



Unique and Ancient Metabolic Strategies:

The Purification of Novel Methanogen Sulfur Trafficking Proteins

James Stanek

July, 2016

*Master's Student of Biochemistry and Molecular Biology
Institute of Environmental Health
Oregon Health and Science University*

Table of Contents

Table of Contents	i
Acknowledgements	iv
List of Figures.....	v
List of Tables	viii
Abstract	ix
Chapter 1: Introduction.....	10
Early Life	10
Unique tRNA Aminoacylation	12
Ancient Strategies For An Anaerobic World	16
Three Protein Families Co-conserved with SepCysS	21
Homocysteine Synthesis in Methanogens.....	23
Connecting Homocysteine Synthesis with SepCysS	24
COG2122; ApbE Protein Superfamily of Flavin Trafficking Proteins	25
The COG1900 Protein Family	28
COG1900 Paralogs	31
A Possible Biochemical Role for COG1900D and Its Implications.....	32
Purpose of Studying the Proteins in This Study	35

Chapter 2: Protein Purification Methods	38
Identifying Proteins of Interest.....	38
Engineered <i>M. acetivorans</i> Strains	39
Engineered <i>E. coli</i> Strains	40
Previous COG1900 Purification Attempts from <i>E. coli</i>	40
Protein Purification Using Overexpressed Systems	44
Refolding.....	46
Protein Tags, Chaperones, and Expression Conditions	47
Chapter 3: Purification of COG1900D Proteins	49
COG1900 Initial Purification Observations	49
COG1900D Solubilization.....	54
COG1900D Expression Conditions in <i>E. coli</i>	62
Purification Hurdles of COG1900 Proteins	65
COG1900D Lysis	68
COG1900D Gel Filtration Chromatography	74
COG1900D Affinity Chromatography	77
MJ1681 Quaternary and Secondary Structure Analysis	82
Discussion	88
Chapter 4: Purification of COG1900A, MA1822 and COG2122 Proteins	89
COG1900A compared to COG1900D purification	89
Expression of MA1821.....	90
COG1900A Lysis.....	90

COG1900A Affinity Chromatography	92
Tlet_1363 COG1900A Fusion Protein Purification	96
MA1822; Ferredoxin Domain Protein	99
Purification of MA1822	100
Use of MA1822 in Future Studies	103
COG2122 Purification	104
Discussion	109
Conclusions	111
References	113
Appendix A: List of Buffers	129
Appendix B: Standards	131

Acknowledgements

One stick breaks easily but a bundle is strong; an old proverb I keep in mind. I sooner give credit for my accomplishments to those who have helped mold me before I give credit to myself.

I would like to thank the members of my lab. My mentor, Dr. John Perona has always been supportive throughout my time as a master's student. I would often take on too much work at one time; John taught me to "focus on one thing and do it very well". Invaluable advice.

The project would not have even been possible without the hard work and dedication of Benjamin Rauch, who spent the substantial portion of his PhD discovering these proteins of interest. Camden Driggers, a post doc in our lab, has taught me to take the slower more careful route in my research which has helped substantially.

The faculty in the OHSU Institute of Environmental Health has provided guidance and support throughout the process. My committee members, Dr. Pierre Moenne-Loccoz and Dr. Bradley Tebo, were always willing to help when I needed it.

Over a summer term we had the help of two high school interns, Emma Weirich and Sharon Zhang, who accelerated the project. They took part in some of the most valuable experiments in this project and it would have taken much longer to accomplish without their assistance.

The entire biology and chemistry departments of Portland State University have always been helpful and provided the tools necessary for this thesis.

My parents, Debra and John Stanek, taught me the importance of perseverance and hard work which are invaluable in graduate school. My sister, Janel Stanek, has always been a much needed motivator and competition during difficult times in college.

My uncle, Terry Peasley, more than anyone else, has taught me the importance of higher education and without him I may not have ever pursued academia. My significant other, Dasuni Wickramaratne, who has put up with my often impatience following long weeks in the lab. Thank you all.

List of Figures

Figure 1.1 Novel SepCysS Pathway	14
Figure 1.2 Persulfide Group Chemistry	15
Figure 1.3 <i>E. coli</i> Sulfur Trafficking Pathways	18
Figure 1.4 Methanogen Sulfur Trafficking	20
Figure 1.5 Protein Families Conserved with SepCysS.....	22
Figure 1.6 COG1900A Reaction	23
Figure 1.7 ApbE Protein Ribbon Diagram.....	27
Figure 1.8 Proposed COG1900A Mechanism ²⁷	30
Figure 1.9 Coenzyme M Biosynthesis Pathway ⁴⁶	33
Figure 1.10 COG1900A and Predicted COG1900D Reactions.....	33
Figure 2.1 COG1900A and COG2122 Protein Domains ²³	43
Figure 3.1 Domain Comparison Between MJ1681 and MA1821.....	51
Figure 3.2 MA1821 Amino Acid Composition	52
Figure 3.3 MJ1681 Amino Acid Composition.....	53
Figure 3.4 Expression Plasmid for MJ1681 Generation	56
Figure 3.5 Induction of Cells with COG1900D Expression Plasmid.....	57

Figure 3.6 Lysis Conditions Screened.....58

Figure 3.7 Gel of Lysis Additives Screen59

Figure 3.8 Evidence of MJ1681 Truncation During Expression63

Figure 3.9 MJ1681 Expression Temperature Comparison64

Figure 3.10 COG1900D Purification Optimization Flow Chart67

Figure 3.11 COG1900D Lysis Method.....69

Figure 3.12 COG1900D Insoluble Pellet Resuspension.....71

Figure 3.13 Insoluble Pellet Containing MJ1681.....72

Figure 3.14 Insoluble Pellet After MJ1681 is Resuspended73

Figure 3.15 MJ1681-SUMO SEC 175

Figure 3.16 MJ1681-SUMO SEC 276

Figure 3.17 MJ1681-SUMO Binding Ni-NTA with Multiple Passes.....78

Figure 3.18 MJ1681 Ni-NTA Chromatography and ULP Cleavage of SUMO80

Figure 3.19 SDS PAGE Gel of Purified MJ168181

Figure 3.20 Analytical SEC of MJ1681.....84

Figure 3.21 CD Spectroscopy of MJ168185

Figure 3.22 MJ1681 CD Spectrum in Guanidinium.....86

Figure 3.23 MJ1681 SEC at pH 887

Figure 4.1 MA1821/22 Expression Plasmid for MA1821/22 Generation	91
Figure 4.2 MA1821 Ni-NTA SDS-PAGE Gel	93
Figure 4.3 MA1821 Gel Filtration Chromatography at Various pH Values	94
Figure 4.4 MA1821 Gel Filtration Chromatography pH 10	95
Figure 4.5 Tlet_1363 Expression Plasmid	97
Figure 4.6 Tlet_1363 Ni-NTA Purification SDS PAGE Gel	98
Figure 4.7 MA1822 SDS-PAGE Gel	102
Figure 4.8 MA1822 Gel Filtration Chromatogram	103
Figure 4.9 MA1715 Expression Vector	105
Figure 4.10 MA1715 Anion Exchange Elutions	108

List of Tables

Table 2.1 Previous COG1900 Purification Attempts from <i>E. coli</i>	42
Table 2.2 Proteins Used in This Study	43
Table 2.3 Purification Strategies Tested.....	45
Table 3.1 COG1900 Protein Profiles	52
Table 3.2 Shift of Deprotonated Phosphate Species by pH.....	60
Table 4.2 MA1822 Profile	99
Table 4.3 MA1715 Profile	105
Table 4.4 Summary of all Protein Purifications.....	110

Abstract

Sulfur trafficking pathways in methanogens are distinct from bacteria because sulfide rather than sulfate, is the primary sulfur source. Many methanogens also lack the genes coding for O-acetylhomoserine sulfhydrylase, O-acetylserine sulfhydrylase, cysteine desulfurase, and cysteinyl-tRNA synthase (CysRS), all essential sulfur trafficking enzymes in bacteria. In the absence of CysRS, Cys-tRNA^{Cys} is aminoacylated in a novel two step pathway which includes the action of the methanogen-specific sulfur-transfer protein, SepCysS. Two protein families are co-conserved with SepCysS, COG2122 and COG1900(A-D). COG2122 rescues cells from starvation under low sulfide conditions. COG1900A synthesizes homocysteine (Hcy) from aspartate-semialdehyde through the conversion of an aldehyde to a thiol. If COG1900D is capable of the same biochemistry, it could be the last enzyme in the pathway of coenzyme M biosynthesis, an essential cofactor in methanogenesis. To date, no successful purification of any COG1900 protein has been reported. Using an unusual high pH buffer that includes phosphate, COG1900 proteins can be solubilized. The purification process is still prone to aggregation making it difficult to achieve a satisfactory level of purity.

Chapter 1: Introduction

Early Life

Life is hypothesized to have arisen during the Archean eon (4-2.5 Gyr). The environment in the Archean was anaerobic and rich in iron and sulfur. It was not until the Proterozoic era (2.5-0.7 Gyr) that the Earth was oxygenated by the metabolic activities of cyanobacteria¹. Upon the evolution of photosynthesis and universal oxygenation, anaerobic organisms would be excluded to niche environments that remained anoxic. The massive increase in dioxygen (O₂) on Earth is sometimes referred to as the “oxygen holocaust” as it drove most anaerobic organisms to extinction².

Despite the near extinction of these organisms, anaerobes were essential to Earth’s early ecosystem and provided the foundation for life as we know it by fostering the biochemical evolution of methane. Methane was a necessary greenhouse gas in a time with reduced solar luminosity³ and methanogens play a continual role in the

carbon cycle on Earth⁴. Surviving anaerobic organisms with ancient genes provide a window into the past and the origins of life on Earth.

Phylogenetic analysis has placed hyperthermophiles as possibly being the last common ancestors of living organisms, thus making them key species in the tree of life⁵. Geological C¹³ fractionation analysis has shown that methanogenesis occurs as far back as 3.46 Gyr⁶. The antiquity of methanogens supports the hypothesis that methanogenesis was essential to support early life on Earth. Methanogens evolved in an environment much different than the present one; many of their metabolic pathways are unique and not well understood.

The Archaean ocean had low oxygen levels making it much less oxidizing than modern oceanic environments; this lack of oxygen resulted in decreased abundance of many oxidized chemical species, such as sulfate⁷. This creates a distinction between metabolic pathways pre and post oxygenation. For example, most extant microbes have transporters for sulfate while methanogens are able to use sulfide as their only sulfur source⁸. The reduction of elemental sulfur (S⁰) to sulfide is common to sulfur reducing bacteria and archaea, all of which are anaerobic. This is due to the large abundance of oxygenated sulfur species now present compared to when methanogens evolved and reduced species were much more prominent.

The enzymes responsible for survival under such drastically different conditions is of great interest. The study of these enzymes is hindered because they now reside primarily in organisms from extreme environments and have many unique adaptations to remain functional. No general approach or strategy has been developed in the purification of proteins from extremophiles.

Unique tRNA Aminoacylation

A remnant of early life metabolism was discovered in a *Methanocaldococcus jannaschii* tRNA aminoacylation pathway. Cysteinyl-tRNA^{Cys} is used in translation of proteins and is typically biosynthesized in one-step by cysteinyl-tRNA synthetase (CysRS). It was discovered that *M. jannaschii* does not contain the gene for CysRS. Instead, a novel two-step pathway is present; aminoacylation of tRNA^{Cys} with phosphoserine using O-phosphoserine-tRNA synthetase (SepRS) followed by conversion to Cys-tRNA^{Cys} via Sep-tRNA-Cys-tRNA synthase (SepCysS)⁹ (Fig 1.1). This “mis-aminoacylation” of tRNA is present for other aminoacyl-tRNA synthetases (AARSs) such as GlnRS and AsnRS that exist in various systems and posed a challenge to the adaptor hypothesis that states there are twenty amino acid synthetases, one for each amino acid¹⁰.

Methanogens almost exclusively utilize the two-step pathway¹¹. Phylogenetic analysis shows the two-step pathway does in fact pre-date the direct pathway and is primarily present in methanogens¹². Nearly all methanogens sequenced contain genes for SepRS and SepCysS homologs⁸. This suggests a selective pressure in methanogens, not seen in other organisms, to maintain this pathway even when CysRS is present in some methanogens through horizontal gene transfer.

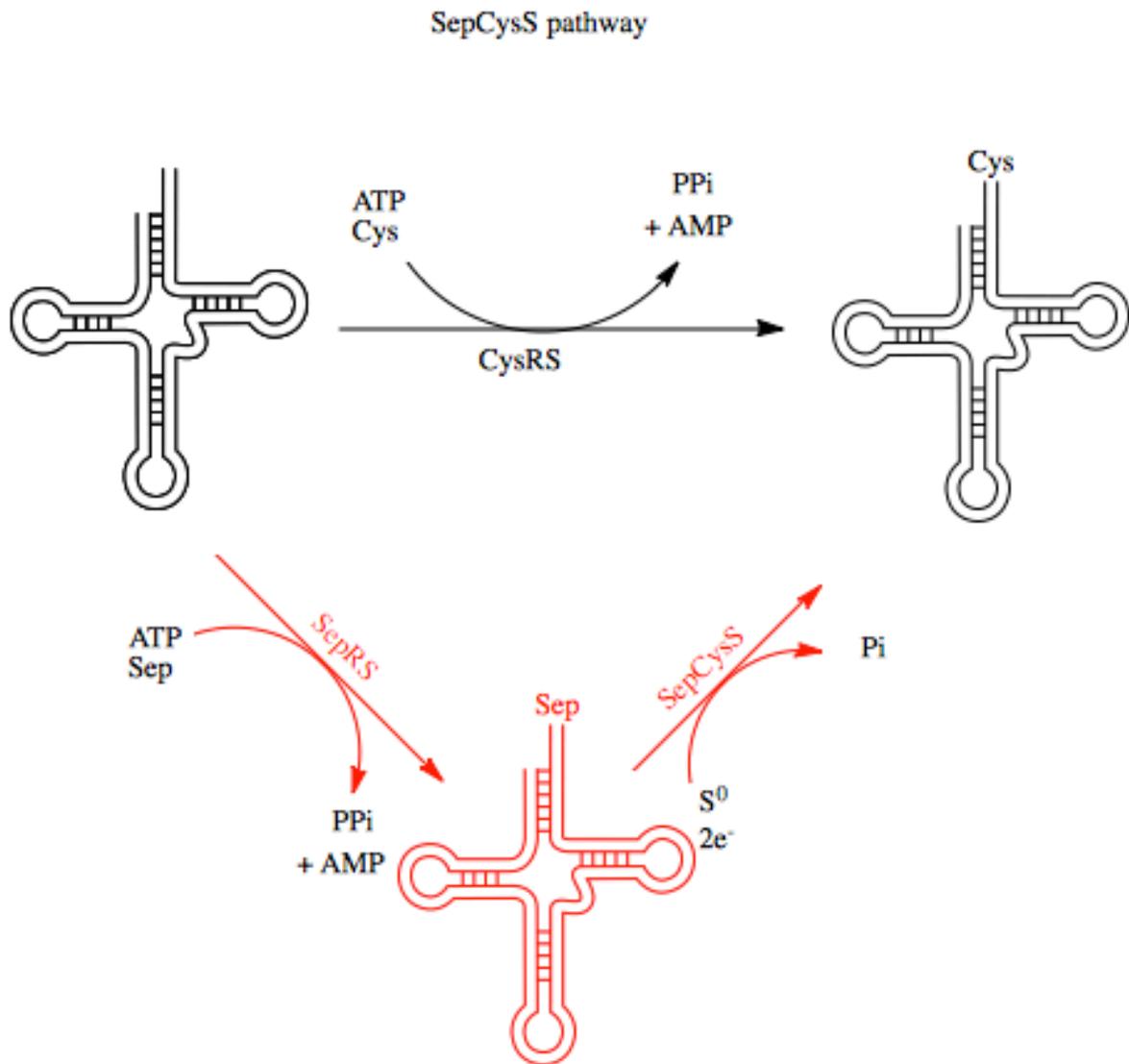


Figure 1.1 Novel SepCysS Pathway

The black pathway is the canonical pathway (CysRS) for aminoacylation of tRNA with cysteine. The pathway in red represents the unique two step aminoacylation process found mainly in methanogens. SepRS mis-aminoacylates tRNA^{Cys} with phosphorylated serine (Sep). Sep-tRNA^{Cys} is then converted to Cys-tRNA^{Cys} by SepCysS.

The trafficking of sulfur in most extant cells is commonly done through persulfide (R-S-S-H) groups¹³. The persulfide sulfur can have three different oxidation states as sulfane (S^0), persulfide (S^{1-}), or sulfide (S^{2-}) acting as an electrophile, nucleophile, or strong nucleophile respectively¹³ (**Fig 1.2**). Transfer of a terminal sulfur group from a protein to another protein or cofactor is done by a family of enzymes known as cysteine desulfurases (CD). CDs use L-cysteine as a substrate to form L-alanine and elemental sulfur¹⁴. Persulfide groups formed by CDs are used to make iron sulfur clusters and are essential in most extant cells¹⁵.

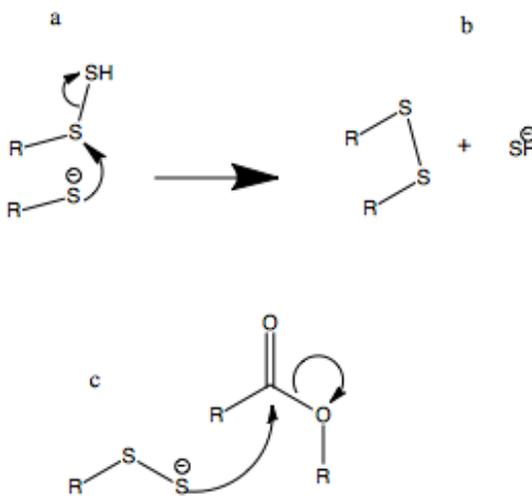


Figure 1.2 Persulfide Group Chemistry

- a) Persulfide atom acting as electrophile (S^0)
- b) Liberated persulfide (S^{1-})
- c) Persulfide sulfur atom acting as nucleophile (S^{2-})

While CDs are essential in most organisms, they are surprisingly absent from methanogens and many nonmethanogenic archaea¹⁶. The novelty of sulfur trafficking in methanogens, highlighted by the uptake of hydrogen sulfide rather than sulfate, the SepCysS pathway, and the lack of canonical cysteine desulfurases presents many unanswered questions about sulfur trafficking in methanogens. These extremophiles present a method of life we are very unfamiliar with. The importance of studying methanogens has never been greater with our recent journey to Mars where life, if any is found, may very well resemble the extremophiles found on Earth¹⁷.

Ancient Strategies For An Anaerobic World

Bacteria use cysteine as a primary sulfur source after taking up sulfate¹⁸. Upon uptake, sulfate can be reduced to sulfide in an ATP-dependent pathway and incorporated into cysteine by O-acetylserine sulfhydrylase (OASS) for further biosynthesis of various other sulfur containing compounds^{19,13} (**Fig 1.3**). The sulfur from cysteine can then be incorporated into proteins through the activities of CDs, or used directly to make Coenzyme A, glutathione and other metabolites.

In bacteria, cysteine can be converted to cystathionine and homocysteine by cystathionine gamma lyase and cystathionine beta lyase respectively, which are needed

for methionine (Met) synthesis, and S-adenosylmethionine (SAM) synthesis.⁸ A

shortcut exists to synthesize homocysteine directly from sulfide through O-

acetylhomoserine sulfhydrylase (OAHs). Cysteine can also be directly incorporated on

to tRNA through CysRS synthesizing Cys-tRNA^{Cys}.

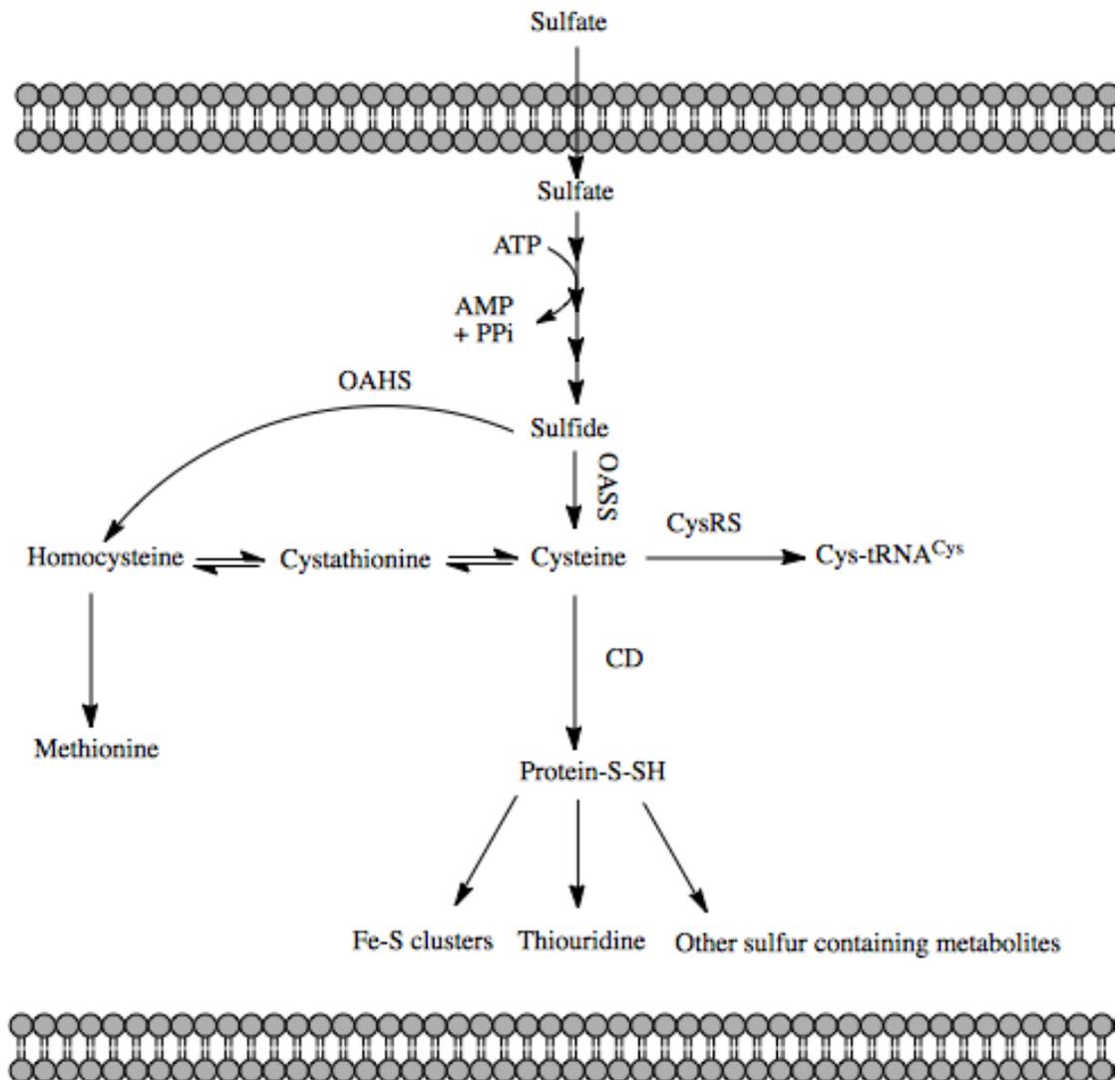


Figure 1.3 *E. coli* Sulfur Trafficking Pathways

Sulfur trafficking in bacteria, *E. coli* modeled above, consists of sulfate uptake for the biosynthesis of cysteine which acts as a sulfur source for other biochemical pathways. O-acetylserine sulphydrylase (OASS) traffics sulfur from sulfide to cysteine while O-acetylhomoserine sulphydrylase (OAHs) can use the sulfur from sulfide to make homocysteine. Cysteine can donate sulfur to persulfide proteins through cysteine desulfurases (CD) or be used to synthesize cysteinyl tRNA through the cysteinyl-tRNA synthase (CysRS) pathway.

Similar sulfur processing mechanisms are not likely to exist in methanogens as none of the necessary genes have been identified in many methanogen genomes. Instead sulfide is directly taken up and the destination immediately thereafter is still unknown. The identification of sulfur donors to unconventional sulfur pathway, such as SepCysS, will help elucidate sulfur assimilation in methanogens.

Sulfide, cysteine, and thiophosphate have all been proposed sulfur donors for SepCysS²⁰. Persulfide is another sulfur donor that has been proposed, based on the primary sequence SepCysS¹². SepCysS has structural similarity with CD²¹ which is known to use persulfide as a sulfur donor. To test if a persulfide was possibly the sulfur donor for SepCysS, the CD IscS from *E. coli* was radiolabeled with ³⁴S and shown to donate its sulfur to SepCysS²². An unknown persulfide carrier protein is likely the native sulfur donor to SepCysS and possibly other sulfur demanding pathways in methanogens (Fig 1.4).

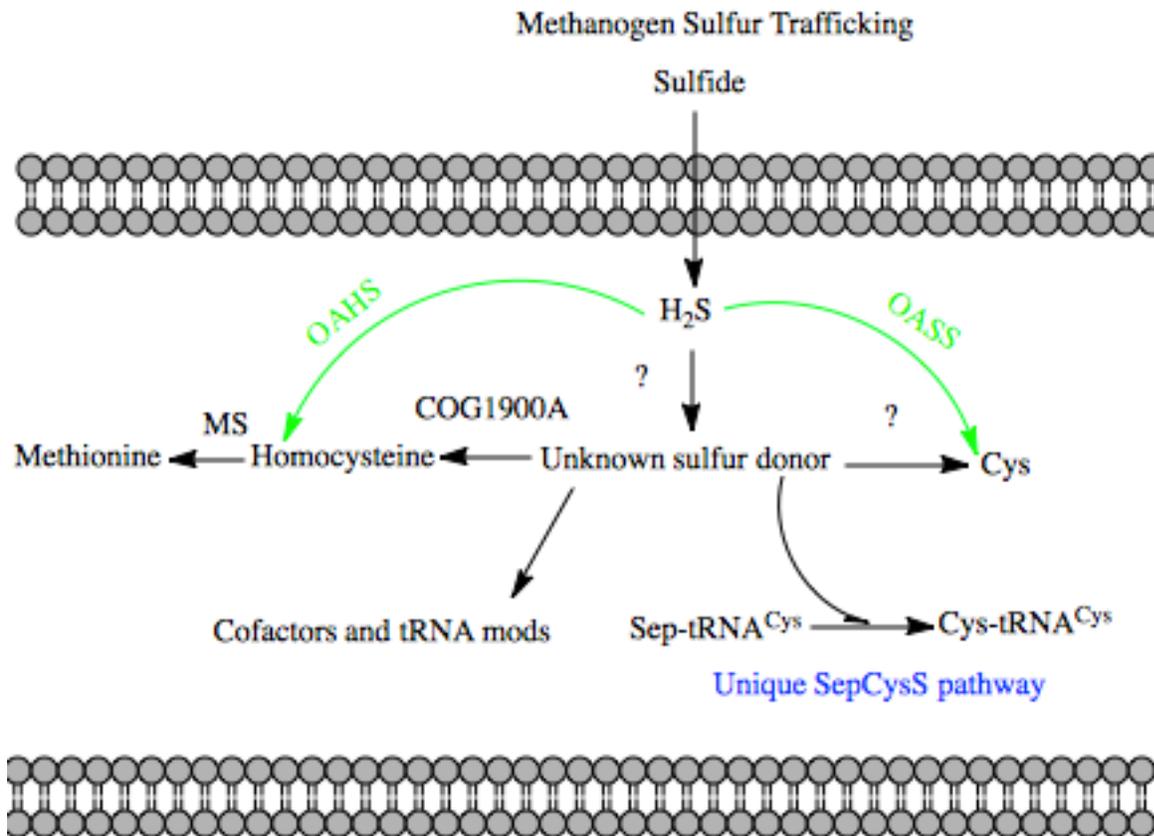


Figure 1.4 Methanogen Sulfur Trafficking

Sulfur trafficking in methanogens. This is a simplified metabolic map for comparison of the major differences between bacterial and methanogen sulfur trafficking. This representation is not all encompassing for simplicity. The green pathways of OAHS and OASS are conventional sulfur pathways in bacteria and are only found in a few methanogens, one being *M. acetivorans*.

Three Protein Families Co-conserved with SepCysS

Bioinformatics identified three conserved orthogonal genes (COG's) conserved in all methanogens that possess SepCysS, COG1900A, COG1900D, and COG2122²³ (Fig 1.5). Conservation with SepCysS suggests that these proteins are involved in sulfur trafficking.

