several independent transformation experiments (n = 6), four mutant strains were obtained (**Table 3**). Native subunit D was replaced with (His)D, (His)DΔD3, (His)DΔFeS1, and (His)DmFeS2, revealing that neither [4Fe-4S] cluster nor the entire FLD region are essential for functional RNAP in *M. acetivorans.*

The mutant strains were first analyzed by comparing their growth with methanol to that of the parent strain (WWM73) (**Fig. 3, Table 5**). Strain DJL51 grew identical to strain WWM73,

indicating the addition of the histidine-tag to the N-terminus of subunit D does not affect RNAP assembly and activity. However, strains DJL52, DJL54, and DJL55 all grew slower than both WWM73 and DJL51, revealing that the FLD region and the ability to bind cluster 1 and 2 are required for the optimal assembly and/or activity of RNAP in *M. acetivorans*. In addition, strains DJL52 and DJL54 took significantly longer to reach the exponential phase of growth, with DJL52 typically having a lag phase six times longer in duration than strain DJL51 (**Table 5**). Despite the longer lag phase duration and slower generation times, strains DJL52, DJL54, and DJL55 each reached a similar final cell density (**Table 5**). These results are consistent with those obtained from Western blot analyses of eluates from the purification of (His)D from the merodiploid strains (**Fig. 2**), which support the importance of the entire FLD region, and in particular cluster 1 to the post D-L heterodimer assembly of RNAP.

Loss of the FLD or each [4Fe-4S] cluster decreases the *in vivo* level of functional RNAP due to impaired assembly and/or stability of RNAP. To determine if the assembly and/or activity of RNAP is affected by the mutations in the FLD regions of subunit D, the levels of subunits D, B', and A" in each mutant strain were compared. Cell-free lysates derived from actively-growing cultures of each strain contained similar levels of subunit D, as determined by Western Blot of normalized total protein (**Fig. 4**). These results reveal that neither the loss of the entire FLD region, nor the inability of subunit D to bind cluster 1 or 2, affects the *in vivo* amount of subunit D. Next, to compare the levels of subunits B' and A'' in each strain, lysate samples containing an equal amount of subunit D (15 ng) for each strain were analyzed by Western blot using anti-B' and anti-A" antibodies. There were no significant differences in the amounts of subunits B' and A'' in lysates from multiple cultures (an example blot is shown in **Fig. 4)**, indicating the *in vivo*

levels of the RNAP catalytic subunits are not impacted by changes to the FLD of subunit D. However, despite each strain containing similar levels of the individual RNAP subunits, lysate derived from strains DJL54, DJL52, and DJL55 exhibited significantly lower non-specific transcription activity compared to strain DJL51, which contains a wild-type FLD region (**Table 6**). In particular, lysates containing (His)DΔD3 and (His)DΔFeS1 contained the lowest activity, consistent with the results obtained from the purification of the (His)D variants from the merodiploid strains (**Fig. 2**), which showed the greater importance of cluster 1 to the assembly of RNAP after the formation of the D-L heterodimer. These results indicate that the assembly and/or the intrinsic activity of RNAP is negatively impacted by the removal of the FLD region or mutation of the cluster 1 or 2 binding sites of subunit D. To specifically examine the impact on assembly, the (His)D variant from strains DJL51, DJL52, DJL54, and DJL55 was purified using Ni^{2+}-affinity chromatography. The level of subunits D, L, B', and A" in each imidazole eluate were determined by Western Blot (**Fig. 5**). Each eluate contained similar concentrations of (His)D (**Table 7**). Western blot analysis of eluate samples normalized to contain the same amount of (His)D revealed a similar level of subunit L in each sample (**Fig. 5**). Thus, the formation of the D-L heterodimer is not altered by mutation of the FLD region of subunit D, consistent with results obtained from the merodiploid strains (**Fig. 2**).

Surprisingly, unlike the partial purification of (His)D variants from the merodiploid strains, which revealed heterodimers containing (His)DΔD3 and (His)DΔFeS1 were defective in assembling with subunit B' (**Fig. 2**), subunit B' co-purified with (His)DΔD3 and (His)DΔFeS1 almost as equally as well as it did with (His)D (**Fig. 5**). However, the co-purification of subunit A" with (His)DΔD3, (His)DΔFeS1, and (His)DmFeS2 was significantly less than that observed with (His)D. Consistent with the diminished co-purification of subunit A", the eluates containing

(His)DΔD3, (His)DΔFeS1, and (His)DmFeS2 exhibited significantly lower RNAP activity compared to the eluate containing (His)D (**Table 7**). These results indicate that (His)DΔD3, (His)DΔFeS1, and (His)DmFeS2 form at least a B'DL subcomplex, but that the loss of the FLD or the inability to bind either cluster results in a subcomplex that is defective in associating with subunit A".

Discussion

The results presented herein provide insight into the role [4Fe-4S] clusters play in specific steps during the assembly of RNAP in *M. acetivorans*. The assembly of RNAP within cells from all three domains of life occurs in stages, starting with the formation of an assembly platform. The assembly of archaeal RNAP starts with the formation of the D-L heterodimer, followed by the step-wise addition of the remaining subunits (Eloranta *et al.*, 1998, Goede *et al.*, 2006). The D-L heterodimer first associates with subunits N and P, followed by the addition of catalytic core subunit B, to form a BDLNP subcomplex. Subunit B is split into two separate proteins (B' and B") in many Euryarchaeota, including *M. acetivorans*. Thus, *M. acetivorans* specifically forms a B'B"DLNP subcomplex. The BDLNP subcomplex is competent in transcription factor interaction and DNA binding, indicating the subcomplex is highly stable (Goede *et al.*, 2006). Subunits A' and A", which make up the other portion of the catalytic core, subsequently associate with the BDLNP complex, followed by auxiliary subunits HKFE to form complete RNAP. Given that the [4Fe-4S] cluster(s) reside in subunit D, which is not part of the catalytic core of RNAP, and that the [4Fe-4S] cluster in the structure of *S. solfataricus* RNAP is a substantial distance from the active site, the cluster(s) likely do not participate in catalysis. A more likely function for the cluster(s) is to serve as a determinant to regulate the *de novo* assembly of RNAP. Alternatively, the cluster(s) may serve as a recognition element for the

specific interaction of RNAP with general or gene-specific transcription factors. The essentiality of the clusters and their importance to specific steps in the assembly of RNAP within *M. acetivorans* was specifically investigated and the results obtained support the model proposed in **Figure 6**.

The purification of (His)D variants from merodiploid strains of *M. acetivorans* revealed that neither the removal of the entire FLD, nor the deletion or mutation of each [4Fe-4S] cluster binding motif, impacts the ability of subunit D to form a heterodimer with subunit L within *M. acetivorans* (**Fig. 2**). These results, combined with the ability to incorporate the [4Fe-4S] clusters *in vitro* into the recombinant D-L heterodimers, indicate insertion of each cluster occurs after the formation of the D-L heterodimer (**Fig. 6**). In contrast, recombinant *S. solfataricus* subunit D with mutations in the [4Fe-4S] cluster binding motif failed to form a heterodimer with subunit L in *E. coli*, indicating the single cluster (analogous to cluster 1 in *M. acetivorans*) is required for D-L heterodimer formation in *S. solfataricus* (Hirata et al., 2008). However, *S. solfataricus* is an extremophile and its subunit D contains a disulfide bond in place of a second [4Fe-4S] cluster, unlike *M. acetivorans*. Mutation of the [4Fe-4S] cluster binding motif in *S. solfataricus* subunit D may result in incorrect disulfide bond formation in the reducing environment of the cytoplasm of *E. coli*. Moreover, subunit D from the methanogen *Methanobrevibacter smithii* contains an FLD predicted to bind only cluster 1, similar to *S. solfataricus* subunit D, and co-expression studies in *E. coli* revealed that [4Fe-4S] cluster incorporation into subunit D is not required for the association of recombinant *M. smithii* subunit D with subunit L (unpublished results, SC Granderson and DJ Lessner). Although the requirement of the cluster(s) for the formation of the D-L heterodimer cannot be ruled out in some species, the clusters are not required for the

association of subunits D and L in two methanogens, and likely the majority of other species containing subunit D predicted to bind [4Fe-4S] clusters.

Results from the purification of (His)D variants from the merodiploid strains of *M. acetivorans* revealed that the removal of the entire FLD and the deletion or mutation of the [4Fe-4S] cluster 1 binding motif severely impacts the ability of the D-L heterodimer to associate with subunit B' (**Fig. 2**). Thus, the formation of the B'B"DLNP subcomplex is likely compromised within *M. acetivorans*. Since the FLD of subunit D is in close proximity to subunit B in the structure of *S. solfataricus* RNAP (Hirata *et al.*, 2008), this result is consistent with the FLD being needed for optimal interaction with subunit B' in *M. acetivorans*. Previous results demonstrated that oxidative loss of the clusters from the FLD resulted in substantial change to the conformational stability of the *M. acetivorans* D-L heterodimer (Lessner *et al.*, 2012). The absence of clusters, in particular cluster 1 shown here, likely alters the confirmation of the FLD which negatively impacts interaction with subunit B', and thus formation of the B'B"DLNP subcomplex within *M. acetivorans* (**Fig. 6**). However, the ability to replace native subunit D with (His)D variants revealed that the FLD, including each cluster, is not an essential determinant for the interaction of the D-L heterodimer with subunit B' and likely the formation of the B'B"DLNP subcomplex in *M. acetivorans*. Unlike in the merodiploid strains, where each (His)D variant must compete with native D for association with native subunit B', each (His)D variant must associate with native subunit B', as well as all other RNAP subunits, in order to obtain a viable gene replacement strain. Purification of (His)DΔD3, (His)DΔDFeS1, and (His)DmFeS2 from the gene replacement strains revealed each variant forms a D-L heterodimer that is competent in associating with subunit B' to likely form a B'B"DLNP subcomplex similar to (His)D (**Fig. 5**). However, all three strains harboring mutations in the FLD of subunit D

exhibited impaired growth and diminished RNAP activity compared to the control strain, with the strain containing (His)DΔDFeS1 having the most severe phenotype (**Fig. 3 and Table 6**). The diminished co-purification of subunit A" with all three (His)D FLD variants (**Fig. 5**) revealed that the observed phenotypes were due, at least in part, to impaired assembly after B'B"DLNP subcomplex formation. Thus, although deletion or mutation of the FLD of subunit D allows formation of a B'B"DLNP subcomplex, the subcomplex is altered such that subunit A", and likely subunit A', does not optimally associate with the complex. It is also possible that RNAP stability is altered such that subunit A" is lost during the purification of each (His)D FLD variant. In either case, these results reveal that the absence of the FLD and each [4Fe-4S] cluster has a substantial effect on the association of the B'B"DLNP subcomplex with subunit A", and likely the remaining subunits (A'HKEF). Thus, the FLD and [4Fe-4S] clusters are critical to the post B'B"DLNP subcomplex formation steps in the assembly of RNAP in *M. acetivorans* (**Fig. 6**).

The results from this study clearly demonstrate that [4Fe-4S] cluster 1 is a more important determinant than [4Fe-4S] cluster 2 for optimal assembly and activity of RNAP in *M. acetivorans*. Subunits B' and A" co-purified with (His)DΔFeS2 and (His)DmFeS2 from the merodiploid strains, but failed to co-purify with (His)DΔFeS1 and (His)DmFeS1 (**Fig. 2**). Moreover, strain DJL52 encoding (His)DΔFeS1 had the most impaired growth of any of the strains (**Fig. 3 and Table 5**). These results are consistent with the [4Fe-4S] cluster 1 motif being more conserved, compared to [4Fe-4S] cluster 2 motif, in subunit D amongst archaea and may explain why the cluster 1 motif, but not the cluster 2 motif, is found in the Rpb3/AC40 subunit of RNAP in some eukaryotes. The FLD and clusters are clearly not essential, since they can be deleted from *M. acetivorans* and are not found in all RNAPs. This supports a regulatory role for

the clusters in the assembly of RNAP. The acquisition of [4Fe-4S] clusters to use as determinants in the assembly of RNAP may allow *M. acetivorans*, and other species, to correlate energy conservation processes (e.g. methanogenesis) that are dependent on iron and Fe-S clusters with biosynthetic processes (e.g. transcription) by altering the levels of functional RNAP. The [4Fe-4S] clusters may be used to sense changes in key environmental factors that affect energy conservation, such as iron availability, oxygen, and reactive oxygen species.

D3/FLD is absent from all species of the order Methanococcales and the phylum Thaumarchaeota, both of which are deeply rooted archaeal lineages (Brochier-Armanet *et al.*, 2011). Since ferredoxin is a critical Fe-S cluster protein in methanogenesis (Ferry, 1999, Thauer *et al.*, 2008), and to the metabolism of most anaerobes, we hypothesize that two [4Fe-4S] cluster ferredoxin was spliced into RNAP in an ancestral anaerobe (e.g., methanogen). This hypothesis is supported by the fact that only strictly anaerobic archaea possess subunit D with two [4Fe-4S] cluster binding motifs, similar to ferredoxin. The FLD was subsequently modified due to diverse selective pressures exerted upon specific lineages, resulting in cluster modification (e.g. oxygen stability) or loss of one or both clusters. Due to the greater importance of cluster 1 to *M. acetivorans* RNAP shown here, cluster 2 was lost in more lineages than cluster 1. Although the results reveal the [4Fe-4S] cluster(s) are key determinants in the assembly of RNAP, it remains unclear whether the [4Fe-4S] cluster(s) also serve as a recognition element for the interaction of RNAP with general and specific transcription factors to specifically alter the expression of certain genes. Importantly, the generation of *M. acetivorans* strains harboring RNAP missing the FLD or each [4Fe-4S] cluster provides an avenue to address the significance of the [4Fe-4S] clusters to the transcription of specific genes in the future.

Acknowledgements
This work was supported in part by grant number P30 GM103450 from the National Institute of General Medical Sciences of the National Institutes of Health (DJL), NSF grant number MCB1121292 (DJL), NASA Exobiology grant number NNX12AR60G (DJL), and the Arkansas Biosciences Institute (DJL), the major research component of the Arkansas Tobacco Settlement Proceeds Act of 2000.

References

Bradford, M.M., (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-254.

Brochier-Armanet, C., P. Forterre & S. Gribaldo, (2011) Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr Opin Microbiol* 14: 274-281.

Cruz, F. & J.G. Ferry, (2006) Interaction of iron-sulfur flavoprotein with oxygen and hydrogen peroxide. *Biochim. Biophys. Acta.* 1760: 858-864.

Darcy, T.J., W. Hausner, D.E. Awery, A.M. Edwards, M. Thomm & J.N. Reeve, (1999) Methanobacterium thermoautotrophicum RNA polymerase and transcription in vitro. *J Bacteriol* 181: 4424-4429.

De Carlo, S., S.-C. Lin, D.J. Taatjes & A. Hoenger, (2010) Molecular basis of transcription initiation in Archaea. *Transcription* 1: 103-111.

Decker, K.B. & D.M. Hinton, (2013) Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu Rev Microbiol* 67: 113-139.

Eloranta, J.J., A. Kato, M.S. Teng & R.O. Weinzierl, (1998) In vitro assembly of an archaeal D-L-N RNA polymerase subunit complex reveals a eukaryote-like structural arrangement. *Nucleic Acids Res.* 26: 5562-5567.

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

Fuss, J.O., C.L. Tsai, J.P. Ishida & J.A. Tainer, (2015) Emerging critical roles of Fe-S clusters in DNA replication and repair. *Biochimica et biophysica acta* 1853: 1253-1271.

Goede, B., S. Naji, O. von Kampen, K. Ilg & M. Thomm, (2006) Protein-protein interactions in the archaeal transcriptional machinery: binding studies of isolated RNA polymerase subunits and transcription factors. *J. Biol. Chem.* 281: 30581-30592.

Guss, A.M., M. Rother, J.K. Zhang, G. Kulkarni & W.W. Metcalf, (2008) New methods for tightly regulated gene expression and highly efficient chromosomal integration of cloned genes for *Methanosarcina* species. *Archaea* 2: 193-203.

Hirata, A., B.J. Klein & K.S. Murakami, (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451: 851-854.

Hirata, A. & K.S. Murakami, (2009) Archaeal RNA Polymerase. *Curr. Opin. Struct. Biol.* 19: 724-731.

Imlay, J.A., (2006) Iron-sulphur clusters and the problem with oxygen. *Mol Microbiol* 59: 1073-1082.

Johnson, D.C., D.R. Dean, A.D. Smith & M.K. Johnson, (2005) Structure, function, and formation of biological iron-sulfur clusters. *Annu. Rev. Biochem.* 74: 247-281.

Klinge, S., J. Hirst, J.D. Maman, T. Krude & L. Pellegrini, (2007) An iron-sulfur domain of the eukaryotic primase is essential for RNA primer synthesis. *Nat Struct Mol Biol* 14: 875-877.

Lessner, F.H., M.E. Jennings, A. Hirata, E.C. Duin & D.J. Lessner, (2012) Subunit D of RNA polymerase from Methanosarcina acetivorans contains two oxygen-labile [4Fe-4S] clusters: implications for oxidant-dependent regulation of transcription. *J Biol Chem* 287: 18510-18523.

Major, T.A., H. Burd & W.B. Whitman, (2004) Abundance of 4Fe-4S motifs in the genomes of methanogens and other prokaryotes. *FEMS Microbiol. Lett.* 239: 117-123.

Metcalf, W.W., J.K. Zhang, E. Apolinario, K.R. Sowers & R.S. Wolfe, (1997) A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. *Proc. Natl. Acad. Sc.i USA* 94: 2626-2631.

Netz, D.J., C.M. Stith, M. Stumpfig, G. Kopf, D. Vogel, H.M. Genau, J.L. Stodola, R. Lill, P.M. Burgers & A.J. Pierik, (2012) Eukaryotic DNA polymerases require an iron-sulfur cluster for the formation of active complexes. *Nat Chem Biol* 8: 125-132.

Rodriguez-Monge, L., C.A. Ouzounis & N.C. Kyrpides, (1998) A ferredoxin-like domain in RNA polymerase 30/40-kDa subunits. *Trends Biochem Sci* 23: 169-170.
Schneider, C.A., W.S. Rasband & K.W. Eliceiri, (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature methods* 9: 671-675.

Sousa, F.L., T. Thiergart, G. Landan, S. Nelson-Sathi, I.A. Pereira, J.F. Allen, N. Lane & W.F. Martin, (2013) Early bioenergetic evolution. *Philos Trans R Soc Lond B Biol Sci* 368: 20130088.

Sowers, K.R., S.F. Baron & J.G. Ferry, (1984) Methanosarcina acetivorans sp. nov., an Acetotrophic Methane-Producing Bacterium Isolated from Marine Sediments. *Applied and Environmental Microbiology* 47: 971-978.

Sowers, K.R., J.E. Boone & R.P. Gunsalus, (1993) Disaggregation of Methanosarcina spp. and Growth as Single Cells at Elevated Osmolarity. *Appl. Environ. Microbiol.* 59: 3832-3839.

Spang, A., J.H. Saw, S.L. Jorgensen, K. Zaremba-Niedzwiedzka, J. Martijn, A.E. Lind, R. van Eijk, C. Schleper, L. Guy & T.J. Ettema, (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521: 173-179.

