

**THE FUNCTIONAL CHARACTERIZATION OF A *HALOARCU*  
*MARISMORTUI* PUTATIVE TRKE**

by

**Martha Giesbrecht**

BSc., University of Northern British Columbia, 2019

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE  
IN  
BIOCHEMSITRY

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

December 2023

© Martha Giesbrecht, 2023

## **Abstract**

Archaea, a domain of organisms possibly linked to the ancestry of eukarya and bacteria, displays a dichotomic evolutionary pattern. *Haloarcua marismortui* (*H. marismortui*), an archaeon discovered in the Dead Sea, has unique traits enabling survival and thriving in hypersaline environments. Research in this study includes investigating the Trk potassium transport system, which may contribute to stability of this archaeon in varying extreme environmental conditions. A primary focus on the TrkE protein role in the system, also known as SapD, which in *Haloarcua marismortui* is OppD1. The Trk system has been studied in several species, and through each it has been consistently found to be homologous to the *E. coli* sapABCDF operon, which has been found to encode an ABC transporter. Within this operon in *E. coli* TrkE is coded for by the SapD gene. The goal of this study was to clone the TrkE homolog from *Haloarcua marismortui* and continue with further research into the characterization of the protein to aid in the prediction of its overall function in potassium transport. Data in this study strongly suggest the size of the *H. marismortui* homolog of TrkE, OppD1, is larger than its *E. coli* homolog by approximately 10,000 Daltons. Further analysis also identified that the mainly alpha-helical protein has a significant sequence identity with SapD. Future considerations include looking at the purpose of the N-terminal extended region of OppD1 not seen in SapD, and further analyzing OppD1 for structure changes in varying salt conditions and possible binding partners. The exceptional survival skills of *H. marismortui* in high-salinity environments, possibly facilitated by the Trk system, make it an excellent model for studying halophiles. These insights offer valuable understanding into how halophiles maintain cellular integrity under

harsh conditions, especially regarding OppD1's potential role within the Trk system for ion transport and osmotic regulation in different environmental settings.

# Table of Contents

<b>THE FUNCTIONAL CHARACTERIZATION OF A <i>HALOARCUA MARISMORTUI</i></b>	
<b>PUTATIVE TRKE .....</b>	<b>I</b>
<b>ABSTRACT .....</b>	<b>II</b>
<b>LIST OF TABLES.....</b>	<b>VII</b>
<b>LIST OF FIGURES.....</b>	<b>VIII</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>XIII</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>XV</b>
<b>CHAPTER ONE: INTRODUCTION TO ARCHAEA AND THE TRK POTASSIUM</b>	
<b>TRANSPORT SYSTEM .....</b>	<b>1</b>
<b>1.1 ARCHAEA .....</b>	<b>2</b>
<b>1.2 ARCHAEL TRANSCRIPTION .....</b>	<b>2</b>
<b>1.3 ARCHAEL TRANSLATION.....</b>	<b>10</b>
<b>1.4 ARCHAEL GENOME AND PROTEOME ADAPTATIONS .....</b>	<b>14</b>
<b>1.5 RCK TRANSPORTERS HKT/Trk/KTR AS SALT CONDUCTORS.....</b>	<b>16</b>
<b>1.6 SCOPE AND SIGNIFICANCE.....</b>	<b>25</b>
<b>1.7 OBJECTIVES .....</b>	<b>26</b>
<b>CHAPTER TWO: CLONING AND EXPRESSION OF OPPD1.....</b>	<b>27</b>
<b>2.1 CHAPTER OBJECTIVES.....</b>	<b>28</b>
<b>2.2 METHODOLOGY .....</b>	<b>28</b>
<b>2.2.1 Expression Vector Purification and Quantification .....</b>	<b>28</b>
<b>2.2.2 Nested PCR Amplification of OppD1 – OppF Gene Segment .....</b>	<b>28</b>

2.2.3 PCR Amplification of <i>OppD1</i> from <i>OppD1-OppF</i> Gene Segment .....	32
2.2.4 Cloning <i>OppD1</i> into the Expression Vector <i>pET21b</i> to Generate the <i>Recombinant Vector pET21b-OppD1</i> .....	34
2.2.5 <i>OppD1</i> Expression .....	35
2.2.6 Western Blotting.....	35
2.2.7 <i>OppD1</i> Purification through FPLC Ion-Exchange Chromatography and Subsequent Concentration of Purified Samples .....	36
<b>2.3 RESULTS</b> .....	36
2.3.1 <i>OppD1</i> Amplification .....	36
2.3.2 <i>OppD1-pET21b(+)</i> Recombinant Vector Construction .....	39
2.3.3 <i>OppD1</i> Expression and Isolation .....	44
<b>2.4 DISCUSSION AND CONCLUSION</b> .....	49
<b>CHAPTER THREE: CHARACTERIZATION OF OPPD1</b> .....	<b>52</b>
3.1 CHAPTER OBJECTIVES.....	53
3.2 METHODOLOGY .....	53
3.2.1 Size Exclusion Chromatography (Gel Filtration).....	53
3.2.2 Circular Dichroism Spectroscopy.....	53
3.2.3 Bioinformatics.....	54
<b>3.3 RESULTS</b> .....	54
3.3.1 Chromatography Analysis .....	54
3.3.2 Circular Dichroism Analysis .....	58
3.3.3 Additional Bioinformatic Analysis .....	62
<b>3.4 DISCUSSION AND CONCLUSION</b> .....	63

<b>CHAPTER 4: CONCLUDING REMARKS .....</b>	<b>69</b>
<b>4.1 SUMMARY .....</b>	<b>70</b>
<b><i>4.1.1 Introduction – Chapter 1 .....</i></b>	<b>70</b>
<b><i>4.1.1.1 Introduction to Archaea and Archaeal Transcription.....</i></b>	<b>70</b>
<b><i>4.1.1.5 Haloarcula marimortui.....</i></b>	<b>71</b>
<b><i>4.1.2 Cloning and Expression of OppD1– Chapter 2 .....</i></b>	<b>73</b>
<b><i>4.1.3 Characterization of OppD1 – Chapter 3 .....</i></b>	<b>73</b>
<b>4.2 SIGNIFICANCE .....</b>	<b>74</b>
<b>4.3 CONCLUSION AND FUTURE CONSIDERATIONS.....</b>	<b>75</b>
<b>REFERENCES .....</b>	<b>78</b>
<b>APPENDIX: .....</b>	<b>89</b>

## List of Tables

Table 1. Trk system proteins with proposed function, and analogs studied in the Gorrell Lab at UNBC. [64]; [4]; [60]; [65]; [54]. .....	21
Table 2. Primers for <i>OppD1</i> amplification from <i>H. marismortui</i> via PCR. Utilized for ligation and verification of recombinant clones. ....	30

## List of Figures

Figure 1. Archaeal transcription initiation adapted from [14].	5
Figure 2. (A) Eukaryotic transcription initiation adapted from [15]. (B) Prokaryotic transcription initiation adapted from [16].	6
Figure 3. Canonical ABC transporter architecture including two transmembrane domains (TMD) and two nucleotide-binding domains (NBD) that function together in membrane transport of the molecule shown (grey hexagon) through the hydrolysis of ATP. Adapted from [54].	18
Figure 4. Proposed gating mechanism of transmembrane protein (TrkH) along with cytosolic membrane bound regulatory protein (TrkA) adapted from [59].	22
Figure 5. The relative position of oppD1 and oppF on Chromosome I of the haloarchaea <i>Haloarcula marismortui</i> . Orange arrows represent forward and reverse primers, OppD1F-NdeI and OppF_R_XhoI (Table 2.) used to produce transcript OppD1-OppF. Annotated genes not IDs elsewhere include: pkn, troR, nifU2, and htID.	31
Figure 6. <i>OppD1</i> PCR Amplification from <i>H. marismortui</i> shown on a 1% agarose gel. Amplification of <i>OppD1</i> was attempted here with a <i>H. marismortui</i> genomic sample and a previously nested product of <i>OppD1-OppF</i> . This agarose gel shows complete amplification of <i>OppD1</i> from the genomic samples with a band sitting at ~1.1kb. No <i>OppD1</i> amplification can be seen in PCR reactions with the nested product. PCR reactions were set up with Phusion GC setup and the <i>NdeI</i> and <i>XhoI</i> primers. Promega's 1kb DNA Ladder served as a reference (Appendix – Figure 1).	37

Figure 7. Colony PCR with the T7 and T7Rev primers of 8 E.coli colonies to determine successful cloning of TrkE homolog *OppD1* into pET21b. pET21b-TrkA ran as a positive control for successful primer binding and amplification. Lanes are numbered according to the colonies used for PCR verification. Promega's 1kb DNA Ladder served as a reference (Appendix – Figure 1.)..... 40

Figure 8. Verification of successful recombinant clones through colony PCR amplification with 3 combinations of the internal and T7 primers as noted (Table 2) (primer combinations are as follows: #-1: *T7For* and *OppD1\_R\_internal1*; #-2: *OppD1\_F\_internal1* and *OppD1\_R\_internal2*; #-3: *OppD1\_F\_internal2* and *T7Rev*). Colonies used here for further verification were initially chosen based on verification from previous figure (Figure 7). PCR reactions performed using the Phusion GC reaction setup with the incorporation of DMSO to promote binding of high GC content templates. Promega's 1kb DNA Ladder served as a reference (Appendix – Figure 1). ..... 42

Figure 9. (A) A sample of the *OppD1\_F\_internal1* primer sequence of pET21b(+) indicating the presence of *OppD1* in the expression vector. Constructed with FinchTV. Remaining sequence data can be found in the appendix (Figure 5 – 8). (B) Segment of DNA alignment of the recombinant *pEt21b(+)-OppD1*. Shown are partial sequence alignments amplified from the internal forward primer 1 (*OppD1\_F\_internal1*) and the internal reverse primer 2 (*OppD1\_R\_internal2*). Full sequencing data can be found in the appendix (Figure 9). ..... 43

Figure 10. Expression analysis of OppD1 in the *E.coli* expression line Rosetta. Expression of the protein can be analyzed here through a 12% SDS-PAGE gel stained with coomassie blue stain. .... 45

Figure 11. Western Blot targeting His-Tag incorporated into OppD1 seen after developing with BCIP. NEB Coloured Pre-Stained Standard Broad Rangel ladder served as a reference (Appendix – Figure 2). OppD1 expression highlighted. .... 46

Figure 12. Chromatogram of NiNTA-column of OppD1. Red and dark blue curves indicate absorbance and 254nm and 280nm respectively. Black and light blue curves indicate conductivity during separation. Green curve indicates gradient of Buffer B %. .... 47

Figure 13. A 12% SDS-PAGE of the His-Tag Affinity Chromatography experiment shown. Samples indicating OppD1 presence can be seen in samples 24-27. NEB Coloured Pre-Stained Standard Broad Rangel ladder served as a reference (Appendix – Figure 2)..... 48

Figure 14. Sample 25 and 26 from the chromatogram on the previous page were concentrated and loaded onto the HiLoad Superdex 16/60 200 Prep Grade GE Column and ran over 1.4 column volumes with the S200 Buffer (50mM Tris; 0.5 M KCl; pH 7.5). Red and Blue indicate absorbance at wavelengths of 280 nm and 254 nm. Orange line gives indication of the concentration of Buffer B which is the S200 buffer used for size exclusion chromatography, and the brown line is conductivity. Numbers shown on peaks indicated retention time, which indicated elution of OppD1 (purple arrow) (See Figure 15 for comparison)..... 55

Figure 15. A 12% SDS-PAGE of HiLoad Superdex 16/60 200 Prep Grade GE Column elution samples from Figure 14. Electrophoresis shows band in elution volume 61 mL of correct size for OppD1. NEB Coloured Pre-Stained Standard Broad Rangel ladder served as a reference (Appendix – Figure 2)..... 56

Figure 16. HiLoad Superdex 16/60 200 Prep Grade GE Column gel filtration run with S200 Buffer (300mM KCl; 50mM Tris; pH 7.5) and samples 59-62 from first SEC run with the HiLoad Superdex 16/60 200 Prep Grade GE Column seen in Figure 14. Buffer used was the S200 Buffer (50mM Tris; 0.5 M KCl; pH 7.5). ..... 57

Figure 17. OppD1 circular dichroism spectra’s from Jasco J-1000 with the following specifications – cell length: 10 mm; scan range: 450-200 nm; data pitch: 1.00 nm; D.I.T: 2 sec; bandwidth: 2 nm; scanning speed: 200 nm/min; accumulation: 3; solvent (300 mM KCl; 50 mM Tris; pH 7.5). Spectra showing appropriate amount of signal to noise ratio with HT levels below 700V, and a lower level of absorption seen across wavelengths. .... 59

Figure 18. OppD1 Circular Dichroism Spectra from Jasco J-1000 with the following specifications – cell length: 10 mm; scan range: 450-200 nm; data pitch: 1.00 nm; D.I.T: 2 sec; bandwidth: 2 nm; scanning speed: 200 nm/min; accumulation: 3; solvent (300 mM KCl; 50 mM Tris; pH 7.5). Spectra data is indicative of significant alpha helical structural components. .... 60

Figure 19 . Single spectrum analysis output from BeStSel online bioinformatic program with maximum usable wavelengths of 200 nm to 250nm. Figure was generated in Microsoft Corporation. (2018). Microsoft Excel. Figure shows percentages of each

<p>categorical secondary structure as approximated by BeStSetl present in OppD1 based on CD spectrum data (Figure 18). .....</p>	61
<p>Figure 20. Complete alignment of <i>E.coli</i> SapD (green) with partial sequence of <i>H. marismortui</i> (Coral) OppD1 (Alpha Fold). Figure 21A shows front view of alignment and 21B shows a 90 degree counter clockwise view around the y-axis.....</p>	64
<p>Figure 21. Depicts the N-terminal unordered sequence of OppD1 shown in wire frame presentation, which is not found in the SapD sequence. Figure 22A shows front view of alignment and 22B shows a 90 degree counter clockwise view around the y-axis. The amino acids shown here are from positions 326 to 393 in the polypeptide chain. Full sequence of amino acids can be seen in the Appendix – Figure 10.....</p>	65
<p>Figure 22. Overlay of OppD1 linear sequence generated from Benchling (green) on SapD (yellow). Figure only shows sequence of OppD1 that aligns with SapD. Figure 23A figure shows front view of alignment and 23B shows a 90 degree counter clockwise view around the y-axis. ....</p>	66

## List of Abbreviations

ABC	ATP-binding cassette
CATH	Class (C), Architecture (A), Topology (T), Homology (H) protein secondary structure organization system
<i>cis</i> -asRNAs	<i>cis</i> -antisenseRNAs
<i>H. marismortui</i>	<i>Haloarcula marismortui</i> (abbreviated)
HRP	Horseradish peroxidase
LrpA	leucine-responsive regulatory protein A
MDR1	Metal dependent repressor
mRNA	messenger RNA – messenger ribonucleic acid
Opp	oligopeptide permease
OppD1	Protein translated from <i>OppD1</i> sequence
<i>OppD1</i>	<i>H. marismortui</i> genetic sequence homologous to <i>TrkE</i> gene in <i>E. coli</i>
RCK	regulator of conductance (K <sup>+</sup> )
RNAP	RNA polymerase – ribonucleic acid polymerase
rRNA	ribosomal RNA – ribosomal ribonucleic acid
Sap	sensitive to antimicrobial peptides
SD	Shine-Dalgarno sequence
TBP	TATA-Binding Protein
TFB	TFIIB Transcription B Factor in archaea
TFIIB	Transcription Factor IIB
<i>TrkE</i>	<i>E. coli</i> gene

TrkE	Protein translated from <i>TrkE</i>
tRNA	transfer RNA – transfer ribonucleic acid
UTR	un-translated region

## **Acknowledgements**

I would like to begin by expressing my heartfelt gratitude to the individual who graciously provided their guidance and supervision for this endeavor. She is the primary reason I have discovered immense joy in the realm of biochemistry. Since my initial exposure to the subject during the "Introductory to Biochemistry" course with Dr. Andrea Gorrell, I have been determined to pursue research in this field throughout my academic career. Therefore, I extend my sincere thanks to Andrea for not only serving as my supervisor but also for being an inspiration, a colleague, and a friend.

Next, I wish to convey my appreciation to Dr. Danie Erasmus for affording me the privilege of serving as his Teaching Assistant. This role has had a profound impact on my self-assurance and public speaking abilities, surpassing even my most optimistic expectations. I am profoundly thankful to Danie for equipping me with the tools to think critically and for imparting the skills necessary to educate the next generation about the marvels of biochemistry. I have greatly treasured our collaboration and my experience as a student in your class. Your insights and wisdom have been invaluable, and I am equally grateful for your willingness to participate in graduate review committee.

I also wish to acknowledge Dr. Lisa Wood, a member of my committee. When it came time to select an additional member for my graduate committee, you immediately came to mind. Before enrolling in one of your courses, I had not previously harbored an affinity for plant biology or related subjects. I extend my gratitude to you for broadening my intellectual horizons and for your willingness to contribute to the evaluation of my graduate work.

Moreover, I extend my appreciation to the Doug Floyd Memorial Scholarship and the Bruno Raeber Student Award for their monetary contributions to my educational journey.

To all my friends, whether from my university days, the workplace, or life in general, I offer my sincere thanks for providing me with the steadfast support that has helped me maintain my equilibrium. Your reminder that the path we chart often requires adjustments and occasional detours is a valuable lesson. You have outlined the importance of utilizing every twist and turn, every obstacle and challenge, as an opportunity for personal growth.

I must also express my gratitude to my students and colleagues at Westside Academy. Your patience with my often-hectic schedule, coupled with your willingness to allow me to share my passion for science with our students, is deeply appreciated.

To my mom and dad, I extend my profound gratitude for your unwavering belief in me, even when my pursuits and objectives may have seemed unclear. You have exemplified the essence of hard work, and I can only hope to be a source of pride because of your tireless efforts in raising me. I am increasingly aware of the sacrifices you have made for my benefit with each passing day, and I want you to know how much I appreciate your efforts. Thank you, and I love you both dearly.

To my awesome siblings, thank you for being a source of comedic relief when things got stressful. Thank you for all the late-night snack runs; venting sessions; and just being there if I needed someone to talk to. You are truly remarkable, and I couldn't ask for a better, or funnier, family.

Lastly, I want to express my gratitude to my husband, Tyler, who entered my life toward the conclusion of my thesis. I am thankful that you have yet to witness me during my most demanding and frantic moments. Your unyielding support is a source of great comfort. I'm extremely thankful for your support to this day, and I'm grateful for your support from this day on.

I understand that some of you may not fully comprehend the profound impact you have had on my life and in bringing this achievement to fruition. However, I hope that I have been able to convey, to some extent, the deep appreciation I hold for each of you and the pivotal roles you have played in my journey. I am immensely grateful for all of this and extend my heartfelt wishes for your continued success and well-being. For all this I appreciate you all and wish you all the best.

**Chapter One: Introduction to Archaea and the Trk Potassium  
Transport System**

## 1.1 Archaea

Archaea is the domain of organisms characterized by a possible ancestry to the more established domains of eukarya and bacteria [1]. After the discovery of the archeal domain of life and its characterization, it soon became apparent this is a dichotomic domain and evolutionarily evolved. Everything from the basic machinery of life to their execution of functions suggests eukaryotic and prokaryotic organisms have a common archaeal ancestor.

*Haloarcula marismortui* (*H. marismortui*) is the organism of interest in this study and was first identified in the Dead Sea of the Middle East, along with *Haloarcula vallismortis* and *Haloferax volcanii* [2]. *H. marismortui*'s genome has been sequenced, and research continues to reveal the unique features that not only allow them to withstand these hypersaline environments but also thrive in these extreme environments [3]. One aspect of these halophiles, which is suggested to be a possible mechanism for survival, is the Trk system [4, 5]. Literature suggests a similar function for stability for these organisms and the increased virulence seen in several other species with similar operons [6].

## 1.2 Archaeal Transcription

The transcriptional machinery of archaeal species loosely resembles the structure found in eukaryotes [7]. It has been found that archaeal promoters contain a TATA box element and a factor B recognition site (Figure 1), which interact respectively with a related TATA-binding protein (TBP) and TFIIB transcription B factor (TFB) in the archaeon *Saccharolobus shibatae* (*S. shibatae*) [8]. The acting polymerase also mirrors the eukaryotic system, with an RNA polymerase (RNAP) consisting of twelve subunits [8, 9]. A comparison of Figures 1, 2A, and 2B shows the basic archaeal machinery resembles a eukaryotic system (Figure 2A), while regulatory mechanisms of transcription resemble that of bacterial systems more closely [8]. Large effects in

messenger ribonucleic acid (mRNA) production, in archaea, are being placed on interactions of binding sites similar to prokaryotes [8].

One example of negative regulation in archaeal expression is seen in the protein leucine-responsive regulatory protein A (LrpA), a homolog of the Lrp/AsnC amino acid metabolism regulators in bacteria, represses its own expression by binding specifically to its own promoter close to the start site of transcription [10]. Similar regulatory mechanisms have been identified in other archaeal proteins, which also result in the repression of gene expression through interference of polymerase docking via recruitment of TBP and TFB [8]. MDR1 is a second example repressor protein found in *Archaeoglobus fulgidus* and it regulates its own transcription through a metal-dependent interaction with a partial overlap of the initiator sequence [11]. MDR1 is a homolog of the bacterial regulator Dtxr, which has also shown to behave in a metal-dependent fashion [11]. MDR1 has been found to interfere with the recruitment of RNAP to the initiator sequence but still allows TBP and TFB to dock, which demonstrates this chimeric architecture seen in archaea, a bacterial regulatory system enforced on a eukaryal transcription setup [11]. Another Lrp/AsnC-like protein that has been found in the archaeal species *Sulfolobus solfataricus* (*S. solfataricus*) is LysM, which is an activator for the transcription of *lysW*XJK [8]. This gene is thought to regulate the biosynthesis of lysine and arginine in response to lysine levels [12]. LysM performs this activation by binding upstream from the *lysW* promoter, thereby stabilizing the promoter for favorable recruitment of the transcriptional machinery [12]

Positive regulation mechanisms also demonstrate both similar bacterial features, and eukaryal resemblance as well [8]. In halophilic archaea the regulation of gas vesicle formation has been found to be through the protein GvpE, which is homologous to the basic leucine zippers (bZIP) family evolutionarily conserved in eukaryotic systems [8]. The exact mechanism by

which activation of the gas vesicle formation is achieved has still not been determined, however it is known that GvpD is the transcriptional repressor likely through direct interaction with GvpE [8].

These regulatory mechanisms discussed have shown changes in the transcription process by influencing the initial recruitment of the transcriptional machinery. However, more recently the first archaeal protein capable of inhibiting the elongation of transcription had been identified in euryarchaeal and was referred to as Eta (euryarchael termination activity) [13]. Transcription termination factors have long since been found in both eukaryote and prokaryote species prior to identifying them in archaea [13]. Studies show that Eta likely functions by binding upstream of the stalled elongation complex, ultimately resulting in RNAP-DNA dissociation [13].

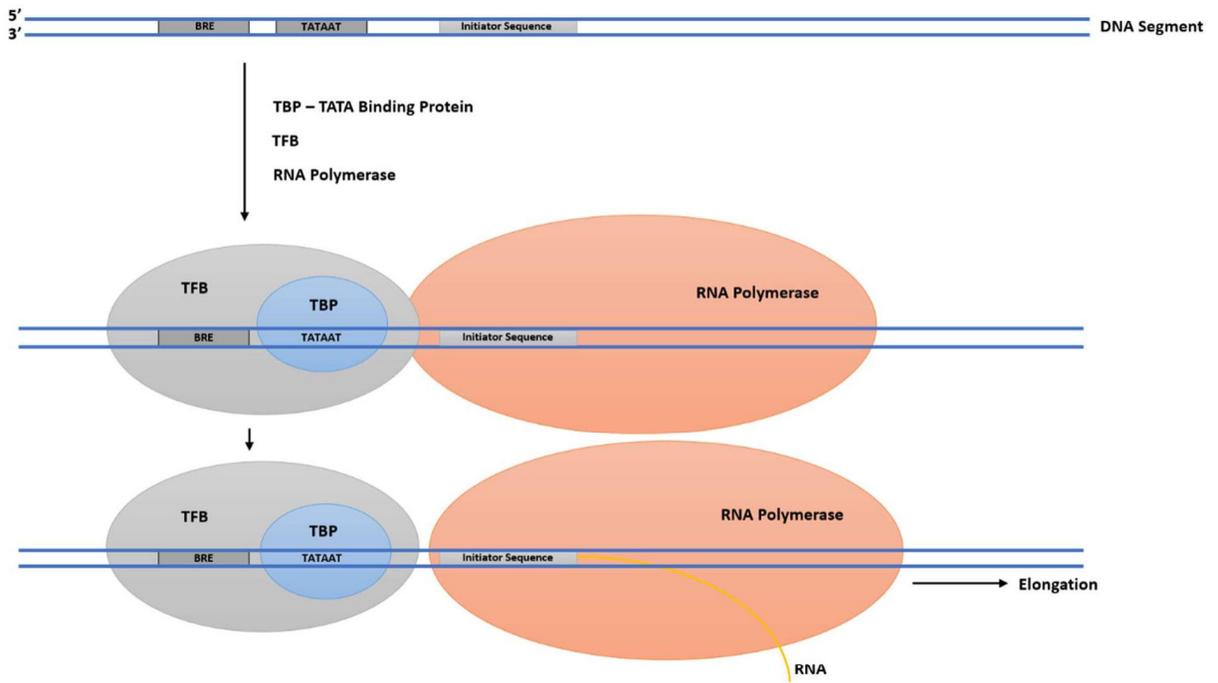


Figure 1. Archaeal transcription initiation adapted from [14].

