UNIVERSITY OF CALIFORNIA, SAN DIEGO

Comparative Analysis of Piezophilic Bacteria: The Search for Adaptations to Life in the Deep Sea

A thesis submitted in partial satisfaction of the requirements for the degree of Master of Science

in

Biology

by

Ian McKenzie Kerman

Committee in charge:

Professor Douglas H. Bartlett, Co-Chair
Professor Eric E. Allen, Co-Chair
Professor Milton H. Saier

2008

The Thesis of Ian McKenzie Kerman is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

Co-Chair

_____

Co-Chair

University of California, San Diego

2008

Epigraph

Here too are dreaming landscapes,
lunar, derelict.
Here too are the masses,
tillers of the soil.
And cells, fighters,
who lay down their lives
for all the world.

Here too are cemeteries,
fame, snow and spates.
And I hear murmuring,
the revolt of immense estates.

Miroslav Holub
In the Microscope

Table of Contents

List of Abbreviations

ΔV	change in volume
BLAST	Basic Local Alignment Search Tool
bp	base pair
COG	cluster of orthologous groups
DHA	docosahexaenoic acid
DNA	deoxyribonucleic acid
EC	Enzyme Commission
EPA	eicosapentaenoic acid
KEGG	Kyoto encyclopedia of genes and genomes
IMG	Integrated Microbial Genomes
ITS	internal transcribed spacer
mRNA	messenger RNA
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
ORF	open reading frame
PCR	polymerase chain reaction
PUFA	polyunsaturated fatty acid
RNA	ribonucleic acid
RP	ribosomal protein
rRNA	ribosomal RNA
RTX	repeat in toxin
SRP	signal recognition particle
tRNA	transfer RNA

List of Tables

List of Graphs

List of Figures

Acknowledgements

This thesis is the accumulation of the efforts of the many people who supported me throughout.

My advisor, Doug Bartlett, taught me to craft a story from the data, rather than simply presenting it to present it. Of all the skills that I have learned during my education, this is likely to be one of the skills that best helps me further my career in science.

I would also like to thank my other committee members. Eric Allen has always been generous with his advice and support and Milton Saier is an excellent teacher who has inspired me since I was an undergraduate.

My mentor upon joining the Bartlett Lab, Federico Lauro, inspired through example and was always supportive, despite my odd working hours.

The other members of the Bartlett Lab, Alix, Emiley, and Taylor, were the spices that added flavor to working in lab. They provided good conversation and support throughout my time in the lab.

Without my close friends, I am confident that I would be nowhere near to where I am today. To Ben, Desiré and Gary, I owe you a beer.

I would lastly like to thank my family who have supported me throughout my life and encouraged me to pursue my passions, whatever they might be.

ABSTRACT OF THE THESIS

Comparative Analysis of Piezophilic Bacteria: The Search for Adaptations to Life in the Deep Sea

by

Ian McKenzie Kerman

Master of Science in Biology

University of California, San Diego, 2008

Professor Douglas H. Bartlett, Co-Chair

Professor Eric E. Allen, Co-Chair

Life in the deep sea presents organisms with a number of challenges that must be overcome through evolutionary adaptation. The most unique of these challenges is that of high hydrostatic pressure. This stress has the potential to affect a number of processes inside the cell and poses a thermodynamic constraint on reactions that involve changes in volume. A number of piezophilic adaptations have been found and more are reported in

this thesis. One such adaptation that has been previously explored is that of sequence insertions in the 16S ribosomal RNA. Here, the rest of the ribosome is analyzed and it is found that the 23S ribosomal RNA also has piezophilic-specific insertions. Mutations are also found in some ribosomal proteins and strong evidence is found for ribosomal RNA operon recombination in *Photobacterium profundum* SS9.

The volume change constraint was explored directly using a method that predicts protein folding volume. Applied to pairs of piezophiles and mesophiles, trends in folding volume for select proteins were examined. It was concluded that the method could be useful, but further changes to the algorithm are necessary before conclusions can be drawn.

Lastly, the genome of *Moritella* sp. PE36 has been sequenced and partially closed. Analysis of the draft genome annotation reveals increased proportions of genes encoding proteins involved in cell motility, metabolism, and secretion relative to that present in related non-piezophilic bacteria. This last gene product category increase is likely to support the surprisingly high number of proteins predicted to have signal peptides.

Introduction

The term "piezophile" has been adopted for microbes that grow optimally at hydrostatic pressures greater than 0.1 MPa, or nearly atmospheric pressure. Piezophile is a combination of the Greek roots "piezo" meaning "to press" and "philo" meaning "to love" (Yayanos *et al.*, 1995). These organisms will be the focus of this thesis as I expand upon our understanding of the unique adaptations necessary for their survival in the deep sea.

Oceans cover approximately 70% of the Earth's surface and about 94% of the ocean is at a depth greater than 200m (Whitman *et al.*, 1998). Although life at these depths faces a number of challenges including low temperature, high hydrostatic pressure and the absence of sunlight (Bartlett, 1992), it is estimated that prokaryotic life at this depth accounts for over 55% of all prokaryotic life on Earth (Whitman *et al.*, 1998). Despite this large occupancy of the Earth's biosphere, life at these depths remains relatively unexplained, which is partially a consequence of the difficulty encountered in the study of life at deep ocean depths (Lauro and Bartlett, 2008).

Sample collection involves navigating collection equipment to depths of up to 10,500 meters and once there, the equipment must be capable of withstanding pressures in excess of 100 MPa (Yayanos *et al.*, 1981). ZoBell and Morita first reported the discovery of deep sea, high hydrostatic pressure-obligate microbes in 1957, but it was not until 1979 that Yayanos et

al. cultured a microbe from the deep sea. Once in the laboratory, high-pressure equipment is also needed to maintain and culture the samples. Meta and single cell genomics is improving the ability for communities and the genomes of more organisms to be studied and compared, but the isolation and culturing of species will still be necessary for genetic and biochemical studies (Lauro and Bartlett, 2008).

High hydrostatic pressure has the potential to affect many aspects of an organism's genome, proteome and cellular structure. In 1954, ZoBell and Morita showed that the activity of the aspartase enzyme of *Escherichia coli* increased in pressures up to 640 atm, but was decreased at higher pressures and inactivated at 1000 atm (1 atm = 0.101 MPa). The succinic dehydrogenase system in *E. coli* was found to be even more pressure sensitive, its activity decreasing at 200 atm (Morita and ZoBell, 1955). Outside of these specific cases, hydrostatic pressure has been shown to interfere with the cellular functions of non-deep sea microbes (Yayanos and Pollard, 1969; Welch *et al.*, 1993; Ishii *et al.*, 2005).

Protein-protein (Silva and Weber, 1993) and protein-DNA (Chilukuri *et al.*, 1997) interactions have also been shown to be affected by hydrostatic pressure. Processes that involve a positive change in volume will be hindered or stopped under high hydrostatic pressure (Somero, 1990). Protein structures that have open cavities can become compressed (Mozhaev *et al.*,

1996) as can the hydration shell surrounding hydrophobic groups in proteins and phospholipids (Kauzmann, 1959).

Yet little is known about the adaptations necessary for microbes to overcome these challenges (Taylor, 2008). It is unknown whether piezophily requires changes in just a few genes, the entire genome, or regulatory modulations (Simonato *et al.*, 2006), however, it has been concluded through comparative genomics that there is no single gene that furnishes a microbe with the ability to survive in the deep sea (Lauro, 2007). Similar to psychrophiles, deep sea microbes have high concentrations of polyunsaturated fatty acids, which have been shown to be important for growth at high hydrostatic pressure (Delong and Yayanos, 1985; Allen *et al.*, 1999; Allen and Bartlett, 2000). There are a number of membrane-associated proteins that have been associated with high hydrostatic pressure adaptation including cytochromes, outer membrane proteins L and H, ToxR, and transport proteins (Simonato *et al.*, 2006). The flagellum of *E. coli*, also a membrane-associated protein, is affected by high hydrostatic pressure in two different ways: (1) rotation of the flagellar filament is hindered or stopped; and (2) the polymerization of the flagellar filament is affected (Meganathan and Marquis, 1973). One piezophile, *Photobacterium profundum* SS9, has an extra flagellar gene cluster compared to its non-piezophilic relative *P. profundum* 3TCK (Campanaro *et al.*, 2005), however, the exact function of this extra gene cluster remains unknown.

Comparative genomics offers an exciting opportunity to examine piezophile genomes and compare them to their closely related shallow water relatives, mesophiles. Using the results of comparative genomics, global and singular differences as well as similarities between strains can be ascertained and marked for genetic or biochemical testing. Such differences and similarities can include homologous genes, gene duplications, genes exclusive to one strain, protein domain comparison and protein family comparisons (Rubin *et al.*, 2000). Genomic hybridization experiments were performed on the SS9 and 3TCK strains, as well as a second piezophile, *P. profundum* DSJ4, to search for genes that were exclusive to the two piezophiles (Simonato *et al.*, 2006). A comprehensive comparison was done between 15 proteins in six pairs of piezophiles and mesophiles, resulting in the conclusion that more genomes are needed to be able to statistically verify any of the findings as being significant (Taylor, 2008).

In this thesis, I expand on the use of comparative genomics in studying piezophiles. In the first chapter, I examine the ribosomal components in five different γ-proteobacteria piezophiles and make comparisons to closely related mesophiles. In the second chapter, a method to model the protein folding volume is introduced. This method is then used to expand upon previous work done by Taylor (2008), using the same 15 proteins that she did for analysis. Lastly, the third chapter describes work done to close the

genome of the piezophile, *Moritella* sp. PE36. A preliminary analysis of the genome is also performed.

Chapter 1:

Comparative Analysis of Piezophilic Ribosomes

Introduction

The ribosome is a ubiquitous macromolecular machine found in all modern living cells. It is composed of both RNA (known as ribosomal RNA or rRNA) and proteins. Although proteins are typically associated with catalyzing reactions in the cell (as enzymes), the ribosome is in fact a ribozyme, with the rRNA catalyzing the formation of the peptide bond between amino acids (Steitz and Moore, 2003).

Catalyzing the peptide bond and thus the formation of proteins, the ribosome is chiefly responsible for the last step in the Central Dogma of molecular biology: the translation of RNA into protein (Crick, 1970). It is not alone in doing this however; a number of factors come together to assist and ensure that translation occurs smoothly (Noller, 1991). A number of components are required to initiate, continue and complete translation, but the main components in translation are the mRNA, which serves as the template for the protein's primary sequence and the tRNAs that decode the mRNA into amino acids through codon–anticodon pairing (Noller, 1991).

The bacterial ribosome is different from the eukaryotic or archaeal ribosome in structure and exact function, though its role in translation is the

same (Wool, 1979; Bell and Jackson, 1998). Bacterial ribosomes are made up of two subunits, designated 30S and 50S, called the small and large subunit, respectively (Steitz and Moore, 2003). The catalytic site of the ribosome is located in the large subunit (Steitz and Moore, 2003). The small subunit serves to bind the mRNA and includes the site where the mRNA is decoded (Steitz and Moore, 2003). The large subunit contains two rRNAs, the 23S and the 5S (Steitz and Moore, 2003). It is the 23S rRNA that contains the catalytic residues necessary for peptide bond formation (Steitz and Moore, 2003). In *Escherichia coli*, the large subunit is made up of 34 proteins in addition to the 23S and 5S rRNAs (Steitz and Moore, 2003), while its small subunit is made up of the 16S rRNA and has 21 proteins associated with it (Steitz and Moore, 2003).

The ribosomal RNAs are usually, though not always, transcribed from DNA as a single, large molecule (Dunn and Studier, 1973; Srivastava and Schlessinger, 1990). This single transcript then goes through a number of processing steps, resulting in the mature 23S, 16S, 5S rRNAs and zero or more tRNAs (Dunn and Studier, 1973; Srivastava and Schlessinger, 1990). The cleavage of the separate products is carried out by a number of ribonucleases including RNase T, RNase E, RNase III, and RNase P, which is itself a ribozyme (Nicholson, 1999). RNase P and RNase T are also involved in the maturation of another RNA molecule, the 4.5S RNA (Nicholson, 1999). The 4.5S RNA is part of the signal recognition particle (SRP), which binds to

a protein on the large subunit and mediates the targeting of the ribosome to the cell membrane (Gu *et al.*, 2003).

The configuration of the rRNA operons can vary from species to species and often even varies within a single species genome if more than one operon is present (Srivastava and Schlessinger, 1990). Typically, however, the 16S rRNA is followed by the 23S rRNA, which is then followed by the one or more copies of the 5S rRNA (Srivastava and Schlessinger, 1990). Zero or more tRNA sequences may be present between the 16S and 23S rRNAs and may also be present preceding or following the 5S rRNA(s) (Srivastava and Schlessinger, 1990). In *E. coli,* upstream of the 16S rRNA by about 190 bp, there are two promoters P2 and P1 (Srivastava and Schlessinger, 1990). These promoters are separated by approximately 120 bp (Srivastava and Schlessinger, 1990) and it is the P1 promoter that is responsible for the majority of transcription (Condon *et al.*, 1995).

The ribosome and the translation of proteins are affected in various ways by both low temperature and high hydrostatic pressure, two environmental conditions encountered by piezophiles (Lauro *et al.*, 2007; Bartlett, 1992). Three proteins produced in response to cold shock are known to associate with the ribosome specifically (Thieringer *et al.*, 1998). There are also a number of proteins that bind mRNA, which are thought to destabilize problematic secondary structures in the mRNA (Thieringer *et al.*, 1998). Under high hydrostatic pressure, reactions causing an overall increase in the

volume of the system are disfavored (Somero, 1990). For proteins under high hydrostatic pressure, there is a positive and therefore unfavorable change in volume when the protein folds (Harpaz *et al.*, 1994). High hydrostatic pressure can also affect other macromolecules and complexes that have internal cavities (Mozhaev *et al.*, 1996).

There is a considerable potential for the ribosome to be negatively affected by the environmental conditions encountered by piezophiles. It has been shown that mesophile ribosomes are sensitive to hydrostatic pressure (Landau, 1967; Gross and Jaenicke, 1990), and the piezosensitivity of the *E. coli* ribosome has been linked to the dissociation of the ribosome (Niven *et al.*, 1999). The ribosome relies extensively on the secondary structure of its rRNA components, which could be affected by the low temperature as evidenced by the proteins of the cold shock response. The ribosome's superstructure has internal cavities where translation takes place and there is potential for these to be compressed by the high hydrostatic pressure. Additionally, the proper folding of the rRNA and ribosomal proteins (RPs) into their respective three-dimensional structures could be inhibited by positive changes in volume. It is with these problems in mind that I examine the various components of the piezophilic ribosome.

Methods

To determine piezophilic-specific mutations, piezophilic sequences were compared to closely related mesophilic sequences. All organisms are

from the γ-proteobacteria class. Table 1.1 shows the piezophile-mesophile pairings that are used throughout this study for comparisons. Ribosomal RNA sequences were extracted from genomic sequences (Table 1.2) using RNAmmer v1.2 (Lagesen *et al.*, 2007). A score threshold of 70 was used to filter the RNAmmer results. Transfer RNA sequences were extracted using tRNAscan-SE v1.4 (Lowe & Eddy, 1997). Ribosomal protein sequences were found using the RP sequences from *Escherichia coli* K12 as the query sequence in a BLAST search of the genomes.

Sequence alignments were done using MUSCLE v3.7 (Edgar, 2004) and two-dimensional structure predictions of RNA were done using UNAFold v3.5 (Markham & Zuker, 2008). Default parameters were used for structure prediction, except for the temperature which was lowered to 15° C (from 37° C) for piezophiles and psychrophiles. Alignments were viewed and phylogenetic trees built with Jalview v2.3 (Clamp *et al.*, 2004). Two-dimensional structure comparisons were done visually and when necessary, were compared to the reference structure diagram of the appropriate molecule from the Comparative RNA Web Site and Project (Cannone *et al.*, 2002) (http://www.rna.ccbb.utexas.edu).

Three-dimensional visualization and analysis was done using the UCSF Chimera software package (Pettersen *et al.*, 2004).

16S Ribosomal RNA

Lauro *et al.* (2007) investigated the 16S rRNA in piezophiles and found that there is a strong correlation between piezophily and elongation of helices in the 16S rRNA. Insertions were found in three helices: 10, 11, and 44[1] (Lauro *et al.*, 2007). *Colwellia* and *Photobacterium* species primarily had insertions in helix 10 whereas *Shewanella* exhibited an insertion in helix 11 (Lauro *et al.*, 2007). *Colwellia* sp. MT41 has an insertion in helix 10 in all of its nine 16S rRNA genes and *Photobacterium profundum* SS9 has the insertion in 12 of its 15 16S rRNA genes. Two copies of *Photobacterium profundum* SS9's 16S rRNA gene had insertions in helix 11. *Shewanella benthica* KT99 has an insertion in both of its identified 16S rRNA genes. The insertion in helix 44 is only found in four of *Photobacterium profundum* SS9's 16S rRNA genes, one of which also has an insertion in helix 11 and all of which have an insertion in helix 10.

Photobacterium profundum SS9 also had an insertion into helix 44, though only in four 16S rRNA genes and not in a pattern corresponding to the other insertions (Lauro *et al.*, 2007). In fact, none of the insertions were strictly correlated with each other in *Photobacterium profundum* SS9 (Lauro *et al.*, 2007). Based on the presence or absence of these insertions, five

[1] Lauro *et al.* (2007) identify the insertion being in helix 49; however Brimacombe (1995) labels that helix as #44, which is the numbering that is used here.

ribotypes were identified in *Photobacterium profundum* SS9 (Lauro *et al.*, 2007). Despite having different ribotypes, it was found that regardless of the pressure the cells are grown at, all of the ribotypes are constitutively expressed (Lauro *et al.*, 2007).

Helices 10, 11, and 44 have all been implicated in interacting with RP S20 (Firpo and Dahlberg, 1998; Cormack and Mackie, 1991; Stern *et al.*, 1989). The S20 protein is necessary for ribosome assembly (Stern *et al.*, 1989) and translation initiation (Götz *et al.*, 1990). Based on these interactions, it is possible that the insertions in the 16S rRNA of piezophilic bacteria stabilize or otherwise modify the interaction between the 16S rRNA and S20.

Using a three-dimensional model of *Escherichia coli*'s 30S ribosomal subunit (Schuwirth *et al.*, 2005; PDB ID: 2AVY), I examined the positioning of the insertions relative to each other and S20.

Figure 1.1 shows a ribbon rendering of the portion of the small subunit containing the helices of interest. Helix 10, 11, and 44 are highlighted as are RPs S16 and S20. On helices 10 and 11, the points of insertion are highlighted in green. From this diagram, the closeness of helix 10 and S20 is evident. Figure 1.2 further shows that residues of helix 10 and S20 come to within 6.107 Å of each other. This measurement is likely incorrect since the resolution of the structure is only 3.46 Å, however it is still useful in illustrating the closeness of the two molecules. The extension of this helix, as

is seen in some piezophiles, is likely to alter its interaction with S20, though to what extent is unknown.

Figure 1 also shows that insertions in helix 11 are unlikely to affect S20 due to their separation in the structure. However, helix 44 is in close proximity to S20 and its extension, though documented in other non-piezophiles (Cannone *et al.*, 2002) (http://www.rna.ccbb.utexas.edu), could have an impact on S20. Figure 1.3 is a surface rendering of the region, and again shows the close association of helix 10 and S20, but also shows potential interaction between helix 44 and S20.

This analysis confirms the determination by Lauro *et al.* (2007) that insertions in helix 10 may alter interactions with RP S20. Evidence towards an alternative interaction between S20 and helix 44 is also demonstrated, though doubt is cast on helix 11's role. Since insertions in helix 11 are found in 16S sequences lacking insertions in helix 10, it is possible that the helix 11 insertions cause the same effect as the helix 10 insertions, but through a different mechanism.

23S Ribosomal RNA

The 23S ribosomal RNA is the main component of the large ribosomal subunit and contains the catalytic peptidyl transferase center. The 23S has been divided into six domains (I – VI) based on secondary structure (Cannone *et al.*, 2002). It is thought, and has been partially shown, that the domains of

the 23S evolved at different times, with domain V, which contains the catalytic center, being the oldest (Hury *et al.*, 2006). These domains have evolved over time to interact with each other and RPs to form the large ribosomal subunit in its present structure. In this study, the 23S rRNAs were examined to determine if, like the 16S rRNA, it has been modified for life in the deep sea.

The alignment of piezophile and mesophile 23S rRNA sequences reveal four regions where the piezophilic sequence differs considerably from the paired mesophilic sequence. Figures 1.4, 1.5, and 1.6 depict the regions of variability in the 23S. These differences were all in the form of insertions into the RNA ranging from 10 to 17 bp. They were found to occur in three piezophiles; *Colwellia* sp. MT41, *Shewanella benthica* KT99 and *Photobacterium profundum* SS9 (hereafter referred to as MT41, KT99, and SS9, respectively). Insertion #4 was also found in the comparison strain *Shewanella frigidimarina.*

For the analysis of these insertions, the conservation diagrams from the Comparative RNA Web Site and Project (CRW) were used (Cannone *et al.*, 2002). Sequence data of γ-proteobacteria for determination of insertion uniqueness was also obtained from the CRW and was supplemented with sequence from version 2.4 of the Integrated Microbial Genomes (IMG) system (Markowitz *et al.*, 2006). Helix numbering is the same as in Mueller *et al.*

(2000). Three-dimensional structure analysis was done with the *Escherichia coli* PDB structure 2AW4 (Schuwirth *et al.*, 2005).

Insertion #1

Figure 1.4 shows two insertions, the first one occurring in SS9 and KT99 which is 14 bp and 13 bp in length, respectively. Relative to their mesophile pairs, the SS9 insertion is at position 537 and the KT99 insertion is at 534. This places the insertion in helix 25, as illustrated in Figures 1.7 and 1.8. In the γ-proteobacteria, helix 25 is 20-28 bp long (10-14 bp per strand) capped off with a tetraloop. With the exception of *Psychromonas ingrahamii*, the helices of the mesophiles and piezophiles in this study without the insertion were predicted not to have any bulges or internal loops (There are two non-canonical G:A pairs at positions 5 and 6 in the helix).

The *E. coli* structure is 28 bp long with a helical length of 12 (Figure 1.9). Helix 25 in *Photobacterium profundum* 3TCK has a length of 27 bp with a helical length of 11. In *Shewanella frigidimarina*, helix 25 is 29 bp long and has a helical length of 12. This is in contrast to SS9 and KT99 whose helices have lengths of 41 and 42 bp, respectively. In an alignment with 281 γ-proteobacteria, insertions of this length in helix 25 were found only in KT99 and SS9. What appeared to be insertions in other species were either much shorter (1-4 bp) or much longer (~100 bp).

As shown in Figure 1.9, both of the piezophiles' helices are elongated and internal loops have been introduced in both structures. Interestingly, there is a single unpaired cytosine residue in the same position relative to the terminal loop in both structures. This residue is part of the internal loop found on both structures, though the loops are of different sizes.

Helix 25 projects from the top 50S subunit and is surrounded by a number of other helices and proteins (Figures 1.10 and 1.11). Figure 1.10 shows the special arrangement of helix 25 relative to helices 2, 46, and 98 and RPs L13, L20 and L21. Through crosslinking studies, helix 25 has been shown to interact with L13 and L21 (Brimacombe, 1995).Helix 46 has also been shown to interact with L21 (Brimacombe, 1995). Figure 11 shows a surface rendering of the same area, from which physical interactions are visible through the color interfaces. Helix 25 is seen touching helix 2, RP L20 and L13. Both helix 98 and RP L21 do not touch helix 25 in the crystal structure, however this may change in a fluid environment.

The point of insertion (in green) is not in direct contact with any other structures, so the insertion is not likely to directly alter any interactions. The current interactions with helix 2, RP L20 and L13 are not likely to be changed. However, through the insertion's elongation of the helix and creation of a bulge in the structure, helix 25 may be brought into contact with helix 98. Interaction with RP L21 is also possible, especially since the helix

would be relatively unrestricted in its motion, allowing it to bend towards L21.

Insertion #1 is found in two piezophiles and is a unique or mostly unique structural feature within the γ-proteobacteria. It has potential to cause increased interaction with other ribosomal components. It is also possible, because it will stick out beyond the surface of the ribosome, that the insertion may interact with other cellular components increasing stability or efficiency of the ribosome. For these reasons, I believe that this feature deserves further exploration into its role in piezophily by experimental methods.

Insertion #2

The second insertion shown in Figure 1.4 occurs only in KT99 amongst our piezophiles and mesophiles, and in fact, is a unique feature amongst 281 γ-proteobacteria examined. The insertion is 13 bp long, and is at position 655, relative to *S. frigidimarina*. KT99's insertion is predicted to form a new helix between helices 27 and 30 (Figure 1.12). The organism *Buchnera aphidicola* has an insertion at the same point as KT99; however it is 5 bp shorter and is not predicted to form the same secondary structure as KT99's insertion.

Though this feature is not found in the other 280 γ-proteobacteria examined, similar structures are found in other classes. For example, in 431 bacterial 23S sequences, 36.2% have an insertion at this same point with a

maximum length of 25 bp (CRW). The archaea also have an extra helix inserted, though it is located between helices 29 and 30 (*E. coli* numbering).

Although this insertion is unique and perhaps interesting in function and origin, it is unlikely that it is a piezophilic adaptation. Its uniqueness to KT99 suggests that it is for a different adaptation, perhaps borrowed from a different class of bacteria.

Insertion #3

Insertion #3 only occurs in SS9 and is an insertion of 17 bp relative to its comparison strain *Photobacterium profundum* 3TCK, as is shown in Figure 1.5. This insertion occurs on helix 45 and results in the elongation of the helix (structure not shown). Helix 45 is a hypervariable region whose length is known to vary by as much as 129 bp in the γ-proteobacteria. The insertion into SS9 falls well within this range, and though its insertion is not found in its comparison strain, *P. profundum* 3TCK, it is unlikely that this mutation is a piezophilic adaptation. This is further compounded since this insertion only occurs in SS9 and none of the other piezophiles.

Insertion #4

Insertion #4 occurs at position 1507 (*E. coli*) and is found in three piezophiles (MT41, KT99 and SS9) and one of the comparison mesophiles (*S.*

frigidimarina). It is also found in two other species of *Shewanella; S. baltica* and *S. denitrificans*.

The insertion is 14 bp long when aligned to the *E. coli* sequence. This occurs in helix 58 at the beginning of a two residue bulge (Figure 1.8). The new sequence is predicted to cause the formation of a new helix, shown in the predicted structures in Figure 1.13. Not shown are the structures of the additional mesophilic *Shewanella* species; however their structures are predicted to be nearly identical to that of *S. frigidimarina*. The helical base ranges from 3-5 bp long (per strand) and the terminal loops are either 4 or 6 bp in length.

On visual examination of the crystal structure, helix 58 is located on the side of the 50S subunit, towards the bottom near the 50S-30S interface (Schuwirth *et al.*, 2005). Helix 58 has been shown to interact with L9 (Brimacombe, 1995), however the crystal structure places considerable distance between these two structures (not shown). Instead, we find that RP L2 is in close proximity to helix 58 (Figures 1.14 and 1.16). Helix 58 is also in close proximity to helices 53, 54, 55, and 63. In Figure 1.15, the distance between a residue on helix 63 and a residue adjacent to the point of insertion on helix 58 is measured to be 6.715 Å. Figure 1.16 also shows the closeness of these two helices as well as the closeness between helix 58 and the other features mentioned.

Insertion #4 has the strongest piezophilic signature of the four large insertions analyzed. This insertion is found in the same three genera as were the insertions noted in the 16S sequence (Lauro *et al.*, 2007). Insertion #4 is also the only insertion found in all 15 of SS9's 23S rRNA sequences (discussed further in the "Ribosomal RNA Operon" section below). Also similar to the insertions into the 16S sequence, the analysis of the three-dimensional models indicates that the insertion would likely alter existing interactions with other components in the ribosome. This is especially true since this insertion results in the creation of a protruding helix, something that could have a profound effect on the existing helix 63.

Evidence against this insertion being a piezophilic adaptation is its presence in various non-piezophilic *Shewanella* species. However, this insertion (and the resulting extra helix) is only found in the *Shewanella*, except for those found in piezophiles MT41 and SS9. This question needs to be answered experimentally, possibly through genetics experiments testing piezophily with different 23S genes.

5S Ribosomal RNA

The smallest of the ribosomal RNAs, the 5S is part of the large ribosomal subunit. The piezophilic 5S is relatively unchanged compared to its mesophilic counterparts. Figure 1.17 shows a sequence alignment of the 5S sequences. As expected, there is variation between species; however there is

little variation between strains. In strains that have more than one 5S ribotype, the same number of ribotypes is present in the comparison strain in all but one species. The number of copies of 5S genes is not significantly different between piezophiles and mesophiles.

Though not a piezophile-specific characteristic, as it was found in all of the 5S rRNAs examined, the structure of the 5S rRNA is different from the *Escherichia coli* reference structure. Figure 1.18 shows that the structural change results from a single base mutation, C31U. The change affects the size of the end loop, which is small, and the physiological significance, if any, is unknown.

Transfer RNAs

Transfer RNAs function is to translate the genetic code into amino acids. Their structure is important for them to properly function, and like the ribosomal RNAs, may be under selective pressure at high-hydrostatic pressure. After extracting the tRNA sequences from the genomes (see Methods), the sequences were separated by amino acid and anticodon type.

In all of the sequence alignments, no significant variations were found between piezophiles and mesophiles. When an Average Distance phylogenetic tree is built from the alignments, the tRNAs cluster by genera, an example of which can be seen in Figure 1.19. The lack of sequence diversity between

piezophiles and mesophiles suggest that tRNAs are not significantly affected by high-hydrostatic pressure.

The prevalence of tRNAs was also investigated, but also did not lead to a suggestion that tRNA count is affected by high-hydrostatic pressure. Graph 1.1 shows the difference in the number of tRNAs sorted by amino acid type. Although there is variation between piezophiles and mesophiles, there is no conclusive pattern to suggest a piezophilic trait. This conclusion does not change if amino acids are grouped by physical or chemical properties, such as side-chain size and hydrophobicity (data not shown).

Ribosomal RNA Operon

The ribosomal RNA operon typically contains the 16S, 23S, and 5S rRNAs and can contain one or more tRNAs (Srivastava and Schlessinger, 1990). The operon is transcribed all at once and is then cleaved into its separate mature RNAs by various RNases (Srivastava and Schlessinger, 1990). There are three non-functional RNA regions of interest in the operon: the upstream regulatory region, the region between the 16S and 23S rRNAs (called the 16S-23S internal transcribed spacer, or ITS) and the region between the 23S and 5S rRNAs, or 23S-5S ITS. The secondary structure formed by the two ITS regions is important for correct processing of the rRNAs and tRNAs in the operon (Brosius *et al.*, 1981). The upstream regulatory region spans approximately 450 bp upstream of the beginning of

the 16S rRNA and contains two sets of promoters as well as a number of other protein binding sites (Condon *et al.*, 1995).

Sequences were extracted from available genomes (Table 1.2) and were aligned with MUSCLE v3.7 (Edgar, 2004) and were analyzed with Jalview v2.3 (Clamp *et al.*, 2004). The coordinates from the RNAmmer results were used in extracting the sequences. Due to unfinished genomes, not all organisms were able to have the operon sequences compared due to incompleteness or lack of sequence.

Upstream Regulatory Region

Only two piezophile-mesophile pairs were able to be compared due to sequence limitations: *Psychromonas* sp. CNPT3/*Psychromonas ingrahamii* 37 (hereafter called CNPT3 and PING, respectively) and *Colwellia* sp. MT41/*Colwellia psychrerythraea* 34H (hereafter called MT41 and 34H, respectively). Both pairs of sequences were aligned separately and then analyzed. *Photobacterium profundum* SS9 is also analyzed, but in the context that it contains multiple ribotypes and, as will be discussed, multiple ribosomal operon types.

The regulatory features of interest are described in Condon *et al.* (1995). Table 1.3 lists the features and the positions of these features in *Escherichia coli* relative to the beginning of the 16S sequence.

One of the more distinctive features in the CNPT3/PING alignment (Figure 1.20) is the presence of large gaps in CNPT3's sequence. Bases -501 to -459 of CNPT3 are part of a variable sequence that presumably lies outside of the operon regulatory region. Beginning at base -460, the two sequences begin to align; however there are two large deletions in CNPT3's sequence, one at base -344 and the other at base -68. These deletions are 25 bp and 15 bp in length, respectively. There are other small gaps in both sequences, but these cancel each other out for the most part.

The main effect of these insertions seems to be that in CNPT3, the P_1 Promoter is about 40 bp closer to the 16S sequence and the P_2 Promoter is about 20 bp closer. Potential Fis binding sites (Hengen *et al.*, 1997) are also affected, with the first one being about 25 bp closer to the P_1 Promoter in PING than in CNPT3. Both of these sequences are different than *E. coli* in that both promoters have predicted UP elements whereas in *E. coli*, only P_1 has it (Condon *et al.*, 1995).

The MT41/34H alignment is less clear in terms of the conclusions that can be drawn because of the sequence diversity in each organism (Figure 1.21). MT41 has four upstream genotypes and 34H has three. There are a number of large deletions/insertions; however they are usually present in at least one genotype in each organism. At -60 from the 16S sequence, there is an insertion into three of the four MT41 genotypes. At this position in the upstream region, it does not affect the promoters; however it may alter the

secondary structure of the transcript. There is also a deletion in three of the four MT41 genotypes (though a different three from previously) that occurs between -253, -265 and -279 bp from the 16S, depending on the genotype.

The two *Colwellia* species have two promoters, however in these sequences, the P_2 Promoter is predicted to be significantly weaker than the P_1 Promoter. Both the -35 and -10 binding sequences are further from consensus sequence and the predicted UP element is significantly weaker, if it even functions as one. This is in contrast to the *Psychromonas* species in which both promoters were of approximately equal strength.