Tamura, K., G. Stecher, D. Peterson, A. Filipski & S. Kumar, (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* 30: 2725-2729.

Thauer, R.K., A.K. Kaster, H. Seedorf, W. Buckel & R. Hedderich, (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* 6: 579-591.

Werner, F. & D. Grohmann, (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* 9: 85-98.

Figures and Tables

Table 1. Plasmids utilized in this study

Plasmid	Description
pRpoDL	Plasmid containing *M. acetivorans rpoD* and *rpoL*(His) (Lessner *et al.*, 2012)
pDL408	*rpoD*ΔD3 in pRpoDL backbone
pDL406	*rpoD*ΔFeS1 in pRpoDL backbone
pDL407	*rpoD*ΔFeS2 in pRpoDL backbone
pDL315	*rpoD*mFeS1 in pRpoDL backbone
pDL314	*rpoD*mFeS2 in pRpoDL backbone
pDL414	*rpoD*ΔCterm in pRpoDL backbone
pJK027A	Plasmid for integration into *M. acetivorans* strain WWM73 chromosome (Guss *et al.*, 2008)
pDL516	(His)*rpoD* cloned into pJK027A
pDL409	(His)*rpoD*ΔD3 cloned into pJK027A
pDL415	(His)*rpoD*ΔFeS1 cloned into pJK027A
pDL416	(His)*rpoD*ΔFeS2 cloned into pJK027A
pDL418	(His)*rpoD*mFeS1 cloned into pJK027A
pDL417	(His)*rpoD*mFeS2 cloned into pJK027A
pDL420	(His)*rpoD*ΔCterm cloned into pJK027A
pJK301	Plasmid for double homolgous recombination into *M. acetivorans* strain WWM73 chromosome (Guss *et al.*, 2008)
pDL517	pJK301 derivative without NdeI site
pDL518	pUC19 with 4 Kb of *M. acetivorans* genomic DNA (region upstream and including rpoD) cloned into BamHI site
pDL519	pDL518 with genomic ApaI site upstream of *rpoD* removed by QuikChange
pDL520	pDL519 with genomic XhoI site upstream of *rpoD* removed by QuikChange
pDL521	pDL517 with 2.8 Kb of genomic DNA upstream of *rpoD* from pDL520 cloned into ApaI/XhoI sites
pDL522	pDL521 with 2.8 Kb of *M. acetivorans* genomic DNA (region downstream of *rpoD*) cloned into BamH1/NotR sites
pDL525	(His)*rpoD* cloned into pDL522
pDL524	(His)*rpoD*ΔD3 cloned into pDL522
pDL523	(His)*rpoD*ΔFeS1 cloned into pDL522
pDL528	(His)*rpoD*ΔFeS2 cloned into pDL522
pDL527	(His)*rpoD*mFeS1 cloned into pDL522
pDL526	(His)*rpoD*mFeS2 cloned into pDL522

Table 2. Primers utilized in this study

Primer	Sequence	Description
RpoDNdeFor	ggggaattaaggcatatgacgatggaag	forward primer to amplify *rpoD* with NdeI site at 5' end
RpoDHindRev	aagctcaaaagcttggcatagg	reverse primer to amplify *rpoD* with HindIII site at 3' end
HisRpoDNdeFor	attaaggcatatgcatcatcatcatcatcatacgatggaagtagacattct	forward primer to amplify *rpoD* with N-terminal His6 tag and NdeI site at 5' end
HisRpoDHindRev	ggtggtaagctttcagagctggtccagaattgc	reverse primer to amplify *rpoD* with HindIII site and stop codon at 3' end
QCRpoDΔFeS1For	gtggacttctatgaaaactcttttg	forward primer to make *rpoD*ΔFeS1 via QuikChange
QCRpoDΔFeS1Rev	ggcaatcttagctccggcctcttc	reverse primer to make *rpoD*ΔFeS1 via QuikChange
QCRpoDΔFeS2For	gaagaggccggagctaagattg	forward primer to make *rpoD*ΔFeS2 via QuikChange
QCRpoDΔFeS2Rev	aatggtaattacaggcatgtttttg	reverse primer to make *rpoD*ΔFeS2 via QuikChange
QCRpoDmFeS2For	accattgaaaactccgatgcctccggacactctgcggca	forward primer to make *rpoD*mFeS2 via QuikChange
QCRpoDmFeS2Rev	tgccgcagagtgtccggaggcatcggagttttcaatggt	reverse primer to make *rpoD*mFeS2 via QuikChange
QCRpoDmFeS1For	aagacatcatgaagtcttccatctccaggctctgtgagca	forward primer to make *rpoD*mFeS1via QuikChange
QCRpoDmFeS1Rev	tgctcacagagcctggagatggaa gacttcatgatgtctt	reverse primer to make *rpoD*mFeS1 via QuikChange
RpoDΔCtermFor	attctggaccagctctgaggatc	forward primer to amplify entire pRpoDL plasmid to generate *rpoD*ΔCterm
RpoDΔCtermRev	ttcatagaagtccactttgatcgcg	reverse primer to amplify entire pRpoDL plasmid to generate *rpoD*ΔCterm
RpoDΔCtermHindRev	ggtggtaagctttcagagctggtccagaatttc	reverse primer to amplify *rpoD*ΔCterm with HindIII site and stop codon at 3' end
C2Achr1	gaagcttccccttgaccaat	Integration screening primer of ΦC31 site in *M. acetivorans* strain WWM73 (Guss *et al.*, 2008)
C2Achr1	ttgattcggataccctgagc	Integration screening primer of ΦC31 site in *M. acetivorans* strain WWM73 (Guss *et al.*, 2008)

Table 2. Cont.

Primer	Sequence	Description
plscreen3	gcaaagaaaagccagtatgga	Integration screening primer of ΦC31 site in *M. acetivorans* strain WWM73 (Guss *et al.*, 2008)
plscreen4	tttttcgtctcagccaatcc	Integration screening primer of ΦC31 site in *M. acetivorans* strain WWM73 (Guss *et al.*, 2008)
QCJK301For	gaatctaaatggaggtttagacacatgc ttgaaagactgaaagactc	Forward primer to remove NdeI site from pJK301 via QuikChange
QCJK301Rev	gagtctttcagtctttcaagcatgtgtcta aacctccatttagattc	Reverse primer to remove NdeI site from pJK301 via QuikChange
RpoDUSNcoFor	gatgatccatggatgttgaaccgcccttt tctg	forward primer to amplify 4 Kb genomic region upstream of *rpoD* with NcoI site at 5' end
RpoDUSNcoRev	ggtggtccatggagaattctctgaataat tcgc	reverse primer to amplify 4 Kb genomic region upstream of *rpoD* with NcoI site at 3' end
RpoDUSBamFor	ggcggcggatccatcatcgactgcgg catatctcccgc	forward primer to amplify 2.8 Kb genomic region upstream of *rpoD* with BamHI site at 5' end
RpoDUSBamRev	agccggatcctcagagctggtccagaa ttgccagc	reverse primer to amplify 2.8 Kb genomic region upstream of *rpoD* with BamHI site at 3' end
QCRpoDUSApaF	gagaagtcccggacccggtgcacag	forward primer to remove ApaI site from genomic DNA via QuikChange
QCRpoDUSApaR	ctgtgcaccgggtccgggacttctc	reverse primer to remove ApaI site from genomic DNA via QuikChange
QCRpoDUSXhoF	caggaaggaggactagagggccacta cag	forward primer to remove XhoI site from genomic DNA via QuikChange
QCRpodUSXhoR	ctgtagtggccctctagtcctccttcctg	reverse primer to remove XhoI site from genomic DNA via QuikChange
RpoDUSApaFor	gctgctgggcccagggcagatgttgaa ccgccttttc	forward primer to amplify 2.8 Kb upstream genomic region with ApaI at 5' end
RpoDUSXhoRev	agccctcgagtcagagctggtccagaa ttgccagc	reverse primer to amplify 2.8 Kb upstream genomic region with XhoI at 5' end

Table 2. Cont.

Primer	Sequence	Description
RpoDDSBamFor	ggcggattctgtcttcttattttgagaactc ttaagg	forward primer to amplify ~2 Kb genomic region downstream of *rpoD* with BamHI at 5' end
RpoDDSNotRev	cgacgacgagcggccgcgcccaccct cactgtggagccggaacc	reverse primer to amplify ~2 Kb genomic region downstream of *rpoD* with NotI at 5' end
TxnAsy90A	atcttaatagttattatttctataaccttt ttaagtatccggtggtggatatctttc ataaatgaaaatatttttcgttgataattata a	Oligonucleotide for non-specific transcription assay
TxnAsy90B	ttataattatcaacgaaaaatattttcatttat gaaagatatccaccaccggatacttaaa aaggttatagaaataataactattaagat	Complimentary oligonucleotide for non-specific transcription assay

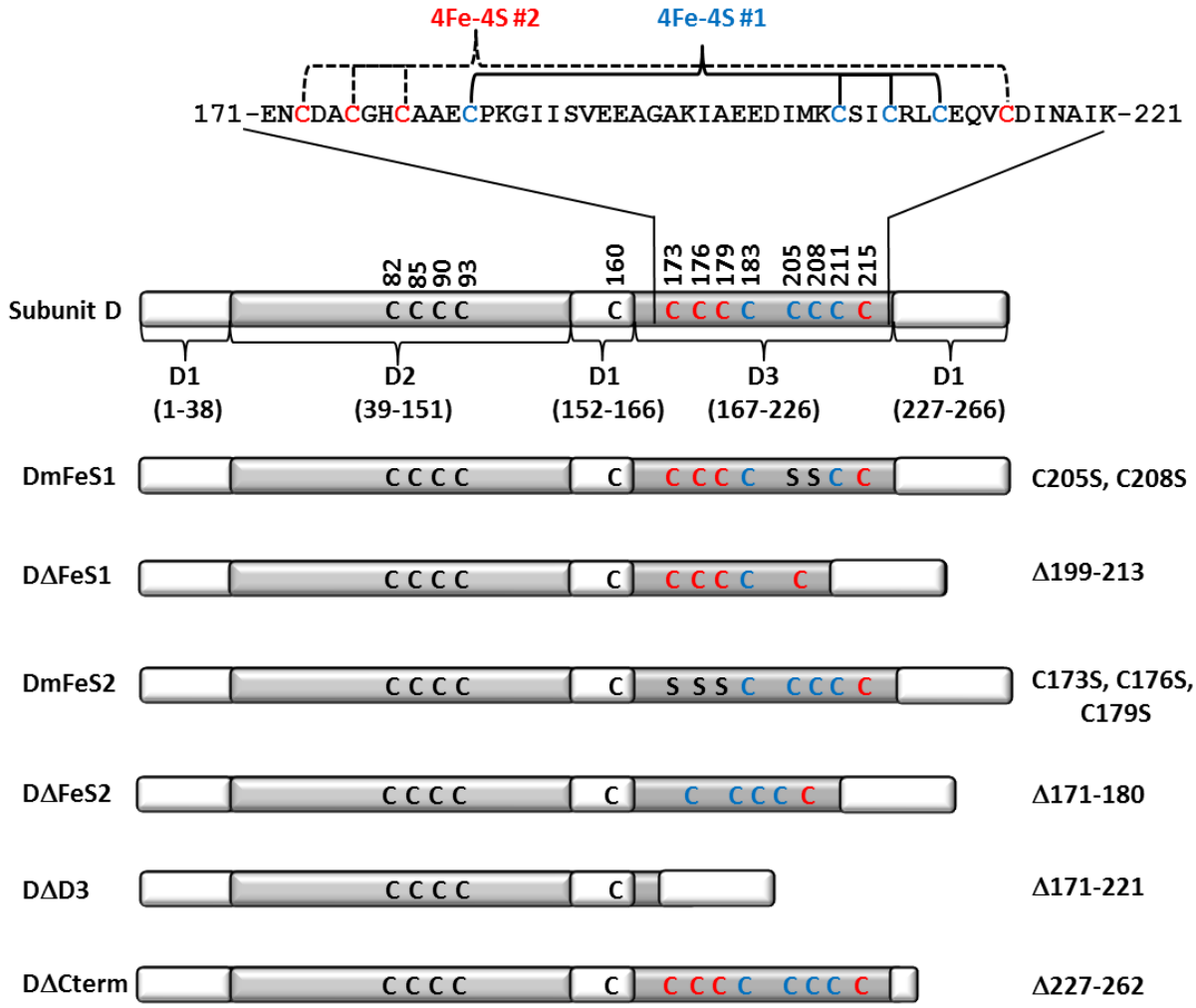

Fig 1. *M. acetivorans* subunit D variants used in this study. Subunit D is depicted to show the three domains (D1-3) and the sequence of D3/FLD, including the cysteines that comprise each [4Fe-4S] cluster binding motif. Each variant is listed below the wild type to show which residues were changed (cysteine (C) to serine (S)) or deleted.

Table 3A. Detailed analysis of the conservation of D3/FLD in subunit D from archaea.

Group Identification:
Group 1: complete motifs #1 and #2
Group 2: complete motif #1
Group 3: complete motif #2
Group 4: partial #1 and/or #2 motifs
Group 5: no #1 and #2 motifs
Group 6: lack domain 3 amino acid residues

Phylum Crenarchaeota							
Order	GI	Accession	Genus	Species	cluster 1	cluster 2	Group Identification
Desulfurococcales	549636204	WP_022541834.1	Aeropyrum	camini	2	2	Group 4
	499168291	WP_010866567.1	Aeropyrum	pernix	2	2	Group 4
	756978731	WP_042666793.1	Desulfurococcus	amylolyticus	4	4	Group 1
	504581108	WP_014768210.1	Desulfurococcus	fermentans	4	4	Group 1
	501638949	WP_012609030.1	Desulfurococcus	kamchatkensis	4	4	Group 1
	503328209	WP_013562870.1	Desulfurococcus	mucosus	4	4	Group 1
	500145705	WP_011821708.1	Hyperthermus	butylicus	2	2	Group 4
	500798811	WP_011998219.1	Ignicoccus	hospitalis	2	4	Group 3
	503067720	WP_013302692.1	Ignisphaera	aggregans	4	4	Group 1
	503792608	WP_014026602.1	Pyrolobus	fumarii	2	2	Group 4
	502908958	WP_013143934.1	Staphylothermus	hellenicus	4	4	Group 1
	500164491	WP_011839107.1	Staphylothermus	marinus	4	4	Group 1
	504550103	WP_014737205.1	Thermogladius	cellulolyticus	4	4	Group 1
	502895271	WP_013130247.1	Thermosphaera	aggregans	4	4	Group 1
Sulfolobales	503541388	WP_013775464.1	Acidianus	hospitalis	4	2	Group 2
	612166521	EZQ03163.1	Candidatus Acidianus	copahuensis	4	2	Group 2
	503502092	WP_013736753.1	Metallosphaera	cuprina	4	2	Group 2
	500929300	WP_012022071.1	Metallosphaera	sedula	4	2	Group 2
	496366102	WP_009075092.1	Metallosphaera	yellowstonensis	4	2	Group 2
	583848448	EWG08139.1	Sulfolobales	archaeon AZ1	4	2	Group 2
	499596277	WP_011277011.1	Sulfolobus	acidocaldarius	4	2	Group 2
	502110198	WP_012712015.1	Sulfolobus	islandicus	4	2	Group 2
	219814435	ACL36491.1	Sulfolobus	shibatae B12	4	2	Group 2
	497674697	WP_009988881.1	Sulfolobus	solfataricus	4	2	Group 2
	499288852	WP_010980142.1	Sulfolobus	tokodaii	4	2	Group 2

Table 3A. Cont.

Phylum Crenarchaeota							
Order	GI	Accession	Genus	Species	cluster 1	cluster 2	Group Identification
Thermoproteales	501137698	WP_012186212.1	Caldivirga	maquilingensis	3	3	Group 4
	145284289	ABP51871.1	Pyrobaculum	arsenaticum	3	3	Group 4
	500176345	WP_011850770.1	Pyrobaculum	calidifontis	3	3	Group 4
	500086438	WP_011762451.1	Pyrobaculum	islandicum	3	3	Group 4
	501319668	WP_012351303.1	Pyrobaculum	neutrophilum	3	3	Group 4
	504114026	WP_014348012.1	Pyrobaculum	oguniense	3	3	Group 4
	356642897	AET33576.1	Pyrobaculum	sp. 1860	3	3	Group 4
	499316920	WP_011007412.1	Pyrobaculum	aerophilum	3	3	Group 4
	500232178	WP_011901774.1	Pyrobaculum	arsenaticum	3	3	Group 4
	742686299	AJB42579.1	Thermofilum	carboxyditrophus	4	0	Group 2
	500075942	WP_011751955.1	Thermofilum	pendens	4	0	Group 2
	778567722	KJR74069.1	Thermoproteus	sp. AZ2	3	3	Group 4
	503893553	WP_014127547.1	Thermoproteus	tenax	3	3	Group 4
	503445973	WP_013680634.1	Thermoproteus	uzoniensis	3	3	Group 4
	503100432	WP_013335225.1	Vulcanisaeta	distributa	3	3	Group 4
	503369677	WP_013604338.1	Vulcanisaeta	moutnovskia	3	3	Group 4
	778553869	KJR71874.1	Vulcanisaeta	sp. AZ3	3	3	Group 4
Acidilobales	503031949	WP_013266925.1	Acidilobus	saccharovorans	2	2	Group 4
	505045107	WP_015232209.1	Caldisphaera	lagunensis	2	2	Group 4
	557076107	ESQ22024.1	uncultured Acidilobus	sp. CIS	2	2	Group 4
	557080029	ESQ25838.1	uncultured Acidilobus	sp. JCHS	1	3	Group 4
	557077206	ESQ23084.1	uncultured Acidilobus	sp. MG	2	2	Group 4
	557078871	ESQ24700.1	uncultured Acidilobus	sp. OSPS	1	3	Group 4
Fervidicoccales	504370501	WP_014557603.1	Fervidicoccus	fontis	4	2	Group 2
Phylum Euryarchaeota							
Thermoplasmatales	765467717	KJE49536.1	Acidiplasma	sp. MBA-1	2	0	Group 4
	518679936	WP_019841629.1	Ferroplasma	acidarmanus fer1	2	0	Group 4
	499491383	WP_011178023.1	Picrophilus	torridus	2	0	Group 4
	10640345	CAC12159.1	Thermoplasma	acidophilum	2	0	Group 4
	499219283	WP_010916823.1	Thermoplasma	volcanium	2	0	Group 4
	472831695	AGI47432.1	Thermoplasmatales	sp. BRNA1	0	0	Group 5
Archaeoglobales	499182231	WP_010879771.1	Archaeoglobus	fulgidus	4	4	Group 1
	502705183	WP_012940424.1	Archaeoglobus	profundus	4	4	Group 1
	505403141	WP_015590243.1	Archaeoglobus	sulfaticallidus	4	4	Group 1
	503448409	WP_013683070.1	Archaeoglobus	veneficus	4	4	Group 1
	502730372	WP_012965356.1	Ferroglobus	placidus	4	4	Group 1

Table 3A. Cont.