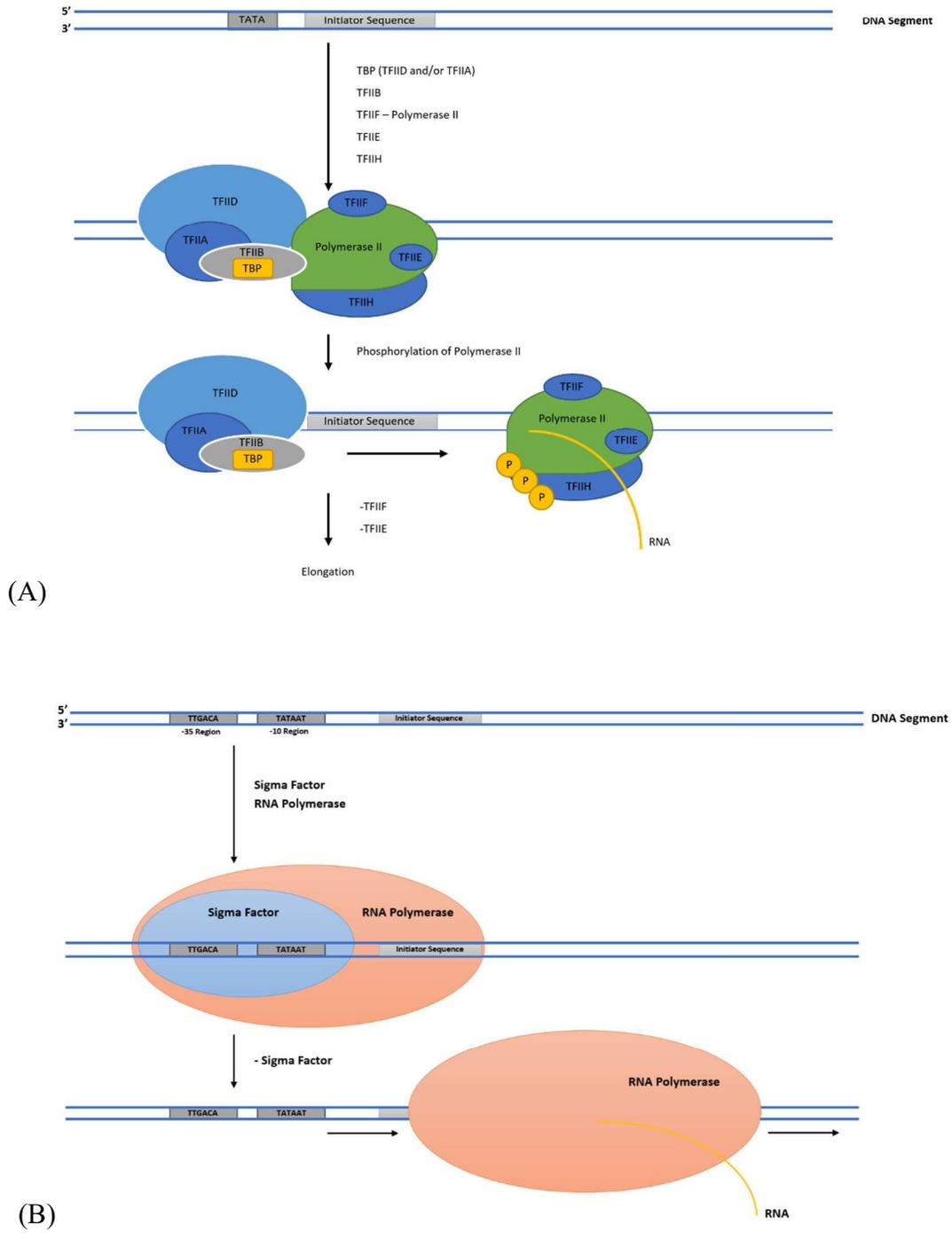


Figure 2. (A) Eukaryotic transcription initiation adapted from [15]. (B) Prokaryotic transcription initiation adapted from [16].

### **1.2.1 Post-Transcriptional Modifications**

Complete regulation of genomic transcription transcends simply repressing or activating transcription initiation, elongation, or even termination. There are examples of post-transcriptional modifications that have been found in archaeal species, some act to stabilize transcripts while others function as indicators for degradation. Across all domains, the function seems to vary most between bacterial and eukaryal transcripts with archaea again demonstrating a dichotic resemblance of both [7-9].

### **1.2.2 Polyadenylation**

Polyadenylation is an important regulatory process that is seen across all domains of life. In eukaryotes, post-transcriptional addition of a polyadenine tail increases the stability of mRNA transcripts, ensures proper translation, and functions in transcript transport [17, 18]. In bacteria, this modification is the initiating factor of the RNA decay pathway [18]. In searching the archaeal domain this post-transcriptional modification is shown to occur in hyperthermophilic archaea, which tend to inhabit severely high temperatures environments [18]. Further correlational analysis shows RNA polyadenylation happens only in archaeal species that encode an exosome [18]. From this finding, research has found the addition of these tails is the product of the archaeal exosome, which produces them in a heteropolymeric fashion where the other nucleotides are added in addition to adenosines [18]. This then makes them poly(A)-rich tails instead of simply a polyadenylation tail, as seen in eukaryotes [18]. This unusual product is also seen in bacterial species where the PNPase is the acting nucleotide-adding protein, and it also adds the 3' tail in a heteropolymeric fashion [18].

Further investigation found that the archaeal exosome consists of homology resembling both eukaryotes and bacteria, which could explain the similarity in nucleotide addition [18]. This

however also raises the question of function with these archaeal transcript tails, and whether they resemble a bacterial or eukaryal method of function. Research behind the function of this poly(A)-rich tail is still rather elusive, however results suggest that it may also be a marker for degradation as seen in bacterial species [18]. These poly(A)-rich tails have since been found in other hyperthermophilic archaea that encode an exosome, which further validates the probability of an exosome-mediated addition [18]. As for why halophilic archaea do not encode an exosome, and as a result do not produce polyadenylation products on their transcripts, still remains unknown [18].

Another post-transcriptional modification, found in both eukaryotes [19] and bacteria [20], recognized in archaeal mRNA recently is methylation [21]. In eukaryotic systems this modification is well known to have roles involving DNA structure mediation, however little is known about its function on RNA [21]. A more recent study done in 2013, found that archaeal mRNAs had several detected 5-methylcytidine ( $m^5C$ ) modification sites, which were found to be sequence-guided modifications [21]. Most known methylations in eukarya and archaea are known to be directed by C/D sRNPs, which constitutes an aFib protein component thought to be the associative methyltransferase element [22]. These methylations are known to occur frequently within ribosomal ribonucleic acids (rRNA) and transfer ribonucleic acids (tRNA) and the function of mRNA methylation is still elusive [21]. A more recently studied C/D sRNPs mediated methylation found in mRNA is the  $N^6$ -methylation of adenine [23]. More specifically, it was found that methylation of a particular transcript adenine results in the unsuccessful binding of a L7Ae family protein [23]. L7Ae proteins are known to help with translational regulation in archaea where they bind a kink-turn structure motif of the RNA species, resulting the regulation of gene expression [24].

### 1.2.3 Interactions with sRNAs

This brings in another crucial and newly discovered mode of possible mRNA regulation, which is the binding and interaction of sRNAs. The first of these to be discovered in the archaeal domain are the small guide RNAs [25]. In both eukaryotes and bacteria it has been identified that a number of small non-coding RNAs require additional proteins from the Sm/Lsm family to moderate interactions with other RNA molecules [25]. So far, these proteins have been identified in *Haloferax volcanii* (*H. volcanii*), but their actual function has not yet been identified [25]. The most highly expressed sRNAs found in archaea are *cis*-asRNAs, *trans*-encoded sRNAs, and *cis* sense RNAs [26]. Of the sRNAs that have been found in *H. volcanii* about 67% of them overlap with the coding sequence of mRNA [26]. This demonstrates the importance of better understanding these differing RNA in different organisms, which have previously shown to have significant regulatory roles in gene expression in other domains.

The first sRNA targeted gene that was found in archaea was MM2441-MM2442 identified in *Methanosarcina mazei* (*M. mazei*) [27, 28]. The authors suggested that the extended 5' UTRs found in *M. mazei* showed an increased importance of regulation in this region of the mRNA [28], however another project found that translation regulation was inhibited in the case that either the 5' or 3' UTR were absent. This was suggested to show a duality in proper function in regulating mRNA [29].

As the discovery of the archaeal domain is still relatively novel in relation to studies of the other two domains, the studies regarding archaeal mRNA regulation are in their infancy. So, it can be conclusively said that little is known about the actual regulation of archaeal mRNA transcripts. However, there are certain regulatory aspects present in archaea that have been previously studied in bacterial and eukaryotic organisms are known [7-9]. With this, there exists

a strong dichotic relationship between bacterial and eukaryal features within archaeal species. With this realization, one can at least speculate functional purposes of homologous findings; however, more research needs to be done to provide a clearer picture of these regulatory mechanisms.

### **1.3 Archaeal Translation**

Research into the translation machinery of the archaeal domain again showed high similarity to eukaryotic ribosomal components with bacterial features, such as SD (Shine-Dalgarno) ribosome binding sequences [30]. More specifically toeprinting and sequencing analysis have shown that ribosomal recruitment in archaea seems to mirror bacteria species, where mRNA/ribosome interaction is mediated by a 5' Shine-Dalgarno sequence [31]. This method easily explains the translation initiation of archaeal mono- and polycistronic transcripts that contain an SD motif, but *Halobacterium salinarum* transcripts with excised SD motifs have shown translation recovery in the case of SD sequence mutations [32]. Similar findings are seen in other studies with *S. solfataricus* [33, 34]. This observation along with the finding that over 50% of mRNA archaeal transcripts are leaderless [35] illustrates the importance in identifying the likely alternative pathway taken in translation initiation with no canonical components.

#### **1.3.1 Leaderless Transcripts**

The definition of a true leaderless transcript seems to vary slightly between studies, but seem to be consistent in defining them containing less than 40 nucleotides prior to the protein encoded region, and no initiating SD sequence [32]. Unlike bacteria, archaea have both monocistronic and polycistronic transcripts, and studies show that in *S. solfataricus* both types are found in a leaderless form [31]. Polycistronic versions demonstrate a structure where the initial encoded gene is leaderless, while subsequent regions usually contain a SD sequence.

A version of this sequence setup has been seen in the gas vesicle operon of *gvpACNO*, where a SD sequence is not present for *gvpA* [32]. Based on this the consensus remains that a scanning method is not employed by archaea, but that ribosomal recruitment is either coordinated by SD binding or some other method of short 5' terminus binding [36].

### **1.3.2 Leaderless Transcript Translation**

Very little is known about translation for these leaderless transcripts, but some of the studies shown here are demonstrating prospective factors involved in ribosomal docking with little interaction or signaling of mRNA prior to translation. Fluorescence studies in combination with density gradient analysis of mRNA and ribosome interactions revealed that in the event of halting translation a little over half of these leaderless transcripts were still associated with the 30S ribosomal unit [31]. Further toeprinting analysis revealed that a initiator tRNA (tRNA<sub>i</sub>) in this binary interaction is likely significant in effective binding of the 30S unit to mRNA [31], illustrating the importance of proper initiating tRNA docking and fidelity of start codon recognition in translation initiation. These results only determined that the 30S subunit was able to bind the leaderless transcript without any other factors, but no indication was given as to how the 50S subunit was then recruited to undergo translation of the transcript [31]. This mechanism is very similar to the canonical method seen in eukaryotic systems. The exception is that in eukaryotic systems the 40S unit is coordinated with the mRNA by a series of initiation factors in a scanning mechanism where the tRNA is incorporated after the start codon is found [37].

Another possible pathway of translation initiation of leaderless mRNA is showcased through filter binding assays showing that leaderless transcripts can be translated by non-dissociated 70S ribosomes in bacteria species [38]. It's also recognized, through ribosome profiling at various temperatures, that this non-dissociated ribosome-mediated translation was

more efficient in higher temperatures where the prevalence of these non-dissociated monosomes, including ribosomal units and necessary cofactors for translation, were seen to be higher [38]. This not only provides a possible inaugural mechanism of leaderless translation, but also sheds light on a likely evolutionary advantage, especially for archaeon thermophiles. Similar findings are found with mammalian translation, in which 80S ribosomes are able to perform direct translation on the condition that Met-tRNA is bound to the 80S ribosomal RNA unit [39], again demonstrating the importance of accurate initiation tRNA docking. A study done in *E. coli* also discovered through binding assays that the bacterial 70S ribosome have a 10-fold increase in affinity for the initiation codon when associated with the fMet-tRNA<sub>i</sub> [40], again eluding to a possible regulation point in this process.

With this seemingly important role of accurate tRNA docking, it becomes important to investigate roles of initiation factors that may prove essential to efficient leaderless translation. Looking further into the factors employed by the archaeal translation system we first look at aIF2, which is the homolog of the eukaryotic translation initiation factor 2 (eIF2) [41]. These units have been known to function in the proper alignment of the methionylated initiator tRNA in the initiation complex of the ribosome in eukarya and archaea [41]. No information is available regarding the expression of aIF2 in the event of leaderless translation, but this additional information along with the finding that leaderless translation via non-dissociated ribosomes require that the initiator tRNA is be properly bound [39, 40], brings to question the possible role aIF2 could play in the mediation of leaderless translation in archaea through Met-tRNA<sub>i</sub> and non-dissociated ribosome coordination.

Studies in *E. coli* also suggest that initiation of leaderless translation is in part decided by the amount of cellular IF2 [40]. Bacterial IF2 also functions in recruiting the fMet-tRNA<sub>i</sub> into

the P site, and as seen previously in eukarya and archaea by eIF2 and e/aIF2 [42]. Although these factors all function in a similar manner their structural consistency has a significant degree of variability between the domains [42]. The confirmed homologs of the bacterial IF2 unit in eukaryotic and archaeal systems are eIF5B and aIF2/2B respectively [42-44]. Surprisingly, studies looking at the aIF2/5B in archaeal leaderless translation have found increased expression of this factor correlates with increased translation of both leaderless and canonical prokaryotic and eukaryotic leadered transcripts [41, 42]. Other studies also looking into this factor have found through luciferase activity that the archaeal form (aIF2/5B) is able to partially replace the loss of function of the eukaryotic factor (eIF5B) [45]. This again demonstrates the conserved function of these factors, and illustrates a likely part involved in the universal translation of leaderless mRNA.

### **1.3.3 5' Capping in mRNA Transcripts**

The capping of the 5' end of transcripts has long been known to function in the stability of the transcript during transport and translation [46]. A more recent study looked at the importance of the 5' terminal structure of mRNA and rRNA binding of leaderless mRNA [47]. The study found through toeprinting that the absence of a 5' phosphate affected accurate ribosome binding [47]. In this same study they found through filter-binding assays that 5' phosphate leaderless mRNA/ribosome interactions have a dissociation rate 4.5 times lower than 5' hydroxyl leaderless mRNA [47]. This finding could both demonstrate the structure around a working leaderless translation setup, and maybe even a method of translation regulation of leaderless transcripts.

Continuing with possible regulatory methods of leaderless translation, a study looking at the haloarchaea heatshock factor (hsp70) mRNA 5' UTR region found that shortening or

deleting this 5' terminus reduced efficient translation. This finding is contrary to the finding found by Sartorius-Neef and colleagues where they found translation function was recovered in the event that the 5' UTR region was excised [32]. There is still need for further research to provide a clearer picture of the extent 5' capping is involved in regulation. These findings, however, may suggest that leaderless translation is in part transcript selective as well, and that leaderless translation occurs only on certain transcripts, which suggests some still unknown factor is responsible for another layer of gene regulation.

#### **1.4 Archaeal Genome and Proteome Adaptations**

Archaea are known for their intense ability to inhabit extreme environments [48], and halophiles specifically have demonstrated the capacity to maintain cellular function in extreme hypersaline environments [2]. It is suggested that these organisms are either simply resistant to these extreme internal and external conditions, or maybe their cellular machinery reaches optimal activity under these hypersaline conditions [2]. The more current question however remains how they maintain this osmotic regulation at an optimal level, and what are the specific mechanisms behind it. *H. marismortui* is the haloarchaea utilized in this study to provide more information regarding this regulation. Various haloarchaea species are found in a range of high ionic stress environments [49]. Sequestration of cations to the intracellular membrane is how these organisms maintain an ionic concentration equivalent to or higher than the surrounding environment [49].

This unusual ability for halophiles to withstand high salinity conditions has been of interest to several scientists who theorize possible adaptations allowing for continued efficient enzyme function. Key possible effects that are a consideration for proteins include solubility, stability, and the overall conformation of the proteins, which encompasses function as well.

Since any halophilic protein is likely subjected to conditions that could influence each of these, research has looked at possible features that could circumvent the lack of function in what would be considered extreme situations.

A key characteristic found to be unique not only to halophilic species but is found in all archaea, is their highly acidic genome [50]. It is hypothesized this adaptation can help prevent protein precipitation by allowing more charged surface amino acids to compete for water or bind cations to maintain a sufficient hydration shell [50]. These presumptions have been verified by the observation of lower intracellular water mobility likely due to increased attraction to protein surfaces aiding in developing an adequate hydration shell [51]. In conjunction with this hydration shell hypothesis, the lack of serine in the proteome is suggested to also be due to its inability to compete for water interaction with inorganic salts [50].

It is found that some species will also synthesize organic molecules simply to ensure osmotic pressure and prevent extracellular salt entry [50]. In producing these molecules, the concentration gradient across the plasma membrane decreases, which can then ensure no water is lost to the surrounding environment due to passive transport [50].

Other halophilic species have shown to instead import high concentrations of these extracellular inorganic salts, again inferring that protein adaptation is necessary in this water-depleted environment [49, 50]. This influx of intracellular salts could alter the electrostatic interactions between amino acids at both the tertiary and the quaternary protein levels, which can affect proper folding and evidently enzyme function [50]. In an environment with high salt which causes competition for protein surface interactions both internally and externally, it poses the question of protein flexibility due to environmental changes. Research suggests a decreased amount of hydrophobic residues within archaeal proteins, especially larger aromatic groups,

could promote more peptide charge repulsion in the protein core [50]. This could allow for significantly more movement, because of the less rigid hydrophobic core mediated internally [50].

Evidence also shows that some halophilic proteins rely on significant salt concentrations to promote folding [50, 51]. The halophile *Halobacterium salinarum* NRC-1 (*H. salinarum* NRC-1) was found to not only rely on high salt levels for protein folding, but also for increased protein stability by circumventing thermal denaturation [50]. It is also believed that halophilic extracellular proteins must be folded prior to extracellular transportation to ensure they remain soluble during transport [52], which would require a relatively consistent salinity level intracellularly to mimic extracellular conditions.

Lastly, research shows that overall halophilicity can be increased simply through highly hydrophobic peptide insertions [50]. This is observed in the serinyl-tRNA synthetase of *H. marismortui* where amino acid mediated hydrophobicity is believed to improve enzyme flexibility [50]. With this understanding of various speculated and observed adaptations in halophilic archaea proteins, it alludes to the delicacy and possible necessity of this solute balance in protein flexibility and function.

### **1.5 RCK transporters HKT/Trk/Ktr as salt conductors**

For haloarchaea species to maintain an ionic gradient they require the ability to sequester ions. This transportation of compounds is mediated through membrane protein transport systems. Membrane transport systems for various organic and inorganic compounds is a large area of study for all three domains of life. One of the superfamilies of transporters is the RCK (Regulator of Conductance of K<sup>+</sup>) transporters found in all three domains of life. In prokaryotic organisms all have RCK transporters for osmoregulation, pH homeostasis, regulation of turgor

pressure and membrane potential [53]. RCK superfamily includes the well-studied HKT/Trk/Ktr transport systems is further umbrellaed under the broader class of well-known ABC transporters [4, 5].

ABC transporters are a broad well-researched group of membrane transporters that have a high affinity for their substrates and function in uptake, export, and osmoregulation [52]. The typical structure of these transporters includes two integral membrane proteins and two cytoplasmic ATPases that promote substrate movement through hydrolysis of ATP [52] (Figure 3). This transport family is also famously connected to severe diseases including cystic fibrosis, cases of multidrug resistance seen in cancer, and are part of the global problem of antibiotic resistance [52]. The basic structural features of these transport systems remain homologous across domains. For example, these ATPases, share three motifs: Walker A and B, ABC signature, and sequence LSGGQ [52]. In archaea this superfamily continues to demonstrate a crucial role in ion transport and osmotic regulation in a variety of environmental conditions [52]

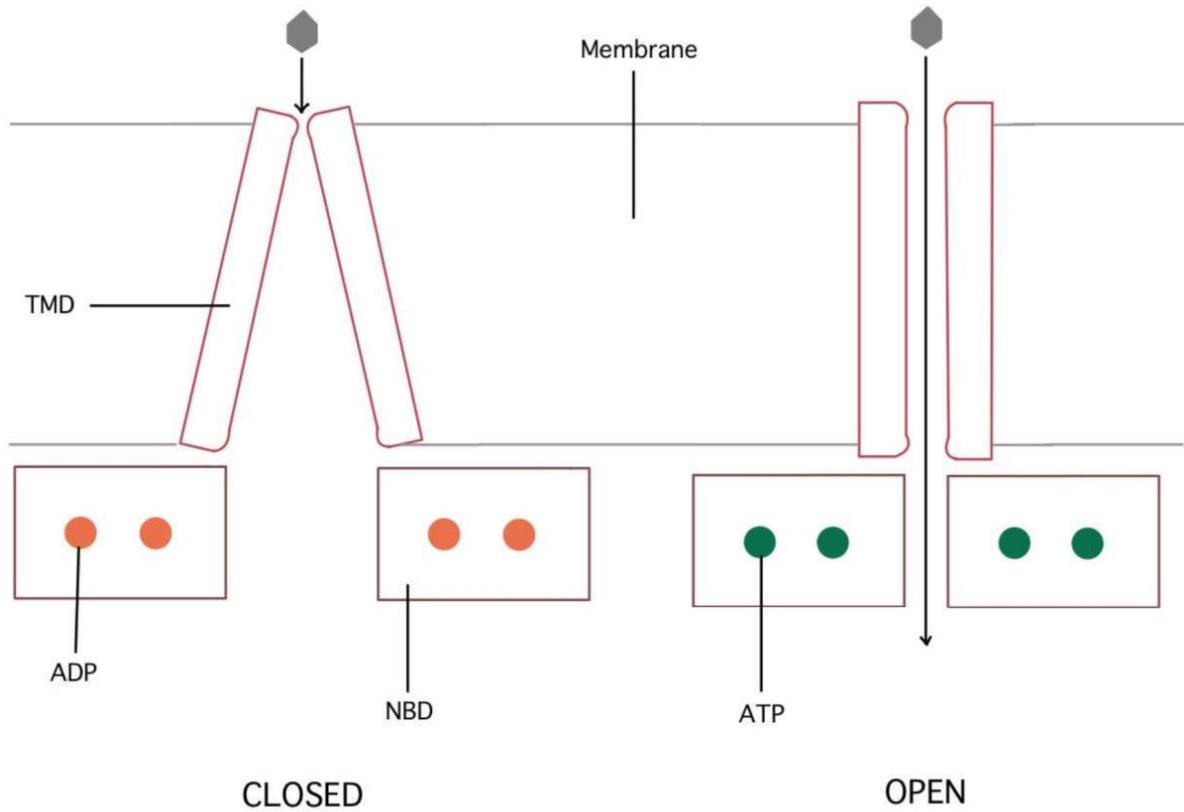


Figure 3. Canonical ABC transporter architecture including two transmembrane domains (TMD) and two nucleotide-binding domains (NBD) that function together in membrane transport of the molecule shown (grey hexagon) through the hydrolysis of ATP. Adapted from [54].

The Trk system, is evolutionarily related to the Ktr system and found both in bacteria and archaea [55] (Table 1), and has been studied in relative depth with regards to the membrane channel responsible for the movement of ions [4, 55-58]. These transport systems all have a characteristic membrane protein with a homo-dimeric membrane domain and an octameric cytoplasmic domain, which is denoted KtrAB in the Ktr system [4]. This membrane channel complex consists of the KtrB dimeric membrane protein that interacts with a cytosolic octameric KtrA ring [4]. Similar findings are shown in the Trk system, in which the dimeric TrkH membrane domain seems to interact with an intracellular TrkA protein (Figure 4) [59].

Due to these systems resembling the canonical ABC transporters, there is some debate about nucleotide binding interaction is required for activation of the channel [60]. With homologous observation of several structural motifs it is hypothesized that the likely nucleotide binding partner is ATP [60]. Although, this statement has not been completely proven because the authors developed some hesitation at the end of the study [60], and ATP has also been alluded to be the binding nucleotide in other papers with no clear evidence again [55]. The hesitation that ATP was the binding nucleotide was due to the authors gathering their comparisons from both intact cells and isolated cells, and so they simply concluded the ATP had some regulatory role in the Trk system but did not conclude with which protein [60].

In the Trk complex studied to date, very little is known about how regulatory proteins function in the role of the Trk complex, but TrkE has shown to be required for the TrkH (membrane protein) function and shown to induce residual activity in TrkG (membrane protein) [60]. Both TrkH and TrkG are found to function in the attachment of the peripheral membrane protein TrkA, which is required for activity of the Trk systems in *E.coli* [60, 61] (Table 1). The product of *TrkE*, in *E.coli*, has also been shown to affect potassium uptake in general but to date

no evidence of which protein it specifically influences in order to produce this function within the system [62]. Research looking specifically at the function of these proteins in combination with TrkE has shown that TrkH system is fully dependent on TrkE, while TrkG system co-dependes on TrkG and TrkE [63]. These findings could be interpreted to suggest that TrkE functions in potassium uptake by some interaction with TrkH and TrkG, which then ensures the membrane protein TrkA remains attached to the membrane and is able to perform the necessary function for potassium transport to occur. Alternatively, TrkE is an interchangeable regulator with TrkA, and so only one or the other is required depending on level potassium transport activity required according to active environmental changes (e.g. changes in turgor pressure [62] or osmolarity [62]).