The "*box*" elements are more difficult to identify and are involved in forming secondary structure necessary for rRNA maturation and possibly in antitermination (Brosius *et al.*, 1981; Condon *et al.*, 1995). In *Psychromonas, boxC* was found to be a near perfect match to *E. coli*'s *boxC* (Figure 1.20). A comparative analysis of the regions where the box elements are likely located is easily done. The box regions typically occur within the first 70 bp after the P_2 transcription start site. In the CNPT3/PING alignment, the two sequences are virtually identical up to the predicted *boxC* site. The *boxC* site occurs approximately 20 bp earlier than in the *E. coli* sequence, suggesting a shortened *box* region. In the MT41/34H alignment, there is a deletion of 3, 4, then 2 bp at positions -203, -198 and -89, respectively, in all of MT41's sequences. These deletions would likely alter the structure of *boxB* and possibly *boxA*. Given the diversity of sequence in the ribosomal RNA operon

for MT41, these consistent deletions are likely significant, though their effect on piezophily is unknown since it is only observed in this genus. However, this could coincide with the shortened box region observed in the *Psychromonas*.

Photobacterium profundum SS9 (hereafter called SS9) has 15 ribosomal RNA operons and it has been reported that all of the operons are expressed all of the time (Lauro *et al.*, 2007). This is of interest because SS9 possess a number of ribotypes; five 16S ribotypes and four 23S ribotypes. Paired together, there are nine unique combinations of 16S and 23S ribotypes found within SS9's rRNA operons. Since they are all expressed, differentiation in regulation would not be expected.

When the aligned sequences are analyzed (not shown), this is in fact what is observed. At approximately -285 bp, depending on the operon, from the 16S sequence, a strong promoter (-35: TTGAC[AT], -10:TAT[AT]AT) can be found on each operon preceded by a relatively strong UP element. It is unknown whether this is the P_1 or P_2 Promoter because a second promoter could not be found. A candidate for a second promoter was found at approximately -167 bp, but it would be a fairly weak promoter (-35: TCGTGA, -10:ATTTAT). This candidate promoter, however is identical throughout all 15 operons whereas the other promoter is somewhat varied.

Even though the regulation regions of the operons are likely identical, other observations about the sequences can be made. A region of interest is

within the -1 to -40 range. For about 150 bp until this point, all 15 sequences are highly conserved. However, in this region, the sequences diverge into three distinct clusters (Figure 1.22). The sequences, when aligned within their clusters, are nearly identical to each other. However, when the sequences are extended to approximately -265 bp, a different clustering set emerges.

For instance, up until the -285 promoter, the sequences were quite varied with the exception of five that have a high degree of sequence similarity (<1 mismatch/100 bp). This occurs all the way up until -71 bp, when one of the sequences suddenly "joins" another cluster, with which it shares near perfect sequence identity. This evidence strongly suggests that there has been recombination within the upstream regions of the ribosomal RNA operons and may also explain the different ribotypes found within SS9.

16S-23S Internal Transcribed Spacer

The 16S-23S ITS is functionally important in the maturation of the ribosomal RNAs (Brosius *et al.*, 1981; Antón *et al.*, 1998). However, even when performing this function, there can be great sequence diversity between and within organisms (Antón *et al.*, 1998; Osorio *et al.*, 2005). Much of the sequence diversity is due to the addition of tRNAs into the sequence (Osorio *et al.*, 2005).

The 16S-23S ITS was examined for all the piezophiles and mesophiles for which it was available. No specific piezophilic traits were found. It was common for a single organism, piezophile or mesophile, to have one or more sequence clusters, usually determined by the number of tRNAs present in that sequence. There was also no correlation between the number of tRNAs or the size of the ITS and piezophily.

23S-5S Internal Transcribed Spacer

The 23S-5S ITS connects the 23S rRNA transcript to the 5S rRNA transcript (Brosius *et al.*, 1981; Antón *et al.*, 1998). Secondary structure again plays an important role in the function of the ITS (Antón *et al.*, 1998). In analyzing the 23S-5S ITS sequences, no significant features were found suggesting piezophilic adaptations. Sequence length was comparable amongst organisms, as was the distribution of tRNAs and tandem 5S sequences.

Ribosomal Proteins

Although the ribosome's catalytic domain is found within its RNA, ribosomal proteins are still necessary for its proper function and assembly (Steitz and Moore, 2003). In this study, we examined the structural implications of insertions that were found in piezophilic ribosomal RNAs. Within the 30S subunit, the S16 and S21 RPs were most likely to be affected

by the insertions. The L2, L9, L13, L20, and L21 RPs were identified as possibly being affected by the insertions in the 23S rRNA.

An overview alignment was done on all ribosomal proteins in the piezophiles and mesophiles used in this study. Based on those alignments, there are no major differences between the piezophile proteins and the mesophile proteins. Both *Moritella* species and *Psychromonas* sp. CNPT3 have an L31B protein, but this is not likely significant for piezophily.

The proteins that were thought to be affected by the ribosomal RNA inserts were examined more closely and revealed a few potential piezophilic adaptations. Ribosomal proteins L20, L21 and S16 revealed nothing significant. Two piezophile-specific substitutions were found in *Colwellia* sp. MT41 and *Shewanella benthica* KT99; one in RP L2 and the other in RP S21. In L2, isoleucine was substituted for valine at position 65 in both species. The substitution in S21 was for asparagines at position 64. *Colwellia* sp. MT41 substituted in a serine whereas *S. benthica* KT99 substituted in a threonine. These substitutions may be significant given that these two organisms are the most piezophilic of the set.

Ribosomal protein L9 has three piezophiles substituting asparagines in at position 73. This gives a total of four piezophiles with asparagines at this position along with one mesophile. One piezophile substituted a serine for an alanine. The placing of asparagines at position 73 for piezophiles is not quite

statistically significant with a calculated p-value of 0.1031 (Fisher's Exact Test).

The last protein to be closely examined is L13. At position 34, three piezophiles substituted serines for threonines. One piezo-meso pair has serines at that position while another has leucines. This substitution is interesting because serine is a smaller residue than threonine, though it is chemically similar. This could be a piezophilic adaptation to living at high pressure or it could be an adaptation to the modifications of the 23S rRNA. Like the previous substitution, this event was not quite statistically significant and also had a calculated p-value of 0.1031 (Fisher's Exact Test).

4.5S RNA (Signal Recognition Particle RNA)

The 4.5S RNA is part of a complex that binds to the ribosome and directs protein translation to the cell membrane (Gu *et al.,* 2003). This complex, the signal recognition particle (SRP) binds to ribosomal protein L23 near the exit point of the new peptide (Gu *et al.,* 2003). Because of its association with the ribosome, it was considered for this study.

Sequences for the 4.5S RNA were found through homology searching using query sequences from *Shewanella frigidimarina*, *Escherichia coli* and *Psychromonas ingrahamii*. This revealed significant matching sequences in all five γ-proteobacteria piezophiles and their mesophile comparison strains. An alignment of the sequences (Figure 1.23) shows significant diversity

between genera, but relatively little diversity within. Based on this analysis, it is unlikely that the 4.5S RNA contributes to an organism's piezotolerance.

Table 1.1: Pairings of organisms for sequence comparison.

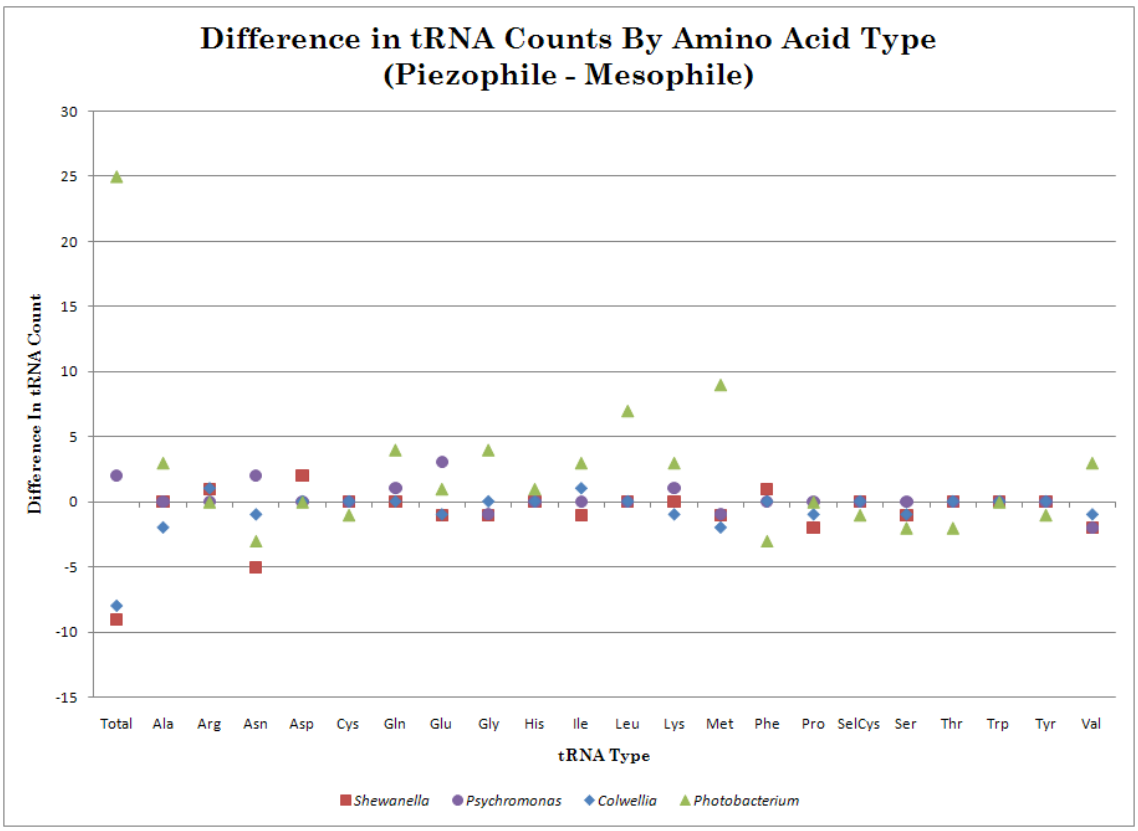
Piezophile	Mesophile
Carnobacterium sp. AT7	*Enterococcus faecalis* V583
Colwellia sp. MT41	*Colwellia psychrerythraea* 34H
Moritella sp. PE36	*Moritella viscosa*
Photobacterium profundum SS9	*Photobacterium profundum* 3TCK
Psychromonas sp. CNPT3	*Psychromonas ingrahamii* 37
Shewanella benthica KT99	*Shewanella frigidimarina* NCIMB 400

Table 1.2: List of organisms and the genome accessions used.

Organism	Genome Accession
Carnobacterium sp. AT7	Unpublished Assembly Latest public assembly: GI:159875372 (Lauro *et al.*, 2007)
Colwellia sp. MT41	Unpublished Assembly (Yayanos *et al.*, 1981)
Colwellia psychrerythraea 34H	Public Assembly: GI:71143482 (Methe *et al.*, 2005)
Enterococcus faecalis V583	Public Assembly: GI:29350190, 29345347, 29345328, & 29345255 (Paulsen *et al.*, 2003)
Moritella sp. PE36	Unpublished Assembly Latest public assembly: GI:149809450 (Yayanos *et al.*, 1986)
Moritella viscosa	Unpublished Assembly (Benedikstoddir *et al.*, 2000)
Photobacterium profundum SS9	Public Assembly: GI:47419843, 47419844, & 46911589 (Delong *et al.*, 1997)
Photobacterium profundum 3TCK	Public Assembly: GI:90329665 (Campanaro *et al.*, 2005)
Psychromonas sp. CNPT3	Unpublished Assembly Latest public assembly: GI:90322500 (Yayanos *et al.*, 1979)
Psychromonas ingrahamii 37	Public Assembly: GI:119862398 (Breezee *et al.*, 2004)
Shewanella benthica KT99	Public Assembly: GI:161332323 (Lauro *et al.*, 2007)
Shewanella frigidimarina NCIMB 400	Public Assembly: GI:114332481 (Bowman *et al.*, 1997)

Table 1.3: List of ribosomal operon upstream regulators found in *Escherichia coli*.

Feature Name	Position Upstream of 16S
boxC	-117
boxA	-143
boxB	-170
P_2 Promoter	-209
P_1 Promoter	-328
UP element	-353
Fis I	-370
Fis II	-401
Fis II	-452



Graph 1.1: Difference in the number of tRNAs in piezophile-mesophile pairs, grouped by amino acid type.

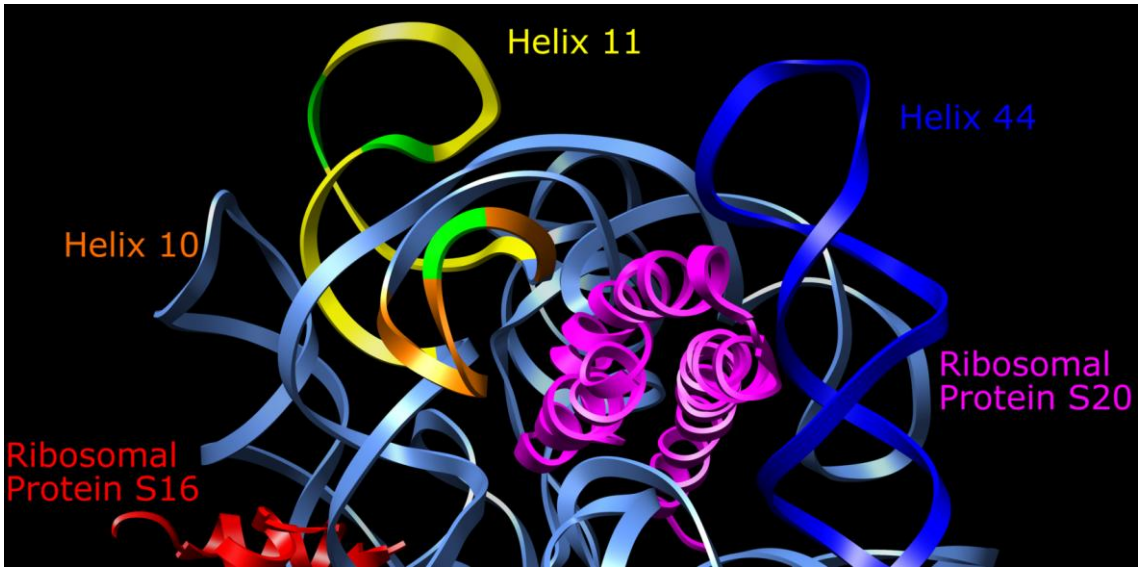


Figure 1.1: Ribbon diagram of the region of the 30S ribosomal subunit surrounding helix 10, rendered by the UCSF Chimera software package. Helices of interest are colored and labeled as are proteins in the region. Points of insertion on helices 10 and 11 are colored green on their respective helices.

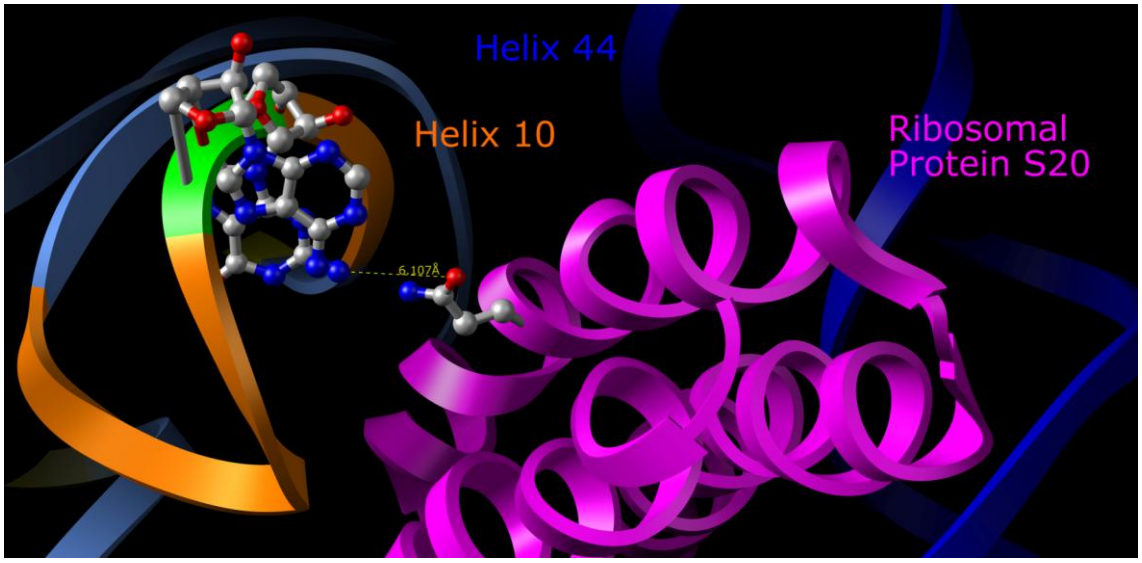


Figure 1.2: Diagram showing the closeness of helix 10 and ribosomal protein S20 as rendered by the UCSF Chimera software package. One residue from each was selected and the distance between them measured, which came to be 6.107 Å.

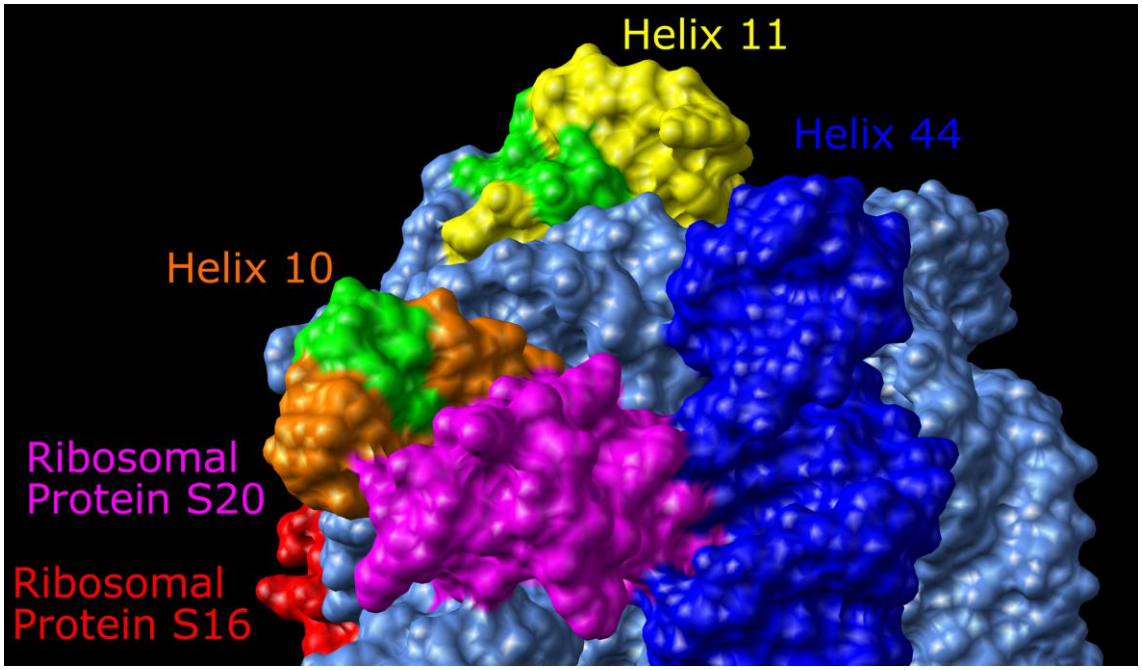


Figure 1.3: Surface rendering of the region of the 30S ribosomal subunit surrounding helix 10 by the UCSF Chimera software package. Contact between helix 10 and S20 and helix 44 and S20 are visible, indicating possible interaction. This interaction could potentially be altered by insertions in either of the helices.

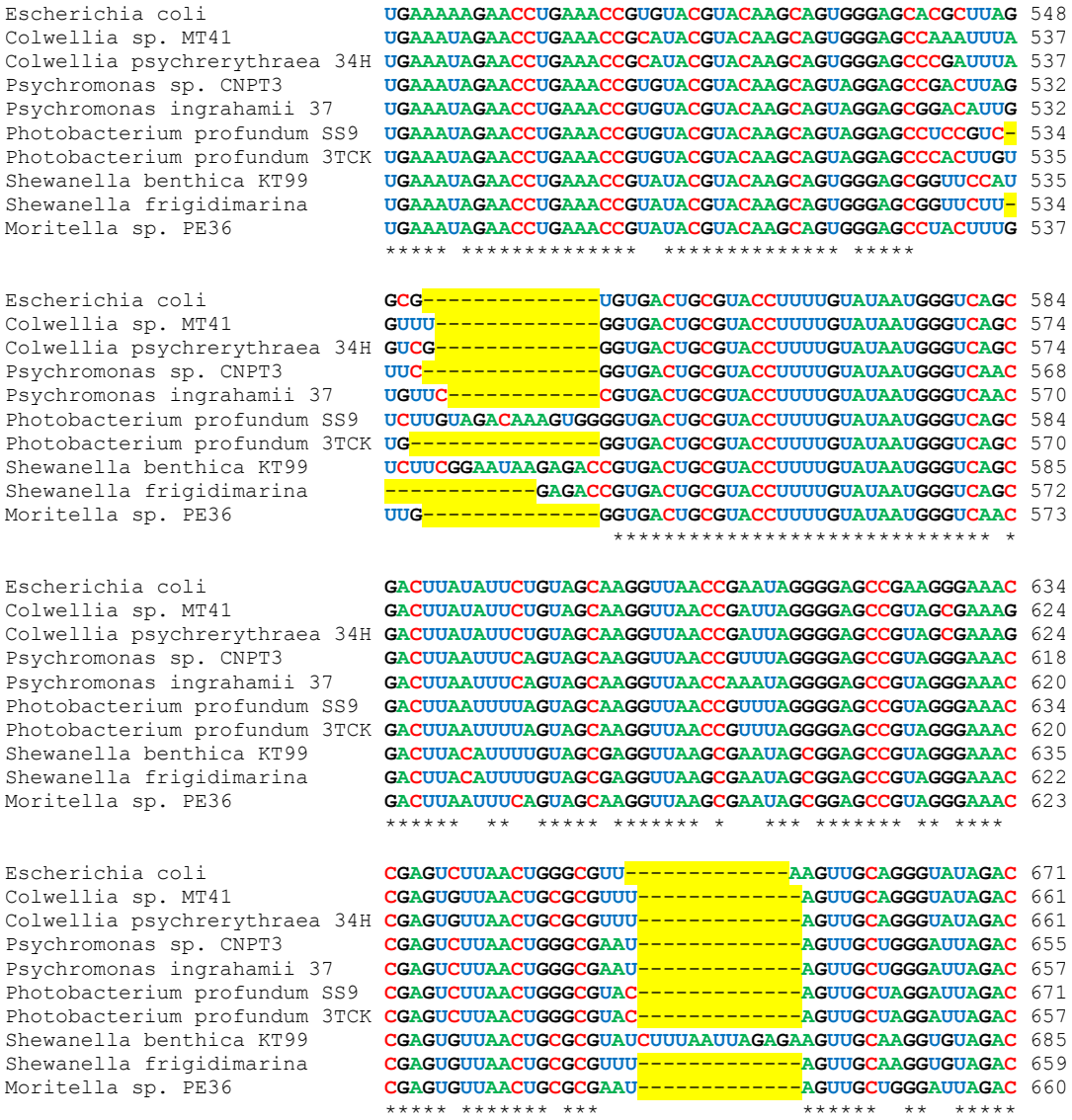


Figure 1.4: Sequence alignment of 23S ribosomal RNA sequences showing insertions #1 and #2.

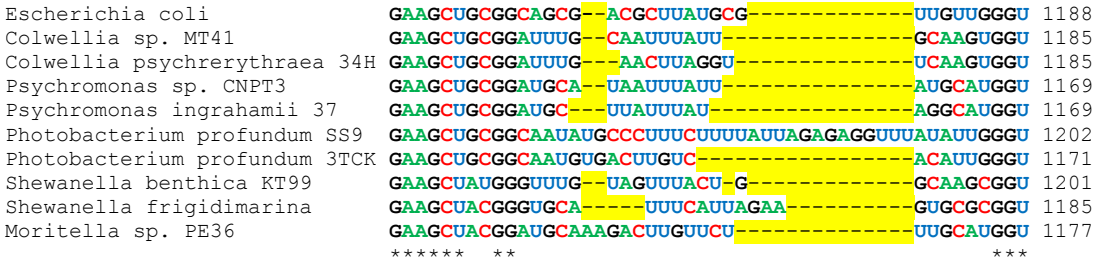


Figure 1.5: Sequence alignment of 23S ribosomal RNA sequences showing insertion #3.

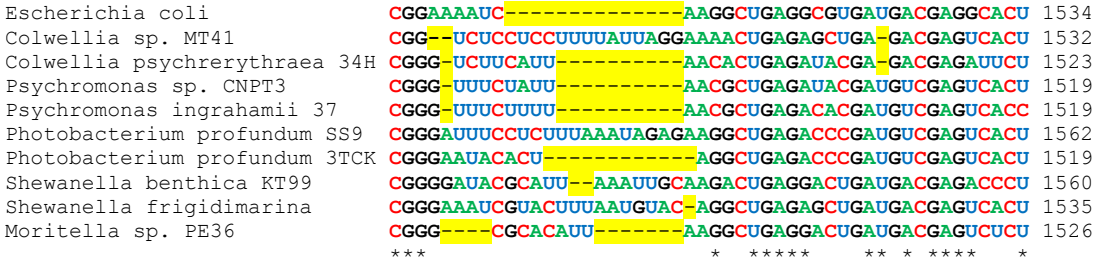


Figure 1.6: Sequence alignment of 23S ribosomal RNA sequences showing insertion #4.

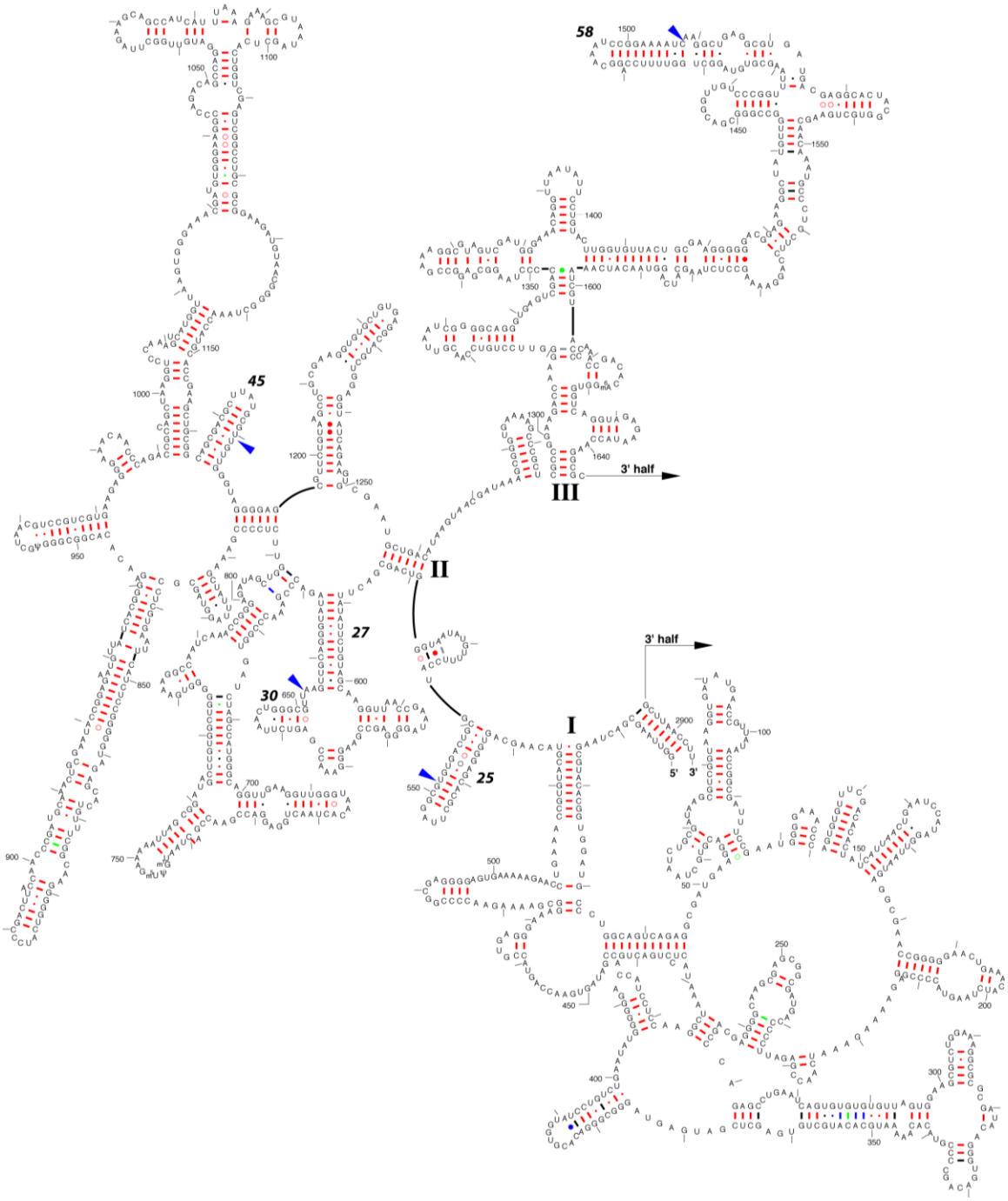


Figure 1.7: Placement of insertions relative to the 5′ half of the reference *Escherichia coli* 23S ribosomal RNA from the Comparative RNA Web Site and Project. The blue arrowheads indicate the point of insertion while the large Arabic numbers indicate the helix numbers. Large Roman numbers label the domains of the 23S rRNA.

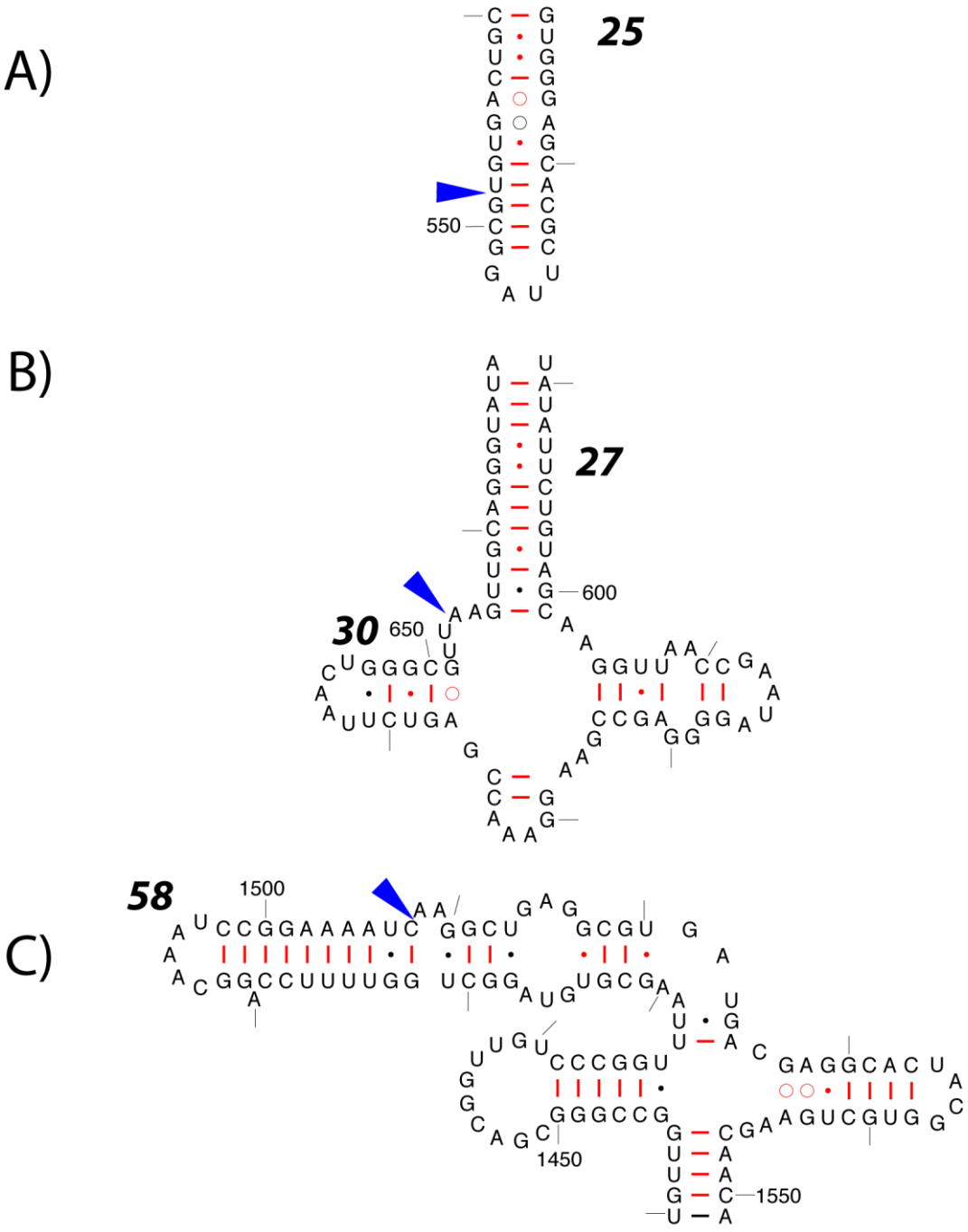


Figure 1.8: Close view of insertions relative to the *Escherichia coli* 23S ribosomal RNA. Blue arrow heads indicate the point of insertion and the large numbers indicate helix numbering. Insertion #3 is not shown. (A) Placement of insertion #1 on helix 25. (B) Placement of insertion #2 between helices 27 and 30. (C) Placement of insertion #4 on helix 58.