Most methanogens lack conventional sulfur trafficking genes such as those for OAHS and OASS which synthesize homocysteine and cysteine respectively; the genome of *Methanococcus jannaschii* revealed many such metabolic gaps in methanogens²⁴. *Methanosarcina acetivorans* has the largest known archaeal genome²⁵ and contain many metabolic redundancies such as CysRS and the SepRS-SepCysS pathway. The *M. acetivorans* genome also contains the genes for OAHS and OASS²³, presumably redundancies of the methanogen genes that are responsible for homocysteine and cysteine synthesis in organisms lacking OAHS and OASS. Such genetic redundancy made *M. acetivorans* an ideal organism for gene knock out studies. By demonstrating that Δ OAHS Δ MA1821 (COG1900A) strains were auxotrophic for homocysteine and adding back MA1821 saved cells from homocysteine auxotrophy, the role of COG1900A as a homocysteine synthesis enzyme was identified²³.

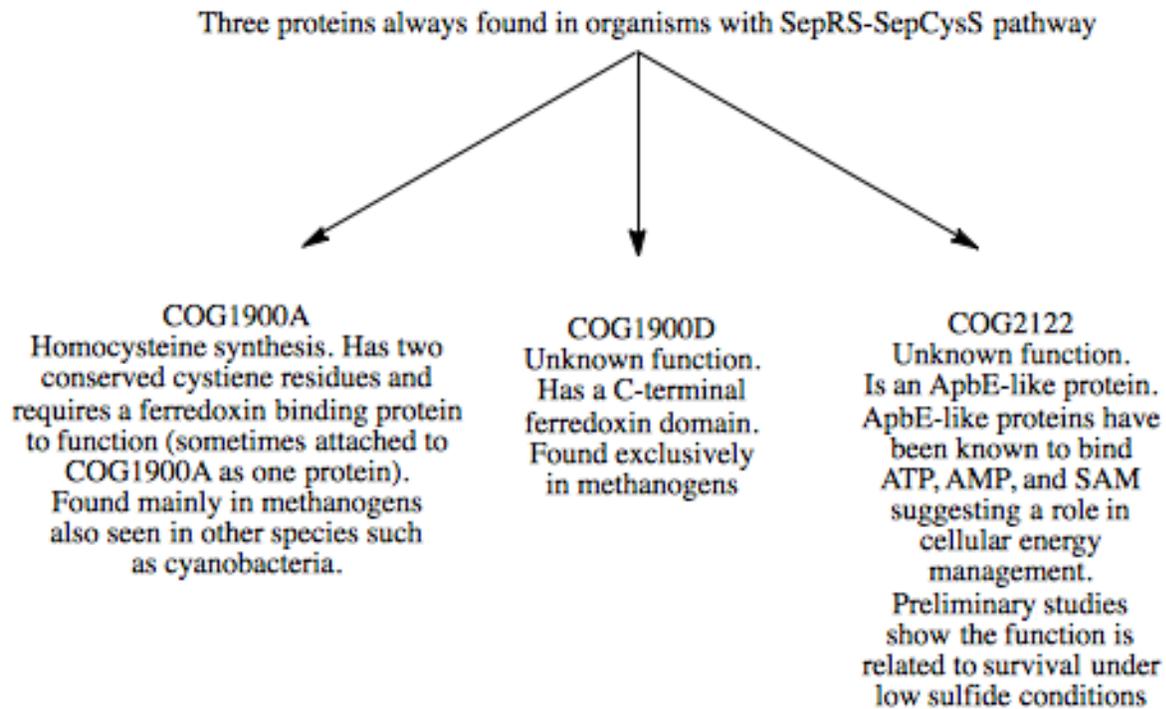


Figure 1.5 Protein Families Conserved with SepCysS

Three protein families are highly conserved in organisms that contain the SepRS SepCysS tRNA aminoacylation pathway. These three protein families might be involved directly in the SepCysS pathway, or could be involved in other important methanogen sulfur trafficking pathways.

Homocysteine Synthesis in Methanogens

Knocking out the OAHs gene in *M. acetivorans* produced a viable strain, suggesting the presence of another enzyme capable of filling the role of homocysteine synthesis. A double knockout of the OAHs gene and COG1900A (MA1821 gene) produces a homocysteine auxotroph, suggesting that COG1900A is capable of supplementing homocysteine synthesis in the absence of OAHs. *M. acetivorans* has a conserved ferredoxin protein (MA1822) that is necessary for MA1821 to function²³ and that is coded directly downstream of MA1821. Some organisms, such as *Thermotoga lettingae*, have both COG1900A and the ferredoxin protein linked as a fusion protein and coded by a single gene. The biochemical role of homocysteine synthesis by COG1900A proteins using aspartate semi-aldehyde has been confirmed with isotopic sulfur and whole cell lysate²⁶ (Fig 1.6).

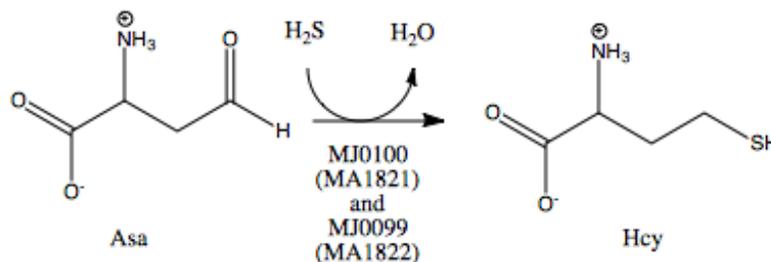


Figure 1.6 COG1900A Reaction

COG1900A catalysis of aspartate semi-aldehyde to homocysteine through conversion of an aldehyde to a thiol group.

M. jannaschii cells were shown to synthesize more homocysteine when the cell extract was supplemented with aspartate semi-aldehyde (Asa). When labeled sulfide (H^{34}S) was also added, almost all (89%) of the homocysteine (Hcy) in the extract was labeled. To further confirm that Asa was the substrate for Hcy, heavy Asa ([3,3- ^2H] Asa) was added which produced deuterium labeled Hcy²⁶.

The same biochemistry was confirmed with sulfide labelling and gene knock outs in *M. acetivorans*. Cell extract with supplemented Asa and labeled sulfide also produced 92% labeled Hcy. A mutated triple knock out strain (Δ OAHS, Δ MA1821 Δ MA1822) produced no labeled Hcy, while a Δ OAHS only strain produces labeled Hcy (67%)²⁶.

Connecting Homocysteine Synthesis with SepCysS

The reason for co-conservation of COG1900A, COG1900D and COG2122 is still unknown. If COG2122 can act as a persulfide donor, this would connect the COG1900A and SepCysS pathways, as both have an unknown sulfur donor and SepCysS is suspected to have a persulfide donor²². The only functional role of COG2122 that has been identified to date is its ability to rescue cells that are supplied with very low concentrations of sulfide²⁷. How it accomplishes this is still not known, but the ability to rescue cells from starvation is significant and suggests that the

COG2122 protein has a central metabolic role in assimilatory sulfur metabolism, connecting multiple pathways.

Under high sulfide conditions, MA1715 (COG2122) was not necessary for adequate growth of *M. acetivorans*. However, under low sulfide conditions (0.2-0.8 mM) growth was rescued with MA1715 present²⁷. This suggests that COG2122 optimizes available cellular sulfide, whether this is enhanced uptake, more efficient intracellular sulfur trafficking, or sulfur donations to certain pathways is still unknown.

Co-conservation of COG2122 and COG1900 proteins suggests a biochemical role between the two that is either necessary or complementary in some way. Knowing the COG1900A's function in homocysteine synthesis and COG2122's ability to rescue cells from sulfide starvation, COG2122's role likely involves sulfur trafficking; either directly supplying pathways such as those with COG1900 proteins, or indirectly through cellular sulfur storage.

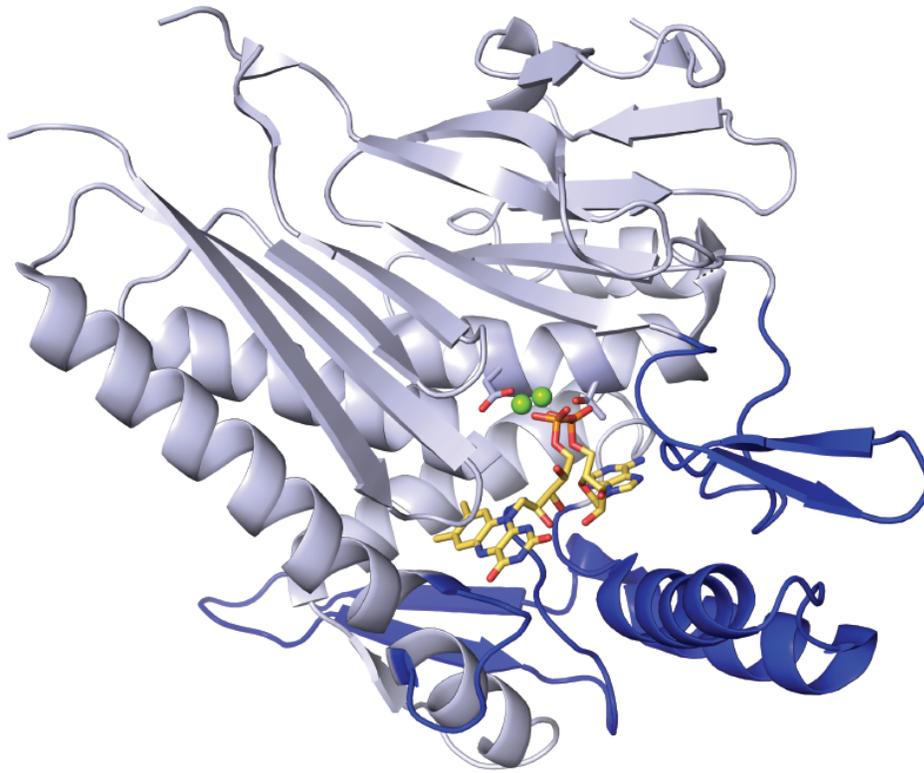
COG2122; ApbE Protein Superfamily of Flavin Trafficking Proteins

COG2122 proteins fall under the ApbE like protein superfamily. The ApbE homolog in *Salmonella typhimurium*, was suspected to be a periplasmic lipoprotein involved in 4-amino-5-hydroxymethyl-2-methyl pyrimidine (HMP) synthesis which ultimately feeds into the formation of thiamine²⁸, a sulfur containing compound.

Strains carrying mutant ApbE enzymes were found to be thiamine deficient. The thiamine biosynthetic pathway is a cytosolic pathway which is not consistent with ApbE proteins being located in the periplasm²⁹.

The crystal structure of *Salmonella enterica* ApbE protein revealed that it is an FAD-binding protein³⁰. The presence of an ApbE homolog in the parasitic species *Treponema pallidum*, which is suspected to lack thiamine biosynthesis pathways, prompted further research into the catalytic reaction of this superfamily. The crystallization and structure determination of *T. pallidum* ApbE protein showed it to be a bimetal-dependent FAD pyrophosphatase³¹.

ApbE proteins cleave the phosphate bond in FAD producing AMP and FMN. There is good reason to believe that COG2122 does not share this biochemical role. The FAD binding domain identified on the *T. pallidum* ApbE protein is conspicuously absent from the COG2122 protein family^{31,32} (**Fig 1.7**). While this is not definitive proof that FAD does not bind COG2122, it raises the possibility of the protein having an entirely different biochemical function.



■ ApbE regions absent from COG2122

Figure 1.7 ApbE Protein Ribbon Diagram

Figure generated by C.M. Driggers (unpublished) based on the ApbE protein from *Treponema pallidum* with FAD-bound, PDB coordinates IVRM³¹.

The COG1900 Protein Family

The COG1900 protein family is one of many protein families in methanogens that harbors an iron sulfur cluster domain. While iron-sulfur clusters are highly protected by intracellular compartments in aerobic cells due to the threat of oxidation, methanogens have a larger fraction of their genome coding for iron sulfur cluster proteins³³ compared to aerobic organisms. In an anaerobic environment this is not a problem as the iron sulfur clusters are not in danger of oxidation. Iron sulfur clusters are known to form spontaneously given anaerobic conditions rich in iron and sulfur^{34,35}; conditions similar to the early Earth³⁶. Given the proper environment of the early Earth and the genomic abundance of iron sulfur cluster binding proteins in methanogens, the iron sulfur world hypothesis has developed³⁷, suggesting that iron sulfur clusters are possibly the first biological catalysts and played a significant role in evolutionary history.

Little is known about COG1900 proteins beyond COG1900A's conversion of aspartate semi-aldehyde to homocysteine²⁷. An aldehyde to thiol conversion has a redox requirement if the mechanism requires the breaking and forming of disulfide bonds (**Fig 1.8**); this explains the necessity of the iron sulfur cluster binding protein gene always present near COG1900A on the genome. Interestingly, COG1900A has two conserved cysteines but only one (Cys54) is essential²³. There are four paralogs of

COG1900 proteins (A-D) as determined by phylogenetic analysis²³. Only COG1900A's function has been identified to date, with the biochemical function of COG1900B-D proteins remaining unknown.

COG1900 Paralogs

COG1900A is the largest of the four paralogs at 500 amino acids in length. The larger size is largely due to the inclusion of a C-terminal cystathionine beta synthase (CBS) domain. CBS domains, like that of inosine-5'-monophosphate (IMP) dehydrogenase

from *Streptococcus pyogenes*, consist of symmetrical, antiparallel β_1 - α_1 - β_2 - β_3 - α_2 structures, and are outside of the catalytic center³⁸. Proteins containing CBS domains range from chloride channels³⁹, nucleotide synthesis pathway enzymes⁴⁰, and even AMP-activated Protein Kinase (AMPK)⁴¹⁻⁴³. CBS domains have been found to bind ATP, AMP, and S-adenosylmethionine which has led to the hypothesis that proteins with CBS domains act as cellular energy sensors and hence play a regulatory role^{44,45}.

All COG1900A proteins have two conserved Cys residues in their catalytic domains; the conserved residues are Cys54 and Cys131 in MA1821. The MA1821 C54A mutation is non-viable while a C131A mutation is. This suggests that Cys54 is directly involved in the mechanism of homocysteine biosynthesis.

COG1900B and C are only found in cyanobacteria²³. Here Cys54 is substituted as Glu, and Asp respectively. COG1900D on the other hand, is exclusive to methanogens. COG1900D possess a [4Fe-4S] domain on the C terminus. The shared need for an iron sulfur cluster between COG1900A and COG1900D suggests the

possibility of a similar biochemical mechanism. COG1900D is exclusively found in methanogens, as opposed to COG1900A which is also found in many other anaerobic bacteria. This suggests COG1900D has a role directly involved in methanogenesis. Reconstruction of phylogenies with respect to COG1900 proteins are in agreement with current methanogen phylogenies, suggesting the proteins were vertically inherited with methanogenesis pathway²³.

A Possible Biochemical Role for COG1900D and Its Implications

COG1900A is capable of converting an aldehyde to a thiol; should COG1900D have a similar biochemical capability, it could be the last missing enzyme in the coenzyme M (CoM) pathway (**Fig 1.9**). A comparison of the known COG1900A biochemistry with the required biochemistry in the final step of CoM synthesis, reveals an analogous reaction taking an aldehyde to a thiol (**Fig 1.10**). CoM is an essential cofactor involved in methane synthesis. An understanding of methanogenesis is of great interest to many fields from environmentalists, those in energy production, evolutionists, astronomers, medical, and many others.

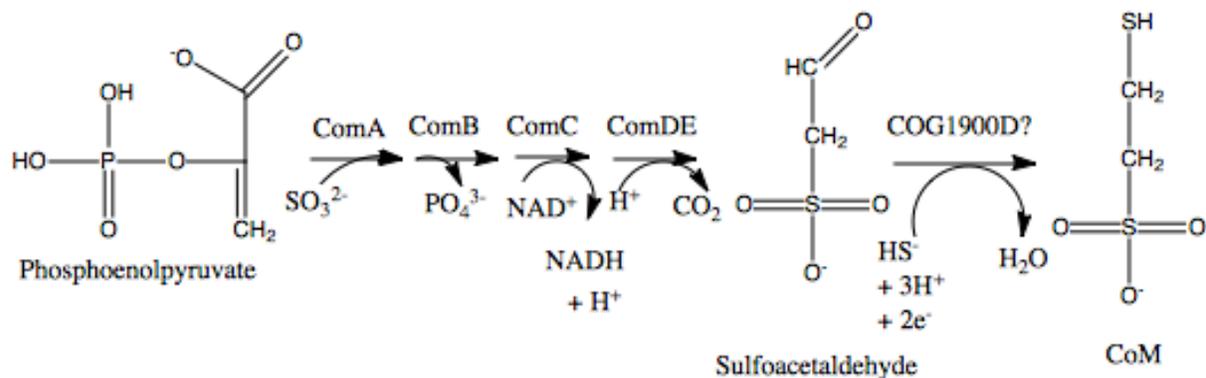


Figure 1.9 Coenzyme M Biosynthesis Pathway⁴⁶

The final step of coenzyme M biosynthesis involves the conversion of an aldehyde to a thiol. COG1900D may be the final enzyme in the pathway, using sulfide to convert sulfoacetaldehyde to coenzyme M.

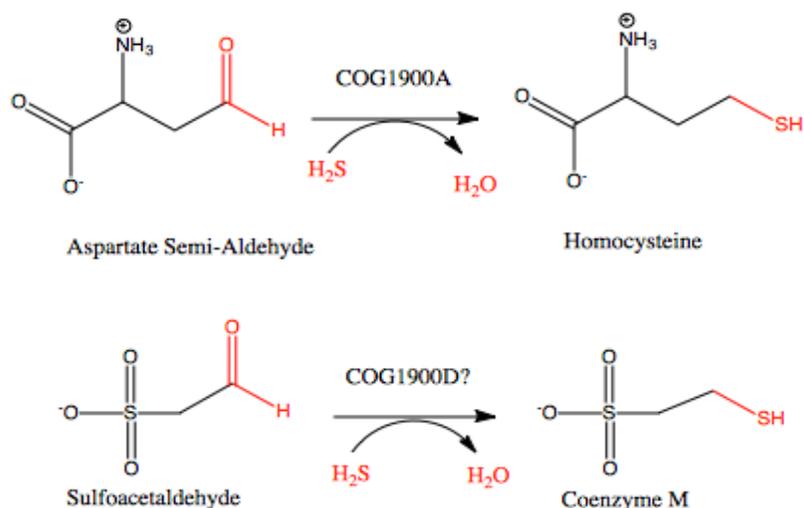


Figure 1.10 COG1900A and Predicted COG1900D Reactions

COG1900A's reaction is known. If COG1900D is the last enzyme in CoM synthesis, the reaction it catalyzes would be analogous to the reaction catalyzed by COG1900A.

Methane is a potent greenhouse gas; while it was essential for developing an atmosphere that could maintain habitable temperatures⁴⁷, it is now contributing to unwanted global warming⁴⁸. An understanding of methane synthesis may reveal means by which we can inhibit its production in methanogens. One possible application of such advancements could be the supplementation of a COG1900D inhibitor in the diet of livestock. Methanogens present in livestock produce large amounts of methane that greatly contributes to global warming⁴⁹. A shift of human's meat intake is not only unlikely, but an increase in agriculture comes with its own problems; at the same time global food demands are at an all-time high and only increasing⁵⁰. A reduction in the methane produced in livestock would have long lasting benefits.

As for more industrial purposes, methane is a potential energy source and biogas which is already being utilized in some nations such as Germany⁵¹. Wastewater treatment is another growing global problem that methanogenesis can help resolve. Anaerobic digestion of wastewater by methanogens is an efficient way to clean up water sources and create methane as an energy source^{52,53}. A better understanding of methanogenesis may also have medical benefits. Beyond the detrimental effects to human health caused by global warming⁵⁴, the human microbiome contains methanogens and the microbiomes effect on health is increasingly apparent⁵⁵.

Methane, while a major contributor to global warming, was essential for creating an environment suitable for present life³⁶. For this reason, methanogenesis is an extremely important process for astronomers looking for potential life on other planets. At the same time it holds a lot of value to evolutionists as a methanogen (species not identified) is suspected of being Earth's last common known ancestor⁵, and the environment created through methanogenesis shaped all subsequent evolution. A clear understanding of methane synthesis and all the enzymes involved in the process, will benefit many diverse fields with important implications. The largest obstacle so far with this regard, has been the identification of the last enzyme responsible for CoM synthesis, and the purification necessary for *in vitro* studies of such sulfur trafficking proteins (COG1900 family) in methanogens.

Purpose of Studying the Proteins in This Study

Many methanogens are extremophiles. The ability of extremophiles to survive in their environments is of interest as they represent the known limits in which life can survive; extraterrestrial life may rely on similar mechanisms for survival⁵⁶. Methanogen proteomics are very poorly understood, one major obstacle being the purification of their proteins. The ability of extremophile proteins to maintain functionality at such high temperatures (45 °C – 110 °C) is currently not well understood beyond the

general pattern of tight packing, ion pairs, salt bridges, and increased hydrogen bonding^{41,57-60}. Extremophile proteins are also known to have higher order oligomeric states and large hydrophobic cores, possible contributing to the challenge involved with their purification⁶¹.

Methanogen sulfur trafficking starting with the uptake of sulfide as opposed to sulfate, is unique and not very well understood. COG2122 proteins are of interest because they have been shown to play a key role in survival under low sulfide conditions²⁷. COG2122 might be involved directly in sulfide uptake, sulfur trafficking, or it could be a signal transduction protein activating a cascade of sulfur assimilation pathways in methanogens; a protein playing any of these roles in methanogens would prove of great interest.

COG1900 proteins are of interest because COG1900A is known to catalyze the synthesis of a thiol from an aldehyde^{23,26} and no COG1900 structure has been solved to date. Purification results to date have also suggested that COG1900 proteins prefer a higher order oligomeric state and are insoluble under most conditions, characteristic of extremophile proteins⁶². A purification method for COG1900 proteins might prove useful for other extremophile proteins that are difficult to purify, possibly for the same reasons. Ideally, a purification strategy for methanogen proteins out of *E. coli* will be developed. *E. coli* are the preferred protein factories because they grow quickly on

inexpensive growth media, there are many expression plasmids designed for them, there are many engineered strains for recombinant protein expression, and *E. coli* metabolism has been extensively studied⁶³.

Chapter 2: Protein Purification Methods

Identifying Proteins of Interest

The identification of COG1900 and COG2122 proteins was accomplished with occurrence profiling. Also referred to as phylogenetic profiling⁶⁴, this method of bioinformatics assumes that genes with interdependent function will be present or absent from genomes of genes with known function. This helps identify potential genes of interest based on co-conservation.

Using the COG database⁶⁵, which is available for free online, a catalogue of all methanogen genes co-conserved with SepCysS was constructed²³; all methanogen genes in genomes without SepCysS were then removed to simplify the list. The group was then narrowed by retaining only genes that were encoded near, or homologous to known sulfur assimilation proteins⁶⁶. This left three genes of interest, COG1900, COG2122, and the gene coding for the small ferredoxin protein always neighboring COG1900A proteins.

Engineered *M. acetivorans* Strains

COG1900A purification has been attempted by Ben Rauch using engineered *M. acetivorans* strains⁶⁶. Genetic manipulations were done using an *M. acetivorans* pseudo wild type strain (WWM75) from the laboratory of William Metcalf⁶⁷, a plasmid for markerless genetic exchange (pMP44)⁶⁸, a shuttle vector (pWM321)⁶⁹, and a promoter for tetracycline-dependent gene expression ($P_{mcrB}(\text{tetO1})$)⁶⁷.

Expression was accomplished using a pBR31-derived, tetracycline-dependent expression plasmid (pBR70) along with an engineered Hcy auxotroph lacking genes for MA1821 and MA1822 (COG1900A); this strain permitted higher expression levels than the pseudo wild type WWM75 and both MA1821 and MA1822 had C-terminal poly-histidine tags⁶⁶. The doubling time of *M. acetivorans* is six hours and growth must be under anaerobic conditions.

Recombinant expression of MA1821-22 from *M. acetivorans* proved to be extremely time consuming and only yielded 0.2 mg of purified protein from one liter of cells⁶⁶. Expression in *E. coli* was also tested⁶⁶. Because of such drawbacks to purification from the native organism, *E. coli* cells are the preferred organism for overexpression; the development of a purification strategy from *E. coli* would be very beneficial.

Engineered *E. coli* Strains

In an effort to work around the impracticality of recombinant protein expression in *M. acetivorans*, *E. coli* strains were engineered to overexpress COG1900, COG2122, and the small ferredoxin proteins. Multiple expression plasmids were constructed to express COG1900A (pBR004, pBR038, pBR096, pBR100 and pBR102)⁶⁶. A MA1821-22 co-expression strain (pBR039)⁶⁶ and a COG1900A-ferredoxin fusion protein from *Thermotoga lettingae* (pBR104)⁶⁶ were engineered to test if COG1900A solubility was dependent on the essential ferredoxin protein. A COG1900D strain (which also fuses the COG1900 and ferredoxin domain as one protein) was also engineered (pBR117)⁶⁶.

A strain for expressing MA1715 (COG2122) was also made (pBR006)⁶⁶. Strains pertinent to this thesis are pBR004, pBR006, pBR039, pBR104 and pBR117, all engineered by Ben Rauch⁶⁶. The plasmid design and any alterations to the protein (poly histidine tag etc.) will be discussed in their respective sections to follow.

Previous COG1900 Purification Attempts from *E. coli*

Previous attempts to purify COG1900 (A and D) proteins as well as the small ferredoxin protein were attempted from *E. coli* with no success (Table 2.1)⁶⁶. None of the conditions tested were able to produce sufficient quantities of soluble protein (>1 mg). Insoluble pellets of COG1900A were resolubilized with ionic detergent (sarkosyl)

and COG2122 proteins were able to be refolded⁶⁶ using the two step refolding method described later in this chapter. Prior purification attempts include the COG1900A protein from *Thermotoga lettingae* (Tlet_1363) which has the COG1900A domain and the required ferredoxin domain attached as one protein. The proteins with purification strategies presented in this thesis are described (**Table 2.2**) and a figure of their domains (**Fig 2.1**) can be found below.

Plasmid	Protein Encoded	Source	Parent Plasmid	Host Strain	Expression Conditions
pBR004	COG1900A-CBS-H6	MA1821	pET-22b(+)	Rosetta2(DE3)pLysS Arctic Express	24 hr at 16 C 5 hr 37 C
pBR005	NIL-Fer-H6	MA1822	pET-22b(+)	Rosetta2(DE3)pLysS	25 hr at 16 C 5 hr 37 C
pBR038	COG1900A-CBS; NIL-Fer-H6	MA1821-22	pET-22b(+)	Rosetta2(DE3)pLysS	5 hr at 37 C 24 hr at RT, anaerobic
pBR096	COG1900A-Fer-H6	Tlet_1363	pET-22b(+)	BL21(DE3)pLysS	24 hr at 16 C 5 hr at 37 C 24 hr at RT, anaerobic
pBR100	H6-SUMO-COG1900A-Fer	Tlet_1363	pET-22b(+)-SUMO	BL21(DE3)pLysS Arctic Express	25 hr at 16 C 5 hr at 37 C 24 hr at RT, anaerobic 24 hr at 10 C
pBR113	H6-SUMO-COG1900D-Fer	MJ1681	pET-22b(+)-SUMO	BL21(DE3)pLysS	5 hr at 37 C 24 hr at RT, anaerobic
pBR006	COG2122-H6	MA1715	pET-22b(+)	Rosetta2(DE3)pLysS	5 hr at 37 C 24 hr at RT, anaerobic
pBR089	H6-SUMO-COG2122	DVU1097	pSGX4	BL21(DE3)pLysS	24 hr at 16 C 24 hr at RT, anaerobic

Table 2.1 Previous COG1900 Purification Attempts from *E. coli*

All purification attempts listed in the table were completed by Ben Rauch and Camden Driggers⁶⁶ prior to this thesis.

Protein	Family	Biological Function	Characteristics
MA1821	COG1900A	Synthesizes homocysteine from aspartate semialdehyde.	500-amino-acids in length. Contains a C terminal CBS domain (380-494) which are known to bind together forming globular domains, the N terminus is the COG1900A domain (4-362).
MA1822	COG1900A	Co-conserved with MA1821, performs the redox chemistry in homocysteine synthesis.	128-amino-acids in length. The N terminal domain contains an NIL domain (5-67), a domain proposed to bind substrate in methionine ABC transporters. The C terminal domain is part of the Fer4_7 superfamily (4Fe-4S). It is almost always co-conserved with MA1821 orthologs.
Tlet_1363	COG1900A	Suspected to synthesizes homocysteine from aspartate semialdehyde.	443-amino-acids in length. N terminal COG1900A domain (5-367). The C terminus of this COG1900A ortholog has the 4Fe-4S domain fused with the COG1900A portion instead of being a separate peptide (394-437).
MJ1681	COG1900D	Suspected to be the last enzyme in coenzyme M synthesis, converting sulfoacetaldehyde to coenzyme M.	380-amino-acids in length. N terminal COG1900D domain (1-292). The C terminal domain has the 4Fe-4S domain fused to it (295-339).
MA1715	COG2122	Likely involved in sulfur trafficking within the cell. Has been demonstrated to support growth during starvation conditions (low sulfide).	253-amino acids in length. The COG2122 domain is uncharacterized, but falls under the ApbE like family. Some ApbE proteins convert FAD to FMN and AMP. The domain that has been identified as the FAD binding domain of the protein is missing from COG2122, suggesting a different purpose in methanogens.