Table 1. Trk system proteins with proposed function, and analogs studied in the Gorrell Lab at UNBC. [64]; [4]; [60]; [65]; [54].

<i>E. coli</i> <i>gene/protein</i>	Functionally Related ABC Subunits	Related Ktr Subunits	Function	Analogous <i>Har</i> <i>mar.</i> gene(s) studied in Gorrell Lab
<i>TrkH</i> / Trk <sup>H</sup>	TMD (Transmembrane Domain)	KtrB	K <sup>+</sup> -translocating subunit	
<i>TrkG</i> / Trk <sup>G</sup>	TMD (Transmembrane Domain)	KtrB	K <sup>+</sup> -translocating subunit	
<i>TrkA</i> / TrkA	NBD (Nucleotide- Binding Domain)	KtrA	Peripheral membrane protein required for system activity	<i>TrkA1/TrkA2</i>
<i>SapD</i> / TrkE			Unknown - Studies give possible indications to involvement in regulation of ion transport	<i>OppD1/OppD2</i>
<i>TrkD</i> / TrkD	TMD (Transmembrane Domain)	KtrB	Low specificity potassium translocating subunit	

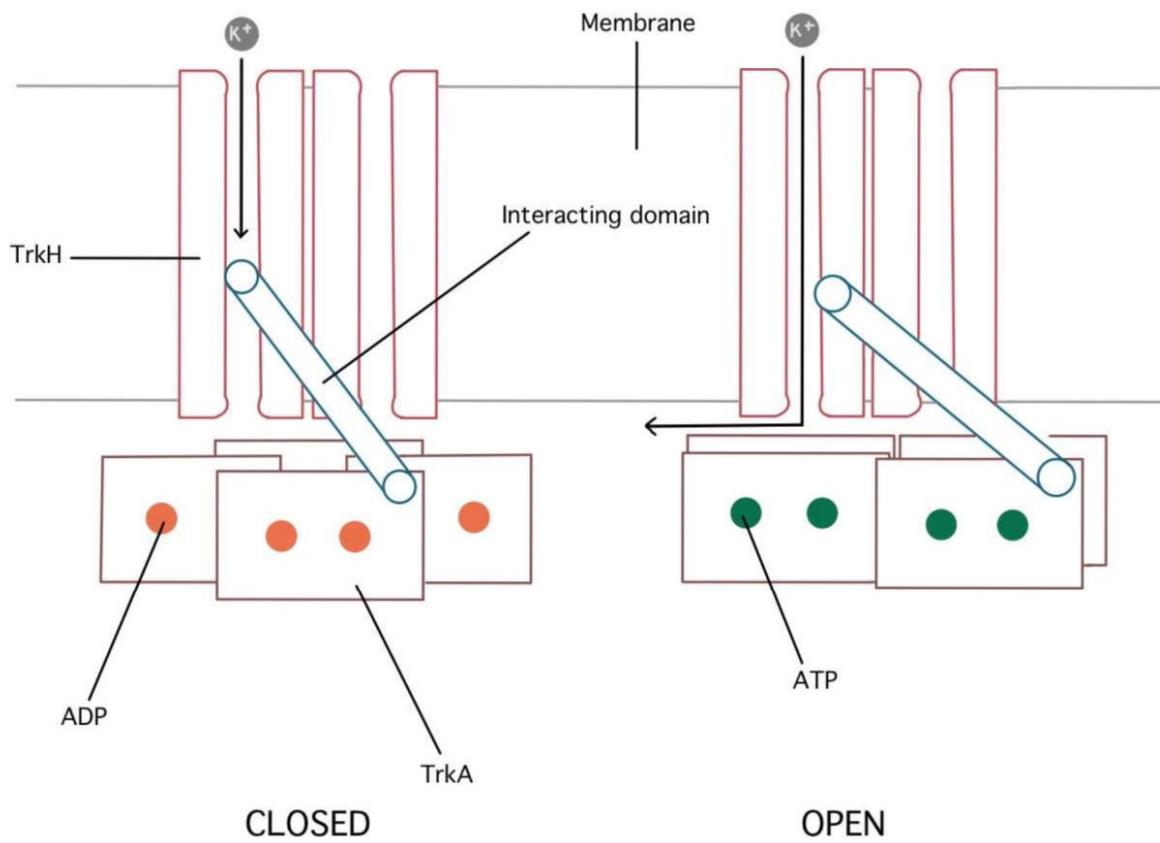


Figure 4. Proposed gating mechanism of transmembrane protein (TrkH) along with cytosolic membrane bound regulatory protein (TrkA) adapted from [59].

TrkE, which appears to be involved in the regulation of the Trk system in some aspect, is shown to be mapped in the *sapABCDF* (*sap*: sensitive to antimicrobial peptide) operon, where the gene is denoted as *SapD* [4]. This component of the operon has previously also shown to encode the ATP binding component for an ABC transporter in *E.coli* with an unknown function [4, 60]. Some studies suggest that one of the many roles of the protein complex encoded by the *sapABCDF* operon is protamine (small oligopeptides) resistance that is seen in *E.coli*, and it was suggested this could be due to the protein TrkE and its role in K<sup>+</sup> transport [60]. Other studies have shown that several virulent bacteria contain this operon and inactivation can result in lack of virulence [6]. Additionally, other proteins encoded within this operon have been found, some of which also include the TrkH and TrkG proteins [62].

Further research into this nuanced *sapABCDF* operon has shown further homological relationships suggesting it could likely function as an oligopeptide permease (Opp) [66]. Some of the findings include the observations that *E.coli SapD* and *SapF* show high similarity to *OppD* and *OppF*. Furthermore, it has also been found that each are significantly homologous such that *SapD* and *SapF* found in *Salmonella typhimurium* and *Bacillus subtilis* are 40% identical and 61% like *OppD* and *OppF* in *Salmonella* [66]. Additionally, the *SapD/SapF* have sequence identity, which is also seen in the *OppD/OppF* pair [66]. Although, *SapD* and *SapF* were significantly different in size and demonstrated a 28% sequence identity, the *OppD/OppF* pair displayed similar-sized proteins and higher sequence identity (41%) [66]. These high levels of sequence identity along with the overlap seen in the well-known Walker motifs suggest multiple functions with various system components [66].

### 1.5.1 *Haloarcula marismortui*

*H. marismortui* is an archaeon that optimally grows in high salinity conditions, where osmotic regulation and salt transportation becomes exceedingly key to its sequence. It is generally believed that the Trk system is a key factor in this [4, 5].

*H. marismortui* is the second halophile species to have its genome completely sequenced allowing observation and investigation of its unique features [3]. *H. marismortui* appears to have more functions than the first sequenced halophile *Haloarcula* NRC-1 and less nutritional requirements [3], which enables it to be better used as a model organism to investigate the cellular processes that allow it to adapt to its environment. This archaeon thrives under extreme salt conditions, and was first identified in the Dead Sea giving rise to its name [3]. It thrives in high salinity conditions with molarities between 3 and 5 and has been found to maintain intracellular total potassium molarities between 1.13 and 2.43 [67]. This sequestration of salts for osmotic balance, in *Har. marismortui* was originally thought to occur through passive transport of potassium across the membrane [68, 69]. However, with these high intercellular salt concentrations, it is unlikely that *Har. marismortui* can attain such high intracellular molarities simply through passive diffusion of potassium across its highly permeable membrane [70]. To endure these extreme salt conditions these organisms must maintain this transmembrane salt gradient by some aspect of their membrane morphology. With this understanding, *H. marismortui* must have specific mechanisms related to its membrane architecture that aid in maintaining balance in such a gradient.

*H. marismortui* has shown to encode two orthologs of *OppD*, which are denoted *OppD1* and *OppD2*. These two genes have been shown to locate areas on chromosome I several kilobases from each other. These organisms have in their genome what has been coined an

ecoparalog, which are paralogs that function in different environments [71]. This defining feature poses the question as to whether *OppD1* and *OppD2* in *H. marismortui* could function as ecoparalogs and could this explain the reason for the need for two locally separate but highly identical transcripts. This other *OppD* ortholog, and other Trk homologs are also shown to undergo differentiated expression in *H. marismortui* under different  $K^+$  concentrations [72].

## 1.6 Scope and Significance

Since its first discovery in the late 1970s, the archeal domain has been found in various locations ranging from extreme aquatic environments to the digestive tract system of mammals, including humans [73]. Species from each branch of the archaeal domain have been isolated and identified in the gut along with a few located in the oral cavity [74]. After identifying these organisms in the human microbiota along with the already known characteristics of archaea as a domain, recent concerns point towards stronger antimicrobial patterns seen from these organisms [75].

All discussions regarding the *OppD1* polypeptide stem from the research conducted on its presumed homolog *TrkE* in the *E. coli* K-12 system [60], and its possible involvement in potassium conductance via the Trk potassium transport system. This study aims to do an initial characterization of the *OppD1* protein from *H. marismortui* in relation to what is hypothesized in prior alignment and functional homological studies outlined in this introduction. Investigating the role of *OppD1* in ion regulation within the context of *H. marismortui* could uncover valuable insights into cellular homeostasis.

## 1.7 Objectives

These two main objectives are attained by utilizing genetic cloning targeting techniques along with an introduction to chromatographic and spectrometric techniques looking at structural and introductory functional features of the polypeptide:

The primary objective is to isolate, clone, express, and purify the target gene/protein responsible for *OppD1*/OppD1 within the archaeon *H. marismortui*.

The secondary objective is an in-depth analysis of the subsequent OppD1 protein's characteristics, with a focus on both its structural and functional aspects. This comprehensive examination is designed to unveil the protein's properties and its potential functions within the archaeal system.

The quest to unravel OppD1's enigmatic nature extends to its characteristic structural features. Detailed structural analyses could yield a deeper understanding of how OppD1 functions and fits into the broader biological landscape of osmotic regulation and ion transport. Upon the successful completion of these two fundamental objectives, the door opens to a realm of potential future hypotheses and investigations.

## **Chapter Two: Cloning and Expression of OppD1**

## 2.1 Chapter Objectives

The primary objective of this chapter is to detail the process of the identification of the *TrkE* homolog *OppD1* in *H. marismortui*, cloning *OppD1* into an expression vector, and subsequent protein expression from this recombinant clone.

## 2.2 Methodology

### 2.2.1 Expression Vector Purification and Quantification

*E. coli* strains DH5 $\alpha$  and NEB turbo competent cells were employed for vector transformation. These bacterial strains were cultured on Luria-Bertani (LB) agar and LB plates.

The cloning procedures used the *pUC19* and *pET21b* vectors. The plasmids underwent transformation and production in the previously mentioned *E. coli* strains. Following that, the plasmids were extracted and purified using the NEB Monarch Plasmid Miniprep Kit according to its protocol.

After purification, the concentration was quantified using a NanoDrop ND-1000 spectrophotometer.

### 2.2.2 Nested PCR Amplification of *OppD1* – *OppF* Gene Segment

*H. marismortui*'s genomic DNA underwent isolation from the cells acquired from Dr. Gorrell's lab at UNBC through osmotic lysis. Due to *H. marismortui*'s typical high salinity environment cell lysis is easily performed simply by the addition of water, which causes an influx of solution into the cells causing them to burst because of the hypotonic conditions. The *OppD1* (seq ID: 12042) and *OppF* (seq ID: 12043) sequences (rrnAC2042 and rrnAC2043 subsequently) from *H. marismortui* were determined from the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/data/view/AY596297>). End sequence primers essential for amplifying *OppD1-OppF* and internal primers (Table 1) for recombinant clone verification were

constructed with and ordered from Integrated DNA Technologies (Coralville, Iowa) after assessing for potential secondary structurization and dimerization via their OligoAnalyzer Tool adjusted for utilized oligo concentrations. Some considerations made during primer design included the incorporation of appropriate restriction sites (end primers); attempting to decrease GC content to control temperature variations between end primers while still maintaining a sufficient length; ensuring product overlap with sequencing primers by choosing appropriate placement along with regulating GC content (internal primers); removing the possibility of non-specific binding through species-specific NCBI BLAST [76] sequence searches. These primers, listed in Table 2 as *OppD1\_F\_NdeI* and *OppF\_R\_XhoI*, played a critical role. Polymerase Chain Reactions (PCR) followed Phusion® High-Fidelity DNA Polymerase-based methods as defined by New England Biolabs (NEB). Each reaction encompassed a final volume of 20 ul, including 4 ul 5X Phusion GC Buffer, 0.4 ul of 10 mM dNTPs, 1 ul of designated primer diluted to a concentration of 10 uM, 0.2 ul of the template, and 0.2 ul Phusion DNA Polymerase. All *OppD1-OppF* (Figure 5) nested PCR products were visualized through 1% agarose gel electrophoresis using ethidium bromide and UV illumination (FluroChem Q). Promega's 1kb DNA Ladder (CAT #: G5711) served as a size reference for all DNA work (Appendix – Figure 1).

Table 2. Primers for *OppD1* amplification from *H. marismortui* via PCR. Utilized for ligation and verification of recombinant clones.

Primer name	Oligonucleotide sequence	T <sub>m</sub> (°C)
<i>OppD1F-NdeI</i>	5'-AAAAAACATATGCGCCTGCTCGAAG-3'	54.4
<i>OppD1R-XhoI</i>	5'-AAAAAACTCGAGGCCATCCTGCTGGCC-3'	54.3
<i>OppF_R_XhoI</i>	5'-AAACTCGAGGCGACTACTGGCCC-3'	49.7
<i>OppD1_R_internal1</i>	5'-AGCCTCTGCCCTCTGTTCG-3'	51.5
<i>OppD1_F_internal1</i>	5'-ATCCGGCACCACCTCGACT-3'	59.7
<i>OppD1_R_internal2</i>	5'-CTCATCAGCCCGACGGTGT-3'	58.6
<i>OppD1_F_internal2</i>	5'-TGCTATCCCGCGTATCGGC-3'	58.9
<i>T7For</i>	5'-TAATACGACTCACTATAGGG-3'	46.8
<i>T7Rev</i>	5'-TATGCTAGTTATTGCTCAG-3'	45.8

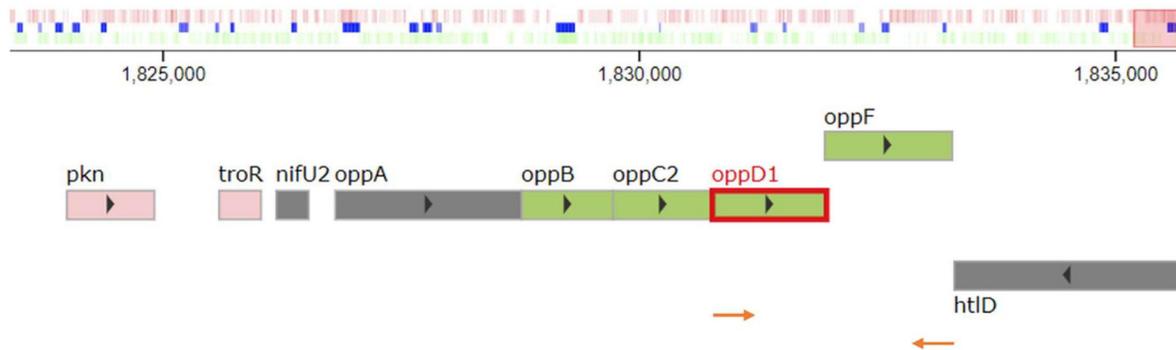


Figure 5. The relative position of *oppD1* and *oppF* on Chromosome I of the haloarchaea *Haloarcula marismortui*. Orange arrows represent forward and reverse primers, *OppD1F-NdeI* and *OppF\_R\_XhoI* (Table 2.) used to produce transcript *OppD1-OppF*. Annotated genes not IDs elsewhere include: *pkn*, *troR*, *nifU2*, and *htID*.

### 2.2.3 PCR Amplification of *OppDI* from *OppDI-OppF* Gene Segment

The gene of interest, *OppDI* (rrnAC2042), had previously designed primers originating from Dr. Gorrell's lab at UNBC (Table 2). The forward and reverse primers included *NdeI* and *XhoI* restriction sites. An additional primer, detailed in Table 2, was designed to further facilitate amplification and ligation through resources from Integrated DNA Technologies. PCR reactions were prepared in 10 µl total volumes, initially utilizing Promega's GoTaq Green Master Mix. This mix contained 10 µl of 2X GoTaq Green MasterMix, 0.5 µM of each primer, and the remaining volume was adjusted with nuclease-free water.

Subsequently, PCR reactions remained at a 10 µl total volume but featured 1 µl of 5X Phusion GC Buffer (NEB), 1mM of dNTPs (NEB), 0.05 µM of each primer, 1 µl of DMSO, 1 unit of Phusion DNA polymerase (NEB), with the remaining volume supplemented with nuclease-free water. For colony PCR reactions, volumes were consistent with the setup, and transformation colonies selected from an LB plate were used to inoculate these reaction volumes.

PCR for amplifying the *OppDI* gene involved temperature cycles commencing with an initial denaturation at 98°C for 2 min. Subsequently, 7 cycles followed with denaturation at 98°C for 10 sec, primary annealing at 60.1°C for 30 sec, and extension at 72°C for 1 min and 30 sec. In continuation, another 20 cycles proceeded with denaturation at 98°C for 10 sec, secondary annealing at 66°C for 30 sec, and extension at 72°C for 1 min and 30 sec, concluding with a final extension at 72°C for 4 min. PCR reactions were performed in a Bio-Rad DNA Engine Dyad Peltier Thermal Cycler.

After visualizing the *OppDI* gene, it was purified from the agarose gel via the E.N.Z.A. Gel Extraction Kit (Omega Bio-tek) following kit defined protocol. The concentration of purified product was quantified utilizing the Nanodrop ND-1000 Spectrophotometer. Furthermore, trials

were executed where *OppDI* was directly extracted and purified from the completed PCR reaction using the PureLink PCR Purification Kit (Invitrogen). A small sample was loaded and visualized through gel electrophoresis and UV transillumination (FluroChem Q). Promega's 1kb DNA Ladder served as a reference (Appendix– Figure 1).

Subsequently, PCR reactions remained at a 10 µl total volume but featured 1 µl of 5X Phusion GC Buffer (NEB), 1mM of dNTPs (NEB), 0.05 µM of each primer, 1 µl of DMSO, 1 unit of Phusion DNA polymerase (NEB), with the remaining volume supplemented with nuclease-free water. For colony PCR reactions, volumes were consistent with the setup, and transformation colonies selected from an LB plate were used to inoculate these reaction volumes.

PCR for amplifying the *OppDI* gene involved temperature cycles commencing with an initial denaturation at 98°C for 2 min. Subsequently, 7 cycles followed with denaturation at 98°C for 10 sec, primary annealing at 60.1°C for 30 sec, and extension at 72°C for 1 min and 30 sec. In continuation, another 20 cycles proceeded with denaturation at 98°C for 10 sec, secondary annealing at 66°C for 30 sec, and extension at 72°C for 1 min and 30 sec, concluding with a final extension at 72°C for 4 min. PCR reactions were performed in a Bio-Rad DNA Engine Dyad Peltier Thermal Cycler.

After visualizing the *OppDI* gene, it was purified from the agarose gel via the E.N.Z.A. Gel Extraction Kit (Omega Bio-tek) following kit defined protocol. The concentration of purified product was quantified utilizing the Nanodrop ND-1000 Spectrophotometer. Furthermore, trials were executed where *OppDI* was directly extracted and purified from the completed PCR reaction using the PureLink PCR Purification Kit (Invitrogen). A small sample was loaded and visualized through gel electrophoresis and UV transillumination (FluroChem Q). Promega's 1kb DNA Ladder served as a reference (Appendix– Figure 1).

#### **2.2.4 Cloning *OppD1* into the Expression Vector *pET21b* to Generate the Recombinant Vector *pET21b-OppD1***

Following the purification of both the *OppD1* gene and vector *pET21b*, the construction of recombinant *pET21b* plasmids containing the gene of interest was undertaken. The objective was to construct the recombinant vector *pET21b-OppD1*. Benefits of using *pET21b(+)* included that it already contains the necessary sites for expression through lactose-induced methods and it contains the necessary gene that confers ampicillin resistance, which allows for isolated growth of recombinant clones. Double digestion of both the insert (*OppD1*) and the vector (*pET21b*) was carried out using *NdeI* and *XhoI* (NEB). The vectors were treated with either Shrimp Alkaline Phosphatase (NEB) or Antarctic Phosphatase (NEB) as per the NEB protocol before setting up the ligation process.

To facilitate recombinant construction, the linearized vector and the gene were mixed with T4 ligase enzyme, following the protocol provided by New England Biolabs for their product. An attempt to expedite and enhance the efficiency of ligation involved using the Instant Sticky-end Ligase Master Mix (NEB) for a trial of the *pET21b-OppD1* ligation. Furthermore, additional attempts at recombinant vector construction were made by adjusting the molar ratio of vector to insert from 1:3 to both 1:5 and 1:10. The total reaction volume remained consistent by modifying the volume of nuclease-free water.

The *E. coli* DH5 $\alpha$  competent cells underwent transformation via the heat-shock method, using the ligation reaction volume, and were cultured on LB-ampicillin plates. When performing the transformation with *pUC19*, the LB-ampicillin plates were prepared for blue-white screening to simplify the identification of potential successes. To verify successful cloning, PCR analysis was performed using *OppD1*-specific primers (end primers and internal primers) (Table 2).

Additionally, the presumed recombinant vectors were digested using a combination of restriction enzymes and visualized through gel electrophoresis with ethidium bromide. The results were compared with Promega's 1kb DNA ladder using UV transillumination (FluroChem Q) (Appendix – Figure 1). Final verification of a likely recombinant clone was performed by DNA sequencing by the UNBC Genetics Lab.

### **2.2.5 OppD1 Expression**

Alpha-lactose 1% w/v @ 4 hours induced OppD1 expression was accomplished in the Sigma- Aldrich Rosetta™(DE3) Competent Cells to indicate optimal protein expression levels. These cells were transformed directly with the *pET21b-OppD1* recombinant clone after the ligation was verified. A 12% SDS-PAGE gel used to visualize OppD1 expression over 4 hours after lactose induced expression (Figure 10).

### **2.2.6 Western Blotting**

The SDS-PAGE proteins were transferred horizontally to a polyvinylidene fluoride (PVDF) membrane using a TE77 Semi-dry System at 60V for 1 hour. The NEB Coloured Pre-Stained Standard Broad Range ladder served as a reference (Appendix – Figure 2) for visualizing protein bands in SDS-PAGE gels and was compatible with Western Blotting techniques to confirm complete transfer to the membrane. The membrane was blocked with Blocking Buffer (TBS; 1% non-fat dry milk (casein)); washed with TBS 3 times (3X); incubated with the primary anti-his antibody (Bio-Rad) 1:10,000 in Blocking Buffer; washed 3X; incubated with the secondary corresponding anti-rabbit horseradish peroxidase (HRP) antibody (Bio-Rad) 1:5,000; and washed 3X. After incubation with 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) to activate the conjugated HRP until the subsequent protein bands were visible as purple.

## **2.2.7 OppD1 Purification through FPLC Ion-Exchange Chromatography and Subsequent Concentration of Purified Samples**

For purifying His-tagged OppD1 from the remaining cell lysate, a Cytiva (formerly GE) HisTrap™ Fast Flow 5 mL nickel ion exchange column was employed. The purification process involved loading 5 mL of lysate through the nickel crosslinked agarose column and washing it with Buffer A (50mM NaH<sub>2</sub>PO<sub>4</sub>; 300mM NaCl, pH 7.5) for 12 column volumes (CV). To elute the desired protein, Buffer B (50mM NaH<sub>2</sub>PO<sub>4</sub>; 300mM NaCl; 125mM imidazole; pH 7.5) was introduced in an 11% step-up gradient for 3 CV, followed by a gradual increase to 100% Buffer B over 2 CV. Subsequently, the column underwent a cleaning phase with an additional 5 CV of 100% Buffer B before being re-equilibrated with Buffer A. All purification steps using the HisTrap™ column were conducted at a flow rate of 1.0 mL/min.

OppD1 fractions were then concentrated using Amicon 10 kDa Millipore EMD centrifuge filter concentrators and analyzed for relative concentration using the Bradford Dye Assay.

## **2.3 Results**

### **2.3.1 *OppD1* Amplification**

In Figure 6, the visualized insert size resulting from the PCR reaction to amplify the *OppD1* gene from the genomic DNA of *H. marismortui* is shown by gel electrophoresis. Following the gel visualization, *OppD1* was purified from the gel as described in the methods. However, considering ligation efficiency, the PCR product was later directly taken and purified for *OppD1* in hopes to increase the overall concentration.