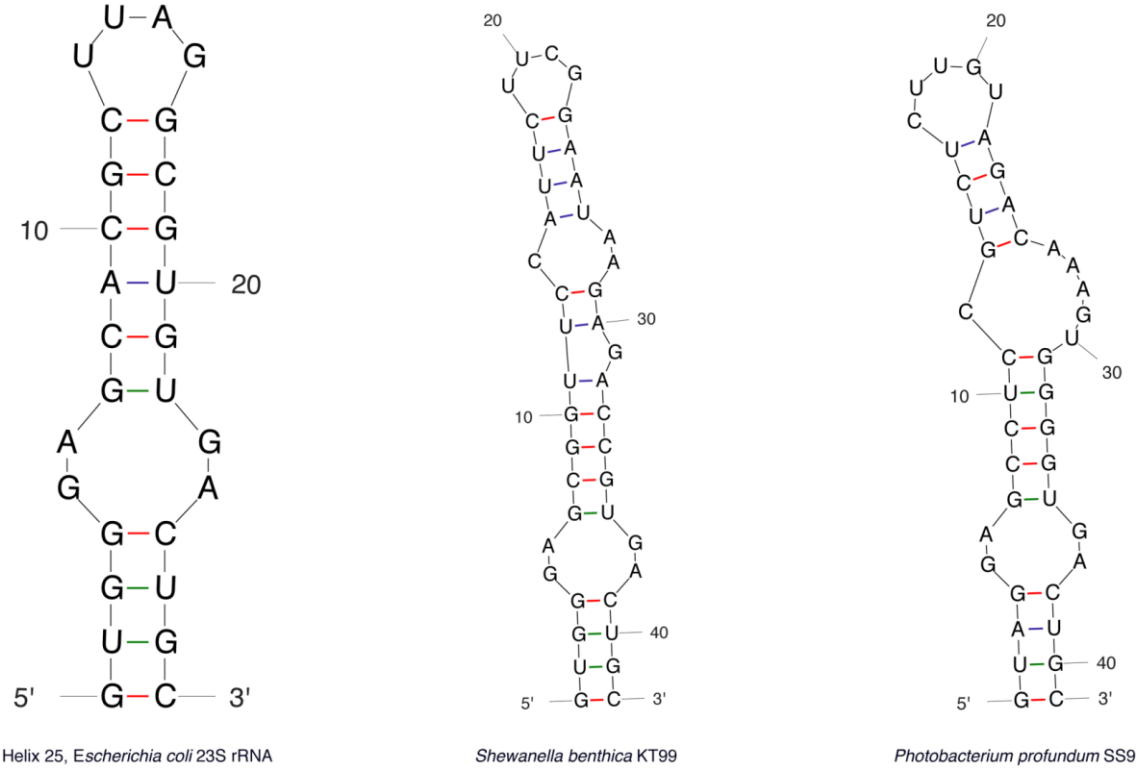


Figure 1.9: Predicted structures of helix 25 in *Escherichia coli*, *Shewanella benthica* KT99, and *Photobacterium profundum* SS9, as predicted by UNAFold v3.5.

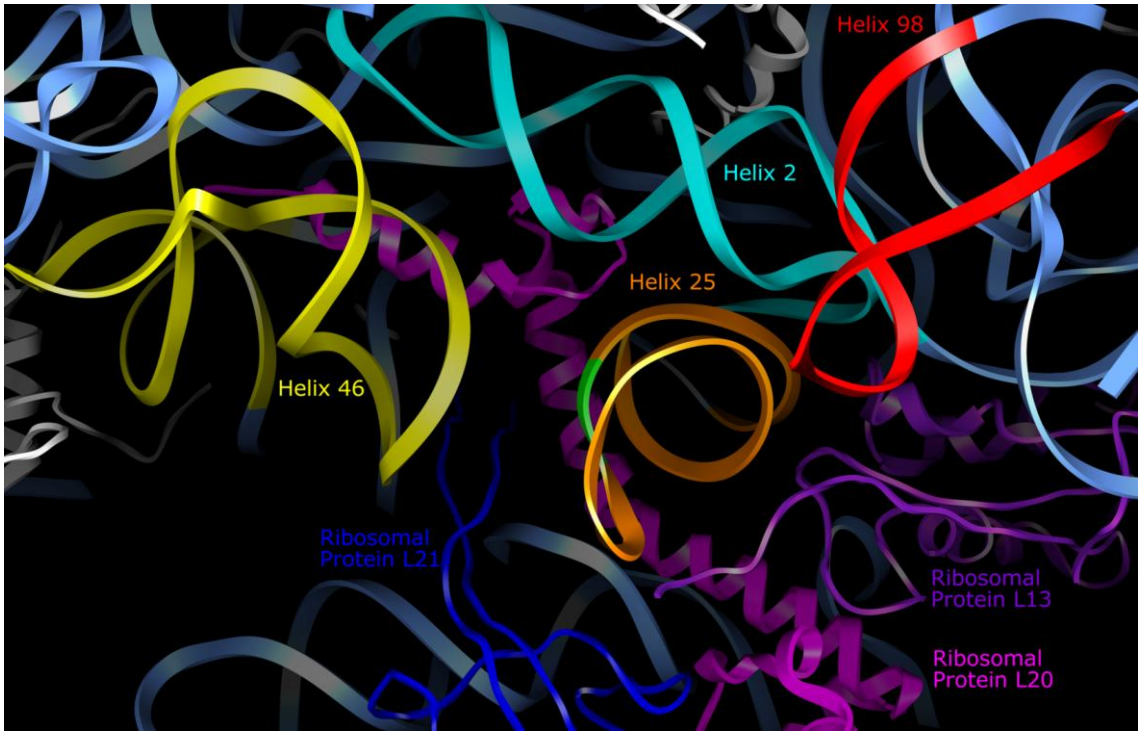


Figure 1.10: Ribbon diagram of the region of the 50S ribosomal subunit surrounding helix 25, rendered by the UCSF Chimera software package. Helices and proteins of interest are colored. The point of insertion on helix 25 is colored green.

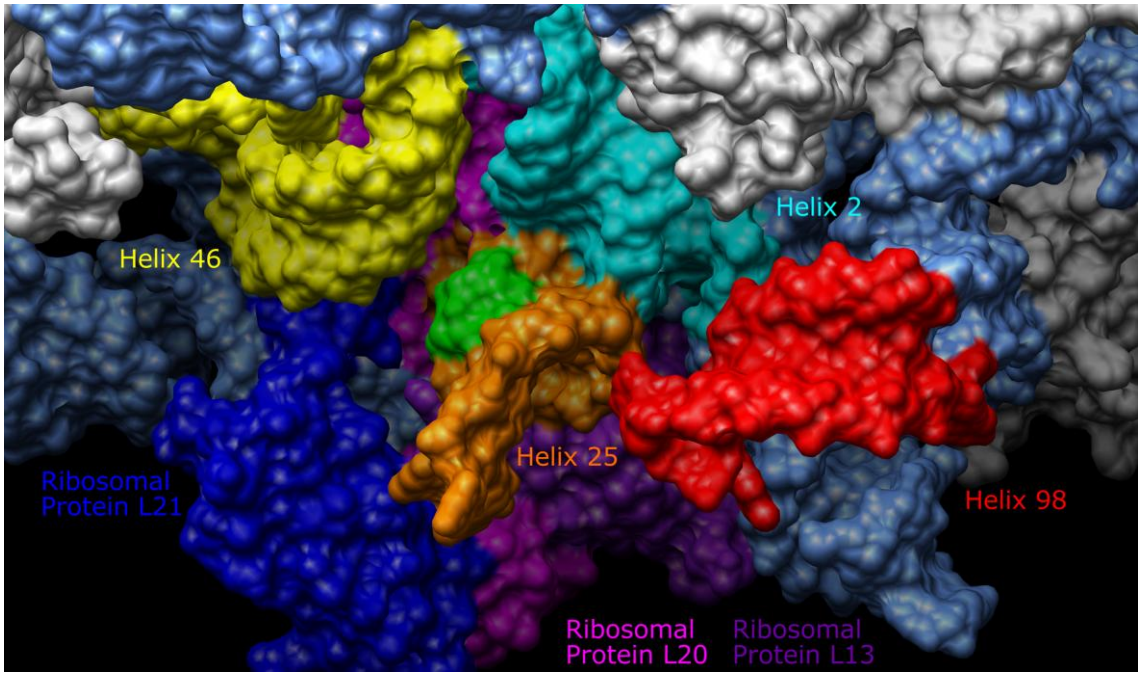


Figure 1.11: Surface rendering of the region of the 50S ribosomal subunit surrounding helix 25 by the UCSF Chimera software package. Contact between helix 25 and helix 2, ribosomal protein L20 and L13 are shown, indicating possible interaction. This interaction could possibly be altered by insertions in helix 25.

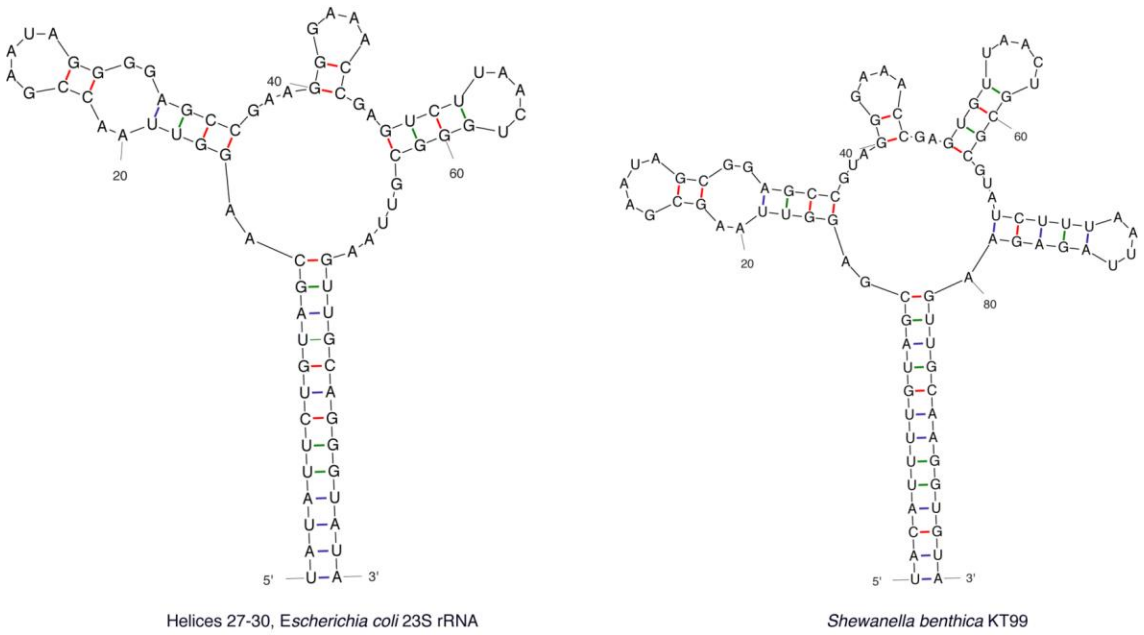


Figure 1.12: Predicted structures of helices 27-30 in *Escherichia coli* and helices 27-30 and Insertion #2 in *Shewanella benthica* KT99, as predicted by UNAFold v3.5.

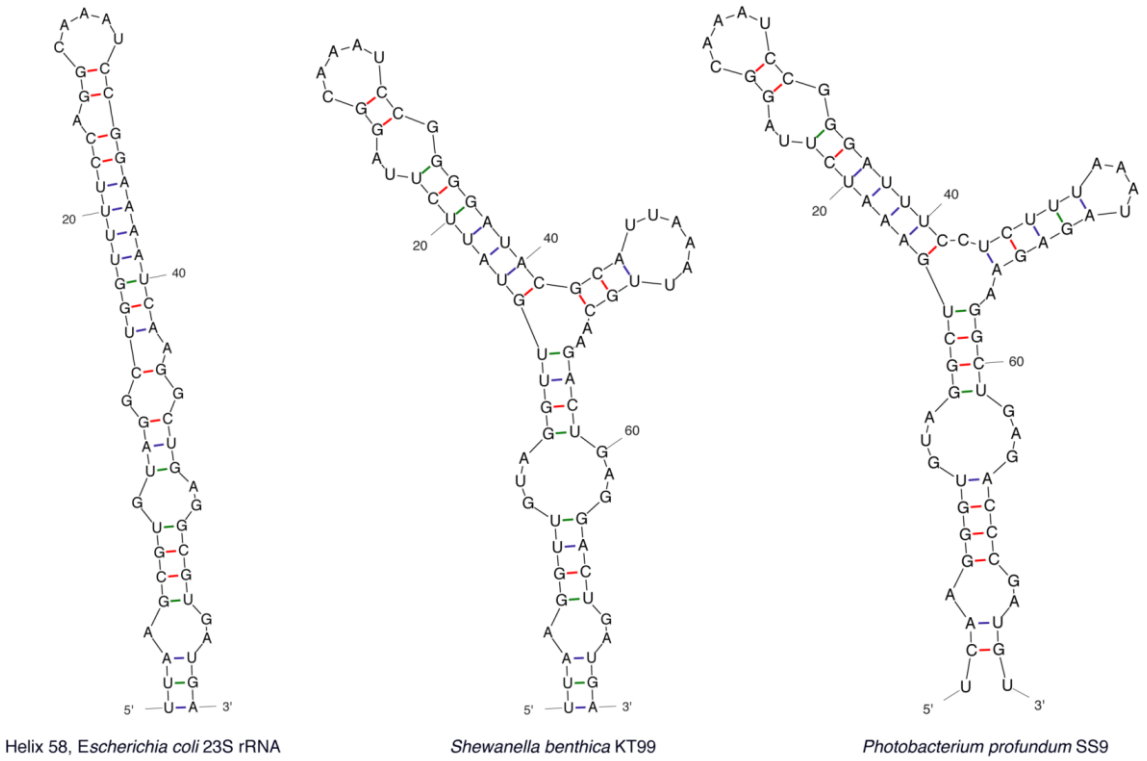


Figure 1.13: Predicted structure helix 58 in *Escherichia coli* and helix 58 with Insertion #4 in *Shewanella benthica* KT99, *Photobacterium profundum* SS9, *Colwellia* sp. MT41 and *Shewanella frigidimarina*.

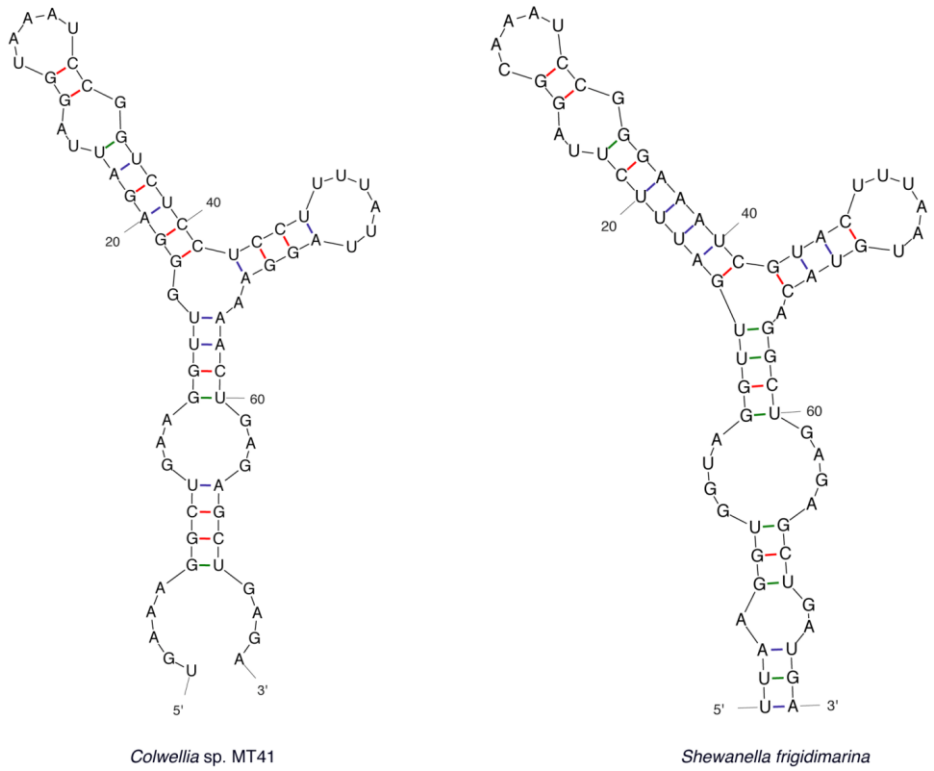


Figure 1.13: Predicted structure helix 58 in *Escherichia coli* and helix 58 with Insertion #4 in *Shewanella benthica* KT99, *Photobacterium profundum* SS9, *Colwellia* sp. MT41 and *Shewanella frigidimarina,* Continued.

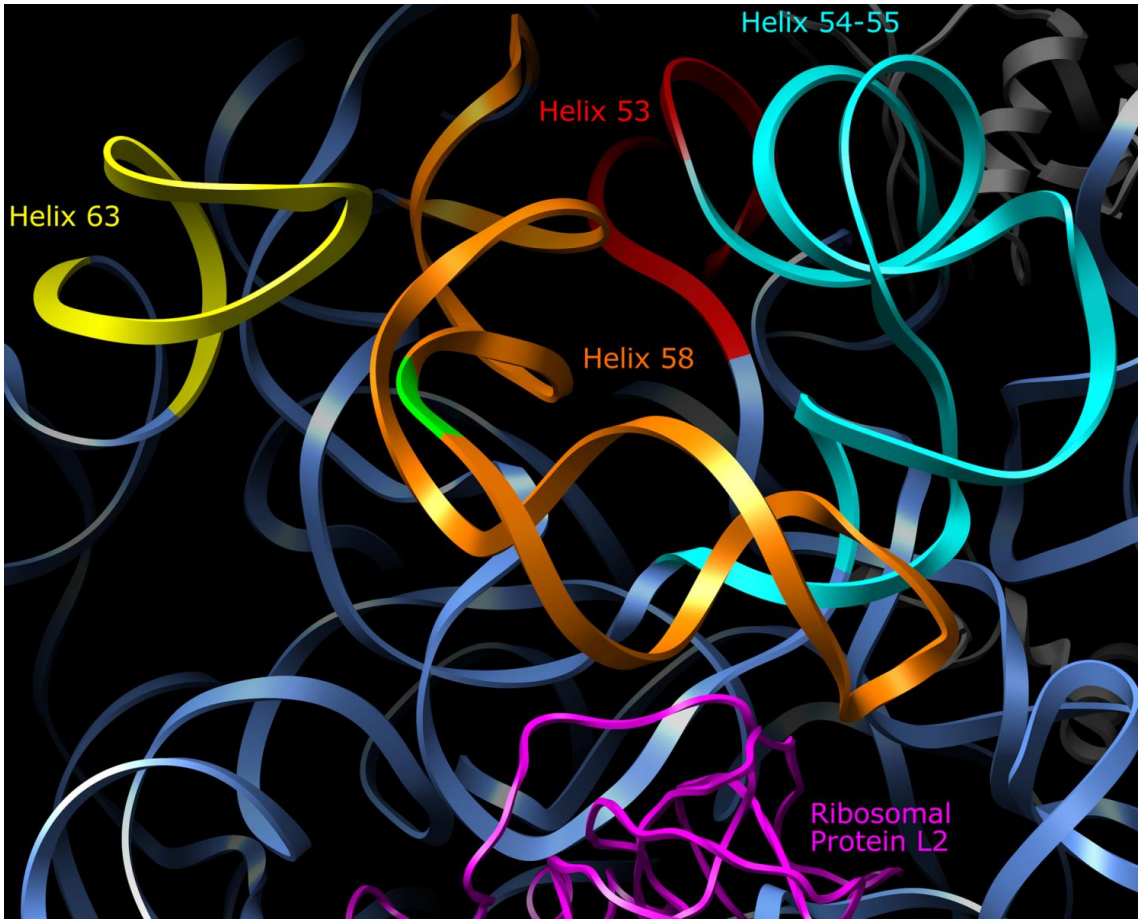


Figure 1.14: Ribbon diagram of the region of the 50S ribosomal subunit surrounding helix 58, rendered by the UCSF Chimera software package. Helices and proteins of interest are colored. The point of insertion on helix 58 is colored green.

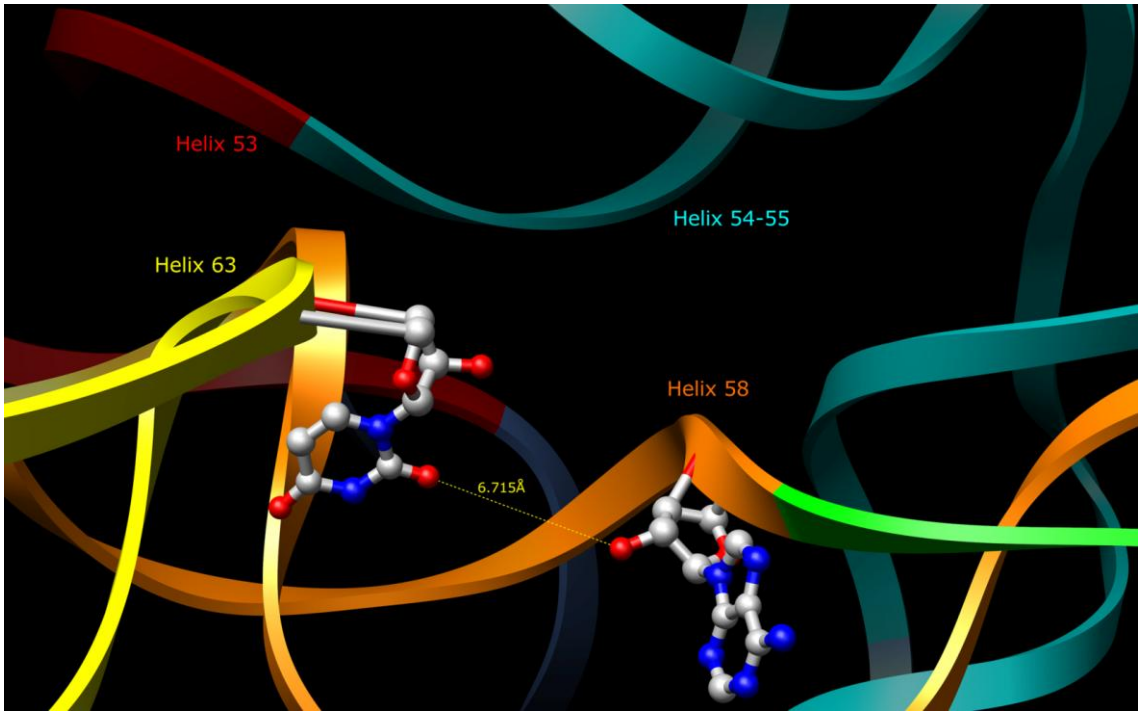


Figure 1.15: Ribbon diagram showing the closeness of helix 58 and helix 63, rendered by the UCSF Chimera software package. The distance between a base pair on helix 63 (uridine) and a base pair on helix 58 (adenosine) adjacent to the point of insertion is shown. The distance is measured to be 6.715 Å.

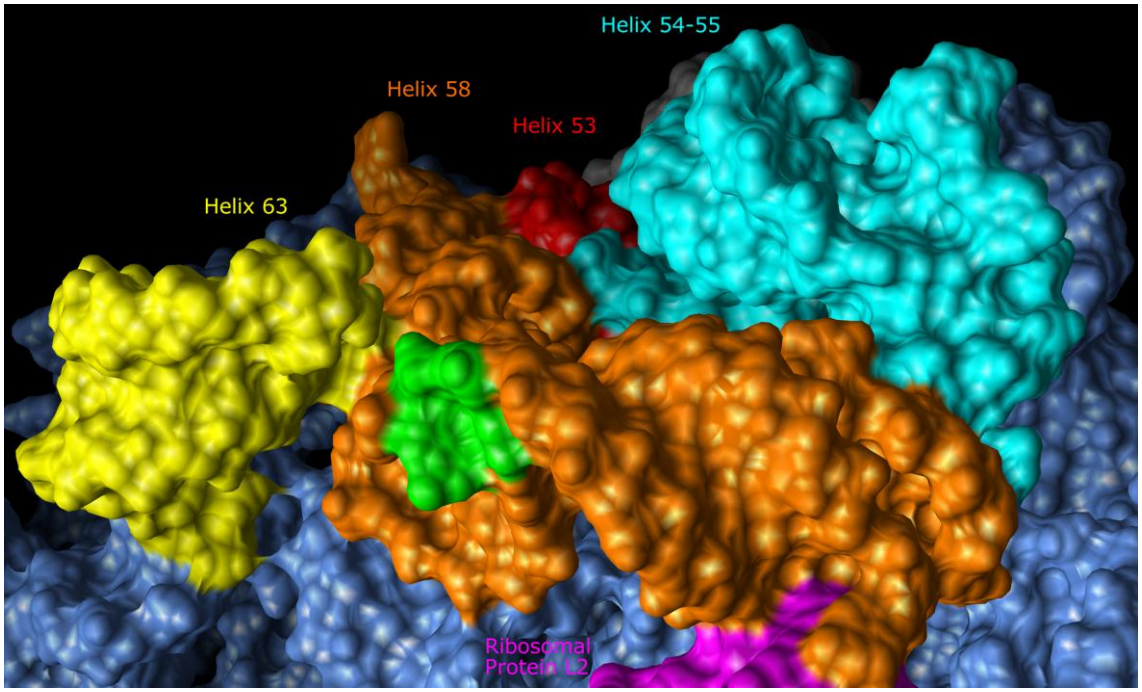


Figure 1.16: Surface rendering of the region of the 50S ribosomal subunit surrounding helix 58 by the UCSF Chimera software package. Contact between helix 58 and helix 63, 53, 54, 55 and ribosomal protein L2 are shown, indicating possible interaction.

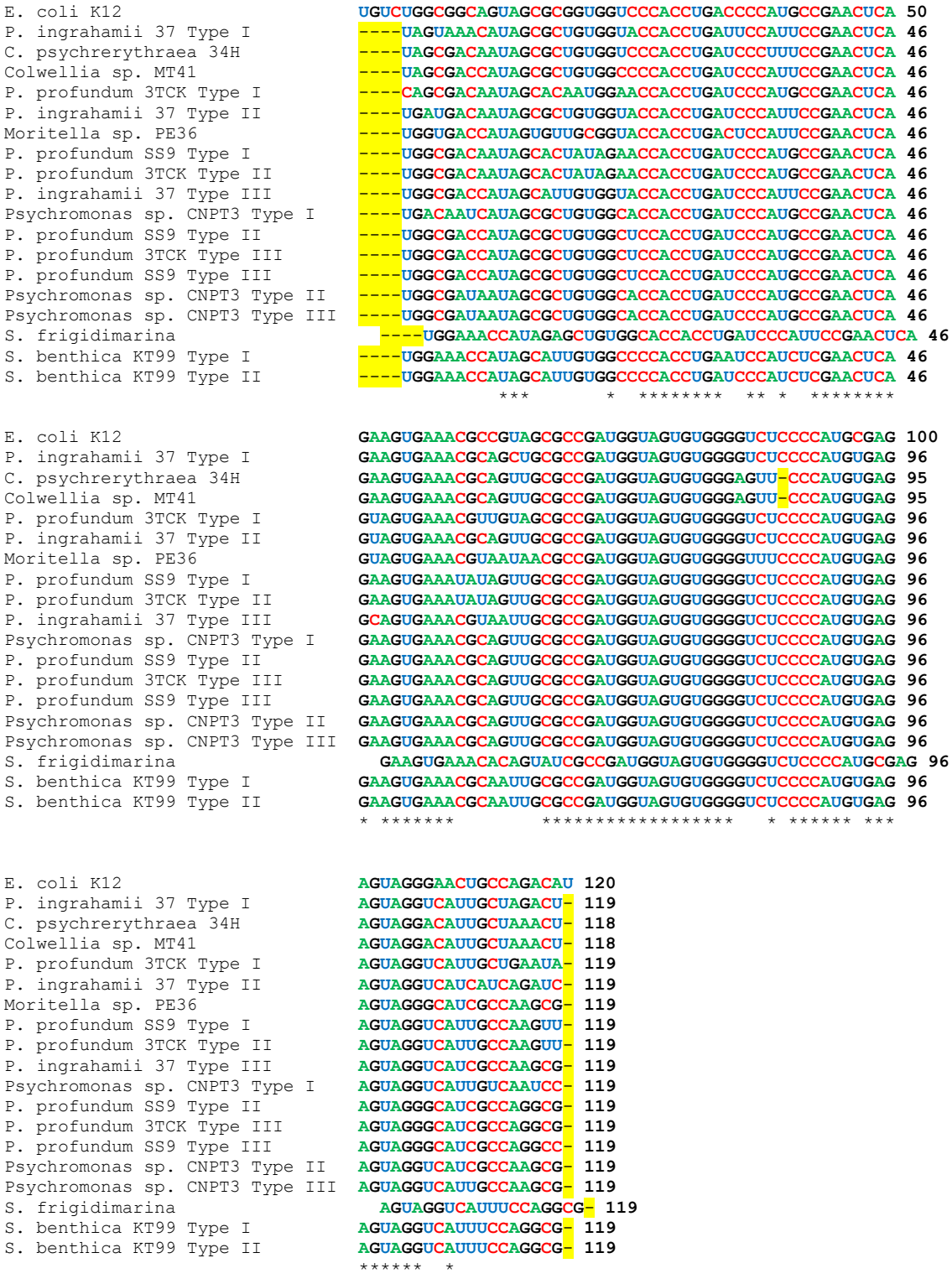


Figure 1.17: Multiple sequence alignment of the 5S ribosomal RNA of piezophiles and mesophiles.

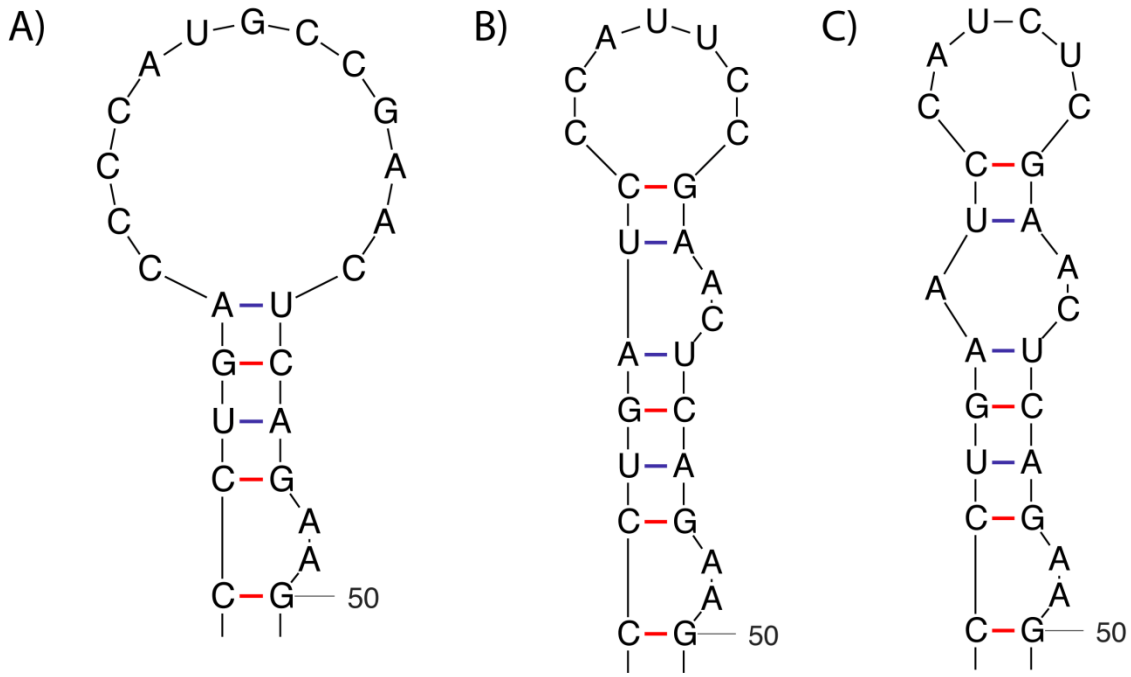


Figure 1.18: Comparison of reference and representative structure of the 5S ribosomal RNA. (A) Loop section of *Escherichia coli* K12 5S rRNA. (B) Loop section of *Colwellia* sp. MT41 5S rRNA, which is representative of most of the 5S rRNAs compared in this study. (C) Loop section of *Shewanella benthica* KT99 5S rRNA. This structure is also found in *Moritella* sp. PE36.