Table 2.2 Proteins Used in This Study

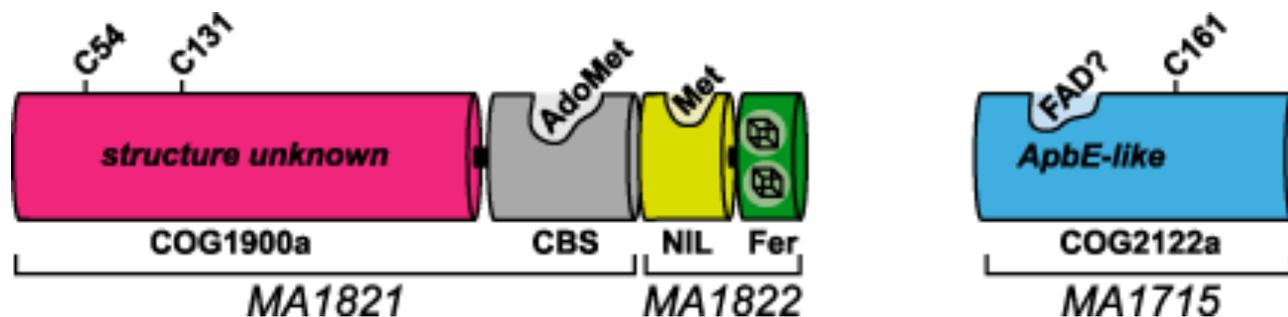


Figure 2.1 COG1900A and COG2122 Protein Domains²³

Protein Purification Using Overexpressed Systems

Purification of desired proteins is essential for *in vitro* studies. *E. coli* has become the organism of choice for recombinant protein production. While substantial quantities of overexpressed proteins may be obtained, many hurdles are present in the overexpression of any non-native protein. The target protein may form inclusion bodies, may not be active, or even be truncated due to transcription errors. To address these issues many techniques have been developed to help stabilize recombinant proteins for purification.

COG2122 and COG1900 proteins are no exception when it comes to purification hurdles. Proteins from both families are susceptible to transcription, solubility, and folding issues. Utilizing a small ubiquitin-related modifier (SUMO) tag system, optimizing expression conditions, the use of an unusual lysis buffer, and even refolding are necessary to obtain proteins from these families (**Table 2.3**). Purification details of each protein will be discussed in their respective chapters. First, a review of the purification techniques used in this thesis is needed.

	MJ1681	MA1821	TL1363	MA1715	MA1822
Refolding		-		+	+
SUMO tag	+				
Arctic cells with Chaperone			-		
37 °C expression	-	-	-	+	+
16 °C expression	+	+	+		+
Freeze/thaw lysis	+				
Sonication lysis	+	+	+	+	+
pH 13 soluble	+	+	+	-	-
pH 13 resuspension	+	-		-	-
Ni-NTA	+	+	+	~	+
Anion IEX	-			+	+
Cation IEX	+				
Gel Filtration	+	-		+	+
Stable in HEPES	~			+	

Table 2.3 Purification Strategies Tested

All protein used was expressed and purified from *E. coli* cells under aerobic conditions. A value of (+) means the method works for the corresponding protein, a (-) value means it does not work, the value (~) means it will work but not well, and a blank means the technique was never tested.

Refolding

Overexpressed proteins are often insoluble. This can be due to the formation of inclusion bodies (amorphous accumulation of proteins)⁷⁰ or because the lysis buffer is missing an essential additive that would support solubility⁷¹. If additives have been screened with no success, the insolubility is likely due to the formation of inclusion bodies. If inclusion bodies are forming, and it is not practical to try expression in another system (native organism for example), the protein must be refolded.

Traditionally this has been accomplished by resuspending an insoluble pellet of inclusion bodies in a chaotropic (disordering) agent such as urea followed by slowly dialyzing out the chaotropic agent to allow the protein to find its proper conformation. Successful refolding of proteins has been enhanced using a two-step denaturing procedure⁷², which will be described below in the purification details of COG1900D. Refolding a recombinant protein is typically the last resort after screening lysis additives. An alternative to screening a plethora of additives or attempting refolding is the use of protein tags and chaperones.

Protein Tags, Chaperones, and Expression Conditions

If a recombinant protein is forming inclusion bodies, it might be possible to prevent such aggregation by cloning a cleavable small ubiquitin modifier (SUMO) tag to either the N or the C terminus. Unlike ubiquitin, SUMO conjugation to peptides is not a label for degradation, rather it is known to help stabilize proteins⁷³⁻⁷⁸.

Deconjugation of SUMO can be accomplished by multiple enzymes, the most popular being ubiquitin like protease (ULP)⁷⁹.

Using a His₆ modified SUMO tag as well as a His₆ modified ULP, stabilization of a target recombinant protein can often be achieved through cloning a modified SUMO tag to the N-terminus of the target protein, purifying the conjugated protein through Ni²⁺ affinity chromatography, cleaving the SUMO tag with the modified ULP, and finishing with a second Ni²⁺ affinity column eluting the desired protein with no tag while removing the SUMO tag and ULP⁸⁰. This purification system is highly desirable as it produces purified target protein with no His₆ tag or large fusion peptide which can alter the proteins structure and or function.

The utilization of chaperone proteins is another means of preventing inclusion body formation or misfolding during overexpression. ArcticExpress *E. coli* cells have been developed that express the chaperonin system Cpn60/10 from *Oleispira*

*antarctica*⁸¹. This chaperone system can help fold some proteins that misfold during overexpression.

A chaperonin system may be necessary for proper expression of recombinant protein, however it is known that expression of recombinant proteins in *E. coli* at lower temperatures improves proper folding⁸². Simply inducing *E. coli* cells at an earlier optical density than the conventional $OD_{600} = 0.6$, using less inducer (IPTG in T7 promoter cells is very common), and expressing at lower temperatures such as 16 °C can save overexpressed proteins from aggregation and inclusion body formation.

Chapter 3: Purification of COG1900D Proteins

COG1900 Initial Purification Observations

To date there has been no reported purification of any COG1900 protein. Attempts have been noted in the literature however, with a group attempting to purify COG1900A from *Methanocaldococcus jannaschii* (MJ0100)⁸³. The CBS domain of MJ0100 was eventually crystallized by removing the COG1900 portion of the protein. Attempts at the full COG1900A purification were too difficult: “Previous studies suggested that the full-length MJ0100 has a strong tendency to aggregate, making physical studies of the protein very difficult”⁸⁴. The previous study mentioned did not give any examples of attempted purification⁸³. The ability to solubilize the CBS domain suggests that the instability and insolubility is due to the COG1900 domain.

Working with COG1900A and COG1900D (MA1821 and MJ1681 respectively), both proved to be completely insoluble in conventional lysis

buffers. Although they are homologues, they have enough variation in their primary sequence to believe their solubilization will not be identical. The two proteins purified have 28% sequence identity and varying domains. MA1821 contains an C-terminal CBS domain while MJ1681 has a C-terminal ferredoxin domain (**Fig 3.1**). The size, predicted pI, and extinction coefficient of the COG1900 proteins purified in this study are all unique (**Table 3.1**).

Analysis of the amino acid composition of MA1821 in comparison to average cytoplasmic and membrane proteins, has the protein aligning primarily with cytoplasmic proteins (**Fig 3.2**). A notable exception is fewer charged residues are seen in MA1821 when compared to the average cytosolic protein (arginine, glutamic acid, and aspartic acid). MJ1681 corresponds more strongly with the average cytosolic protein with respect to Arg, Glu, and Asp, but strongly differs with a much higher frequency in lysine residues⁸⁵ (**Fig 3.3**). The high lysine frequency is very common in the *M. jannaschii* proteome⁸⁶. This is consistent with the observation that proteins from hyperthermophiles have a high frequency of charged, hydrophobic, and large residues⁸⁷. The increased

frequency of hydrophobic residues may contribute to the challenge in purifying COG1900 proteins.

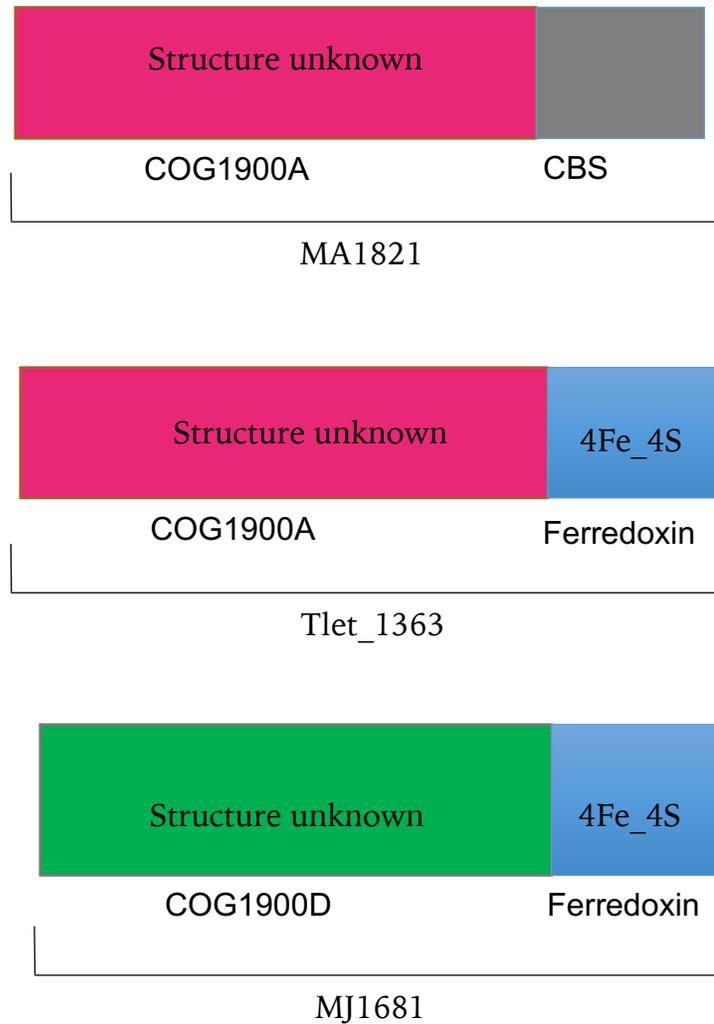


Figure 3.1 Domain Comparison Between MJ1681 and MA1821

	MJ1681	Tlet_1363	MA1821
Amino Acids	380	443	500
Molecular Weight (kDa)	42.58	48.74	54.34
Theoretical pI	7.39	6.84	5.92
Extinction Coefficient ($M^{-1}cm^{-1}$)	24130	38445	48945
Absorbance 0.1% (ϵ 0.1%)	0.567	0.789	0.901

Table 3.1 COG1900 Protein Profiles

The parameters were calculated using an ExPasy Prediction tool⁸⁸.

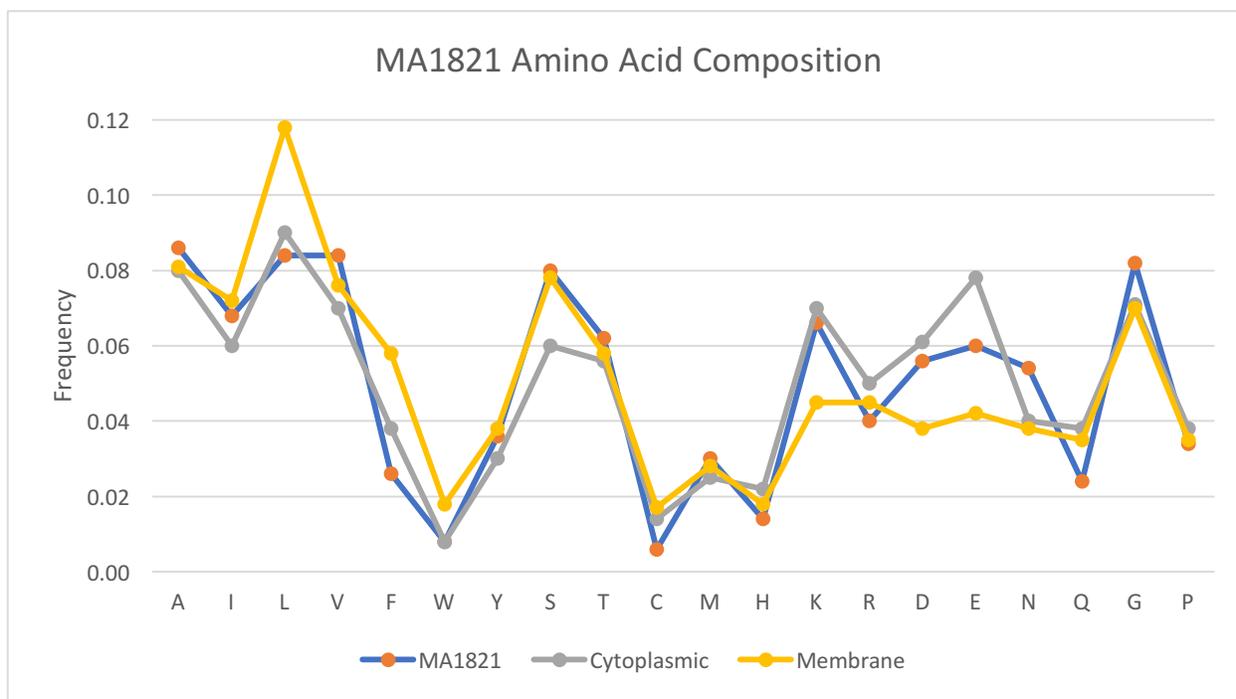


Figure 3.2 MA1821 Amino Acid Composition

The frequency of each amino acid in the primary structure of MA1821 compared to the averages of cytoplasmic and membrane proteins. Cytoplasmic and membrane proteins averages include representation from all three branches of life.

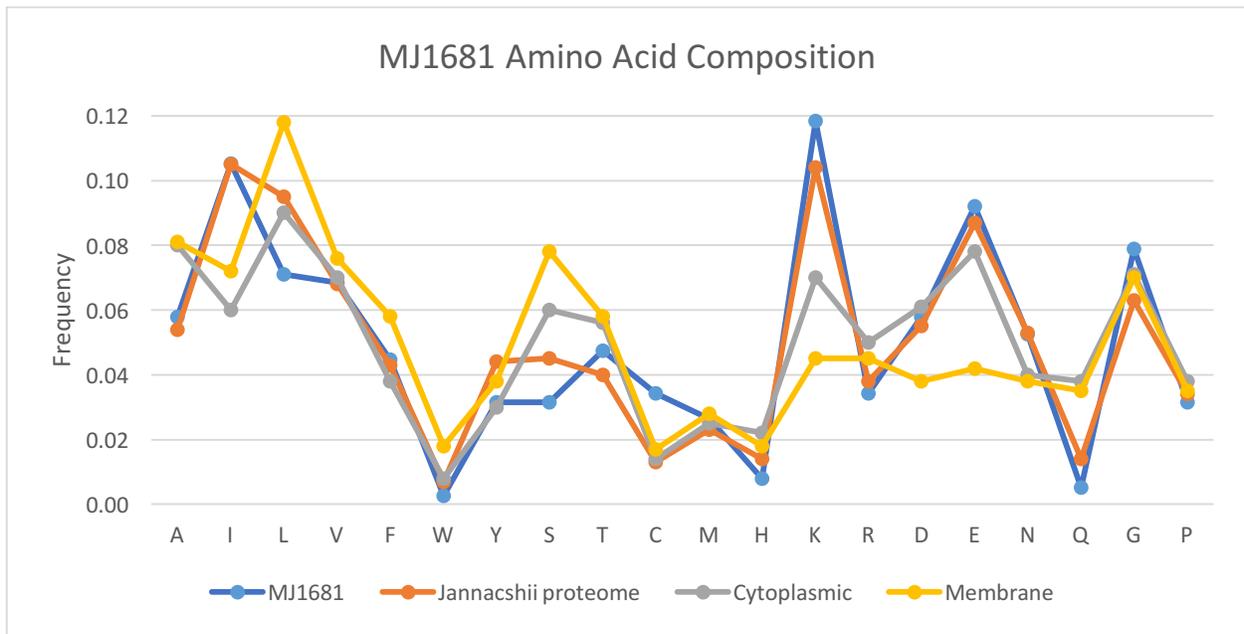


Figure 3.3 MJ1681 Amino Acid Composition

The frequency of amino acids in the primary structure of MJ1681 compared to the average for cytoplasmic, membrane, and jannaschii proteins. Cytoplasmic and membrane proteins averages have representation from all three branches of life.

COG1900D Solubilization

MJ1681 protein was purified from *E. coli* cells under aerobic conditions using strain pBR117⁶⁶. An pET22b(+) vector had the MJ1681 gene with an N-terminal SUMO (N-terminal poly histidine tagged SUMO) tag inserted using BamHI and XhoI restriction sites; the plasmid was transformed into BL21 (DE3) cells (**Fig 3.4**).

Identification of MJ1681 was accomplished by inducing cells containing the pBR117 with various amounts of Isopropyl β -D-1-thiogalactopyranoside (IPTG) (**Fig 3.5**).

Conventional lysis buffers failed to produce any soluble MJ1681 (COG1900D) based on previous attempts and buffer screening including buffers with DTT to reduce any unwanted disulfide bonds (data not shown). A screen of lysis conditions was done to test a wide array of pH, stabilizers, and kosmotropic agents (**Fig 3.6**). It is hypothesized that kosmotropic agents are capable of helping stabilize proteins in solution, however it should be noted that kosmotropic agents are still poorly understood⁸⁹. Of forty tested lysis conditions, just one solubilized MJ1681. The pH of the lysis buffers was adjusted before the additives were presented causing some pH fluctuation. The successful lysis buffer consisted of 50mM HEPES pH 8, 200mM NaCl, 100mM potassium phosphate tribasic (the phosphate adjusted the final pH of the buffer to 12.7) (**Fig 3.7**).

Cells used in screening conditions were grown to an OD₆₀₀ of 0.6-0.7 and MJ1681 expression was induced using 0.25mM IPTG and expression was done for 2h at 37 °C. The solubilizing ability of the buffer might be attributed to the kosmotropic (order-making) properties of phosphate⁴⁷. It is hypothesized that molecules with a high ionic density help stabilize proteins⁹¹. A kosmotropic agent reduces the density of water molecules while a chaotropic agent increases the density. Kosmotropic agent's ability to help stabilize proteins can be understood then as reducing the interaction of water with sensitive regions of the protein such as exposed hydrophobic regions. The smaller and hence denser the ion is, the better it is as a kosmotropic agent⁹². At a pH of 12.7, 57.43% of phosphate molecules are completely deprotonated, creating a very dense negative charge with strong kosmotropic properties (**Table 3.2**). The alkalinity of the solution can cause dehydro reactions and cross linking of amino acids on proteins and this is a point of concern when preparing a protein under such extreme pH conditions⁹³.

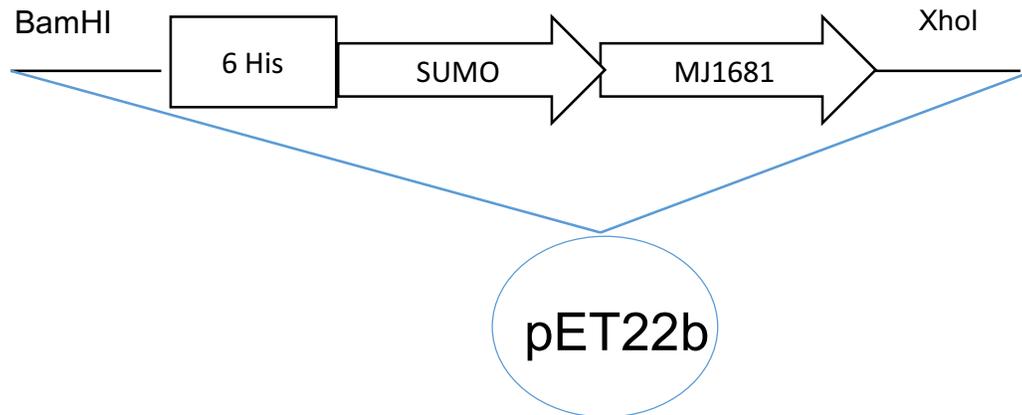


Figure 3.4 Expression Plasmid for MJ1681 Generation

MJ1681 was expressed aerobically in *E. coli* using a SUMO tag on the N-terminus with a six histidine residue tag attached to the N-terminus of the SUMO tag. The genes were transformed into BL21 (DE3) cells on a pET22b plasmid inducible with IPTG.

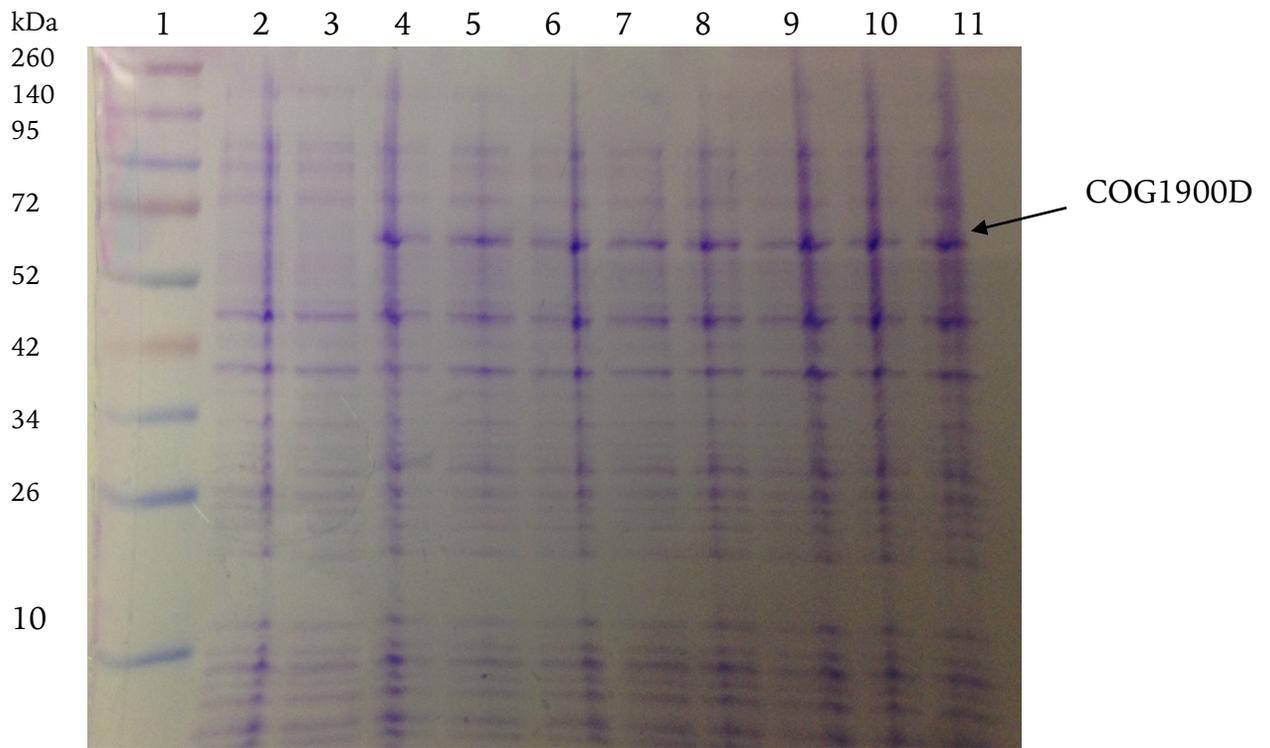


Figure 3.5 Induction of Cells with COG1900D Expression Plasmid

COG1900D is only expressed when IPTG is added to growth media. A value of (-) indicates no iron was added to the growth media, while (+) indicates iron was added.

All lanes are whole cell lysate. Lanes are as follows,

- 1- Standard ladder
- 2- 0 mM IPTG -
- 3- 0 mM IPTG +
- 4- 0.25 mM IPTG -
- 5- 0.25 mM IPTG +
- 6- 0.5 mM IPTG -
- 7- 0.5 mM IPTG +
- 8- 0.75 mM IPTG -
- 9- 0.75 mM IPTG +
- 10- 1 mM IPTG -
- 11- 1 m

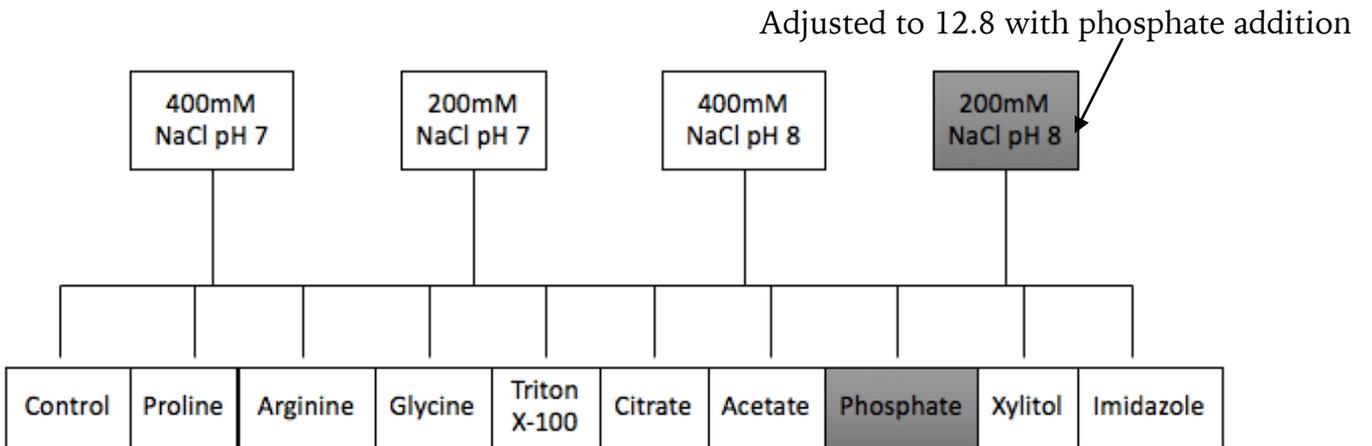


Figure 3.6 Lysis Conditions Screened

Forty lysis conditions tested yielding only one successful test. All pH values were prior to adding the stabilizing agent in the bottom row. The only successful conditions have a grayed background. The phosphate added was potassium phosphate tribasic, making a final pH of 12.8 without adjustments at 25 °C.

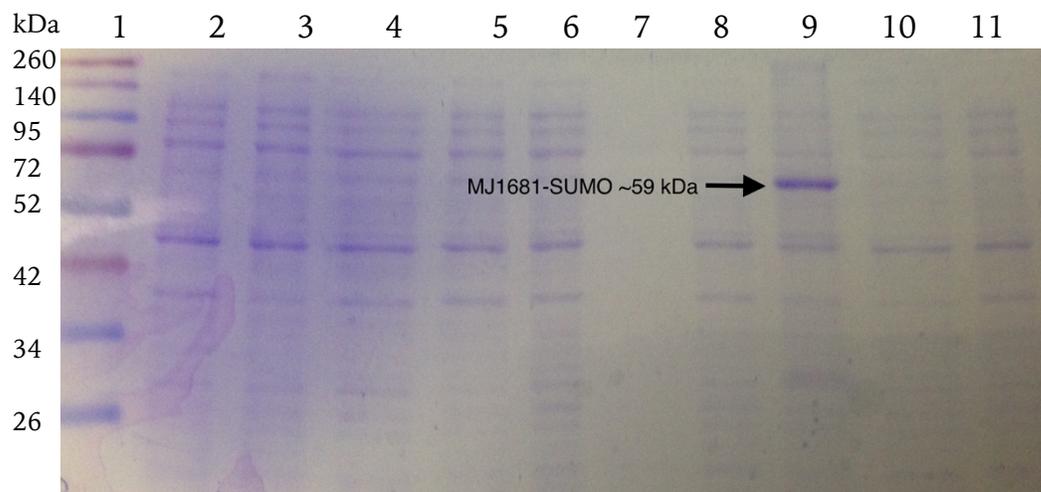


Figure 3.7 Gel of Lysis Additives Screen

Lane additives from left to right are as follows:

- 1 - Ladder
- 2 - Control
- 3 - Proline
- 4 - Arginine
- 5 - Glycine
- 6 - Triton X-100
- 7 - Citrate
- 8 - Acetate
- 9 - Phosphate
- 10 - Xylitol
- 11 - Imidazole

MJ1681 appears in the phosphate lane between ladder positions of 52kDa and 72kDa.

pH	% PO ₄ ³⁻
7	0.00%
8	0.00%
9	0.02%
10	0.21%
11	2.09%
12	17.61%
12.5	40.34%
12.7	57.43%
13	68.13%
13.3	81.01%

Table 3.2 Shift of Deprotonated Phosphate Species by pH

Using the Henderson Hasselbalch equation ($\text{pH} = \text{pK}_a + \log(\text{Base}/\text{Acid})$), the amount of PO₄³⁻ in solution can be calculated. A pK_a value of 12.67 was used for the reaction of HPO₄²⁻ \rightleftharpoons PO₄³⁻.