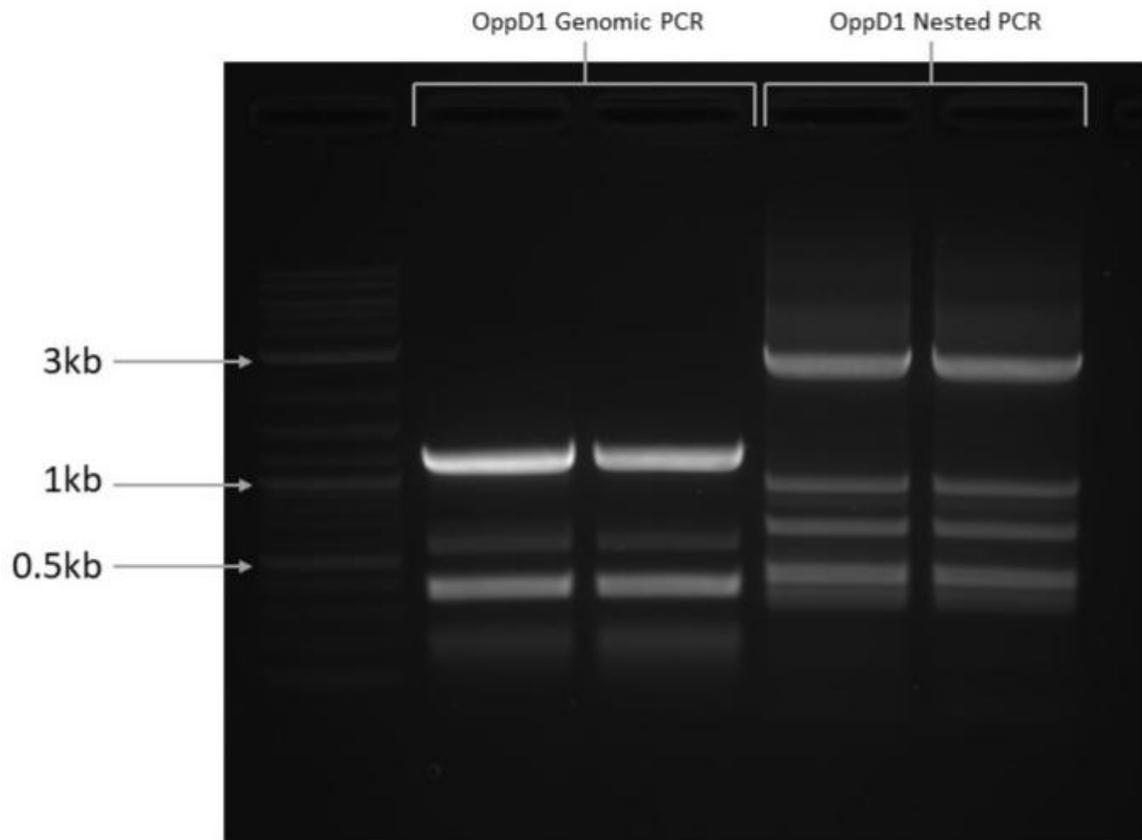


Figure 6. *OppD1* PCR Amplification from *H. marismortui* shown on a 1% agarose gel. Amplification of *OppD1* was attempted here with a *H. marismortui* genomic sample and a previously nested product of *OppD1-OppF*. This agarose gel shows complete amplification of *OppD1* from the genomic samples with a band sitting at ~1.1kb. No *OppD1* amplification can be seen in PCR reactions with the nested product. PCR reactions were set up with Phusion GC setup and the *NdeI* and *XhoI* primers. Promega's 1kb DNA Ladder served as a reference (Appendix – Figure 1.).

The vector pUC19 was chosen for recombinant structure construction due to its suitability for visualizing potential successful ligations. Initial attempts at constructing the recombinant vector yielded no growth of transformed colonies or no further growth of observed positive colonies through colony PCR. To rule out instrument errors, thorough equipment review was conducted, and new media and plates were prepared. Subsequent ligation attempts resulted in adequate colony growth for positive vector verification. However, although multiple trials showed positive colonies through colony PCR (Figure 7), further verification through restriction endonuclease activity failed.

A potential factor not previously considered was self-ligation of the vector after the restriction endonuclease reaction. To mitigate this possibility, the 5' end of the digested plasmid was dephosphorylated. After implementing these protocol adjustments and varying insert-to-vector ratios, a consistent pattern of positive ligation identification during colony PCR emerged. Nevertheless, additional verification through digestion reactions still did not confirm successful ligation.

Upon closer examination of New England Biolabs 2X GoTaq Polymerase master mix, it was discovered that the Taq polymerase commonly adds extra adenine bases during extension. Given that the *OppD1* and *pUC19* ligation setup involved blunt ends, this is likely to have contributed to the ligation difficulties. To address this, *OppD1* amplifications from *H. marismortui* genomic DNA were conducted using the Phusion Polymerase setup, which ensures blunt end extension.

Despite these method modifications, the results continued to show a consistent pattern of positive colonies in colony PCR, with no indication of successful ligation during restriction endonuclease activity.

### 2.3.2 *OppD1*-pET21b(+) Recombinant Vector Construction

After multiple attempts to construct the *OppD1*-*pUC19* vector, for the purposes of increasing the concentration of *OppD1*, a change in focus led to a direct ligation into the expression vector *pET21b*. In this sticky-end ligation, both the vector and the insert were digested with *XhoI* and *NdeI*. The choice of polymerase for *OppD1* amplification was not a critical factor since the PCR-produced gene was digested on both ends following amplification. Similar to the *pUC19* vector, the *pET21b* vector was dephosphorylated after the restriction endonuclease reaction to prevent self-ligation in future attempts.

Comparing the *pUC19* recombinant vector construction with the *pET21b* recombinant vector, similar results were observed at each step. Visual indications of positive ligation colonies were seen through colony PCR and electrophoresis with the respective primers (Figure 7). However, further confirmation of successful ligation through restriction endonuclease activity showed no recombinant ligated product. To rule out potential enzyme malfunction, several restriction enzyme combinations were tested and all produced similar results.

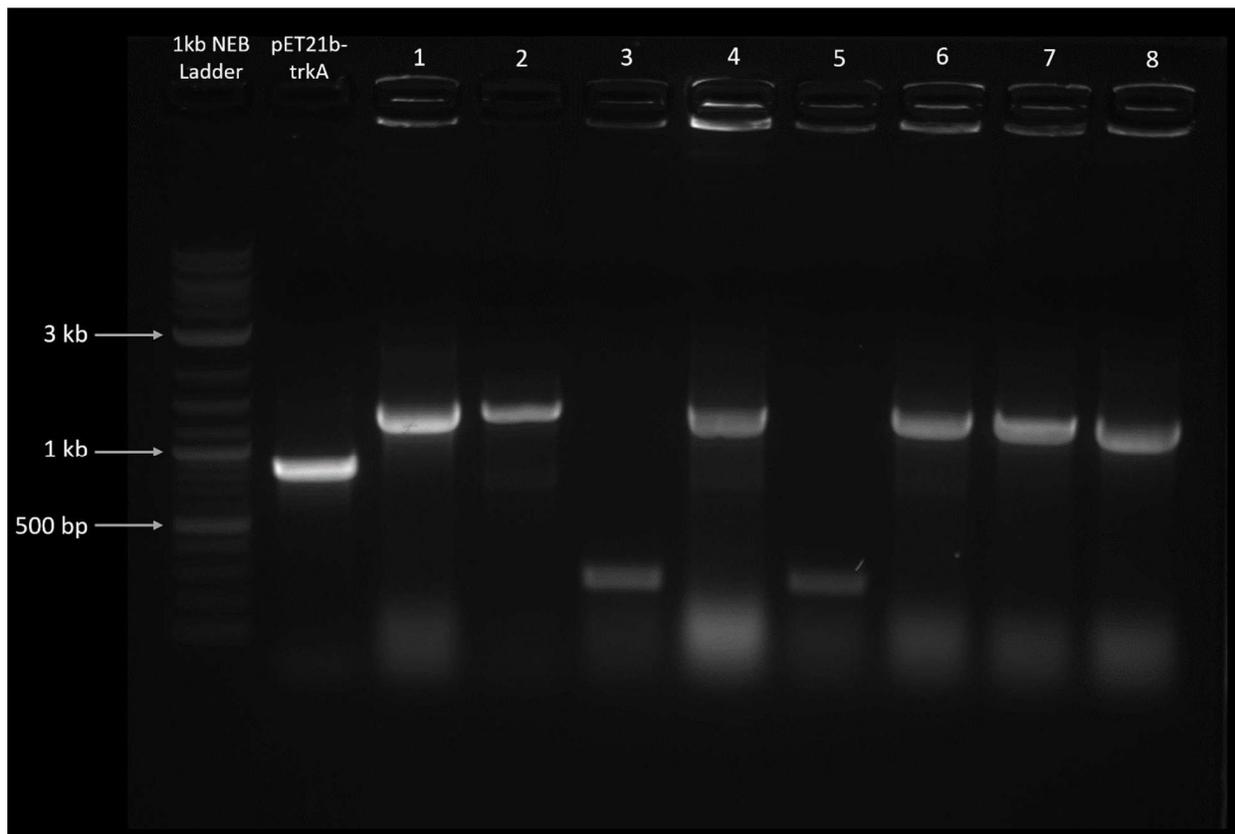


Figure 7. Colony PCR with the T7 and T7Rev primers of 8 *E.coli* colonies to determine successful cloning of *TrkE* homolog *OppD1* into *pET21b*. *pET21b-TrkA* ran as a positive control for successful primer binding and amplification. Lanes are numbered according to the colonies used for PCR verification. Promega's 1kb DNA Ladder served as a reference (Appendix – Figure 1.).

### 2.3.3 OppD1 Internal Primer Design and Utilization

After multiple attempts to ensure the complete ligation of OppD1, an additional PCR method was employed to confirm the correct order and complete sequence ligation of OppD1. Four internal primers were developed and optimized for the further amplification of OppD1 through nested PCR methods (Table 2.). This method of PCR was to troubleshoot the possibility of non-specific binding to homologous genes in *E. coli* during colony PCR. These internal primers also aimed to provide insights into the insert's orientation, a task typically achieved through restriction endonuclease activity. Additionally, this nested PCR was conducted using the common T7 primers to confirm the proper setup for expression. Once a successful recombinant clone was further verified (Figure 8), the plasmid was extracted and sent for sequencing (Figure 9), which served as the final confirmation of a successful clone and allowed the project to progress to expression and further analysis.

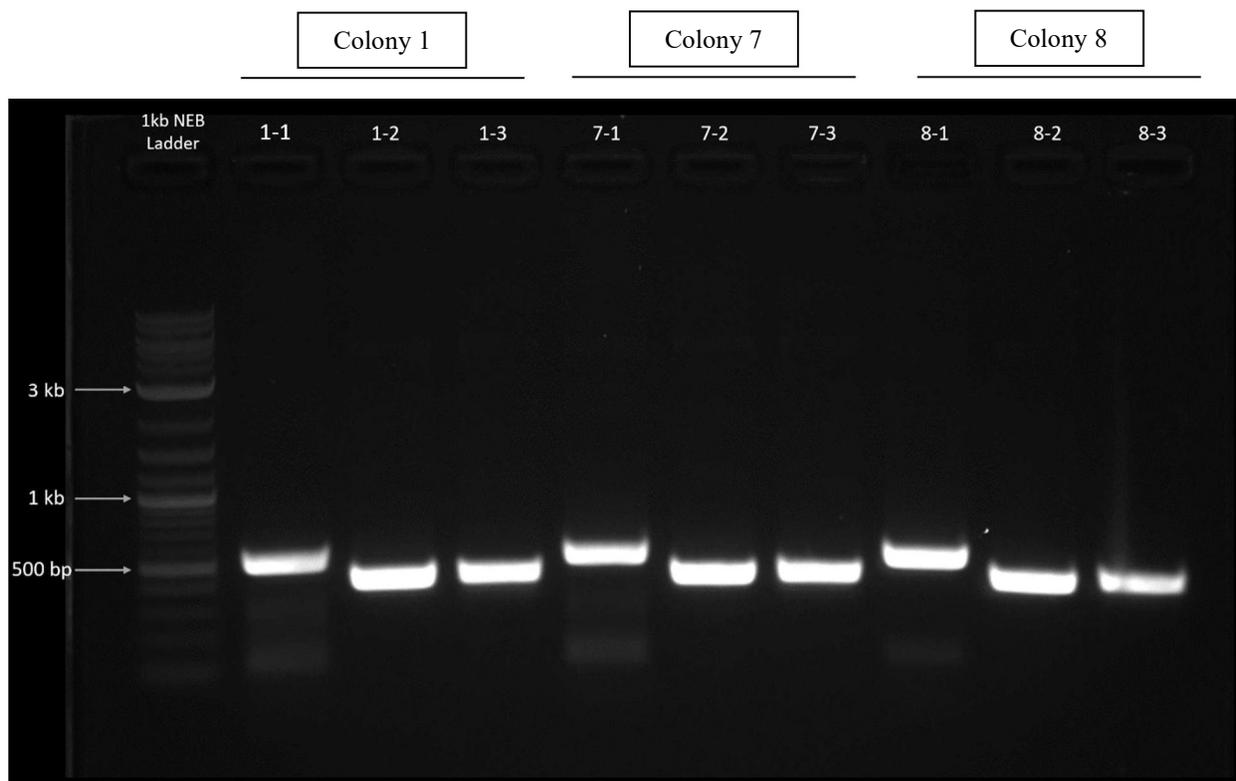


Figure 8. Verification of successful recombinant clones through colony PCR amplification with 3 combinations of the internal and T7 primers as noted (Table 2) (primer combinations are as follows: #-1: *T7For* and *OppD1\_R\_interal1*; #-2: *OppD1\_F\_interna11* and *OppD1\_R\_interal2*; #-3: *OppD1\_F\_interna2* and *T7Rev*). Colonies used here for further verification were initially chosen based on verification from previous figure (Figure 7). PCR reactions performed using the Phusion GC reaction setup with the incorporation of DMSO to promote binding of high GC content templates. Promega's 1kb DNA Ladder served as a reference (Appendix – Figure 1).

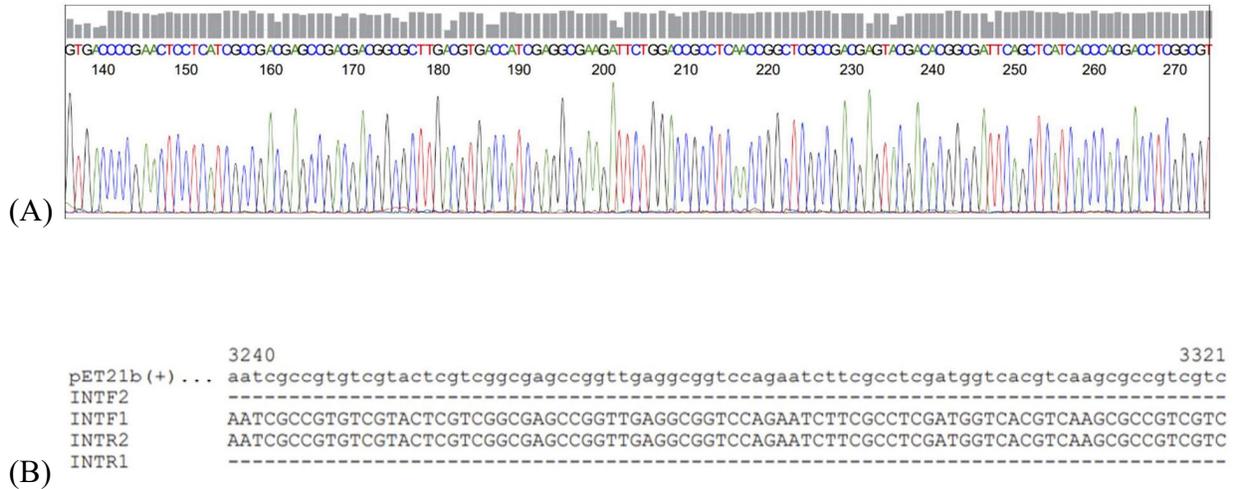


Figure 9. (A) A sample of the *OppDI\_F\_internal1* primer sequence of pET21b(+) indicating the presence of *OppDI* in the expression vector. Constructed with FinchTV. Remaining sequence data can be found in the appendix (Figure 5 – 8). (B) Segment of DNA alignment of the recombinant *pEt21b(+)-OppDI*. Shown are partial sequence alignments amplified from the internal forward primer 1 (*OppDI\_F\_internal1*) and the internal reverse primer 2 (*OppDI\_R\_internal2*). Full sequencing data can be found in the appendix (Figure 9).

### 2.3.3 OppD1 Expression and Isolation

Following successful verification through sequencing, the recombinant clone underwent transformation into the Sigma-Aldrich Rosetta™(DE3) competent expression cell line.

Expression was induced by lactose, and optimal expression levels were achieved at approximately 2 hours after induction (Figure 10). To further confirm these initial observations, Western Blotting was conducted, targeting the encoded His-Tag (Figure 11).

Affinity Chromatography (Figure 12) revealed a notable amount of protein elution during affinity binding to the column, with absorbance levels plateauing, suggesting no further protein release. After the addition of wash buffer to cleanse the column of bound components, absorbance increased, showing two distinct peaks. SDS-PAGE electrophoresis validated the substantial protein presence in the lysate and the flow-through (Figure 13), indicating that a significant portion of additional protein passed through during the affinity binding process. Electrophoresis also depicted proteins binding to the column washing through upon initial introduction of the washing buffer, with OppD1 visible in samples 24 through 27.

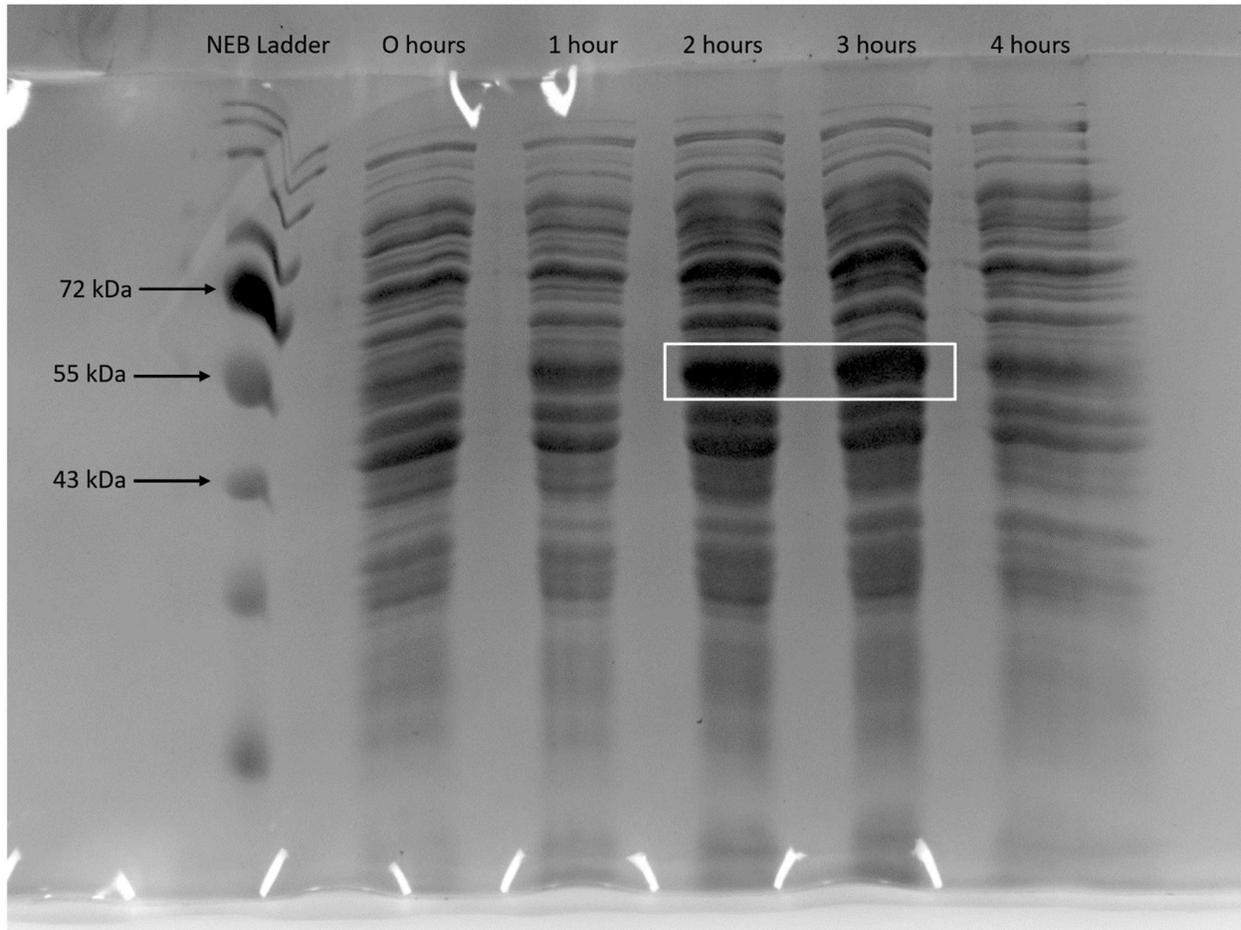


Figure 10. Expression analysis of OppD1 in the *E.coli* expression line Rosetta. Expression of the protein can be analyzed here through a 12% SDS-PAGE gel stained with coomassie blue stain.

NEB Coloured Pre-Stained Standard Broad Range ladder served as a reference (Appendix – Figure 2).

OppD1 expression highlighted.



Figure 11. Western Blot targeting His-Tag incorporated into OppD1 seen after developing with BCIP. NEB Coloured Pre-Stained Standard Broad Rangel ladder served as a reference (Appendix – Figure 2). OppD1 expression highlighted.

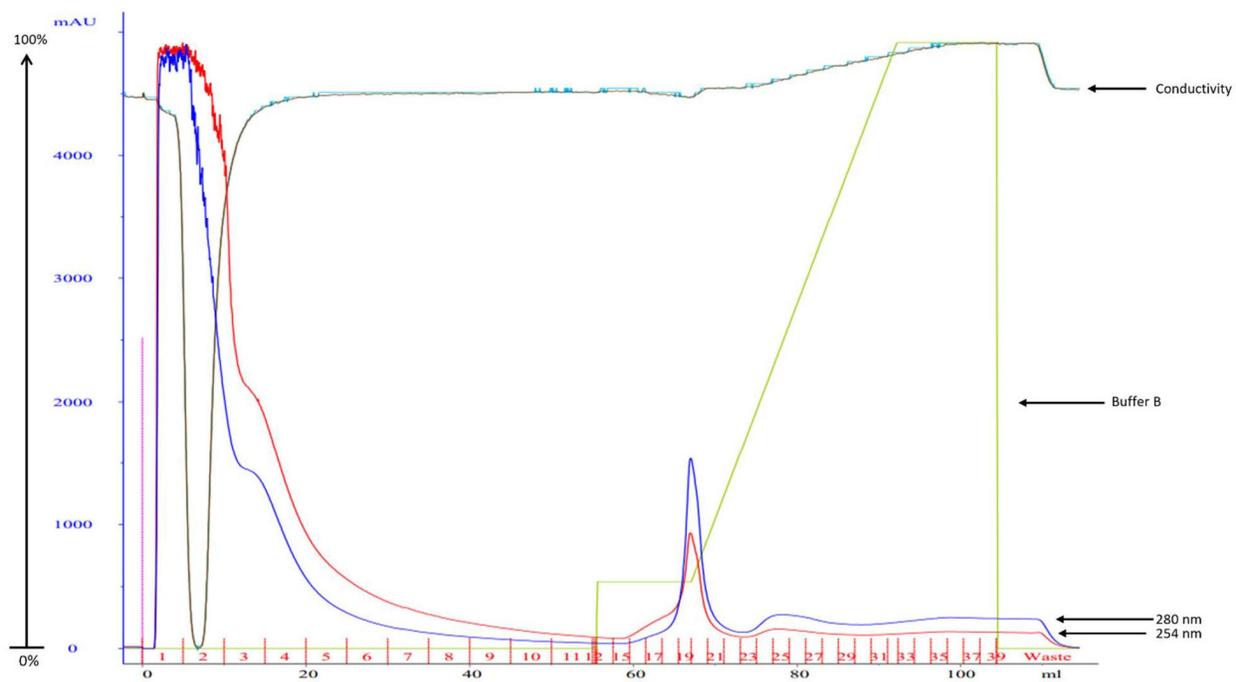


Figure 12. Chromatogram of NiNTA-column of OppD1. Red and dark blue curves indicate absorbance and 254nm and 280nm respectively. Black and light blue curves indicate conductivity during separation. Green curve indicates gradient of Buffer B %.