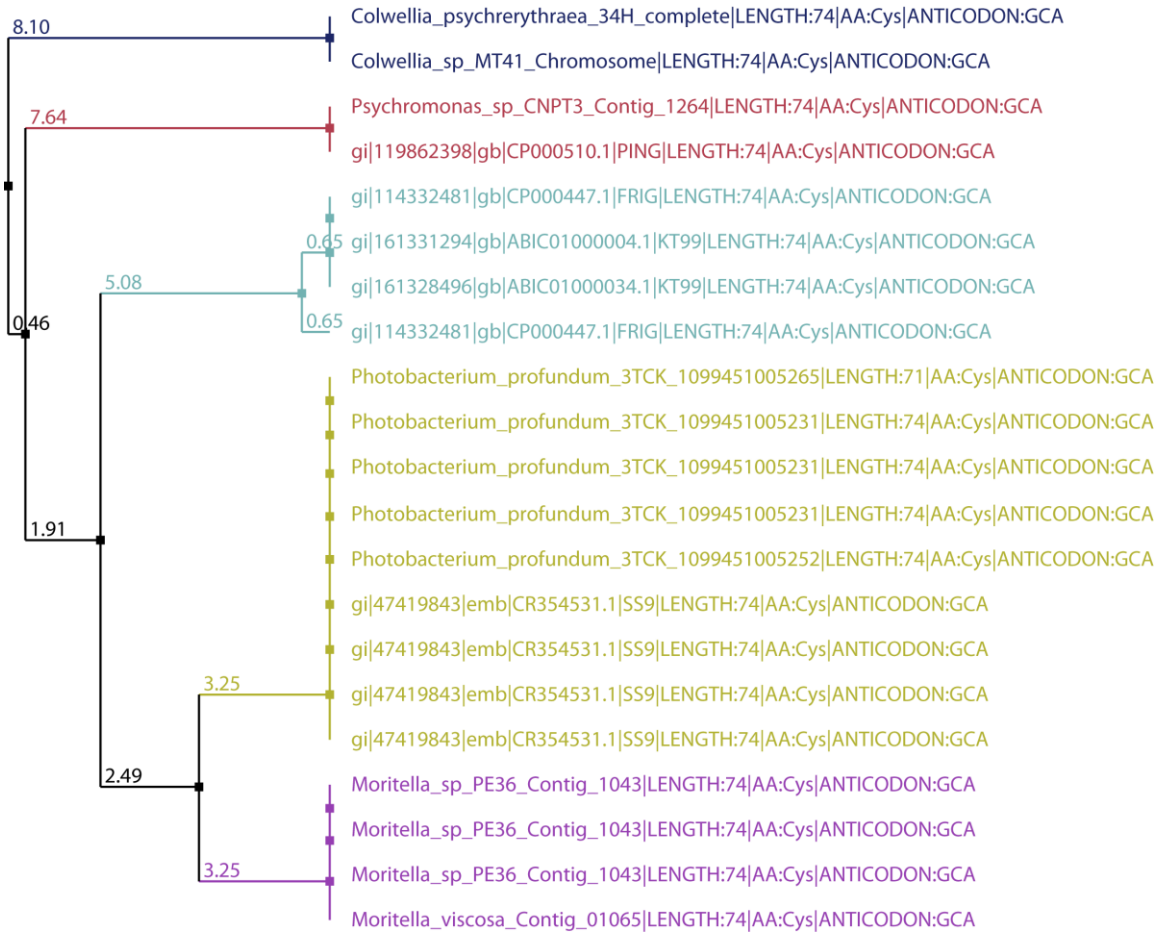


Figure 1.19: Phylogenetic relationship between cysteine transfer RNAs. The clustering of tRNAs into their derived genera is highlighted.

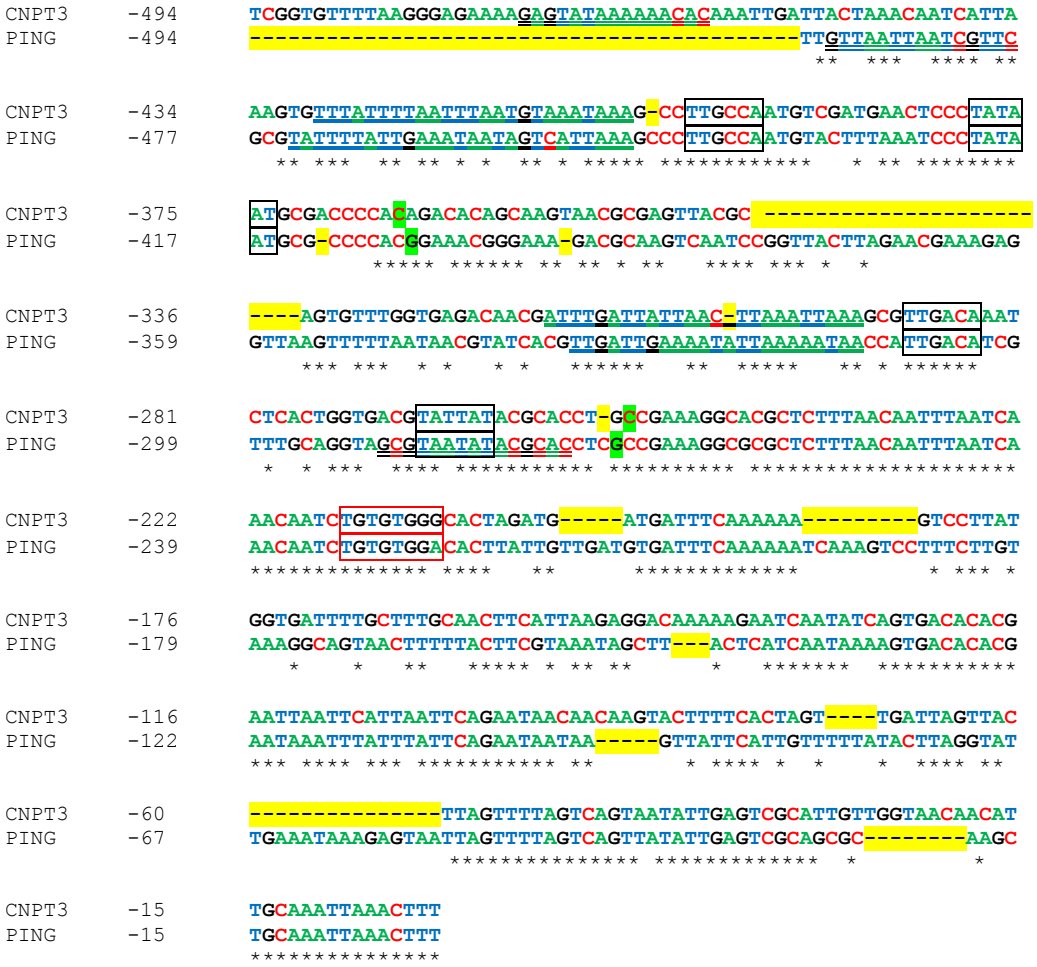


Figure 1.20: Sequence alignment of two representative upstream regions of *Psychromonas* sp. CNPT3 and *Psychromonas ingrahamii*. The '-1' position is immediately upstream from the 16S sequence. Sequences that are underlined are potential UP elements whereas those sequences that are double underlined are potential Fis sites. Elements in black boxes are promoter sequences (either -35 or -10) and bases with a green background are the start of transcription. The *boxC* element is boxed in red.

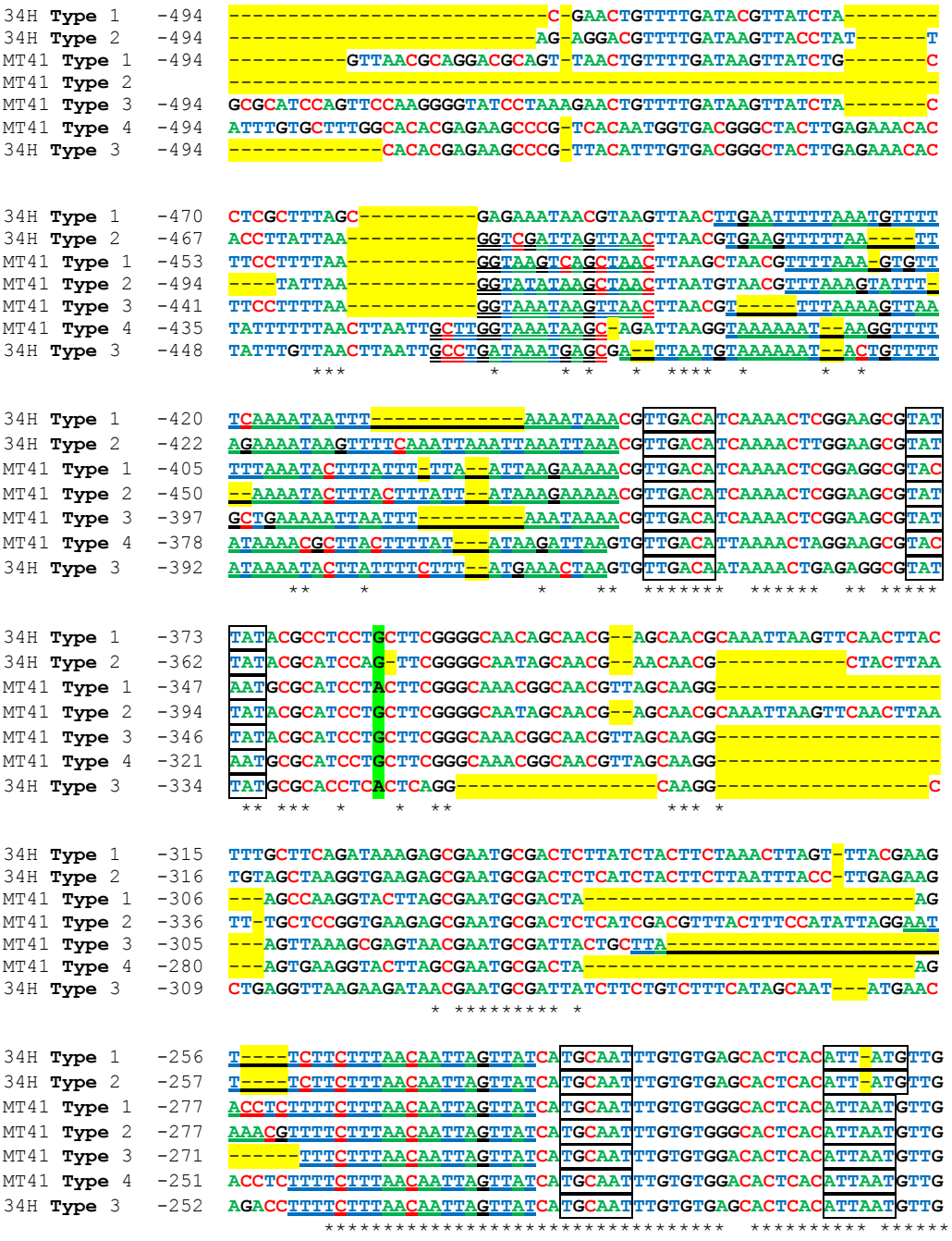


Figure 1.21: Sequence alignment of upstream regions of *Colwellia* sp. MT41 and *Colwellia psychrerythraea* 34H. The '-1' position is immediately upstream from the 16S sequence. Sequences that are underlined are potential UP elements whereas those sequences that are double underlined are potential Fis sites. Boxed elements are promoter sequences (either -35 or -10) and bases with a green background are the start of transcription.

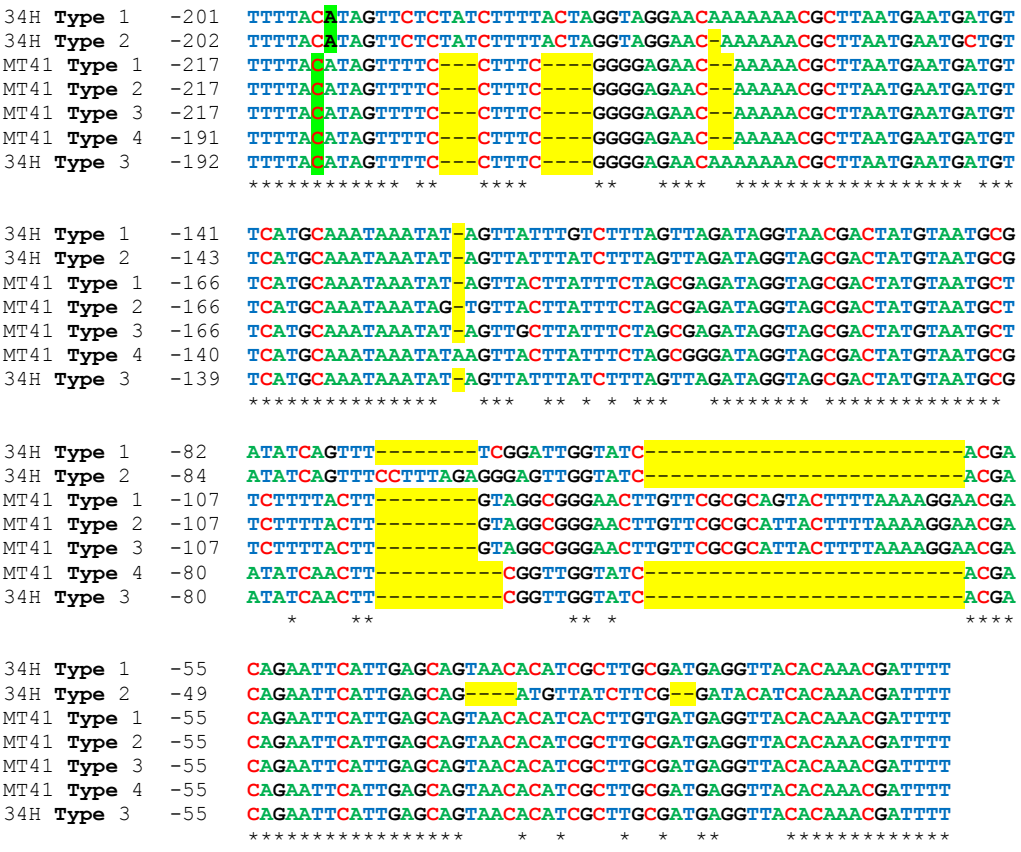


Figure 1.21: Sequence alignment of upstream regions of *Colwellia* sp. MT41 and *Colwellia psychrerythraea* 34H. The '-1' position is immediately upstream from the 16S sequence. Sequences that are underlined are potential UP elements whereas those sequences that are double underlined are potential Fis sites. Boxed elements are promoter sequences (either -35 or -10) and bases with a green background are the start of transcription, Continued.

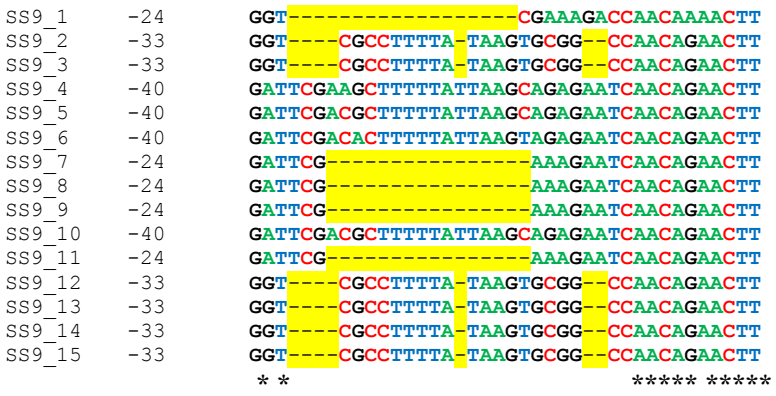


Figure 1.22: Alignment of the region of *Photobacterium profundum* SS9's rRNA operons immediately upstream of the 16S sequence.

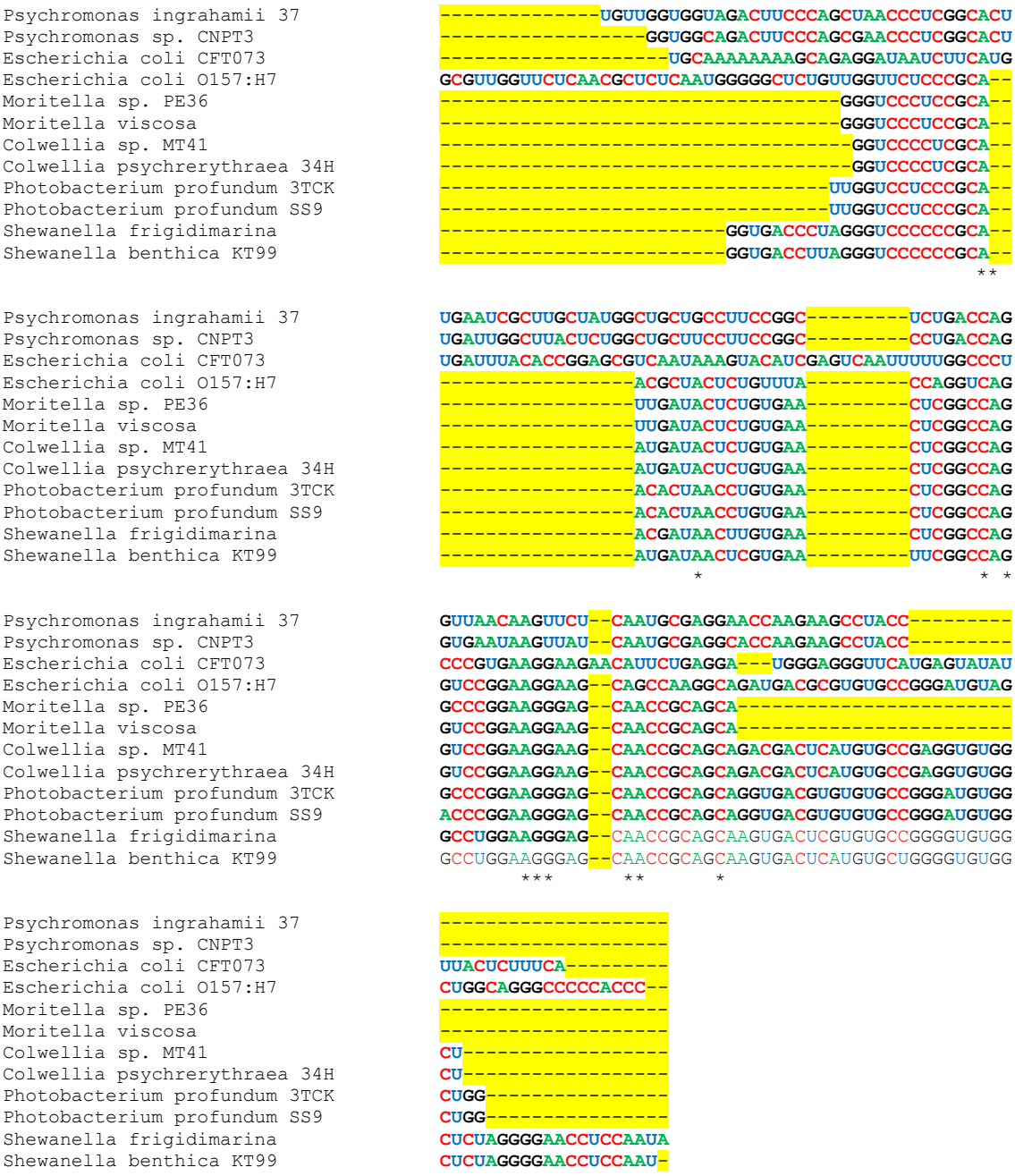


Figure 1.23: Alignment of 4.5S RNA sequences.

Chapter 2:

Modeling of Protein Folding Volume

Introduction

Living in the deep sea, high hydrostatic pressure is one of the environmental conditions that organisms have had to adapt to in order to survive (Bartlett, 1992). The effects of hydrostatic pressure on a protein can be significant, leading to local or global changes to the proteins structure and with high enough pressure, may cause protein denaturation (Mozhaev *et al.*, 1996). The compressibility of a protein is responsible for many of the changes a protein may undergo and is largely determined by the presence of internal cavities formed through protein folding (Mozhaev *et al.*, 1996). Pressures of 1-2 kbar are able to disrupt the binding of oligomeric proteins (Heremans, 1982). It has been shown that regions of the lysozyme protein are considerably compressible (Kundrot and Richards, 1987). Molecular dynamics simulations have been performed that show changes in the protein's hydration shell when subjected to 10kbar of pressure (Kitchen *et al.*, 1992).

In the presence of high hydrostatic pressure, processes with a negative change in volume will occur more rapidly and processes with a positive change in volume will be protracted (Somero, 1990). This effect can be described by Equation 2.1, with K_p being the rate constant at the given

pressure P, K_1 is the rate at atmospheric pressure, ΔV is the process' change in volume, R is the universal gas constant and T is temperature (Johnson *et al.*, 1974).

$$K_p = K_1 e^{-\left(\frac{P\Delta V}{RT}\right)}$$

Equation 2.1

It has been found that an unfolded protein under high hydrostatic conditions actually occupies a smaller volume than the folded protein (Harpaz *et al.*, 1994). This phenomenon is explained by the compressibility of the interface between hydrophobic side-chains and water molecules (Kauzmann, 1959). When a protein folds, there are two separate volume changing events, one in the hydrophobic amino acids and one in the hydrophilic amino acids (Harpaz *et al.*, 1994). As mentioned above, there is a decrease in volume among the hydrophobic amino acids because of the compressibility of the water interface. In the hydrophilic amino acids, there is an increase in volume (Harpaz *et al.*,, 1994). Under atmospheric pressure, these two changes balance each other out and the overall change in volume is less than 0.5% (Brandts *et al.*, 1970; Hawley, 1971; Zipp and Kauzmann, 1973). However, under high hydrostatic pressure, the volume change occurring among the hydrophobic amino acids is not as pronounced and therefore results in an overall increase in volume (Harpaz *et al.*, 1994).

There are many biophysical aspects that must be taken into account when examining protein folding (Matthews, 1993). Equation 1 suggests that protein folding at high hydrostatic pressure occurs more slowly than at atmospheric pressure. If this is true, then the converse would also be true; that protein unfolding would occur faster at high hydrostatic pressure than at atmospheric pressure. With regards to organisms adapted to life in the deep sea, their systems would not only have to be adapted to favor protein folding, but also to protect against protein unfolding. However, it has been shown that only minor changes in the amino acid sequence (Siebenaller, 1984; Mozhaev *et al.*, 1996) are necessary, as even a single mutation in the amino acid sequence can render a protein functional at high hydrostatic pressure (Siebenaller, 1984).

A program developed by Federico Lauro (unpublished, source available upon request) calculates the change in volume upon protein folding of buried residues. This program is based on multiple works (Cohn and Edsall, 1994; Mishra and Ahluwalia, 1983; Cohn *et al.*, 1934; Rao *et al.*, 1984; Jolicoeur *et al.*, 1986) a summary of which is given in Harpaz *et al.* (1994). This summary consists of calculated changes in folding volume for buried amino acids. Using these values, a protein's primary sequence, and the protein's predicted solvent accessibility (see Methods for more detail), Lauro's program calculates the predicted change in volume for buried amino acids upon protein folding. The predicted change in volume was calculated for fifteen

proteins in six organism pairs. From this, I will help elucidate some of the evolutionary adaptations of deep sea microbes.

Methods

For comparison, fifteen proteins were selected to be analyzed. These proteins have been previously found to be involved in pressure sensitivity. See Table 2.1 for the list of proteins and relevant references. The sequences were extracted from the latest versions of each organism's genome (Table 2.2). Not all proteins were found in all organisms, possibly attributable to incomplete genomic sequences (Table 2.3). When annotations of a protein were not available, the *Escherichia coli* protein sequence was used. The BLAST program (Altschul *et al.*, 1997) was used to search the genomic sequence and the best match from the search was used, with a minimum identity of 35%.

Using the primary sequence of each protein, the secondary structure and solute accessibility were computed using SSpro 4.01 and ACCpro 4.01, respectively (Cheng *et al.*, 2005). The solute accessibility prediction classifies each amino acid of a protein as either exposed, intermediate or buried based upon predicted positioning within the three-dimensional protein. Exposed amino acids are the most accessible to solvent and are at or near the protein's surface, whereas buried amino acids are the least accessible to solvent and are in the protein's core. For the change in volume calculation, I counted only

those amino acids determined to be buried. The buried amino acids were chosen as they are thought to be the main factor in the difference of folding volumes in normal and high hydrostatic pressures (Harpaz *et al.*, 1994). The total volume of the buried amino acids was tallied based on values as determined by Harpaz *et al.* (1994) and are reproduced in Table 2.4.

With the folding volumes calculated, the shared proteins of the organism pairs (Table 2.5) were compared to each other and the differences in volume calculated. The average and standard deviation of the differences for each protein were computed using standard statistical methods.

Results

The calculated change in volume (hereafter referred to as ΔV) varied widely in the proteins and organisms analyzed. *Colwellia* sp. MT41 had the largest positive ΔV with AnsA at 71.2 $Å^3$. *Colwellia psychrerythraea 34H* had the largest negative ΔV with CydD having a ΔV of -451.7 $Å^3$ (Table 2.6). The vast majority of volume changes upon folding were negative, with only seven positive volume changes, four in FabB, two in AnsA and one in FabF.

The ΔV values are plotted in Graphs 2.1 and 2.2. Each point represents a piezophile-mesophile protein pair. In Graph 2.1, the points are labeled by protein and in Graph 2.2, they are labeled by organism pair. The points are primarily located around they y = x axis. For both Graphs 2.1 and 2.2, a point that lies above the y = x-axis have a smaller (or more negative) ΔV in the

piezophile protein than the mesophile protein. The points are best clustered by protein, with the exception of CydD, which shows considerable range.

Differences in the ΔV were calculated as the piezophile ΔV – the mesophile ΔV. A positive difference in ΔV indicates that the piezophile protein had a greater ΔV than the mesophile protein and a negative difference in ΔV indicates that the mesophile protein had a greater ΔV. The differences in ΔV between organism pairs were quite varied, ranging from 164.6 $Å^3$ to -83.7 $Å^3$ (Table 2.6, Graph 2.3). When averaged by protein, the range of values fell between 36.55 $Å^3$ and -38.80 $Å^3$ (Table 2.7). These averages are plotted in Graph 2.4 with high-low bars showing one standard deviation.

Graph 2.5 shows the ΔV values for the two most piezophilic organisms of the six tested, *Colwellia* sp. MT41 (Yayanos *et al.*, 1981) and *Shewanella benthica* KT99 (Lauro *et al.*, 2007). In eleven of the proteins, at least one of the organisms differed from the average by 10 $Å^3$, with seven of the proteins having piezophile-mesophile pairs that differ by at least that much. There is also much variance between the two organisms with differences in ΔV as high as 135.5 $Å^3$.

Discussion

From thermodynamics it is known that a process that causes a decrease in volume will occur faster than a process that causes an increase in volume at high hydrostatic pressure (Somero, 1990). Therefore, when this is

applied to protein folding, a protein that when folded occupies less volume than its unfolded form will be more likely to fold and will do so more quickly. This is the biological ideal since a protein's function is often dependent upon its folded form.

It has also been shown that a protein folded under high hydrostatic pressure will increase in volume due to interactions between the protein's hydrophobic amino acids and water. For an organism living under high hydrostatic pressure, this apparently presents a barrier that the organism must overcome. To elucidate how this barrier has been overcome, I examined six piezophiles and fifteen of their proteins that have been associated with piezophily.

Looking at the data from the simulations, it is evident that there is not a clear-cut conclusion that can be drawn. We find that in the proteins being studied, there are many cases in which the piezophile protein has a larger change in folding volume, while in other cases there is a smaller change in folding volume (Graph 2.3, Graph 2.4).

In Graph 2.5, we see that amid the most piezophilic of the organisms, there is a protein, AnsA, which has a larger folding volume than the mesophile pair of one piezophile and a smaller folding volume than the other piezophile. However, Graph 2.5 does show that for the two organisms, their proteins correlate on being larger or smaller than their respective mesophile's protein in nine of thirteen proteins. In two of the non-correlated proteins, the

proteins are less than 2 $Å^3$ from correlating. These proteins may warrant further investigation into their structural characteristics to ascertain if there is a significant structural difference between the piezophilic protein and the mesophilic protein that may be an adaptation to living in the deep sea.

The technique used provides a model for predicting the interior folding volume of a protein, but only an experimentally derived structure would allow us to determine the actual volumes. Although its accuracy is not known, this technique may still provide useful data to guide future analysis.

For future work, one important question to be asked is "What is the folding volume of the exposed amino acids?" This would provide greater insight as to the overall folding volume of the protein. This determination is likely to be more valuable than solely determining the folding volume of the buried residues. If the proteins have been adapted to decrease folding volume in the same manner as non-adapted proteins, the folding volumes of the buried and exposed residues may balance each other out, causing a very small change in overall folding volume.

Another interesting facet that could be examined is the hydration shell of the protein. The hydration shell of proteins under high hydrostatic pressure has been shown to become more ordered and have higher energy (Mozhaev *et al.*, 1996). Examined energetically, this change in hydration may make up for thermodynamic differences relating to protein folding or contribute to protein folding in other aspects.

Living in the deep sea comes with an abundance of challenges to which piezophiles have had to adapt. Infrequent carbon sources make the organisms resourceful while high hydrostatic pressure has significant effects on the biophysical properties of proteins and membranes. The frigid temperature of the water does the same, reducing molecular dynamics and Brownian motion. Yet there is life that lives in, and even thrives, in these conditions.

Table 2.1: List of the fifteen proteins selected for folding volume comparison. References indicated papers that demonstrate the protein's importance in pressure sensitivity.

Protein	Description	Reference
AnsA	L-asparaginase	Lauro, 2007
CydD	ATP-binding/permease protein. Formation of cytochrome D oxidase	Bebbington and Williams, 1993; Kato *et al.*, 1996
DiaA	DnaA initiator-associating protein	Lauro, 2007
DnaK	Heat shock protein; HSP70	Welch *et al.* 1993
FabB	3-oxoacyl-[acyl-carrier-protein] synthase I	Lauro, 2007
FabF	3-oxoacyl-[acyl-carrier-protein] synthase II	Allen and Bartlett, 2000
FtsZ	Cell division protein	Ishii *et al.*, 2004
HNS	Histone-like protein	Dersch *et al.*, 1994; Ishii *et al.* 2005
Mdh	Malate dehydrogenase	Saito and Nakayama, 2004; Saito *et al.*, 2006; Welch and Bartlett, 1997
MreB	Rod shape-determining protein	Jones *et al.* 2001
RecD	Exodeoxyribonuclease V, alpha chain	Bidle and Bartlett, 1999; Finch *et al.*, 1986
RpoE	RNA polymerase sigma-E regulator	Lauro, 2007
RseB	Sigma-E factor regulatory protein	Lauro, 2007
SeqA	Negative regulator of replication initiation	Lauro, 2007
ToxR	Piezo-sensor and transcriptional regulator	Welch and Bartlett, 1998

Table 2.2: List of organisms and the genome accessions used.

Organism	Genome Accession
Carnobacterium sp. AT7	Unpublished Assembly Latest public assembly: GI:159875372 (Lauro *et al.*, 2007)
Colwellia sp. MT41	Unpublished Assembly (Yayanos *et al.*, 1981)
Colwellia psychrerythraea 34H	Public Assembly: GI:71143482 (Methe *et al.*, 2005)
Enterococcus faecalis V583	Public Assembly: GI:29350190, 29345347, 29345328, & 29345255 (Paulsen *et al.*, 2003)
Moritella sp. PE36	Unpublished Assembly Latest public assembly: GI:149809450 (Yayanos *et al.*, 1986)
Moritella viscosa	Unpublished Assembly (Benedikstoddir *et al.*, 2000)
Photobacterium profundum SS9	Public Assembly: GI:47419843, 47419844, & 46911589 (Delong *et al.*, 1997)
Photobacterium profundum 3TCK	Public Assembly: GI:90329665 (Campanaro *et al.*, 2005)
Psychromonas sp. CNPT3	Unpublished Assembly Latest public assembly: GI:90322500 (Yayanos *et al.*, 1979)
Psychromonas ingrahamii 37	Public Assembly: GI:119862398 (Breezee *et al.*, 2004)
Shewanella benthica KT99	Public Assembly: GI:161332323 (Lauro *et al.*, 2007)
Shewanella frigidimarina NCIMB 400	Public Assembly: GI:114332481 (Bowman *et al.*, 1997)

Table 2.3: Intersections marked with a '●' indicate that that protein was found in that organism. Cells shaded in red indicate proteins that were found in one organism, but not its comparative organism.

	Carnobacterium sp. AT7	*Enterococcus faecalis* V583	*Colwellia* sp. MT41	*Colwellia psychrerythraea* 34H	*Moritella* sp. PE36	*Moritella viscosa*	*Photobacterium profundum* SS9	*Photobacterium profundum* 3TCK	*Psychromonas* sp. CNPT3	*Psychromonas ingrahamii* 37	*Shewanella benthica* KT99	*Shewanella frigidimarina* NCIMB 400
AnsA	●	●	●	●	●	●	●	●	●	●	●	●
CydD	●	●	●	●	●	●	●	●	●	●	●	●
DiaA	●	●	●	●	●	●	●	●	●	●	●	●
DnaK	●	●	●	●	●	●	●	●	●	●	●	●
FabB			●	●	●	●	●	●	●	●	●	●
FabF	●	●	●	●	●	●	●	●	●	●	●	●
FtsZ	●	●	●	●		●	●	●	●	●	●	●
H-NS					●	●	●	●			●	●
Mdh			●	●	●	●	●	●	●	●	●	●
MreB			●	●	●	●	●	●	●	●	●	●
RecD	●	●	●	●	●	●	●	●	●	●	●	●
RpoE		●	●	●	●	●	●	●	●	●	●	●
RseB			●	●	●	●	●	●	●	●	●	●
SeqA			●	●	●	●	●	●	●	●	●	●
ToxR							●	●	●	●		

Table 2.4: Changes in amino acid side chain volume upon protein folding.

Amino Acid	Changes in Side Chain Volume on Folding (Å^3)
Ala	-2.6
Arg	7.9
Asn	7.0
Asp	11.9
Cys	-1.0
Gln	1.3
Glu	8.5
Gly	0
His	3.3
Ile	-2.6
Leu	-6.2
Lys	7.6
Met	0.7
Phe	-0.9
Pro	-6.2
Ser	1.4
Thr	0.3
Trp	0.6
Tyr	-0.3
Val	-3.6

Table 2.5: Pairings of organisms for genomic comparison.

Piezophile	Mesophile
Carnobacterium sp. AT7	*Enterococcus faecalis* V583
Colwellia sp. MT41	*Colwellia psychrerythraea* 34H
Moritella sp. PE36	*Moritella viscosa*
Photobacterium profundum SS9	*Photobacterium profundum* 3TCK
Psychromonas sp. CNPT3	*Psychromonas ingrahamii* 37
Shewanella benthica KT99	*Shewanella frigidimarina* NCIMB 400

Table 2.6: Calculated change in volume for specified proteins in Å^3. Organisms with two values for a given protein had two separate copies of that protein.