Phylum Euryarchaeota							
Order	GI	Accession	Genus	Species	cluster 1	cluster 2	Group Identification
Halobacteriales	557372300	WP_023393369.1	Candidatus Halobonum	tyrrellensis	0	0	Group 5
	495254127	WP_007978882.1	Haladaptatus	paucihalophilus	0	0	Group 5
	495692844	WP_008417423.1	Halalkalicoccus	jeotgali	0	0	Group 5
	519065695	WP_020221570.1	Halarchaeum	acidiphilum	0	0	Group 5
	544639648	WP_021074204.1	Haloarchaeon	3A1 DGR	0	0	Group 5
	445770088	EMA21156.1	Haloarcula	amylolytica JCM 13557	0	0	Group 5
	445772692	EMA23737.1	Haloarcula	argentinensis	0	0	Group 5
	445770767	EMA21825.1	Haloarcula	californiae ATCC 33799	0	0	Group 5
	343782465	AEM56442.1	Haloarcula	hispanica ATCC 33960	0	0	Group 5
	490731190	WP_004593560.1	Haloarcula	japonica	0	0	Group 5
	445761375	EMA12623.1	Haloarcula	sinaiiensis	0	0	Group 5
	749701601	AJF26395.1	Haloarcula	sp. CBA1115	0	0	Group 5
	490651807	WP_004516801.1	Haloarcula	vallismortis	0	0	Group 5
	6172230	BAA85898.1	Halobacterium	salinarum	0	0	Group 5
	573485772	AHG04178.1	Halobacterium	sp. DL1	0	0	Group 5
	10580674	AAG19520.1	Halobacterium	sp. NRC-1	0	0	Group 5
	494241630	WP_007143644.1	Halobiforma	lacisalsi	0	0	Group 5
	493722633	WP_006672048.1	Halobiforma	nitratireducens	0	0	Group 5
	494969291	WP_007695317.1	Halococcus	hamelinensis	0	0	Group 5
	490156745	WP_004055417.1	Halococcus	morrhuae	0	0	Group 5
	492978452	WP_006076166.1	Halococcus	saccharolyticus	0	0	Group 5
	491184899	WP_005043260.1	Halococcus	salifodinae	0	0	Group 5
	495016319	WP_007742329.1	Halococcus	thailandensis	0	0	Group 5
	445734852	ELZ86408.1	Haloferax	alexandrinus	0	0	Group 5
	445753371	EMA04788.1	Haloferax	denitrificans	0	0	Group 5
	445731483	ELZ83067.1	Haloferax	elongans	0	0	Group 5
	491116063	WP_004974519.1	Haloferax	gibbonsii	0	0	Group 5
	445726991	ELZ78607.1	Haloferax	larsenii JCM 13917	0	0	Group 5
	445727170	ELZ78784.1	Haloferax	lucentense DSM 14919	0	0	Group 5
	490161608	WP_004060268.1	Haloferax	mediterranei	0	0	Group 5
	495596002	WP_008320581.1	Haloferax	mucosum	0	0	Group 5
	495370238	WP_008094951.1	Haloferax	prahovense	0	0	Group 5
	811259724	CQR49584.1	Haloferax	sp. Arc-Hr	0	0	Group 5
	445714096	ELZ65863.1	Haloferax	sp. ATCC BAA-644	0	0	Group 5
	432199583	ELK55744.1	Haloferax	sp. BAB2207	0	0	Group 5

Table 3A. Cont.

Phylum Euryarchaeota							
Order	GI	Accession	Genus	Species	cluster 1	cluster 2	Group Identification
Halobacteriales	445747833	ELZ99287.1	Haloferax	sulfurifontis	0	0	Group 5
	490142684	WP_004043026.1	Haloferax	volcanii	0	0	Group 5
	492944738	WP_006054069.1	Halogeometricum	borinquense	0	0	Group 5
	495658536	WP_008383115.1	Halogeometricum	pallidum	0	0	Group 5
	496825878	WP_009374959.1	Halogranum	salarium	0	0	Group 5
	399238157	EJN59086.1	Halogranum	salarium B-1	0	0	Group 5
	517069552	WP_018258370.1	Halomicrobium	katesii	0	0	Group 5
	506243765	WP_015763540.1	Halomicrobium	mukohataei	0	0	Group 5
	541197318	ERH05481.1	Halonotius	sp. J07HN4	0	0	Group 5
	503644206	WP_013878282.1	Halopiger	xanaduensis	0	0	Group 5
	541189943	ERG98331.1	Haloquadratum	sp. J07HQX50	0	0	Group 5
	499891421	WP_011572155.1	Haloquadratum	walsbyi	0	0	Group 5
	495799143	WP_008523722.1	Halorhabdus	tiamatea	0	0	Group 5
	506270475	WP_015790250.1	Halorhabdus	utahensis	0	0	Group 5
	495275191	WP_007999946.1	Halorubrum	aidingense	0	0	Group 5
	445818080	EMA67947.1	Halorubrum	arcis JCM 13916	0	0	Group 5
	495717573	WP_008442152.1	Halorubrum	californiense	0	0	Group 5
	493052949	WP_006112314.1	Halorubrum	coriense	0	0	Group 5
	445704112	ELZ56030.1	Halorubrum	distributum JCM 10118	0	0	Group 5
	495859120	WP_008583699.1	Halorubrum	hochstenium	0	0	Group 5
	496123800	WP_008848307.1	Halorubrum	kocurii	0	0	Group 5
	506390814	WP_015910533.1	Halorubrum	lacusprofundi	0	0	Group 5
	495283665	WP_008008419.1	Halorubrum	lipolyticum	0	0	Group 5
	445807636	EMA57719.1	Halorubrum	litoreum JCM 13561	0	0	Group 5
	490146119	WP_004046449.1	Halorubrum	saccharovorum	0	0	Group 5
	568635820	CDK38680.1	Halorubrum	sp. AJ67	0	0	Group 5
	515913811	WP_017344394.1	Halorubrum	sp. T3	0	0	Group 5
	493677827	WP_006628062.1	Halorubrum	tebenquichense	0	0	Group 5
	445680510	ELZ32953.1	Halorubrum	terrestre JCM 10247	0	0	Group 5
	493939513	WP_006883727.1	Halosimplex	carlsbadense	0	0	Group 5
	573481298	AHF99708.1	Halostagnicola	larsenii XH-48	0	0	Group 5
	635282901	KDE59326.1	Halostagnicola	sp. A56	0	0	Group 5
	495288157	WP_008012911.1	Haloterrigena	limicola	0	0	Group 5
	496170037	WP_008894544.1	Haloterrigena	salina	0	0	Group 5
	445656947	ELZ09779.1	Haloterrigena	thermotolerans	0	0	Group 5
	502708525	WP_012943705.1	Haloterrigena	turkmenica	0	0	Group 5

Table 3A. Cont.

Phylum Euryarchaeota							
Order	GI	Accession	Genus	Species	cluster 1	cluster 2	Group Identification
Halobacteriales	494978720	WP_007704744.1	Halovivax	asiaticus	0	0	Group 5
	505113880	WP_015300982.1	Halovivax	ruber	0	0	Group 5
	505222569	WP_015409671.1	Natronomonas	moolapensis	0	0	Group 5
	499642399	WP_011323133.1	Natronomonas	pharaonis	0	0	Group 5
Methanomassiliicoccales	521284377	WP_020448645.1	Candidatus Methanomassiliicoccus	intestinalis	4	3	Group 2
	505317145	WP_015504247.1	Candidatus Methanomethylophilus	alvus	1	0	Group 4
	731480881	AIZ56060.1	Candidatus Methanoplasma	termitum	4	1	Group 2
Methanosarcinales	630829973	KCZ71686.1	Candidatus Methanoperedens	nitroreducens	4	4	Group 1
	499817493	WP_011498227.1	Methanococcoides	burtonii	4	4	Group 1
	695945606	KGK99555.1	Methanococcoides	methylutens	4	4	Group 1
	502959317	WP_013194293.1	Methanohalobium	evestigatum	4	4	Group 1
	502802712	WP_013037688.1	Methanohalophilus	mahii	4	4	Group 1
	504865935	WP_015053037.1	Methanolobus	psychrophilus	4	4	Group 1
	564600724	WP_023845824.1	Methanolobus	tindarius	4	4	Group 1
	505138223	WP_015325325.1	Methanomethylovorans	hollandica	4	4	Group 1
	503483865	WP_013718526.1	Methanosaeta	concilii	4	4	Group 1
	504399265	WP_014586367.1	Methanosaeta	harundinacea	4	4	Group 1
	500016048	WP_011696766.1	Methanosaeta	thermophila	4	4	Group 1
	503664411	WP_013898487.1	Methanosalsum	zhilinae	4	4	Group 1
	805383795	AKB79672.1	Methanosarcina	horonobensis HB-1	4	4	Group 1
	805377871	AKB73943.1	Methanosarcina	lacustris Z-7289	4	4	Group 1
	499344550	WP_011034089.1	Methanosarcina	mazei	4	4	Group 1
	805357625	AKB37696.1	Methanosarcina	siciliae C2J	4	4	Group 1
	814874231	KKH49661.1	Methanosarcina	sp. 1.H.A.2.2	4	4	Group 1
	805395774	AKB46606.1	Methanosarcina	sp. Kolksee	4	4	Group 1
	805345072	AKB25912.1	Methanosarcina	sp. MTP4	4	4	Group 1
	805339923	AKB20948.1	Methanosarcina	sp. WH1	4	4	Group 1
	805336272	AKB17554.1	Methanosarcina	sp. WWM596	4	4	Group 1
	805334855	AKB16488.1	Methanosarcina	thermophila CHTI-55	4	4	Group 1
	805392031	AKB43129.1	Methanosarcina	vacuolata Z-761	4	4	Group 1
	300681098	P0CG28.1	Methanosarcina	acetivorans	4	4	Group 1
	499624383	WP_011305117.1	Methanosarcina	barkeri	4	4	Group 1
	757130681	WP_042684954.1	Methermicoccus	shengliensis	4	4	Group 1

Table 3A. Cont.

Phylum Euryarchaeota							
Order	GI	Accession	Genus	Species	cluster 1	cluster 2	Group Identification
Methanobacteriales	325329997	ADZ09059.1	Methanobacterium	lacus	4	4	Group 1
	333825789	AEG18451.1	Methanobacterium	paludis	4	4	Group 1
	410598430	EKQ53003.1	Methanobacterium	sp. Maddingley MBC34	4	4	Group 1
	557946297	CDG64904.1	Methanobacterium	sp. MB1	4	4	Group 1
	490129971	WP_004030361.1	Methanobacterium	formicicum	4	4	Group 1
	757147281	WP_042701481.1	Methanobrevibacter	arboriphilus	4	0	Group 2
	757140303	WP_042694545.1	Methanobrevibacter	oralis	4	1	Group 2
	502720726	WP_012955710.1	Methanobrevibacter	ruminantium	4	0	Group 2
	490135086	WP_004035446.1	Methanobrevibacter	smithii	4	1	Group 2
	509084877	AGN17123.1	Methanobrevibacter	sp. AbM4	4	2	Group 2
	757153864	WP_042707958.1	Methanobrevibacter	wolinii	4	1	Group 2
	499725722	WP_011406456.1	Methanosphaera	stadtmanae	4	0	Group 2
	503060369	WP_013295345.1	Methanothermobacter	marburgensis	4	4	Group 1
	427188567	BAM69285.1	Methanothermobacter	sp. CaT2	4	4	Group 1
	499178138	WP_010875678.1	Methanothermobacter	thermautotrophicus	4	4	Group 1
	503179301	WP_013413962.1	Methanothermus	fervidus	4	4	Group 1
Methanococcales	506271274	WP_015791049.1	Methanocaldococcus	fervens	0	0	Group 6
	502865135	WP_013100111.1	Methanocaldococcus	infernus	0	0	Group 6
	499172100	WP_010869687.1	Methanocaldococcus	jannaschii	0	0	Group 6
	502744715	WP_012979699.1	Methanocaldococcus	FS406-22	0	0	Group 6
	490729706	WP_004592097.1	Methanocaldococcus	villosus	0	0	Group 6
	506213931	WP_015733706.1	Methanocaldococcus	vulcanius	0	0	Group 6
	500684523	WP_011973807.1	Methanococcus	aeolicus	0	0	Group 6
	499484626	WP_011171266.1	Methanococcus	maripaludis	0	0	Group 6
	500678650	WP_011972442.1	Methanococcus	vannielii	0	0	Group 6
	502944629	WP_013179605.1	Methanococcus	voltae	0	0	Group 6
	503633084	WP_013867160.1	Methanothermococcus	okinawensis	0	0	Group 6
	750400916	WP_040682822.1	Methanothermococcus	thermolithotrophicus	0	0	Group 6
	494103325	WP_007044112.1	Methanotorris	formicicus	0	0	Group 6
	503564932	WP_013799008.1	Methanotorris	igneus	0	0	Group 6
Methanocellales	500970307	WP_012034945.1	Methanocella	arvoryzae	4	4	Group 1
	504219015	WP_014406117.1	Methanocella	conradii	4	4	Group 1
	502665036	WP_012901004.1	Methanocella	paludicola	4	4	Group 1

Table 3A. Cont.

Phylum Euryarchaeota							
Order	GI	Accession	Genus	Species	cluster 1	cluster 2	Group Identification
Methanomicrobiales	757145617	WP_042699825.1	Methanocorpusculum	bavaricum	4	4	Group 1
	500158946	WP_011833616.1	Methanocorpusculum	labreanum	4	4	Group 1
	504679381	WP_014866483.1	Methanoculleus	bourgensis	4	4	Group 1
	500170213	WP_011844638.1	Methanoculleus	marisnigri	4	4	Group 1
	524837464	CDF31353.1	Methanoculleus	sp. CAG:1088	1	0	Group 4
	635277975	KDE55429.1	Methanoculleus	sp. MH98A	4	4	Group 1
	490137557	WP_004037915.1	Methanofollis	liminatans	4	4	Group 1
	503093880	WP_013328703.1	Methanolacinia	petrolearia	4	4	Group 1
	494524364	WP_007313817.1	Methanolinea	tarda	4	4	Group 1
	757151336	WP_042705527.1	Methanomicrobium	mobile	4	4	Group 1
	490177901	WP_004076527.1	Methanoplanus	limicola	4	4	Group 1
	501056331	WP_012107796.1	Methanoregula	boonei	4	4	Group 1
	505097418	WP_015284520.1	Methanoregula	formicica	4	4	Group 1
	501694025	WP_012618985.1	Methanosphaerula	palustris	4	4	Group 1
	499769110	WP_011449844.1	Methanospirillum	hungatei	4	4	Group 1
Methanopyrales	499329350	WP_011019842.1	Methanopyrus	kandleri	2	0	Group 4
Natrialbales	493716495	WP_006666034.1	Natrialba	aegyptia	0	0	Group 5
	493046023	WP_006108385.1	Natrialba	asiatica	0	0	Group 5
	445644783	ELY97792.1	Natrialba	chahannaoensis	0	0	Group 5
	445642008	ELY95079.1	Natrialba	hulunbeirensis	0	0	Group 5
	490324485	WP_004213956.1	Natrialba	magadii	0	0	Group 5
	493878387	WP_006824724.1	Natrialba	taiwanensis	0	0	Group 5
	494169723	WP_007109452.1	Natrinema	altunense	0	0	Group 5
	445629988	ELY83258.1	Natrinema	gari JCM 14663	0	0	Group 5
	493192676	WP_006184685.1	Natrinema	pallidum	0	0	Group 5
	433305757	AGB31569.1	Natrinema	pellirubrum DSM 15624	0	0	Group 5
	397681963	AFO56340.1	Natrinema	sp. J7-2	0	0	Group 5
	493477541	WP_006432487.1	Natrinema	versiforme	0	0	Group 5
	491746571	WP_005579704.1	Natronobacterium	gregoryi	0	0	Group 5
	491712494	WP_005557953.1	Natronococcus	amylolyticus	0	0	Group 5
	495699887	WP_008424466.1	Natronococcus	jeotgali	0	0	Group 5
	505133951	WP_015321053.1	Natronococcus	occultus	0	0	Group 5
	494471810	WP_007261288.1	Natronolimnobius	innermongolicus	0	0	Group 5
	492956238	WP_006064323.1	Natronorubrum	bangense	0	0	Group 5
	495434800	WP_008159495.1	Natronorubrum	sulfidifaciens	0	0	Group 5
	493008974	WP_006088587.1	Natronorubrum	tibetense	0	0	Group 5

Table 3A. Cont.

Phylum Euryarchaeota							
Order	GI	Accession	Genus	Species	cluster 1	cluster 2	Group Identification
Thermococcales	664800581	AIF68814.1	Palaeococcus	pacificus DY20341	0	0	Group 5
	499169378	WP_010867654.1	Pyrococcus	abyssi	0	0	Group 5
	499322302	WP_011012794.1	Pyrococcus	furiosus	0	0	Group 5
	499188167	WP_010885707.1	Pyrococcus	horikoshii	0	0	Group 5
	331033327	AEC51139.1	Pyrococcus	sp. NA2	0	0	Group 5
	388250246	AFK23099.1	Pyrococcus	sp. ST04	0	0	Group 5
	503670933	WP_013905009.1	Pyrococcus	yayanosii	0	0	Group 5
	507915398	AGN00996.1	Salinarchaeum	sp. Harcht-Bsk1	0	0	Group 5
	503231745	WP_013466406.1	Thermococcus	barophilus	0	0	Group 5
	390519068	AFL94800.1	Thermococcus	sp. CL1	0	0	Group 5
	390960760	YP_006424594.1	Thermococcus	cleftensis	0	0	Group 5
	700302974	AIU69793.1	Thermococcus	eurythermalis	0	0	Group 5
	506339853	WP_015859572.1	Thermococcus	gammatolerans	0	0	Group 5
	499569671	WP_011250454.1	Thermococcus	kodakarensis	0	0	Group 5
	490170111	WP_004068752.1	Thermococcus	litoralis	0	0	Group 5
	589910733	AHL23782.1	Thermococcus	nautili	0	0	Group 5
	501566623	WP_012571063.1	Thermococcus	onnurineus	0	0	Group 5
	573024089	AHF79623.1	Thermococcus	paralvinellae	0	0	Group 5
	506328909	WP_015848628.1	Thermococcus	sibiricus	0	0	Group 5
	340809685	AEK72842.1	Thermococcus	sp. 4557	0	0	Group 5
	214033184	EEB74012.1	Thermococcus	sp. AM4	0	0	Group 5
	757147190	WP_042701390.1	Thermococcus	sp. PK	0	0	Group 5
	498164860	WP_010479016.1	Thermococcus	zilligii	0	0	Group 5
Phylum Thaumarchaeota							
Nitrosopumilales	494644810	WP_007402754.1	Candidatus Nitrosoarchaeum	limnia	0	0	Group 6
	407045575	AFS80328.1	Candidatus Nitrosopumilus	koreensis AR1	0	0	Group 6
	495576771	WP_008301350.1	Candidatus Nitrosopumilus	salaria	0	0	Group 6
	407047503	AFS82255.1	Candidatus Nitrosopumilus	sp. AR2	0	0	Group 6
	756793731	AJM93144.1	Candidatus Nitrosopumilus	sp. D3C	0	0	Group 6
	770480626	AJW71423.1	Candidatus Nitrosopumilus	sp. NF5	0	0	Group 6
	501170936	WP_012214811.1	Nitrosopumilus	maritimus	0	0	Group 6
	648398193	WP_026089944.1	Nitrosopumilus	sp. AR2	0	0	Group 6
Nitrososphaerales	665993646	AIF83406.1	Candidatus Nitrososphaera	evergladensis SR1	0	0	Group 6
	504832748	WP_015019850.1	Candidatus Nitrososphaera	gargensis	0	0	Group 6
	647810504	AIC14205.1	Nitrososphaera	viennensis EN76	0	0	Group 6

Table 3A. Cont.