As unusual as a pH 13 buffer is, it may be necessary in providing enough strong kosmotropic phosphate species to solubilize difficult proteins such as COG1900. Such a powerful stabilizing agent does generate other problems. Phosphate, likely because of kosmotropic properties, has been known to help solubilize membrane proteins. This could explain some of the contaminants seen when resuspending an insoluble pellet containing MJ1681 (data not shown).

When cells containing COG1900D are lysed in a buffer that COG1900 proteins are insoluble in and centrifuged, the pellet contains COG1900D as well as the cells' membrane proteins. Upon discarding the soluble fraction, and thereby removing the cells non-membrane proteins, the insoluble pellet can be resuspended in the high pH phosphate buffer that solubilizes COG1900D. Preparing COG1900D through resuspension has the benefit of eliminating most of the contaminant proteins in the cell as well as increasing the yield of COG1900D. Aerobic lysis will certainly remove the iron sulfur clusters in COG1900D and render the protein inactive.

The same high pH buffer that solubilizes MJ1681 (COG1900D) also solubilizes MA1821 (COG1900A). This buffer was used in an attempt to solubilize MA1822, the ferredoxin protein neighboring MA1821 on the genome, and was unable to produce any soluble protein; MA1822 can be solubilized through refolding, see chapter 4. This suggests a unique interaction between COG1900 and the completely deprotonated

phosphate species, and also alleviates some of the concern that such extreme conditions could be solubilizing through a general denaturing effect like chaotropic agents such as guanidinium hydrochloride. A residual concern is the possibility of unwanted amino acid side chain reactions on COG1900D due deprotonation of every side chain at pH 13.

COG1900D Expression Conditions in *E. coli*

While much more COG1900D is expressed at 37 °C than lower temperatures, it is apparent that many of the contaminants in the preparation were actually truncated pieces of MJ1681. *E. coli* cells that over express MJ1681-SUMO at 37 °C produce multiple proteins that are cleaved by ULP determined by unexpected band migrations after adding ULP a purified sample (**Fig 3.8**). ULP is an extremely specific protease⁹⁴ and is not expected to cleave any proteins without a SUMO tag. The specificity of ULP, multiple band migration on SDS PAGE gel after ULP is added to the sample, and the prominence of the contaminants suggests they are truncated MJ1681-SUMO.

This is likely either due to transcription errors in bacterial *E. coli* trying to transcribe an archaeal protein or unwanted protease activity in the cell during

expression. Overexpression of proteins at lower temperatures has been shown to help reduce transcription errors⁸² and indeed, expression of MJ1681 at 16 °C reduces the amount of contaminants produced compared to at 37 °C; the difference in purity produced through the two expression temperatures is easy to verify after a gel filtration column (Fig 3.9).

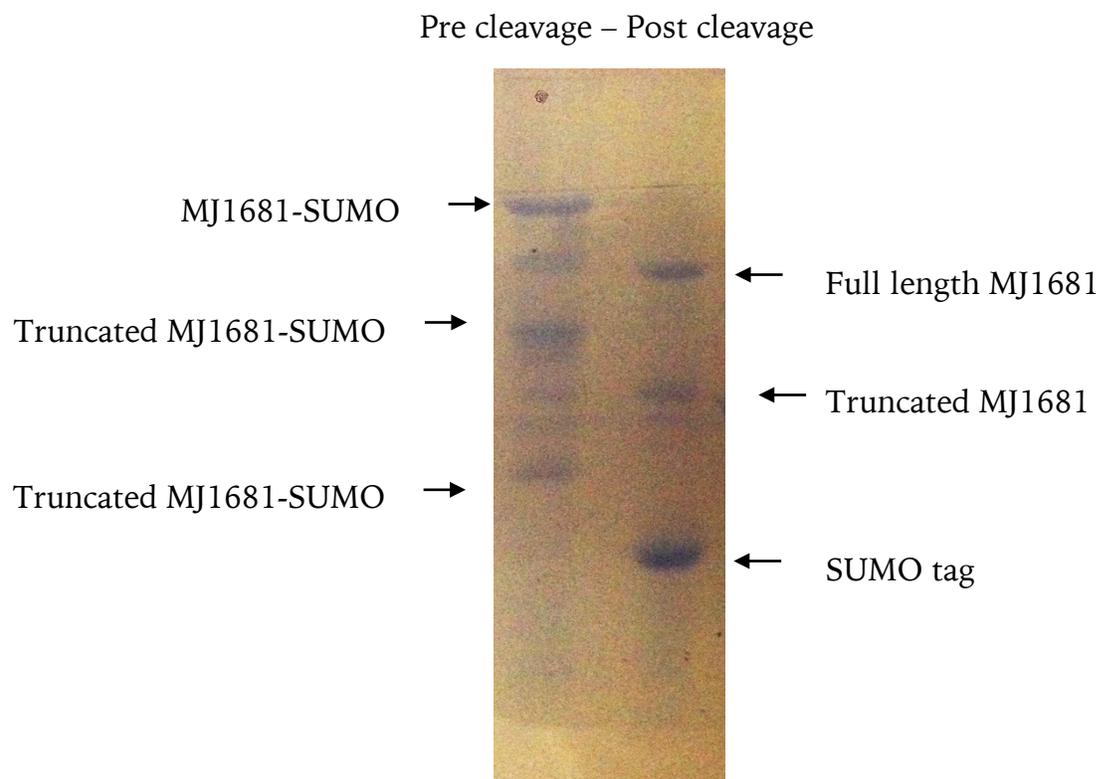


Figure 3.8 Evidence of MJ1681 Truncation During Expression

When expressed at 37 °C, MJ1681 appears truncated through migration of unexpected bands upon SUMO tag cleavage. Bands that were present before ULP addition are absent in the post cleaved sample of the same dilution. The smallest truncated product does not appear because it is small enough it likely ran off the gel.

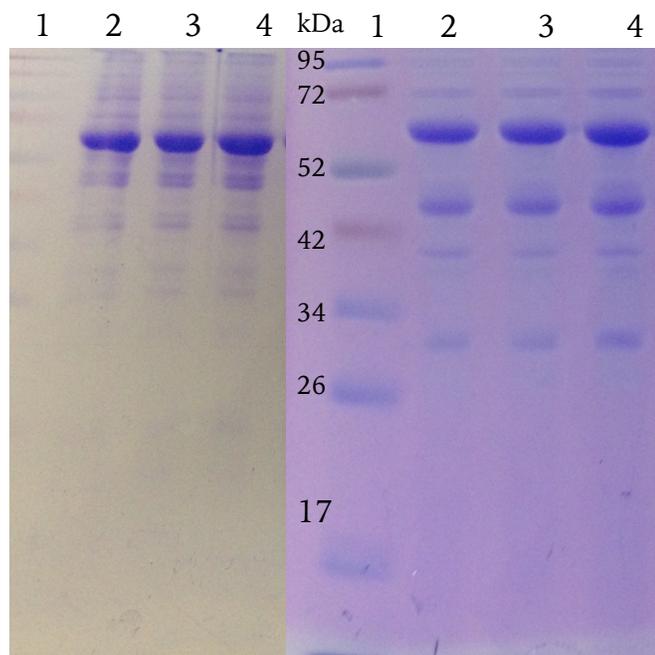


Figure 3.9 MJ1681 Expression Temperature Comparison

The gel on the left is MJ1681 expressed aerobically from *E. coli* at 37 °C for two hours while the figure on the right is the same strain expressed overnight at 16 °C. Both samples were elutions from gel filtration chromatography of comparable concentration and elution volume. Lanes for both gels are as follows,

- 1- Standard ladder
- 2- Elution fraction (85 mL)
- 3- Elution fraction (88 mL)
- 4- Elution fraction (91 mL)

Purification Hurdles of COG1900 Proteins

While finding conditions that permitted soluble COG1900 proteins was the biggest hurdle, the purification still presents many challenges with each step. The first obvious hurdle in this purification is the unusually high pH. Protein purification has been optimized for a more canonical pH range of 6-8. Stepping out of that range possibly jeopardizes conventional purification methods. Common affinity columns such as Ni²⁺ used with histidine tagged proteins become less efficient as more proteins will have an overall negative charge.

The answer to these problems is then to move the protein into a buffer that maintains stability and is within an ideal operating range for common purification methods, unless a new buffer that can solubilize the protein is pursued. While the phosphate and high pH buffer works in solubilizing all COG1900A and COG1900D proteins tested, the dialysis into new buffers is where the two protein groups begin to show different behaviors. While COG1900D appears to tolerate large shifts in pH and dialysis into morpholino buffers, COG1900A will either form a visible precipitate suspension or reveal precipitate upon centrifugation.

It is apparent therefore, that while COG1900 domains share a common favorability for deprotonated phosphate groups (possibly high pH) the protein families are unique enough to require individual purification protocols. The optimization of the

COG1900D purification required testing many different chromatography methods and the order of such methods does matter (**Fig 3.10**).

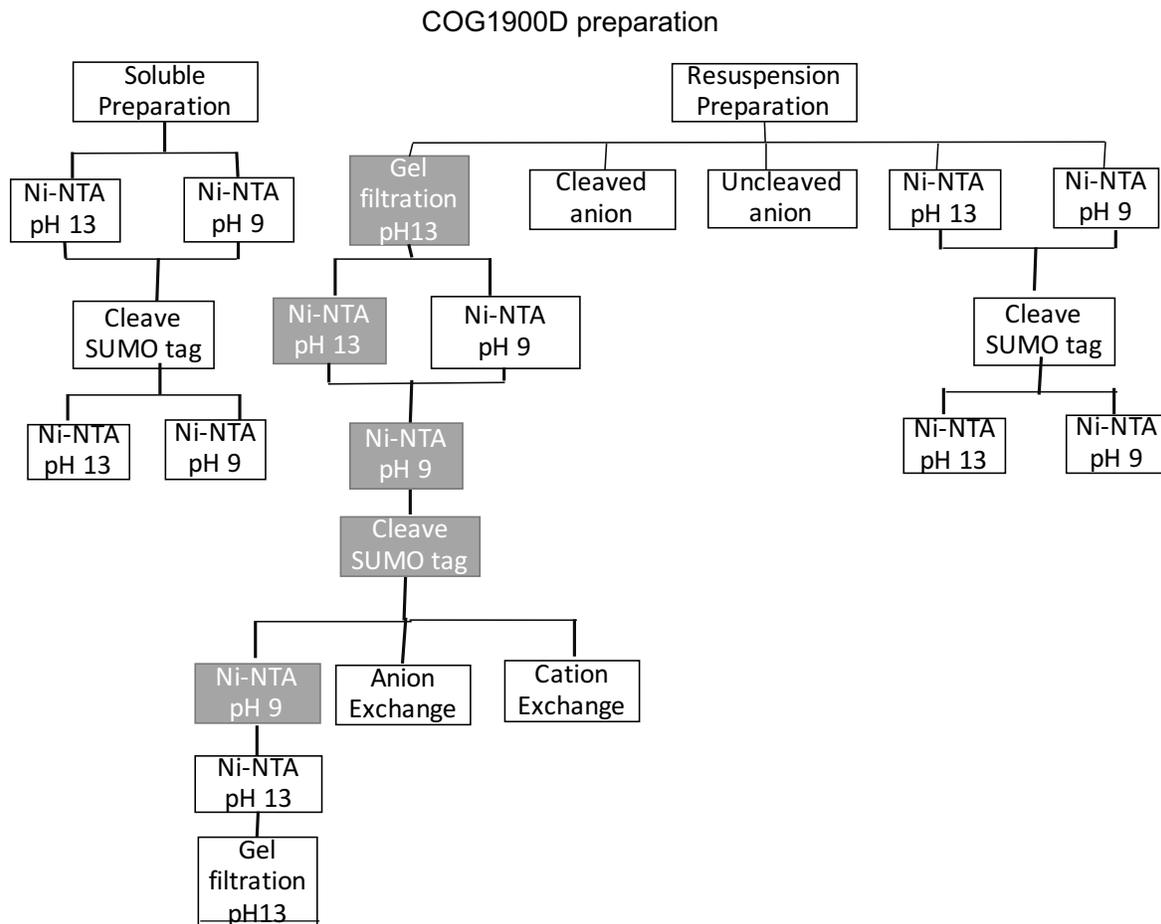


Figure 3.10 COG1900D Purification Optimization Flow Chart

Many different chromatography columns, conditions, and order of purification steps have been tested to maximize purity and yield of COG1900D. The most effective approach is highlighted by the grey boxes. The remaining purification strategies listed were attempted but were either unnecessary or proved worse than the highlighted method. A soluble preparation refers to lysing cells containing the target protein in a buffer that will produce target protein in the soluble fraction following lysis. A resuspension preparation refers to lysing the cells containing the target protein in a buffer that will not produce target protein in the soluble region, instead placing it in the insoluble pellet. The insoluble pellet is then “resuspended” in a suitable buffer that does keep the target protein in the soluble fraction.

COG1900D Lysis

Crystallization of either COG1900A or COG1900D will allow us to visualize a new protein structure never before analyzed. COG1900D has the iron sulfur cluster binding domain attached to the protein and could therefore be biochemically active upon purification while COG1900A would require the additional purification of a separate iron sulfur cluster binding protein. For these reasons the bulk of this thesis' work was done on the purification of MJ1681 (COG1900D), so that will be discussed at length before addressing the purification of COG1900A and the ferredoxin protein.

Lysis by sonication, instead of freeze thaw cycles, was determined to be the preferred method for MJ1681 (**Fig 3.11**). While the cells will be lysed given lysozyme and freeze thaw cycles, MJ1681 does not appear any less soluble or stable after sonication which is much more thorough and faster. For the results of this thesis, all cells were all lysed by sonication over ice for four minutes cycling on and off every thirty seconds (eight minutes total time, four minutes 'on'). Clearing the lysate can be done by centrifugation at 30,000g for 30 minutes.

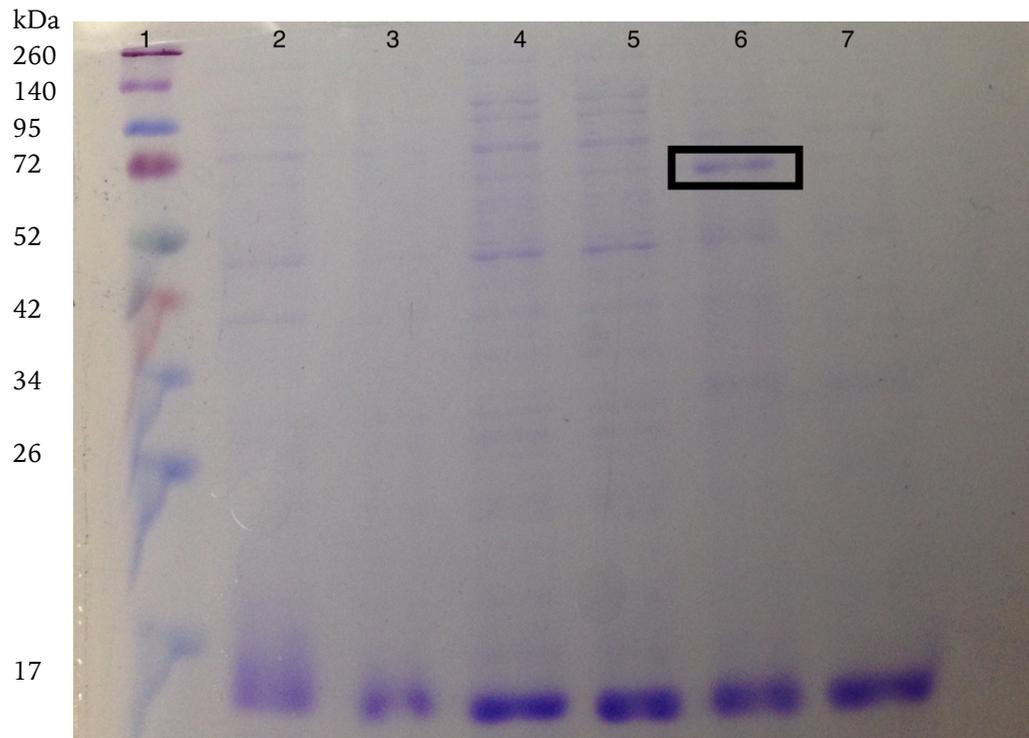


Figure 3.11 COG1900D Lysis Method

MJ1681 is boxed in the image for easy identification. Addition of ULP confirms the identity as MJ1681 by cleavage of the SUMO tag. Lanes are as follows:

- 1- Standard ladder
- 2- Arginine additive lysed by sonication
- 3- Arginine additive lysed by freeze thaw cycles
- 4- Triton X-100 additive lysed by sonication
- 5- Triton X-100 additive lysed by freeze thaw cycles
- 6- Phosphate additive lysed by sonication
- 7- Phosphate additive lysed by freeze thaw cycles

Initial experiments were conducted by lysing the cells in the phosphate high pH buffer, immediately solubilizing MJ1681. Since then, a resuspension technique has been adopted that still permits solubility while removing many, if not all, soluble *E. coli* contaminants. Instead of lysing the MJ1681 containing cells in the high pH phosphate buffer, a more canonical buffer such as 20mM Tris-HCl pH 8 with 100mM sodium chloride produces completely insoluble MJ1681 that will form a pellet when the lysate is cleared.

Discarding the soluble fraction, the insoluble pellet can be resuspended in the high pH phosphate buffer and centrifuged again. Upon the second centrifugation step, MJ1681 was in the soluble fraction (**Fig 3.12**). The first step therefore removed all *E. coli* contaminants that are soluble in Tris-HCl, NaCl buffer. This not only greatly improves the yield of MJ1681 but also the purity right from the lysis step.

Without anaerobic conditions, the iron sulfur cluster domain of MJ1681 does not appear to remain intact during preparation; although the pellet retains a unique red hue (**Fig 3.13**), there is no distinct shoulder in a visible spectrum at 450nm⁹⁵, which would be indicative of iron sulfur clusters (data not shown). This red hue is one way to track the protein during preparation. After resuspending the insoluble pellet in the high pH phosphate buffer and a second centrifugation step, the soluble fraction will

now contain the red hue while the new insoluble pellet should be white/pale (**Fig 3.14**).

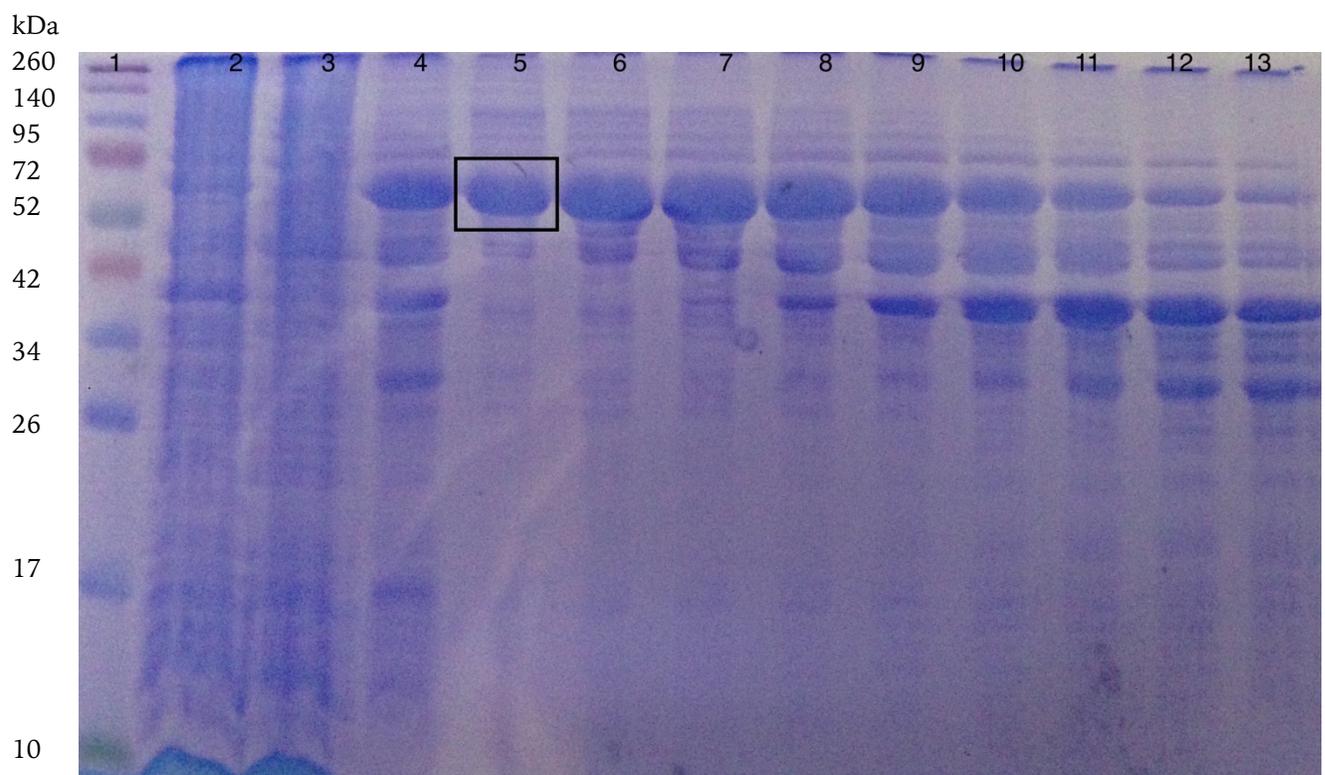


Figure 3.12 COG1900D Insoluble Pellet Resuspension

MJ1681 is boxed for identification. Lanes are as follows:

- 1- Standard ladder
- 2- Whole cell lysate
- 3- Soluble lysis
- 4- Resuspended insoluble pellet
- 5-13 – SEC fractions

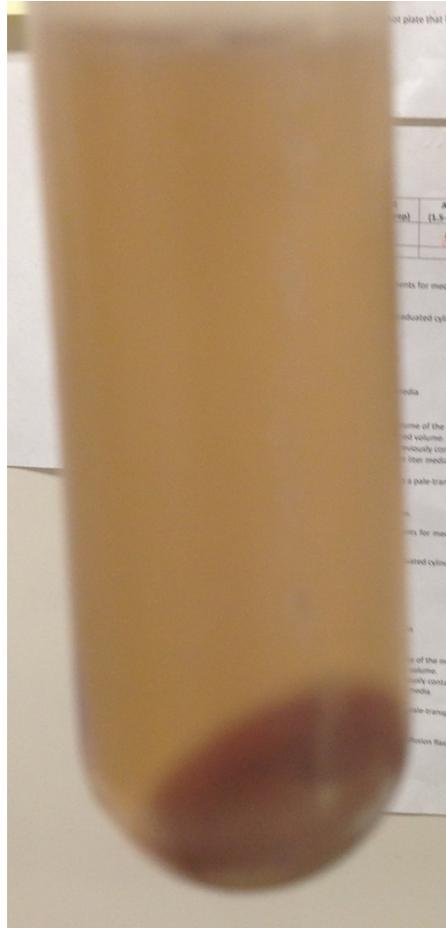


Figure 3.13 Insoluble Pellet Containing MJ1681

After lysis in a buffer not capable of solubilizing MJ1681 and subsequent centrifugation, the insoluble pellet has a red hue. This is one way to track the protein during preparation.



Figure 3.14 Insoluble Pellet After MJ1681 is Resuspended

After resuspending an MJ1681 pellet into the soluble fraction, the pellet formed after centrifugation completely loses the red hue (see figure 3.12) seen in a pellet containing MJ1681.

COG1900D Gel Filtration Chromatography

Decanting the supernatant will remove the soluble *E. coli* proteins, some insoluble proteins and cell debris will have been eliminated in the second white pellet, and the remaining solution will be mostly MJ1681 (**Fig 3.14**). Phosphate's distinct ability to stabilize membrane proteins coupled with the high pH is enough to solubilize other *E. coli* contaminants along with MJ1681 which still need to be removed.

Following resuspension of the insoluble pellet, the best purification step to follow with is gel filtration chromatography. The purification size gel filtration column used in all preparations is a hand packed 230 mL sephadex s-200. The column is prepared the night before by allowing 10 column volumes of running buffer (10 mM potassium phosphate, pH 13, 20 mM NaCl) to pass over the column by gravity.

MJ1681-SUMO will elute over a wide range of 80 mL to 120 mL (**Fig 3.15 and Fig 3.16**). The A_{280} chromatogram produced is one large elution peak from 80 mL to 160 mL (data not shown). The wide range of MJ1681-SUMO elution is likely due to MJ1681 oligomerization into a variety of different sized assemblies of full length and truncated protein. The early elutions still separate much of the full length MJ1681-SUMO from the contaminants and can produce fairly pure samples from this chromatography step alone. Only the purest samples are pooled (i.e. lane 7 and earlier in figure 3.12).