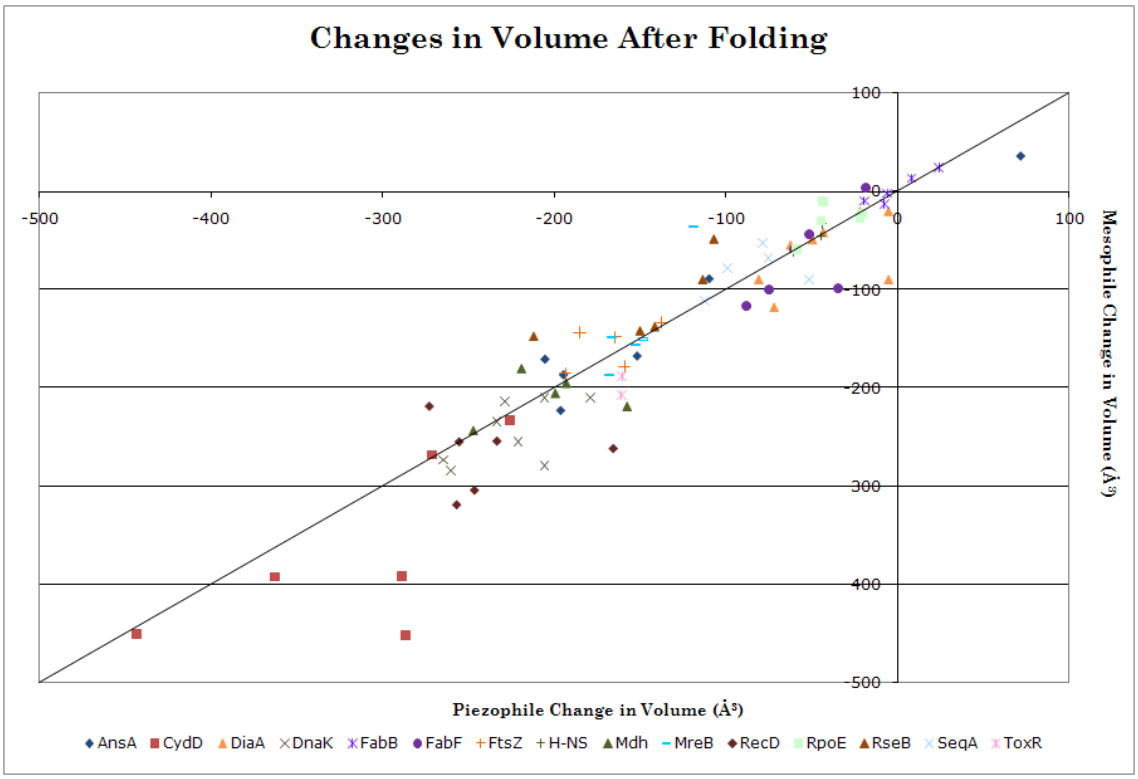
	***Carnobacterium* sp. AT7**	***Enterococcus faecalis* V583**	**Difference in Change in Volume**	***Colwellia* sp. MT41**	***Colwellia psychrerythraea* 34H**	**Difference in Change in Volume**	***Moritella* sp. PE36**	***Moritella viscosa***	**Difference in Change in Volume**
AnsA	-110	-89	-21	71.2	35.7	35.5	-194.7	-186.6	-8.1
CydD	-226.2	-232.5	6.3	-287.1	-451.7	164.6	-443.8	-450.5	6.7
DiaA	-72.1	-117.6	45.5	-43.5	-41.5	-2	-62.6	-54.3	-8.3
DnaK	-221.4	-254.5	33.1	-206.2	-278.4 -210	72.2 3.8	-229.2	-213.4	-15.8
FabB	-	-	-	-19.4	-9.1	-10.3	-7.9	-12.7	4.8
FabF	-88.1	-116.3	28.2	-75.1	-99.3	24.2	-51.6	-43	-8.6
FtsZ	-159.3	-178.1	18.8	-193.7	-185.4	-8.3	-	-	-
H-NS	-	-	-	-	-	-	-44.6	-44.7	0.1
Mdh	-	-	-	-247.4	-243.1	-4.3	-219.4	-180.4	-39
MreB	-	-	-	-148.3	-150.4	2.1	-118.9	-35.2	-83.7
RecD	-272.8	-217.7	-55.1	-233.6	-253.6	19.7	-246.6	-303.5	56.9
RpoE	-	-	-	-20.5	-22.1	1.6	-21.8	-26.7	4.9
RseB	-	-	-	-150.2	-140.9	-9.3	-107.2	-47.6	-59.6
SeqA	-	-	-	-112.6	-110.8	-1.8	-99.1	-77.4	-21.7
ToxR	-	-	-	-	-	-	-	-	-

Table 2.6: Calculated change in volume for specified proteins in Å^3. Organisms with two values for a given protein had two separate copies of that protein, Continued.

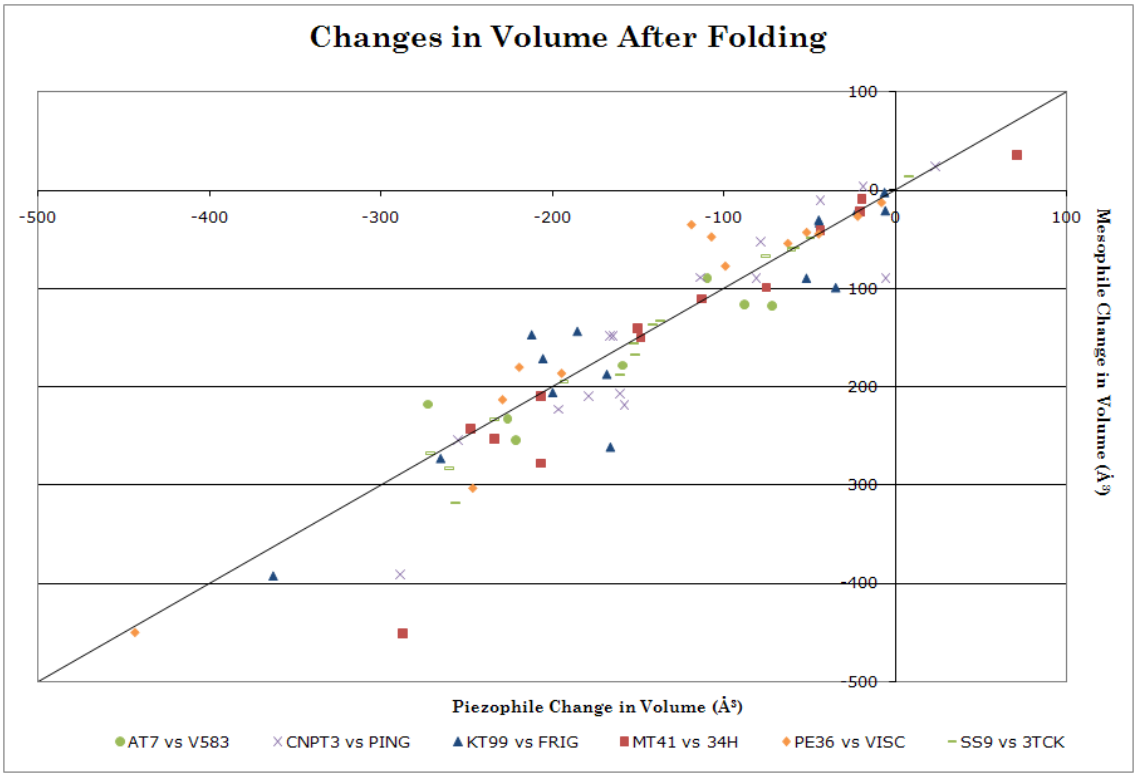
	***Photobacterium profundum* SS9**	***Photobacterium profundum* 3TCK**	**Difference in Change in Volume**	***Psychromonas* sp. CNPT3**	***Psychromonas ingrahamii* 37**	**Difference in Change in Volume**	***Shewanella benthica* KT99**	***Shewanella frigidimarina* NCIMB 400**	**Difference in Change in Volume**
AnsA	-151.9	-167.5	15.6	-196.4	-222.9	26.5	-205.5	-170.8	-34.7
CydD	-271.6	-268	-3.6	-289.2	-391.4	102.2	-363.1	-392.2	29.1
DiaA	-49.8	-48.8	-1	-5.4 -81.1	-89.3	83.9 8.2	-22.1 -5.5	-20.1	-2 14.6
DnaK	-234 -260.7	-233.4 -283.6	-0.6 22.9	-179.2	-209.5	30.3	-265.3	-272.5	7.2
FabB	8	14	-6	23.6	25	-1.4	-6	-1.6	-4.4
FabF	-60.4	-58.8	-1.6	-18.6	4	-22.6	-34.5	-98.5	64
FtsZ	-137.5	-132.7	-4.8	-164.8	-148	-16.8	-185.4	-143	-42.4
H-NS	-60.8	-60.8	0	-	-	-	-44	-32.4	-11.6
Mdh	-193.5	-195.1	1.6	-158.1	-218.7	60.6	-199.8	-205.4	5.6
MreB	152.7	-156.3	3.6	-166.6	-148.4	-18.2	-168.2	-186.8	18.6
RecD	-257	-318.4	61.4	-255.5	-254	-1.5	-166.1	-260.9	94.8
RpoE	-59	-59	0	-43.6	-9.4	-34.2	-44.4	-29.6	-14.8
RseB	-141.6	-136.6	-5	-113.7	-88.7	-25	-212.1	-146.5	-65.6
SeqA	-75.6	-67.1	-8.5	-78.6	-51.9	-26.7	-51.6	-88.9	37.3
ToxR	-160.7	-187.9	27.2	-160.8	-206.7	45.9	-	-	-

Table 2.7: Average differences in the change in volume and the standard deviation for each protein.

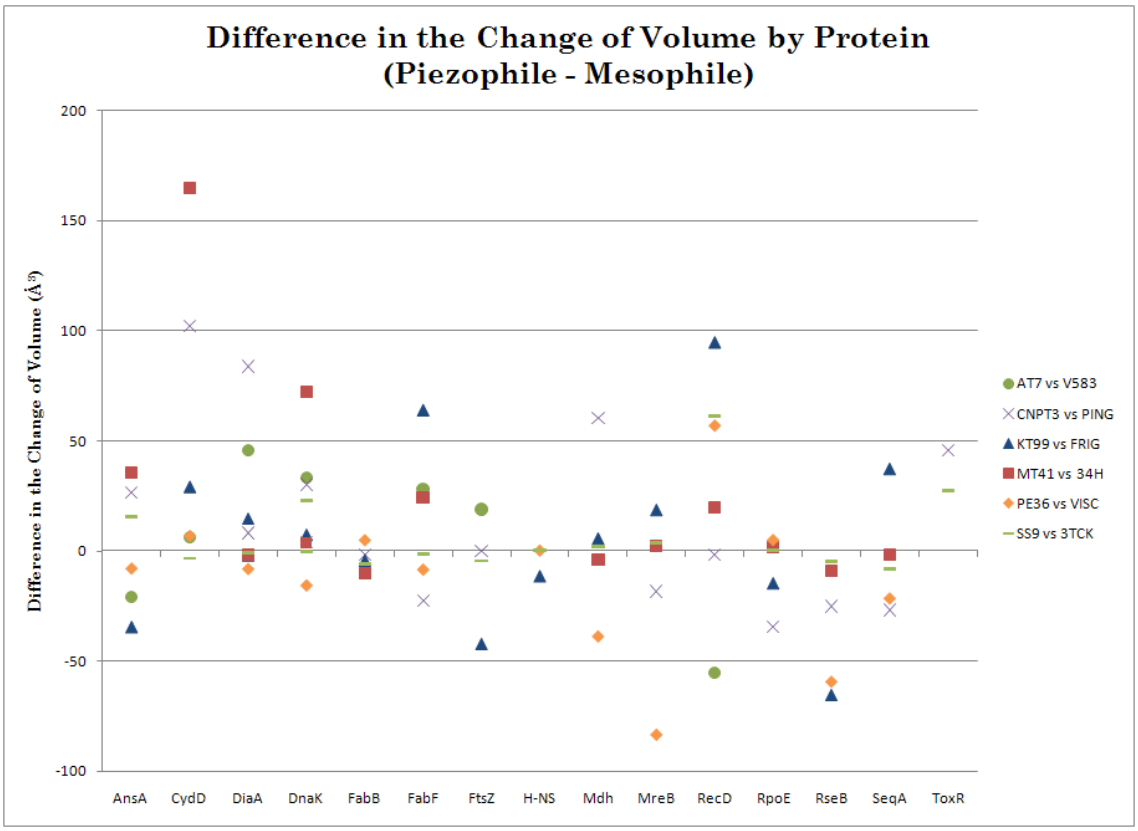
	Average Difference in the Change in Volume	Standard Deviation
AnsA	2.30	27.87
CydD	50.88	67.75
DiaA	17.36	31.75
DnaK	19.14	27.09
FabB	-3.46	5.63
FabF	13.93	31.30
FtsZ	-10.70	22.10
H-NS	-3.83	6.73
Mdh	4.90	35.80
MreB	-15.52	40.30
RecD	29.37	53.38
RpoE	-8.50	16.23
RseB	-32.90	28.20
SeqA	-4.28	25.29
ToxR	36.55	13.22



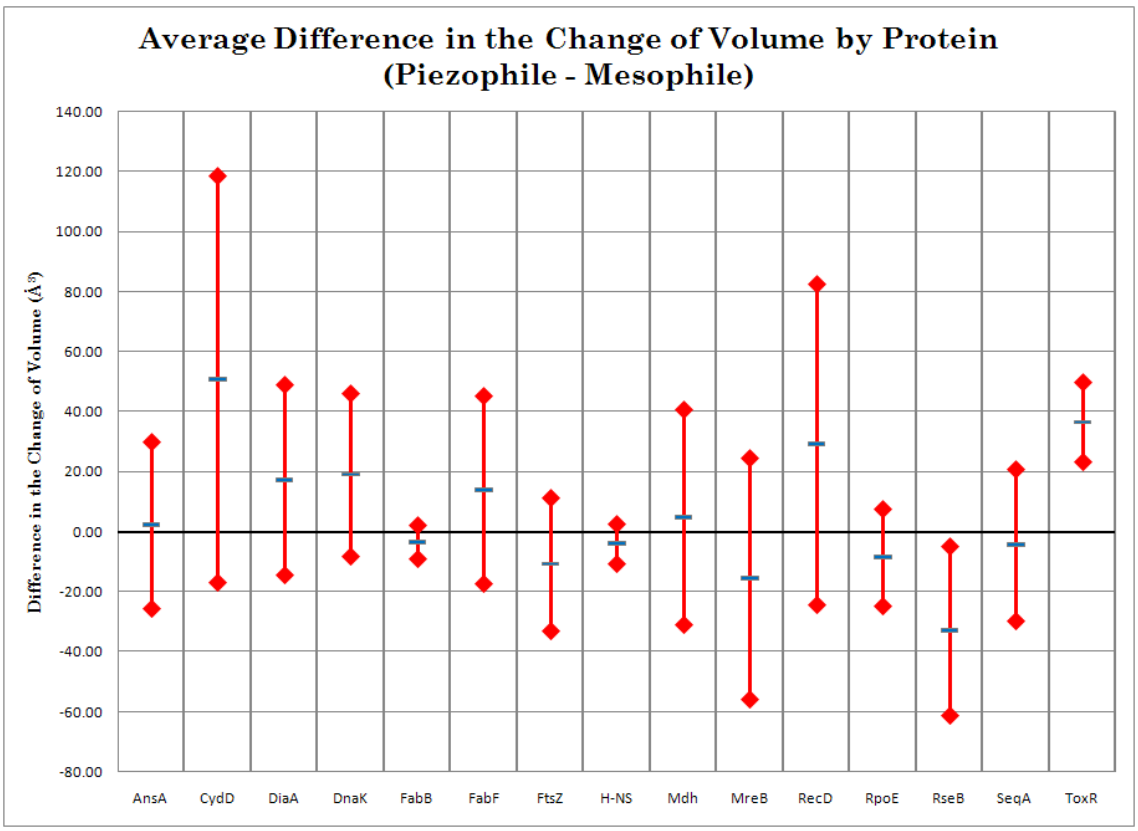
Graph 2.1: Plot of the changes in volume of the buried amino acids, grouped by protein. Each point represents one piezophile - mesophile pair. The diagonal line is the y = x axis.



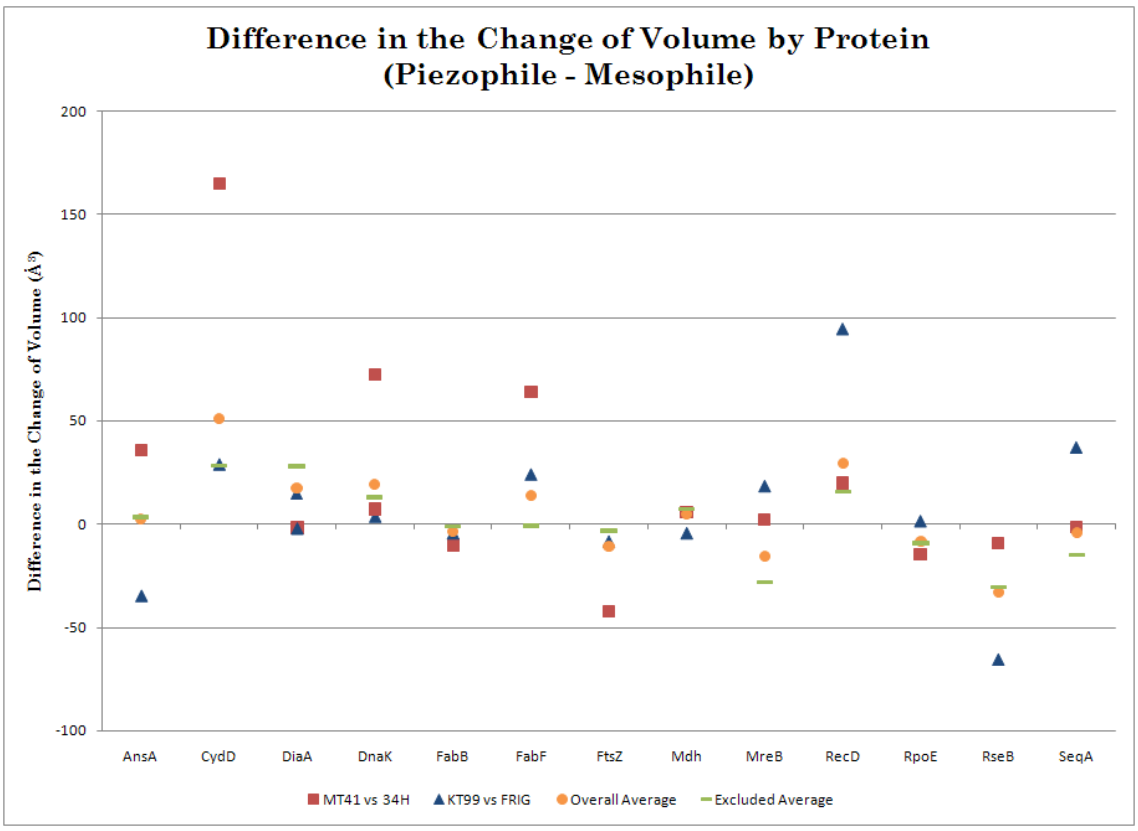
Graph 2.2: Plot of the changes in volume of the buried amino acids, grouped by piezophile - mesophile pair. The diagonal line is the y = x axis.



Graph 2.3: Plot of the difference in the change in volume of the buried amino acids.



Graph 2.4: Plot of the average change in volume of the buried amino acids upon protein folding. The blue dash indicates the average and the high-low bars represent one standard deviation.



Graph 2.5: Plot of the difference in the change in volume of the buried amino acids for the two most piezophilic organisms. The overall average of the difference in the change of volume is also plotted for each protein.

Chapter 3:

Genomic Sequencing and Analysis of *Moritella* sp. PE36

Introduction

In 1990, it was suggested that the two genera *Vibrio* and *Photobacterium* be split up into five genera based on molecular analysis: *Vibrio*, *Photobacterium*, *Listonella*, *Colwellia* and *Moritella* (Steven, 1990). The genus *Moritella* was named after marine microbiologist Richard Y. Morita, Professor Emeritus at Oregon State University (Urakawa *et al.*, 1998). The genus was separated from the others and was created within the order Alteromonadales, family Alteromonadaceae, remaining in the Gamma Proteobacteria class (Steven, 1990). The family Moritellaceae was later suggested and created to include the single genus *Moritella* (Ivanova, 2004), though it now contains a second genus, *Paramoritella* (Hosoya *et al.*, 2008).

Moritella are gram-negative, rod-shaped bacteria that form convex circular colonies, which are opaque and cream in color (Urakawa *et al.*, 1998). They are motile via a polar flagellum (Urakawa *et al.*, 1998). The *Moritella* are chemo-organotrophic, facultatively anaerobic, and halophilic (Urakawa *et al.*, 1998). In addition, most strains of *Moritella* are able to reduce nitrate to nitrite (Urakawa *et al.*, 1998). No gas is produced from the utilization of D-glucose or D-fructose, although acid is (Urakawa *et al.*, 1998). All strains

produce negative results for H_2S production tests and do not contain arginine dihydrolase, lysine decarboxylase, or ornithine decarboxylase (Urakawa *et al.*, 1998). The genus has been isolated from seawater, marine sediments, the deep sea, and fish farms (Kim *et al.*, 2008). One species, *Moritella viscosa,* is a fish pathogen, causing "winter ulcer" in salmonid fish (Benediktsdóttier *et al.*, 2000). Psychrophilicity is a characteristic of the genus, with some members also being piezophilic (Kim *et al.*, 2008).

There are currently seven species of *Moritella*: *Moritella marina* (originally classified as *Vibrio marinus* (Russel, 1892; Colwell and Morita, 1964)) (Urakawa *et al.*, 1998); *Moritella japonica* (Nogi *et al.*, 1998); *Moritella yayanosii* (Nogi and Kato, 1999); *Moritella viscosa* (originally classified as *Vibrio viscosus* (Lunder *et al.*, 2000)) (Benediktsdóttier *et al.*, 2000); *Moritella profunda* (Xu *et al.*, 2003); *Moritella abyssi* (Xu *et al.*, 2003); and *Moritella dasanensis* (Kim *et al.*, 2008). Of these, *japonica*, *yayanosii, profunda, and abyssi* are piezophilic (Nogi *et al.*, 1998; Nogi and Kato, 1999; Xu *et al.*, 2003). Thirty separate *Moritella* strains have been isolated from deep sea sediment and the intestinal contents of deep sea fish (Saito and Nakayama, 2004; Nakayama *et al.*, 2005). Of these, 18 have been determined to be facultatively piezophilic and six as obligately piezophilic (Nakayama *et al.*, 1994; Yano *et al.*, 1995; Yano, 1996; Saito and Nakayama, 2004).

At Scripps Institution of Oceanography, *Moritella* sp. strain PE36 (hereafter referred to as PE36) is maintained as part of a collection of

piezophilic bacteria. PE36 was isolated beyond the Patton Escarpment in the North Pacific Ocean, approximately 288 km off the coast of San Diego, California at a depth of 3584 m (DeLong and Yayanos, 1986). PE36 was isolated from the water surrounding a trapped amphipod (DeLong and Yayanos, 1986). PE36 is classified as a piezopsychrophile (Yayanos, 1995), but it is actually facultatively piezophilic; able to grow weakly at atmospheric pressure (DeLong *et al.*, 1997). Its optimum growth pressure is approximately 41.4 MPa, close to the pressure of its isolation depth (a ten meter increase in depth increases the pressure by 0.1 MPa) (DeLong *et al.*, 1997). PE36's optimum growth temperature (≈10°C) is significantly higher than its isolation temperature (≈2°C) (Yayanos *et al.*, 1986). Growth of PE36 is arrested at a temperature of 20°C at any pressure and at 13°C at atmospheric pressure (Yayanos *et al.*, 1986).

Like the other *Moritella*, PE36 is a chemo-organotroph and in support of this is the finding of a Phosphenolpyruvate (PEP):sugar phosphotransferase system (PTS) for glucose, which was found to have a higher glucose uptake rate at high hydrostatic pressure than at atmospheric pressure (DeLong and Yayanos, 1987). However, glucose, as well as other organic energy sources, is not always available to microbes in cold, deep sea environments (Lauro and Bartlett, 2007). One possible way PE36 has adapted to this condition is through the use of phosphorylated guanidines, including arginine and creatine, which are used to store high-energy

phosphates and then are utilized to generate ATP during episodes of high-metabolic demand (Andrews *et al.*, 2008). *Moritella* sp. PE36 was found to possess a putative arginine kinase (Andrews *et al.*, 2008) and is not present in other piezophiles or *Moritella viscosa*.

Polyunsaturated fatty acids, or PUFAs, have been found to comprise a large percentage of the membranes of PE36. At 34.5 MPa, the ω-6 fatty acid docosahexaenoic acid (DHA) accounted for almost 24.7% (wt %) of PE36's membrane (DeLong and Yayanos, 1986). This percentage dropped to 15.7% at atmospheric pressure, while at an increased pressure of 62.0 MPa, the composition rose to 29.7% (DeLong and Yayanos, 1986). As with psychrophiles and other piezophiles, the presence of PUFAs in the membrane is thought to help maintain membrane fluidity and functionality (Simonato *et al.*, 2006).

Sequencing and Assembly

Whole Genome Shotgun Sequencing

To achieve a first draft of PE36's genome, a clone library of the genome was made. The library consisted of small, 4kb plasmid inserts and large, 40kb fosmid inserts and was constructed as described in Goldberg *et al.* (2006). Fosmids are similar to cosmids in size and the delivery to bacterial cells is via phage, but are based on the bacterial F-factor, which makes the vector more stable and less prone to recombination (Kim *et al.*, 1992; Shizuya *et al.*, 1992).

The pCC1FOS vector used is a single-copy fosmid under control of the F-factor origin of replication (ORI), but can be induced to multi-copy through an arabinose-controlled ORI. The vector also contains a chloramphenicol resistance gene, allowing for selective growth.

Whole genome shotgun sequencing (Fleischmann *et al.*, 1995) was then performed on the library, sequencing the 3′ and 5′ ends of each library insert. The sequencing was done by the J. Craig Venter Institute's Joint Technology Center with Applied Biosystems 3730XL DNA sequencers (Applied Biosystems, Foster City, CA) with funding provided by the Gordon and Betty Moore Foundation's Marine Microbial Genome Sequencing Project.

Initial Assembly

Knowing which pair of reads went together, the end sequences, or reads, were assembled into contiguous segments, or contigs. A contig is a gapless stretch of sequence formed by overlapping reads. The actual assembly of the reads was done computationally using the Celera assembler (Myers, 2000). The draft assembly has been deposited in GenBank under the accession ID ABCQ00000000 and contains 131 contigs.

However, due to bias in building the genomic library and in sequencing, some regions may be left without any sequence coverage. Using the pairwise data from the reads, contigs can be linked together into scaffolds, with gaps forming between the two contigs. This was done using

pairs of reads that had reads on separate contigs. Commonly, more than one scaffold will be left due to a lack of connecting read pairs between some contigs, resulting in a "hard stop" at the end of each scaffold. The scaffolds allow the creation of a contig map, which shows the orientation of each contig relative to others in the same scaffold. This information is then used to close the genome.

Closing the Genome

The initial assembly of 131 contigs and 10 scaffolds was imported into Consed, a program for visualizing and editing genome assemblies (Gordon *et al.*, 1998). Scaffolds and contig orientation were visualized using Hawkeye (Schatz *et al.*, 2007). For base-calling of new sequence, Phred was used (Ewing *et al.*, 1998) and Phrap was used for any additional assembly work (Ewing and Green, 1998). Using these tools, strategies were devised for closing the genome. Three different situations were encountered, each needing its own sequencing strategy.

For most of the sequencing, DNA was sequenced from the fosmid libraries. When an insert was needed from the library, the appropriate clone was grown from a glycerol stock in 6 ml of lysogeny broth (LB) with 30 µg/ml chloramphenicol. After incubating overnight at 37°C in a shaker, 600 µl of the culture was transferred to fresh LB with chloramphenicol. This sample

was incubated for an hour after which the multicopy induction solution was added (1% L-arabinose). This was again grown overnight in a 37°C shaker.

DNA from the libraries was isolated in two different manners. The first was done using a standard alkaline lysis. This was followed by isopropanol precipitation and two wash steps in 70% ethanol. The pelleted DNA was dried using a SpeedVac (Savant) for 15 minutes and then stored at -2°C. Samples were rehydrated and concentrated in 10 mM Tris buffer and evaluation of the DNA was done using agarose gel electrophoresis.

For isolating a large quantity of insert DNA, the PureLink HiPure Plasmid Midiprep system was used (Invitrogen, Carlsbad, CA). Purified DNA was precipitated, concentrated, and evaluated as described above.

The first situation encountered was the sequencing of the ribosomal RNA operons. In microbial genomes, there are often multiple rRNA operons, usually very similar in sequence (Nomura *et al.*, 1977). PE36 has a predicted 12-14 rRNA operons. The range comes from 12 predicted 23S rRNA genes, which each indicate an operon, and 14 5S rRNA genes, which may be present in one or more copies per operon. To sequence these, ribosome specific primers were designed and sequencing was done off of library clones that contained only one operon to ensure accurate sequencing. The reads generated from this sequencing were used to identify reads from the genomic assembly that belonged to this operon. Additional reads of the operon were sequenced as needed. All of the reads associated with the operons were then

assembled separately from the genome and collapsed into a surrogate sequence consisting of a single "read." The surrogate sequence was then inserted into the genomic assembly.

The second situation encountered in closing the genome was filling the gaps between contigs on the same scaffold. These gaps are covered by one or more library clones, from which sequencing was done. Sequencing primers for both ends of the gap were designed using Consed and appropriate library clones were identified using Hawkeye. Appropriate clones were those that spanned the gap entirely, allowing for sequencing of both ends of the gap from the same clone. Also, when possible, three clones were selected for sequencing. The process was repeated as needed, using the new sequence to design new primers for further sequencing, a process known as "primer walking" (Voss *et al.*, 1993).

The last closing situation dealt with the hard stops at the end of each scaffold. The hard stops are formed as a consequence of bias in library formation and in sequencing, both of which may be due to large repeats or secondary structure (Goldbert *et al.*, 2006). Scaffolds were linked through combinatorial PCR (Kapitonov *et al.*, 1999) in which a PCR primer was designed for each of the scaffold ends. Through the PCR results, some scaffolds were able to be linked together by the presence of a successful PCR reaction between two end primers. The PCR reaction is repeated multiple times to account for errors produced by the *Taq* polymerase (Tindall and

Kunkel, 1988). Sequencing to fill the gap is done off of these PCR products, using primer walking when needed.

The remaining scaffolds were unable to be linked together, either because of ambiguous results (multiple PCR reactions for one primer) or lack of a PCR reaction. A lack of a reaction may be indicative of a prohibitory gap size, strong secondary structure, or other factor.

Current Status

The initial assembly of the genome consisted of 131 contigs and 10 scaffolds. The total sequence length was 5,164,151 bp; 50,709 bp of which are in contig 595, a predicted plasmid. Since the original assembly, 123,673 bp of new sequence has been added, bringing the total sequence length to 5,287,824 bp. The assembly has been condensed to 15 contigs and five scaffolds, one of which is the predicted plasmid whose length is 49,993 bp. The average contig size is 374,024 bp with three contigs being greater than 1 Mbp, accounting for 3,970,972 bp.

The remaining gaps in the genome have proven difficult to sequence. A number of the gaps are spanned by ribosomal or repeated sequences. The remaining hard stops are the result of difficult-to-PCR regions, examples of which are covered in the previous section. Figure 3.1 is a contig map showing the current assembly and indicates gaps spanned by ribosomal or repeated sequences.

Results and Discussion

Annotation on the draft genome assembly was performed by the Joint Genome Institute at the Department of Energy. The annotation is hosted in and available in version 2.4 of the Integrated Microbial Genomes (IMG) system (Markowitz *et al.*, 2006). Although no other *Moritella* genome has been sequenced and released to the public, sequence comparisons to PE36 were made using an unpublished draft genome (175 contigs) of *Moritella viscosa* (Benedikstoddir *et al.*, 2000). No annotation is available for *Moritella viscosa*, so for comparison of annotations, the psychrophilic *Shewanella frigidimarina* NCIMB 400 will be used (Bowman *et al.*, 1997). Both *Moritella* and *Shewanella* are members of the order Alteromonadales and the closeness of the two genera has been noted by comparison of their 16S rRNA sequence (Urakawa *et al.*, 1998). *Shewanella frigidimarina* NCIMB 400 was chosen because it is a psychrophile, which is a characteristic of the *Moritella* genus and because it has been used as a comparison organism for another piezophile, *Shewanella benthica* KT99 (Taylor, 2008).

Genome Statistics

PE36's genome consists of two circular molecules, one large chromosome and one plasmid. The genome has a GC content of 41.02% and has a predicted 4,781 protein coding genes, 33 rRNA genes and 129 tRNA

genes. Of the 4,781 protein coding genes, 2,930 have a predicted function while 254 are hypothetical proteins. The remaining 1,597 genes are "conserved hypothetical genes" or genes that share sequence similarity with genes in other organisms, but have not been experimentally verified. Table 3.1 summarizes much of the information presented in this paragraph and some of the data that is presented in the following paragraphs.

Using the predicted open reading frames (ORFs) from IMG, codon counts and frequencies were determined using a custom Perl script. When this was done, 263 ORFs were identified as either having a frameshift or being truncated as the gene did not have 3n nucleotides. Table 3.2 contains the list of ORFs that were identified. ATG (AUG) was the most common start codon accounting for 90.5% of the genes. The start codons GTG (GUG) and TTG (UUG) accounted for 6.4% and 3.1%, respectively. Table 3.3 shows the full statistics for all of the codons and the resulting amino acids.

Ribosomal RNA prediction was performed on the most recent assembly of the PE36's genome using the program RNAmmer (Lagesen *et al.*, 2007). RNAmmer's prediction yielded nine 16S genes, ten 23S genes and fourteen 5S genes; results that differed from the original IMG prediction. The rRNA genes are located in ten operons, two of which have two 5S genes encoded. There are two additional 5S genes that are not associated with other rRNA genes, so these may represent an error in prediction or may have some biological significance. There is also one operon that is lacking a predicted

16S rRNA gene and this too may represent an error in prediction or it may represent an obsolete operon that is no longer expressed.