Phylum Thaumarchaeota							
Order	GI	Accession	Genus	Species	cluster 1	cluster 2	Group Identification
Cenarchaeales	503248130	WP_013482791.1	Cenarchaeum	symbiosum	0	0	Group 6
Phylum Nanoarchaeota							
	490715555	WP_004578204.1	Nanoarchaeote	Nst1	0	0	Group 5
	499466795	WP_011153435.1	Nanoarchaeum	equitans Kin4-M	0	0	Group 5
Phylum Korarchaeota							
	501267206	WP_012310224.1	Candidatus Korarchaeum	cryptofilum	2	4	Group 3
Phylum Lokiarchaeota							
	816395085	KKK46447.1	Lokiarcheota		4	4	Group 1
Phylum Parvarchaeota							
	255513546	EET89812.1	Candidatus Micrarchaeum	acidiphilum ARMAN-2	0	0	Group 6
	269986616	EEZ92898.1	Candidatus Parvarchaeum	acidiphilum ARMAN-4	4	0	Group 2
	290559364	EFD92697.1	Candidatus Parvarchaeum	acidophilus ARMAN-5	4	0	Group 2

Table 3B. D3/FLD group totals by phylum and percentages by group.

Phylum	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Total
Crenarchaeota	9	14	1	25	0	0	49
Euryarchaeota	58	9	0	8	119	14	208
Thaumarchaeota	0	0	0	0	0	12	12
Korarchaeota	0	0	1	0	0	0	1
Nanoarchaeota	0	0	0	0	2	0	2
Lokiarchaeota	1	0	0	0	0	0	1
Parvarchaeota	0	2	0	0	0	1	3
Total	68	25	2	33	121	27	276
Percentage	24.6	9.1	0.7	12.0	43.8	9.8	

Table 4. Summary of the conservation of D3/FLD in subunit D from archaea.

Phylum/Order	D3/FLD	[4Fe-4S] cluster binding motif	
		Cluster #1	Cluster #2
Crenarchaeota			
Desulfurococcales	Yes	X[a]	X
Sulfolobales	Yes	X	
Thermoproteales	Yes		
Euryarchaeota			
Acidilobales	Yes		
Fervidicoccales	Yes	X	
Thermoplasmatales	Yes		
Archaeoglobales	Yes	X	X
Halobacteriales	Yes		
Methanomassiliicoccales	Yes	X	
Methanosarcinales	Yes	X	X
Methanobacteriales[b]	Yes	X	X
Methanococcales	No		
Methanocellales	Yes	X	X
Methanomicrobiales	Yes	X	X
Methanopyrales	Yes		
Natrialbales	Yes		
Thermococcales	Yes		
Thaumarchaeota			
Nitrosopumilales	No		
Nitrososphaerales	No		
Cenarchaeales	No		
Nanoarchaeota	Yes		
Korarchaeota	Yes		X
Parvarchaeota	Yes	X	
Lokiarchaeota	Yes	X	X

[a] an X designates at least 60% of sequences belonging to species within the order possess the binding motif

[b] Subunit D in all sequenced Methanobacteriales contains cluster #1, but 44% lack cluster #2.

Table 5. Iron and sulfide content of purified recombinant D-L(His) heterodimers.

D/L heterodimer	Iron[a]	Sulfide[b]
D-L(His)[c]	8.3 ± 0.2	7.7 ± 0.8
DΔD3-L(His)[c]	BDL	BDL
DΔFeS1-L(His)	4.6 ± 0.1	3.3 ± 0.2
DmFeS1-L(His)	4.7 ± 0.2	3.5 ± 0.7
DΔFeS2-L(His)	3.8 ± 0.1	3.0 ± 0.1
DmFeS2-L(His)	4.9 ± 0.2	3.3 ± 0.2

[a]nmol of iron/nmol of D-L heterodimer
[b]nmol of sulfide/nmol of D-L heterodimer
[c]results from reference (Lessner *et al.*, 2012)

Table 6. *M. acetivorans* strains utilized in this study.

Strain designation	Relevant Genotype
DJL30[a]	*rpoD* merodiploid: contains tetracycline inducible (His)D
DJL31[a]	*rpoD* merodiploid: contains tetracycline inducible (His)DΔD3
DJL32	*rpoD* merodiploid: contains tetracycline inducible (His)DΔFeS1
DJL33	*rpoD* merodiploid: contains tetracycline inducible (His)DΔFeS2
DJL34	*rpoD* merodiploid: contains tetracycline inducible (His)DmFeS2
DJL35	*rpoD* merodiploid: contains tetracycline inducible (His)DmFeS1
DJL40	*rpoD* merodiploid: contains tetracycline inducible (His)DΔCterm
DJL51	Native *rpoD* replaced with *rpoD* encoding (His)D
DJL52	Native *rpoD* replaced with *rpoD* encoding (His)DΔFeS1
DJL54	Native *rpoD* replaced with *rpoD* encoding (His)DΔD3
DJL55	Native *rpoD* replaced with *rpoD* encoding (His)DmFeS2

[a]previously generated (Lessner *et al.*, 2012)

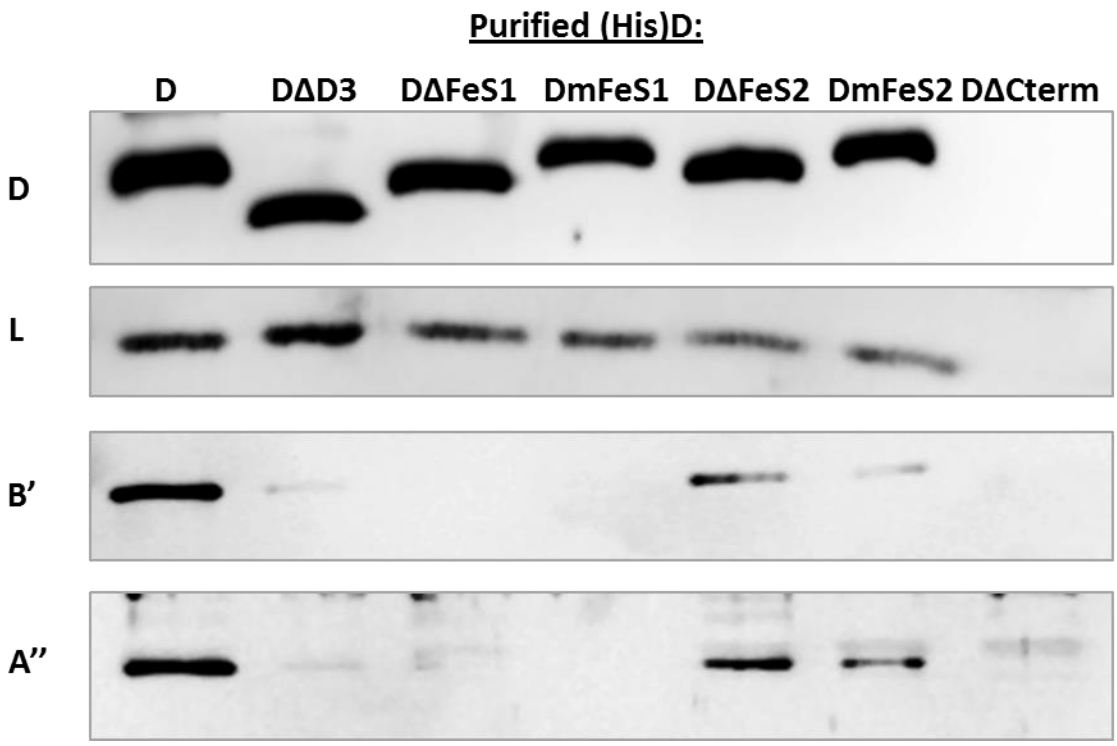

Fig 2. Analysis of the co-purification of endogenous subunits L, B', and A" with (His)D variants expressed in *M. acetivorans*. Separate SDS-PAGE gels were loaded with samples of each imidazole eluate containing 15 ng of each (His)D. The gels were analyzed by Western blot with antibodies specific for subunit D, L, B' or A". For the imidazole eluate from the (His)DΔCterm purification the gels were loaded with 1.5 μg of total protein since (His)DΔCterm was not detected in the eluate.

Table 7. Purification of (His)D from merodiploid strains of *M. acetivorans*.

Strain/D subunit	D concentration in eluate	Non-specific transcription assay activity[a]
DJL30/(His)D	15.9 ng/μL	2668 ± 412
DJL31/(His)DΔD3	7.3 ng/μL	BDL[b]
DJL32/ (His)DΔFeS1	6.7 ng/μL	BDL[b]
DJL35/(His)DmFeS1	1.2 ng/μL	BDL[b]
DJL33/(His)DΔFeS2	7.3 ng/μL	BDL[b]
DJL34/(His)DmFeS2	0.8 ng/μL	BDL[b]

[a]counts per minute μg^{-1} subunit D
[b]Below detection limit of assay

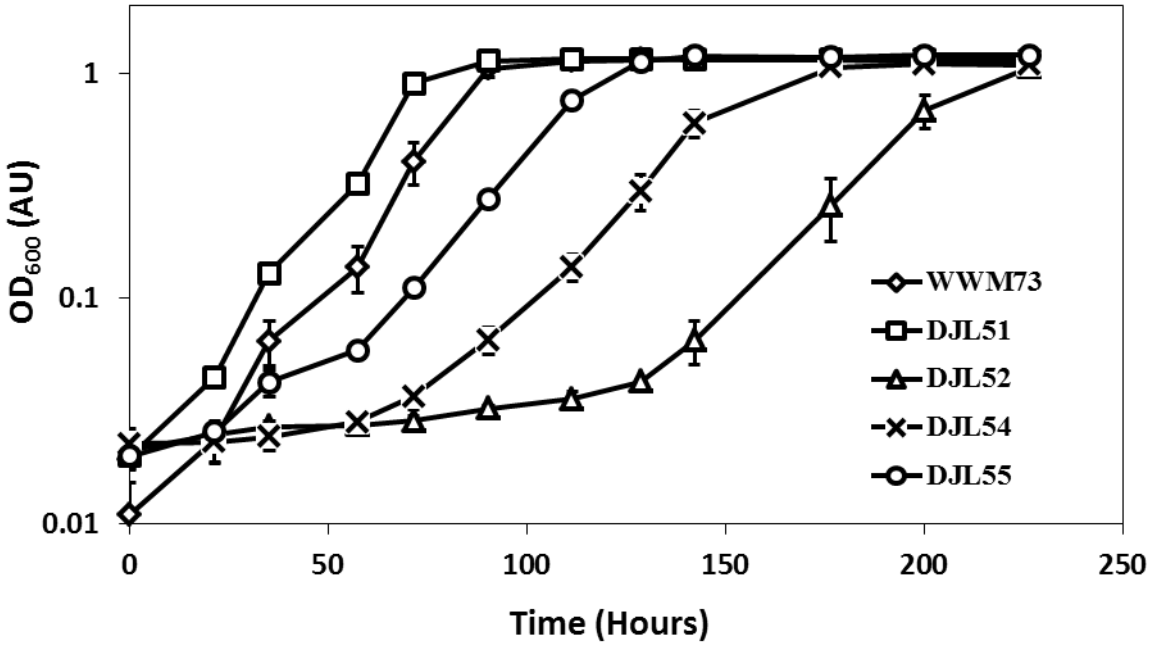

Fig. 3. Comparison of the growth of *M. acetivorans* mutants with methanol. Data points represent the mean ± standard deviation of triplicate cultures.

Table 8. Growth parameters of *M. acetivorans* strains with methanol.

Strain/D subunit	Lag phase duration (hours)	Generation time (hours)	Maximum OD_{600}
WWM73/WT-D	≤ 20	9.9 ± 0.7	1.2 ± 0.03
DJL51/(His)D	≤ 20	9.4 ± 0.5	1.2 ± 0.05
DJL54/(His)DΔD3	≥ 85	16.5 ± 0.3*	1.1 ± 0.1
DJL52/ (His)DΔFeS1	≥ 135	17.1 ± 0.4*	1.1 ± 0.02
DJL55/(His)DmFeS2	≥ 60	14.5 ± 0.2*	1.2 ± 0.03

Values are averages ± standard deviation of triplicate cultures

*Significant difference in generation time versus DJL51 ($P < 0.05$)

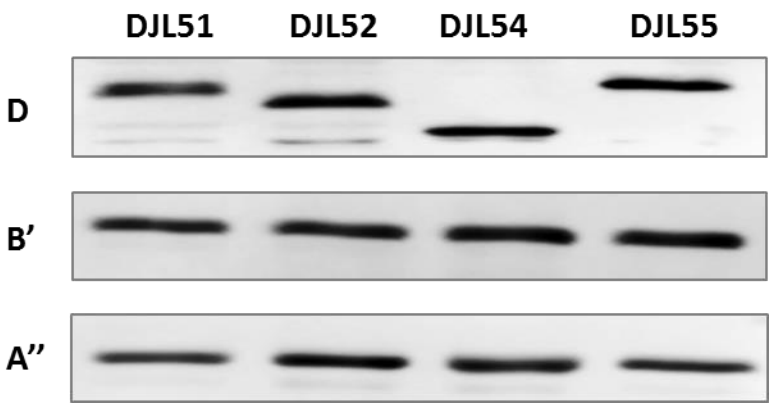

Fig. 4. Comparison of the levels of RNAP subunits D, B', and A" in cell lysate from *M. acetivorans* mutants by Western blot. Separate SDS-PAGE gels were loaded with samples containing 20 μg of total protein (D blot), or normalized to contain 15 ng of subunit D (B' and A" blots).

Table 9. RNAP activity in lysate from *M. acetivorans rpoD* replacement strains.

Strain/D subunit	D concentration in lysate	Non-specific RNAP transcription assay activity[a]
DJL51/(His)D	3.9 ng/µL	8182 ± 165
DJL54/(His)DΔD3	5.5 ng/µL	2216 ± 128 (27%[b])*
DJL52/ (His)DΔFeS1	5.2 ng/µL	2765 ± 253 (34% [b])*
DJL55/(His)DmFeS2	6.5 ng/µL	4505 ± 586 (55% [b])*

[a]counts per minute ng^{-1} subunit D

[b]Percent activity of (His)D lysate

*Significant difference in activity versus DJL51 lysate ($P < 0.05$)

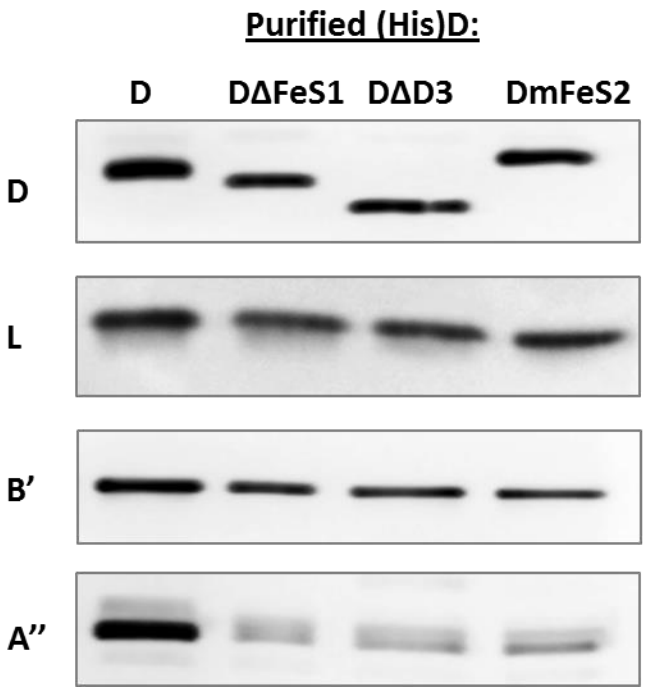

Fig. 5. Analysis of the co-purification of subunits L, B', and A" with (His)D variants purified from *M. acetivorans* mutant strains. Separate SDS-PAGE gels were loaded with samples of each imidazole eluate normalized to subunit D; 15 ng (D blot), 20 ng (L blot), 10 ng (B' blot), and 15 ng (A" blot).

Table 10. Purification (His)D from *rpoD* replacement strains of *M. acetivorans.*

Strain/D subunit	D concentration in eluate	Non-specific RNAP transcription assay activity[a]
DJL51/(His)D	6.7 ng/μL	210 ± 14
DJL54/(His)DΔD3	6.6 ng/μL	90 ± 20 (43% [b])*
DJL52/ (His)DΔFeS1	9.8 ng/μL	33 ± 9 (16% [b])*
DJL55/(His)DmFeS2	5.5 ng/μL	94 ± 10 (45% [b])*

[a] counts per minute ng^{-1} subunit D

[b]Percent activity of (His)D lysate

*Significant difference in activity versus DJL51 eluate ($P < 0.05$)

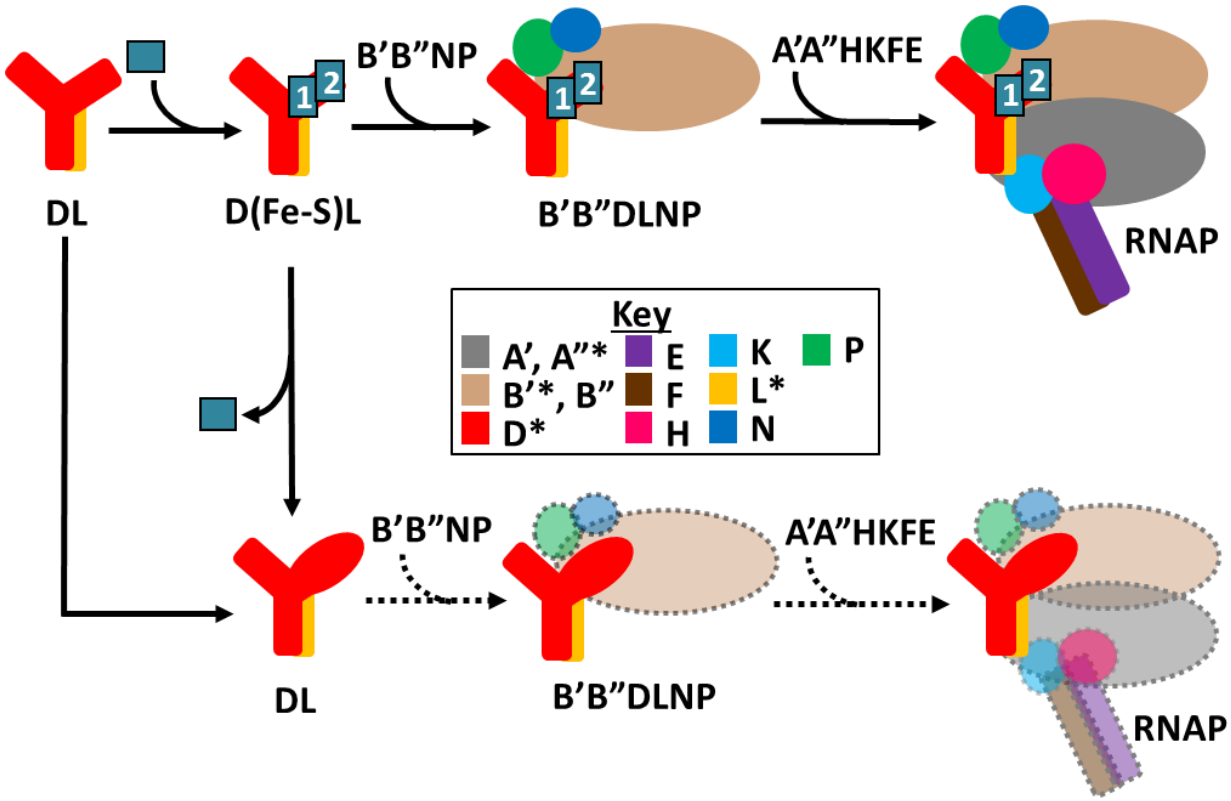

Fig. 6. Model of the impact of the [4Fe-4S] clusters in the FLD of subunit D on the assembly of RNAP in *M. acetivorans*. Assembly of the two [4Fe-4S] clusters (boxes labeled 1 and 2) occurs after formation of the D-L heterodimer. The lack of cluster incorporation into or cluster loss from the D-L heterodimer negatively impacts assembly and/or stability B'B"DLNP subcomplex. The absence of the cluster(s) alters the conformation of the B'B"DLNP subcomplex and impacts the association with subunit A', decreasing assembly and/or stability of complete RNAP. Decreased assembly and/or stability is indicated by the transparent subunits and dashed lines.