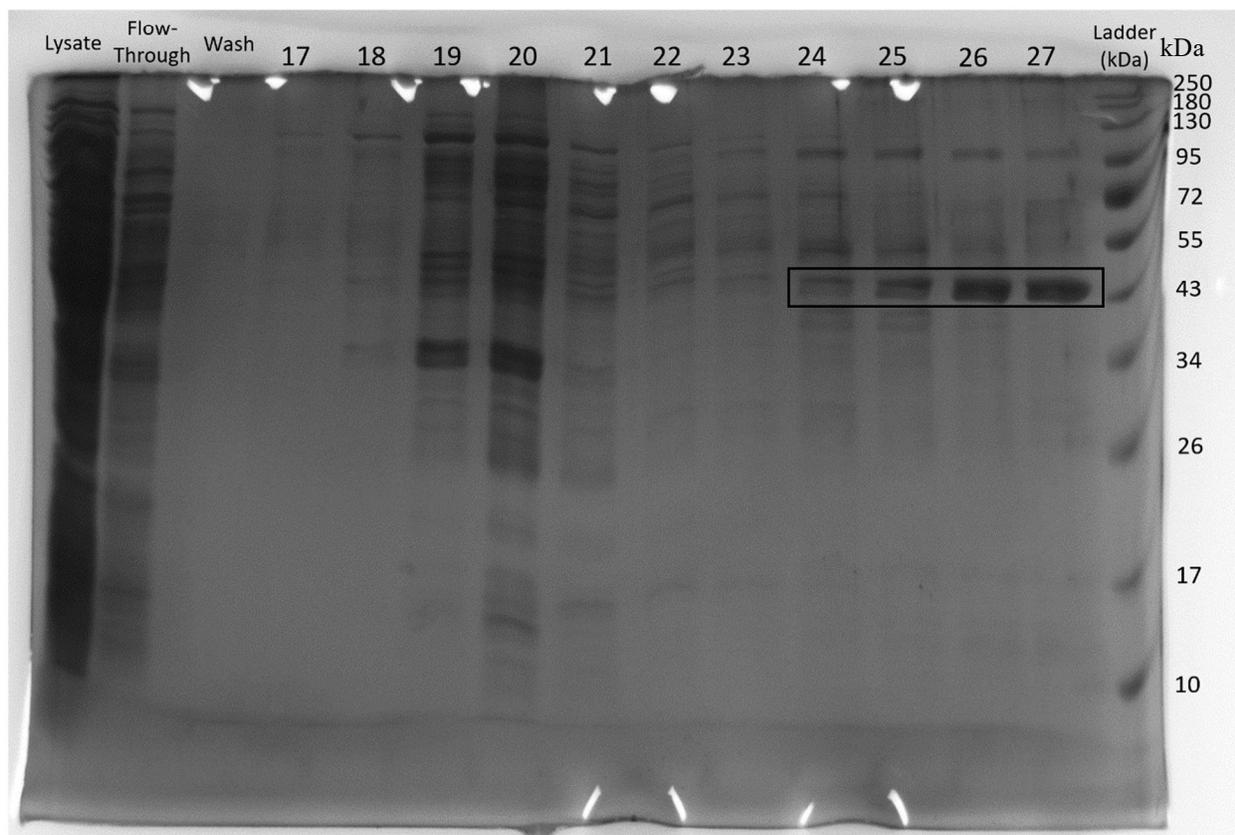


Figure 13. A 12% SDS-PAGE of the His-Tag Affinity Chromatography experiment shown. Samples indicating OppD1 presence can be seen in samples 24-27. NEB Coloured Pre-Stained Standard Broad Rangel ladder served as a reference (Appendix – Figure 2).

## 2.4 Discussion and Conclusion

In this chapter, the primary goal was to amplify and clone the *OppD1* gene from the genomic DNA of *H. marismortui*, followed by the expression and purification of the OppD1 protein. The results and observations obtained throughout the experiment revealed several key findings.

During the *OppD1* amplification phase, initial attempts to construct the recombinant vector *pUC19-OppD1* encountered difficulties, with no growth of transformed colonies or further development of positive colonies through colony PCR. To address these issues, various modifications were introduced, including dephosphorylation to prevent self-ligation [77] and the use of Phusion Polymerase for blunted *OppD1* amplification [78]. Despite these adjustments, the results consistently showed positive colonies in colony PCR by indicating correct sized amplicons, without successful confirmation of ligation through restriction endonuclease activity.

A change in focus led to the direct ligation of *OppD1* into the expression vector *pET21b* using sticky-end ligation. The results in this phase closely mirrored those of the *pUC19* vector construction, with positive ligation colonies observed. However, further verification through restriction endonuclease activity again did not confirm the successful ligation, despite attempts with multiple restriction enzyme combinations.

To address these ongoing challenges, an additional PCR method was employed, involving the use of internal primers to further confirm the correct order and complete sequence ligation of *OppD1*. Until now PCR methods indicated successful ligations, but restriction enzyme activity suggest otherwise. This would require further analysis of sequence similarity between *H. marismortui* and *E. coli* because there remains a possibility that the *OppD1* sequences in both species have high sequence identity at the 5' and 3' ends, which could produce

false positives when using *E. coli* [4, 66]. To attempt to mitigate this possibility internal primers were designed to provide insights into the orientation and size of the targeted *OppD1*, a task typically achieved through restriction endonuclease activity. The nested PCR also utilized common T7 primers to confirm the proper setup for expression. Successful verification of a recombinant clone finally allowed for sequencing, which served as the final confirmation of a successfully cloned *H. marismortui OppD1*.

The *OppD1* expression and isolation phase was accomplished, as the recombinant clone was transformed into the Rosetta expression cell line and induced with lactose. Optimal protein expression was achieved at approximately 2 hours, as confirmed by Figure 10. Additional validation was conducted through Western Blotting, targeting the encoded His-Tag (Figure 11).

The confirmation of *OppD1* binding the column during Affinity Exchange Chromatography targeting the incorporated His-tag are visually confirmed in Figure 13. This chromatographic technique successfully isolated *OppD1* from a plethora of cellular proteins in *E. coli*.

In summary, this study faced initial challenges in the cloning phases due to similarity in sequences even with primer optimizations, resulting in limited success in the construction of recombinant vectors. However, through further optimization and a shift in focus, the project ultimately achieved success in expressing and purifying *OppD1*. While challenges were encountered in the early stages, these findings provide valuable insights into the nuances of protein behavior and emphasize the importance of optimizing cloning procedures for successful recombinant protein expression and purification. Future studies of similar nature may consider working towards generating an accurate molar extinction coefficient by performing more various experiments. This would allow for a more precise quantitative analysis of protein samples prior

to characterization. Another area that needs more attention to possible interactions between restriction endonuclease activity and archaeal transcripts, with respect to further cloning attempts.

## **Chapter Three: Characterization of OppD1**

### **3.1 Chapter Objectives**

In this chapter the primary objective is the further characterization of OppD1 through Size Exclusion Chromatography and Circular Dichroism Spectroscopy in terms of protein size and secondary structure composition. Through this encompassed analysis hypotheses can be made to its overall function in *H. marismortui*, and more broadly ion transport in the domain of archaeal species.

### **3.2 Methodology**

#### **3.2.1 Size Exclusion Chromatography (Gel Filtration)**

To achieve further purification and quantify the size of OppD1, size exclusion chromatography (SEC; gel filtration) was conducted using the Cytiva HiLoad 16/60 Superdex 200 Prep grade 120 mL column, which has the capability to separate 10 kDa to 600 kDa (globular) proteins. This polishing procedure was carried out with a flow rate of 0.5 ml/min, utilizing the modified S200 buffer (50mM Tris; 0.5 M KCl; pH 7.5), spanning 1.4 column volumes (CV).

#### **3.2.2 Circular Dichroism Spectroscopy**

Characterizing OppD1 regarding its overall secondary structure composition involved employing Circular Dichroism techniques. The samples utilized in this procedure were maintained in the original S200 polishing buffer (50mM Tris; 0.5 M KCl; pH 7.5). Sample concentrations were estimated using Bradford Assay methods, and they were determined to be approximately 1 mg/ml. A cell with a 10 mm path length was utilized to scan the range of 200-450 nm, ensuring optimal absorbance with minimal buffer interference. Each scan consisted of three accumulations conducted at a speed of 200 nm/min. The analysis of resultant spectrum was

performed using the BeStSel (Beta Structure Selection) secondary structure determination and fold recognition bioinformatic program [79].

### **3.2.3 Bioinformatics**

To generate visual representations of OppD1's secondary structures, a partial alignment of OppD1 was computed in DeepView Swiss PDB Viewer [80] and comparisons were made using SapD as a reference. The crystal structure of SapD provided an appropriate alignment starting point because the gene, *SapD*, contains the sequence for the homologous TrkE protein. Additional structure generation and secondary checks were conducted with AlphaFold [81]. Energy minimizations were as defined in the programs with default parameters.

## **3.3 Results**

### **3.3.1 Chromatography Analysis**

The chromatogram from the HiLoad Superdex 16/60 200 Prep Grade GE Column (Figure 14) illustrates a successful gel filtration, effectively separating proteins extracted from *E.coli* following OppD1 expression based on size. The SDS-PAGE (Figure 15) of fractions reveals OppD1 eluted in sample 61. Comparing the elution volumes and the electrophoresis gel, it becomes evident that OppD1, in its probable monomeric state, has an approximate molecular weight of ~44,000 Daltons.

Further purification of OppD1 was achieved through additional gel filtration runs involving samples 59-62, as shown in Figure 16. No adjustments were made to the parameters involved in this chromatography technique. This chromatogram demonstrates improved separation, allowing the isolation and concentration of the peak around 60 ml for subsequent CD spectrum analysis.

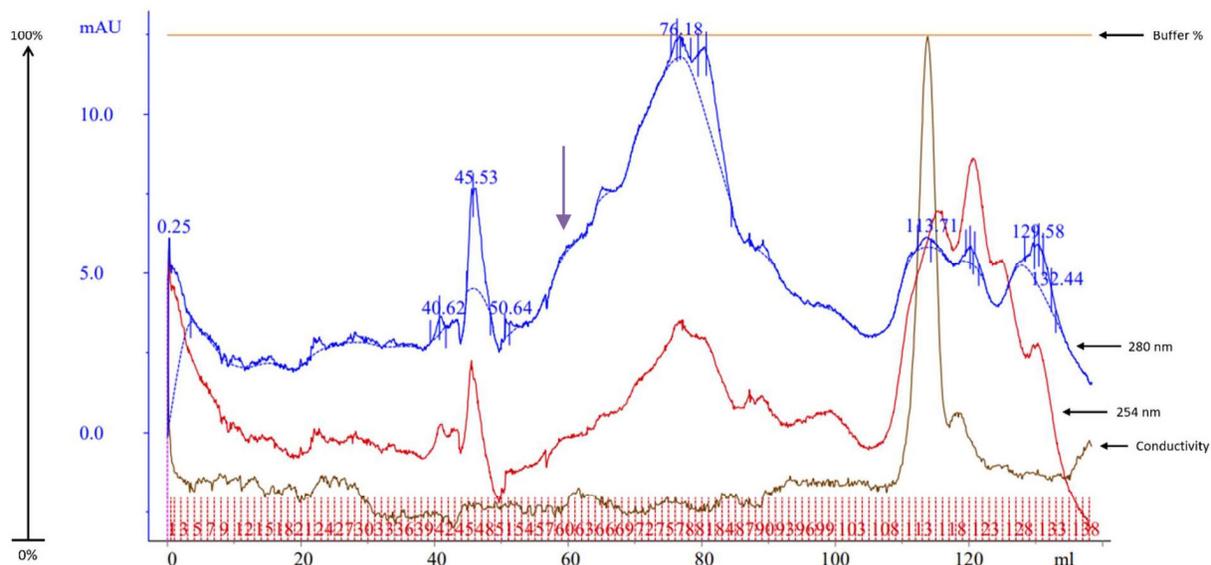


Figure 14. Sample 25 and 26 from the chromatogram on the previous page were concentrated and loaded onto the HiLoad Superdex 16/60 200 Prep Grade GE Column and ran over 1.4 column volumes with the S200 Buffer (50mM Tris; 0.5 M KCl; pH 7.5). Red and Blue indicate absorbance at wavelengths of 280 nm and 254 nm. Orange line gives indication of the concentration of Buffer B which is the S200 buffer used for size exclusion chromatography, and the brown line is conductivity. Numbers shown on peaks indicated retention time, which indicated elution of OppD1 (purple arrow) (See Figure 15 for comparison)

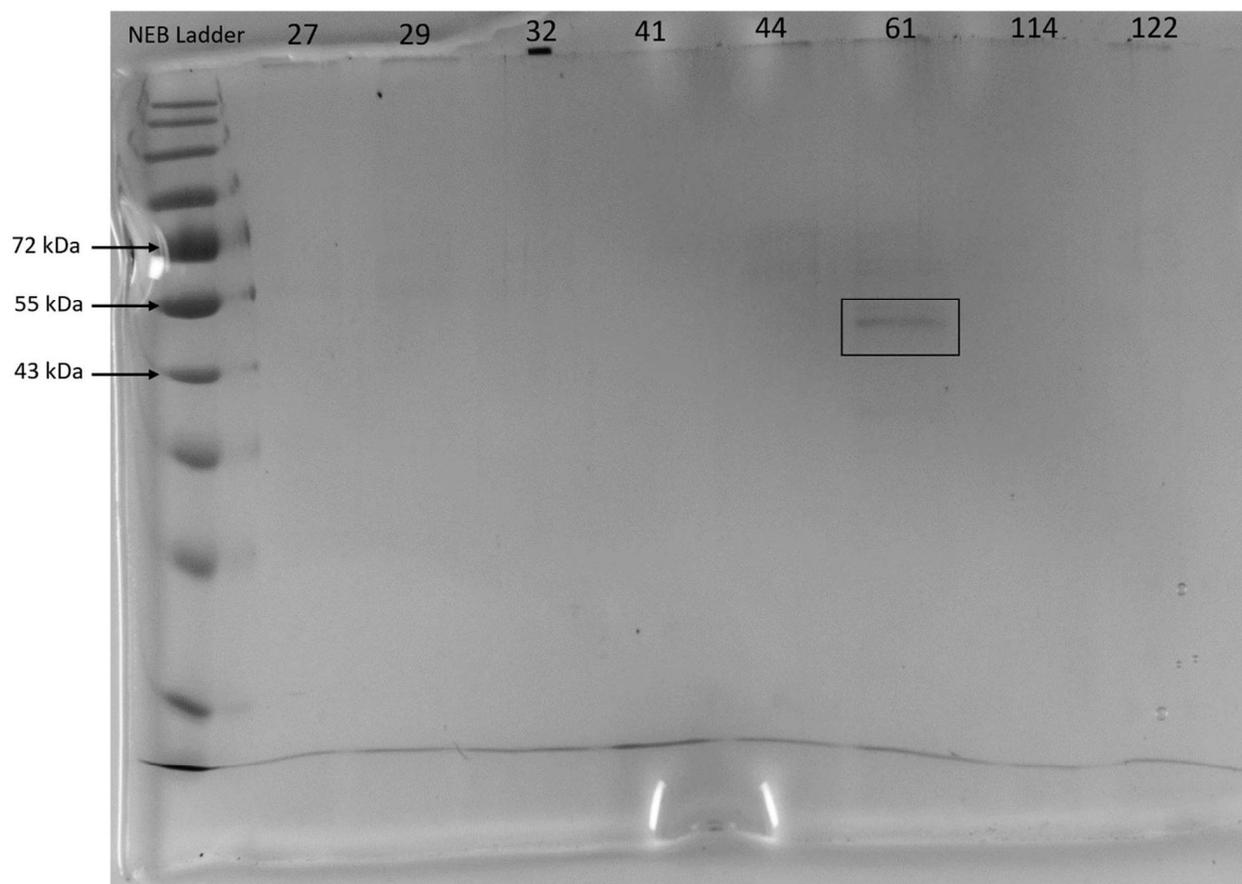


Figure 15. A 12% SDS-PAGE of HiLoad Superdex 16/60 200 Prep Grade GE Column elution samples from Figure 14. Electrophoresis shows band in elution volume 61 mL of correct size for OppD1. NEB Coloured Pre-Stained Standard Broad Rangel ladder served as a reference (Appendix – Figure 2).

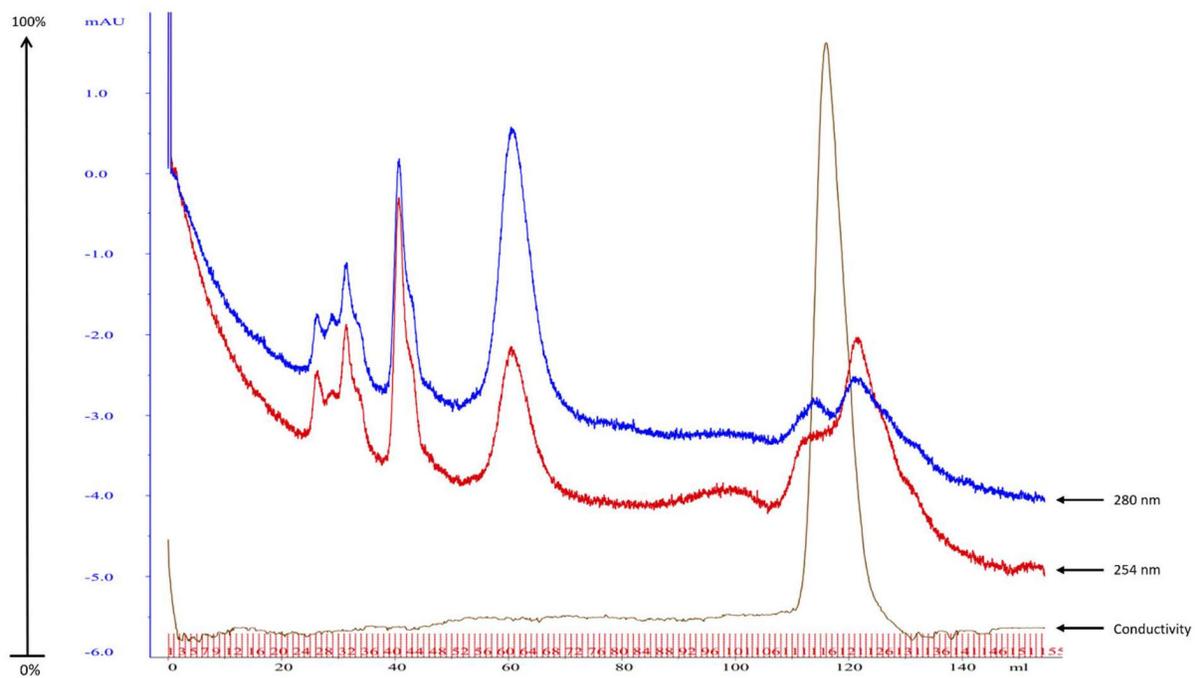


Figure 16. HiLoad Superdex 16/60 200 Prep Grade GE Column gel filtration run with S200 Buffer (300mM KCl; 50mM Tris; pH 7.5) and samples 59-62 from first SEC run with the HiLoad Superdex 16/60 200 Prep Grade GE Column seen in Figure 14. Buffer used was the S200 Buffer (50mM Tris; 0.5 M KCl; pH 7.5).

### 3.3.2 Circular Dichroism Analysis

The CD Spectrum experiments consistently reveal HT voltages below 700 V, which indicated reasonable buffer interference in absorption. These results were consistent even with the elevated salt content in the buffer, signifying an appropriate signal-to-noise ratio (S/N), as demonstrated in Figure 17. In this figure, the 3<sup>rd</sup> panel displays a reduced protein absorption level while still depicting absorption within the wavelength range of 210 to 240 nm. Preliminary analysis of proposed amino acid composition through Benchling (2023) ability to analyze and *in silico* translation, showed a significantly low amount of fluorescent amino acid composition. Based on these initial analyses, it is apparent that the provided CD spectrum for OppD1 absorption holds validity, as depicted in Figure 18.

An initial visual assessment of the spectrum in Figure 18 suggests the presence of significant alpha helical secondary structures, characterized by negative bands at 222 nm and 208 nm, as well as a positive band at 193 nm, aligning with the observations described by Greenfield in 2006 [82]. Further detailed secondary structure information is obtained through bioinformatic analysis using BeStSel (Beta Structure Selection), as displayed in Figure 19. This relatively novel bioinformatic program has the capability to predict a protein's folding at the topology and homology levels in alignment with the CATH classification system, which looks at protein class (C), architecture (A), topology (T), and homologous superfamily (H) [79].

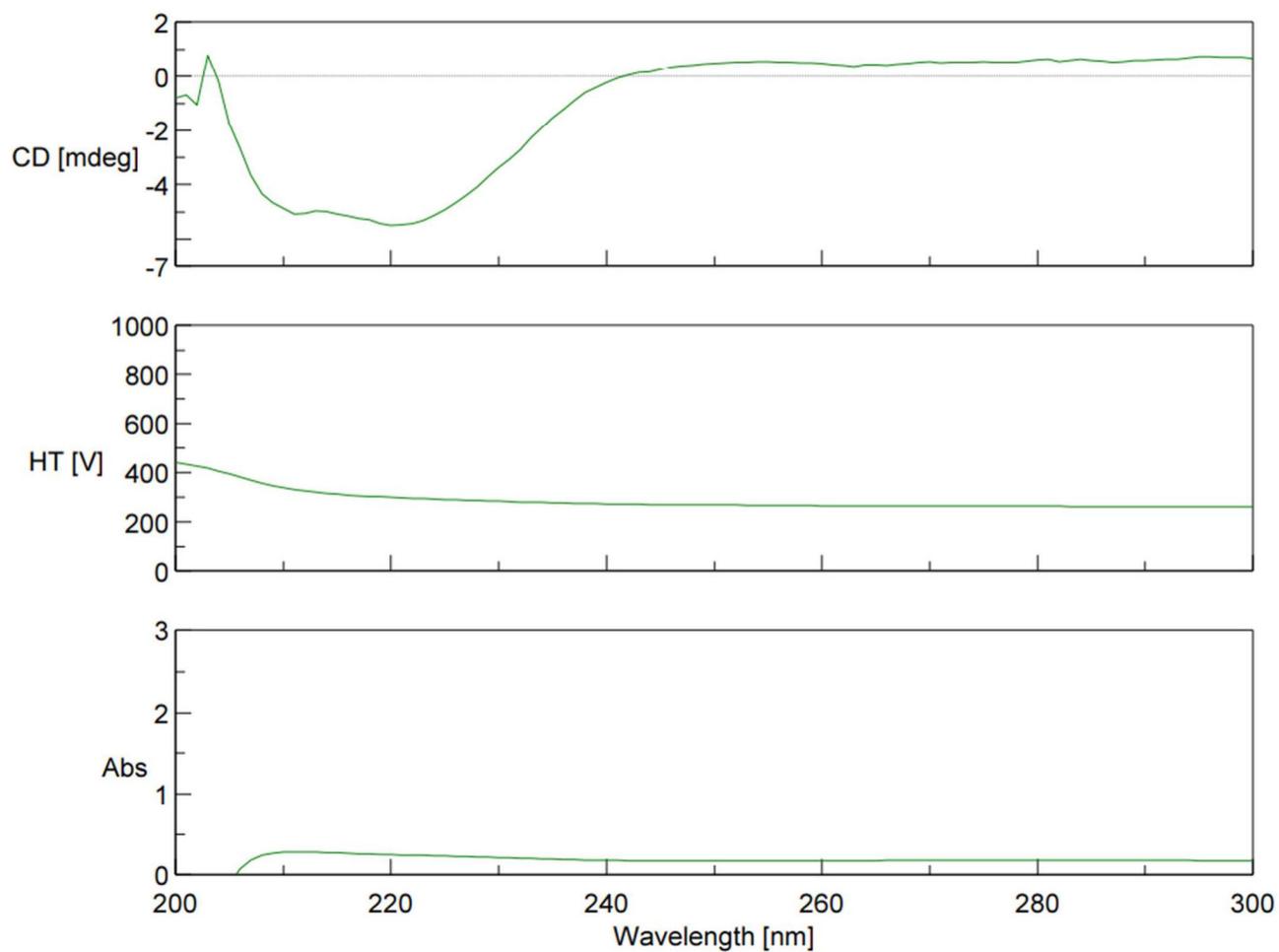


Figure 17. OppD1 circular dichroism spectra's from Jasco J-1000 with the following specifications – cell length: 10 mm; scan range: 450-200 nm; data pitch: 1.00 nm; D.I.T: 2 sec; bandwidth: 2 nm; scanning speed: 200 nm/min; accumulation: 3; solvent (300 mM KCl; 50 mM Tris; pH 7.5). Spectra showing appropriate amount of signal to noise ratio with HT levels below 700V, and a lower level of absorption seen across wavelengths.

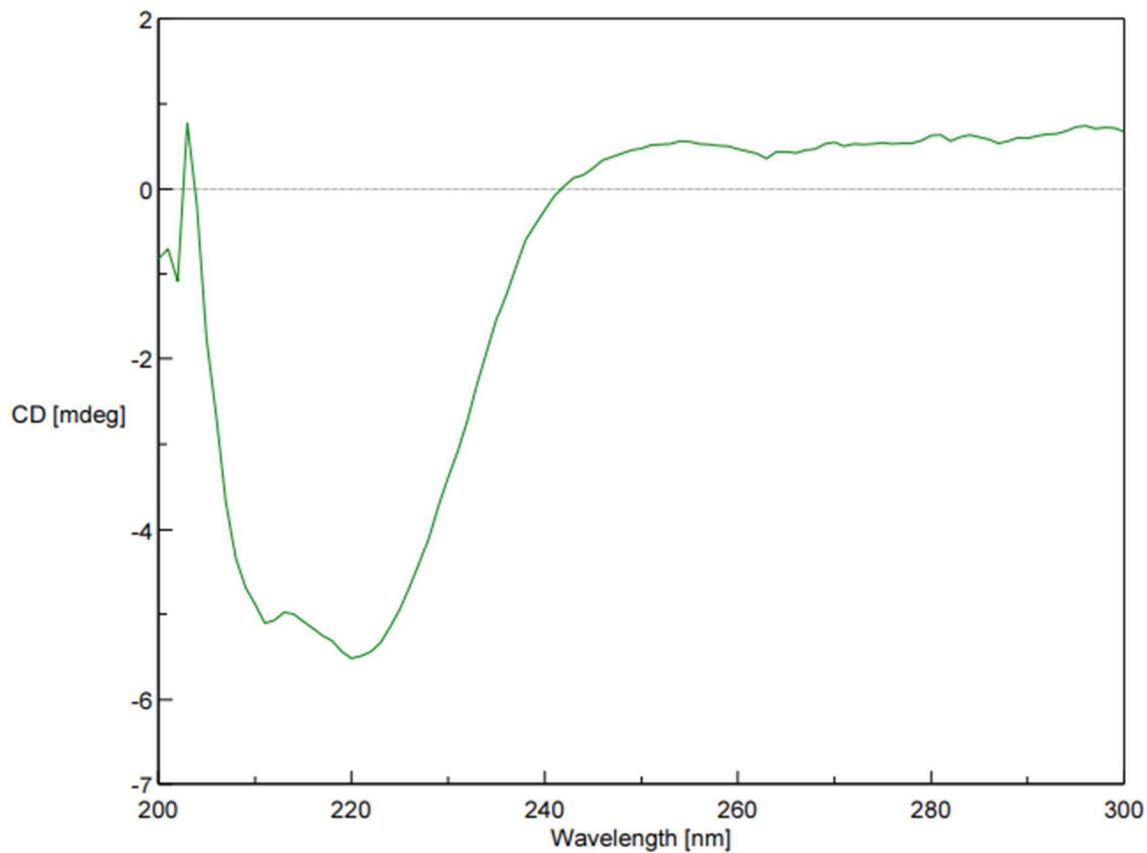


Figure 18. OppD1 Circular Dichroism Spectra from Jasco J-1000 with the following specifications – cell length: 10 mm; scan range: 450-200 nm; data pitch: 1.00 nm; D.I.T: 2 sec; bandwidth: 2 nm; scanning speed: 200 nm/min; accumulation: 3; solvent (300 mM KCl; 50 mM Tris; pH 7.5). Spectra data is indicative of significant alpha helical structural components.