As with the ribosomal RNAs, tRNA prediction was also performed on the most recent assembly of PE36's genome. To do this, we used the program tRNAscan-SE (Lowe and Eddy, 1997). The original IMG prediction identified 121 tRNA genes while the new prediction identified 129. Two of these tRNAs are present on the plasmid, however tRNAscan-SE was unable to determine the anticodon of one of genes. Overall, the methionine tRNA occurred most often with thirteen genes accounting for over 10% of all tRNAs. Also of interest is that only 33 anticodons are represented in the genome, though all codons are used extensively throughout the genes. This may indicate that PE36 relies heavily on the wobble base pairing to translate proteins. A full list of all tRNAs found and corresponding anticodons are shown in Table 3.4. Graph 3.1 shows the correlation between amino acid usage and tRNA presence.

The chromosome's current size is 5,236,340 bp, though it is not closed, so the size is currently underreported. Of the 5.2 Mbp in the chromosome, 86.4% are predicted to be within protein coding genes. On the chromosome, there is a density of 0.907 genes per kilobase with an average of 966 bp per gene.

The plasmid, which is closed, is 49,993 bp in length and is made up of one contig (contig 595). Contig 595 has a size 50,709 bp, but 716 bp is

overlapping with itself on the ends. The contig was trimmed by 383 bp on the 5′ end and 333 bp on the 3′ end. The plasmid has a high density of genes at 1.140 genes per kb. 91.9% of the plasmid is coding sequence.

COG and KEGG

With the large number of genomes being sequenced, it has become necessary to start classifying genes and proteins into category levels higher than gene name and exact function. The first major introduction of a classification scheme was clusters of orthologous groups, or COGs (Tatusov *et al.*, 1997). COGs are groups of proteins that share orthologous relationships based on sequence similarity (Tatusov *et al.*, 1997).

Another classification scheme was developed in 1958 by Dixon and Webb. This scheme was specifically for enzymes and classified the enzymes by the type of reaction they catalyzed (Tipton and Boyce, 2000). Each enzyme was assigned four numbers; each separated by a period and collectively called its EC number (Tipton and Boyce, 2000). Each number represents a different tier in the classification system, the first number being the highest and broadest tier (Tipton and Boyce, 2000). The EC numbers have been further grouped into functional and pathway categories in the Kyoto encyclopedia of genes and genomes, or KEGG (Kanehisa and Goto, 2000).

The IMG annotation of PE36 includes assigned COG and KEGG categories for a number of the predicted genes (Markowitz *et al.*, 2006). Of the

4,781 predicted protein coding genes, 3,445 were assigned at least one COG and 550 were assigned an EC number, 474 which were part of a KEGG pathway. The distribution of proteins amongst categories is not even and when compared to the distribution in another organism, it can lend some insight into the evolutionary significance of the distribution.

PE36 has approximately 9.4% more genes assigned to a COG than does *Shewanella frigidimarina* NCIMB 400 (hereafter referred to as FRIG). Because of this disparity, I will highlight only the differences that represent at least a 20% differentiation between PE36 and FRIG. Additionally, both organisms have approximately the same number of proteins categorized in a KEGG pathway.

The COG with the greatest divergence between the two organisms is cell motility with PE36 having 45% more genes than FRIG in this category. More than half of the extra proteins PE36 has are flagellum-related proteins; in fact PE36 possesses two flagellar loci, one putatively for a polar flagellum and one putatively for a lateral flagellum. However, a BLAST search reveals that *Moritella viscosa* ostensibly has two flagellar loci as well, and therefore the number of flagellar systems is not likely an adaptation for piezophily.

PE36 also has almost twice as many chemotaxis proteins as FRIG, though it is unknown if this is a species-specific or a piezophile-specific trait. Taylor notes in her 2008 thesis that *Psychromonas* sp. CNPT3 does have a number of chemotaxis genes more than its comparison strain, *Psychromonas*

ingrahamii 37. Taylor also notes that *Psychromonas* sp. CNPT3 has 19 methyl-accepting chemotaxis proteins with *Psychromonas ingrahamii* 37 having none. PE36 has 37 of these proteins, while FRIG has 24. The similarity in findings in PE36 as compared to those in *Psychromonas* sp. CNPT3 would afford evidence towards an increase in chemotaxis genes as being an adaptation to the deep sea, but the evidence is not definitive. PE36 would also need to be compared directly to a closer relative, such as *Moritella viscosa*.

PE36 has 37% more genes than FRIG in the category of intracellular trafficking, secretion and vesicular transport. Nine of the 39 extra genes are flagellum-related proteins and eleven are pilus-related proteins, mainly type IV. The remaining 19 extra proteins are distributed amongst different functions. Having this extra infrastructure for protein trafficking and secretion is not surprising when it is noted that PE36 has 865 more genes, or 169% more, than FRIG with a predicted signal peptide. This increase in the occurrence of signal peptides is not seen in other piezophiles such as *Photobacterium profundum* SS9 and *Psychromonas* sp. CNPT3, so this may be a singular trait of *Moritella* and not of piezophily.

The other areas in which FRIG has a significant number of proteins less than PE36 are amino acid transport and metabolism, carbohydrate transport and metabolism, and transcription. The increase in the number of proteins that PE36 has in these areas could be reflective of the environment

in which it lives. In the deep sea there is not always a steady supply of nutrients and therefore, by having these extra proteins and the transcriptional regulators to control their expression, PE36 may be able to take advantage of a broader variety of nutrients, utilizing whatever nutrients are available to it.

PE36 showed a large reduction in the number of genes dedicated to the replication, recombination and repair of DNA when compared to FRIG. Of the 52 extra proteins in FRIG, 45 can be attributed to an increase in the number of annotated transposases (+24) and integrases (+21). PE36 has only one annotated DNA-lyase, whereas FRIG has four, including two photo-lyases. The lack of a photo-lyase in PE36 is expected since the deep sea is mostly devoid of light radiation.

For a more general analysis, Graph 3.2 shows the percentage distribution of each COG category in the genome and Graph 3.3 shows a comparison of the COG distribution in PE36 and FRIG. Table 3.4 lists the exact number of proteins in each category for both PE36 and FRIG.

KEGG categories are fewer in number than COG categories (12 vs. 25) and they focus solely on enzymes. Within PE36, 474 proteins were assigned to a KEGG category, with some having multiple designations. 465 proteins were assigned a KEGG category in FRIG, resulting in a difference of less than 2% between the two organisms.

There are three KEGG categories that show a significant increase in proteins for PE36. These are amino acid metabolism (+26), metabolism of co-factors and vitamins (+20) and xenobiotics degradation and metabolism (+21). Graph 3.4 depicts the distribution of KEGG categories in PE36 and Graph 3.5 compares the number of proteins in each category for PE36 and FRIG. The total counts for each KEGG category are given in Table 3.5.

In the amino acid metabolism category, a number of the enzymes are duplicated in the genome available within different operons. Some of these "duplications" are the result of a gene being split between two contigs and being annotated as two separate genes. However, there are a number of genes that, if correctly annotated, are actually duplicated in PE36's genome. This occurs in FRIG's genome as well, but with fewer genes. A duplicated gene is considered suspect if there are two identically annotated genes that occur on the edge of a contig. Table 3.6 gives a list of genes that have been duplicated in PE36 but have not been duplicated in FRIG or have been duplicated in fewer numbers.

PE36 and FRIG are evenly balanced in the metabolism of vitamins and co-factors except in ubiquinone biosynthesis. In this category, PE36 has fifteen proteins and FRIG has only two. PE36 also exhibits gene duplication in this category. There are four genes that are duplicated in PE36 and not in FRIG: phosphoserine aminotransferase; NADH-ubiquinone oxidoreductase;

Riboflavin synthase, alpha subunit; and cobyrinic acid a,c-diamide adenosyltransferase.

The last KEGG category to be looked at in detail is xenobiotics degradation and metabolism. According to the annotation, PE36 contains genes that are in pathways to degrade a wide variety of materials; with 17 genes in total and nine more than FRIG. Some of the compounds listed are 1- and 2-Methylnaphthalene, atrazine, benzoate, caprolactam, alcohol, and nitrobenzene. Although actual degradation of these compounds would have to be experimentally verified, the presence of these degradation pathways would present a very interesting addition to life in the deep sea.

Many of these compounds are toxic and are industrial compounds. Is it possible that pollutants from human activities have reached the deep seas and that piezophilic bacterium, ever in search of nutrients, have adapted to utilize these wastes? This possibility is called into question due to the hydrophobicity of many of the compounds in question. This makes it unlikely that the compounds would make it to the deep sea in high enough concentration to influence PE36's evolutionary pathway. However, the presence of these genes in PE36's genome could reflect PE36's ability to metabolize other organic compounds found in the deep sea such as humic substances.

Plasmid Annotation

The plasmid in PE36's genome is fully closed and is very close to being finished, having an error rate of one base in 12,500 bases. Given the size of the plasmid, we would expect there to only be four incorrectly determined bases.

The auto-annotation performed by IMG identified 59 genes. The annotation is drawn onto the plasmid in Figure 3.2, which also identifies the genes by their annotation. Two of the genes identified were tRNAs. tRNAscan-SE was unable to identify the anticodon of one of the tRNA genes, but structural prediction by the Vienna RNAfold WebServer (Hofacker, 2003) shows that the anticodon may be "AAA" (Figure 3.3). If the tRNA is expressed and its anticodon is "AAA," it would be the only such tRNA in PE36's genome.

Of the 57 protein coding genes, 22 are hypothetical with no similar genes in other genomes. These genes have an average predicted length of 108 amino acids with half of the genes being less than 100 amino acids in length. Twenty-nine genes are designated conserved-hypothetical and they have an average length of 308 amino acids, though this average drops to 251 amino acids when the largest gene, with 1906 amino acids, is removed from the set. In the conserved-hypothetical set, there is a "Hypothetical Protease" (Gene Object ID: 641152986; Locus Tag: PE36_00090) which has a strong Pfam match to a Caudovirus prohead protease and other evidence which points towards it being a phage prohead protease in the HK97 family, of which

Caudovirus belongs to. One of the functionally predicted proteins is annotated as a portal protein from an HK97 family phage, offering evidence that this protein should be annotated as an HK97 family prohead protease.

The remaining six protein-coding genes have a functional prediction, which was determined using homology-based evidence. One of the genes is *parA*, a plasmid stability protein associated with partitioning of the plasmid during cell division. ParA is part of the bacteriophage P1 prophage and normally needs a second protein, ParB, to correctly partition (Davey and Funnell, 1994). ParB, however, has not been found in the genome through homology searching. This may mean that *parA* was incorrectly annotated (it shares 33% identity with *E. coli parA* genes), that *parA* is a remnant of an obsolete partitioning mechanism that has been replaced by a different mechanism in the genome, or the functional ParB has differentiated to the point in PE36 that it is not detectable by homology searching. An experiment deleting *parA* from the plasmid and then studying its propagation through different generations would be interesting to determine the requirement of the *parA* gene for plasmid partitioning in PE36.

The other five functionally predicted genes are also all phage-related. Table 3.7 lists all of these genes along with their COG category and length. There is one phage terminase protein, a tail component, a phage portal protein, a transcriptional regulator and a phage-type restriction exonuclease. The phage portal protein is of the HK97-family of phages.

Unfortunately, few of the genes on PE36's plasmid have functional annotations. It is possible that the plasmid has its origins as a prophage, which it may still function as, based upon the presence of the *parA* gene and the fact that all of the functionally annotated proteins are phage-related. On the other hand, this could simply be an indication that PE36 has been exposed to many phages and has integrated them into its plasmid along with the other proteins that it requires. No significant homology in *Moritella viscosa* is found with the plasmid raising the possibility that it may be related to piezophily, though no evidence for this other than its exclusion in a non-piezophile can be found. An interesting test, other than determining the function of all of the genes on the plasmid, would be to test the plasmid's necessity in PE36's growth and see if it affects PE36's piezophily.

Large Tandem Repeats

PE36 has two large sets of tandem repeats in its genome. The full sequence for the repeated region has not been sequenced due to the large average size of the repeats and their tandem structure. The first repeat lies between contigs 1043 and 1044 (hereafter referred to as the "43-44 repeat") and the second repeat lies between contigs 1044 and 984 (hereafter referred to as the "44-84 repeat"). Figure 3.1 illustrates the placement and orientation of the repeats and associated contigs. Each sequence is actually made up of two different "families" of repeated sequences and Figure 3.4 shows an

alignment of all four repeat families. These repeat families are unique to PE36 both amongst sequenced piezophiles and all microbes via NCBI's Genomic BLAST (Cummings *et al.*, 2002). The two 44-84 repeat families are more similar to each other than are the two 43-44 repeat families. This may indicate that the 44-84 repeat is a younger repeat, whose tandem repeats have had less time to diverge.

The distance between these two repeats is over 1.65 Mbp and all four repeats share about 17% identity with each other. The 43-44 repeat is at least 4,220 bp in length and the 44-84 repeat is at least 7,059 bp in length, but since they have not been sequenced in full these values are likely to change with a closed genome.

In addition to the two large tandem repeats, there is also a pair of nearly identical (86.4% identical) sequences downstream from the 44-84 repeats. This pair, whose alignments can be seen in Figure 3.5, are separated by 58 bp of sequence. Based on a BLAST query, it appears that these two sequences are part of, or have been inserted into a microcystin-dependent protein or phage collar protein. The actual repeated sequences do not significantly match any other sequence, indicating that these sequences were inserted into and are not a part of the gene(s); which would need to be substantiated through experimental evidence. Such an experiment could be done to remove the repeated sequence and compare protein product to protein

product with the repeated sequence intact. Further downstream is a gene annotated to encode the enzyme methionyl-tRNA formyltransferase.

The upstream and downstream regions of the 43-44 repeats are annotated with single genes on either contig that incorporate the tandem repeat into the sequence. The gene on contig 1043 has been annotated as the iron-regulated protein FrpC. FrpC is a secreted repeat in toxin (RTX) protein found in *Neisseria meningitidis*, though its biological activity is not known (Osicka *et al.*, 2004). RTX proteins are found in a number of pathogenic organisms, including members of the genera *Neisseria*, *Vibrio,* and *Escherichia* (Osicka *et al.*, 2004). FrpC has also been found to have the interesting property of engaging in calcium ion-dependent autocatalytic activity and cross-linking itself to other FrpC proteins (Osicka *et al.*, 2004).

On contig 1044, the gene has been annotated as a hypothetical protein, but has been given the COG category of an autotransporter adhesin (COG5295). However, because both of the predicted genes incorporate the tandem repeat into their sequence, there is a strong possibility that once the gap is filled in between the contigs, the two genes will end up being part of one gene.

There are genes that immediately follow the hypothetical protein in an operon-like manner by genes that are involved in toxin secretion. A diagram of the region is shown in Figure 3.6. The toxin secreting genes along with the

annotated *frpC* gene lends strong evidence that PE36 has a dedicated operon to secreting a toxin.

The immediate upstream and downstream regions of the 44-84 repeats are also each annotated with a single gene, a fibronectin type III domain protein (8854 bp) and a probable aggregation factor core protein MAFp3 (3059 bp), respectively. Both of these predicted genes incorporate the tandem repeats on either contig. In addition, both genes have a COG match of an "RTX toxin and related Ca^{2+}-binding proteins" (COG2931). However, when a BLAST search is done using the sequence of these two genes, very few matches are found, and all matches are considerably smaller than the query. Given that through the BLAST search no other genes were found that are subsequences of the annotated genes and because both incorporate the tandem repeat into their sequence, it is likely that the two genes are really one large RTX protein.

PE36 potentially has two large RTX toxins encoded in its genome. *Moritella viscosa* is a known fish pathogen and contains an RTX toxin closely related to that of *Vibrio vulnificus* and *Vibrio cholerae* O1, but not to either of PE36's potential toxins. *M. viscosa* also only appears to have one RTX toxin, though it remains unclear due to the draft state of its genome. It is possible that PE36, in the deep sea, has adapted to be pathogenic to multiple types of fish, as opposed to *M. viscosa*, which typically infects only salmonid-type fish. This could be particularly advantageous to PE36 as it may encounter fish less

frequently in the deep sea than *M. viscosa* does in shallower waters and consequently needs to be able to infect whatever type of fish crosses its path.

Interesting Genes

This section focuses primarily on an assortment of genes and gene products that speak to PE36's environment and evolutionary past. PE36 has twice the number of fatty acid-related genes as does FRIG. PE36 appears to have a duplication of the *fab* genes used for fatty acid biosynthesis. Uncertainty exists because most of the fatty acid biosynthesis proteins lack specific annotation. A BLAST search using the *E. coli* sequence of FabF (RefSeq: NP_415613.1) returns one very good match and one weak match in PE36. In addition to the *fab* genes, PE36 also has a full complement the *pfa* genes used for polyunsaturated fat production. PE36 is known to produce DHA (DeLong and Yayanos, 1986). Similar to a study done on *Moritella marina,* which also produces DHA, *pfaB* is located in the same operon as *pfaA, pfaC,* and *pfaD* (Allen and Bartlett, 2002). PE36's PfaB also possesses a ß-ketoacyl-ACP synthase domain, which is one indication of DHA production (Allen and Bartlett, 2002).

Compared to FRIG, PE36 has few annotated integrases and transposases. PE36 only has 20 of these genes whereas FRIG has 77. It is possible when the genome is completed and reannotated, that this number could go up significantly. These proteins may be duplicated around the

genome and as a result, in the initial draft copy of the genome, may not be all accounted for due to assembly. Based on this initial set, it appears that there are at least five transposase families present: IS150, ISSod8, IS91, IS3/IS911 and Tn7.

The number of proteins for RNA modification is approximately equal in PE36 and FRIG. Proteins for pseudouridine synthase are present for both the large and small subunits of the ribosome and tRNAs. Also present is a pseudouridylate synthase protein in each genome.

FRIG has seventeen phage-related genes, but only six of them are non-integrase genes. FRIG also has four phage shock proteins and one phage repressor. PE36 on the other hand, has 41 phage-related genes, with six of them being integrase genes and two phage shock proteins. Although not all of the proteins specify a source of the phage protein, based upon the ones that do, PE36 has genes for at least four phages: bacteriophage f237, prophage ps3, prophage PSPPH06 and an HK97-family phage. PSPPH06 actually has two complete sets of genes in PE36's genome, each with a P2 family major capsid protein and a TP901 family tail tape measure protein.

Lastly, PE36 has a number of "survival" genes that may give insight into its living environment. Interestingly, none of these genes are encoded on PE36's plasmid, which is typically where resistance and related genes are found. To fend off other organisms, bacterial or otherwise, PE36 carries an armory of at least two antibiotics and three toxins. The three toxins include

the two RTX toxins discussed earlier, as well as a third toxin that is an insecticidal toxin. The two antibiotics are colicin V and bacitracin, both amino acid-based antibiotics (Cascales *et al.*, 2007; Katz and Demain, 1977). Colicin V is generally active against *Enterobacteriaceae* and related strains (Cascales *et al.*, 2007), whereas bacitracin is a broad-spectrum antibiotic against gram-positive bacteria (Johnson *et al.*, 1945).

PE36 has a large repertoire of resistance genes including those protecting against antibiotics, heavy metals, and organic compounds. Table 3.8 contains a list of all of the resistance genes annotated in PE36. Compared to FRIG, PE36 has twice as many multidrug resistance proteins and also has more chemical resistance genes (4 to 1). Some of the specific resistances in PE36 include those against bicyclomycin, camphor, acriflavin, fusaric acid, arsenic, copper and tellurite. FRIG has more acriflavin and heavy metal resistance genes and has nine genes for glyoxalase/bleomycin resistance. Both copper and tellurium resistance has been found in the gram-positive piezophile *Carnobacterium* sp. AT7 (Taylor, 2008). FRIG does not have any annotated toxin genes, though it does have one anti-toxin gene.

Conclusion

The sequencing of *Moritella* sp. PE36 and its eventual closure will provide an exciting look into piezophilic lifestyles. Initial analysis of the draft genome suggests that PE36, like its close relative *Moritella viscosa*, may be a

fish pathogen. PE36 also has one of the larger genomes of the sequenced piezophiles at greater than 5.2Mbp. This large genome seems to contain a large number of duplicated genes, the originals remaining in their native operons with the copies being dispersed amongst different proteins. It would be exciting to determine if these duplicates provide novel functionality within PE36.

With 45% more genes involved in cell motility than FRIG, PE36 adds to the hypothesis that adaptation to life in the deep sea requires adjustments in how the cell moves around (Taylor, 2008). This is especially evident with PE36 since it has three motility systems: a polar flagella system, a lateral flagella system, and a twitching motility system. Existing in low nutrient conditions in the deep sea, PE36 seems to have adapted to this by having more proteins involved with amino acid, vitamin/cofactor and xenobiotic metabolism.

Although unknown as to their role in piezophily, if any, PE36 has a number of genes that are predicted to have a signal peptide sequence. Along with the increase in signal peptide-containing proteins, PE36 has an increased inventory of secretion and trafficking proteins.

As a potential fish pathogen, PE36 makes for an interesting case study not only for learning more about piezophily, but also for learning more about the deep sea ecosystem. When PE36's genome is completed, it will add to our accumulation of knowledge and allow more comparative techniques to be

used that will help elucidate the specific adaptations necessary for life in the deep sea.

Table 3.1: Summary of the genome of *Moritella* sp. PE36.

	Chromosome	Plasmid	Total
Size (bp)	5,236,340	49,993	5,287,824
GC Content	41.03%	40.1%	41.02%
Protein Coding Genes	4,726	57	4,781
Genes with Function	2,929	6	2,930
Conserved Hypothetical	1,568	29	1,597
Hypothetical	229	22	254
Genes with Signal Peptides	1,366	8	1,376
Genes with TM Regions	1,214	4	1,218
rRNA Genes	33	0	33
23S rRNA	10	0	10
16S rRNA	9	0	9
5S rRNA	14	0	14
tRNA Genes	127	2	129
Average Gene Size (bp)	951.5	805.8	
Gene Density (genes/kb)	0.907	1.140	
Percent Coding	86.4%	91.9%	
Percent Non-Coding	13.6%	8.1%	

Table 3.2: List of ORFs identified as having a frameshift or being truncated.

Gene OID	Gene OID	Gene OID	Gene OID	Gene OID
641149695	641149695	641152215	641152802	641153320
641149773	641149773	641152230	641152803	641153406
641149774	641149774	641152271	641152804	641153407
641149775	641149775	641152301	641152805	641153445
641149777	641149777	641152302	641152806	641153478
641149811	641149811	641152303	641152807	641153479
641149850	641149850	641152304	641152808	641153516
641149852	641149852	641152375	641152809	641153619
641149854	641149854	641152376	641152810	641153620
641149856	641149856	641152414	641152811	641153621
641149858	641149858	641152416	641152812	641153622
641149870	641149870	641152417	641152813	641153686
641149871	641149871	641152418	641152814	641153687
641149872	641149872	641152420	641152815	641153722
641149873	641149873	641152422	641152837	641153751
641149874	641149874	641152423	641152889	641153752
641149979	641149979	641152424	641152890	641153778
641150045	641150045	641152498	641152911	641153779
641150046	641150046	641152500	641152938	641153810
641150126	641150126	641152555	641152976	641153841
641150301	641150301	641152575	641152977	641153866
641150302	641150302	641152576	641153001	641153890
641150486	641150486	641152577	641153036	641153931
641150506	641150506	641152578	641153041	641153932
641150507	641150507	641152580	641153042	641153972
641150508	641150508	641152581	641153084	641153988
641150509	641150509	641152582	641153085	641154006
641150510	641150510	641152616	641153169	641154046
641150511	641150511	641152617	641153214	641154047
641150512	641150512	641152618	641153252	641154068
641150667	641150667	641152661	641153253	641154091
641150675	641150675	641152662	641153294	641154092
641150676	641150676	641152664	641153313	641154114
641150677	641150677	641152666	641153314	641154167
641150678	641150678	641152707	641153315	641154188
641150679	641150679	641152708	641153316	641154189
641150680	641150680	641152762	641153317	641154205
641150681	641150681	641152763	641153318	641154222
641150686	641150686	641152801	641153319	641154223

Table 3.2: List of ORFs identified as having a frameshift or being truncated, Continued.

Gene OID	Gene OID	Gene OID	Gene OID	Gene OID
641154224	641154350	641154475	641154542	641154614
641154241	641154364	641154491	641154543	641154615
641154242	641154365	641154515	641154549	641154616
641154259	641154401	641154517	641154553	641154617
641154260	641154402	641154521	641154556	641154619
641154277	641154432	641154527	641154557	641154621
641154278	641154433	641154529	641154561	641154622
641154291	641154434	641154530	641154564	641154623
641154292	641154435	641154531	641154569	641154626
641154293	641154445	641154533	641154586	641154627
641154294	641154446	641154534	641154598	641154628
641154311	641154455	641154537	641154600	641154629
641154321	641154456	641154540	641154604	641154614
641154337	641154467	641154541	641154605	641154615

Table 3.3: Summary of Codon Usage in *Moritella* sp. PE36.

Amino Acid	Codon	Number of Times Used	% of Total Usage
Alanine		**127,142**	**8.7%**
	GCA	39,062	2.7%
	GCC	23,761	1.6%
	GCG	31,134	2.1%
	GCT	33,185	2.3%
Arginine		**60,084**	**4.1%**
	AGA	6,039	0.4%
	AGG	1,989	0.1%
	CGA	6,828	0.5%
	CGC	13,069	0.9%
	CGG	3,264	0.2%
	CGT	28,895	2.0%
Asparagine		**69,806**	**4.8%**
	AAC	25,628	1.7%
	AAT	44,178	3.0%
Aspartate		**82,470**	**5.6%**
	GAC	21,818	1.5%
	GAT	60,652	4.1%
Cysteine		**15,635**	**1.1%**
	TGC	4,942	0.3%
	TGT	10,693	0.7%
Glutamate		**82,088**	**5.6%**
	GAA	60,166	4.1%
	GAG	21,922	1.5%
Glutamine		**66,780**	**4.5%**
	CAA	45,895	3.1%
	CAG	20,885	1.4%
Glycine		**95,260**	**6.5%**
	GGA	8,908	0.6%
	GGC	26,239	1.8%
	GGG	9,183	0.6%
	GGT	50,930	3.5%
Histidine		**31,618**	**2.2%**
	CAC	10,863	0.7%
	CAT	20,755	1.4%
Isoleucine		**103,921**	**7.1%**
	ATA	16,346	1.1%
	ATC	33,272	2.3%
	ATT	54,303	3.7%

Table 3.3: Summary of Codon Usage in *Moritella* sp. PE36, Continued.

Amino Acid	Codon	Number of Times Used	% of Total Usage
Leucine		**154,618**	**10.5%**
	CTA	17,080	1.2%
	CTC	10,145	0.7%
	CTG	19,806	1.3%
	CTT	20,583	1.4%
	TTA	64,535	4.4%
	TTG	22,469	1.5%
Lysine		**78,735**	**5.4%**
	AAA	59,418	4.0%
	AAG	19,317	1.3%
Methionine		**37,731**	**2.6%**
	ATG	37,731	2.6%
Phenylalanine		**61,580**	**4.2%**
	TTC	18,532	1.3%
	TTT	43,048	2.9%
Proline		**54,087**	**3.7%**
	CCA	19,695	1.3%
	CCC	5,937	0.4%
	CCG	12,347	0.8%
	CCT	16,108	1.1%
Serine		**99,781**	**6.8%**
	AGC	16,625	1.1%
	AGT	25,962	1.8%
	TCA	22,484	1.5%
	TCC	4,921	0.3%
	TCG	11,900	0.8%
	TCT	17,889	1.2%
Threonine		**84,716**	**5.8%**
	ACA	22,521	1.5%
	ACC	24,117	1.6%
	ACG	16,799	1.1%
	ACT	21,279	1.4%
Tryptophan		**16,783**	**1.1%**
	TGG	16,783	1.1%
Tyrosine		**45,972**	**3.1%**
	TAC	15,037	1.0%
	TAT	30,935	2.1%
Valine		**99,760**	**6.8%**
	GTA	25,686	1.7%
	GTC	14,930	1.0%
	GTG	25,690	1.7%
	GTT	33,454	2.3%

Table 3.3: Summary of Codon Usage in *Moritella* sp. PE36, Continued.

Amino Acid	Codon	Number of Times Used	% of Total Usage
Methionine (Start Codon)		**4636**	**100%**
	ATG	4195	90.5%
	GTG	297	6.1%
	TTG	144	3.4%

Table 3.4: List of COG categories and the occurrence of each in *Moritella* sp. PE36 and *Shewanella frigidimarina* NCIMB 400.

COG Category	PE36	FRIG
[A] RNA processing and modification	1	1
[B] Chromatin structure and dynamics	0	1
[C] Energy production and conversion	255	261
[D] Cell cycle control, cell division, chromosome partitioning	39	34
[E] Amino acid transport and metabolism	336	248
[F] Nucleotide transport and metabolism	89	75
[G] Carbohydrate transport and metabolism	163	122
[H] Coenzyme transport and metabolism	180	159
[I] Lipid transport and metabolism	123	125
[J] Translation, ribosomal structure and biogenesis	187	201
[K] Transcription	316	252
[L] Replication, recombination and repair	150	202
[M] Cell wall/membrane/envelope biogenesis	236	204
[N] Cell motility	162	112
[O] Posttranslational modification, protein turnover, chaperones	173	162
[P] Inorganic ion transport and metabolism	201	195
[Q] Secondary metabolites biosynthesis, transport and catabolism	67	74
[R] General function prediction only	398	375
[S] Function unknown	334	314
[T] Signal transduction mechanisms	263	269
[U] Intracellular trafficking, secretion, and vesicular transport	144	105
[V] Defense mechanisms	54	59
[W] Extracellular structures	0	0
[Y] Nuclear structure	0	0
[Z] Cytoskeleton	1	0

Table 3.5: List of KEGG categories and the occurrence of each in *Moritella* sp. PE36 and *Shewanella frigidimarina* NCIMB 400.

KEGG Category	PE36	FRIG
[1] Amino Acid Metabolism	191	165
[2] Biosynthesis of Polyketides and Non-Ribosomal Peptides	15	9
[3] Biosynthesis of Secondary Metabolites	22	22
[4] Carbohydrate Metabolism	135	129
[5] Energy Metabolism	70	72
[6] Glycan Biosynthesis and Metabolism	29	20
[7] Lipid Metabolism	52	41
[8] Metabolism of Co-Factors and Vitamins	92	72
[9] Metabolism of Other Amino Acids	42	36
[10] Nucleotide Metabolism	78	78
[11] Xenobiotics Biodegradation and Metabolism	56	35
[12] Other	73	62

Table 3.6: List of amino acid metabolism genes duplicated in *Moritella* sp. PE36, but not, or less so, in *Shewanella frigidimarina* NCIMB 400.

Protein	Genes in PE36	Genes in FRIG
Enoyl-CoA Hydratase	3	0
Dihydrolipoamide Dehydrogenase	3	1
Acetyl-CoA Acetyltransferase	3	2
Alanine racemase	2	1
Aspartate Carbamoyltransferase	2	1
2-amino-3-ketobutyrate Coenzyme A Ligase	2	1
L-threonine 3-dehydrogenase	2	1
Phosphoserine Aminotransferase	2	1
Dihydrodipicolinate Synthase	2	1

Table 3.7: List of genes functionally annotated on *Moritella* sp. PE36's plasmid.

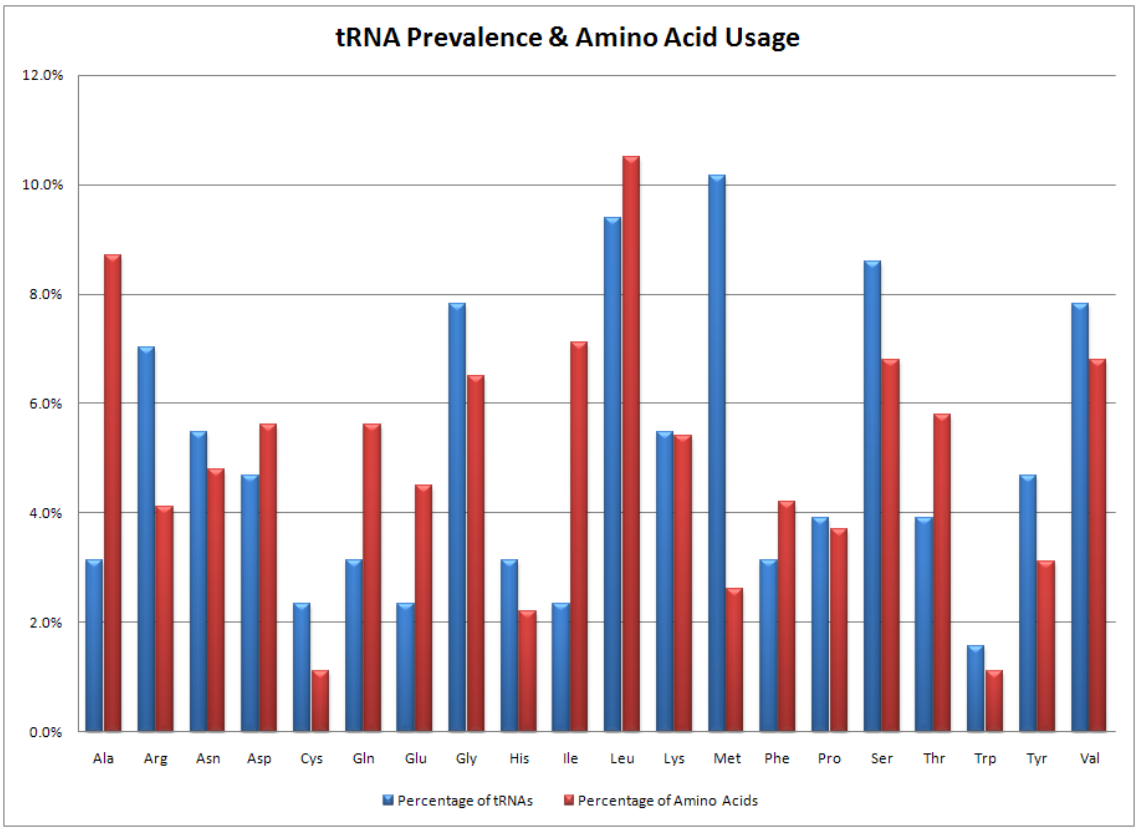
Gene Name	COG Category	Size (aa)
Exonuclease	Replication, Recombination and Repair [L]	212
Phage Portal Protein, HK97 Family	Function Unknown [S]	427
Phage Related Protein	Transcription [K]	555
Phage-Related Protein VCBS, Tail Component	General Function Prediction Only [R]	1635
Plasmid Stability Protein ParA	Cell Cycle Control, Cell Division, Chromosome Partitioning [D]	229
Putative Phage Terminase, Large Subunit	General Function Prediction Only [R]	566

Table 3.8: List of resistance genes found in the genome of *Moritella* sp. PE36.