Appendix 2.1: Lead Author Confirmation Letter for Chapter II

J. William Fulbright College of Arts and Sciences
Department of Biological Sciences

Chapter II, titled "The [4Fe-4S] clusters of subunit D are key determinants in the post D-L heterodimer assembly of RNA polymerase in *Methanosarcina acetivorans*" of M. E. Jennings's dissertation was submitted for publication in Molecular Microbiology in 2016 with coauthors E. A. Karr, F.H. Lessner, and D.J. Lessner.

I, Dr. Daniel J. Lessner, advisor of Matthew Edward Jennings, confirm Matthew Edward Jennings was first author and completed at least 51% of the work for this manuscript.

Dr. Daniel J. Lessner
Associate Professor
Department of Biological Sciences
University of Arkansas

Date

Science Engineering, Room 601 · Fayetteville, AR 72701-1201 · 479-575-3251 · Fax: 575-4010

www.uark.edu

Chapter III

Expression of a bacterial catalase in a strictly anaerobic methanogen significantly increases tolerance to hydrogen peroxide but not oxygen

Matthew E. Jennings

Department of Biological Sciences, University of Arkansas,

Fayetteville, AR 72701 USA

Abstract

Haem-dependent catalase is an antioxidant enzyme that degrades H_2O_2, producing H_2O and O_2, and is common in aerobes. Catalase is present in some strictly anaerobic methane-producing archaea (methanogens), but the importance of catalase to the antioxidant system of methanogens is poorly understood. We report here that a survey of the sequenced genomes of methanogens revealed that the majority of species lack genes encoding catalase. Moreover, *Methanosarcina acetivorans* is a methanogen capable of synthesizing haem and encodes haem-dependent catalase in its genome; yet, *Methanosarcina acetivorans* cells lack detectable catalase activity. However, inducible expression of the haem-dependent catalase from *Escherichia coli* (EcKatG) in the chromosome of *Methanosarcina acetivorans* resulted in a 100-fold increase in the endogenous catalase activity compared with uninduced cells. The increased catalase activity conferred a 10-fold increase in the resistance of EcKatG-induced cells to H_2O_2 compared with uninduced cells. The EcKatG-induced cells were also able to grow when exposed to levels of H_2O_2 that inhibited or killed uninduced cells. However, despite the significant increase in catalase activity, growth studies revealed that EcKatG-induced cells did not exhibit increased tolerance to O_2 compared with uninduced cells. These results support the lack of catalase in the majority of methanogens, since methanogens are more likely to encounter O_2 rather than high concentrations of H_2O_2 in the natural environment. Catalase appears to be a minor component of the antioxidant system in methanogens, even those that are aerotolerant, including *Methanosarcina acetivorans*. Importantly, the experimental approach used here demonstrated the feasibility of engineering beneficial traits, such as H_2O_2 tolerance, in methanogens.

Introduction

Methane-producing archaea (methanogens) are strictly anaerobic microorganisms, which are only capable of growth by methanogenesis. Methanogenesis requires specific coenzymes and enzymes, many of which contain metal cofactors (such as Fe-S clusters) and function at low redox potentials (Thauer *et al.*, 2008). Exposure of anaerobes to molecular oxygen (O_2) results in autoxidation of cellular components, including flavoenzymes and metalloenzymes, leading to the production of reactive oxygen species (ROS), including superoxide (O_2^-) and hydrogen peroxide (H_2O_2). O_2 and the produced ROS cause oxidative damage of enzymes, cofactors, coenzymes, and general macromolecules, and may ultimately lead to cell death (Imlay, 2002). Thus, methanogens are sensitive to O_2, and are only capable of growing and producing methane under anaerobic conditions. Nonetheless, methanogens are transiently exposed to O_2, which would necessitate antioxidant and repair enzymes to facilitate O_2 tolerance (Angel *et al.*, 2011; Angel *et al.*, 2012; Fetzer *et al.*, 1993). Indeed, the majority of methanogen species can tolerate O_2 exposure, and there is evidence that some methanogens associated with termite guts can produce methane in the presence of low levels of O_2 (Tholen *et al.*, 2007). A detailed understanding of the molecular mechanisms underlying the oxidant tolerance of methanogens is limited. Methanogens are of significant environmental and biotechnological importance, and an understanding of the antioxidant mechanisms may lead to development of methods to enhance or inhibit methanogenesis.

Recent evidence suggests that strictly anaerobic bacteria and archaea contain antioxidant enzymes that differ from those found in aerobes and facultative organisms. For example, superoxide dismutase (SOD) and catalase are prevalent in aerobes and facultative microbes, but are found less frequently in strict anaerobes. It is hypothesized that anaerobes lack SOD and catalase because each enzyme produces O_2 as an end product, which would serve to further

propagate the production of ROS in anaerobes (Imlay, 2002). Instead, strict anaerobes contain enzymes such as superoxide reductase, peroxidase, and rubrerythrin, which degrade O_2^- and H_2O_2 without producing O_2 (Jenney *et al.*, 1999; Lumppio *et al.*, 2001). Nonetheless, there is evidence that catalase is part of the antioxidant system in some anaerobes. Catalase has been shown to be important to the tolerance to O_2 and H_2O_2 by some sulfate-reducing bacteria, *Bacteroides* spp., and acetogens (Brioukhanov & Netrusov, 2004). There is limited evidence that indicates catalase contributes to the oxidant tolerance of methanogens. *Methanobrevibacter arboriphilus*, a methanogen which lacks cytochromes and is incapable of synthesizing heme, surprisingly possesses a heme-dependent catalase (Shima *et al.*, 2001). *M. arboriphilus* is thought to acquire heme from the external environment and convert apo-catalase to the active form. Supplementation of growth medium with hemin results in a substantial increase in *M. arboriphilus* intracellular catalase activity, which increases the tolerance of this methanogen to H_2O_2 and O_2 (Brioukhanov & Netrusov, 2012). *Methanosarcina barkeri*, a cytochrome-containing species that is capable of synthesizing heme, also contains intracellular catalase activity due to a heme-dependent catalase (Shima *et al.*, 1999). *M. barkeri* and *M. arboriphilus* are both tolerant to O_2 and millimolar levels of H_2O_2. The catalase gene in *M. barkeri* is transcriptionally up-regulated upon exposure of cells to sub-lethal concentrations of H_2O_2 (Brioukhanov & Netrusov, 2004). Recent evidence also reveals that methanogens are prevalent in aerated soils and once anaerobic conditions are restored, methanogenesis ensues (Angel *et al.*, 2011). Among methanogens, members of the genera *Methanosarcina* and *Methanocella* dominate aerated soils, indicating these methanogens are extremely aerotolerant (Angel *et al.*, 2012). Moreover, active transcription of the gene encoding a catalase (*katE*) was identified in

samples of aerated soils, indicating that catalase is a potential component of the antioxidant system in aerotolerant *Methanosarcina* and *Methanocella* (Angel *et al.*, 2011).

To ascertain the importance of catalase to the antioxidant system of methanogens, we are using the cytochrome-containing species *Methanosarcina acetivorans* as a model, because it is aerotolerant, its metabolism has been extensively investigated, and it has a robust genetic system (Ferry & Lessner, 2008; Guss *et al.*, 2008; Horne & Lessner, 2013; Lessner *et al.*, 2006). We have recently developed methods to assess the viability *M. acetivorans* after challenge with oxidants. Importantly, the exogenous addition of catalase conferred a significant increase in the tolerance to H_2O_2, indicating catalase could be important to the oxidant tolerance of *M. acetivorans* (Horne & Lessner, 2013). The goal of the present study was to examine the prevalence of catalase in methanogens and to ascertain the importance of endogenous catalase to the H_2O_2 and O_2 tolerance of *M. acetivorans*. We also wanted to test the feasibility of engineering methanogen strains with increased oxidant tolerance. Therefore, we used a novel approach which employed the controlled expression of a bacterial catalase within *M. acetivorans*. Because the expression of the bacterial catalase could be tightly controlled, the approach used allowed for the specific assessment of the importance of catalase activity to the oxidant tolerance of *M. acetivorans*.

Materials and Methods

Growth of *Methanosarcina acetivorans.* *M. acetivorans* strains were grown in high-salt (HS) medium supplemented with 125 mM methanol as a carbon and energy source, and 0.025 % Na_2S as a reductant as previously described (Sowers *et al.*, 1984). Growth was monitored spectrophotometrically as optical density at 600 nm (OD_{600}) using a Genesys 10 Bio

spectrophotometer (Thermo Scientific). The inducer tetracycline was added to a final concentration of 100 μg ml^{-1} where indicated.

Construction of an EcKatG-expression strain of *M. acetivorans*. PCR was used to amplify *katG* from *E. coli* DH5α genomic DNA. The forward primer for the amplification contained the sequence for an *Nde*I restriction site (5'-GGTGGTCATATGAGCACGTCAGACGATATCCATAAC -3'), while the reverse primer contained a *Hind*III restriction site (5'-GGGGTAAGCTTTTACAGCAGGTCGAAACGGTCGAGG-3'). The PCR product was digested with *Nde*I and *Hind*III and ligated with similarly digested pJK027A (Guss *et al.*, 2008), generating plasmid pDL329. pDL329 contains *E. coli katG* fused to the $P_{mcrB(tetO1)}$ promoter in pJK027A. *M. acetivorans* strain WWM73 was transformed with pDL329 and transformants selected as previously described (Guss *et al.*, 2008). Successful integration of the plasmid into the chromosome of strain WWM73 was determined as described (Guss *et al.*, 2008), and the resulting strain was named DJL20. *M. acetivorans* strain DJL20 is capable of tetracycline-inducible expression of EcKatG.

Determination of Catalase activity. Cell lysates of *M. acetivorans* strains were prepared by harvesting cells (50 ml) by centrifugation at 16,000 x *g* at 4 °C. The cell pellets were frozen at -20 °C, thawed, and resuspended in 0.5 ml of 50 mM Tris-HCl pH 8.0. The cells were lysed by sonication and clarified lysates were obtained by centrifugation at 16,000 x *g* at 4 °C. The catalase activity in cell lysates was determined spectrophotometrically (Beckman DU-7400) by monitoring the decrease in absorbance at 240 nm of 13 mM H_2O_2 in 50 mM Tris-HCl, pH 8.0.

The amount of H_2O_2 consumed was determined using $\varepsilon_{240} = 39.4\ M^{-1}\ cm^{-1}$. One unit of activity is defined as 1 µmol of H_2O_2 consumed min^{-1}. The protein concentration of cell lysates was determined using the Bradford method (Bradford, 1976).

Oxidant challenge of *M. acetivorans*. The tolerance of *M. acetivorans* cells to H_2O_2 was assessed using recently developed methods (Horne & Lessner, 2013). Specifically, cells of strain DJL20 were challenged with various concentrations of H_2O_2 for one hour under anaerobic conditions and viability determined using the microtiter-plate method (Horne & Lessner, 2013). To assess the ability of cells to grow when challenged with H_2O_2 or O_2, *M. acetivorans* strains were grown in 10 ml of HS medium, devoid of sulfide and resazurin to avoid the abiotic reduction of H_2O_2 and O_2 by sulfide and the interference of oxidized resazurin with OD_{600} measurements. Mid-exponential phase cultures were challenged by the direct addition of H_2O_2 or O_2 to the culture tubes. Solutions of H_2O_2 were freshly prepared and 0.2 ml aliquots were added to mid-exponential phase cells using a syringe and needle. Pure O_2 was added to the desired percentage (vol/vol) of the headspace volume of each tube by using a syringe and needle. To promote uniform O_2 exposure, the tubes were incubated on the side and then mixed by inverting the tubes every hour.

Statistical analysis. Viability determinations by microtiter-plate method were independently replicated a minimum of three times. Growth inhibition experiments and catalase activity assays were replicated three times. All data are presented as mean ± SD. Plotting and calculation of the standard deviation were performed in Microsoft Excel®.

Results

Distribution of monofunctional catalase (KatE) and catalase-peroxidase (KatG) in methanogens. Heme-containing monofunctional catalases (KatE) have catalatic activity, but not peroxidatic activity, and are widely distributed within the three domains of life (Zamocky *et al.*, 2008). Using *E. coli* KatE as a query, homologous proteins encoded in the genomes of several methanogens were identified (Table 1). Of the 57 sequenced methanogen genomes currently within the database (http://img.jgi.doe.gov), only 11 contain a putative KatE homolog, comprising approximately 19% of the genomes. The presence of KatE is not restricted to certain methanogen orders or genera. However, all sequenced species within the order Methanococcales lack a putative KatE, whereas KatE is more prevalent in species of the orders Methanomicrobiales and Methanosarcinales. KatE from *Methanosarcina barkeri* and *Methanobrevibacter arboriphilus* have been characterized with each demonstrated to have catalatic activity (Brioukhanov & Netrusov, 2004; Brioukhanov & Netrusov, 2012; Shima *et al.*, 1999; Shima *et al.*, 2001).

Heme-containing bifunctional catalases have catalatic activity and peroxidatic activity and are also distributed within the three domains of life (Passardi *et al.*, 2007). Of the 57 sequenced methanogen genomes within the database (http://img.jgi.doe.gov), only 9 contain a putative KatG homolog, approximately 16% of the genomes (Table 1). KatG appears restricted to species within the orders Methanobacteriales, Methanomicrobiales, and Methanosarcinales. A KatG homolog from a methanogen has not been experimentally characterized. Only two methanogens possess a putative KatE and KatG, *Methanolobus psychrophilus* R15 and *M. acetivorans*, both members of the Methanosarcinales.

The genome of *M. acetivorans* encodes catalase, but cells lack catalase activity. The genome of *M. acetivorans* encodes a KatG (MA0972) homolog and a non-functional KatE (MA2081) homolog. Although the amino acid sequence of the putative KatE from *M. acetivorans* is 88% identical to characterized *M. barkeri* KatE (Shima *et al.*, 1999), sequencing of *ma2081* confirms that the gene contains a frameshift causing a nonsense mutation which results in synthesis of a truncated protein (173 of 496 predicted amino acids). Since half of the active site residues identified in KatE (Diaz *et al.*, 2012) are missing in the truncated protein (Fig. 1), *M. acetivorans* likely does not possess a functional KatE, unlike *M. barkeri*.

M. acetivorans KatG (MaKatG) is 77% identical to *Burkholderia pseudomallei* KatG, for which the structure has been solved (Carpena *et al.*, 2003) and 64% identical to well-characterized *E. coli* KatG (Diaz *et al.*, 2012). Moreover, the active site residues are conserved in MaKatG (Fig. 2), indicating MaKatG may have catalatic and peroxidatic activities. MaKatG has not been detected in several proteomic analyses of *M. acetivorans* (Lessner *et al.*, 2006; Li *et al.*, 2005a; Li *et al.*, 2005b), suggesting that it may not be expressed or is expressed at a low level under non-stress conditions. The level of catalase activity in lysates derived from wild-type *M. acetivorans* cells grown with methanol, harvested at mid-exponential phase or stationary phase, was below the detection limit. The lack of catalase activity suggests MaKatG is not constitutively expressed and is not induced upon entry into stationary phase. Since catalase activity in *M. barkeri* was shown to be induced by the addition of ROS (Brioukhanov *et al.*, 2006), catalase activity was also measured in lysates from *M. acetivorans* cells exposed to sub-lethal concentrations of H_2O_2 (1.5 mM) and O_2 (5%) for one or four hours. The level of catalase activity in H_2O_2- or O_2- challenged *M. acetivorans* cells was also below the detection limit. Finally, to ascertain if MaKatG encodes a functional catalase, MaKatG was expressed in *E. coli*

and catalase activity measure in *E. coli* lysates. *E. coli* cells expressed soluble MaKatG, indicative of proper folding, but did not exhibit an increase in catalase activity (data not shown), suggesting MaKatG lacks catalase activity. Overall, these results suggest MaKatG is non-functional and therefore does not serve a role in the antioxidant system in *M. acetivorans*. However, we have previously demonstrated that exogenous addition of catalase does protect *M. acetivorans* from H_2O_2 toxicity (Horne & Lessner, 2013), indicating catalase is a potential protective enzyme for *M. acetivorans*. Therefore, we set out to construct a *M. acetivorans* strain capable of inducible expression of recombinant catalase from a bacterium in order to determine if endogenous catalase can protect *M. acetivorans* from H_2O_2 and O_2 toxicity and whether it is possible to engineer a methanogen strain with increased oxidant tolerance.