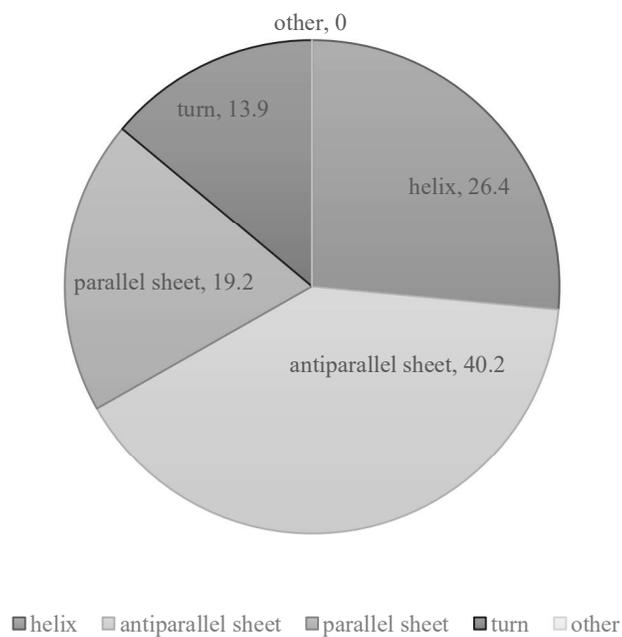


Figure 19 . Single spectrum analysis output from BeStSel online bioinformatic program with maximum usable wavelengths of 200 nm to 250nm. Figure was generated in Microsoft Corporation. (2018). Microsoft Excel. Figure shows percentages of each categorical secondary structure as approximated by BeStSetl present in OppD1 based on CD spectrum data (Figure 18).

### 3.3.3 Additional Bioinformatic Analysis

Initial attempts to superimpose, via modeling, OppD1's unordered sequence onto the monomeric SapD (A13W\_04843: AlphaFold) were unsuccessful, despite the high sequence identity [4], raising questions as to why. Subsequent analysis involved accessing the cataloged *H. marismortui* OppD1 protein's PDB file from AlphaFold and conducting a fitting protocol using SapD (A13W\_04843: AlphaFold) as the reference in Swiss PDB Viewer (DeepView/Swiss-PDBViewer @ 1995-2012; Guex and Peitsch 1997). This operation enabled the complete alignment of SapD to a partial segment of OppD1 (Figure 20), with a significant portion of the N-terminal region remaining unordered (Figure 21). This finding aligns with articles indicating SapD's approximate size as 37,661 Da [83] and the estimated size of OppD1 from SDS PAGE electrophoresis at around ~44,000 Da. The discrepancy in size and structure alignment suggests that *H. marismortui*'s OppD1 likely possesses additional domains serving a distinct function in the archaeal version of this putative ATP-binding protein compared to its *E. coli* homolog. This could possibly include interactions with other system proteins isolated in the Gorrell Lab, such as TrkA, OppF and OppD2. Another possibility may be that this additional unordered domain functions in some aspect of solubility or overall stability of the archaeal protein.

The percentages derived from the BeStSet program, in conjunction with the sequence alignment of OppD1/SapD, imply that the unaligned portion of OppD1 is likely more than random unordered sequence (uncoiled or unfolded) but likely folded into a secondary structure that serves an unidentified function. Considering this discovery, OppD1, generated from the cloned OppD1 transcript, was aligned with SapD using only the identical sequence portions of OppD1, as illustrated in Figure 22. This alignment demonstrated the complete correspondence of the backbone of each protein without revealing any secondary structural components for OppD1.

### 3.4 Discussion and Conclusion

In this chapter, we conducted a comprehensive analysis of OppD1, focusing on its protein size and secondary structure composition using Size Exclusion Chromatography (SEC), Circular Dichroism Spectroscopy, and bioinformatics tools. The objective was to gain a deeper understanding of OppD1's characteristics and its relationship with the homologous protein, SapD [4, 66].

The chromatogram results displayed in Figure 10 exemplify a successful gel filtration experiment, effectively segregating proteins extracted from *H. marismortui* following OppD1 expression based on their sizes. The SDS-PAGE in Figure 16 highlights the absorbance peak, indicating OppD1's elution point at sample 61. Comparison of elution volumes and the electrophoresis gel provides valuable insights, revealing that OppD1, in its likely monomeric form, possesses an approximate molecular weight of ~44,000 Daltons. Furthermore, additional gel filtration runs, as depicted in Figure 17, showed enhanced separation, enabling the concentration of the peak around 60 ml for subsequent Circular Dichroism (CD) spectrum analysis.

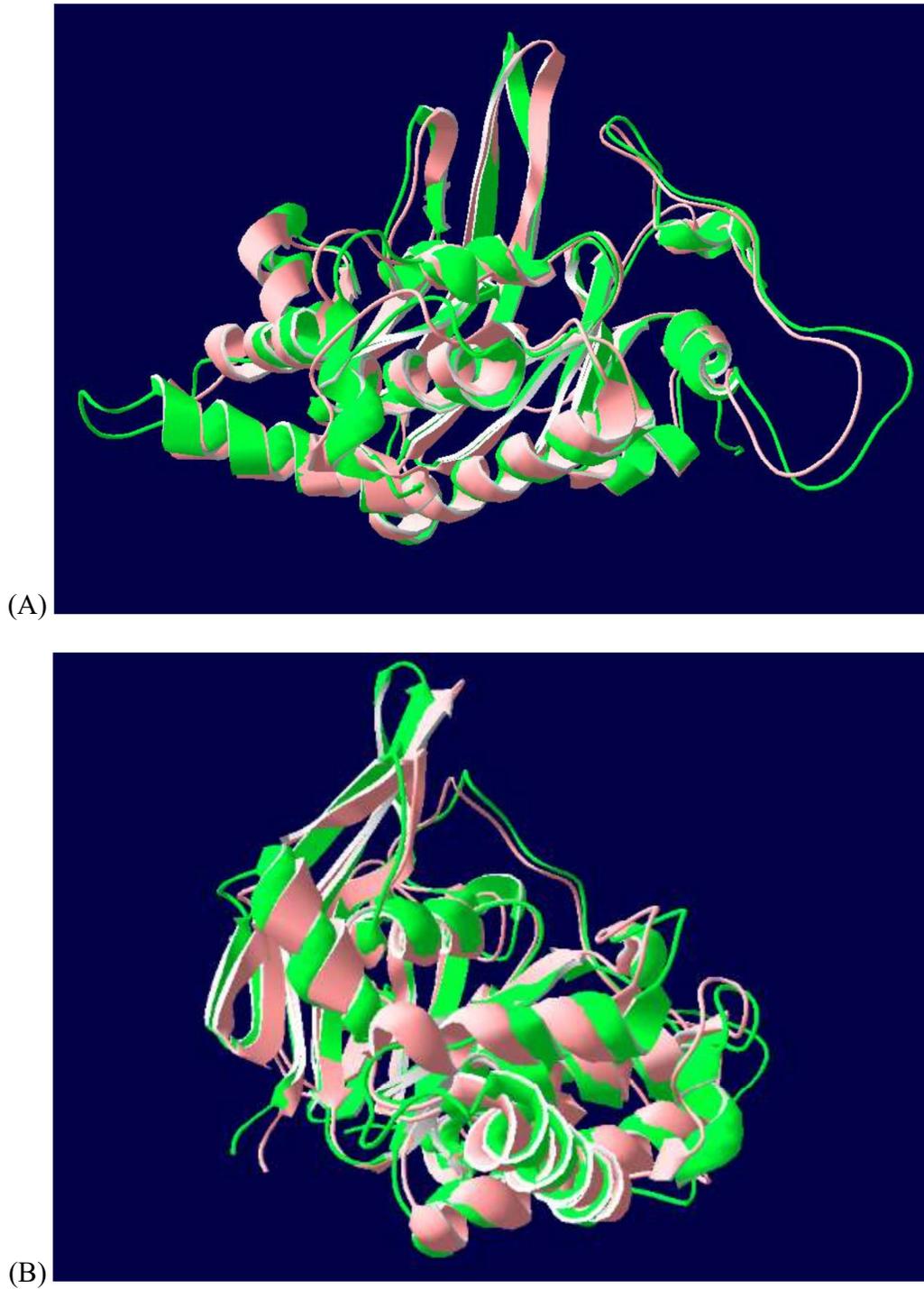


Figure 20. Complete alignment of *E. coli* SapD (green) with partial sequence of *H. marismortui* (Coral) OppD1 (Alpha Fold). Figure 21A shows front view of alignment and 21B shows a 90 degree counter clockwise view around the y-axis.

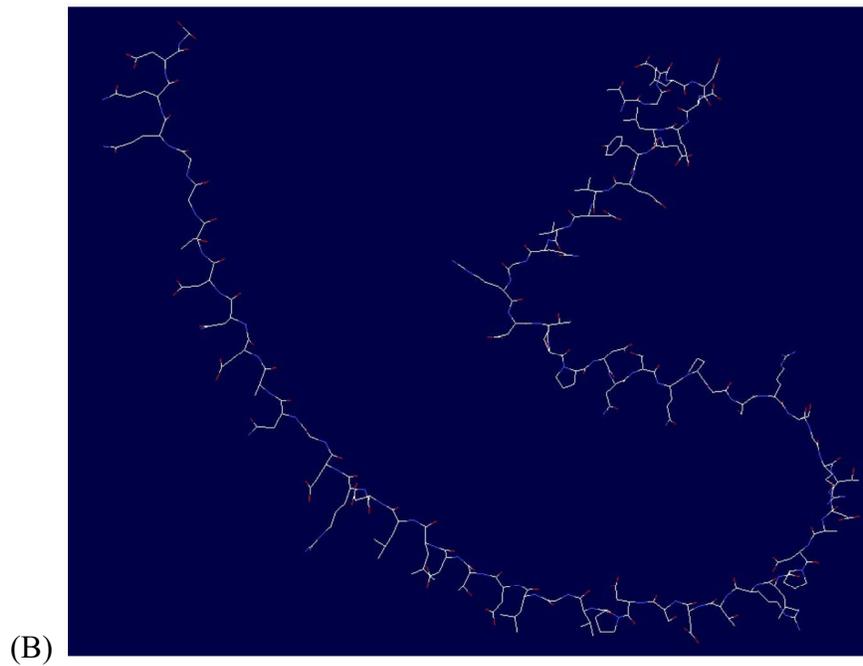
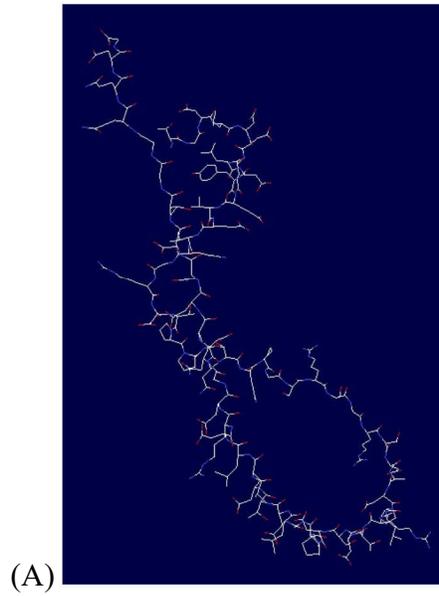


Figure 21. Depicts the N-terminal unordered sequence of OppD1 shown in wire frame presentation, which is not found in the SapD sequence. Figure 22A shows front view of alignment and 22B shows a 90 degree counter clockwise view around the y-axis. The amino acids shown here are from positions 326 to 393 in the polypeptide chain. Full sequence of amino acids can be seen in the Appendix – Figure 10.

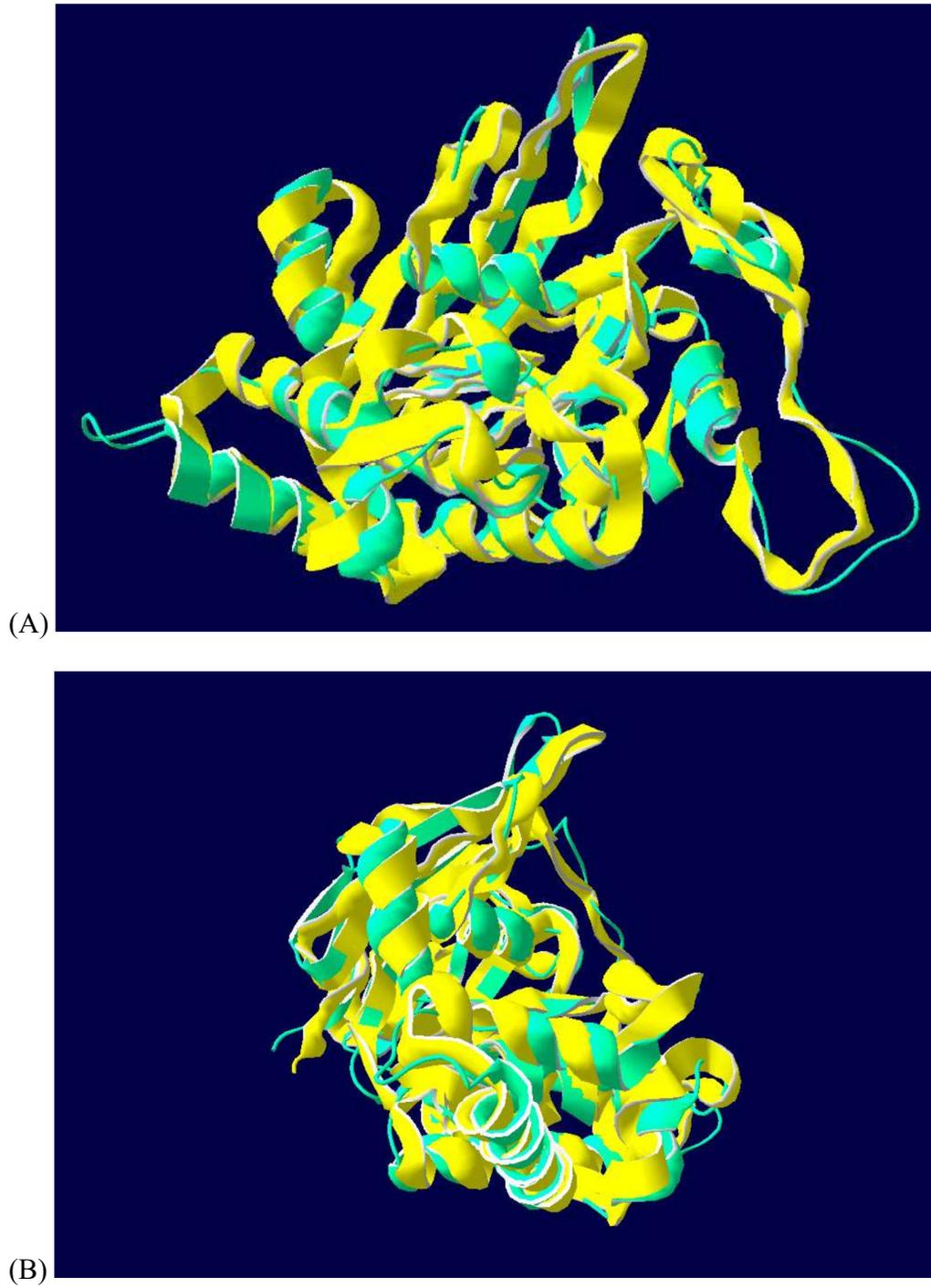


Figure 22. Overlay of OppD1 linear sequence generated from Benchling (green) on SapD (yellow). Figure only shows sequence of OppD1 that aligns with SapD. Figure 23A figure shows front view of alignment and 23B shows a 90 degree counter clockwise view around the y-axis.

Circular Dichroism Spectroscopy consistently depicted HT voltages below 700 V, even when challenged by elevated salt content in the buffer. This demonstrates an appropriate signal-to-noise ratio (S/N), as evident in Figure 18. Notably, the third spectrum in Figure 18 exhibited a reduced level of protein absorption while still maintaining absorption within the wavelength range of 210 to 240 nm. These low levels of absorption will also be because of the low level of aromatic amino acids present in OppD1. These initial analyses confirm the relative validity of the CD spectrum for OppD1 absorption, as illustrated in Figure 19. Initial visual assessment of this spectrum suggested the presence of significant alpha helical secondary structures, identified by negative bands at 222 nm and 208 nm, along with a positive band at 193 nm, consistent with Greenfield's observations in 2006 [82]. Furthermore, detailed secondary structure information was derived through bioinformatic analysis using BeStSel (Beta Structure Selection), as portrayed in Figure 20. This innovative bioinformatic program possesses the capability to predict a protein's folding at the topology and homology levels in alignment with the CATH classification system [79].

Initial attempts to superimpose OppD1's linear sequence onto SapD were met with challenges, despite their high sequence identity, raising intriguing questions. Subsequent analyses included accessing the cataloged OppD1 protein's PDB file from AlphaFold and conducting a fitting protocol using SapD as the reference, as represented in Figure 16. This operation resulted in the complete alignment of SapD to a partial segment of OppD1, leaving a significant portion of the N-terminal region unordered, as depicted in Figure 21. This finding aligns with earlier articles indicating SapD's approximate size as 37,661 Da [83] and the estimated size of OppD1 from SDS PAGE electrophoresis at around ~44,000 Da. This discrepancy in size and structural alignment suggests that *H. marismortui*'s OppD1 may harbor

additional domains serving distinct functions in the archaeal context, in contrast to its *E. coli* homolog. These additional domains could potentially interact with other system proteins within the Gorrell Lab, such as TrkA, or be involved in aspects related to solubility or overall stability of the archaeal protein.

The percentages derived from the BeStSet program [79], coupled with the sequence alignment of OppD1/SapD, suggest that the unaligned portion of OppD1 is not merely random unordered sequence, but rather is likely folded into a secondary structure that serves an unidentified function. Consequently, a comparison between OppD1 and SapD, focusing exclusively on their identical sequence portions [4, 66], as illustrated in Figure 23, demonstrates the complete alignment of the backbone of each protein without revealing any secondary structural components for OppD1.

In summary, this chapter's results significantly enhance our understanding of OppD1, its structural characteristics, and its potential differences and similarities from its homolog, SapD. These findings open new avenues for further exploration and research into OppD1's functional role in *H. marismortui* and its interactions with other proteins in the archaeal system.

## **Chapter 4: Concluding Remarks**

## **4.1 Summary**

### **4.1.1 Introduction – Chapter 1**

#### **4.1.1.1 Introduction to Archaea and Archaeal Transcription**

Archaea, as a distinct domain of life, holds potential ancestral connections to the more established domains of Eukarya and Bacteria [1]. In this comprehensive investigation, the essential role of ion transport in maintaining ionic gradients in haloarchaea species was examined [53]. To achieve this, membrane protein transport systems were explored, with a particular focus on the RCK (Regulator of Conductance of K<sup>+</sup>) transporters, a superfamily found across all three domains of life. These RCK transporters play a vital role in processes such as osmoregulation, pH homeostasis, turgor pressure regulation, and membrane potential control in prokaryotic organisms.

Furthermore, this superfamily is subsumed under the broader category of ABC transporters, a well-studied group of membrane transporters involved in substrate uptake, export, and osmoregulation [52]. These transporters typically consist of two integral membrane proteins and two cytoplasmic ATPases, which facilitate substrate movement through ATP hydrolysis [52]. Notably, ABC transporters have been linked to significant health concerns, including cystic fibrosis, multidrug resistance in cancer, and antibiotic resistance [52].

Despite their structural homology across domains, ABC transporters continue to play a critical role in ion transport and osmotic regulation in archaeal environments [52]. This study particularly delved into the Trk system, an evolutionarily related system to the Ktr system and prevalent in both bacteria and archaea [55]. The membrane channel complex of these systems includes a dimeric membrane protein, such as KtrB in the Ktr system or TrkH in the Trk system, which interacts with a cytosolic octameric protein, KtrA [4, 55, 57, 84, 85].

While these systems resemble canonical ABC transporters, there is some debate regarding the necessity of nucleotide binding interactions for channel activation [60]. While ATP is hypothesized to be the likely binding partner, conclusive evidence remains elusive [55, 60].

Despite the wealth of knowledge on these transport systems, limited information is available on the role of regulatory proteins within the Trk complex [60]. TrkE has shown to be crucial for the function of TrkH and to induce residual activity in TrkG [60]. Both TrkH and TrkG participate in the attachment of the peripheral membrane protein TrkA, essential for Trk system activity [60, 61]. The product of TrkE, found in *E. coli*, has been associated with potassium uptake, although the specific mechanism remains unclear [62]. Research has suggested that TrkE plays a role in potassium uptake through interactions with TrkH and TrkG. These findings imply that TrkE might serve as a modulator of TrkA, and its requirement may vary according to environmental conditions and potassium transport activity levels [63].

In addition, *TrkE*, which has been identified in *SapD* shows homology to oligopeptide permease systems (Opp) [4]. Some identified Opp proteins have shown high sequence similarity to genes mapped in the *sapABCDF* operon that contains *SapD*, which could mean likely overlap in certain features such as architecture and function [66]. Due to the variability in sequences and subsequent proteins that all contain some sequence and function similarities it further highlights the importance of looking into the possibility that TrkE may function in other systems and serve in a variety of different cellular processes.

#### **4.1.1.5 *Haloarcula marismortui***

In this study, *H. marismortui* is an organism for investigating remarkable adaptation to high-salinity environments, making it an invaluable subject for halophilic research [3]. As an archaeon that thrives in extreme salt conditions, it is critical for *H. marismortui* to maintain

osmotic regulation and effective salt transportation [4, 5]. Due to its complete genome sequence and remarkable ability to adapt to high-salinity environments *H. marismortui* can also be used as a model organism to investigate the cellular processes that allow for this adaptation.

*H. marismortui* is notable for its unique ability to withstand and even flourish in hypersaline environments, discovered in the Dead Sea, one of the saltiest natural bodies of water on Earth [3]. The complete sequencing of *H. marismortui*'s genome has provided valuable insights into the mechanisms that allow it to not only survive but also thrive in such extreme conditions [3]. One of the key factors contributing to its salt tolerance is the Trk system, which is believed to play a role in stabilizing the organism in these environments [4, 5]. Interestingly, similar systems have been associated with enhanced virulence in other species [6].

*H. marismortui* employs a strategy of salt sequestration to maintain osmotic balance. Initially, it was thought that this involved passive transport of potassium ions [68, 69]. However, considering the extremely high intracellular salt concentrations found in *H. marismortui* [67], it is improbable that such high potassium molarities could be achieved solely through passive diffusion across its permeable membrane [70]. Therefore, *H. marismortui* must employ specific mechanisms related to its membrane morphology to maintain the critical transmembrane salt gradient essential for its survival. Unpublished research shows that these membrane mechanisms likely utilize various Trk homologs, that can show differentiated expression in *H. marismortui* under different  $K^+$  concentrations [72].

In conclusion, *H. marismortui*'s ability to adapt and thrive in high-salinity environments makes it an excellent model for studying halophilic organisms. Its utilization of salt sequestration and its potential involvement of the Trk system highlight the fascinating strategies that enable this extremophile to flourish in hypersaline conditions. These findings shed light on the unique

adaptations of halophiles and their ability to maintain cellular integrity in such challenging environments.

#### **4.1.2 Cloning and Expression of OppD1– Chapter 2**

The process of amplifying and cloning the *OppD1* gene from *H. marismortui*'s genomic DNA and subsequently expressing and purifying the target protein was the focus of chapter 2. The initial cloning phase faced challenges, resulting in difficulties with recombinant vector construction and successful ligation confirmation. The study then shifted focus to the direct ligation of *OppD1* into the *pET21b* expression vector, but similar issues persisted.

To further address these challenges, internal primers were employed to confirm proper sequence ligation. These primers along with the technique of nested PCR were used to conclude that a successful recombinant clone was generated and ensured that initial colony PCR positives were not due to non-specific binding of the primers to likely *E. coli* transcripts similar to *OppD1*. Successful verification allowed sequencing of the cloned gene. Subsequent *OppD1* expression and purification demonstrated optimal protein expression and affinity exchange chromatography's effectiveness.

In summary, while initial cloning posed challenges, the project ultimately succeeded in expressing and purifying OppD1. These findings highlight the importance of optimizing cloning at PCR steps for purposes beyond simply adjusting parameters, but also adjusting primers relative to host cells.