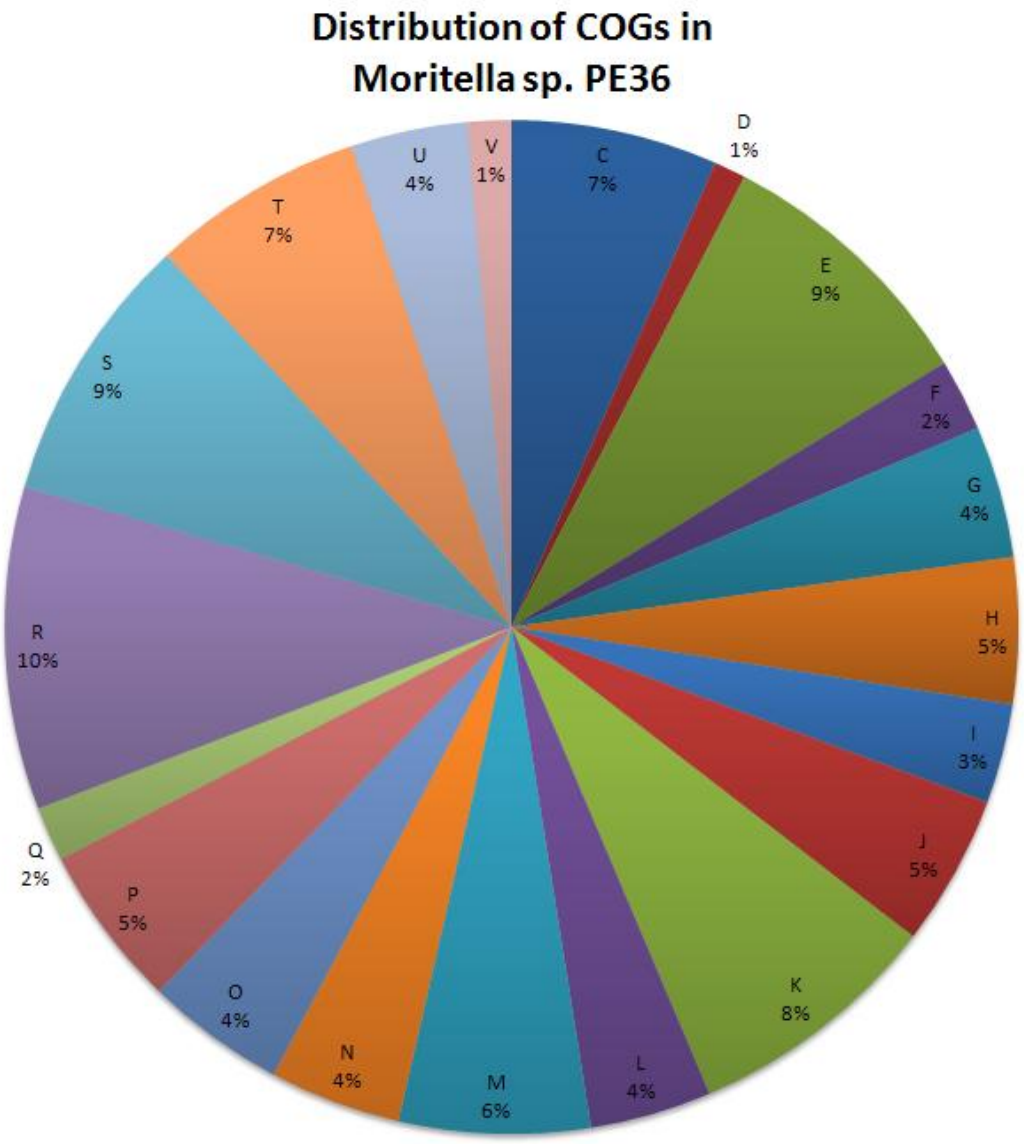
Gene OID	Protein Name
641153919	ABC-type multidrug transport system, ATPase and permease component
641150196	acriflavin resistance plasma membrane protein
641153475	acriflavin resistance protein d
641151737	arsenical resistance operon repressor
641151429	bicyclomycin resistance protein
641153932	bicyclomycin resistance protein
641153132	bicyclomycin resistance protein, putative
641150693	camphor resistance protein CrcB
641150710	copper resistance protein A
641150796	drug resistance transporter, Bcr/CflA family protein
641153165	fusaric acid resistance protein FusE
641152483	Heavy metal response regulator
641154226	heavy metal-transporting ATPase
641150375	MFS family multidrug transport protein, bicyclomycin resistance protein
641151009	multidrug resistance protein D
641151661	multiple antibiotic resistance protein MarC
641152230	Na(+) driven multidrug efflux pump
641149783	organic solvent tolerance protein
641153381	Predicted permease of the drug/metabolite transporter (DMT) superfamily
641154412	putative ABC-type multidrug transport system, ATPase and permease component
641153454	putative AcrB, Cation/multidrug efflux pump
641151738	putative arsenical-resistance protein
641154227	putative heavy metal transcriptional repressor
641149878	Putative Multidrug resistance efflux pump
641150264	putative multidrug resistance protein

Table 3.8: List of resistance genes found in the genome of *Moritella* sp. PE36, Continued.

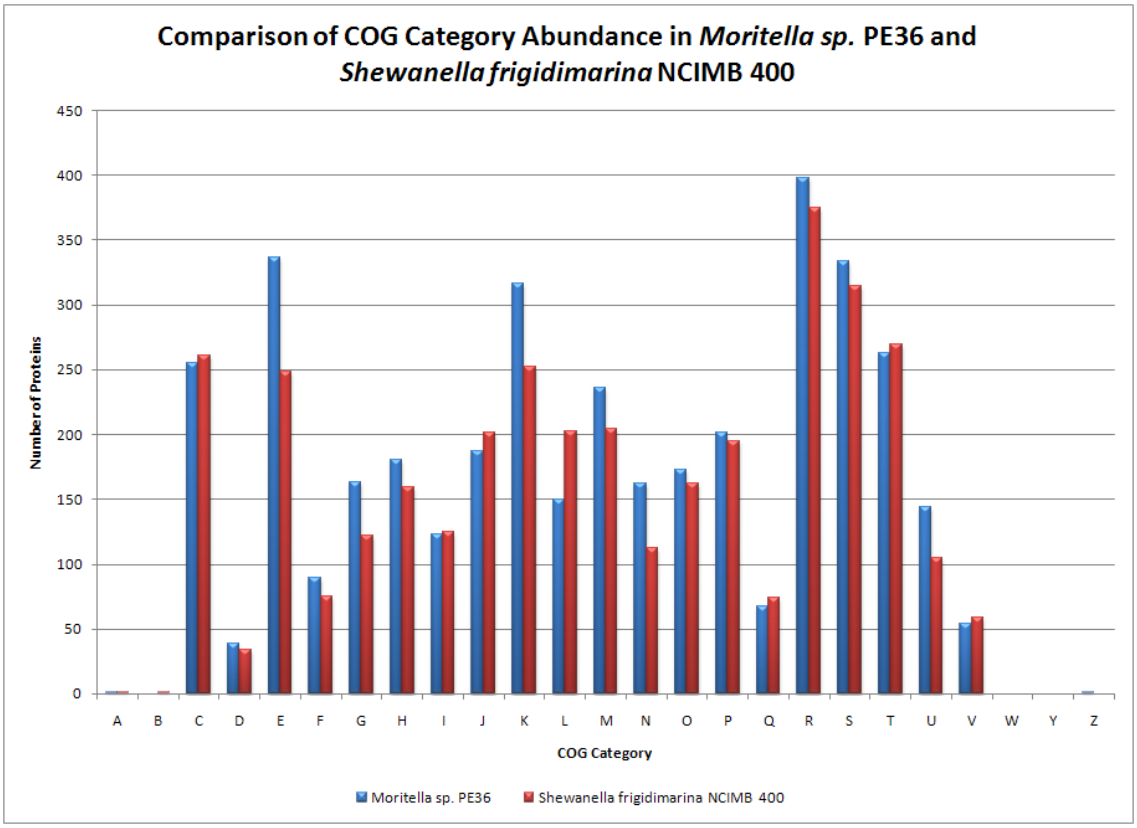
Gene OID	Protein Name
641150525	putative multidrug resistance protein
641153062	putative multidrug resistance protein
641151292	putative multidrug resistance protein D
641154236	putative multidrug resistance protein(AcrB/AcrD/AcrF family)
641151255	Putative Na+-driven multidrug efflux pump
641151798	Putative Na+-driven multidrug efflux pump
641152628	putative tellurite resistance protein
641149929	quaternary ammonium compound-resistance protein
641150399	RND multidrug efflux membrane fusion protein MexE
641150398	RND multidrug efflux transporter MexF
641153210	tellurite resistance protein TehB
641153578	transcriptional regulator, ArsR family
641151332	transporter, drug/metabolite exporter family
641152479	Twin-arginine translocation pathway signal: Copper-resistance protein CopA



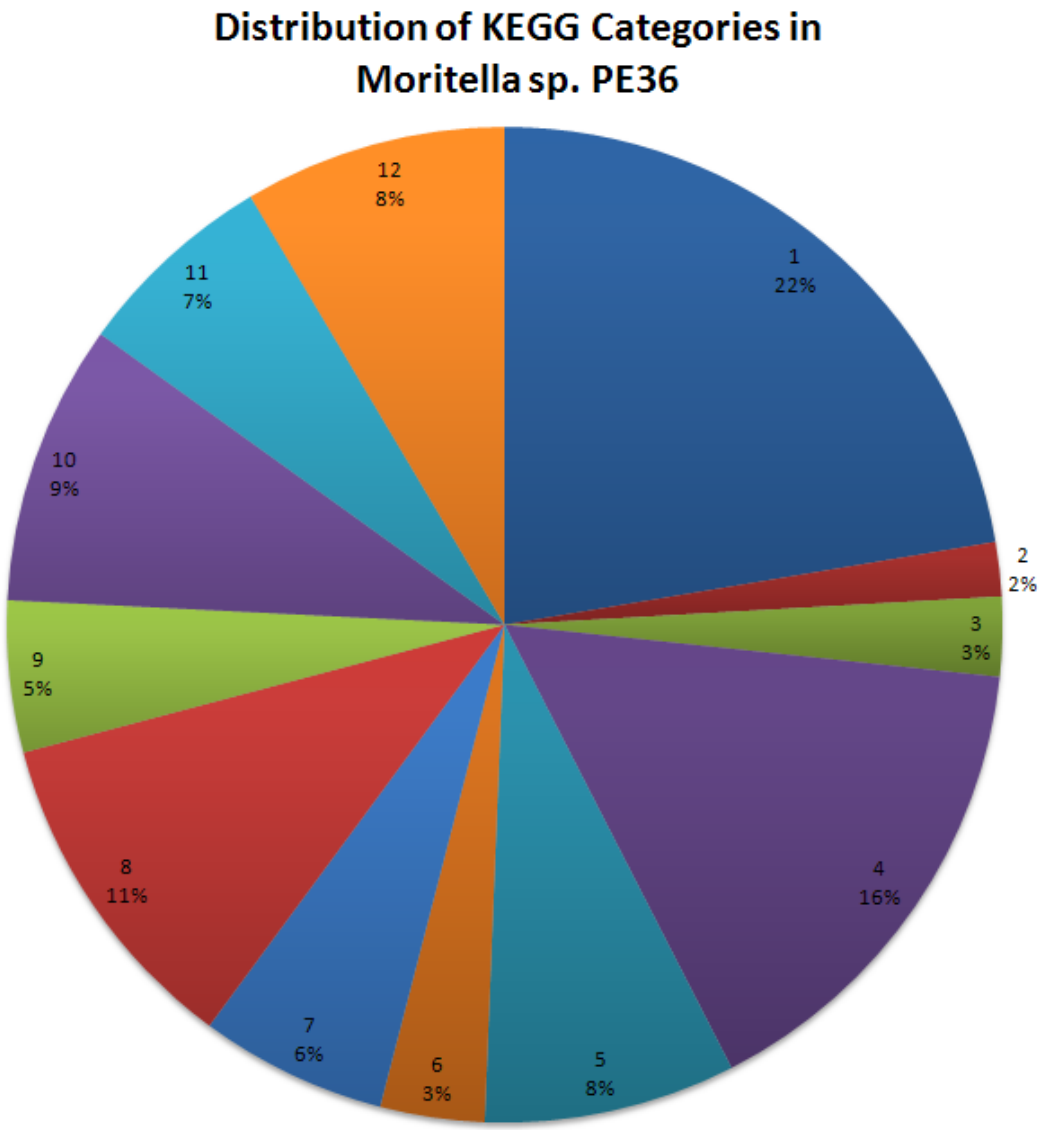
Graph 3.1: Histogram showing the usage of amino acids in the *Moritella* sp. PE36 genome and the prevalence of tRNA genes for those amino acids.



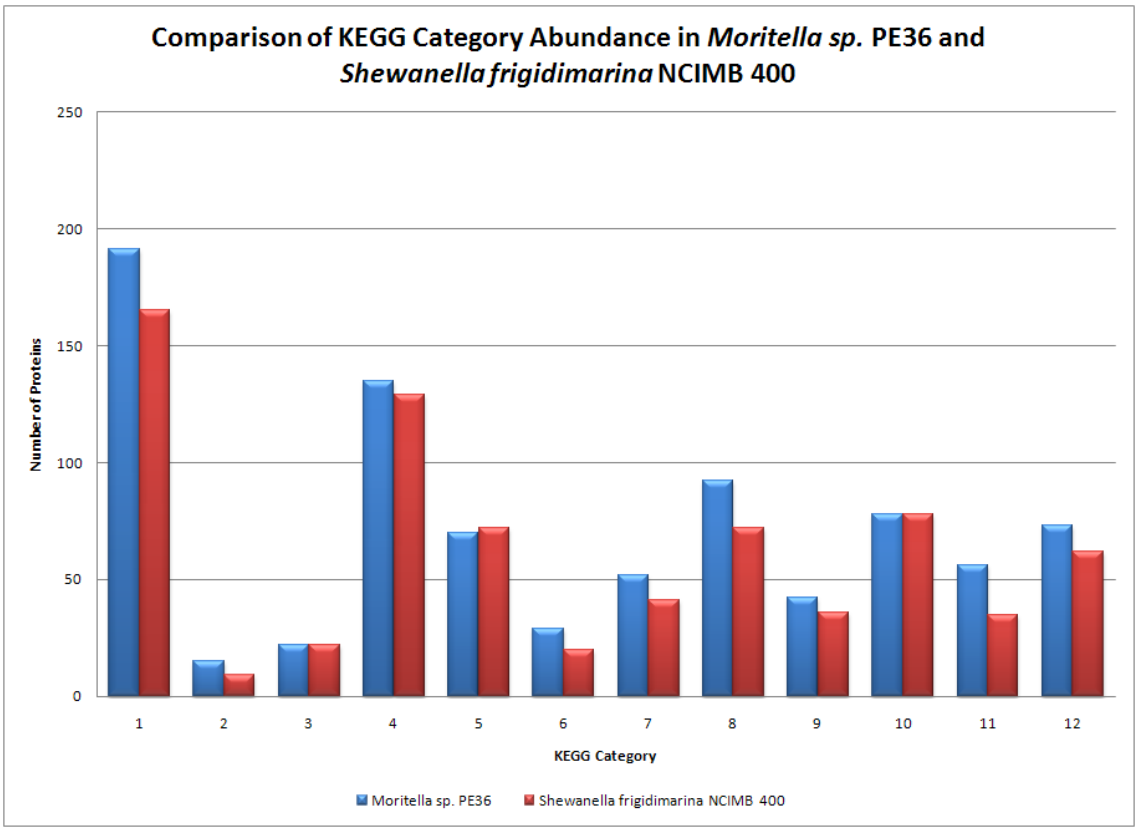
Graph 3.2: Distribution of COGs in *Moritella* sp. PE36 as a percentage of the total number of proteins assigned to at least one COG category. The categories are: Energy Production and Conversion, C; Cell Cycle Control, Cell Division, and Chromosome Partitioning, D; Amino Acid Transport and Metabolism, E; Nucleotide Transport and Metabolism, F; Carbohydrate Transport and Metabolism, G; Coenzyme Transport and Metabolism, H; Lipid Transport and Metabolism, I; Translation, Ribosomal Structure and Biogenesis, J; Transcription, K; Replication, Recombination and Repair, L; Cell Wall/Membrane/Envelope biogenesis, M; Cell Motility, N; Post-Translational Modification, Protein Turnover, Chaperones, O; Inorganic Ion Transport and Metabolism, P; Secondary Metabolites Biosynthesis, Transport and Catabolism, Q; General Function Prediction Only, R; Function Unknown, S; Signal Transduction Mechanisms, T; Intracellular Trafficking, Secretion and Vesicular Transport, U; and Defense Mechanisms, V. Five categories are not shown because they had only one or no proteins in the category.



Graph 3.3: Histogram comparing the COG category abundances in Moritella sp. PE36 and Shewanella frigidimarina NCIMB 400. The categories are: RNA Processing and Modification, A; Chromatin Structure and Dynamics, B; Energy Production and Conversion, C; Cell Cycle Control, Cell Division, and Chromosome Partitioning, D; Amino Acid Transport and Metabolism, E; Nucleotide Transport and Metabolism, F; Carbohydrate Transport and Metabolism, G; Coenzyme Transport and Metabolism, H; Lipid Transport and Metabolism, I; Translation, Ribosomal Structure and Biogenesis, J; Transcription, K; Replication, Recombination and Repair, L; Cell Wall/Membrane/Envelope biogenesis, M; Cell Motility, N; Post-Translational Modification, Protein Turnover, Chaperones, O; Inorganic Ion Transport and Metabolism, P; Secondary Metabolites Biosynthesis, Transport and Catabolism, Q; General Function Prediction Only, R; Function Unknown, S; Signal Transduction Mechanisms, T; Intracellular Trafficking, Secretion and Vesicular Transport, U; Defense Mechanisms, V; Extracellular Structures, W; Nuclear Structure, Y; and Cytoskeleton, Z.



Graph 3.4: Distribution of KEGG categories in *Moritella* sp. PE36 as a percentage of the total number of proteins assigned to at least one KEGG category. The categories are: Amino Acid Metabolism, 1; Biosynthesis of Polyketides and Non-Ribosomal Peptides, 2; Biosynthesis of Secondary Metabolites, 3; Carbohydrate Metabolism, 4; Energy Metabolism, 5; Glycan Biosynthesis and Metabolism, 6; Lipid Metabolism, 7; Metabolism of Co-Factors and Vitamins, 8; Metabolism of Other Amino Acids, 9; Nucleotide Metabolism, 10; Xenobiotics Biodegradation and Metabolism, 11; and Other, 12.



Graph 3.5: Histogram comparing the KEGG category abundances in Moritella sp. PE36 and Shewanella frigidimarina NCIMB 400. The categories are: Amino Acid Metabolism, 1; Biosynthesis of Polyketides and Non-Ribosomal Peptides, 2; Biosynthesis of Secondary Metabolites, 3; Carbohydrate Metabolism, 4; Energy Metabolism, 5; Glycan Biosynthesis and Metabolism, 6; Lipid Metabolism, 7; Metabolism of Co-Factors and Vitamins, 8; Metabolism of Other Amino Acids, 9; Nucleotide Metabolism, 10; Xenobiotics Biodegradation and Metabolism, 11; and Other, 12.

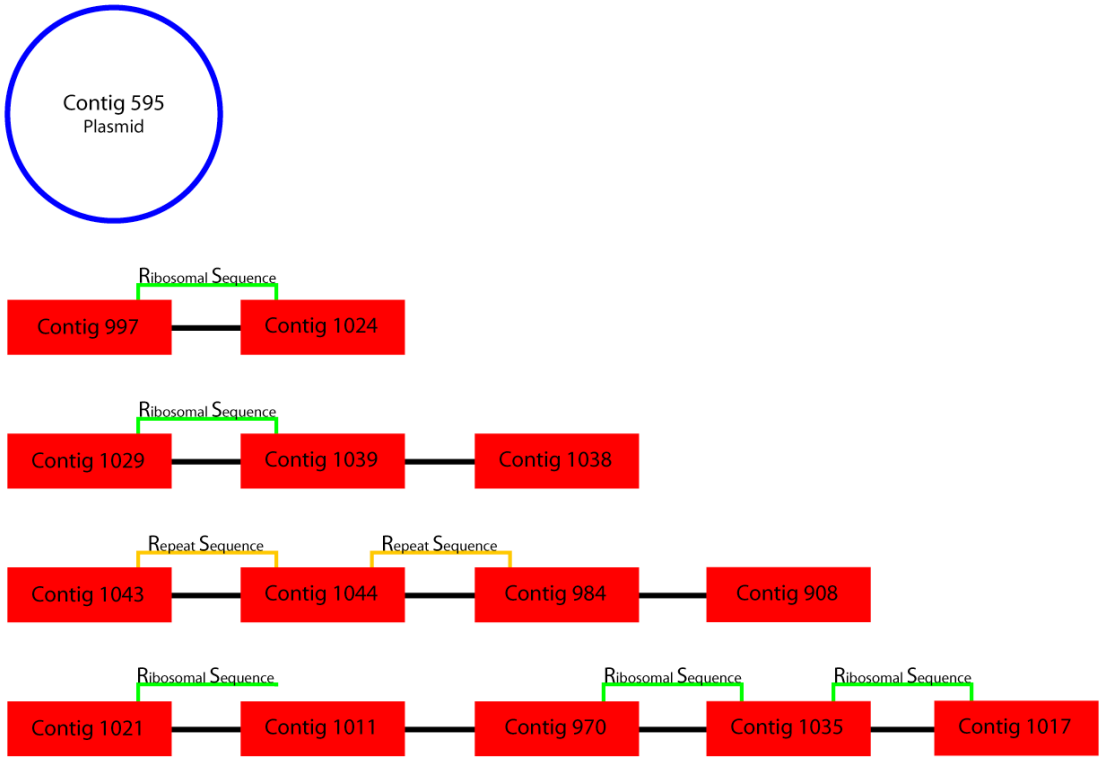


Figure 3.1: Contig and scaffold map of the current assembly of Moritella sp. PE36. There are currently 15 contigs assembled into four scaffolds and one plasmid. Gaps spanned by ribosomal or repeated sequences are marked. Ribosomal sequence that is only present on one end of the gap is indicated by not having the marker touch the contigs where the sequence is not present.

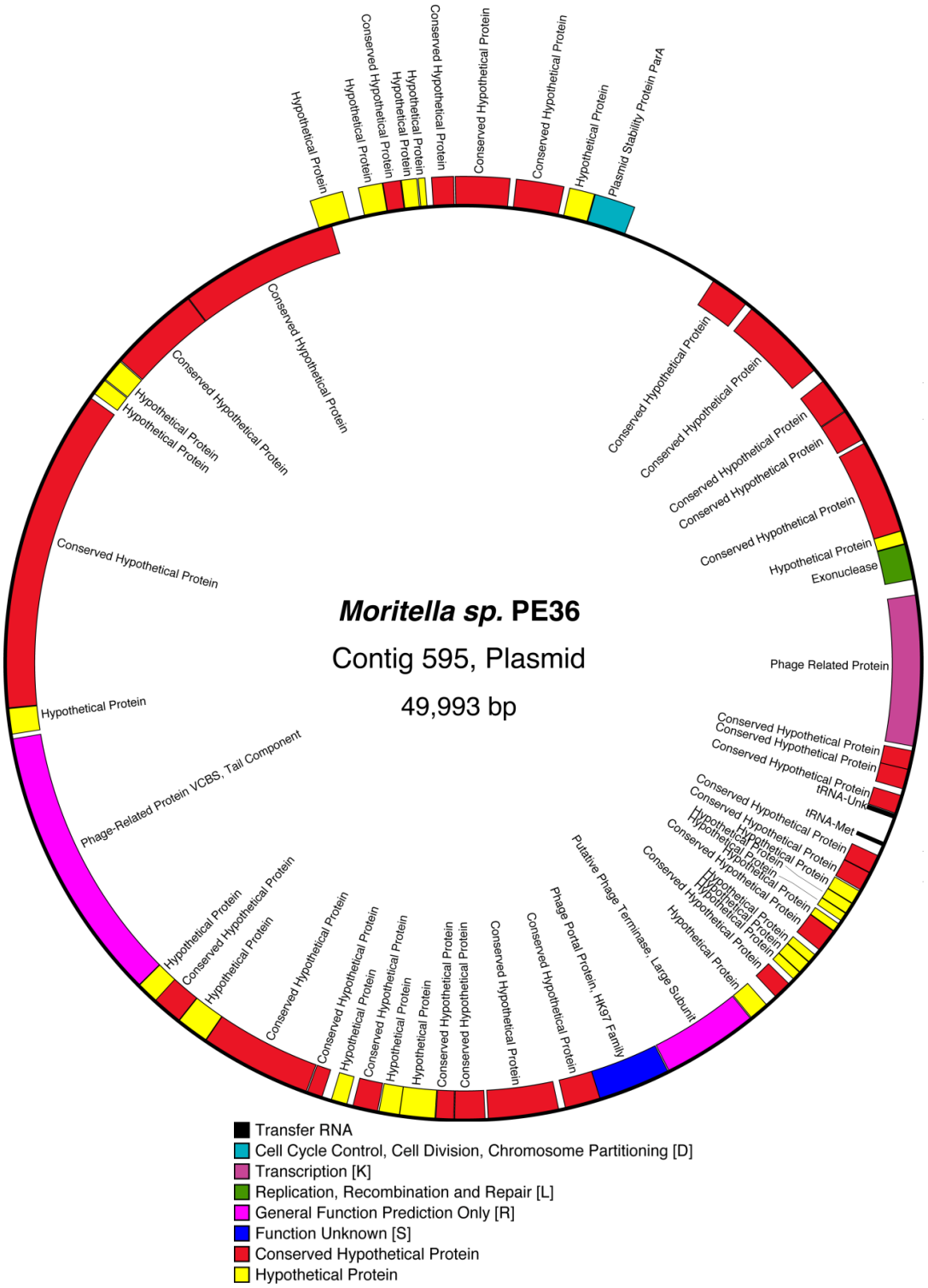


Figure 3.2: Physical map of *Moritella* sp. PE36's plasmid as drawn by OGDRAW. Genes on the inside of the circle are transcribed clockwise and genes on the outside of the circle are transcribed counterclockwise. Gene colors indicate the COG category of that gene.

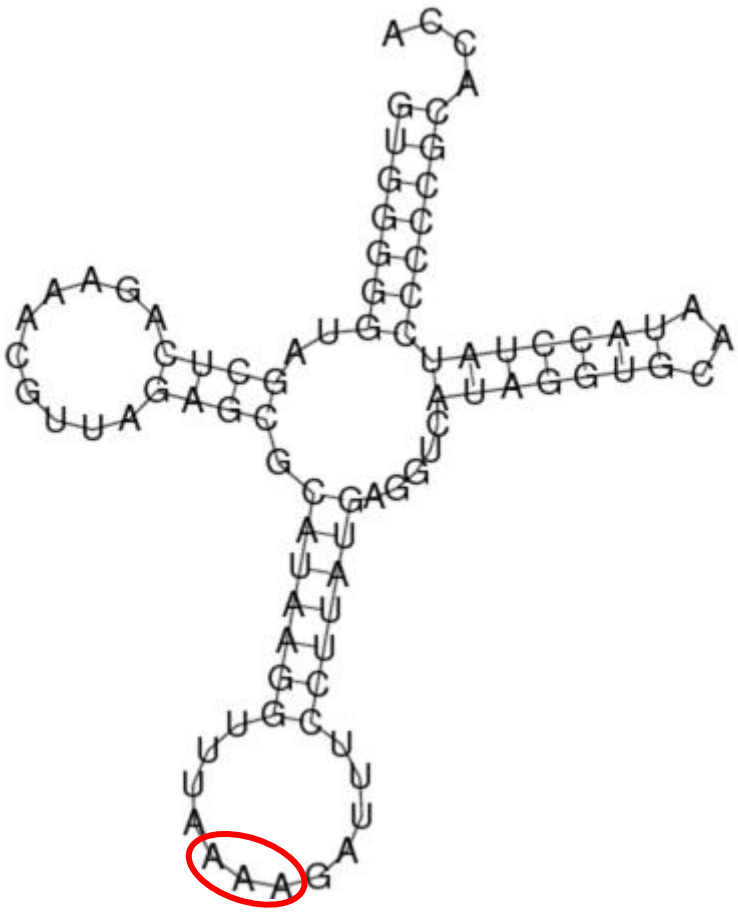


Figure 3.3: Structure of the unknown tRNA on *Moritella* sp. PE36's plasmid (Gene Object ID: 641153002; Locus Tag: PE36_t00332) as predicted by the Vienna RNAfold WebServer. The structure shows a possible anticodon of "AAA," located on the bottom loop, circled in red.

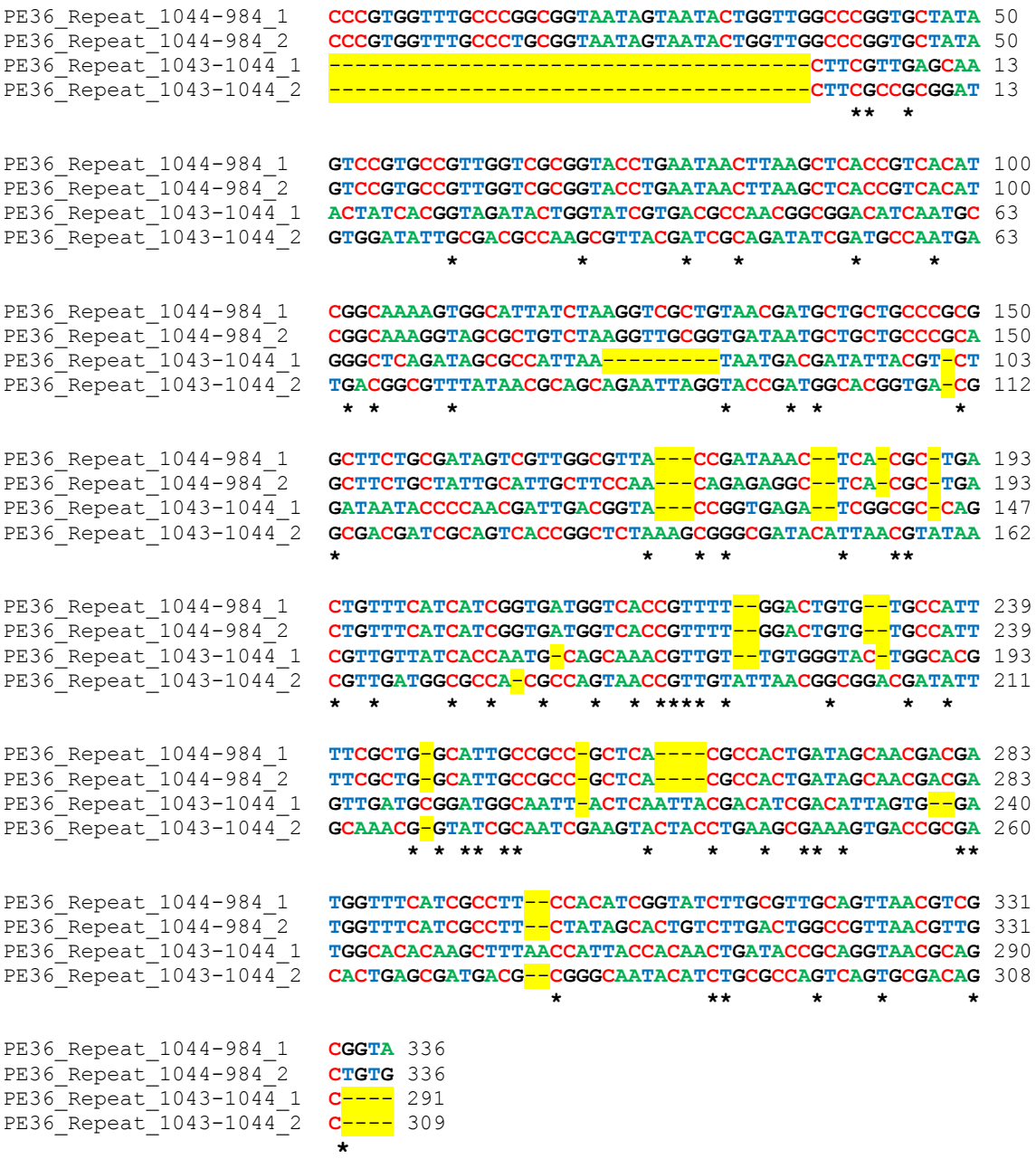


Figure 3.4: Sequence alignment of the repeat families in *Moritella* sp. PE36's two large tandem repeats. PE36_Repeat_1044-984_# represents the two repeat families found in the repeat spanning contigs 1044 and 984. The other two repeat families are found in the repeat spanning contigs 1043 and 1044.