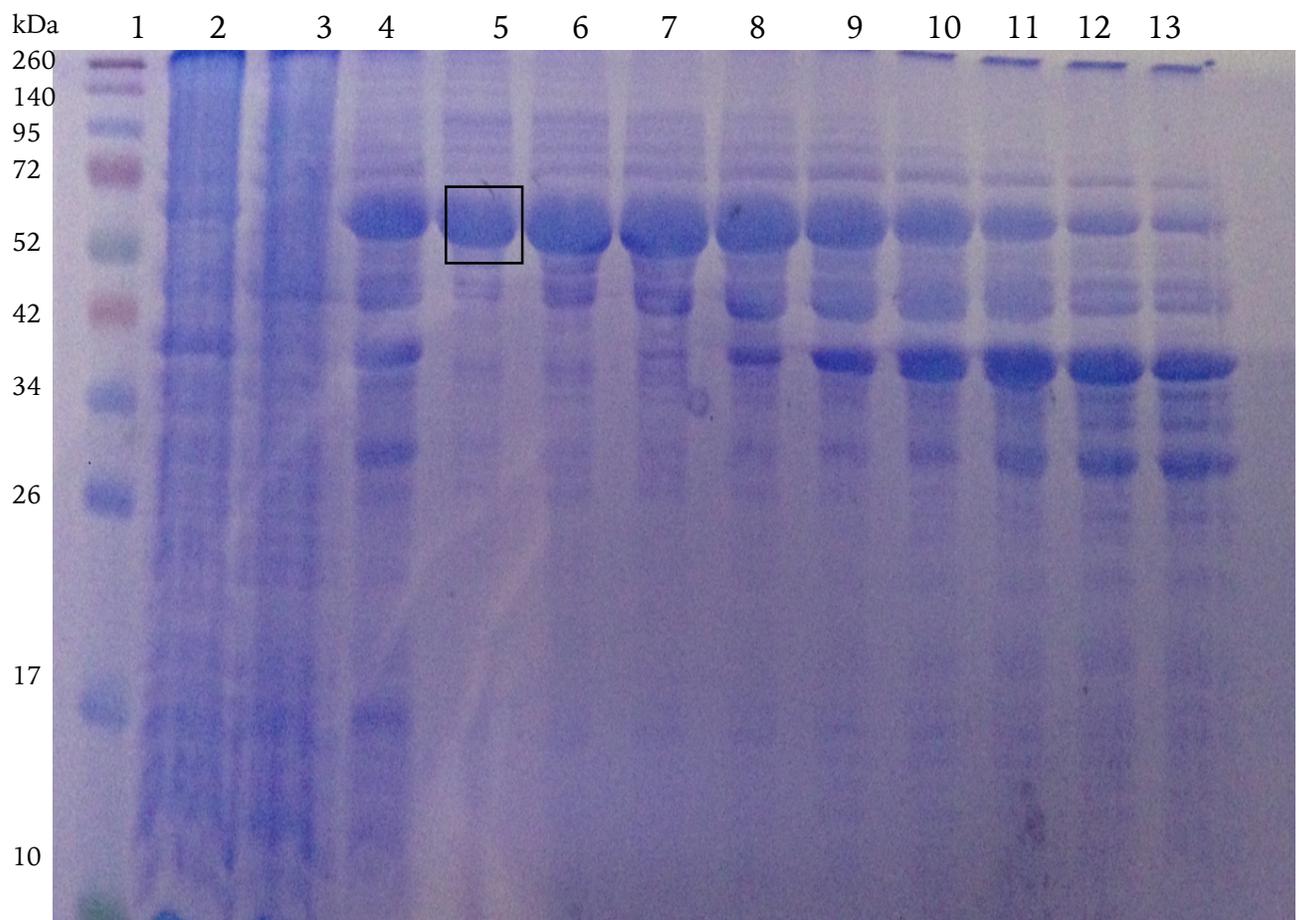


Figure 3.15 MJ1681-SUMO SEC 1

MJ1681-SUMO is boxed for easy identification

Lanes

1 - Ladder

2 - Dirty lysis (not centrifuged)

3 - Clean lysis (centrifuged)

4 - Resuspended insoluble pellet

5-13 - SEC elutions starting at 94 mL and going to 116 mL

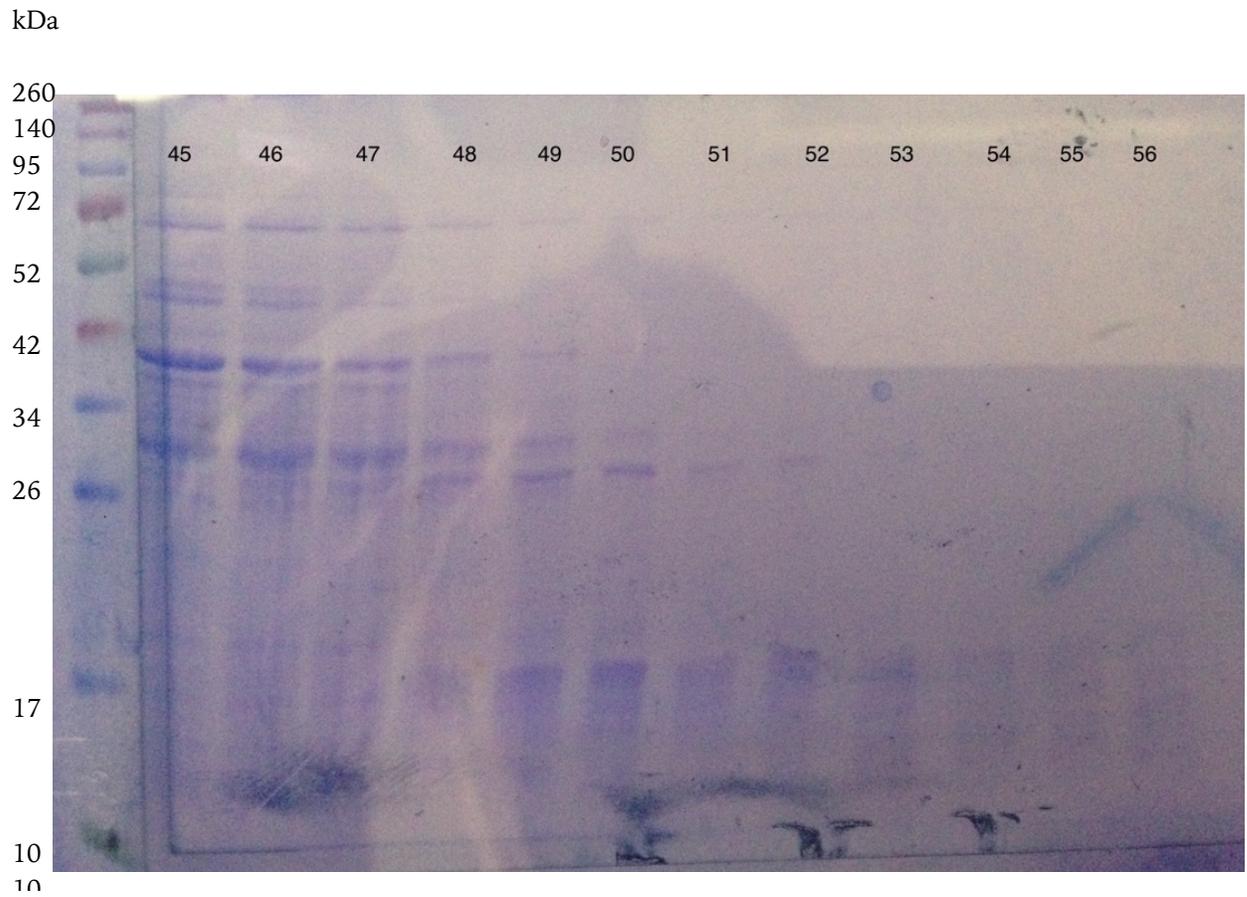


Figure 3.16 MJ1681-SUMO SEC 2

Lanes are elution volume of 120 mL to 148 mL (continuation of figure 3.12)

COG1900D Affinity Chromatography

The next purification step when using a poly histidine tagged protein is Ni-NTA affinity chromatography. As mentioned before, a Ni²⁺ affinity column has complications at such an extreme pH. Most, if not all, proteins at pH 13 will have a net negative charge. This appears to permit unwanted affinity from proteins in general to the Ni²⁺ resin. There is also very poor recovery of protein under such conditions, with about 1% recovery on average. The low yield can be addressed through an overnight incubation on the Ni-NTA resin. 1 mL of Ni-NTA resin is equilibrated with 10 column volumes of running buffer (10 mM potassium phosphate, pH 13, 20 mM NaCl) before being loaded with the sample.

We hypothesize that there is some equilibrium between quaternary states of MJ1681 with only one favorable state for binding the Ni-NTA resin. As the binding occurs the bound protein should be removed from the solutions initial equilibrium, and by Le Chatelier's principle drive the equilibrium towards the favorable binding state. Instead of doing an overnight incubation, repeated Ni-NTA columns can be performed in the same day by letting the flow through from each pass over the Ni-NTA beads incubate for 1 hour before passing the flow through over the beads again (**Fig 3.17**). This can be repeated until the A₂₈₀ of the flow through does not change between passes.

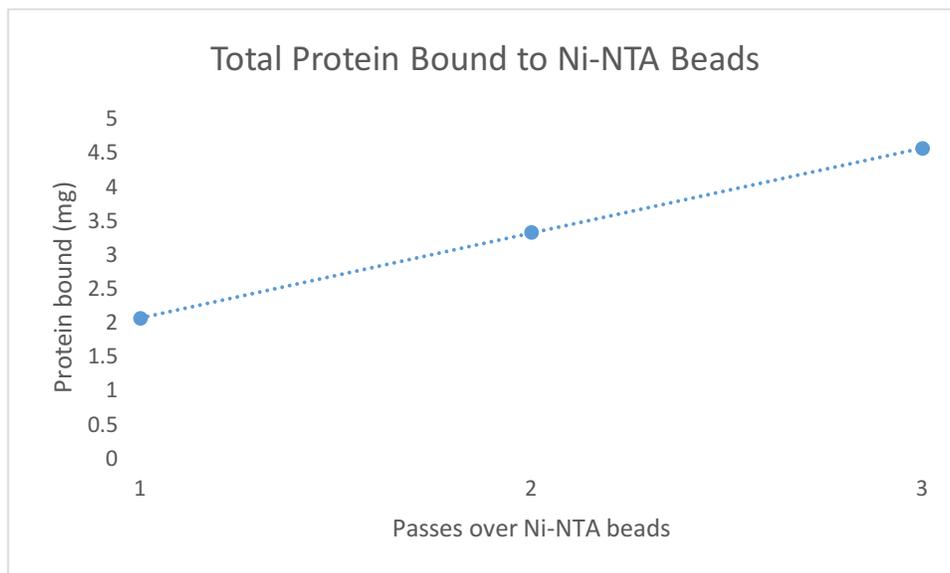


Figure 3.17 MJ1681-SUMO Binding Ni-NTA with Multiple Passes

A sample of MJ1681-SUMO-His₆ was passed over 1 mL of Ni-NTA beads and the flow through set aside and incubated 1 hour before passing it over the column again. The amount of bound protein between each pass was calculated as the difference in the A_{280} of the flow through after each pass over the column. Each pass will continue to bind more protein until the beads are either saturated or all target protein is bound.

Once the column is saturated it is washed with 5 column volumes of running buffer followed by elution with 5 column volumes of elution buffer (10 mM potassium phosphate, pH 13, 20 mM NaCl, 1 M imidazole) (**Fig 3.18**). The elutions are pooled and buffer exchange is done using spin concentrators with a molecular weight cutoff of 30 kDa (MJ1681 is 42 kDa) into Ni-NTA pH 8 buffer (10 mM potassium phosphate, pH 8, 20 mM NaCl).

Once the imidazole is removed by buffer exchange, the sample is passed over 1 mL of Ni-NTA beads that have been equilibrated with 10 column volumes of Ni-NTA pH 8 buffer. The column is washed with 5 column volumes of Ni-NTA pH 8 buffer and the protein is eluted with 5 column volumes of Ni-NTA pH 8 elution buffer (10 mM potassium phosphate, pH 8, 20 mM NaCl, 1 M imidazole). The elutions are pooled and 0.03 mg of ULP is added to the sample to remove the SUMO tag from MJ1681 before being dialyzed overnight against 1 L of Ni-NTA pH 8 buffer. Cleavage completion can be assessed using SDS PAGE gel (**Fig 3.19**).

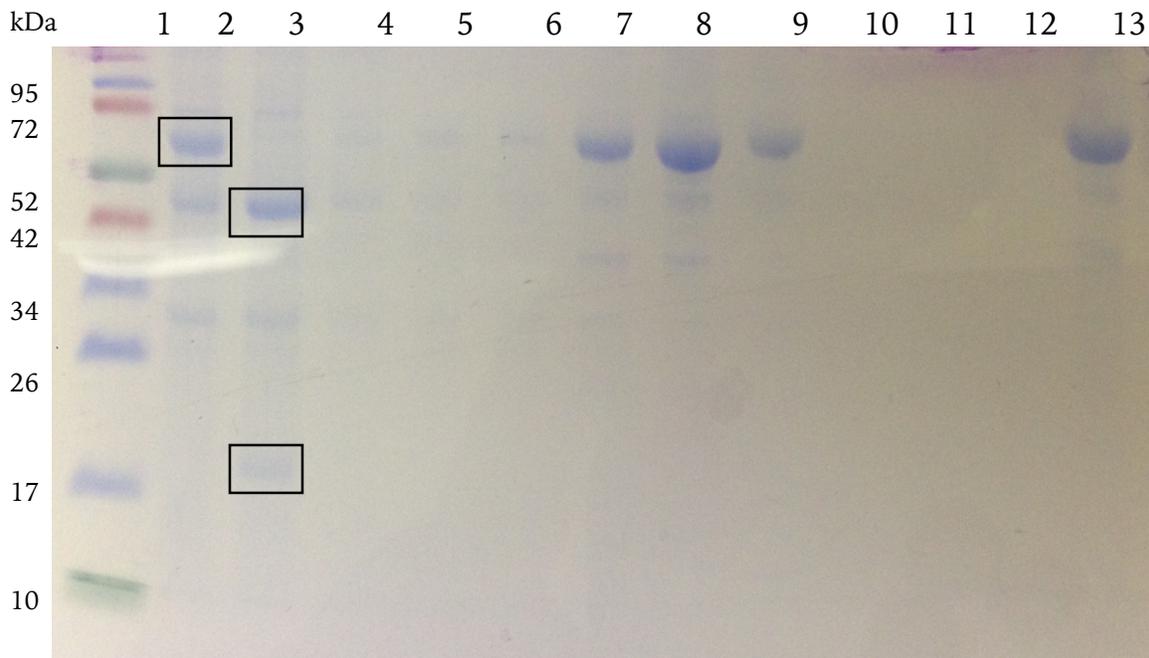


Figure 3.18 MJ1681 Ni-NTA Chromatography and ULP Cleavage of SUMO Tag

Lane 2 has MJ1681-SUMO boxed and lane 3 has full length MJ1681 without SUMO and cleaved SUMO (lowest) boxed for identification. Note, the gel is not in preparation order as the cleavage occurs after the Ni-NTA columns as in the text. Lanes are:

- 1- Ladder
- 2- Pre SUMO tag cleavage
- 3- Post SUMO tag cleavage
- 4- Ni-NTA pH 13 wash fraction 1
- 5- Ni-NTA pH 13 wash fraction 2
- 6- Ni-NTA pH 13 wash fraction 3
- 7- Ni-NTA pH 13 elution fraction 1
- 8- Ni-NTA pH 13 elution fraction 2
- 9- Ni-NTA pH 13 elution fraction 3
- 10- Ni-NTA pH 9 wash fraction 1
- 11- Ni-NTA pH 9 wash fraction 2
- 12- Ni-NTA pH 9 wash fraction 3
- 13- Ni-NTA pH 9 elution fraction 1

Once cleavage has been completed, the sample is passed over 1 mL of Ni-NTA beads that have been equilibrated with 10 column volumes of Ni-NTA pH 8 buffer. The flow through is collected and purity estimated using SDS PAGE gel (Fig 3.19). 1 L of cells produces about 3 mg of 80% pure MJ1681.

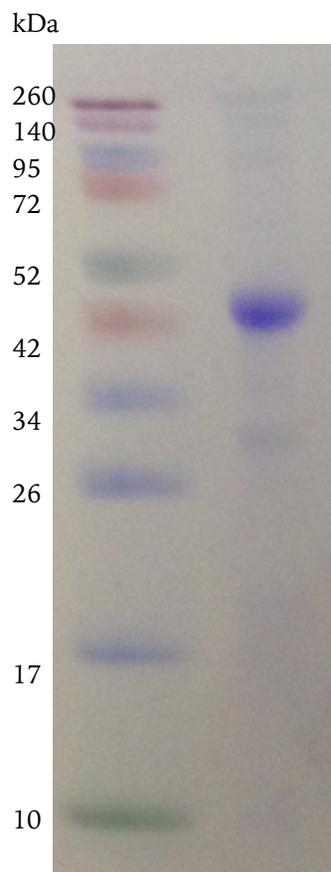


Figure 3.19 SDS PAGE Gel of Purified MJ1681

Maximum purity of MJ1681 achieved is roughly 80% based on SDS PAGE gel estimation. -

MJ1681 Quaternary and Secondary Structure Analysis

Purified MJ1681 at pH 13 forms a hexamer as determined by analytical SEC using a GE superdex 200 column (**Fig 3.20**) (see appendix B for a chromatogram of standard proteins with the GE superdex 200 column). All analytical gel filtration columns were done in the presence of DTT. If a hexamer is the appropriate quaternary structure, then the peak at the exclusion limit is presumed to be aggregation. The preparation leading to the analytical SEC data presented had Triton X-100 in the solution. While working under the reported critical micelle limit, it was determined that micelles indeed were forming, verified by fractions that absorbed at 280 nm⁹⁶ yet had no detectable protein on SDS PAGE gel (data not shown). Because Triton X-100 absorbs at 280 nm, a peak presented at the size of Triton X-100 micelles (~120 kDa).

The presence of a dominant MJ1681 species smaller than the exclusion limit of the column alleviates some concern that the pH 13 buffer is denaturing. Further evidence that the protein is not being denatured at pH 13 is the presence of secondary structure elements observed using circular dichroism (CD) spectroscopy. At pH 13 alpha helices can be detected (**Fig 3.21**), which is unique compared to MJ1681 in 6M guanidinium HCl (**Fig 3.22**) which shows no secondary structure (see appendix B for an example of CD spectroscopy interpretation and the spectrum of lysozyme for

comparison). All CD spectroscopy was done using an 'Aviv circular dichroism spectrometer Model 202'.

The remaining secondary structure cannot be determined using CD spectroscopy at this time because the pH 13 buffer absorbs strongly at 200 nm for reasons unknown. Buffer containing only phosphate and sodium ions at pH 13 absorbs very strongly at 200 nm, chloride in the protein buffer is therefore not the problem, rather it is either the phosphate or the sodium at pH 13 causing the problem (data not shown).

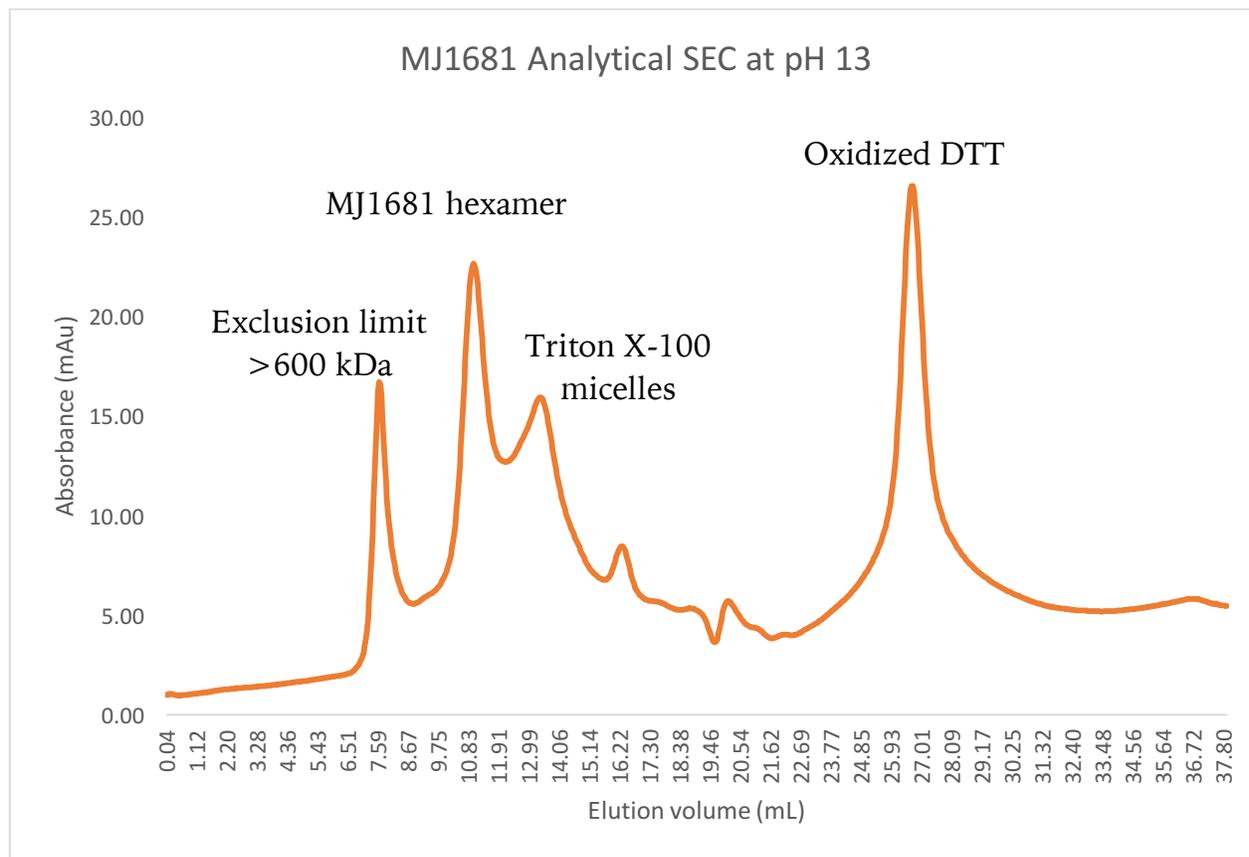


Figure 3.20 Analytical SEC of MJ1681

A GE superdex 200 column was used to analyze purified MJ1681. The exclusion limit of the column is 600 kDa and the first peak is at the exclusion limit. It is presumed to be aggregate unless the native quaternary structure of MJ1681 is determined to be over 600 kDa. The second peak contains MJ1681 and is a hexamer (see appendix B for standard and size calculation). The neighboring peak is Triton X-100 micelles which are ~120 kDa. The smaller peaks are not MJ1681, they were too dilute to observe on SDS PAGE gel, however any remaining SUMO tag in the sample would run around 16 mL and might explain that peak. The final peak is oxidized DTT.

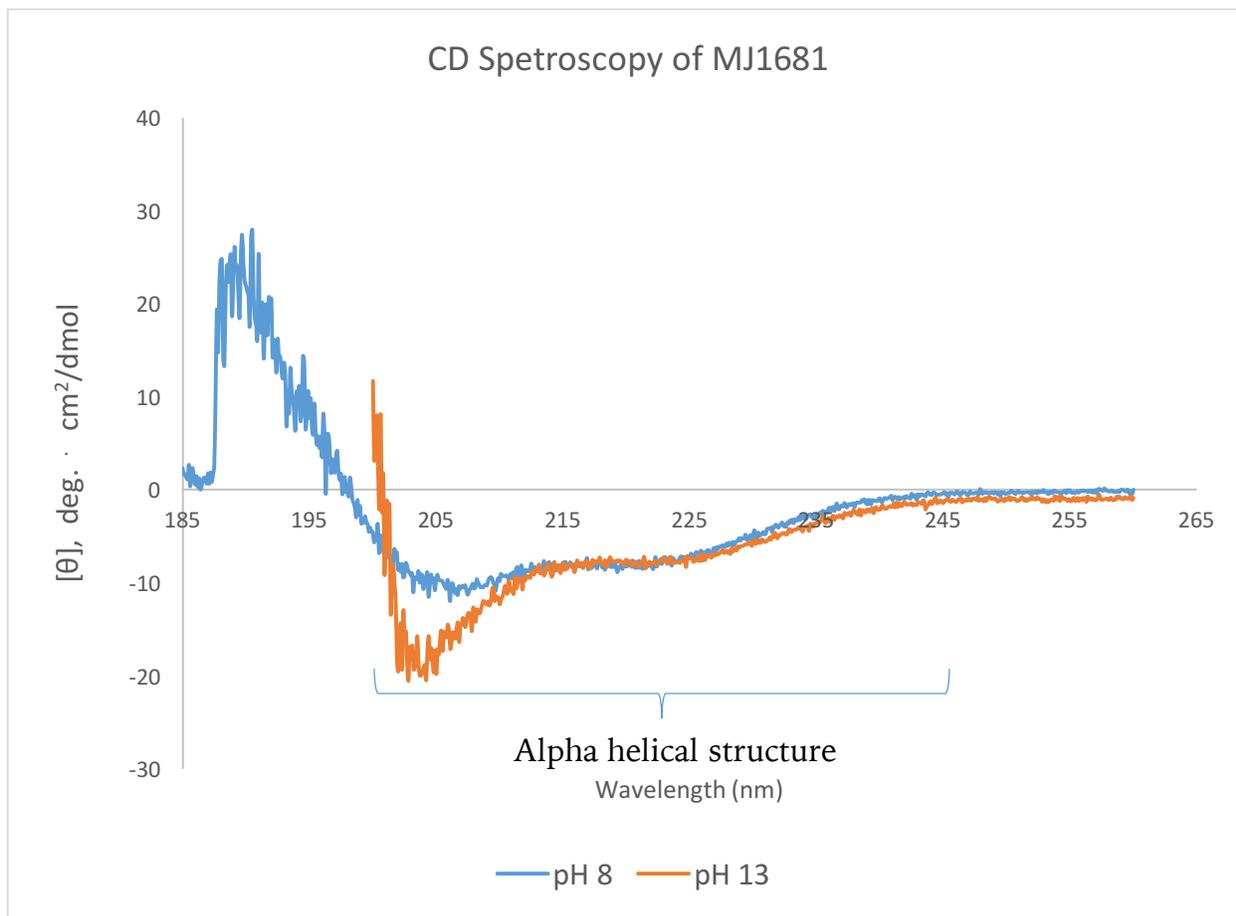


Figure 3.21 CD Spectroscopy of MJ1681

Purified MJ1681 CD Spectroscopy results at pH 8 and pH 13. The pH 13 sample cannot have the complete spectrum analyzed at this time due to the limitations of the pH 13 buffer. The buffer absorbs very strongly at 200 nm for unknown reasons. This strong absorption renders any data collected below 200 nm unreliable. MJ1681 at both pH 8 and pH 13 has alpha helical structures. Interestingly, there is a much stronger alpha helical signal at pH 13 as seen in the spectrum have a much stronger signal near 200 nm. A figure for interpreting CD spectroscopy and a sample of lysozyme as a positive control for comparison can be found in appendix B.

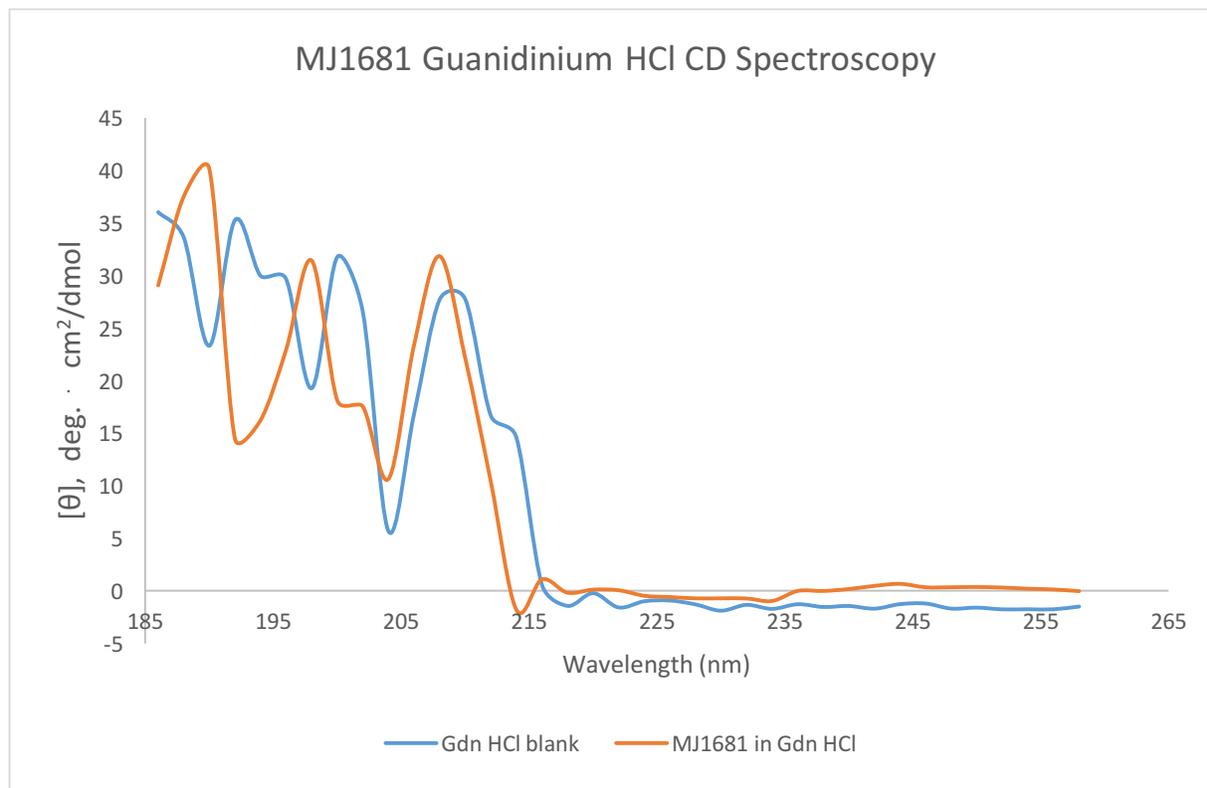


Figure 3.22 MJ1681 CD Spectrum in Guanidinium

MJ1681 in 6 M guanidinium HCl has no detectable secondary structure. When compared to the CD spectrum of MJ1681 in pH 13 buffer, this demonstrates that the pH 13 buffer is not denaturing, at least not to the extent of a chaotropic agent like guanidinium HCl. The peak at 215 nm is due to guanidinium HCl.

Differences in the secondary and quaternary structure of MJ1681 are observed between pH 8 and pH 13, neither of which is physiological for *M. jannaschii* (pH 6⁹⁷). MJ1681 appears to have more alpha helical content at pH 13 than at pH 8 based on CD analysis. A shift in quaternary structure is also observed through analytical SEC at pH 8 (Fig 3.22). A dominant hexameric peak is no longer observed, instead MJ1681 is dispersed between the aggregate peak and the monomer relatively equally, the aggregate peak being the most represented of the species.