Expression of recombinant E. coli KatG increases endogenous catalase activity in *M. acetivorans*. To specifically assess the ability of KatG to protect *M. acetivorans* from oxidants, the gene encoding KatG from *E. coli* (EcKatG) was fused to a tetracycline-inducible methanogen promoter ($P_{mcrBTetO1}$), and the gene fusion ($P_{mcrBTetO1}$-*EckatG*) was moved into the chromosome of *M. acetivorans* strain WWM73 (Guss *et al.*, 2008). The resulting strain, DJL20, exhibited similar growth rates and yields when grown with methanol in the presence or absence of tetracycline (data not shown), indicating induction of heme-dependent EcKatG is not inhibitory or detrimental to *M. acetivorans*. Lysate derived from mid-exponential phase cells of DJL20 grown in the presence of tetracycline (EcKatG-induced) exhibited a 100-fold increase in catalase activity (88 ± 13 U mg protein^{-1}), compared to activity in cells grown in the absence of tetracycline (uninduced), which was close to the detection limit (0.8 ± 0.2 U mg protein^{-1}). The induced catalase activity (~90 U mg protein^{-1}) in *M. acetivorans* strain DJL20 is comparable to

the intrinsic catalase activity (5-300 U mg protein^{-1}) observed in other anaerobes (Brioukhanov & Netrusov, 2004). These results show that EcKatG is expressed in an active form within *M. acetivorans*, supporting that the machinery in *M. acetivorans* recognizes EcKatG as a hemoprotein and facilitates the proper incorporation of heme into the active site of EcKatG. Importantly, to our knowledge this is the first example of the heterologous expression of a bacterial heme-dependent enzyme within a methanogen, indicating that methanogens (archaea) and bacteria likely contain similar mechanisms for heme incorporation.

An increase in endogenous catalase activity confers increased resistance of *M. acetivorans* to H_2O_2. The effect of increased catalase activity in strain DJL20 on the tolerance to H_2O_2 was assessed. Strain DJL20 was grown with methanol in the presence or absence of tetracycline to mid-exponential phase. Harvested cells were challenged with H_2O_2 for one hour and viability was assessed using a recently developed microtiter assay (Horne & Lessner, 2013). EcKatG-induced cells exhibited no loss of viability when challenged with a concentration of H_2O_2 (6 mM) that is lethal to uninduced cells (Fig. 3A). These results demonstrate that expression of EcKatG within *M. acetivorans* not only increases endogenous catalase activity, but that the cells are protected from the toxic effects of H_2O_2. Using the microtiter assay, the maximum dose of H_2O_2 resulting in a complete loss of viability of EcKatG-induced cells was 60 mM H_2O_2 (Fig. 3B), a 10-fold increase in the lethal H_2O_2 concentration over the uninduced cells. Thus, expression of EcKatG increases the tolerance of *M. acetivorans* to high concentrations of H_2O_2.

The effect of H_2O_2 on actively growing cultures of *M. acetivorans* was also examined. In order to test the intrinsic tolerance of strain DJL20 to H_2O_2, the cells were grown in medium devoid of sulfide, because we have shown that sulfide can protect *M. acetivorans* from H_2O_2

(Horne & Lessner, 2013). *M. acetivorans* exhibits a similar doubling time (~9 h) when grown with methanol in medium containing or lacking sulfide (data not shown). However, in medium devoid of sulfide, a longer doubling time was observed for EcKatG-induced cells (~18 h) compared to the uninduced cells (~9 h). The slower growth rate was specific to the absence of sulfide and induction of EcKatG, indicating that expression of a heme-dependent catalase in the absence of sulfide is inhibitory. Despite the slower growth rate, actively growing cultures of EcKatG-induced cells were more resistant to the addition of 1.5 or 3 mM H_2O_2 at mid-exponential phase, compared to the uninduced cells (Fig. 4). The uninduced cells immediately stopped growing upon the addition of 1.5 mM H_2O_2, a concentration which only resulted in a 10-fold loss of viability according to the microtiter assay (Fig. 3). The addition of 3 mM H_2O_2 not only stopped growth of the uninduced cells, but resulted in cell lysis as evidenced by the decrease in optical density (Fig. 4A). This result is consistent with the significant loss of viability of the uninduced cells when exposed to 3 mM H_2O_2 for one hour as assessed by the microtiter assay (Fig. 3). However, EcKatG-induced cells were able to overcome the addition of 1.5 mM H_2O_2 and continue to grow, albeit slower and with lower yield than the H_2O_2-free control (Fig. 3). The addition of 3 mM H_2O_2 did result in a cessation of growth of the EcKatG-induced cells; however, there was no apparent cell lysis, unlike the uninduced cells. Moreover, there were still a substantial number of viable cells remaining 40 hours post-H_2O_2 addition, because growth was observed when fresh medium was inoculated with an aliquot of the 3 mM H_2O_2-exposed EcKatG-induced cultures (data not shown). These results demonstrate that expression of EcKatG can protect *M. acetivorans* from acute H_2O_2 toxicity. To our knowledge this is the first example of engineered H_2O_2 resistance within a strict anaerobe, including methanogens.

An increase in endogenous catalase activity does not confer increased resistance of *M. acetivorans* to O_2. In the majority of natural environments, methanogens are exposed to O_2, rather than high levels of H_2O_2. However, methanogens contain many enzymes, including low-potential flavoenzymes, similar to those found in other strict anaerobes known to reduce O_2 to H_2O_2 and/or O_2^- (Imlay, 2003). Therefore, catalase may serve a role in degrading endogenously-produced H_2O_2 when cells are exposed to O_2. Although, the addition of atmospheric levels of O_2 (20%) results in a cessation of growth of *M. acetivorans*, exposure to this O_2 concentration for one hour does not decrease viability (Horne & Lessner, 2013). Therefore, we investigated the effect of EcKatG expression on the ability of *M. acetivorans* to grow in the presence of lower concentrations of O_2, similar to studies with *M. arboriphilus* (Brioukhanov & Netrusov, 2012). EcKatG-induced and uninduced cells of strain DJL20 were grown with methanol in medium devoid of sulfide. At mid-exponential phase, cells were challenged by the addition 1% or 5% O_2 (vol/vol) to the headspace of the cultures (Fig. 5). Under these culture conditions the cells would be exposed to a maximum of ~ 190 μM O_2, which is the maximum dissolved O_2 concentration at 35 °C (Colt, 1984). Both the uninduced and the EcKatG-induced cells were initially inhibited by the addition of 1% O_2 but eventually were able to resume growth. The addition of 5% O_2 to the headspace of uninduced and EcKatG-induced cells resulted in a cessation of growth. This result indicates that a significant increase in endogenous catalase activity does not confer an increase in the resistance of growing *M. acetivorans* cells to O_2. In contrast, increased endogenous catalase activity in *M. arboriphilus* was observed to increase the ability of this methanogen to overcome the addition of O_2 to growing cultures (Brioukhanov & Netrusov, 2012). However, a slight increase in the resistance of *M. acetivorans* strain to O_2 by increased catalase activity could be masked by the slower growth rate of the EcKatG-induced cells in medium devoid of sulfide.

Therefore, we attempted to determine the cause of the slower growth rate of the EcKatG-induced cells when grown in the absence of sulfide.

The effect of the addition of exogenous hemin on growth and aerotolerance of *M. acetivorans*. *M. acetivorans* is capable of synthesizing heme necessary for incorporation into endogenous cytochromes, as well as recombinant EcKatG. However, in the absence of sulfide expression of EcKatG could cause a depletion of heme, resulting in limitation of heme to metabolic enzymes, causing a slower growth rate. To determine if heme was a limiting factor during growth of EcKatG-induced cells in the absence of sulfide, growth was monitored in medium supplemented with hemin. The addition of 30μM hemin fully restored the growth rate of the EcKatG-induced cells with methanol in the absence of sulfide to that observed for EcKatG-induced cells grown in the presence of sulfide (Fig. 6). Although the molecular connection between sulfide and heme levels is not apparent, the ability of exogenous hemin to restore the normal growth rate indicates that EcKatG-induced *M. acetivorans* cells have decreased levels or synthesis rates of heme when grown in the absence of sulfide.

The exogenous addition of hemin was shown to positively affect the resistance of *M. arboriphilus* to oxidants. *M. arboriphilus* lacks cytochromes and is incapable of synthesizing heme. However, the addition of exogenous hemin results in an increase in endogenous catalase activity, due to conversion of the heme-dependent catalase (KatE) to the holo-form (Brioukhanov & Netrusov, 2012). The increased endogenous catalase activity was postulated to account for a significant increase in the resistance of *M. arboriphilus* to both H_2O_2 and O_2. Therefore, we assessed the effect of exogenous hemin on the endogenous catalase activity in EcKatG-induced and uninduced cells of *M. acetivorans* strain DJL20. An increase in the

endogenous catalase activity was not observed in either EcKatG-induced or uninduced cells when grown in the presence of sulfide and/or hemin, compared to cells grown in the absence of sulfide and/or hemin (data not shown). This result indicates that *M. acetivorans* does not contain an endogenous heme-inducible catalase, unlike *M. arborphilus*, and that EcKatG is not limited for heme when expressed in *M. acetivorans* cells grown in medium without sulfide or hemin. However, the exogenous addition of 30µM hemin increased the resistance of both uninduced and EcKatG-induced cells of *M. acetivorans* strain DJL20 to O_2, albeit only slightly, as no inhibition was observed by the addition of 1% O_2 (Fig. 7) compared to cells grown in medium without hemin (Fig. 5). Hemin may cause a change in the expression of other antioxidant enzymes in *M. acetivorans*, which could account for the increased resistance of strain DJL20 to O_2. Hemin may also provide additional protection from O_2 and/or H_2O_2 toxicity. For example, in a buffered solution, 30µM hemin is able to decompose 3 mM H_2O_2 within 10 min (Fig. 8). Taken together, these data indicate that exogenous hemin can provide additional protection from oxidants and that the 100-fold increase in the endogenous catalase activity in *M. acetivorans* does not increase the tolerance of growing cultures of *M. acetivorans* to O_2, at least under the experimental conditions tested here. This result is in contrast to the observed protection by increased hemin-dependent catalase activity in *M. arboriphilus*, which was demonstrated to significantly increase the tolerance of this methanogen to O_2, as well as H_2O_2 (Brioukhanov & Netrusov, 2012).

Discussion

The results from this study provide insight into the role of catalase in the antioxidant system of methanogens and the feasibility of engineering beneficial traits into methanogens. The limited number of methanogens that encode catalase suggests that it is not a core component of the antioxidant system in methanogens. *M. acetivorans* encodes homologs of the heme-

dependent catalases KatE and KatG, yet one contains a frameshift mutation (KatE) and the other does not appear to encode a functional catalase. In contrast, functional KatE in *M. barkeri* is up-regulated upon exposure of the cells to H_2O_2 and the O_2^--generating chemical paraquat (Brioukhanov *et al.*, 2006). Phylogenetic evidence suggests that catalase genes in methanogens were acquired by lateral gene transfer (Zamocky *et al.*, 2012). The disparity in catalase activity between the two *Methanosarcina* species may be a result of environmental differences exerting varying selective pressure. For example, the genome of the marine species *M. acetivorans* encodes hydrogenase, yet *M. acetivorans* lacks detectable hydrogenase activity and does not consume or produce hydrogen during growth. In contrast, the freshwater-species *M. barkeri* possesses hydrogenase activity and has the ability to consume and produce hydrogen. It is postulated that *M. acetivorans* has lost the inability to use hydrogen because in the marine environment it would have to compete with sulfate-reducers for hydrogen (Ferry & Lessner, 2008; Guss *et al.*, 2009). Similarly, differences between freshwater and marine environments, such as the solubility of oxygen and microbial community composition, could alter the levels of O_2 and ROS exposure to methanogens in marine and freshwater environments and hence the selective pressure for utilizing specific antioxidant enzymes. It appears *M. acetivorans* does not use the acquired *katE* or *katG* genes. It is unclear how widespread this phenomenon is in methanogens, because catalase activity has not been examined in the remaining methanogens that encode catalase (Table 1). Nonetheless, it should be noted that the presence of a gene encoding catalase in the genome of an individual species is not sufficient evidence to conclude that catalase plays a role in the antioxidant system of the organism.

To determine if catalase provides an advantage to *M. acetivorans* during exposure to O_2 and H_2O_2, we employed a novel approach using the established *M. acetivorans* genetic system. Heterologous expression of *E. coli* KatG in *M. acetivorans* resulted in endogenous catalase activity within the range of activities observed for *M. barkeri* and *M. arboriphilus*. This increase in catalase activity improved the tolerance of *M. acetivorans* to H_2O_2, but did not provide an advantage when cells were exposed to O_2. Exposure of anaerobes to O_2 results in the endogenous production of H_2O_2 and O_2^-; however, the rates of synthesis and levels of each ROS have not been determined in *M. acetivorans*. H_2O_2 may not be the primary ROS produced during O_2 exposure, which could explain the lack of additional tolerance by increased catalase activity. Alternatively, endogenously produced H_2O_2 may be sufficiently scavenged by a number of other antioxidant enzymes encoded in the genome of *M. acetivorans*, including rubrerythrin (MA0639), peroxiredoxin (MA4103), and several iron-sulfur flavoproteins (Isf), which have been shown in other species to scavenge H_2O_2 (Cruz & Ferry, 2006; Lumppio *et al.*, 2001). *M. acetivorans* is more likely to come in contact with O_2, rather than high concentrations of H_2O_2, and may have evolved H_2O_2-scavenging capabilities that do not involve catalase, even though catalase genes were acquired.

EcKatG is a heme-dependent enzyme and was active when expressed in *M. acetivorans* cells grown in medium lacking hemin, revealing that heme synthesized in *M. acetivorans* is properly inserted into EcKatG. However, when EcKatG was expressed in *M. acetivorans* cells grown in medium lacking sulfide, a reduced growth rate was observed. Supplementation with hemin restored normal growth and did not induce additional catalase activity. The lack of catalase induction by hemin is similar to results seen with *M. barkeri* (Brioukhanov & Netrusov, 2004). However, hemin induces catalase activity in *M. arboriphilus*, a methanogen that is

incapable of synthesizing heme. Addition of heme converts the apo-catalase to the active form (heme-containing) in *M. arboriphilus* (Brioukhanov & Netrusov, 2012). Also, unlike *M. acetivorans*, an increase in catalase activity in *M. arboriphilus* correlated with increased tolerance of growing cultures to O_2. A correlation with catalase activity and O_2 tolerance has not been documented with *M. barkeri*. Among methanogens, the use of catalase as a primary H_2O_2 scavenging enzyme may be unique to *M. arboriphilus*. Interestingly, none of the sequenced *Methanobrevibacter* genomes within the database (http://img.jgi.doe.gov) encode KatE or KatG, indicating use of KatE maybe specific to *M. arboriphilus* strains. However, since the levels of catalase activity in *M. arboriphilus* are dependent on the concentration of hemin in the growth medium under the experimental conditions examined (Brioukhanov & Netrusov, 2012), it is difficult to distinguish protection solely from catalase from that by hemin and possibly hemin-induced factors.

Conclusions

In summary, the results suggest catalase is not a key antioxidant enzyme in methanogens. Despite the fact that the majority of methanogens are tolerant to some O_2, the key enzymes and factors that contribute to the observed aerotolerance are not clear. More detailed characterization is required to identify these enzymes and factors. Importantly, the recombinant approach described here could be used to identify and assess the importance of other enzymes (e.g. superoxide dismutase, superoxide reductase, peroxidase, etc.) to the oxidant tolerance of methanogens. Finally, the results described here, along with previous studies (Lessner *et al.*, 2010), highlight the ability to design methanogen strains with beneficial traits, which may aid in the development of methanogens as biological catalysts.

Acknowledgements

This work was supported in part by grant number P30 GM103450 from the National Institute of General Medical Sciences of the National Institutes of Health (DJL), NSF grant number MCB1121292 (DJL), NASA Exobiology grant number NNX12AR60G (DJL), and the Arkansas Biosciences Institute (DJL), the major research component of the Arkansas Tobacco Settlement Proceeds Act of 2000.

References

Angel, R., Matthies, D. & Conrad, R. (2011). Activation of methanogenesis in arid biological soil crusts despite the presence of oxygen. *PLoS One* 6, e20453.

Angel, R., Claus, P. & Conrad, R. (2012). Methanogenic archaea are globally ubiquitous in aerated soils and become active under wet anoxic conditions. *ISME J* 6, 847-862.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72, 248-254.

Brioukhanov, A. L. & Netrusov, A. I. (2004). Catalase and superoxide dismutase: distribution, properties, and physiological role in cells of strict anaerobes. *Biochemistry (Mosc)* 69, 949-962.

Brioukhanov, A. L., Netrusov, A. I. & Eggen, R. I. (2006). The catalase and superoxide dismutase genes are transcriptionally up-regulated upon oxidative stress in the strictly anaerobic archaeon Methanosarcina barkeri. *Microbiology* 152, 1671-1677.

Brioukhanov, A. L. & Netrusov, A. I. (2012). The positive effect of exogenous hemin on a resistance of strict anaerobic archaeon *Methanobrevibacter arboriphilus* to oxidative stresses. *Curr Microbiol* 65, 375-383.

Carpena, X., Loprasert, S., Mongkolsuk, S., Switala, J., Loewen, P. C. & Fita, I. (2003). Catalase-peroxidase KatG of *Burkholderia pseudomallei* at 1.7A resolution. *J Mol Biol* 327, 475-489.

Colt, J. (1984). *Computation of Dissolved Gas Concentrations in Water as Functions of Temperature, Salinity, and Pressure*. Bethesda, Maryland: American Fisheries Society Special Publication.

Cruz, F. & Ferry, J. G. (2006). Interaction of iron-sulfur flavoprotein with oxygen and hydrogen peroxide. *Biochim Biophys Acta* 1760, 858-864.

Diaz, A., Loewen, P. C., Fita, I. & Carpena, X. (2012). Thirty years of heme catalases structural biology. *Arch Biochem Biophys* 525, 102-110.

Ferry, J. G. & Lessner, D. (2008). Methanogenesis in Marine Sediments. *Ann N Y Acad Sci* 1125, 147-157.

Fetzer, S., Bak, F. & Conrad, R. (1993). Sensitivity of methanogenic bacteria from paddy soil to oxygen and desiccation. *FEMS Microbiol Ecol* 12, 107-115.

Guss, A. M., Rother, M., Zhang, J. K., Kulkarni, G. & Metcalf, W. W. (2008). New methods for tightly regulated gene expression and highly efficient chromosomal integration of cloned genes for *Methanosarcina* species. *Archaea* 2, 193-203.

Guss, A. M., Kulkarni, G. & Metcalf, W. W. (2009). Differences in hydrogenase gene expression between Methanosarcina acetivorans and Methanosarcina barkeri. *J Bacteriol* 191, 2826-2833.

Horne, A. J. & Lessner, D. J. (2013). Assessment of the oxidant tolerance of *Methanosarcina acetivorans*. *FEMS Microbiol Lett* 343, 13-19.

Imlay, J. A. (2002). How oxygen damages microbes: oxygen tolerance and obligate anaerobiosis. *Adv Microb Physiol* 46, 111-153.

Imlay, J. A. (2003). Pathways of oxidative damage. *Annu Rev Microbiol* 57, 395-418.

Jenney, F. E., Jr., Verhagen, M. F., Cui, X. & Adams, M. W. (1999). Anaerobic microbes: oxygen detoxification without superoxide dismutase. *Science* 286, 306-309.

Lessner, D. J., Li, L., Li, Q. & other authors (2006). An unconventional pathway for reduction of CO2 to methane in CO-grown *Methanosarcina acetivorans* revealed by proteomics. *Proc Natl Acad Sci U S A* 103, 17921-17926.