#### **4.1.3 Characterization of OppD1 – Chapter 3**

In this chapter, a comprehensive analysis of OppD1 is performed, focusing on its protein size and secondary structure composition using Size Exclusion Chromatography (SEC), Circular Dichroism Spectroscopy, and bioinformatics tools. The objective was to gain a deeper

understanding of OppD1's characteristics and its relationship with the homologous protein, SapD.

Through SEC, OppD1 was successfully purified from *E.coli* after expression based on their sizes. SDS-PAGE analysis of samples; comparisons to bioinformatic data, and comparisons to literature of SapD suggest OppD1 is likely a monomeric protein of approximately 44 kDa.

Circular Dichroism Spectroscopy consistently revealed the presence of alpha helical secondary structures in OppD1. The analysis indicates most of the secondary structure components are alpha helical. These findings align with observations made in a previous study by Greenfield in 2006 [82].

However, efforts to align OppD1's sequence with that of SapD for modeling raised questions, as they exhibited sequence variations. It was found after several alignment attempts that *H. marismortui*'s OppD1 may contain additional domains that likely serve unique functions or allow interaction in comparison to its *E. coli* homolog.

In summary, this analysis provided valuable insights into OppD1's structural characteristics and its potential differences from SapD. These findings open new avenues for further exploration and research into OppD1's functional role in *H. marismortui* and its interactions with other proteins in the archaeal system.

## **4.2 Significance**

Since its initial discovery in the late 1970s, the archaeal domain has been encountered in diverse environments, spanning from extreme aquatic ecosystems to the digestive tracts of mammals, including humans [73]. Various branches of the archaeal domain have been identified in the gut, and a few have been found in the oral cavity [86]. Recent concerns have emerged

about heightened antimicrobial tendencies exhibited by these microorganisms, adding to the intrigue [87].

The discussions surrounding the OppD1 polypeptide predominantly stem from research conducted on its presumed counterpart, TrkE, in the *E.coli* K-12 system [60]. This study aims to provide an initial characterization of the OppD1 protein from *H. marismortui*, in the context of the hypotheses generated by previous alignment and functional homology studies detailed in this introduction. By delving into OppD1's role in ion regulation within the framework of *H. marismortui*, valuable insights into cellular homeostasis may be unveiled.

### **4.3 Conclusion and Future Considerations**

In the context of future research on archaeal cloning and the characteristics of OppD1, and halophilic genes and proteins several key considerations warrant attention:

First, combining evolutionary couplings with NMR data [88] should offer a more precise insight into OppD1's secondary and tertiary structure, enhancing our understanding of this protein's architecture. The size and monomeric nature of OppD1 lends itself to NMR structure determination [89].

Second, working towards a more streamlined purification of OppD1, should allow for even comprehensive experiments with circular dichroism and NMR. In addition, dialyzing samples into differing salts, concentration of salts to analyze the optimal archaeon protein conformational changes [50].

Third, investigating potential complications between restriction endonuclease activity with archaeal transcripts during cloning attempts is also crucial for enhancing cloning success for future archaeal protein characterization objectives.

Fourth, utilizing the sizing chromatographic techniques can shed light on the potential of OppD1 to bind with protein partners, providing further valuable information regarding its structural organization in the overall transporter.

Fifth, a pertinent aspect of future research is to examine the potential interactions between OppD1 and previously studied Gorrell Lab Trk proteins, including TrkA1/TrkA2, OppD2, and OppF. Investigating the possible collaborative functions among these proteins may offer a comprehensive view of the intricate workings of the archaeal system, providing a rich ground for further exploration and discovery. Also, the unaligned portions of OppD1 present an intriguing area for investigation. Analyzing their potential functionality, including their involvement in protein solubility and stability, can expand our knowledge of OppD1's role in archaeal systems. These considerations collectively contribute to advancing our understanding of OppD1 and its significance in archaeal biology. With this research it may then be possible to unravel an additional layer revealing further the impact of this membrane system in the archaea *H. marismortui*, and more broadly to the large protein family of ATP Transporters and their role in the global concern of antibiotic resistance.

Lastly, *H. marismortui*'s adeptness in high-salinity habitats and its utilization of salt sequestration, potentially through the Trk system, establish it as a prime model for studying halophilic organisms. These revelations shed light on halophiles' unique adaptations for maintaining cellular integrity in challenging environments.

These preliminary findings and possible future considerations can bring further understanding in the overall functional mechanism of OppD1 as part of the Trk system, and its role in ion transport with the goal of osmotic regulation in response to varying environmental conditions.



## References

1. Barry, E.R. and S.D. Bell, *DNA replication in the archaea*. Microbiol Mol Biol Rev, 2006. **70**(4): p. 876-87.
2. DasSarma, S. and P. Arora, *Halophiles*. e LS, 2001.
3. Baliga, N.S., et al., *Genome sequence of Haloarcula marismortui: a halophilic archaeon from the Dead Sea*. Genome Res, 2004. **14**(11): p. 2221-34.
4. Vieira-Pires, R.S., A. Szollosi, and J.H. Morais-Cabral, *The structure of the KtrAB potassium transporter*. Nature, 2013. **496**(7445): p. 323-8.
5. Corratge-Faillie, C., et al., *Potassium and sodium transport in non-animal cells: the Trk/Ktr/HKT transporter family*. Cell Mol Life Sci, 2010. **67**(15): p. 2511-32.
6. Lupp, C., R.E.W. Hancock, and E.G. Ruby, *The Vibrio fischeri sapABCDF locus is required for normal growth, both in culture and in symbiosis*. Archives of Microbiology, 2002. **179**(1): p. 57-65.
7. Bell, S.D. and S.P. Jackson, *Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features*. Trends Microbiol, 1998. **6**(6): p. 222-8.
8. Ouhammouch, M., *Transcriptional regulation in Archaea*. Curr Opin Genet Dev, 2004. **14**(2): p. 133-8.
9. Kramm, K., U. Endesfelder, and D. Grohmann, *A Single-Molecule View of Archaeal Transcription*. J Mol Biol, 2019. **431**(20): p. 4116-4131.

10. Brinkman, A.B., et al., *An Lrp-like transcriptional regulator from the archaeon Pyrococcus furiosus is negatively autoregulated*. J Biol Chem, 2000. **275**(49): p. 38160-9.
11. Bell, S.D., et al., *Transcriptional regulation of an archaeal operon in vivo and in vitro*. Mol Cell, 1999. **4**(6): p. 971-82.
12. Brinkman, A.B., et al., *The Sulfolobus solfataricus Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability*. J Biol Chem, 2002. **277**(33): p. 29537-49.
13. Walker, J.E., O. Luyties, and T.J. Santangelo, *Factor-dependent archaeal transcription termination*. Proc Natl Acad Sci U S A, 2017. **114**(33): p. E6767-E6773.
14. Cavicchioli, R., *Archaea: molecular and cellular biology*. 2007: ASM Press.
15. Nelson, D.L. and M.M. Cox, *Lehninger Principles of Biochemistry 6th Edition (2013)*.
16. Clark, M.A., M. Douglas, and J. Choi, *Biology 2e via Openstax. com*. 2018.
17. Munoz-Tello, P., et al., *Polyuridylation in Eukaryotes: A 3'-End Modification Regulating RNA Life*. Biomed Res Int, 2015. **2015**: p. 968127.
18. Portnoy, V. and G. Schuster, *RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R*. Nucleic Acids Res, 2006. **34**(20): p. 5923-31.
19. Martienssen, R.A. and E.J. Richards, *DNA methylation in eukaryotes*. Current Opinion in Genetics & Development, 1995. **5**(2): p. 234-242.

20. Sánchez-Romero, M.A., I. Cota, and J. Casadesús, *DNA methylation in bacteria: from the methyl group to the methylome*. *Current Opinion in Microbiology*, 2015. **25**: p. 9-16.
21. Edelheit, S., et al., *Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs*. *PLoS Genet*, 2013. **9**(6): p. e1003602.
22. Aittaleb, M., et al., *Structure and function of archaeal box C/D sRNP core proteins*. *Nat Struct Biol*, 2003. **10**(4): p. 256-63.
23. Huang, L., et al., *Control of box C/D snoRNP assembly by N(6)-methylation of adenine*. *EMBO Rep*, 2017. **18**(9): p. 1631-1645.
24. Huang, L., S. Ashraf, and D.M.J. Lilley, *The role of RNA structure in translational regulation by L7Ae protein in archaea*. *RNA*, 2019. **25**(1): p. 60-69.
25. Fischer, S., et al., *Regulatory RNAs in Haloferax volcanii*. *Biochem Soc Trans*, 2011. **39**(1): p. 159-62.
26. Gelsinger, D.R. and J. DiRuggiero, *The Non-Coding Regulatory RNA Revolution in Archaea*. *Genes (Basel)*, 2018. **9**(3).
27. Babski, J., et al., *Small regulatory RNAs in Archaea*. *RNA Biol*, 2014. **11**(5): p. 484-93.
28. Prasse, D., et al., *Regulatory RNAs in archaea: first target identification in Methanoarchaea*. *Biochem Soc Trans*, 2013. **41**(1): p. 344-9.
29. Brenneis, M. and J. Soppa, *Regulation of translation in haloarchaea: 5'- and 3'-UTRs are essential and have to functionally interact in vivo*. *PLoS One*, 2009. **4**(2): p. e4484.

30. Srivastava, A., et al., *In silico analysis of 5'-UTRs highlights the prevalence of Shine-Dalgarno and leaderless-dependent mechanisms of translation initiation in bacteria and archaea, respectively*. J Theor Biol, 2016. **402**: p. 54-61.
31. Benelli, D., E. Maone, and P. Londei, *Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation*. Mol Microbiol, 2003. **50**(2): p. 635-43.
32. Sartorius-Neef, S. and F. Pfeifer, *In vivo studies on putative Shine-Dalgarno sequences of the halophilic archaeon Halobacterium salinarum*. Mol Microbiol, 2004. **51**(2): p. 579-88.
33. Condo, I., et al., *Cis-acting signals controlling translational initiation in the thermophilic archaeon Sulfolobus solfataricus*. Mol Microbiol, 1999. **34**(2): p. 377-84.
34. Tolstrup, N., et al., *Two different and highly organized mechanisms of translation initiation in the archaeon Sulfolobus solfataricus*. Extremophiles, 2000. **4**(3): p. 175-9.
35. Benelli, D. and P. Londei, *Translation initiation in Archaea: conserved and domain-specific features*. Biochem Soc Trans, 2011. **39**(1): p. 89-93.
36. Schmitt, E., et al., *Start Codon Recognition in Eukaryotic and Archaeal Translation Initiation: A Common Structural Core*. Int J Mol Sci, 2019. **20**(4).
37. Akulich, K.A., et al., *Four translation initiation pathways employed by the leaderless mRNA in eukaryotes*. Sci Rep, 2016. **6**: p. 37905.
38. Moll, I., et al., *Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs*. Nucleic Acids Res, 2004. **32**(11): p. 3354-63.

39. Andreev, D.E., et al., *A leaderless mRNA can bind to mammalian 80S ribosomes and direct polypeptide synthesis in the absence of translation initiation factors*. Mol Cell Biol, 2006. **26**(8): p. 3164-9.
40. Moll, I., et al., *Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control*. Mol Microbiol, 2002. **43**(1): p. 239-46.
41. Arkhipova, V., et al., *Binding of the 5'-Triphosphate End of mRNA to the gamma-Subunit of Translation Initiation Factor 2 of the Crenarchaeon Sulfolobus solfataricus*. J Mol Biol, 2015. **427**(19): p. 3086-95.
42. Maone, E., et al., *Functional analysis of the translation factor aIF2/5B in the thermophilic archaeon Sulfolobus solfataricus*. Mol Microbiol, 2007. **65**(3): p. 700-13.
43. Yatime, L., et al., *Functional molecular mapping of archaeal translation initiation factor 2*. J Biol Chem, 2004. **279**(16): p. 15984-93.
44. Pedulla, N., et al., *The archaeal eIF2 homologue: functional properties of an ancient translation initiation factor*. Nucleic Acids Res, 2005. **33**(6): p. 1804-12.
45. Lee, J.H., et al., *Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation initiation factor IF2*. Proc Natl Acad Sci U S A, 1999. **96**(8): p. 4342-7.
46. Mauer, J., et al., *Reversible methylation of m6Am in the 5' cap controls mRNA stability*. Nature, 2017. **541**(7637): p. 371-375.

47. Giliberti, J., et al., *A 5'-terminal phosphate is required for stable ternary complex formation and translation of leaderless mRNA in Escherichia coli*. RNA, 2012. **18**(3): p. 508-18.
48. Schafer, G., M. Engelhard, and V. Muller, *Bioenergetics of the Archaea*. Microbiol Mol Biol Rev, 1999. **63**(3): p. 570-620.
49. Thombre, R.S., et al., *Biology and survival of extremely halophilic archaeon Haloarcula marismortui RR12 isolated from Mumbai salterns, India in response to salinity stress*. Sci Rep, 2016. **6**: p. 25642.
50. Reed, C.J., et al., *Protein Adaptations in Archaeal Extremophiles*. Archaea, 2013. **2013**: p. 373275.
51. Tehei, M., et al., *Neutron scattering reveals extremely slow cell water in a Dead Sea organism*. Proceedings of the National Academy of Sciences, 2007. **104**(3): p. 766-771.
52. Albers, S.V., et al., *Insights into ABC transport in archaea*. J Bioenerg Biomembr, 2004. **36**(1): p. 5-15.
53. Rocha, R., et al., *Characterization of the molecular properties of KtrC, a second RCK domain that regulates a Ktr channel in Bacillus subtilis*. J Struct Biol, 2019. **205**(3): p. 34-43.
54. Wilkens, S., *Structure and mechanism of ABC transporters*. F1000Prime Rep, 2015. **7**: p. 14.

55. Guo, Y., et al., *Characterization and function analysis of a Halo-alkaline-adaptable Trk K<sup>+</sup> uptake system in Alkalimonas amylolytica strain N10*. Science in China Series C: Life Sciences, 2009. **52**(10): p. 949-957.
56. Cotsaftis, O., et al., *A two-staged model of Na<sup>+</sup> exclusion in rice explained by 3D modeling of HKT transporters and alternative splicing*. PLoS One, 2012. **7**(7): p. e39865.
57. Hamamoto, S., et al., *HKT transporters mediate salt stress resistance in plants: from structure and function to the field*. Curr Opin Biotechnol, 2015. **32**: p. 113-120.
58. Marc Mosimann, S.G., Tanja Wenzler, Alexandra Luscher, Nobuyuki, and Pascal Maser, *A Trk/HKT-Type K<sup>+</sup> Transporter from Trypanosoma brucei*. Eukaryotic Cell, 2010. **9**(4): p. 539-546.
59. Cao, Y., et al., *Gating of the TrkH ion channel by its associated RCK protein TrkA*. Nature, 2013. **496**(7445): p. 317-322.
60. Harms, C., et al., *Identification of the ABC protein SapD as the subunit that confers ATP dependence to the K<sup>+</sup>-uptake systems TrkH and TrkG from Escherichia coli K-12*. Microbiology, 2001. **147**(11): p. 2991-3003.
61. Buurman, E.T., et al., *Multiple Paths for Nonphysiological Transport of K<sup>+</sup> in Escherichia coli*. Journal of Bacteriology, 2004. **186**(13): p. 4238-4245.

62. Dosch, D.C., et al., *Genetic analysis of potassium transport loci in Escherichia coli: evidence for three constitutive systems mediating uptake potassium*. Journal of Bacteriology, 1991. **173**(2): p. 687-696.
63. Schlösser, A., et al., *TrkH and its homolog, TrkG, determine the specificity and kinetics of cation transport by the Trk system of Escherichia coli*. Journal of Bacteriology, 1995. **177**(7): p. 1908-1910.
64. Szollosi, A., et al., *Dissecting the Molecular Mechanism of Nucleotide-Dependent Activation of the KtrAB K<sup>+</sup> Transporter*. PLoS Biol, 2016. **14**(1): p. e1002356.
65. Bossemeyer, D., A. Schlosser, and E.P. Bakker, *Specific cesium transport via the Escherichia coli Kup (TrkD) K<sup>+</sup> uptake system*. J Bacteriol, 1989. **171**(4): p. 2219-21.
66. Parra-Lopez, C., M.T. Baer, and E.A. Groisman, *Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in Salmonella typhimurium*. EMBO J, 1993. **12**(11): p. 4053-62.
67. Jensen, M.W., et al., *Potassium stress growth characteristics and energetics in the haloarchaeon Haloarcula marismortui*. Extremophiles, 2015. **19**(2): p. 315-25.
68. Oren, A., et al., *Haloarcula marismortui (Volcani) sp. nov., nom. rev., an Extremely Halophilic Bacterium from the Dead Sea*. International Journal of Systematic and Evolutionary Microbiology, 1990. **40**(2): p. 209-210.

69. Meury, J. and M. Kohiyama, *ATP is required for K<sup>+</sup> active transport in the archaeobacterium Haloferax volcanii*. Archives of Microbiology, 1989. **151**(6): p. 530-536.
70. Ginzburg, M., L. Sachs, and B.Z. Ginzburg, *Ion metabolism in aHalobacterium*. The Journal of Membrane Biology, 1971. **5**(1): p. 78-101.
71. Alexey S. Syutkin, et al., *Haloarcula marismortui archaeellin genes as ecoparalogs*. Extremophiles, 2014. **18**: p. 341-349.
72. Short, L., M.W. Jensen, and A. Gorrell, *Transport Gene Expression Profiling in Haloarcula marismortui Under Potassium Stress Conditions*.
73. Gribaldo, S. and C. Brochier-Armanet, *The origin and evolution of Archaea: a state of the art*. Philos Trans R Soc Lond B Biol Sci, 2006. **361**(1470): p. 1007-22.
74. Gaci, N.N., et al., *Archaea and the human gut: New beginning of an old story*. World journal of gastroenterology : WJG, 2014. **20**(43): p. 16062-16078.
75. Dridi, B., et al., *The antimicrobial resistance pattern of cultured human methanogens reflects the unique phylogenetic position of archaea*. Journal of antimicrobial chemotherapy, 2011. **66**(9): p. 2038-2044.
76. Sayers, E.W., et al., *Database resources of the national center for biotechnology information*. Nucleic Acids Res, 2022. **50**(D1): p. D20-d26.
77. Ish-Horowicz, D. and J.F. Burke, *Rapid and efficient cosmid cloning*. Nucleic Acids Research, 1981. **9**(13): p. 2989-2898.

78. Manual, I., *Phusion® High-Fidelity PCR Kit*. 2017.
79. Micsonai, A., et al., *BeStSel: webservice for secondary structure and fold prediction for protein CD spectroscopy*. *Nucleic Acids Res*, 2022. **50**(W1): p. W90-W98.
80. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling*. *ELECTROPHORESIS*, 1997. **18**(15): p. 2714-2723.
81. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold*. *Nature*, 2021. **596**(7873): p. 583-589.
82. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure*. *Nat Protoc*, 2006. **1**(6): p. 2876-90.
83. Welch, R.A., et al., *Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli*. *Proc Natl Acad Sci U S A*, 2002. **99**(26): p. 17020-4.
84. Mosimann, M., et al., *A Trk/HKT-type K<sup>+</sup> transporter from Trypanosoma brucei*. *Eukaryot Cell*, 2010. **9**(4): p. 539-46.
85. Olivier Cotsaftis, D.P., Neil Shirley, Mark Tester, Maria Hrmova, *A Two-Staged Model of Na<sup>+</sup> Exclusion in Rice Explained by 3D Modeling of HKT Transporters and Alternative Splicing*. *PLoS ONE*, 2012. **7**(7): p. e39865.
86. Gaci, N., et al., *Archaea and the human gut: new beginning of an old story*. *World J Gastroenterol*, 2014. **20**(43): p. 16062-78.

87. Dridi, B., et al., *The antimicrobial resistance pattern of cultured human methanogens reflects the unique phylogenetic position of archaea*. *J Antimicrob Chemother*, 2011. **66**(9): p. 2038-44.
88. Tang, Y., et al., *Protein structure determination by combining sparse NMR data with evolutionary couplings*. *Nature methods*, 2015. **12**(8): p. 751-754.
89. Kay, L.E., *NMR studies of protein structure and dynamics*. *Journal of magnetic resonance*, 2011. **213**(2): p. 477-491.

**Appendix:**

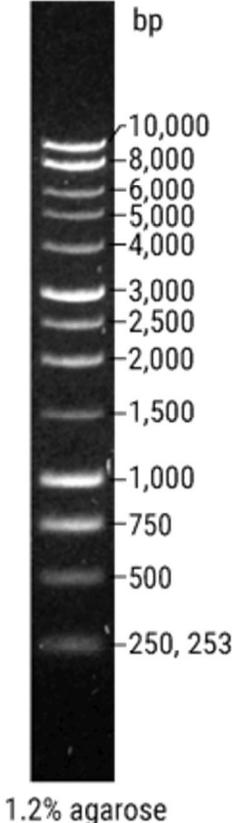


Figure 1. Promega 1kb DNA ladder (CAT#: G5711) shown on a 1.2% agarose gel used for gene identification and size referencing.

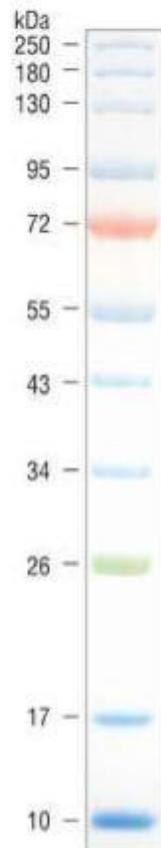


Figure 2. NEB Color Pre-stained Protein Standard, Broad Range (10-250 kDa) (CAT#: P7719L) used for protein identification and size recognition.

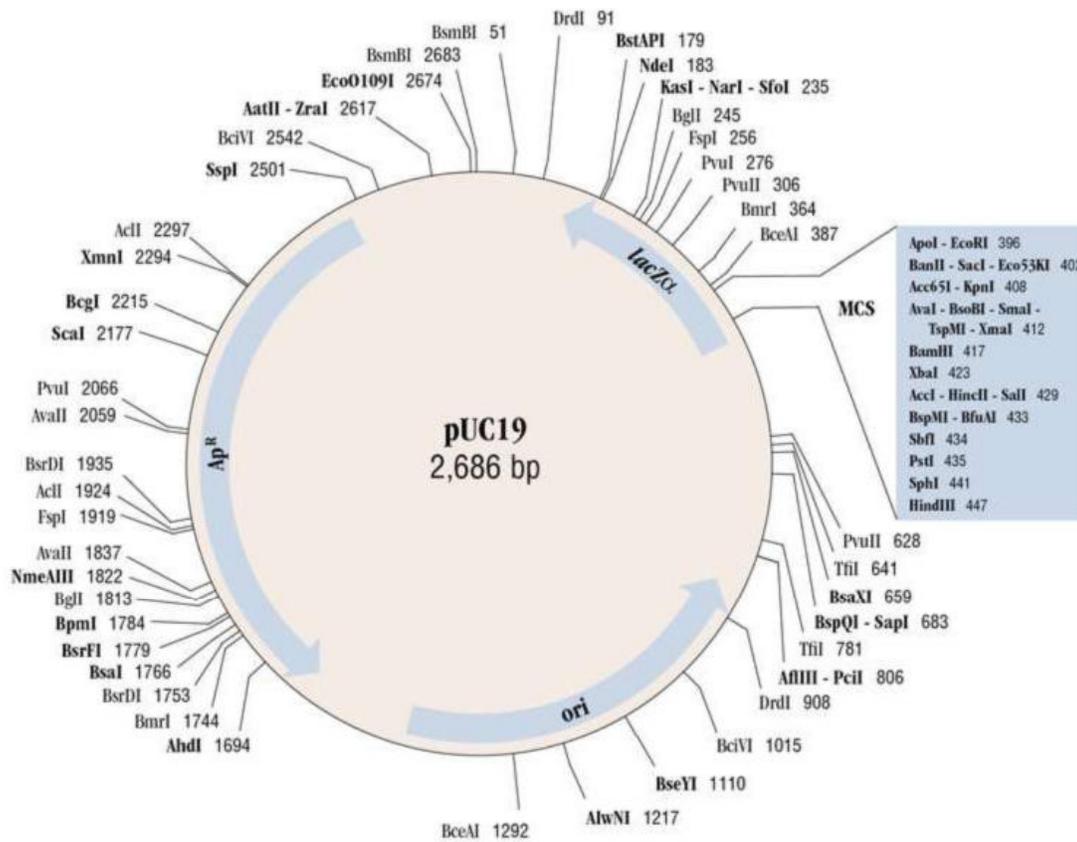


Figure 3. pUC19 (New England Biolabs) vector map showing size of the amplification vector; restriction endonuclease sites; and sequence landmarks.

pET-21a(+) sequence landmarks	
T7 promoter	311-327
T7 transcription start	310
T7•Tag coding sequence	207-239
Multiple cloning sites ( <i>Bam</i> H I - <i>Xho</i> I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
<i>lac</i> I coding sequence	714-1793
pBR322 origin	3227
<i>bla</i> coding sequence	3988-4845
f1 origin	4977-5432

The maps for pET-21b(+), pET-21c(+) and pET-21d(+) are the same as pET-21a(+) (shown) with the following exceptions: pET-21b(+) is a 5442bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-21c(+) is a 5441bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198. pET-21d(+) is a 5440bp plasmid; the *Bam*H I site is in the same reading frame as in pET-21c(+). An *Nco* I site is substituted for the *Nde* I site with a net 1bp deletion at position 238 of pET-21c(+). As a result, *Nco* I cuts pET21d(+) at 234, and *Nhe* I cuts at 229. For the rest of the sites, subtract 3bp from each site beyond position 239 in pET-21a(+). *Nde* I does not cut pET-21d(+). Note also that *Sty* I is not unique in pET-21d(+).