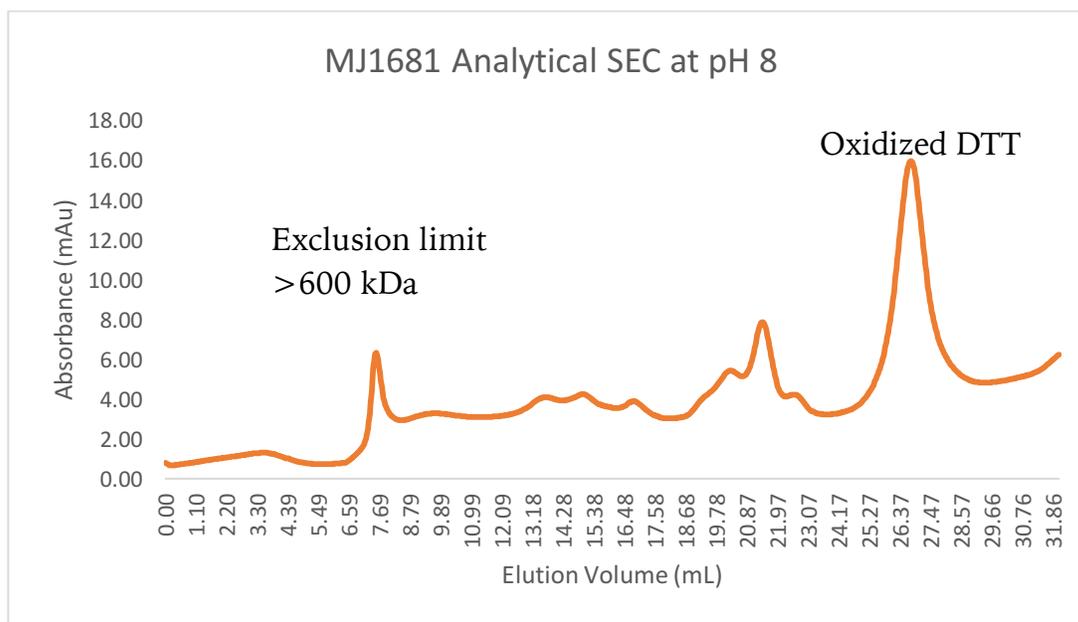


Figure 3.23 MJ1681 SEC at pH 8

The largest peak is oxidized DTT. The first peak is aggregated MJ1681 and the remainder of MJ1681 is dispersed between the aggregate peak and 15 mL which would be near monomer size. The remaining peaks are contaminants from a less purified preparation of MJ1681 than the pH 13 SEC column.

Discussion

The purification of COG1900D can be achieved using a non-conventional lysis buffer. COG1900D can be moved from the abnormally high pH of 13 used in lysis, to more canonical pH ranges including pH 6 and pH 8. COG1900D has not been found to be soluble in any other condition than the high pH phosphate buffer

Given the current purity achieved, while not ideal, is enough to start crystallization trials with. It is clear through CD spectroscopy that some structure is retained even at pH 13. It is not clear if the native fold is retained at pH 13; however, crystal tray screens may help determine a more suitable buffer for the protein should pH 13 not be conducive for the native fold. Reconstitution of the iron sulfur clusters would allow testing of the proteins activity. If the protein is active and synthesizes CoM, the function of COG1900D and the retention of the native fold under extreme purification conditions will be confirmed.

Ultracentrifugation studies can be done at various pH values to accurately determine the shift in oligomeric state. This would help elucidate the odd behavior observed on gel filtration chromatography at different pH values. Other species of phosphate, such as pyrophosphate or polyphosphate, may also be able to stabilize COG1900D proteins without the need for buffers with such extreme pH. This might also help achieve purity above the current 80%.

Chapter 4: Purification of COG1900A, MA1822 and COG2122 Proteins

COG1900A compared to COG1900D purification

COG1900A is responsible for synthesizing homocysteine and is not strictly conserved to methanogens. Another small ferredoxin binding protein is essential for function and is either directly attached to the protein or coded separately as a second protein.

The first major difference in COG1900A purification compared COG1900D is that insoluble pellets containing COG1900A cannot be resuspended (moved into the soluble fraction) with a suitable high pH phosphate buffer. This initial step in purification of COG1900D is essential for recovering large quantities of protein and isolates the target protein from most soluble *E. coli* contaminants. When attempted with a COG1900A containing insoluble pellet, the pellet forms a clear, gelatinous coating that is impervious to attempted solubilization in an appropriate buffer through stir bar mixing, manual mixing, vortexing, and even attempts at physical shearing (data not shown). This may be attributed to the lack of a SUMO tag, the presence of a CBS

domain, the lack of a ferredoxin domain, or simply the differences due to primary structure such as pI (see table 3.1).

Expression of MA1821

All MA1821 was generated aerobically in pBR004 or pBR039 (for coexpression of MA1821/22) *E. coli* cells using a T7 IPTG inducible expression pET22b(+) plasmid (Fig 4.1). In pBR004 strains, MA1821 contains a poly histidine tag. In pBR039 MA1822 contains a poly histidine tag while MA1821 does not. Induction was done with 1 mM IPTG and expression of MA1821 was overnight at 16 °C.

COG1900A Lysis

Without the option to resuspend the insoluble pellet containing COG1900A that results from lysis at physiological pH, the lysis must then be carried out in 30 mL of the COG1900 solubilization buffer (10 mM potassium phosphate, pH 13, 20 mM NaCl). The 40 conditions used to screen COG1900D solubility (see figure 3.7) were not retested with COG1900A, only the confirmation that the same buffer that solubilizes COG1900D also solubilizes COG1900A.

Lysis was done by sonicating for four minutes, with thirty second intervals of on/off cycling (eight minutes' total time) over ice. The lysate was cleared by

centrifugation at 20,000 g for 15 minutes. The soluble fraction was collected and insoluble pellet was discarded.

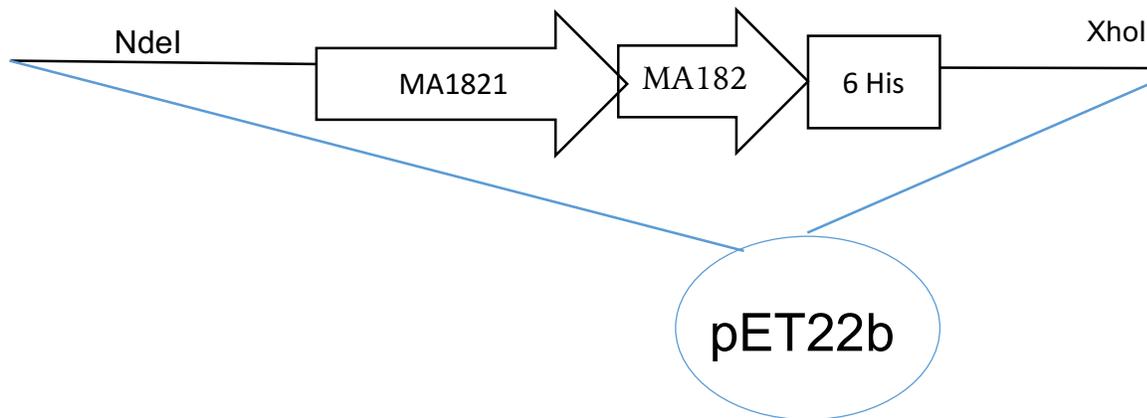


Figure 4.1 MA1821/22 Expression Plasmid for MA1821/22 Generation

MA1821 was expressed aerobically in *E. coli* Rosetta2(DE3)pLysS cells with chloramphenicol resistance. The plasmid was inducible using IPTG and was ampicillin resistant. The gene for MA1822 was included on pBR039 and not included on pBR004, in pBR004 the poly histidine tag is attached to MA1821. pBR039 (with MA1822) was used in the purification of MA1822 discussed later.

COG1900A Affinity Chromatography

Solubilized MA1821 was incubated on 1 mL of Ni-NTA with shaking for 30 minutes at 4 °C. Flow through over a column was collected and set aside. The Ni-NTA resin was washed with ten column volumes of COG1900A lysis buffer and eluted in ten column volumes of COG1900 elution buffer (10 mM potassium phosphate, pH 9.0, 20 mM NaCl, 1 M imidazole). A substantial portion of the protein is lost in the flow through with only about 1% of MA1821 binding the Ni-NTA (**Fig 4.2**). Passing the flow through over the Ni-NTA resin multiple times, as was done with COG1900D, was never tested with COG1900A.

The sample was then loaded to a gel filtration column for further separation. A GE superdex 200 column was used at 0.5 mL/min. The elution profile of MA1821 varies greatly depending on the pH (**Fig 4.3**), with an elution peak above the exclusion limit of the column (>600 kDa). Running the sample at pH 10 seems to produce the smallest peak at the exclusion limit (**Fig 4.4**). The elution profile of MA1821 may be dependent of the presence of MA1822. 1 L of cells produces about 1.5 mg of 75% pure MA1821.

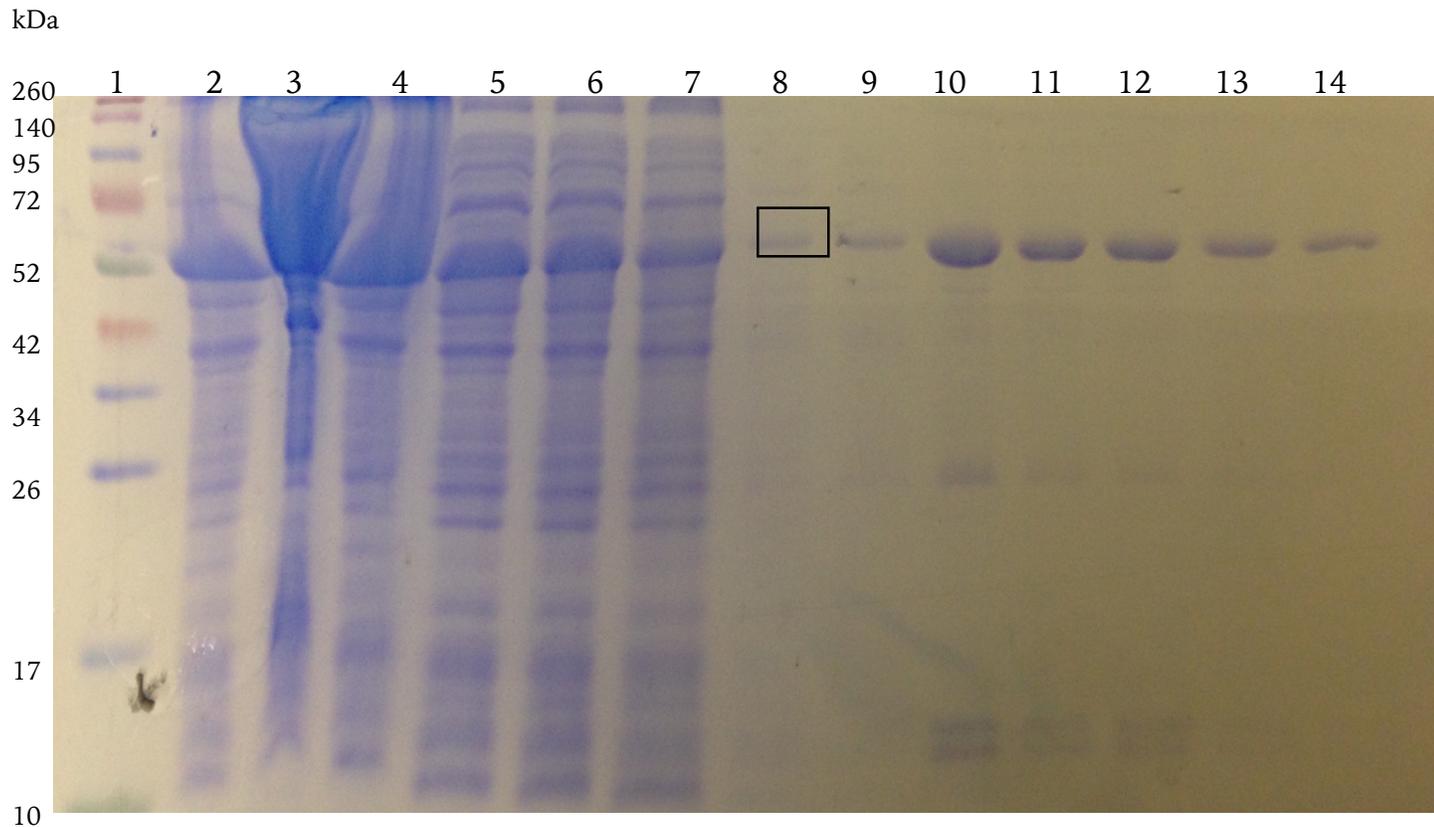


Figure 4.2 MA1821 Ni-NTA SDS-PAGE Gel

MA1821 is boxed for identification. Lanes are as follows,

- 1- Ladder
- 2- Whole cell lysate
- 3- Insoluble pellet
- 4- Dirty lysis (not centrifuged)
- 5- Cleared lysis (centrifuged)
- 6- Flow through (~30 mL)
- 7- Wash 1
- 8- Wash 3
- 9- Wash 5
- 10-Elution 1 (each fraction ~1.5 mL)
- 11-Elution 2
- 12-Elution 3
- 13-Elution 4
- 14-Elution 5

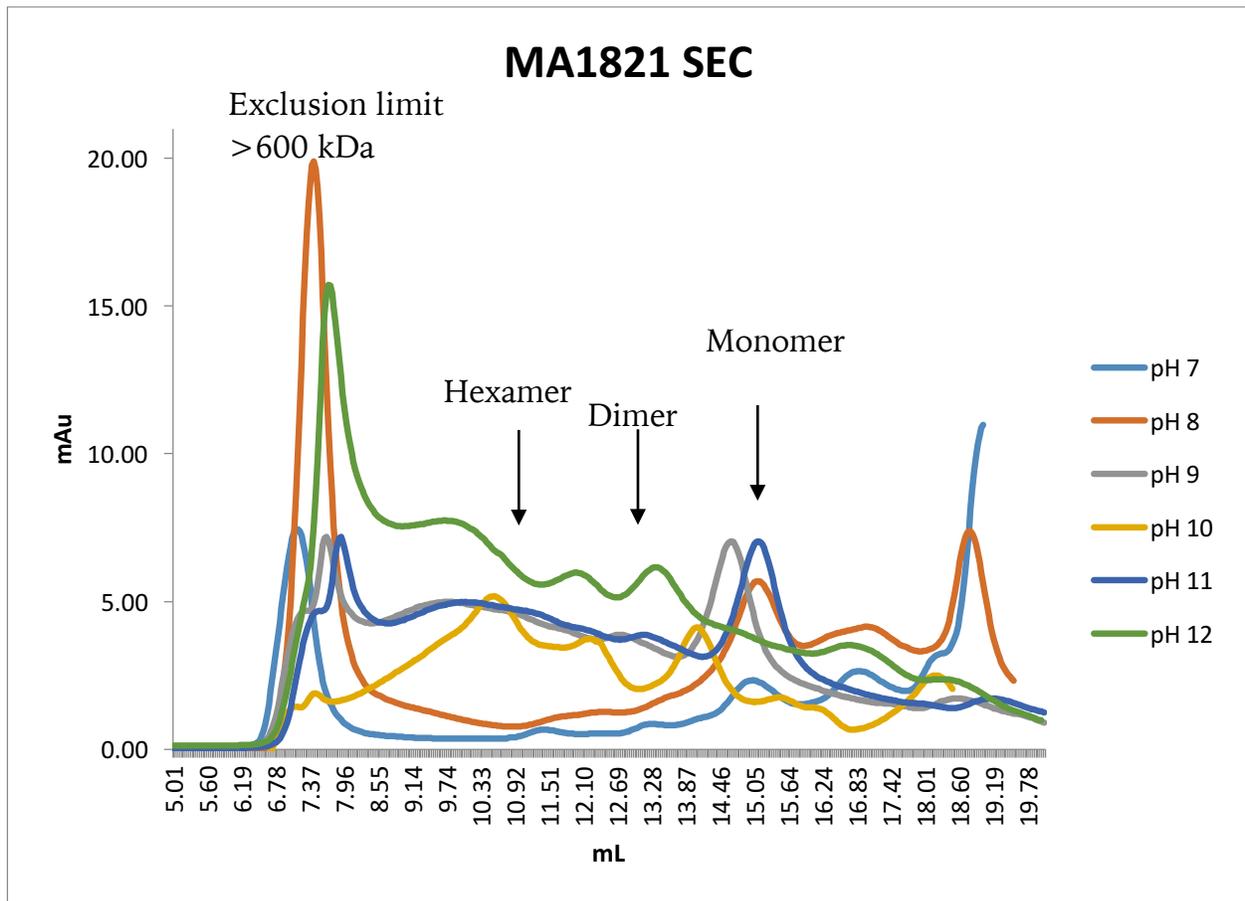


Figure 4.3 MA1821 Gel Filtration Chromatography at Various pH Values

The elution profile of MA1821 over a GE superdex 200 column. The exclusion limit is 600 kDa and any complex larger than this will elute around 7 mL as seen by the most prominent peaks. A monomeric form of MA1821 should elute at 15 mL, a dimer around 13 mL, and a hexamer near 11 mL.

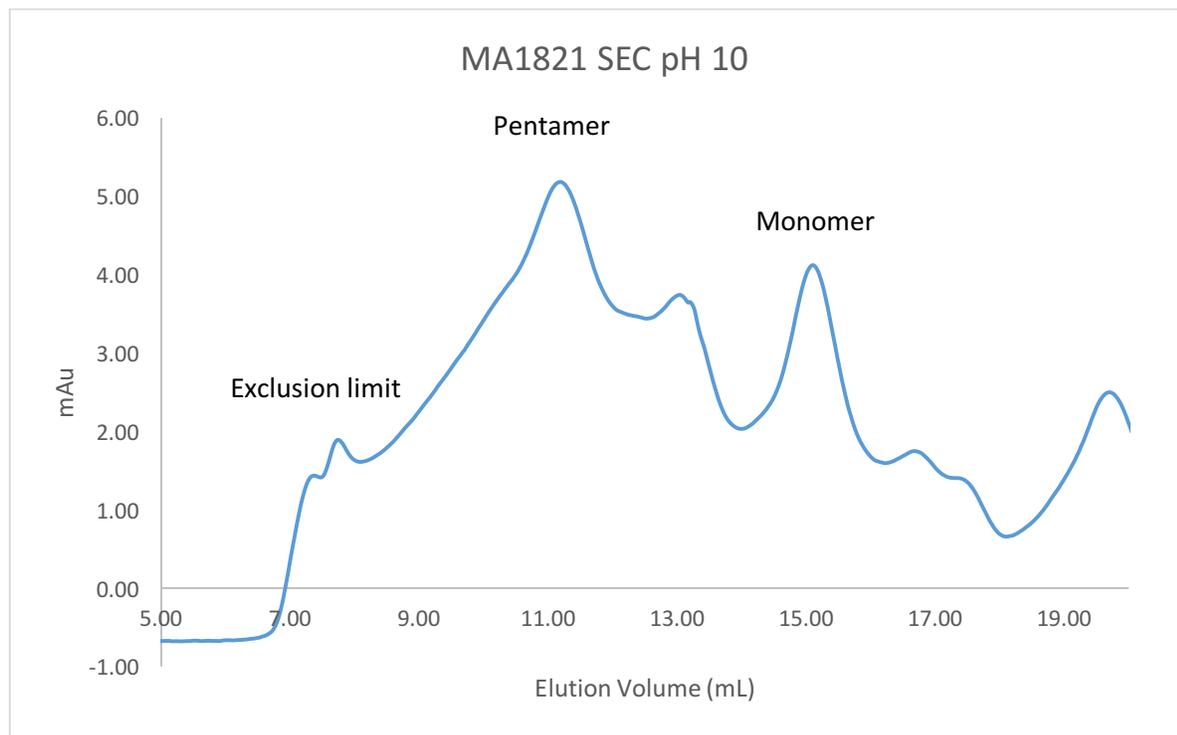


Figure 4.4 MA1821 Gel Filtration Chromatography pH 10

Running MA1821 over a gel filtration column at pH 10 produces the smallest exclusion limit peak (near 7 mL). It is the only condition discovered that produces the largest peak after the exclusion limit, in this case near 11 mL which would be a pentamer of MA1821.

Tlet_1363 COG1900A Fusion Protein Purification

Tlet_1363 is a COG1900A fusion protein. Unlike MA1821, Tlet_1363 has the necessary ferredoxin domain (MA1822 in *M. acetivorans*) attached at the C-terminus. Tlet_1363 was generated through aerobic expression in *E. coli* that were transformed with the pBR104⁶⁶ plasmid (Fig 4.5). Tlet_1363 has an N-terminal SUMO tag attached which has an N-terminal poly histidine tag.

The lysis and affinity chromatography purification steps of Tlet_1363 are exactly the same as that of MA1821 above. 10 mL of cells containing Tlet_1363 at $OD_{600} = 0.6$ were lysed in COG1900 lysis buffer (10 mM potassium phosphate, pH 13, 20 mM NaCl) to test solubility. Tlet_1363 is soluble in the COG1900 lysis buffer and can be purified through Ni-NTA affinity with the given construct (Fig 4.6).

No further purification tests were done with Tlet_1363. Tlet_1363 and MJ1681 were both good representatives to try and purify from COG1900A and COG1900D families respectively given they both have the necessary ferredoxin domain already attached. MJ1681 was studied much more thoroughly because initial success and optimization of growth was done with MJ1681 and its biochemistry has not been confirmed, while COG1900A is already known to synthesize homocysteine.

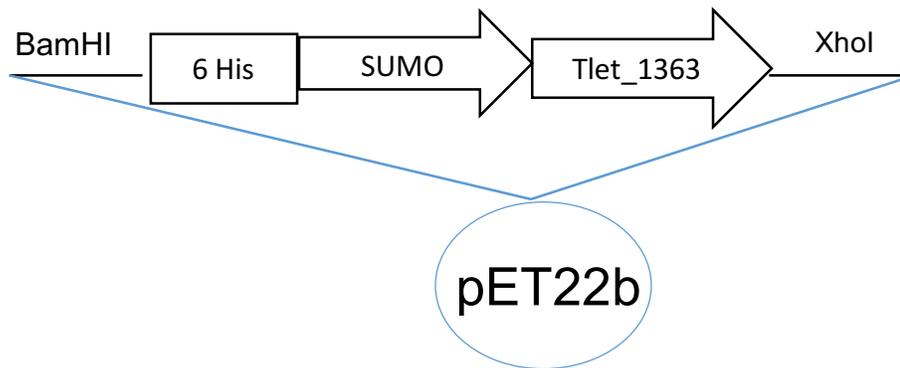


Figure 4.5 Tlet_1363 Expression Plasmid

Tlet_1363 was expressed aerobically in *E. coli* Rosetta2(DE3)pLysS cells with chloramphenicol resistance. The plasmid was inducible using IPTG and was kanamycin resistant. Tlet_1363 had a C-terminal SUMO tag and the SUMO tag had a C-terminal polyhistidine tag.

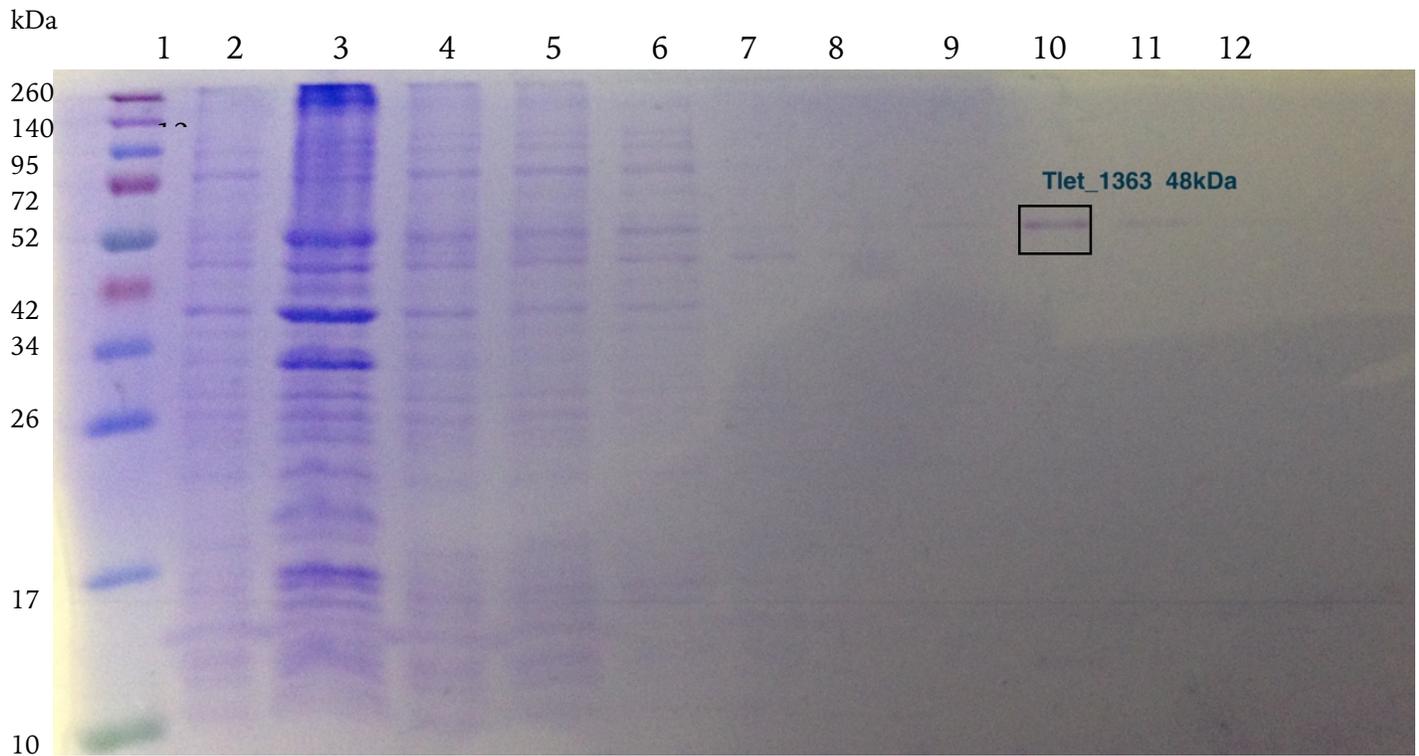


Figure 4.6 Tlet_1363 Ni-NTA Purification SDS PAGE Gel

Tlet_1363 is boxed for identification purposes. Lanes are as follows,

- 1- Ladder
- 2- Whole cell lysis
- 3- Insoluble pellet
- 4- Dirty lysis
- 5- Cleared lysis
- 6- Wash 1
- 7- Wash 3
- 8- Wash 5
- 9- Elution 1
- 10-Elution 2
- 11-Elution 3
- 12-Elution 4
- 13-Elution 5

MA1822; Ferredoxin Domain Protein

MA1822 is 128-amino-acids in length. It is a small ferredoxin protein that neighbors COG1900A proteins, MA1821 in *M. acetivorans*⁶⁶ (Table 4.2). It is essential in organism's dependent on COG1900A for homocysteine synthesis. Considering the lack of a redox domain on MA1821, and the need for redox chemistry in the conversion of an aldehyde to a thiol, it is likely that MA1822 fills this role. Like MA1821, MA1822 appears completely insoluble in conventional lysis buffers when purified aerobically from *E. coli* and is even insoluble in the COG1900 lysis buffer (10 mM potassium phosphate, pH 13, 20 mM NaCl).

Amino Acids	130
Molecular Weight	14.46 kDa
Theoretical pI	5.20
Extinction Coefficient (ϵ_{molar})	9105 M ⁻¹ cm ⁻¹
Absorbance 0.1% ($\epsilon_{0.1\%}$)	0.629

Table 4.2 MA1822 Profile

Parameters were calculated using an ExPasy prediction tool⁸⁸.

Purification of MA1822

Cleaning Inclusion Bodies

All MA1822 with a His₆ tag was purified from Carbenicillin resistant Rossetta2(DE3)pLysS cells containing an ampicillin resistant and IPTG inducible pET22b vector with MA1822 inserted via NdeI and XhoI restriction sites courtesy of Benjamin Rauch (**Fig 4.1**). Successful purification of MA1822 was achieved through a two-step denaturing and refolding procedure⁷². Cells were grown to OD₆₀₀ = 0.6-0.7 and expression was induced with 1 mM IPTG. Expression was carried out for 4 hours at 37 °C. The cells were harvested at 5,000 g at 4 °C for 15 minutes. The cell pellet was washed twice with 150 mL of washing buffer 1 (20 mM Tris, pH 8.0), centrifuged 30,000 g 4 °C for 30 minutes, and then lysed in 80 mL of buffer C (50 mM Tris, pH 8.0) by sonication over ice.

The lysate was cleared by centrifugation at 30,000 g for 30 minutes at 4 °C. The supernatant was discarded and the pellets cleaned twice with 150 mL of washing buffer 2 (50 mM Tris, pH 8.0, 50 mM NaCl, 2% Triton X-100 (v/v), 1.5 mM β-mercaptoethanol 1.6 M urea) by resuspension through stirring for 20 minutes and centrifuging the solution at 30,000 g for 30 minutes at 4 °C. Removal of the detergent was accomplished by cleaning the pellet two more times with wash buffer 1. The cleaned inclusion bodies were stored at -20 °C.