Lessner, D. J., Lhu, L., Wahal, C. S. & Ferry, J. G. (2010). An engineered methanogenic pathway derived from the domains *Bacteria* and *Archaea*. *mBio* 1, e000243-000210.

Li, Q., Li, L., Rejtar, T., Karger, B. L. & Ferry, J. G. (2005a). Proteome of *Methanosarcina acetivorans* Part II: comparison of protein levels in acetate- and methanol-grown cells. *J Proteome Res* 4, 129-135.

Li, Q., Li, L., Rejtar, T., Karger, B. L. & Ferry, J. G. (2005b). Proteome of *Methanosarcina acetivorans* Part I: an expanded view of the biology of the cell. *J Proteome Res* 4, 112-128.

Lumppio, H. L., Shenvi, N. V., Summers, A. O., Voordouw, G. & Kurtz, D. M., Jr. (2001). Rubrerythrin and rubredoxin oxidoreductase in *Desulfovibrio vulgaris*: a novel oxidative stress protection system. *J Bacteriol* 183, 101-108.

Passardi, F., Zamocky, M., Favet, J., Jakopitsch, C., Penel, C., Obinger, C. & Dunand, C. (2007). Phylogenetic distribution of catalase-peroxidases: are there patches of order in chaos? *Gene* 397, 101-113.

Shima, S., Netrusov, A., Sordel, M., Wicke, M., Hartmann, G. C. & Thauer, R. K. (1999). Purification, characterization, and primary structure of a monofunctional catalase from *Methanosarcina barkeri*. *Arch Microbiol* 171, 317-323.

Shima, S., Sordel-Klippert, M., Brioukhanov, A., Netrusov, A., Linder, D. & Thauer, R. K. (2001). Characterization of a heme-dependent catalase from *Methanobrevibacter arboriphilus*. *Appl Environ Microbiol* 67, 3041-3045.

Sowers, K. R., Baron, S. F. & Ferry, J. G. (1984). *Methanosarcina acetivorans* sp. nov., an Acetotrophic Methane-Producing Bacterium Isolated from Marine Sediments. *Appl Environ Microbiol* 47, 971-978.

Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W. & Hedderich, R. (2008). Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* 6, 579-591.

Tholen, A., Pester, M. & Brune, A. (2007). Simultaneous methanogenesis and oxygen reduction by *Methanobrevibacter cuticularis* at low oxygen fluxes. *FEMS Microbiol Ecol* 62, 303-312.

Zamocky, M., Furtmuller, P. G. & Obinger, C. (2008). Evolution of catalases from bacteria to humans. *Antioxid Redox Signal* 10, 1527-1548.

Zamocky, M., Gasselhuber, B., Furtmuller, P. G. & Obinger, C. (2012). Molecular evolution of hydrogen peroxide degrading enzymes. *Arch Biochem Biophys* 525, 131-144.

Figures and Tables

Table 1. Catalases characterized from methanogens and putative catalases encoded in the genomes of sequenced methanogens.

Methanogen	KatE[a] (Monofunctional)	KatG[b] (Bifunctional)
Methanoculleus marisnigri JR1	Memar_0875, 707aa (52%)	-[c]
Methanosarcina mazei Go1	MM_1950, 710aa (51%) MM_2557, 565aa (40%)	-
Methanosarcina barkeri str. Fusaro	Mbar_A0814[d], 505aa (50%)	-
Methanosalsum zhilinae DSM 4017	Mzhil_1695, 505aa (50%)	-
Methanococcoides burtonii DSM 6242	Mbur_0523, 504aa (50%)	-
Methanosaeta concilii GP6	MCON_0161, 489aa (42%)	-
Methanocorpusculum labreanum Z	Mlab_1574, 501aa (47%)	-
Methanocella sp. RC-I	RCIX619, 715aa (49%)	-
Methanofollis liminatans GKZPZ, DSM 4140	Metli_1702, 485aa (40%)	-
Methanomethylovorans hollandica DSM 15978	Metho_1739, 491aa (40%)	-
Methanoplanus limicola M3, DSM 2279	Metlim_1586, 482aa (47%)	-
Methanobrevibacter arboriphilus	AJ300838[e], 503aa (41%)	-
Methanoregula boonei 6A8	-	Mboo_0959, 733aa (63%)
Methanosphaerula palustris E1-9c	-	Mpal_2625, 710aa, (66%)
Methanohalophilus mahii DSM 5219	-	Mmah_1997, 725aa (63%)
Methanospirillum hungatei JF-1	-	Mhun_2433, 718aa (63%)
Methanobacterium sp. SWAN-1	-	MSWAN_0630, 724aa (62%)
Methanosaeta harundinacea 6Ac	-	Mhar_0135, 720aa (62%)
Methanobacterium formicicum DSM 3637	-	A994_08946, 721aa (61%)
Methanolobus psychrophilus R15	Mpsy_0349, 708aa (50%) Mpsy_1225, 505aa (41%)	Mpsy_0021, 732aa (58%)
Methanosarcina acetivorans C2A	MA2081, 173aa	MA0972, 736aa (64%)

[a]query *E. coli* KatE 753 aa, sequence identity in parentheses
[b]query *E. coli* KatG 726 aa, sequence identity in parentheses
[c]dash indicates genome lacks homolog.
[d]previously characterized (Shima et al., 1999))
[e]previously characterized (Shima et al., 2001))

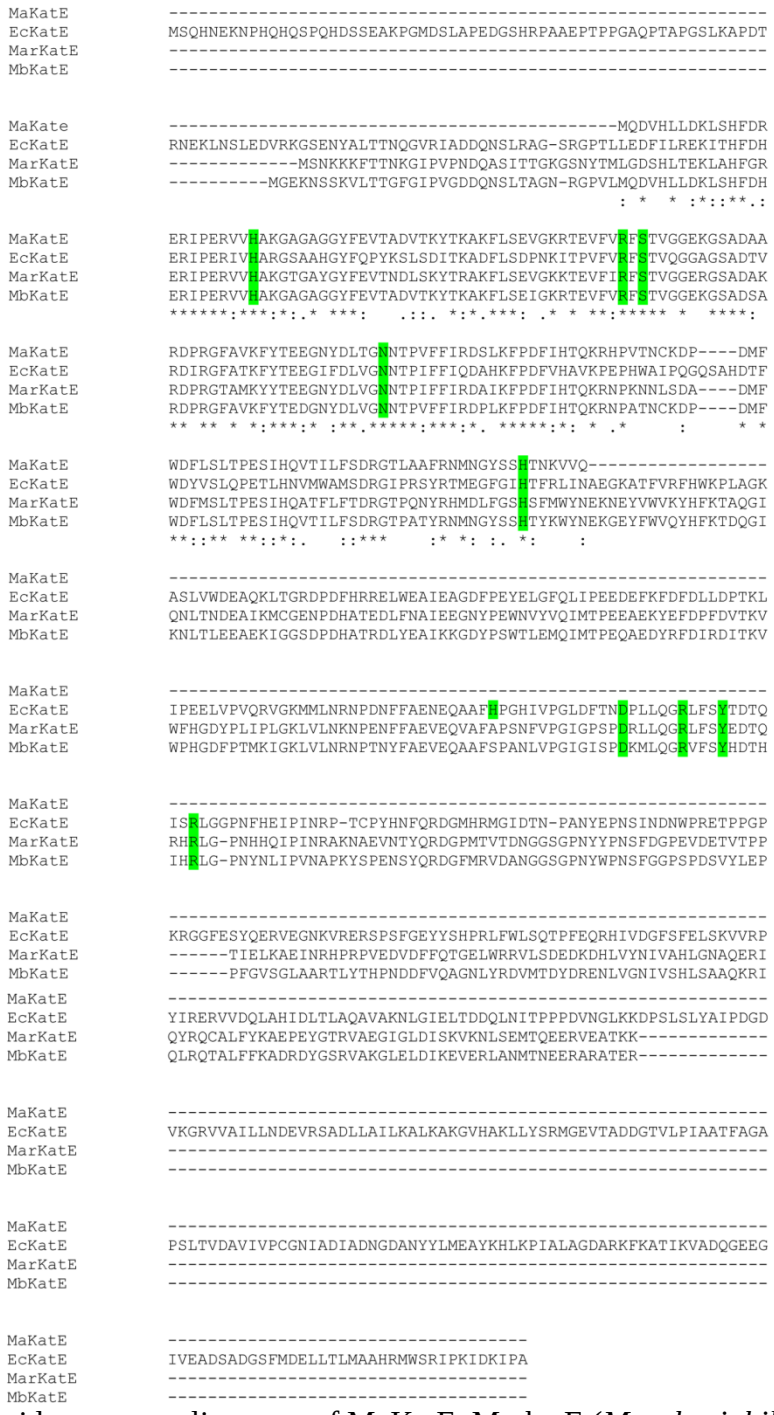

Figure 1. Amino acid sequence alignment of MaKatE, MarkatE (*M. arboriphilus*), MbKatE (*M. barkeri*), and EcKatE. Identical amino acid residues are indicated with an asterisk and similar residues are indicated by a colon or period. The active site residues of EcKatE are highlighted: R165, S167, R125, R422, H128, N201, R411, Y415, H392, H275, and D405 (From Diaz et al., 2012).

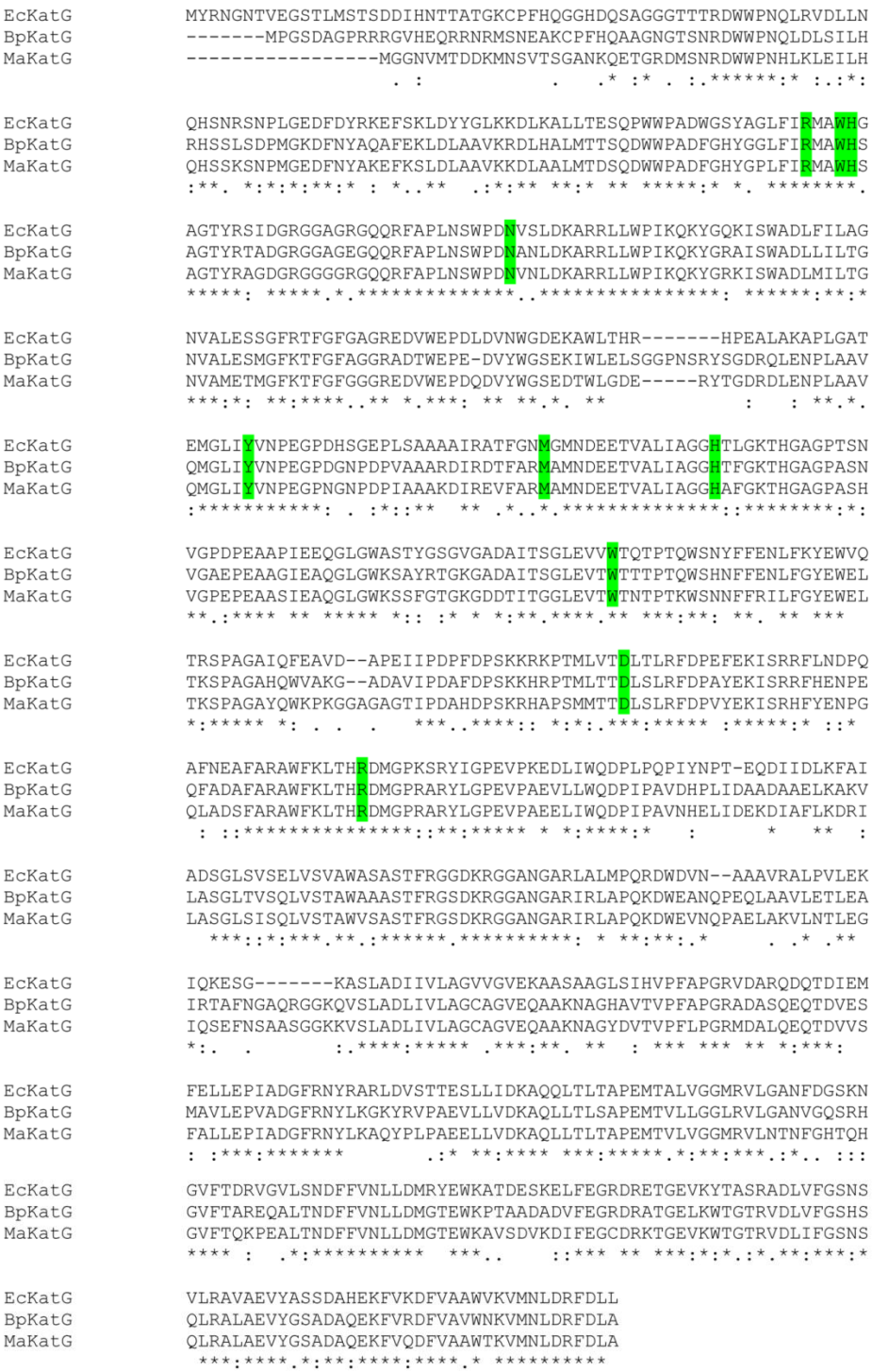

Figure 2. Amino acid sequence alignment of MaKatG, EcKatG, and BpKatG (*Burkholderia pseudomallei*). Identical amino acid residues are indicated with an asterisk and similar residues are indicated by a colon or period. The active site residues of BpKatG are highlighted: R108, W111, H112, Y238, M264, H279, W330, D389, and R426 (From Diaz et al., 2012).

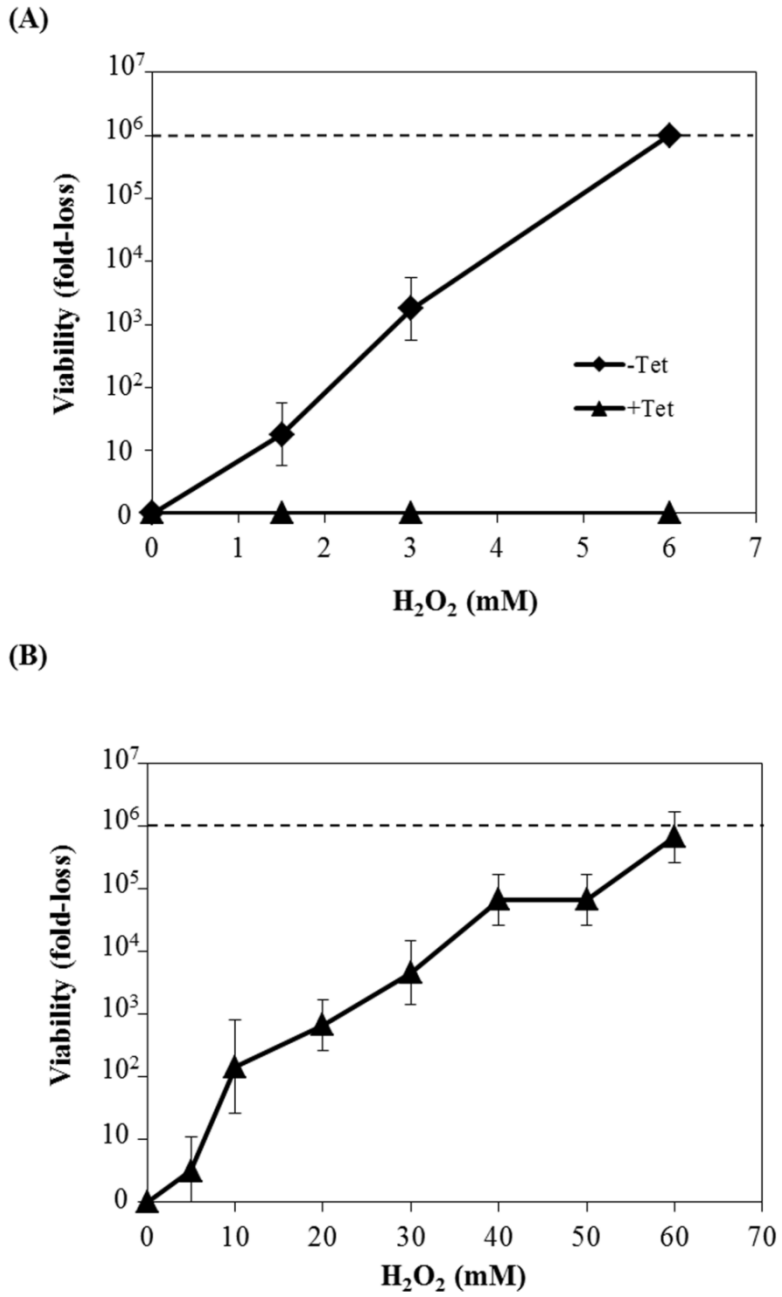

Figure 3. (A) Comparison of H_2O_2 tolerance of EcKatG-induced cells to uninduced cells of *M. acetivorans* strain DJL20. Cells were grown in the absence of tetracycline (-Tet) or in the presence of tetracycline (+Tet) and challenged with 0, 1.5, 3.0, and 6.0 mM H_2O_2 for one hour. (B) H_2O_2 tolerance of EcKatG-induced cells of strain DJL20 assessed by the microtiter-assay. In each graph the line depicts the highest fold loss that results in a complete absence of viable cells and the data plotted are the mean ± SD from three independent experiments.

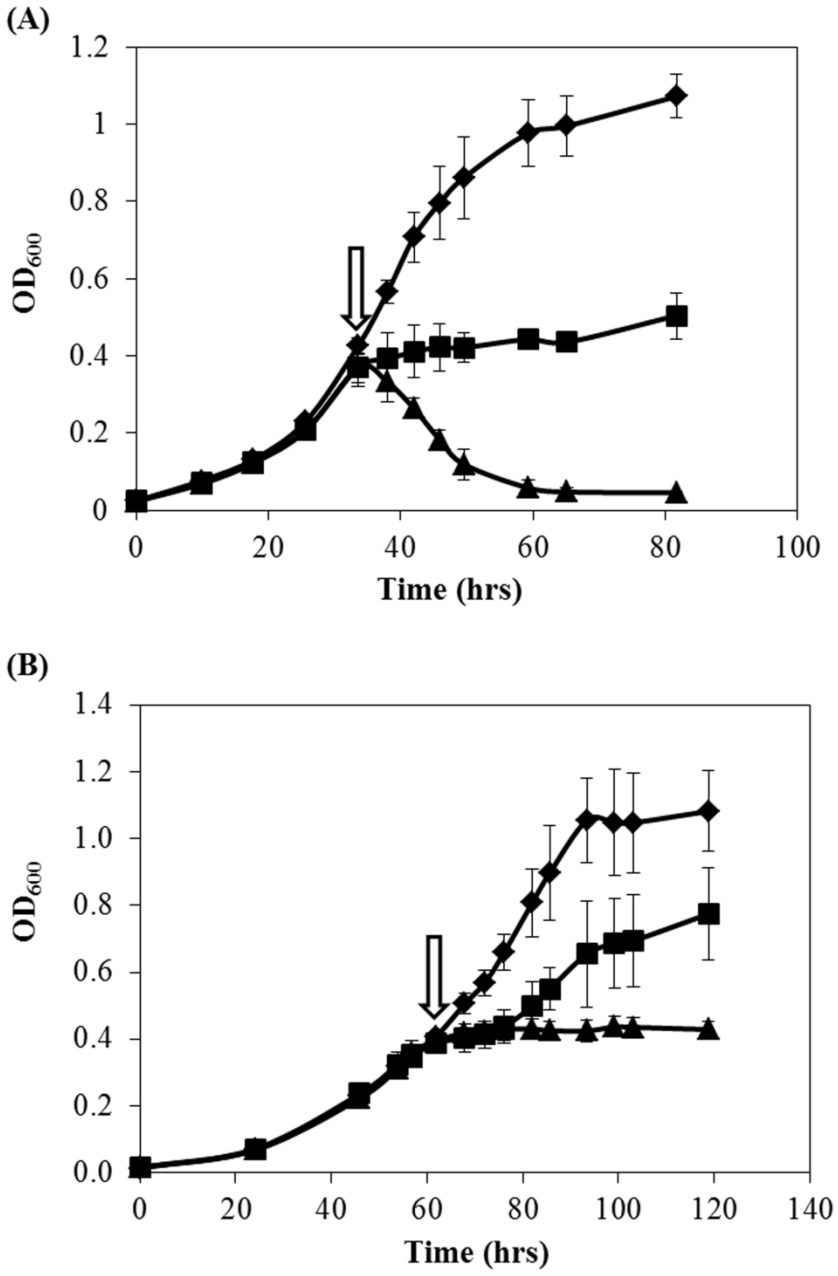

Figure 4. Effect of H_2O_2 on growth of EcKatG-induced cells compared to uninduced cells of *M. acetivorans* strain DJL20. Mid-exponential phase cultures of DJL20 grown in the absence (A) or presence (B) of tetracycline were challenged with 0 mM (diamonds), 1.5 mM (squares), or 3 mM (triangles) H_2O_2 at the time point indicated by the arrow. The data plotted are the mean ± SD from three independent experiments.