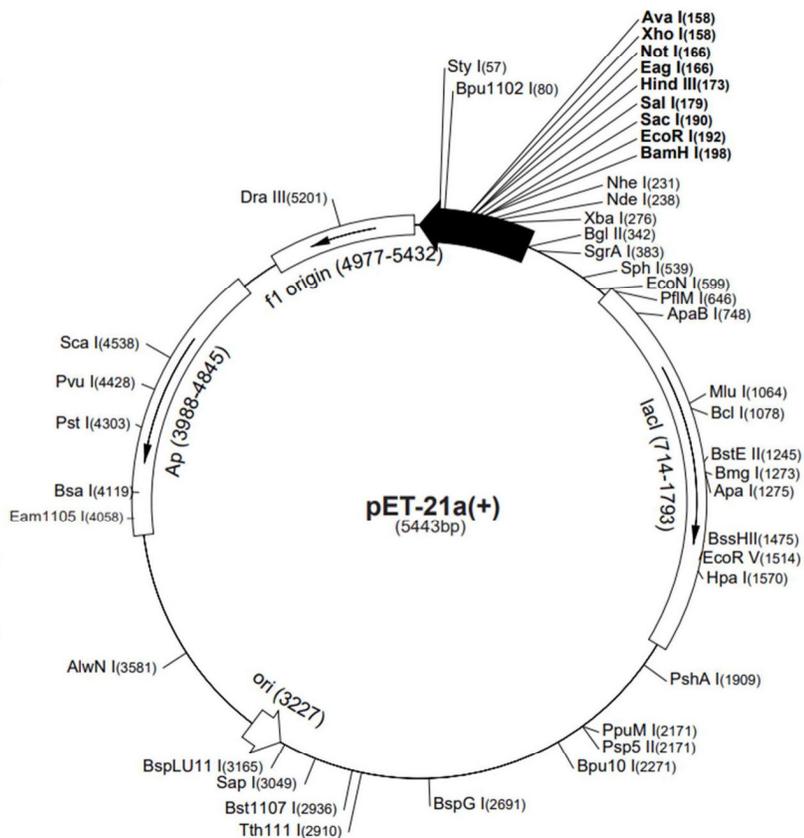


Figure 4. pET21b(+) (Novagen) expression vector map. Exceptions for pET21b(+) listed in terms of the original vector pET21a(+). Vector map shows size of the amplification vector; restriction endonuclease sites; and sequence landmarks.

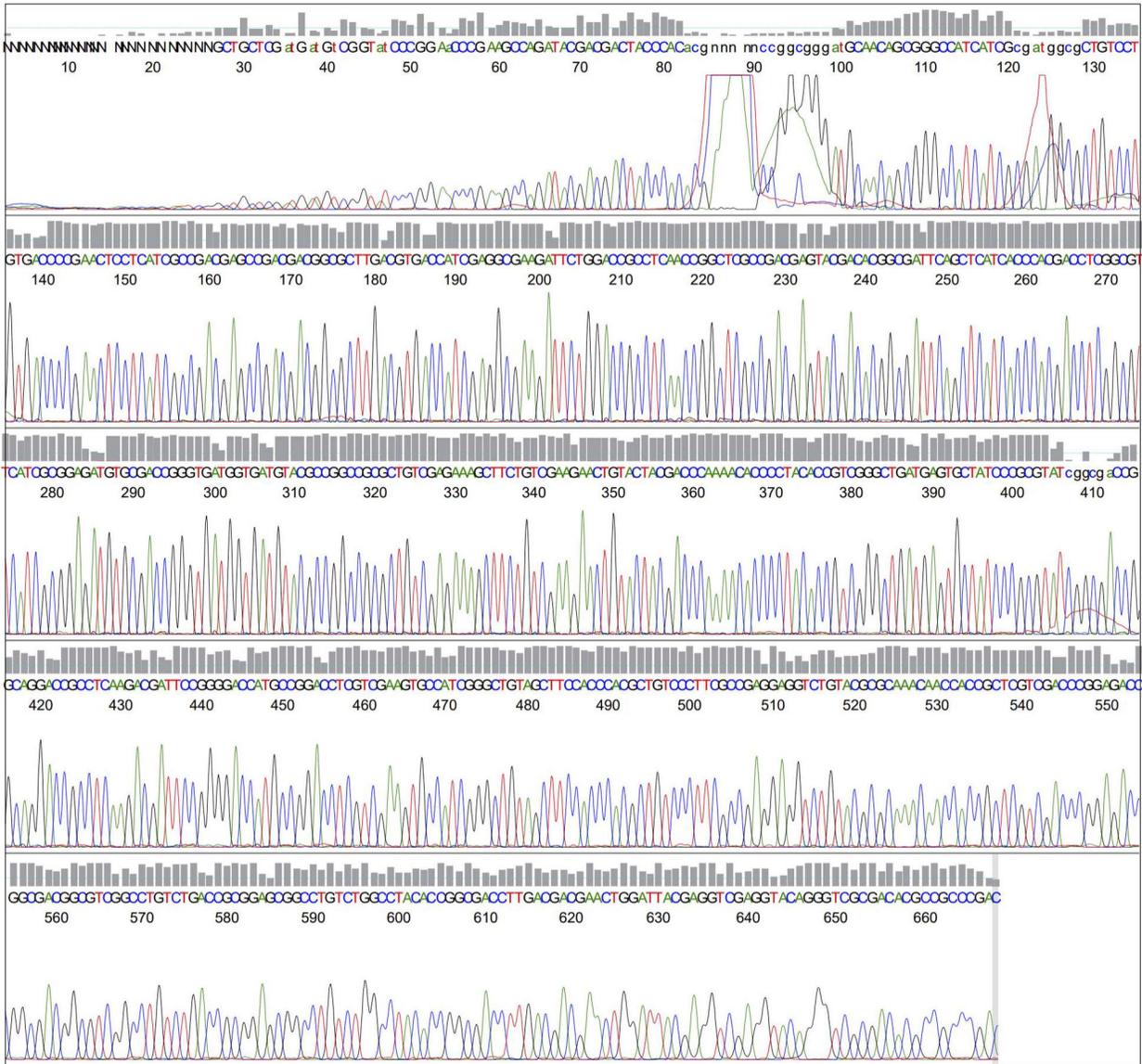


Figure 5. The *OppD1\_F\_internall* primer sequence of pET21b(+) indicating the presence of *OppD1* in the expression vector. Constructed with FinchTV.

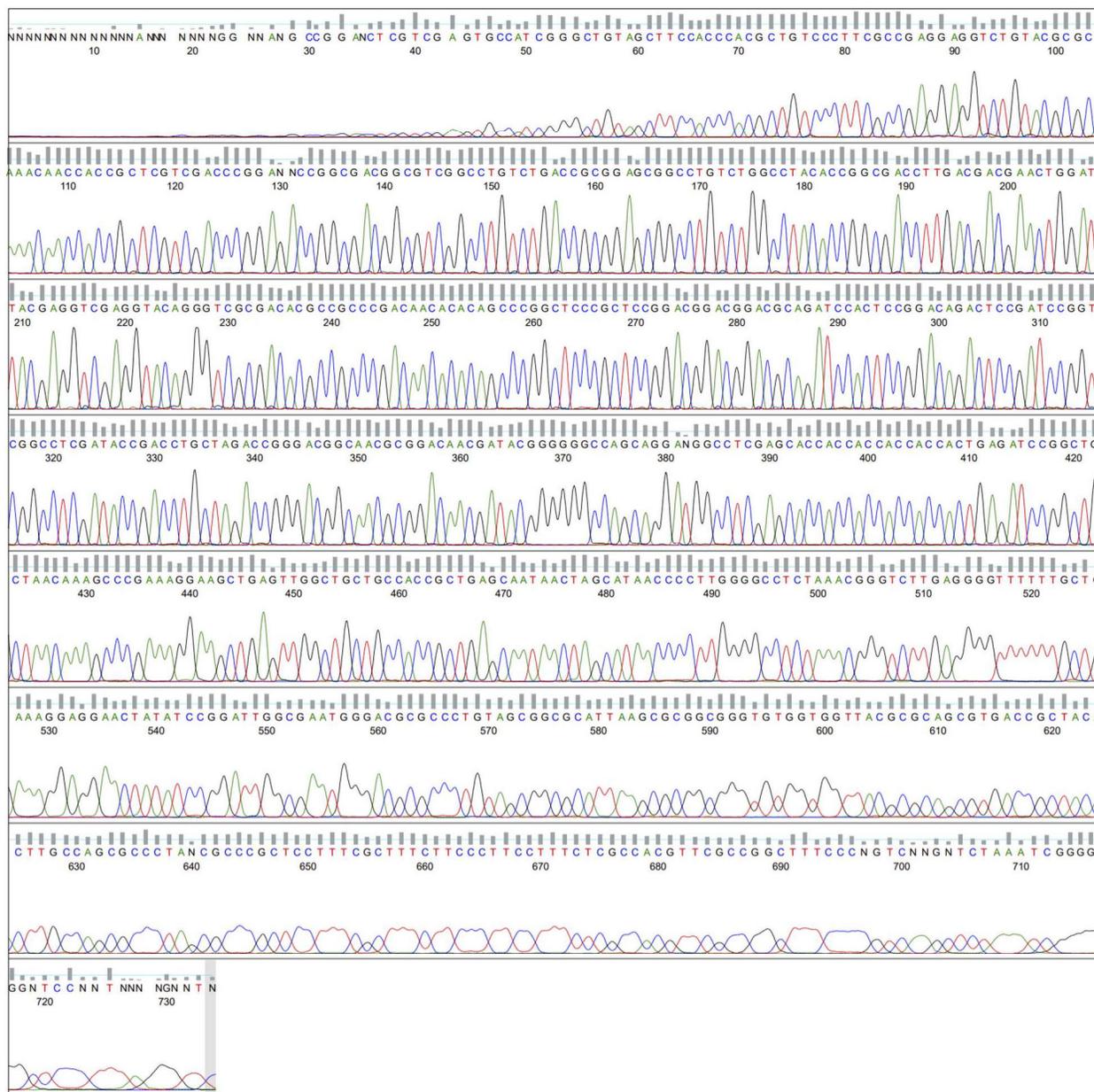


Figure 6. The *OppD1\_F\_internal2* primer sequence of pET21b(+) indicating the presence of *OppD1* in the expression vector. Constructed with FinchTV.



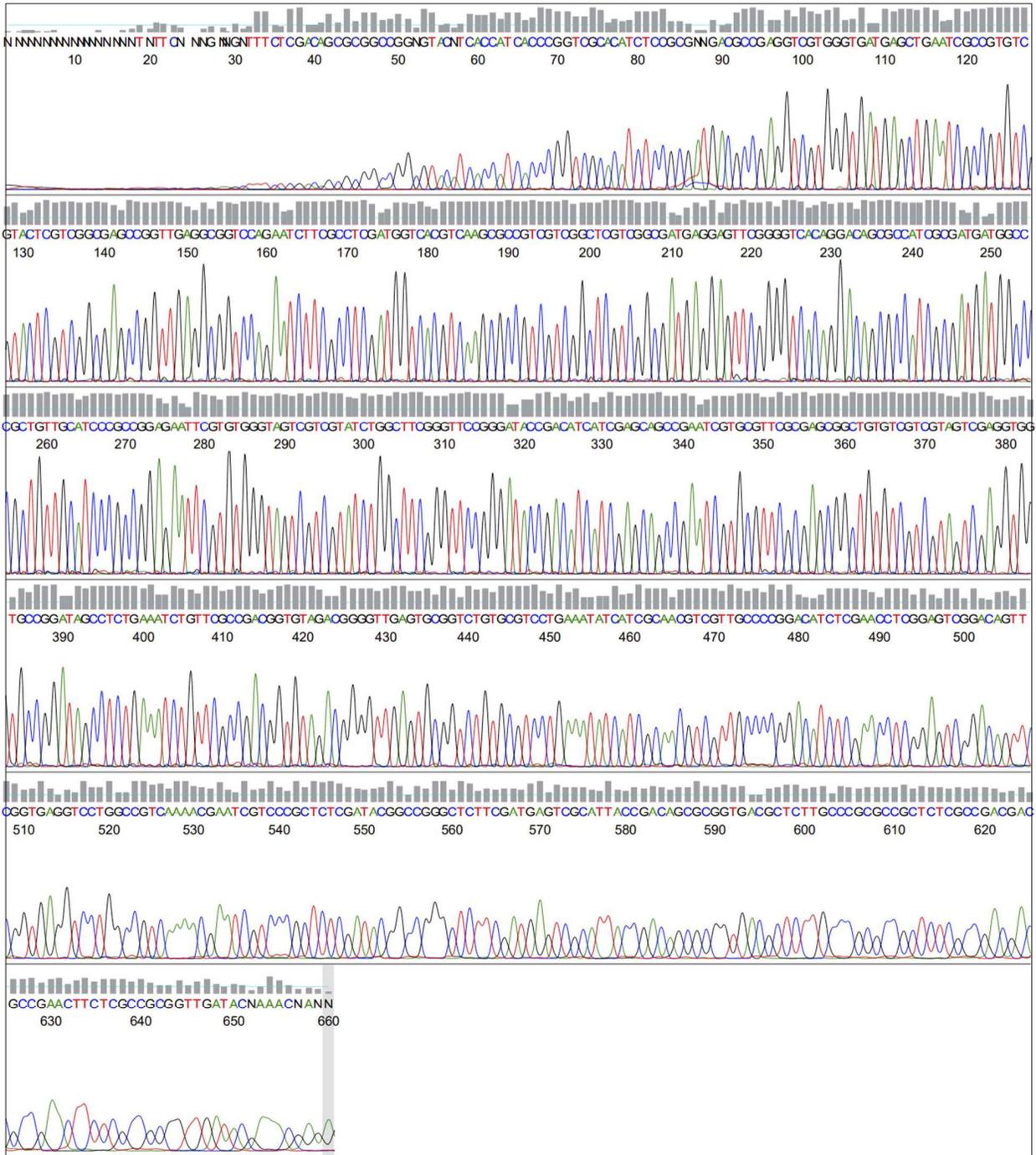


Figure 8. The *OppD1\_R\_internal2* primer sequence of pET21b(+) indicating the presence of *OppD1* in the expression vector. Constructed with FinchTV.

Figure 9. The following four pages contain the raw sequence data for the alignment of *OppDI* internal primers with the transcript inside *pET21b(+)-OppDI*.

2338 2419  
pET21b (+) ... ccctaaagggagccccgatttagagcttgacgggaaagccggcgaacgtggcgagaaaggaaggaagaaagcgaagga  
INTF2 ---aAaGGCAGCCCCCGATTTAGAgCttGACgGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGA  
INTF1 -----  
INTR2 -----  
INTR1 -----

2420 2501  
pET21b (+) ... gcggcgctagggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgcccgcgcttaatgcgcccgtacagg  
INTF2 GCGGGCGcTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCTTAATGCGCCGTACAGG  
INTF1 -----  
INTR2 -----  
INTR1 -----

2502 2583  
pET21b (+) ... gcgcgctcccattcgccaatccggatatagttcctcctttcagcaaaaaacccctcaagaccggttttagaggccccaaggggt  
INTF2 GCGCGTCCCATTGCGCAATCCGGATATAGTTCTCTCTTTCAGCAAAAAACCCCTCAAGACCCGTTTtagaggccccaaggggt  
INTF1 -----  
INTR2 -----  
INTR1 -----

2584 2665  
pET21b (+) ... tatgctagttattgctcagcgggtggcagcagccaactcagcttcctttcgggctttgttagcagccggatctcagtggtggt  
INTF2 TATGCTAGTTATTGCTCAGCGGTGGCAGCAGCCAACCTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGTTGGT  
INTF1 -----  
INTR2 -----  
INTR1 -----

2666 2747  
pET21b (+) ... ggtggtggtgctcagggccatcctgctggccccccgtatcgttgtccggttgcggtcccggctctagcaggtcggtatcgag  
INTF2 GGTGCTGGTGTCTCGAGGCCaTCCCTGCTGGCCCCCGTATCGTTGTCGCGGTTGCCGTTCCCGTCCCAGGCTAGCAGGTCTCGGTATCGAG  
INTF1 -----  
INTR2 -----  
INTR1 -----

2748 2829  
pET21b (+) ... gccgaccggatcggagctctgtccggagtgatctgcgtccgtccgctccggagcgggagccgggctgtgtgtgctggggcggc  
INTF2 GCCGACCCGGATCGGAGTCTGTCCGGAGTGGATCTGCGTCCGTCCGTCCGGAGCGGGAGCCGGGCTGTGTGTGTCGGGCGGC  
INTF1 -----GTCGGGCGGC  
INTR2 -----  
INTR1 -----

2830 2911  
pET21b (+) ... gtgtcgcgaccctgtacctcgacctcgtaatccagttcgtcgtcaaggtcgcgggtgtaggccagacagggcgtccgcggt  
INTF2 GTGTGCGGACCCGTACCTCGACCTCGTAATCCAGTTCGTGTCGTCAGGTCGCGGTTAGGCCAGACAGGCCGCTCCGCGGT  
INTF1 GTGTGCGGACCCGTACCTCGACCTCGTAATCCAGTTCGTGTCGTCAGGTCGCGGTTAGGCCAGACAGGCCGCTCCGCGGT  
INTR2 -----  
INTR1 -----

```

                2912                                                    2993
pET21b (+) ... cagacaggccgacgccgtcgccggtctccgggtcgacgagcgggtggttgtttgcgcgtaacagacctcctcgccgaagggaca
INTF2          CAGACAGGCCGACGCCGTCGCCGGTcTCCGGGTCGACGAGCGGTGGTTGTTTGC GCGTACAGACCTCCTCGGCGAAGGGACA
INTF1          CAGACAGGCCGACGCCGTCGCCGGTCTCCGGGTCGACGAGCGGTGGTTGTTTGC GCGTACAGACCTCCTCGGCGAAGGGACA
INTR2          -----
INTR1          -----

```

```

                2994                                                    3075
pET21b (+) ... gcgtgggtggaagctacagcccgatggcacttcgacgaggtccggcatgggtccccggaatcgctcttgaggcggctcctgccgg
INTF2          GCGTGGGTGGAAGCTACAGCCC GATGGCAC TCGACGAGgTCCGGC-----
INTF1          GCGTGGGTGGAAGCTACAGCCC GATGGCAC TCGACGAGGTCCGGCATGGTCCCCGGAATCGTCTTGAGGCGGTCCCTGCCGG
INTR2          -----
INTR1          -----

```

```

                3076                                                    3157
pET21b (+) ... tcgccgatacgcgggatagcactcatcagcccagcggtaggggtgttttgggtcgtagtacagttcttcgacagaagcct
INTF2          -----
INTF1          tcgccgATACGCGGGATAGCACTCATCAGCCC GACGGTGTAGGGGTGTTTGGGTCGTAGTACAGTCTTCGACAGAAGCTT
INTR2          -----GTT
INTR1          -----

```

```

                3158                                                    3239
pET21b (+) ... tctcgacagcgcggccggcgtacatcaccatcaccgggtcgacatctccgcatgacgcccagggtcgtgggtgatgagctg
INTF2          -----
INTF1          TCTCGACAGCGCGCCGGCGGTACATCACCATCACC GGTTCGCACATCTCCGCGATGACGCCGAGGTCGTGGGTGATGAGCTG
INTR2          TCTCGACAGCGCGCCGGcGTAC TACCATCACC GGTTCGCACATCTCCGCGatGACGCCGAGGTCGTGGGTGATGAGCTG
INTR1          -----

```

```

                3240                                                    3321
pET21b (+) ... aatcgccgtgctgactcgtcgccgagccggttgaggcgggtccagaatcttcgcctcgatggtcacgtaagcggcctcgtc
INTF2          -----
INTF1          AATCGCCGTGTCGTACTCGTCGGCGAGCCGGTTGAGGCGGTCCAGAATCTTCGCCTCGATGGTCACGTCAAGCGCCGTCGTC
INTR2          AATCGCCGTGTCGTACTCGTCGGCGAGCCGGTTGAGGCGGTCCAGAATCTTCGCCTCGATGGTCACGTCAAGCGCCGTCGTC
INTR1          -----

```

```

                3322                                                    3403
pET21b (+) ... ggctcgctcggcgatgaggagttcggggtcacaggacagcggcatcgcgatgatggcccgtggtgcatcccggcggagaatt
INTF2          -----
INTF1          GGCTCGTCGGCGATGAGGAGTTCGGGGTCACAGGACAGcggcatcgCGATGATGGCCCGCTGTTGCatcccggcgg nnnnn
INTR2          GGCTCGTCGGCGATGAGGAGTTCGGGGTCACAGGACAGCggcatCGCGATGATGGCCCGCTGTTGCATCCCggcggagaatt
INTR1          -----

```

```

                3404                                                    3485
pET21b (+) ... cgtgtgggtagtcgctgatctggttcgggttcgggataccgacatcatcgagcagccgaatcgctgcttcgagagcggc
INTF2          -----
INTF1          cgtGTGGGTAGTCGTCGTATCTGGCTTCGGGTTCGGGatACCGaCatCatCGAGCAGC-----
INTR2          CGTGTGGGTAGTCGTCGTATCTGGCTTCGGGTTCGGGATACCGACATCATCGAGCAGCCGAATCGTGCgtTCGCGAGCGGC
INTR1          -----

```

3486 3567  
pET21b (+) ... tgtgtcgtcgtagtcgaggtggtgccggatagcctctgaaatctggttcgccgacggtgtagacggggtgagtgccggtctgt  
INTF2 -----  
INTF1 -----  
INTR2 TGTGTCGTCGTAGTCGAGGTGGTGCCGGATAGCCTCTGAAATCTGTTCCGCCGACGGTGTAGACGGGGTTGAGTGGCGTCTGT  
INTR1 -----CTGT

3568 3649  
pET21b (+) ... gcgctctgaaatatcatcgcaacgctggtgccccggacatctcgaacctcggagtcggacagttcgggtgaggtcctggccgt  
INTF2 -----  
INTF1 -----  
INTR2 GCGTCTGAAATATCATCGCAACGTCGTTGCCCGGACATCTCGAACCTCGGAGTCGGACAGTTCGGTGGAGTCTGGCCGT  
INTR1 gcgTCCTGAAATaTcATCGCaacGTCGTTGCCCGGACATCTCGAACCTCGGAGTCGGACAGTTCGGTGGAGTCTGGCCGT

3650 3731  
pET21b (+) ... caaaacgaatcgtcccgtctcgcgatacggccgggctcttcgatgagtcgcattaccgacagcgcgggtgacgctcttgccccg  
INTF2 -----  
INTF1 -----  
INTR2 CAAAACGAATCGTCCCGTCTCGATACGGCCGGGCTCTTCGATGAGTCGCATTACCACAGCGCGGTGACGCTCTTGGCCGC  
INTR1 CAAAACGAATCGTCCCGTCTCGATACGGCCGGGCTCTTCGATGAGTCGCATTACCACAGCGCGGTGACGCTCTTGGCCGC

3732 3813  
pET21b (+) ... gccgtctcgcgccgacgacgccgaactctcgcgccggttgatacgaacgagaggtcgtcgacggcggtgacaacgcctcc  
INTF2 -----  
INTF1 -----  
INTR2 GCCGCTCTCGCCGACGACGCCGAACCTCTCGCCCGGTTGATACNAACNAga-----  
INTR1 GCCGCTCTCGCCGACGACGCCGAACCTCTCGCCCGGTTGATACGAAACGAGAGGTCGTCGACGGCGGTGACAACGCCTCC

3814 3895  
pET21b (+) ... tcgggtgtagaagttcaccgtgaggttctcgacttcgagcagcgcattatgtatatctccttcttaaagttaaacaaaattat  
INTF2 -----  
INTF1 -----  
INTR2 -----  
INTR1 TCGGTGTAGAAGTTCACCGTGAGGTTCTCGACTTCGAGCAGCGCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTAT

3896 3977  
pET21b (+) ... ttctagaggggaattggtatccgctcacaattcccctatagtgagtcgtattaatttcgcgggatcgagatctcgatcctct  
INTF2 -----  
INTF1 -----  
INTR2 -----  
INTR1 TTCTAGAGGGGAATTGTTATCCGCTCACAATCCCCTATAGTGAGTCGTATTAATTCGCGGGATCGAGATCTCGATCCTCT

3978 4059  
pET21b (+) ... acgccggacgcatcgtggccgcatcaccggcgccacaggtgcggttgctggcgcctatatcgccgacatcaccgatgggga  
INTF2 -----  
INTF1 -----  
INTR2 -----  
INTR1 ACGCCGACGACATCGTGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCTATATCGCCGACATCACCAGTGGGGA

Page 4 INTR1

```
4060                                     4141
pET21b (+) ... agatcgggctcgccacttcgggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtggccgggggactggtg
INTF2 -----
INTF1 -----
INTR2 -----
INTR1 AGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTG
```

.....

```
4142                                     4219
pET21b (+) ... ggcgccatctccttgcatgcaccattccttgcgggcggtgctcaacggcctcaacctactactgggctgcttccta
INTF2 -----
INTF1 -----
INTR2 -----
INTR1 GCGCCATCTCCTTGCATGCACCATTCTTTCGGCGGGCGGTGCTCAACGGCCTCAACCTACTACTg-----
```

.....

326 – TGDLDELDY EVEVQGRDTP PDNTQPGSRS GRTDADPLRT  
DSDPVGLDTD LLDRDGNADN DTGGQQDG – 393

Figure 10. Amino acid sequence of the unordered portion of OppD1 not aligned with sapD. Sequence shows amino acids # 326 to 393, which are separated into groups of 10 amino acids and the last grouping containing 8 amino acids.