Two-Step Denaturing and Refolding

The cleaned inclusion bodies were resuspended in 5 mL of extraction buffer 1 (50 mM Tris, pH 8.0, 50 mM NaCl, 10 mM β -mercaptoethanol, 7 M Gdn HCl) followed by centrifugation at 30,000 g for 30 minutes at 4 °C. The supernatant, now containing MA1822, was quickly added to 200 mL of dilution buffer (50 mM Tris, pH 8.0, 1 mM EDTA, 50 mM NaCl, 10 mM β -mercaptoethanol) to rapidly precipitate the protein and centrifuged at 30,000 g for 30 minutes at 4 °C. The pellet was collected and resuspended in 5 mL of extraction buffer 2 (50 mM Tris, pH 8.0, 50 mM NaCl, 10 mM β -mercaptoethanol, 8 M urea).

The supernatant was collected and pellet discarded. The solution was then added drop wise to 400 mL of refolding buffer (20 mM Tris, pH 8.0, 1 mM EDTA, 1 mM GSH, 0.1 mM GSSG) while being stirred slowly with a magnetic stir bar. The solution was moved to 4 °C and allowed to stir slowly for two days. The solution was loaded to 1 mL of Sepharose-Q anion exchange resin at a rate of 0.5 mL/min using a peristaltic pump. The protein was eluted using 20 mL of buffer D (20 mM Tris, pH 8.0, 1 mM EDTA, 1 M NaCl) over a linear NaCl gradient of 0-1 M. The fractions containing MA1822 were confirmed using SDS-PAGE gel analysis (**Fig 4.7**). Oligomeric state analysis was done using a GE superdex 200 gel filtration column at a flow rate of 0.5 mL/min (**Fig 4.8**). 1 L of cells produces about 1 mg of 95%+ pure MA1822.

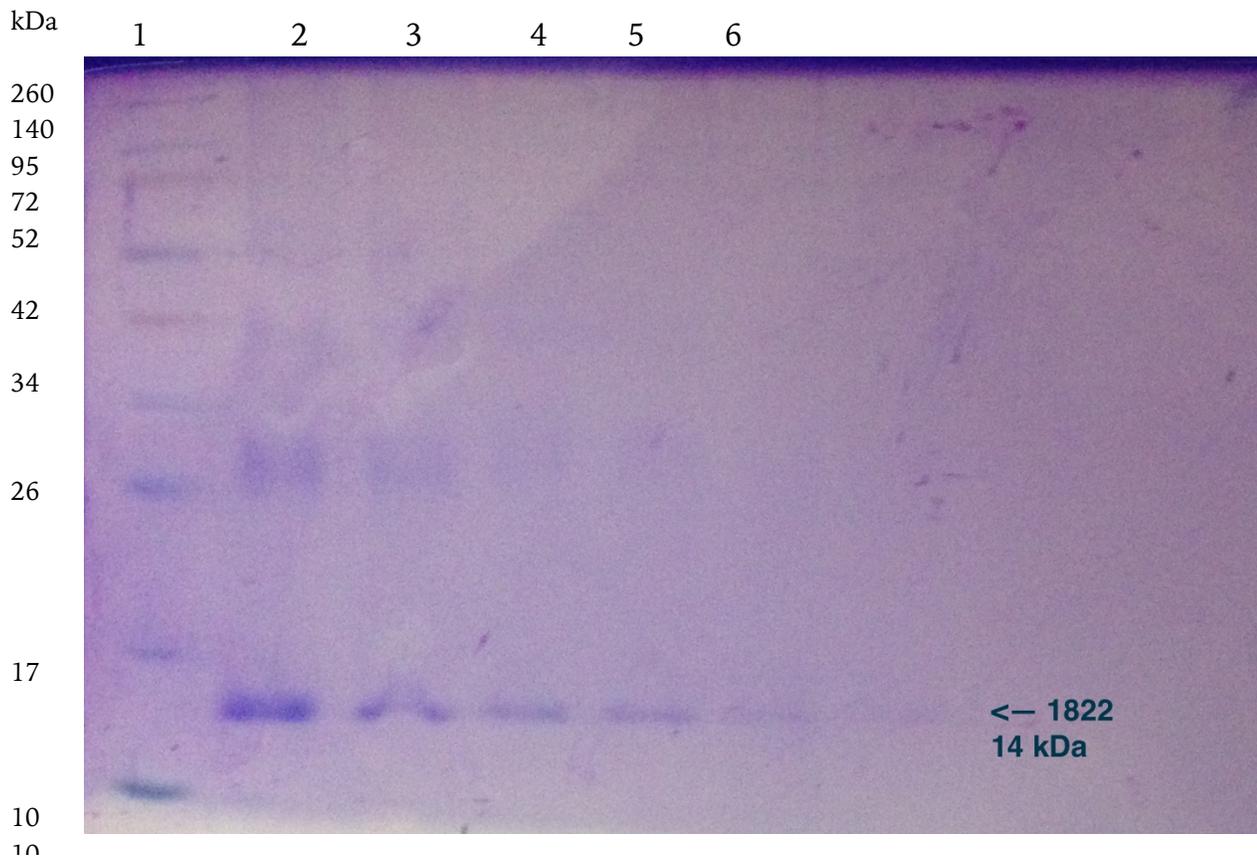


Figure 4.7 MA1822 SDS-PAGE Gel

Successful purification of MA1822 verified by SDS-PAGE gel. The labeled protein band eluted between the 10 kDa and 17 kDa ladder markers which is expected for MA1822.

Lanes are as follows,

1- Standard ladder

2-6 elution fractions from anion exchange chromatography

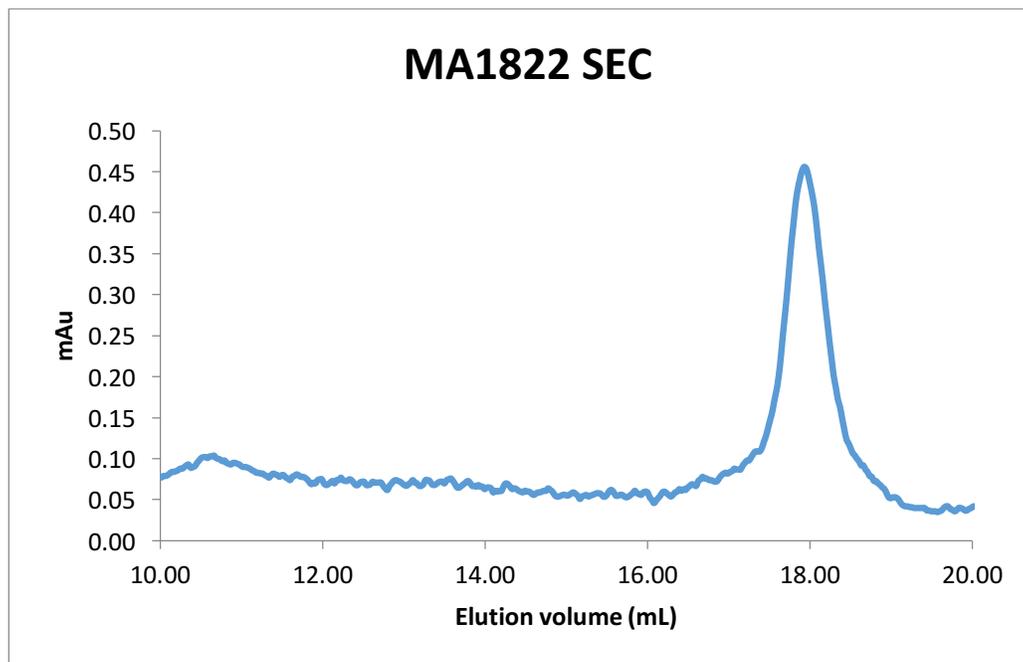


Figure 4.8 MA1822 Gel Filtration Chromatogram

The quality of MA1822 that purified was tested using a GE superdex 200 column run at 0.5 mL/min. The elution profile presented a monodispersed peak within the range for monomeric MA1822.

Use of MA1822 in Future Studies

MA1822 runs solely as a monomer on size exclusion chromatography. As a small ferredoxin protein suspected in assisting COG1900A proteins, a monomer is likely the native oligomeric state. Given the harsh conditions of denaturing and refolding the protein, the iron sulfur clusters are lost in the preparation. It is possible to reconstitute apoenzymes with iron sulfur clusters even in aerobic conditions using

β -mercaptoethanol^{34,35}. Reconstitution would be essential for *in vitro* studies with MA1821.

Should iron sulfur clusters be reconstituted to refolded MA1822, any kinetic studies would still require additional materials beyond MA1821. For example, the reducing agent needed to regenerate the ferredoxin is still unknown. The sulfur donor, while possibly MA1715, is also currently unknown. Without these essential components no kinetic studies can be accomplished *in vitro*. The purification of MA1822 as outlined above will provide a foundation for such future studies.

COG2122 Purification

The 253 amino acid MA1715 protein is the COG2122 homologue in *M. acetivorans* (Table 4.3). All MA1715 with a His₆ tag was purified from carbenicillin resistant Rossetta2(DE3)pLysS cells containing an ampicillin resistant and IPTG inducible pET22b vector with MA1715 inserted via NdeI and XhoI restriction sites courtesy of Benjamin Rauch (Fig 4.9). Successful purification of MA1715 was achieved through the same two-step denaturing and refolding procedure⁷² used to purify MA1822. Cells were grown to OD₆₀₀ = 0.6-0.7 and expression was induced with 1 mM IPTG. Expression was carried out for 4 hours at 37 °C. The cells were harvested at 5,000 g at 4 °C for 15 minutes. The cell pellet was washed twice with 150 mL of

washing buffer 1 (20 mM Tris, pH 8.0), centrifuged 30,000 g 4 °C for 30 minutes, and then lysed in 80 mL of buffer C (50 mM Tris, pH 8.0) by sonication over ice.

Amino Acids	253
Molecular Weight	26.76 kDa
Theoretical pI	5.49
Extinction Coefficient (ϵ_{molar})	$16055 \text{ M}^{-1} \text{ cm}^{-1}$
Absorbance 0.1% ($\epsilon_{0.1\%}$)	0.600

Table 4.3 MA1715 Profile

The parameters were calculated using an ExPasy Prediction tool⁸⁸.

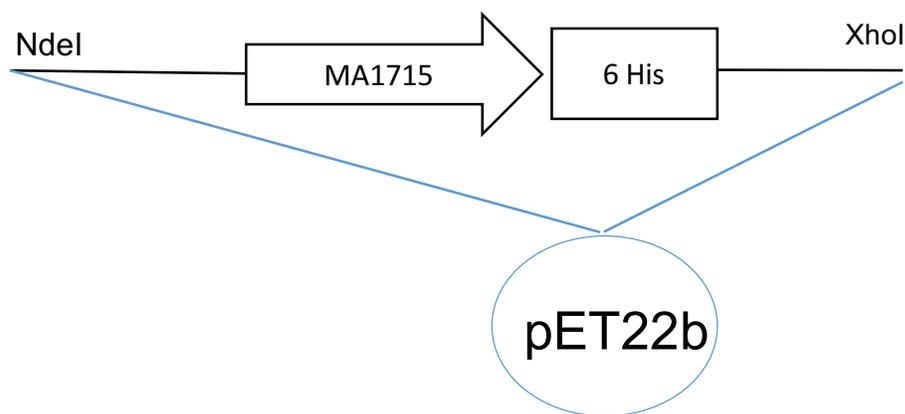


Figure 4.9 MA1715 Expression Vector

MA1715 was expressed aerobically in *E. coli* Rosetta2(DE3)pLysS cells with chloramphenicol resistance. The plasmid was inducible using IPTG and was ampicillin resistant.

The lysate was cleared by centrifugation at 30,000 g for 30 minutes at 4 °C. The supernatant was discarded and the pellets cleaned twice with 150 mL of washing buffer 2 (50 mM Tris, pH 8.0, 50 mM NaCl, 2% Triton X-100 (v/v), 1.5 mM β -mercaptoethanol 1.6 M urea) by resuspension through stirring for 20 minutes and centrifuging the solution at 30,000 g for 30 minutes, 4 °C. Removal of the detergent was accomplished by cleaning the pellet two more times with wash buffer 1. The cleaned inclusion bodies were stored at -20 °C.

Two-Step Denaturing and Refolding

The cleaned inclusion bodies were resuspended in 5 mL of extraction buffer 1 (50 mM Tris, pH 8.0, 50 mM NaCl, 10 mM β -mercaptoethanol, 7 M Gdn HCl) followed by centrifugation at 30,000 g for 30 minutes at 4 °C. The supernatant, now containing MA1715, was quickly added to 200 mL of dilution buffer (50 mM Tris, pH 8.0, 1 mM EDTA, 50 mM NaCl, 10 mM β -mercaptoethanol) to rapidly precipitate the protein and centrifuged at 30,000 g for 30 minutes at 4 °C. The pellet was collected and resuspended in 5 mL of extraction buffer 2 (50 mM Tris, pH 8.0, 50 mM NaCl, 10 mM β -mercaptoethanol, 8 M urea).

The supernatant was collected and pellet discarded. The solution was then added drop wise to 400 mL of refolding buffer (20 mM Tris, pH 8.0, 1 mM EDTA, 1

mM GSH, 0.1 mM GSSG) while being stirred slowly with a magnetic stir bar. The solution was moved to 4 °C and allowed to stir slowly for two days. The solution was loaded to 1 mL of Sepharose-Q anion exchange resin at a rate of 0.5 mL/min using a peristaltic pump. The protein was eluted using a step gradient in MA1715 elution buffer (10 mM potassium phosphate, pH 8, 1 M NaCl) (**Fig 4.10**). Fractions were pooled and stored at 4 °C. 1 L of cells yields roughly 1 mg of 95%+ pure MA1715 after refolding.

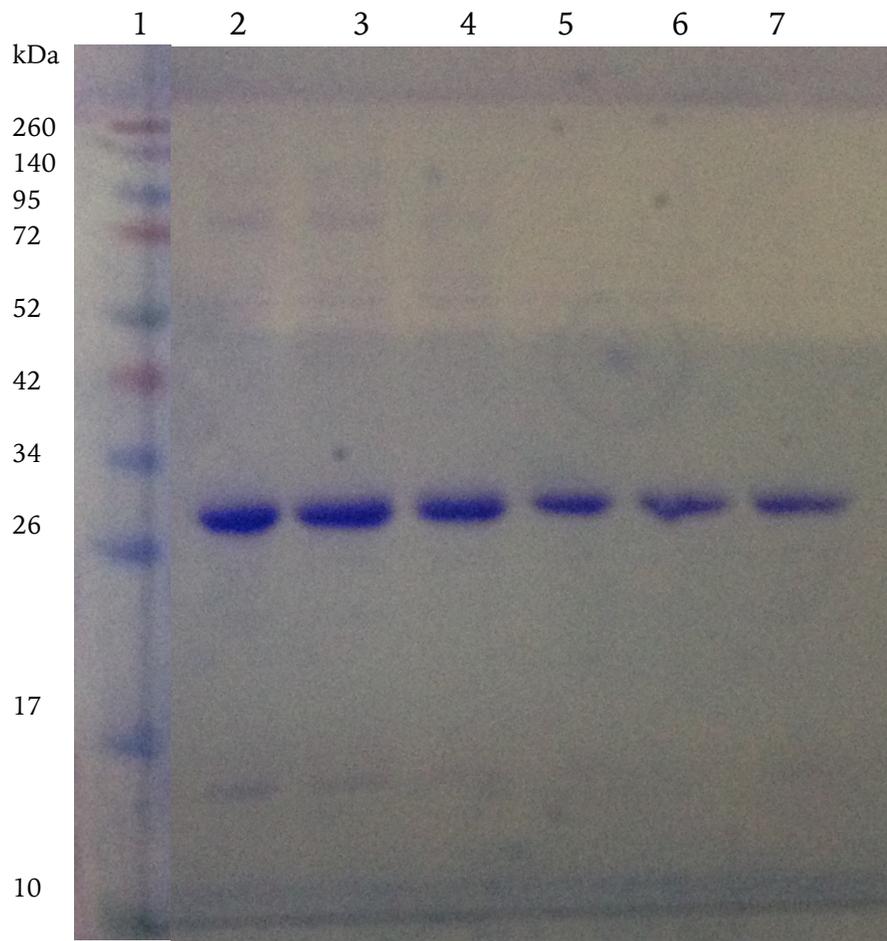


Figure 4.10 MA1715 Anion Exchange Elutions

Protein was bound to 1 mL resin volume of Q-Sepharose beads. Protein was eluted via a step elution from 0 to 1 M NaCl. Lanes are as follows,

1- Standard ladder

2-7 Elution fractions from anion exchange column

Discussion

Roughly 1 mg of MA1715 can be prepared at a purity level of at least 95% after refolding. This preparation provides a foundation for future work with MA1715 such as studying affinity for other co-conserved SepCysS proteins and crystallography for structural studies. Purification of other co-conserved SepCysS proteins will be necessary to continue studies of MA1715 beyond structural analysis.

While there is no SUMO tagged construct to provide evidence for transcription errors of MA1821 in *E. coli*, it is likely that the prominent remaining contaminants are truncated pieces of MA1821 because *E. coli* produces truncated MJ1681. Overcoming expression issues may require purification from the native organism. Tagged MA1821 could feasibly be obtained using the above lysis and Ni-NTA purification strategy when lysed from *M. acetivorans*.

MA1822 is not solubilized when lysed under the same high pH phosphate conditions. Knowing that this lysis buffer is not some universal solubilizing agent, and the variable MA1821 SEC results at differing pH values, it is reasonable to assume some un-aggregated MA1821 complex is formed. Without a monodispersed gel filtration column, and no further evidence from other methods, it cannot be assumed that MA1821 is folded in an active, native state, however. Future work with MA1821 should explore purification from different organisms to reduce transcription error, and

follow up on the pH dependent quaternary state through methods such as circular dichroism.

Tlet_1363 is soluble in the COG1900 lysis buffer but no further work was done beyond one affinity column. Tlet_1364 may prove more cooperative than COG1900D or MA1821 considering it has the advantage of a ferredoxin domain already attached, the lack of which may contribute to the difficulty in purifying MA1821. Further experimentation in purifying Tlet_1363 should be pursued as MJ1681 purification has been nearly exhausted and MA1821 is lacking the essential ferredoxin domain. If MA1821 purification is continued, the refolding of MA1822 and anaerobic reconstitution of the iron sulfur clusters may prove to be a viable preparation strategy.

A summary of all purifications is below (**Table 4.4**)

	MJ1681	Tlet_1363	MA1821	MA1822	MA1715
Cell volume	1 L	1 L	1 L	1 L	1 L
COG1900 lysis buffer soluble	Yes	Yes	Yes	No	No
COG1900 buffer resuspension	Yes	?	No	No	No
Refoldable	?	?	No	Yes	Yes
Yield (mg)	3	?	1.5	1	1
Purity (%)	80	?	75	95+	95+

Table 4.4 Summary of all Protein Purifications

Conclusions

The purification of COG1900 proteins is a difficult, yet very important hurdle to overcome. The most obvious benefit is a greater understanding of sulfur assimilation in methanogens, which is very poorly understood. But the purification of COG1900 proteins would prove beneficial beyond a better understanding of their metabolic functions. This work aims at providing a foundation for further research on COG1900 and COG2122 through developed purification methods.

Methanogenesis is an ancient way of life and a better understanding of the proteins involved provides significant evolutionary history. A better understanding of methanogenesis has practical purposes in both astrobiology and energy production. The biochemical function of taking an aldehyde to a thiol is of interest to biochemists, and the structure of COG1900 may not only elucidate how such a reaction occurs, but no COG1900 structure has ever been studied.

Archaea, and extremophiles more specifically, have many unstudied facets because the environment they survive in requires such unique biological adaptations which are often not compatible with canonical *in vitro* purification strategies. The purification strategies presented in this thesis may prove useful beyond the proteins tested. Unconventional methods, such as the high pH phosphate buffer, may prove

useful for studying other extremophile proteins. While the high pH buffer could be causing unwanted amino acid side chain reactions, it is not a general solubilizing agent and does have a unique effect on COG1900 proteins not seen on other proteins tested in this study.

Purification of COG2122 and the small ferredoxin protein is accomplished using a new approach to refolding proteins and both refold as monomers. COG1900 protein purification strategies have been exhausted through testing forty lysis conditions and optimizing the yield and purity of every chromatography step in the purification process. This work provides a viable purification of COG1900 proteins which will permit future work that is currently impossible due to any purification strategy in the literature. The purification of COG1900 could elucidate the structure of an unknown protein domain, the kinetics of homocysteine and coenzyme M synthesis, and a better understanding of ancient sulfur trafficking techniques.

References

1. Anbar, A. D. & Knoll, A. H. Proterozoic ocean chemistry and evolution: a bioinorganic bridge? *Science* **297**, 1137–1142 (2002).
2. Margulis, L. *Microcosmos : four billion years of evolution from our microbial ancestors /*. (Summit Books, c1986).
3. Kasting, J. F. Earth's early atmosphere. *Science* **259**, 920–926 (1993).
4. Ström, L., Mastepanov, M. & Christensen, T. R. Species-specific Effects of Vascular Plants on Carbon Turnover and Methane Emissions from Wetlands. *Biogeochemistry* **75**, 65–82
5. Miller, S. L. & Lazcano, A. The origin of life—did it occur at high temperatures? *J. Mol. Evol.* **41**, 689–692 (1995).
6. Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516–519 (2006).
7. Habicht, K. S., Gade, M., Thamdrup, B., Berg, P. & Canfield, D. E. Calibration of sulfate levels in the archaean ocean. *Science* **298**, 2372–2374 (2002).
8. Liu, Y., Beer, L. L. & Whitman, W. B. Methanogens: a window into ancient sulfur metabolism. *Trends Microbiol.* **20**, 251–258 (2012).

9. Sauerwald, A., Zhu, W., Major, T. A., Roy, H., Palioura, S., Jahn, D., Whitman, W. B., Yates, J. R., Ibba, M. & Söll, D. RNA-dependent cysteine biosynthesis in archaea. *Science* **307**, 1969–1972 (2005).
10. Ibba, M., Becker, H. D., Stathopoulos, C., Tumbula, D. L. & Söll, D. The adaptor hypothesis revisited. *Trends Biochem. Sci.* **25**, 311–316 (2000).
11. Tumbula, D., Vothknecht, U. C., Kim, H. S., Ibba, M., Min, B., Li, T., Pelaschier, J., Stathopoulos, C., Becker, H. & Söll, D. Archaeal aminoacyl-tRNA synthesis: diversity replaces dogma. *Genetics* **152**, 1269–1276 (1999).
12. O'Donoghue, P., Sethi, A., Woese, C. R. & Luthey-Schulten, Z. A. The evolutionary history of Cys-tRNA^{Cys} formation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 19003–19008 (2005).
13. Mueller, E. G. Trafficking in persulfides: delivering sulfur in biosynthetic pathways. *Nat. Chem. Biol.* **2**, 185–194 (2006).
14. Zheng, L., White, R. H., Cash, V. L., Jack, R. F. & Dean, D. R. Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 2754–2758 (1993).
15. Schwartz, C. J., Djaman, O., Imlay, J. A. & Kiley, P. J. The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9009–9014 (2000).

16. Liu, Y., Sieprawska-Lupa, M., Whitman, W. B. & White, R. H. Cysteine Is Not the Sulfur Source for Iron-Sulfur Cluster and Methionine Biosynthesis in the Methanogenic Archaeon *Methanococcus maripaludis*. *J. Biol. Chem.* **285**, 31923–31929 (2010).
17. Rothschild, L. J. & Mancinelli, R. L. Life in extreme environments. *Nature* **409**, 1092–1101 (2001).
18. Sekowska, A., Kung, H. F. & Danchin, A. Sulfur metabolism in *Escherichia coli* and related bacteria: facts and fiction. *J. Mol. Microbiol. Biotechnol.* **2**, 145–177 (2000).
19. Kessler, D. Enzymatic activation of sulfur for incorporation into biomolecules in prokaryotes. *FEMS Microbiol. Rev.* **30**, 825–840 (2006).
20. Hauenstein, S. I. & Perona, J. J. Redundant synthesis of cysteinyl-tRNA^{Cys} in *Methanosarcina mazei*. *J. Biol. Chem.* **283**, 22007–22017 (2008).
21. Fukunaga, R. & Yokoyama, S. Structural insights into the first step of RNA-dependent cysteine biosynthesis in archaea. *Nat. Struct. Mol. Biol.* **14**, 272–279 (2007).
22. Liu, Y., Santos, P. C. D., Zhu, X., Orlando, R., Dean, D. R., Söll, D. & Yuan, J. Catalytic Mechanism of Sep-tRNA:Cys-tRNA Synthase SULFUR TRANSFER IS MEDIATED BY DISULFIDE AND PERSULFIDE. *J. Biol. Chem.* **287**, 5426–5433 (2012).

23. Rauch, B. J., Gustafson, A. & Perona, J. J. Novel proteins for homocysteine biosynthesis in anaerobic microorganisms. *Mol. Microbiol.* **94**, 1330–1342 (2014).
24. Selkov, E., Maltsev, N., Olsen, G. J., Overbeek, R. & Whitman, W. B. A reconstruction of the metabolism of *Methanococcus jannaschii* from sequence data. *Gene* **197**, GC11–26 (1997).
25. Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A., Guss, A. M., Hedderich, R., Ingram-Smith, C., Kuettner, H. C., Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J. N., Smith, K., Springer, T. A., Umayam, L. A., White, O., White, R. H., Macario, E. C. de, Ferry, J. G., Jarrell, K. F., Jing, H., Macario, A. J. L., Paulsen, I., Pritchett, M., Sowers, K. R., Swanson, R. V., Zinder, S. H., Lander, E., Metcalf, W. W. & Birren, B. The Genome of *M. acetivorans* Reveals Extensive Metabolic and Physiological Diversity. *Genome Res.* **12**, 532–542 (2002).
26. Allen, K. D., Miller, D. V., Rauch, B. J., Perona, J. J. & White, R. H. Homocysteine Is Biosynthesized from Aspartate Semialdehyde and Hydrogen Sulfide in Methanogenic Archaea. *Biochemistry (Am. Chem. Soc.)* **54**, 3129–3132 (2015).

27. Rauch, B. J. & Perona, J. J. Efficient sulfide assimilation in *Methanosarcina acetivorans* is mediated by the MA1715 protein. *J. Bacteriol.* JB.00141–16 (2016).
doi:10.1128/JB.00141-16
28. Beck, B. J. & Downs, D. M. The *apbE* Gene Encodes a Lipoprotein Involved in Thiamine Synthesis in *Salmonella typhimurium*. *J. Bacteriol.* **180**, 885–891 (1998).
29. Beck, B. J. & Downs, D. M. A periplasmic location is essential for the role of the ApbE lipoprotein in thiamine synthesis in *Salmonella typhimurium*. *J. Bacteriol.* **181**, 7285–7290 (1999).
30. Boyd, J. M., Endrizzi, J. A., Hamilton, T. L., Christopherson, M. R., Mulder, D. W., Downs, D. M. & Peters, J. W. FAD binding by ApbE protein from *Salmonella enterica*: a new class of FAD-binding proteins. *J. Bacteriol.* **193**, 887–895 (2011).
31. Deka, R. K., Brautigam, C. A., Liu, W. Z., Tomchick, D. R. & Norgard, M. V. The TP0796 Lipoprotein of *Treponema pallidum* Is a Bimetal-dependent FAD Pyrophosphatase with a Potential Role in Flavin Homeostasis. *J. Biol. Chem.* **288**, 11106–11121 (2013).
32. Han, G. W., Sri Krishna, S., Schwarzenbacher, R., McMullan, D., Ginalski, K., Elsliger, M.-A., Brittain, S. M., Abdubek, P., Agarwalla, S., Ambing, E., Astakhova, T., Axelrod, H., Canaves, J. M., Chiu, H.-J., DiDonato, M., Grzechnik, S. K., Hale, J., Hampton, E., Haugen, J., Jaroszewski, L., Jin, K. K., Klock, H. E., Knuth, M. W.,

- Koesema, E., Kreusch, A., Kuhn, P., Miller, M. D., Morse, A. T., Moy, K., Nigoghossian, E., Oommachen, S., Ouyang, J., Paulsen, J., Quijano, K., Reyes, R., Rife, C., Spraggon, G., Stevens, R. C., van den Bedem, H., Velasquez, J., Wang, X., West, B., White, A., Wolf, G., Xu, Q., Hodgson, K. O., Wooley, J., Deacon, A. M., Godzik, A., Lesley, S. A. & Wilson, I. A. Crystal structure of the ApbE protein (TM1553) from *Thermotoga maritima* at 1.58 Å resolution. *Proteins* **64**, 1083–1090 (2006).
33. Major, T. A., Burd, H. & Whitman, W. B. Abundance of 4Fe–4S motifs in the genomes of methanogens and other prokaryotes. *FEMS Microbiol. Lett.* **239**, 117–123 (2004).
34. Malkin, R. & Rabinowitz, J. C. The reconstitution of clostridial ferredoxin. *Biochem. Biophys. Res. Commun.* **23**, 822–827 (1966).
35. Külzer, R., Pils, T., Kappl, R., Hüttermann, J. & Knappe, J. Reconstitution and characterization of the polynuclear iron-sulfur cluster in pyruvate formate-lyase-activating enzyme. Molecular properties of the holoenzyme form. *J. Biol. Chem.* **273**, 4897–4903 (1998).
36. Urey. On The Early Chemical History Of The Earth And The Origin Of Life. *Geophysics* **38**, (1952).