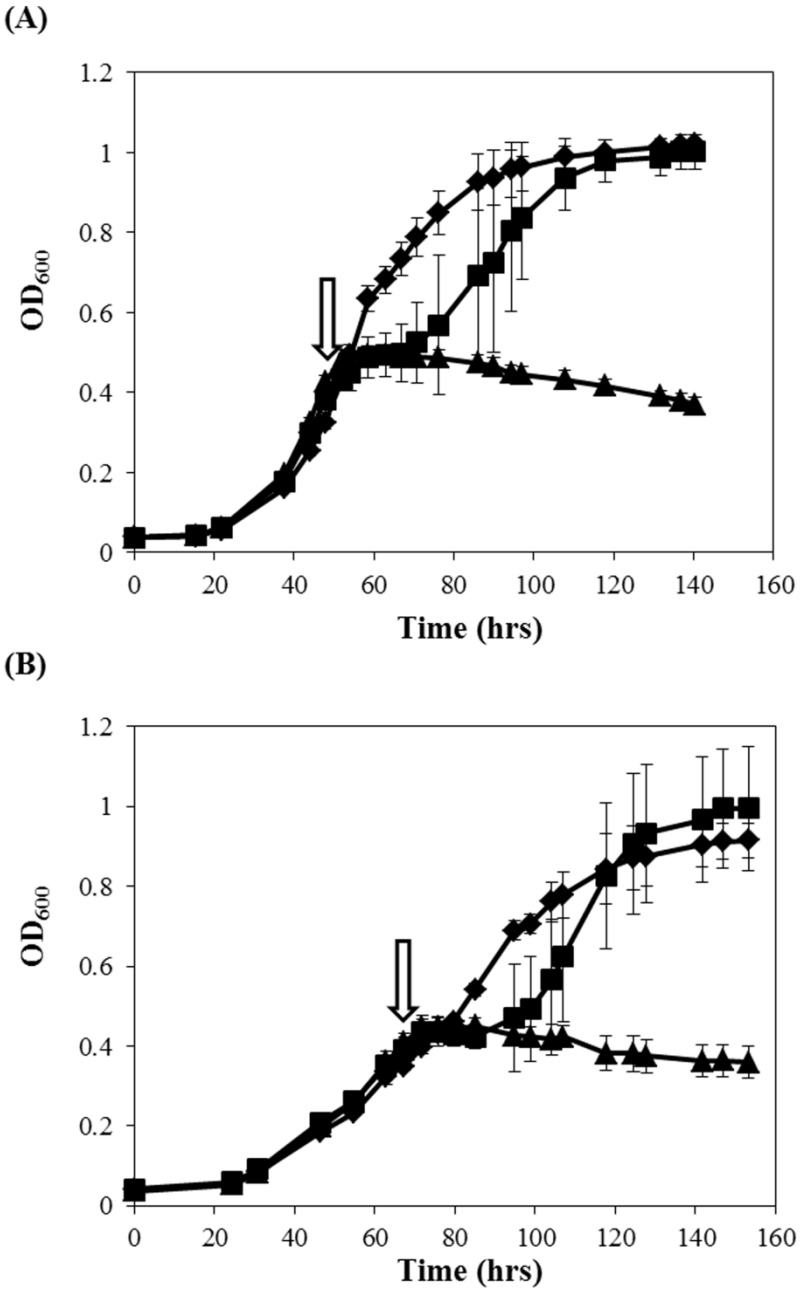

Figure 5. Effect of O_2 on growth of EcKatG-induced cells compared to uninduced cells of *M. acetivorans* strain DJL20. Mid-exponential phase cultures of DJL20 grown in the absence (A) or presence (B) of tetracycline were not challenged with O_2 (diamonds) or were challenged with 1% (squares), or 5% (triangles) O_2 at the time point indicated by the arrow. The data plotted are the mean ± SD from three independent experiments.

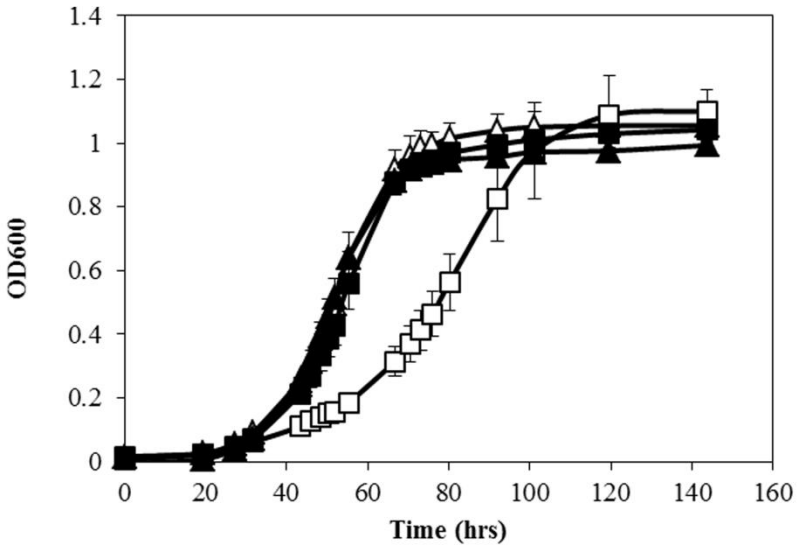

Figure 6. Growth of EcKatG-induced cells compared to uninduced cells of *M. acetivorans* strain DJL20 in medium supplemented with hemin. Cultures of DJL20 were grown in the absence (triangles) or presence (squares) of tetracycline in medium lacking (open symbols) or containing (filled symbols) 30 μM hemin. The data plotted are the mean ± SD from three independent experiments.

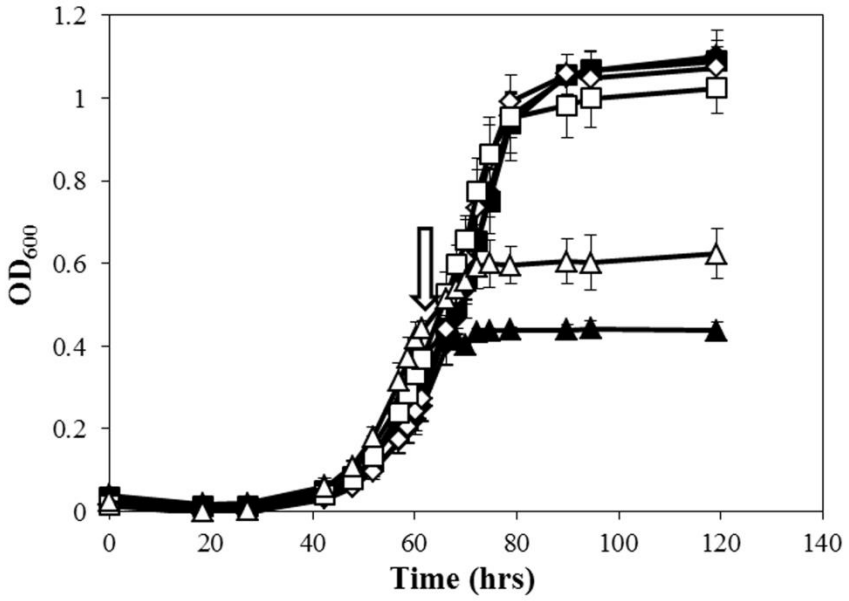

Figure 7. Effect of O_2 on growth of EcKatG-induced cells compared to uninduced cells of *M. acetivorans* strain DJL20 in medium supplemented with 30 µM hemin. Mid-exponential phase cultures of DJL20 grown in the absence (open symbols) or presence (filled symbols) of tetracycline were not challenged with O_2 (diamonds) or were challenged with 1% (squares), or 5% (triangles) O_2 at the time point indicated by the arrow. The data plotted are the mean ± SD from three independent experiments.

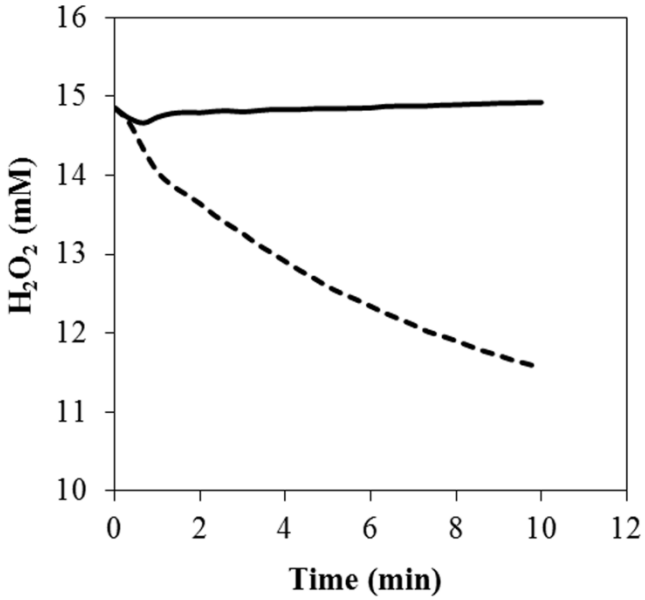

Figure 8. H2O2 degradation by hemin**.** Degradation of H2O2 in 50 mM Tris-HCl, pH 8.0 (solid line) or 50 mM Tris-HCl, pH 8.0 supplemented with 30 μM hemin (dashed line). H2O2 concentration was determined by monitoring absorbance at 240 nm (ε240 = 39.4 M-1 cm-1).

Appendix 3.1: Lead Author Confirmation Letter for Chapter III

UNIVERSITY OF ARKANSAS

J. William Fulbright College of Arts and Sciences
Department of Biological Sciences

Chapter IV, titled "Expression of a bacterial catalase in a strictly anaerobic methanogen significantly increases tolerance to hydrogen peroxide but not oxygen," of M. E. Jennings's dissertation was published in Microbiology in 2014 with coauthors C.W. Schaff, A.J. Horne, F.H. Lessner, and D.J. Lessner.

I, Dr. Daniel J. Lessner, advisor of Matthew Edward Jennings, confirm Matthew Edward Jennings was first author and completed at least 51% of the work for this manuscript.

Dr. Daniel J. Lessner
Associate Professor
Department of Biological Sciences
University of Arkansas

Date

Science Engineering, Room 601 · Fayetteville, AR 72701-1201 · 479-575-3251 · Fax: 575-4010

www.uark.edu

CONCLUSIONS

Methanogens are a group of anaerobic archaea which share a unique metabolic pathway; methanogenesis. This pathway, the only mechanism by which these organisms conserve energy, results in the production of methane. In fact, methanogens are the only known source of biologically produced methane. The potential of methane as an agent of climate change (methane is a more potent absorber of heat than carbon dioxide) and as a potential fuel source means understanding the metabolism and physiology of methanogens can help address two important issues facing human civilization.

As obligate anaerobes, methanogens grow poorly if at all in the presence of oxygen (O_2). Methanogens possess a large number of Fe-S proteins, more so than any known group, and the prosthetic groups in these proteins are extremely sensitive to O_2. Oxygen not only disrupts existing clusters but inhibits the production of new clusters, and Fe-S clusters can facilitate the production of more ROS through the Fenton reaction which can be lethal if left unchecked. Despite all this many methanogen species can survive prolonged exposure to aerobic conditions and resume growth once anaerobic conditions have resumed, suggesting they possess mechanisms for dealing with oxidative stress. Studies of methanogens species have shown the presence of enzymes with detectable levels of ROS detoxification activity, and transcriptional regulators which modulate expression in response to redox conditions. However, there is still much yet unknown about the oxidative stress responses of methanogens.

This dissertation presented two scientific inquiries into the oxidative stress response of methanogens using the model organism *Methanosarcina acetivorans*. A potential new mechanism of transcriptional regulation was investigated through the examination of two [4Fe-4S] clusters bound to RNAP subunit D, and their role in assembly and activity of RNAP. The activity of two putative catalases, and their importance to the oxidative stress response of *M.*

acetivorans was also investigated. Together these results provide a clearer picture of how *M. acetivorans* senses and responds to oxidative stress. These results can be used as a guide for additional studies done using *M. acetivorans*, other methanogens, and other anaerobic organisms.

[4Fe-4S] clusters in RNAP may constitute a new regulatory mechanism in transcription regulation

Fe-S clusters have been adapted to perform a sensory function in certain transcription factors through their sensitivity to O_2, regulating gene expression in response to the redox state of the cell. The presence of Fe-S clusters in the D subunit of RNA polymerase (RNAP) suggested the prosthetic group may have also been adapted to perform a regulatory role in the transcriptional machinery itself. Iron-sulfur clusters could regulate assembly and activity of RNAP through modulation of D-L heterodimer formation, since formation of the D-L heterodimer is the first step in RNAP assembly. The D subunit of *M. acetivorans* contains a ferredoxin-like domain (FLD) with two [4Fe-4S] clusters bound to it, which so far is a feature found exclusively in anaerobic archaea. The results presented in this dissertation provide strong support for the [4Fe-4S] clusters playing an important role in the assembly of *M. acetivorans* RNAP, especially during assembly of subunit A'' with the B'B''DLNP subcomplex. The mechanism by which loss of cluster inhibits RNAP assembly was not determined, though the evidence suggests a misfolding of the FLD interferes with the interaction of subunit D with subunit B' and more severely later during interaction with subunit A'' at minimum. The evidence presented shows the clusters and FLD are important for RNAP assembly and activity, and are sensitive to oxygen, which suggests the clusters within the FLD could play a role the oxidative stress response of *M. acetivorans.* The results presented in this dissertation not only provide evidence as to the role of the cluster in *M. acetivorans*, but also provides evidence that

the clusters may be playing a previously unrecognized regulatory role in transcription from a wide variety of organisms from two different lineages.

The absence of [4Fe-4S] clusters and the FLD from the D subunit homolog in bacteria (α), suggests that the domain and clusters were gained after the divergence of the archaea/eukaryote and bacteria lineages. Archaea are thought to have evolved early in Earth's history, and methanogen ancestors, which are specifically thought to have evolved early, would have had no problem in the oxygen-free atmosphere of early Earth. Likely, the acquisition of domain 3 (D3) occurred during this time and originally contained two [4Fe-4S] clusters. Since oxygen and ROS were much rarer during that time, there was not an intense selective pressure preventing the incorporation of Fe-S clusters into RNAP. The other subunits of RNAP co-evolved to accommodate the new domain, and the clusters may have been adapted to perform a regulatory role in RNAP. When oxygen began to accumulate in the atmosphere, establishing a more oxidizing environment, this was a huge selective pressure to modify the [4Fe-4S] clusters of D3 so that they were no longer a liability in the presence of oxygen. Organisms which persisted in an anaerobic lifestyle, such as the ancestors of modern methanogens, would have faced a lower selective pressure to get rid of oxygen sensitive components of RNAP, and could explain why today only obligate anaerobic archaea possess two [4Fe-4S] clusters in subunit D. The results in this dissertation prove that cluster two is less important for assembly and activity compared to cluster one, which would explain why this cluster was lost first in most lineages which possess only a single cluster. The second cluster may have been lost in the ancestor of eukaryotes, which could explain why only cluster one is present among that domain. The specific role of the cluster(s) was likely also turned through evolution as either one or both of the clusters were lost in certain lineages. This would explain the observed differences in the D subunits from

M. acetivorans and *Sulfolobus solfataricus*; the single cluster in *S. solfataricus* subunit D is required for D-L heterodimer formation, but the D subunit from *M. acetivorans* can form a heterodimer without either cluster. The Fe-S clusters in RNAP may be a previously unrecognized mechanism of transcription control; one that evolved very early and has been adapted to different roles depending on the evolutionary history of the organism.

Catalase is not an important component of the *M. acetivorans* oxidative stress response

Organisms express ROS detoxification enzymes to convert ROS into less reactive molecules. A few ROS detoxification enzymes from methanogens have been characterized, including a functional catalase from *Methanosarcina barkeri.* Catalase produces O_2 as one of its terminal products, so they are less common but not unheard of in anaerobes. The genome sequence of *M. acetivorans* contains two putative catalase genes, which were chosen for characterization. One gene MA2081, contained a point mutation that introduced a stop codon resulted in a truncated protein product and was not characterized further. The protein product from gene MA0972 was expressed recombinantly in *E. coli* but no activity was detected. Cell lysate of *M. acetivorans* did not contain detectable catalase activity under either aerobic or anaerobic conditions. The catalase KatG from *E. coli* was cloned into *M. acetivorans* to determine if an active catalase could confer resistance to oxidative stress. Cells expressing catalase were able to tolerate higher concentrations of H_2O_2, but no difference was detected when cells were exposed to O_2.

The results indicate catalase is not an important component of the antioxidant response of *M. acetivorans.* These results were surprising, since the closely related *M. barkeri* possesses a functional catalase, suggesting the antioxidant response differs between the two organisms. *M. acetivorans* is a salt water organism which *M. barkeri* lives in fresh water, and this difference of

environment could be what drove these two species to evolve alternate oxidative stress responses. *M. barkeri* may encounter H_2O_2 much more frequently than *M. acetivorans,* which would be a pressure for the organism to maintain a functional catalase. Other organisms coexisting in the same environment as *M. acetivorans* may effectively reduce H_2O_2, eliminating the need for *M. acetivorans* to maintain its own catalase. The results from the experiments conducted for this dissertation suggest that even among closely related methanogens, the pathways used to detoxify ROS are different. However, since the oxidative stress response of only a few methanogen species have been extensively characterized it is still difficult to determine if certain ROS response pathways are favored by methanogen lineages. As the ROS responses from additional methanogens are investigated, a better understanding of methanogen ROS responses will emerge.