37. Brack, A. *The Molecular Origins of Life: Assembling Pieces of the Puzzle*. (Cambridge University Press, 1998).
38. Zhang, R., Evans, G., Rotella, F. J., Westbrook, E. M., Beno, D., Huberman, E., Joachimiak, A. & Collart, F. R. Characteristics and crystal structure of bacterial inosine-5'-monophosphate dehydrogenase. *Biochemistry (Am. Chem. Soc.)* **38**, 4691–4700 (1999).
39. Jentsch, T. J., Stein, V., Weinreich, F. & Zdebik, A. A. Molecular structure and physiological function of chloride channels. *Physiol. Rev.* **82**, 503–568 (2002).
40. Wong, P. C. L. & Henderson, J. F. Purine ribonucleotide biosynthesis, interconversion and catabolism in mouse brain in vitro. *Biochem. J.* **129**, 1085–1094 (1972).
41. Bräsen, C., Urbanke, C. & Schönheit, P. A novel octameric AMP-forming acetyl-CoA synthetase from the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *FEBS Lett.* **579**, 477–482 (2005).
42. Cheung, P. C., Salt, I. P., Davies, S. P., Hardie, D. G. & Carling, D. Characterization of AMP-activated protein kinase gamma-subunit isoforms and their role in AMP binding. *Biochem. J.* **346**, 659–669 (2000).

43. Stapleton, D., Mitchelhill, K. I., Gao, G., Widmer, J., Michell, B. J., Teh, T., House, C. M., Fernandez, C. S., Cox, T., Witters, L. A. & Kemp, B. E. Mammalian AMP-activated protein kinase subfamily. *J. Biol. Chem.* **271**, 611–614 (1996).
44. Scott, J. W., Hawley, S. A., Green, K. A., Anis, M., Stewart, G., Scullion, G. A., Norman, D. G. & Hardie, D. G. CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J. Clin. Invest.* **113**, 274–284 (2004).
45. Ignoul, S. & Eggermont, J. CBS domains: structure, function, and pathology in human proteins. *Am. J. Physiol. - Cell Physiol.* **289**, C1369–C1378 (2005).
46. Graham, D. E., Xu, H. & White, R. H. Identification of Coenzyme M Biosynthetic Phosphosulfolactate Synthase A NEW FAMILY OF SULFONATE-BIOSYNTHESIZING ENZYMES. *J. Biol. Chem.* **277**, 13421–13429 (2002).
47. Pavlov, A. A., Kasting, J. F., Brown, L. L., Rages, K. A. & Freedman, R. Greenhouse warming by CH₄ in the atmosphere of early Earth. *J. Geophys. Res. Planets* **105**, 11981–11990 (2000).
48. US EPA, C. C. D. Methane Emissions. Available at: <https://www3.epa.gov/climatechange/ghgemissions/gases/ch4.html>. (Accessed: 30th April 2016)

49. Johnson, K. A. & Johnson, D. E. Methane emissions from cattle. *J. Anim. Sci.* **73**, 2483–2492 (1995).
50. Tilman, D., Balzer, C., Hill, J. & Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci.* **108**, 20260–20264 (2011).
51. Weiland, P. Production and energetic use of biogas from energy crops and wastes in Germany. *Appl. Biochem. Biotechnol.* **109**, 263–274 (2003).
52. Lettinga, G. Anaerobic digestion and wastewater treatment systems. *Antonie Van Leeuwenhoek* **67**, 3–28 (1995).
53. Sekiguchi, Y., Kamagata, Y. & Harada, H. Recent advances in methane fermentation technology. *Curr. Opin. Biotechnol.* **12**, 277–282 (2001).
54. Health, N. C. for E. CDC - Climate Change and Public Health - Climate Effects on Health. Available at: <http://www.cdc.gov/climateandhealth/effects/>. (Accessed: 20th July 2016)
55. Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C., Knight, R. & Gordon, J. I. The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* **449**, 804–810 (2007).
56. Cavicchioli, R. Extremophiles and the search for extraterrestrial life. *Astrobiology* **2**, 281–292 (2002).

57. Karshikoff, A. & Ladenstein, R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a 'traffic rule' for hot roads. *Trends Biochem. Sci.* **26**, 550–557 (2001).
58. Elcock, A. H. The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **284**, 489–502 (1998).
59. Xiao, L. & Honig, B. Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol.* **289**, 1435–1444 (1999).
60. Vieille, C. & Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev. MMBR* **65**, 1–43 (2001).
61. Cavicchioli, R. *Archaea: Molecular and Cellular Biology*. (ASM Press, 2007).
62. Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L. & Engel, P. C. Protein thermostability in extremophiles. *Biochimie* **80**, 933–941 (1998).
63. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **5**, (2014).
64. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4285–4288 (1999).

65. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
66. Rauch, B. Sulfide Assimilation in Anaerobes: Novel Proteins for an Ancient Strategy. (2015).
67. Guss, A. M., Rother, M., Zhang, J. K., Kulkarni, G. & Metcalf, W. W. New methods for tightly regulated gene expression and highly efficient chromosomal integration of cloned genes for *Methanosarcina* species. *Archaea Vanc. BC* **2**, 193–203 (2008).
68. Pritchett, M. A., Zhang, J. K. & Metcalf, W. W. Development of a Markerless Genetic Exchange Method for *Methanosarcina acetivorans* C2A and Its Use in Construction of New Genetic Tools for Methanogenic Archaea. *Appl. Environ. Microbiol.* **70**, 1425–1433 (2004).
69. Metcalf, W. W., Zhang, J. K., Apolinario, E., Sowers, K. R. & Wolfe, R. S. A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 2626–2631 (1997).
70. Bowden, G. A., Paredes, A. M. & Georgiou, G. Structure and Morphology of Protein Inclusion Bodies in *Escherichia Coli*. *Nat. Biotechnol.* **9**, 725–730 (1991).

71. Leibly, D. J., Nguyen, T. N., Kao, L. T., Hewitt, S. N., Barrett, L. K. & Van Voorhis, W. C. Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. *PloS One* **7**, e52482 (2012).
72. Yang, Z., Zhang, L., Zhang, Y., Zhang, T., Feng, Y., Lu, X., Lan, W., Wang, J., Wu, H., Cao, C. & Wang, X. Highly Efficient Production of Soluble Proteins from Insoluble Inclusion Bodies by a Two-Step-Denaturing and Refolding Method. *PLOS ONE* **6**, e22981 (2011).
73. Boddy, M. N., Howe, K., Etkin, L. D., Solomon, E. & Freemont, P. S. PIC 1, a novel ubiquitin-like protein which interacts with the PML component of a multiprotein complex that is disrupted in acute promyelocytic leukaemia. *Oncogene* **13**, 971–982 (1996).
74. Mannen, H., Tseng, H.-M., Cho, C. & Li, S. S.-L. Cloning and Expression of Human Homolog HSMT3 to Yeast SMT3 Suppressor of MIF2 Mutations in a Centromere Protein Gene. *Biochem. Biophys. Res. Commun.* **222**, 178–180 (1996).
75. Matunis, M. J., Coutavas, E. & Blobel, G. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.* **135**, 1457–1470 (1996).
76. Okura, T., Gong, L., Kamitani, T., Wada, T., Okura, I., Wei, C. F., Chang, H. M. & Yeh, E. T. Protection against Fas/APO-1- and tumor necrosis factor-mediated cell

- death by a novel protein, sentrin. *J. Immunol. Baltim. Md 1950* **157**, 4277–4281 (1996).
77. Winkel, L. P. F., Van den Hout, J. M. P., Kamphoven, J. H. J., Disseldorp, J. A. M., Remmerswaal, M., Arts, W. F. M., Loonen, M. C. B., Vulto, A. G., Van Doorn, P. A., De Jong, G., Hop, W., Smit, G. P. A., Shapira, S. K., Boer, M. A., van Diggelen, O. P., Reuser, A. J. J. & Van der Ploeg, A. T. Enzyme replacement therapy in late-onset Pompe's disease: a three-year follow-up. *Ann. Neurol.* **55**, 495–502 (2004).
78. Mahajan, R., Delphin, C., Guan, T., Gerace, L. & Melchior, F. A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88**, 97–107 (1997).
79. Hochstrasser, M. Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405–439 (1996).
80. Lee, C.-D., Sun, H.-C., Hu, S.-M., Chiu, C.-F., Homhuan, A., Liang, S.-M., Leng, C.-H. & Wang, T.-F. An improved SUMO fusion protein system for effective production of native proteins. *Protein Sci. Publ. Protein Soc.* **17**, 1241–1248 (2008).
81. ArcticExpress Competent Cells. Available at:
http://www.genomics.agilent.com/article.jsp?pageId=468&_requestid=840573.
(Accessed: 31st May 2016)

82. Shirano, Y. & Shibata, D. Low temperature cultivation of *Escherichia coli* carrying a rice lipoxygenase L-2 cDNA produces a soluble and active enzyme at a high level. *FEBS Lett.* **271**, 128–130 (1990).
83. Lucas, M., Kortazar, D., Astigarraga, E., Fernández, J. A., Mato, J. M., Martínez-Chantar, M. L. & Martínez-Cruz, L. A. Purification, crystallization and preliminary X-ray diffraction analysis of the CBS-domain pair from the *Methanococcus jannaschii* protein MJ0100. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **64**, 936–941 (2008).
84. Lucas, M., Encinar, J. A., Arribas, E. A., Oyenarte, I., García, I. G., Kortazar, D., Fernández, J. A., Mato, J. M., Martínez-Chantar, M. L. & Martínez-Cruz, L. A. Binding of S-Methyl-5'-Thioadenosine and S-Adenosyl-l-Methionine to Protein MJ0100 Triggers an Open-to-Closed Conformational Change in Its CBS Motif Pair. *J. Mol. Biol.* **396**, 800–820 (2010).
85. Schwartz, R., Ting, C. S. & King, J. Whole Proteome pI Values Correlate with Subcellular Localizations of Proteins for Organisms within the Three Domains of Life. *Genome Res.* **11**, 703–709 (2001).
86. *Non-coding Genomics of Methanocaldococcus Jannaschii: A Survey of Promoters, Non-coding RNA Genes, and Repetitive DNA Elements.* (ProQuest, 2007).

87. Haney, P. J., Badger, J. H., Buldak, G. L., Reich, C. I., Woese, C. R. & Olsen, G. J. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc. Natl. Acad. Sci.* **96**, 3578–3583 (1999).
88. ExPASy ProtParam tool. Available at: <http://web.expasy.org/cgi-bin/protparam/protparam>. (Accessed: 8th April 2016)
89. Zangi, R. Can Salting-In/Salting-Out Ions be Classified as Chaotropes/Kosmotropes? *J. Phys. Chem. B* **114**, 643–650 (2010).
90. Tadeo, X., López-Méndez, B., Castaño, D., Trigueros, T. & Millet, O. Protein Stabilization and the Hofmeister Effect: The Role of Hydrophobic Solvation. *Biophys. J.* **97**, 2595–2603 (2009).
91. Moelbert, S., Normand, B. & De Los Rios, P. Kosmotropes and chaotropes: modelling preferential exclusion, binding and aggregate stability. *Biophys. Chem.* **112**, 45–57 (2004).
92. Yang, Z. Hofmeister effects: an explanation for the impact of ionic liquids on biocatalysis. *J. Biotechnol.* **144**, 12–22 (2009).
93. Friedman, M. Chemistry, Biochemistry, Nutrition, and Microbiology of Lysinoalanine, Lanthionine, and Histidinoalanine in Food and Other Proteins. *J. Agric. Food Chem.* **47**, 1295–1319 (1999).

94. Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D. & Butt, T. R. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **5**, 75–86 (2004).
95. Agar, J. N., Krebs, C., Frazzon, J., Huynh, B. H., Dean, D. R. & Johnson, M. K. IscU as a Scaffold for Iron–Sulfur Cluster Biosynthesis: Sequential Assembly of [2Fe-2S] and [4Fe-4S] Clusters in IscU. *Biochemistry (Am. Chem. Soc.)* **39**, 7856–7862 (2000).
96. Available at: https://www.sigmaaldrich.com/content/dam/sigmaaldrich/docs/Sigma/Product_Information_Sheet/1/t8532pis.pdf. (Accessed: 21st July 2016)
97. Jones, W. J., Leigh, J. A., Mayer, F., Woese, C. R. & Wolfe, R. S. *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch. Microbiol.* **136**, 254–261 (1983).
98. 4.2.1 Circular dichroism spectroscopy. Available at: http://www.cryst.bbk.ac.uk/PPS2/course/section8/ss-960531_21.html. (Accessed: 23rd June 2016)

Appendix A: List of Buffers

Washing Buffer I

20 mM Tris, pH 8.0

Washing Buffer II

50 mM Tris, pH 8.0, 50 mM NaCl, 2% Triton X-100 (v/v), 1.5 mM β -mercaptoethanol,

1.6 M urea

Buffer C

50 mM Tris, pH 8.0

Buffer D

20 mM Tris, pH 8.0, 1 mM EDTA, 1 M NaCl

Extraction Buffer I

50 mM Tris, pH 8.0, 50 mM NaCl, 10 mM β -mercaptoethanol, 7 M Gdn HCl

Extraction Buffer II

50 mM Tris, pH 8.0, 50 mM NaCl, 10 mM β -mercaptoethanol, 8 M urea

Dilution Buffer

50 mM Tris, pH 8.0, 1 mM EDTA, 50 mM NaCl, 10 mM β -mercaptoethanol

Refolding Buffer

20 mM Tris, pH 8.0, 1 mM EDTA, 1 mM GSH, 0.1 mM GSSG

COG1900A Lysis Buffer

10 mM potassium phosphate, pH 13.0, 20 mM NaCl

COG1900D Lysis Buffer

20 mM Tris, pH 8.0

COG1900D Resuspension Buffer

10 mM potassium phosphate, pH 13.0, 20 mM NaCl

COG1900 Ni-NTA Elution Buffer

10 mM potassium phosphate, pH 9.0, 20 mM NaCl, 1 M imidazole

MA1715 Elution Buffer

10 mM potassium phosphate, pH 8, 1 M NaCl

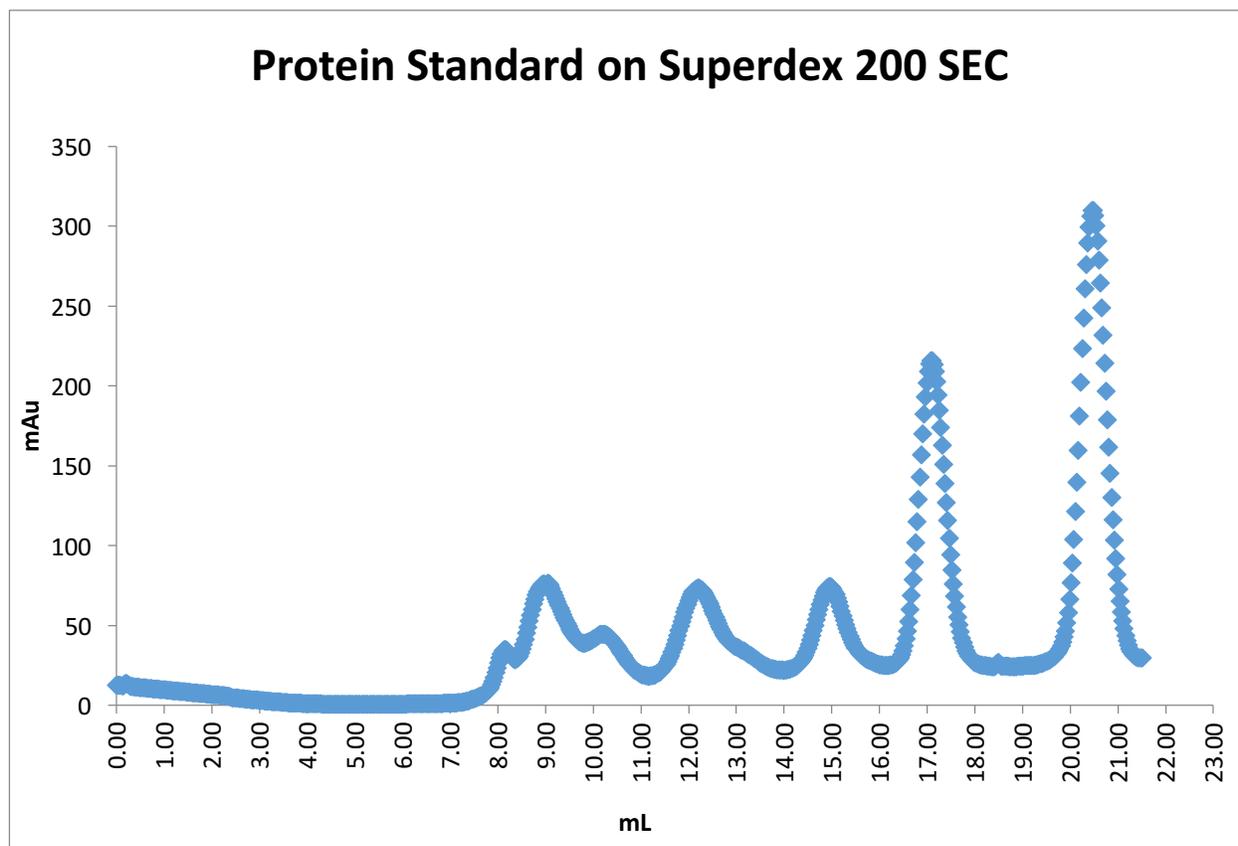
Ni-NTA pH 8 buffer

10 mM potassium phosphate, pH 8, 20 mM NaCl

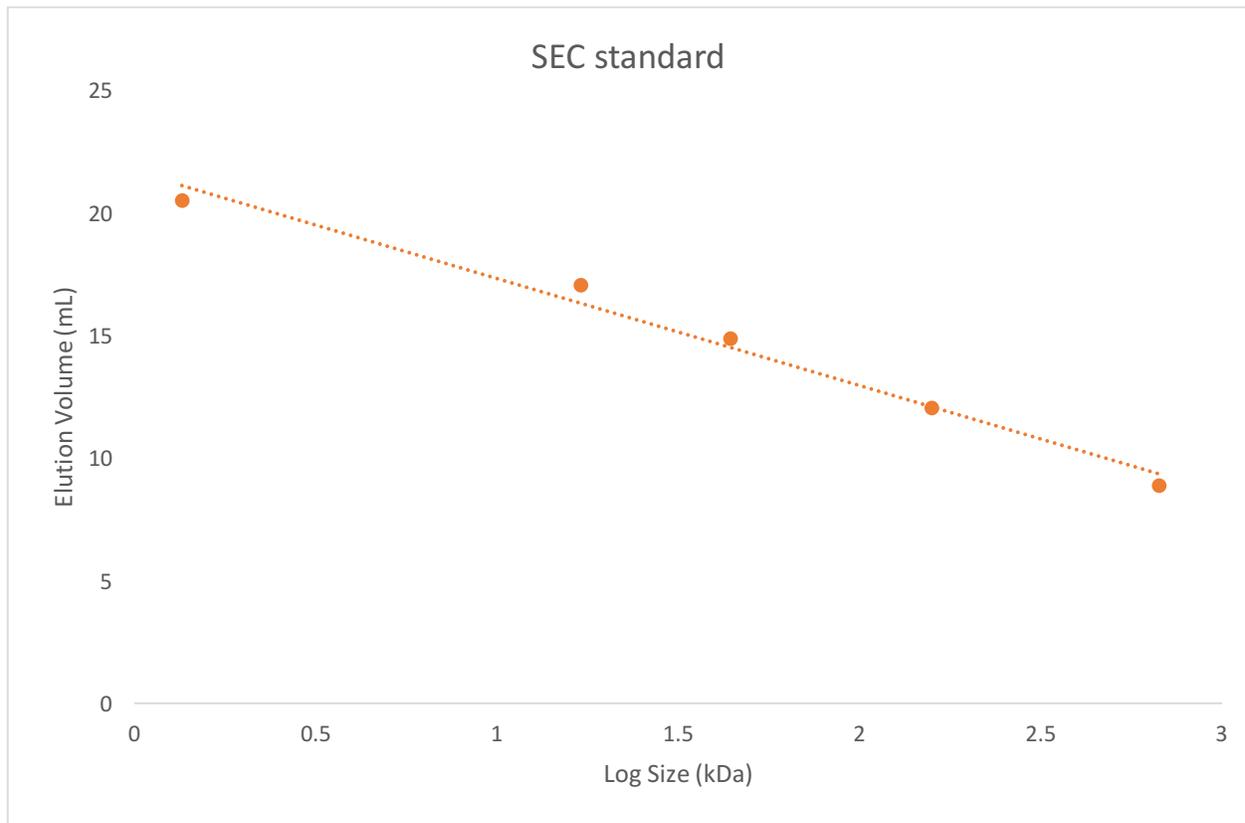
Ni-NTA pH 8 elution buffer

10 mM potassium phosphate, pH 8, 20 mM NaCl, 1 M imidazole

Appendix B: Standards



The standard sample used on the GE superdex 200 column. The five peaks from left to right are thyroglobulin (670 kDa), γ -globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), vitamin B₁₂ (1.35 kDa).



Predicting the size of an eluted molecule from the superdex 200 can be computed using this standard plot. A plot of the elution volume against the log of the standard elutant size produces a standard linear curve

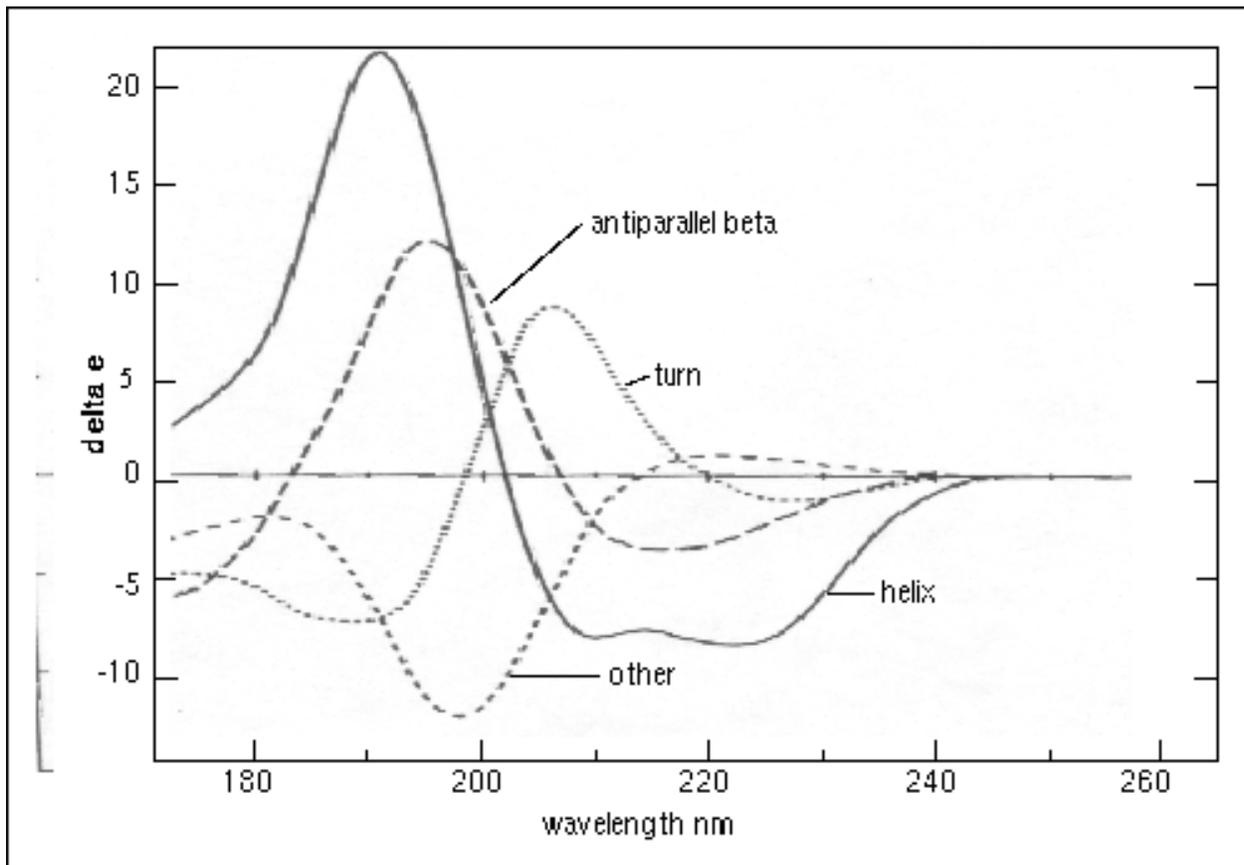
$$y = -4.3601x + 21.661$$

Upon solving for x based on a species elution volume (y), the log can be back calculated as 10^x .

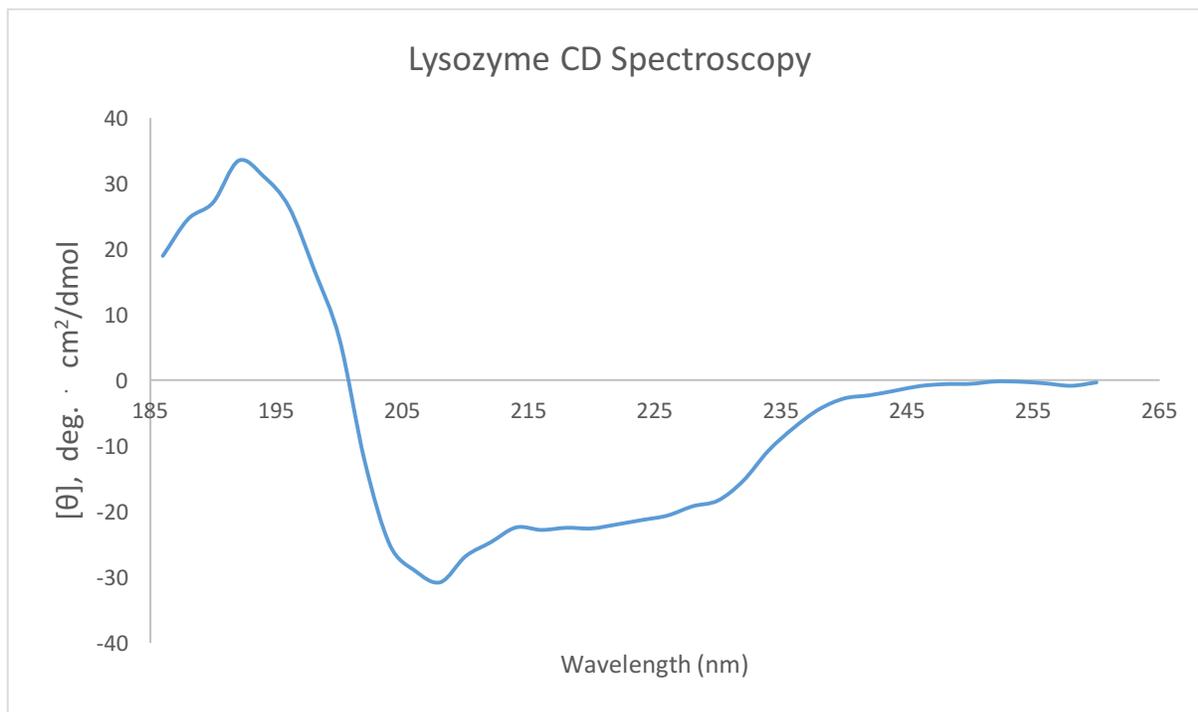
Therefore, solving

$$10^{(y-21.661)/-4.3601}$$

predicts the eluted species size in kDa.



Interpretation of CD spectroscopy data⁹⁸



CD Spectroscopy of lysozyme as a positive control. The sample was concentrated at 4.5 mg/mL. Lysozyme contains many alpha helices and the negative signal between 200 and 240 nm is indicative of alpha helical structure (seen in interpretation on previous page). COG1900D presented a strong signal for alpha helical structure and so lysozyme was used as a positive control to confirm that a similar chromatogram is produced.