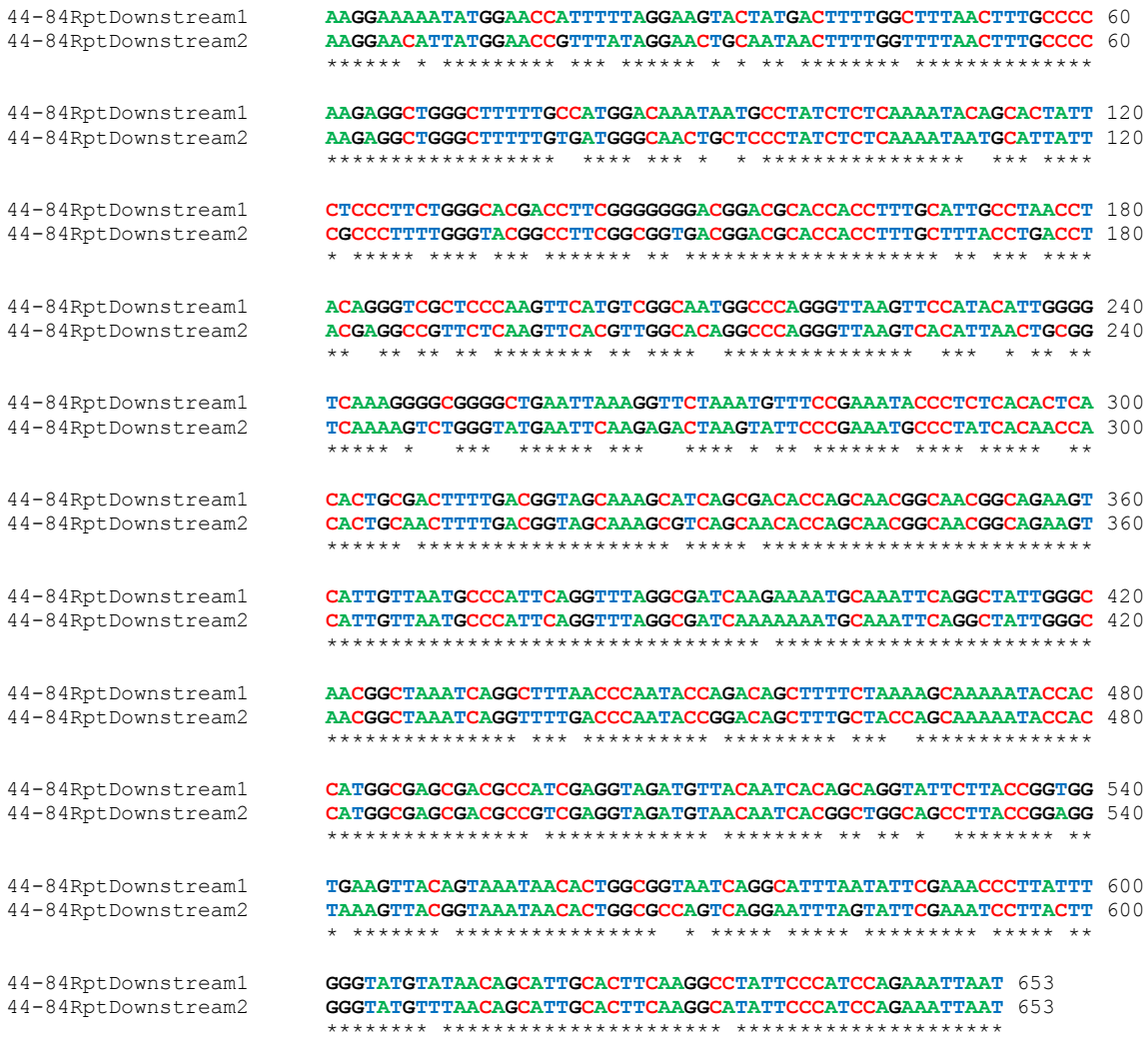
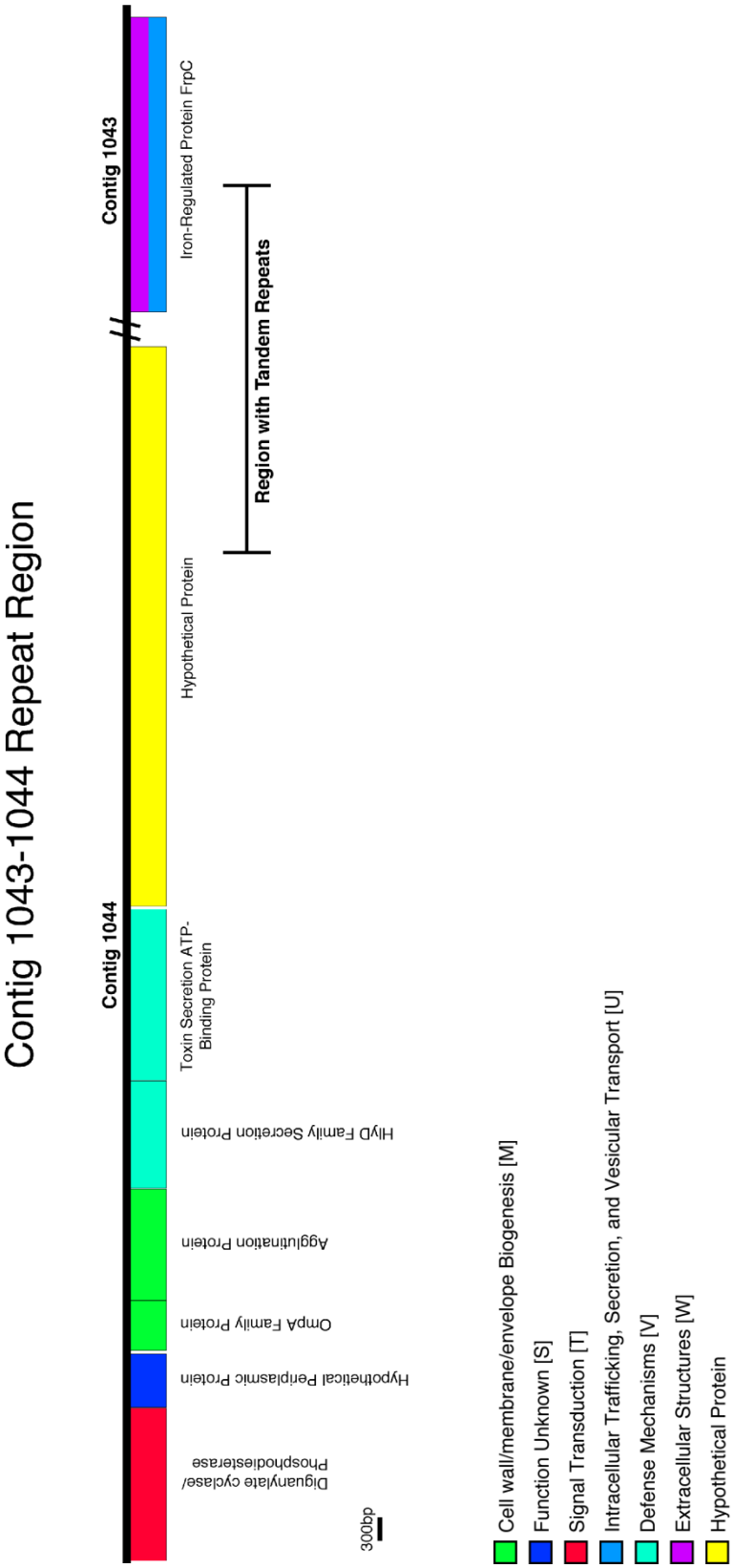


Figure 3.5: Alignment of two highly similar sequences on contig 984 which occur 914 bp upstream of the tandem repeats that span contigs 1044 and 984.

Figure 3.6: Schematic showing the annotated genes in the area of the 43-44 repeats. Sequenced repeats are included in the Iron-Regulated Protein FrpC and in the Hypothetical Protein. The '/ /' represents regions of the genome that have not been sequenced. Translation occurs from left-to-right.


Moritella sp. PE36
Contig 1043-1044 Repeat Region
Contig 1044
Contig 1043
Diguanylate cyclase/ Phosphodiesterase
Hypothetical Periplasmic Protein
OmpA Family Protein
Agglutination Protein
HlyD Family Secretion Protein
Toxin Secretion ATP-Binding Protein
Hypothetical Protein
Iron-Regulated Protein FrpC
Region with Tandem Repeats
300bp
Cell wall/membrane/envelope Biogenesis [M]
Function Unknown [S]
Signal Transduction [T]
Intracellular Trafficking, Secretion, and Vesicular Transport [U]
Defense Mechanisms [V]
Extracellular Structures [W]
Hypothetical Protein

References

Allen, E. E. and Bartlett, D. H. (2000). FabF is required for piezoregulation of cis-vaccenic acid levels and piezophilic growth of the deep-Sea bacterium Photobacterium profundum strain SS9. Journal of bacteriology, 182 (5), 1264-1271.

Allen, E. E. and Bartlett, D. H. (2002). Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium Photobacterium profundum strain SS9. Microbiology (Reading, England), 148 (Pt 6), 1903-1913.

Allen, E. E., Facciotti, D. and Bartlett, D. H. (1999). Monounsaturated but not polyunsaturated fatty acids are required for growth of the deep-sea bacterium Photobacterium profundum SS9 at high pressure and low temperature. *Applied and environmental microbiology*, *65* (4), 1710-1720.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W.*, et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research.*, *25* (17), 3389-3402.

Andrews, L.D., Graham, J., Snider, M.J. and Fraga, D. (2008). Characterization of a novel bacterial arginine kinase from Desulfotalea psychrophila. *Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology*, *150* (3), 312-319.

Anton, A. I., Martinez-Murcia, A. J. and Rodriguez-Valera, F. (1998). Sequence diversity in the 16S-23S intergenic spacer region (ISR) of the rRNA operons in representatives of the Escherichia coli ECOR collection. *Journal of molecular evolution*, *47* (1), 62-72.

Baldi, P., Brunak, S., Frasconi, P., Soda, G. and Pollastri, G. (1999). Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics (Oxford, England)*, *15* (11), 937-946.

Bartlett, D. H. (1992). Microbial life at high pressures. *Science progress*, *76* (301-302 Pt 3-4), 479-496.

Bebbington, K. J. and Williams, H. D. (1993). Investigation of the role of the cydD gene product in production of a functional cytochrome d oxidase in Escherichia coli. *FEMS microbiology letters*, *112* (1), 19-24.

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

Benediktsdottir, E., Verdonck, L., Sproer, C., Helgason, S. and Swings, J. (2000). Characterization of Vibrio viscosus and Vibrio wodanis isolated at different geographical locations: a proposal for reclassification of Vibrio viscosus as Moritella viscosa comb. nov. *International journal of systematic and evolutionary microbiology*, *50 Pt 2*, 479-488.

Bidle, K. A. and Bartlett, D. H. (1999). RecD function is required for high-pressure growth of a deep-sea bacterium. *Journal of bacteriology*, *181* (8), 2330-2337.

Bowman, J. P., McCammon, S. A., Nichols, D. S., Skerratt, J. H., Rea, S. M., Nichols, P. D.*, et al.* (1997). Shewanella gelidimarina sp. nov. and Shewanella frigidimarina sp. nov., novel Antarctic species with the ability to produce eicosapentaenoic acid (20:5 omega 3) and grow anaerobically by dissimilatory Fe(III) reduction. *International journal of systematic bacteriology*, *47* (4), 1040-1047.

Brandts, J. F., Oliveira, R. J. and Westort, C. (1970). Thermodynamics of protein denaturation. Effect of pressu on the denaturation of ribonuclease A. *Biochemistry*, *9* (4), 1038-1047.

Breezee, J., Cady, N. and Staley, J. T. (2004). Subfreezing growth of the sea ice bacterium "Psychromonas ingrahamii". *Microbial ecology*, *47* (3), 300-304.

Brettar, I., Christen, R. and Hofle, M. G. (2002). Shewanella denitrificans sp. nov., a vigorously denitrifying bacterium isolated from the oxic-anoxic interface of the Gotland Deep in the central Baltic Sea. *International journal of systematic and evolutionary microbiology*, *52* (Pt 6), 2211-2217.

Brimacombe, R. (1995). The structure of ribosomal RNA: a three-dimensional jigsaw puzzle. *European journal of biochemistry / FEBS*, *230* (2), 365-383.

Brosius, J., Dull, T. J., Sleeter, D. D. and Noller, H. F. (1981). Gene organization and primary structure of a ribosomal RNA operon from Escherichia coli. *Journal of molecular biology, 148* (2), 107-127.

Campanaro, S., Vezzi, A., Vitulo, N., Lauro, F. M., D'Angelo, M., Simonato, F.*, et al.* (2005). Laterally transferred elements and high pressure adaptation in Photobacterium profundum strains. *BMC genomics*, *6*, 122.

Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y.*, et al.* (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC bioinformatics*, *3*, 2.

Cascales, E., Buchanan, S. K., Duche, D., Kleanthous, C., Lloubes, R., Postle, K.*, et al.* (2007). Colicin biology. *Microbiology and molecular biology reviews : MMBR*, *71* (1), 158-229.

Cheng, J., Randall, A. Z., Sweredoski, M. J. and Baldi, P. (2005). SCRATCH: a protein structure and structural feature prediction server. *Nucleic acids research*, *33* (Web Server issue), W72-6.

Chilukuri, L. N., Fortes, P. A. and Bartlett, D. H. (1997). High pressure modulation of Escherichia coli DNA gyrase activity. *Biochemical and biophysical research communications*, *239* (2), 552-556.

Clamp, M., Cuff, J., Searle, S. M. and Barton, G. J. (2004). The Jalview Java alignment editor. *Bioinformatics (Oxford, England)*, *20* (3), 426-427.

Cohn, E. and Edsall, J. (1943). Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions. *Reinhold, New York*, 370-381.

Cohn, E., McMeekin, T., Edsall, J. and Blanchard, M. (1934). Studies in the physical chemistry of amino acids, peptides and related substances. 1. The apparent molal volume and the electrostriction of the solvent. *Journal of the American Chemical Society*, *56*, 784-794.

Colwell, R. and Morita, R. (1964). Reisolation and emendation of description of Vibrio marinus (Russell) ford. *Journal of bacteriology*, *88*, 831-837.

Condon, C., Squires, C. and Squires, C. L. (1995). Control of rRNA transcription in Escherichia coli. *Microbiological reviews*, *59* (4), 623-645.

Cormack, R. S. and Mackie, G. A. (1991). Mapping ribosomal protein S20-16 S rRNA interactions by mutagenesis. *The Journal of biological chemistry*, *266* (28), 18525-18529.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227* (5258), 561-563.

Cummings, L., Riley, L., Black, L., Souvorov, A., Resenchuk, S., Dondoshansky, I.*, et al.* (2002). Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes. *FEMS microbiology letters.*, *216* (2), 133-138.

Davey, M. J. and Funnell, B. E. (1994). The P1 plasmid partition protein ParA. A role for ATP in site-specific DNA binding. *The Journal of biological chemistry*, *269* (47), 29908-29913.

DeLong, E. F. and Yayanos, A. A. (1985). Adaptation of the membrane lipids of a deep-sea bacterium to changes in hydrostatic pressure. *Science (New York, N.Y.)*, *228* (4703), 1101-1103.

Delong, E. F. and Yayanos, A. A. (1986). Biochemical Function and Ecological Significance of Novel Bacterial Lipids in Deep-Sea Procaryotes. *Applied and environmental microbiology*, *51* (4), 730-737.

Delong, E. F. and Yayanos, A. A. (1987). Properties of the Glucose Transport System in Some Deep-Sea Bacteria. *Applied and environmental microbiology*, *53* (3), 527-532.

Delong, E. F., Franks, D. G. and Yayanos, A. A. (1997). Evolutionary Relationships of Cultivated Psychrophilic and Barophilic Deep-Sea Bacteria. *Applied and environmental microbiology*, *63* (5), 2105-2108.

Dersch, P., Kneip, S. and Bremer, E. (1994). The nucleoid-associated DNA-binding protein H-NS is required for the efficient adaptation of Escherichia coli K-12 to a cold environment. *Mol Gen Genet*, *245* (2), 255-259.

Dixon, M. and Webb, E. C. (1958). *Enzymes.* London: Longmans Green.

Dunn, J. J. and Studier, F. W. (1973). T7 early RNAs and Escherichia coli ribosomal RNAs are cut from large precursor RNAs in vivo by ribonuclease 3.

Proceedings of the National Academy of Sciences of the United States of America, *70* (12), 3296-3300.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research.*, *32* (5), 1792-1797.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*, *8* (3), 186-194.

Ewing, B., Hillier, L., Wendl, M. C. and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*, *8* (3), 175-185.

Finch, P. W., Storey, A., Brown, K., Hickson, I. D. and Emmerson, P. T. (1986). Complete nucleotide sequence of recD, the structural gene for the alpha subunit of Exonuclease V of Escherichia coli. *Nucleic acids research*, *14* (21), 8583-8594.

Firpo, M. A. and Dahlberg, A. E. (1998). The importance of base pairing in the penultimate stem of Escherichia coli 16S rRNA for ribosomal subunit association. *Nucleic acids research*, *26* (9), 2156-2160.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R.*, et al.* (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science (New York, N.Y.)*, *269* (5223), 496-512.

Goldberg, S. M., Johnson, J., Busam, D., Feldblyum, T., Ferriera, S., Friedman, R.*, et al.* (2006). A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *103* (30), 11240-11245.

Gordon, D., Abajian, C. and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res*, *8* (3), 195-202.

Gotz, F., Dabbs, E. R. and Gualerzi, C. O. (1990). Escherichia coli 30S mutants lacking protein S20 are defective in translation initiation. *Biochimica et biophysica acta*, *1050* (1-3), 93-97.

Gross, M. and Jaenicke, R. (1990). Pressure-induced dissociation of tight couple ribosomes. *FEBS letters*, *267* (2), 239-241.

Gross, M. and Jaenicke, R. (1994). Proteins under pressure. The influence of high hydrostatic pressure on structure, function and assembly of proteins and protein complexes. *European journal of biochemistry / FEBS*, *221* (2), 617-630.

Gu, S.-Q., Peske, F., Wieden, H.-J., Rodnina, M. V. and Wintermeyer, W. (2003). The signal recognition particle binds to protein L23 at the peptide exit of the Escherichia coli ribosome. *RNA (New York, N.Y.)*, *9* (5), 566-573.

Harpaz, Y., Gerstein, M. and Chothia, C. (1994). Volume changes on protein folding. *Structure (London, England : 1993)*, *2* (7), 641-649.

Hawley, S. A. (1971). Reversible pressure--temperature denaturation of chymotrypsinogen. *Biochemistry*, *10* (13), 2436-2442.

Hengen, P. N., Bartram, S. L., Stewart, L. E. and Schneider, T. D. (1997). Information analysis of Fis binding sites. *Nucleic Acids Res*, *25* (24), 4994-5002.

Heremans, K. (1982). High pressure effects on proteins and other biomolecules. *Annual review of biophysics and bioengineering*, *11*, 1-21.

Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic acids research*, *31* (13), 3429-3431.

Hosoya, S., Suzuki, S., Adachi, K., Matsuda, S. and Kasai, H. (2008). Paramoritella alkaliphila gen. nov., sp. nov., a member of the family Moritellaceae isolated in the Republic of Palau. *Int J Syst Evol Microbiol (in press)* .

Hury, J., Nagaswamy, U., Larios-Sanz, M. and Fox, G. E. (2006). Ribosome origins: the relative age of 23S rRNA Domains. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life*, *36* (4), 421-429.

Ishii, A., Oshima, T., Sato, T., Nakasone, K., Mori, H. and Kato, C. (2005). Analysis of hydrostatic pressure effects on transcription in Escherichia coli by DNA microarray procedure. *Extremophiles : life under extreme conditions*, *9* (1), 65-73.

Ishii, A., Sato, T., Wachi, M., Nagai, K. and Kato, C. (2004). Effects of high hydrostatic pressure on bacterial cytoskeleton FtsZ polymers in vivo and in vitro. *Microbiology (Reading, England)*, *150* (Pt 6), 1965-1972.

Ivanova, E., Flavier, S. and Christen, R. (2004). Phylogenetic relationships among marine Alteromonas-like proteobacteria: emended description of the family Alteromonadaceae and proposal of Pseudoalteromonadaceae fam. nov., Colwelliaceae fam. nov., Shewanellaceae fam. nov., Moritellaceae fam. nov., Ferri. *International Journal of Systematic and Evolutionary Microbiology*, *54* (5), 1773-1788.

Johnson, B., Anker, H. and Meleney, F. (1945). Bacitracin: A New Antibiotic Produced By A Member Of The B. subtilis Group. *Science (New York, N.Y.)*, *102* (2650), 376-377.

Johnson, F., Eyring, H. and Stover, B. (1974). The theory of rate processes in biology and medicine. *Wiley, New York* .

Jolicoeur, C., Riedl, B., Desrochers, D., Lemelin, L., Zamojska, R. and Enea, O. (1986). Solvation of amino acid residues in water and urea-water mixtures: Volumes and heat capacities of 20 amino acids in water and in 8 molar urea at 25C. *Journal of Solution Chemistry*, *15* (2), 109-128.

Jones, L. J., Carballido-Lopez, R. and Errington, J. (2001). Control of cell shape in bacteria: helical, actin-like filaments in Bacillus subtilis. *Cell*, *104* (6), 913-922.

Kanehisa, M. and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, *28* (1), 27-30.

Kapitonov, D., Bieberich, E. and Yu, R. (1999). Combinatorial PCR approach to homology-based cloning: Cloning and expression of mouse and human GM3-synthase. *Glycoconjugate Journal*, *16* (7), 337.

Kato, C., Tamegai, H., Ikegami, A., Usami, R. and Horikoshi, K. (1996). Open reading frame 3 of the barotolerant bacterium strain DSS12 is

complementary with cydD in Escherichia coli: cydD functions are required for cell stability at high pressure. *Journal of biochemistry*, *120* (2), 301-305.

Katz, E. and Demain, A. L. (1977). The peptide antibiotics of Bacillus: chemistry, biogenesis, and possible functions. *Bacteriological reviews*, *41* (2), 449-474.

Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Advances in protein chemistry*, *14*, 1-63.

Kim, H. J., Park, S., Lee, J. M., Park, S., Jung, W., Kang, J.-S.*, et al.* (2008). Moritella dasanensis sp. nov., a psychrophilic bacterium isolated from the Arctic ocean. *International journal of systematic and evolutionary microbiology*, *58* (Pt 4), 817-820.

Kim, U. J., Shizuya, H., de Jong, P. J., Birren, B. and Simon, M. I. (1992). Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic acids research*, *20* (5), 1083-1085.

Kitchen, D. B., Reed, L. H. and Levy, R. M. (1992). Molecular dynamics simulation of solvated protein at high pressure. *Biochemistry*, *31* (41), 10083-10093.

Kundrot, C. E. and Richards, F. M. (1987). Crystal structure of hen egg-white lysozyme at a hydrostatic pressure of 1000 atmospheres. *Journal of molecular biology*, *193* (1), 157-170.

Lagesen, K., Hallin, P., Rodland, E., Staerfeldt, H., Rognes, T. and Ussery, D. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35* (9), gkm160v1-3108.

Landau, J. V. (1967). Induction, transcription and translation in Escherichia coli: a hydrostatic pressure study. *Biochimica et biophysica acta*, *149* (2), 506-512.

Lauro, F. (2007). Evolutionary and functional genomics of bacteria from the cold deep sea. *PhD Dissertation, University of California, San Diego* .

Lauro, F. (2007). Evolutionary and functional genomics of bacteria from the cold deep sea. *Ph.D. Dissertation, University of California, San Diego* .

Lauro, F. M., Chastain, R. A., Blankenship, L. E., Yayanos, A. A. and Bartlett, D. H. (2007). The unique 16S rRNA genes of piezophiles reflect both phylogeny and adaptation. *Applied and environmental microbiology*, *73* (3), 838-845.

Lauro, F. and Bartlett, D. (2007). Prokaryotic lifestyles in deep sea habitats. *Extremophiles : life under extreme conditions*, *12* (1), 15-25.

Lohse, M., Drechsel, O. and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current genetics*, *52* (5-6), 267-274.

Lowe, T. M. and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research.*, *25* (5), 955-964.

Lunder, T., Sorum, H., Holstad, G., Steigerwalt, A. G., Mowinckel, P. and Brenner, D. J. (2000). Phenotypic and genotypic characterization of Vibrio viscosus sp. nov. and Vibrio wodanis sp. nov. isolated from Atlantic salmon (Salmo salar) with 'winter ulcer'. *International journal of systematic and evolutionary microbiology*, *50 Pt 2*, 427-450.

Markham, N. R. and Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization. *Methods in molecular biology (Clifton, N.J.)*, *453*, 3-31.

Markham, N. and Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization.

Markowitz, V. M., Korzeniewski, F., Palaniappan, K., Szeto, E., Werner, G., Padki, A.*, et al.* (2006). The integrated microbial genomes (IMG) system. *Nucleic acids research.*, *34* (Database issue), D344-8.

Matthews, C. R. (1993). Pathways of protein folding. *Annual review of biochemistry*, *62*, 653-683.

Meganathan, R. and Marquis, R. E. Loss of bacterial motility under pressure. *Nature*, *246* (5434), 525-527.

Methe, B. A., Nelson, K. E., Deming, J. W., Momen, B., Melamud, E., Zhang, X., *et al.* (2005). The psychrophilic lifestyle as revealed by the genome sequence of Colwellia psychrerythraea 34H through genomic and proteomic analyses. *Proceedings of the National Academy of Sciences of the United States of America*, *102* (31), 10913-10918.

Mishra, A. and Ahluwalia, J. (1984). Apparent molal volumes of amino acids, N-acetyl amino acids and peptides in aqueous solutions. *Journal of Physical Chemistry*, *88*, 86-92.

Morita, R. Y. and ZoBell, C. E. (1956). Effect of hydrostatic pressure on the succinic dehydrogenase system in Escherichia coli. *Journal of bacteriology*, *71* (6), 668-672.

Morita, R. and Morita, C. (1955). Occurrence of bacteria in pelagic sediments collected during the mid-Pacific expedition. *Deep Sea Research Part II: Topical Studies in Oceanography*, *3*, 66-73.

Mozhaev, V. V., Heremans, K., Frank, J., Masson, P. and Balny, C. (1996). High pressure effects on protein structure and function. *Proteins*, *24* (1), 81-91.

Mueller, F., Sommer, I., Baranov, P., Matadeen, R., Stoldt, M., Wohnert, J., *et al.* (2000). The 3D arrangement of the 23 S and 5 S rRNA in the Escherichia coli 50 S ribosomal subunit based on a cryo-electron microscopic reconstruction at 7.5 A resolution. *Journal of molecular biology*, *298* (1), 35-59.

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., *et al.* (2000). A whole-genome assembly of Drosophila. *Science (New York, N.Y.)*, *287* (5461), 2196-2204.

Nakayama, A., Saito, R., Matsuzaki, M., Yano, Y. and Yoshida, K. (2005). Phylogenetic analysis based on 16S rRNA gene sequences of deep-sea bacteria isolated from intestinal contents of deep-sea fishes retrieved from the abyssal zone. *The Journal of general and applied microbiology*, *51* (6), 385-394.

Nakayama, A., Yano, Y. and Yoshida, K. (1994). New Method for Isolating Barophiles from Intestinal Contents of Deep-Sea Fishes Retrieved from the Abyssal Zone. *Applied and environmental microbiology*, *60* (11), 4210-4212.

Nicholson, A. W. (1999). Function, mechanism and regulation of bacterial ribonucleases. *FEMS microbiology reviews.*, *23* (3), 371-390.

Niven, G. W., Miles, C. A. and Mackey, B. M. (1999). The effects of hydrostatic pressure on ribosome conformation in Escherichia coli: and in vivo study using differential scanning calorimetry. *Microbiology (Reading, England)*, *145 (Pt 2)*, 419-425.

Nogi, Y. and Kato, C. (1999). Taxonomic studies of extremely barophilic bacteria isolated from the Mariana Trench and description of Moritella yayanosii sp. nov., a new barophilic bacterial isolate. *Extremophiles : life under extreme conditions*, *3* (1), 71-77.

Nogi, Y., Kato, C. and Horikoshi, K. (1998). Moritella japonica sp. nov., a novel barophilic bacterium isolated from a Japan Trench sediment. *The Journal of general and applied microbiology*, *44* (4), 289-295.

Noller, H. F. (1991). Ribosomal RNA and translation. *Annual review of biochemistry*, *60*, 191-227.

Nomura, M., Morgan, E. and Jaskunas, S. (1977). Genetics of Bacterial Ribosomes. *Annual Review of Genetics*, *11* (1), 297-347.

Osicka, R., Prochazkova, K., Sulc, M., Linhartova, I., Havlicek, V. and Sebo, P. (2004). A novel "clip-and-link" activity of repeat in toxin (RTX) proteins from gram-negative pathogens. Covalent protein cross-linking by an Asp-Lys isopeptide bond upon calcium-dependent processing at an Asp-Pro bond. *The Journal of biological chemistry*, *279* (24), 24944-24956.

Osorio, C. R., Collins, M. D., Romalde, J. L. and Toranzo, A. E. (2005). Variation in 16S-23S rRNA intergenic spacer regions in Photobacterium damselae: a mosaic-like structure. *Applied and environmental microbiology*, *71* (2), 636-645.

Parsell, D. A. and Lindquist, S. (1993). The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annual review of genetics*, *27*, 437-496.

Paulsen, I., Banerjei, L., Myers, G., Nelson, K., Seshadri, R., Read, T., *et al.* (2003). Role of Mobile DNA in the Evolution of Vancomycin-Resistant Enterococcus faecalis. *Science*, *299* (5615), 2071-2074.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., *et al.* (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25* (13), 1605-1612.

Rao, M., Atreyi, M. and Rajeswari, M. (1984). Partial molar volumes of alpha-amino acids with ionogenic side chains in water. *Journal of Physical Chemistry*, *88*, 3129-3131.

Rubin, G., Yandell, M., Wortman, J., Gabor, G., Nelson, C., Hariharan, I., *et al.* (2000). Comparative Genomics of the Eukaryotes. *Science*, *287* (5461), 2204-2215.

Russell, H. (3792). Untersuchungen uber im Golf von Neapel lebende Bacterien. *Medical Microbiology and Immunology*, *11* (1), 165-206.

Saito, R. and Nakayama, A. (2004). Differences in malate dehydrogenases from the obligately piezophilic deep-sea bacterium Moritella sp. strain 2D2 and the psychrophilic bacterium Moritella sp. strain 5710. *FEMS microbiology letters*, *233* (1), 165-172.

Saito, R., Kato, C. and Nakayama, A. (2006). Amino acid substitutions in malate dehydrogenases of piezophilic bacteria isolated from intestinal contents of deep-sea fishes retrieved from the abyssal zone. *The Journal of general and applied microbiology*, *52* (1), 9-19.

Schatz, M. C., Phillippy, A. M., Shneiderman, B. and Salzberg, S. L. (2007). Hawkeye: an interactive visual analytics tool for genome assemblies. *Genome Biology*, *8* (3), R34.

Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., *et al.* (2005). Structures of the bacterial ribosome at 3.5 A resolution. *Science.*, *310* (5749), 827-834.

Sengupta, J., Agrawal, R. K. and Frank, J. (2001). Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA.

Proceedings of the National Academy of Sciences of the United States of America, *98* (21), 11991-11996.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y.*, et al.* (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proceedings of the National Academy of Sciences of the United States of America*, *89* (18), 8794-8797.

Siebenaller, J. F. (1984). Structural comparison of lactate dehydrogenase homologs differing in sensitivity to hydrostatic pressure. *Biochimica et biophysica acta*, *786* (3), 161-169.

Silva, J. L. and Weber, G. (1993). Pressure stability of proteins. *Annual review of physical chemistry*, *44*, 89-113.

Simonato, F., Campanaro, S., Lauro, F. M., Vezzi, A., D'Angelo, M., Vitulo, N.*, et al.* (2006). Piezophilic adaptation: a genomic point of view. *Journal of biotechnology*, *126* (1), 11-25.

Somero, G. (1990). Life at Low Volume Change: Hydrostatic Pressure as a Selective Factor in the Aquatic Environment. *American Zoologist*, *30* (1).

Srivastava, A. K. and Schlessinger, D. (1990). Mechanism and regulation of bacterial ribosomal RNA processing. *Annual review of microbiology*, *44*, 105-129.

Steitz, T. A. and Moore, P. B. (2003). RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends in biochemical sciences*, *28* (8), 411-418.

Stern, S., Powers, T., Changchien, L. M. and Noller, H. F. (1989). RNA-protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. *Science (New York, N.Y.)*, *244* (4906), 783-790.

Steven, S. (1990). Molecular systematics of Vibrio and Photobacterium. *Ph.D. Dissertation. University of Maryland, College Park* .

Stratton, T. K. (2008). Genomic Analysis of High Pressure Adaptation in Deep Sea Bacteria. *Masters Thesis, University of California, San Diego* .

Tatusov, R. L., Koonin, E. V. and Lipman, D. J. (1997). A genomic perspective on protein families. *Science.*, *278* (5338), 631-637.

Thieringer, H. A., Jones, P. G. and Inouye, M. (1998). Cold shock and adaptation. *BioEssays : news and reviews in molecular, cellular and developmental biology*, *20* (1), 49-57.

Tindall, K. and Kunkel, T. (1988). Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. *Biochemistry*, *27* (16), 6008.

Tipton, K. and Boyce, S. (2000). History of the enzyme nomenclature system. *Bioinformatics (Oxford, England)*, *16* (1), 34-40.

Urakawa, H., Kita-Tsukamoto, K., Steven, S. E., Ohwada, K. and Colwell, R. R. (1998). A proposal to transfer Vibrio marinus (Russell 1891) to a new genus Moritella gen. nov. as Moritella marina comb. nov. *FEMS microbiology letters*, *165* (2), 373-378.

Vezzi, A., Campanaro, S., D'Angelo, M., Simonato, F., Vitulo, N., Lauro, F. M.*, et al.* (2005). Life at depth: Photobacterium profundum genome sequence and expression analysis. *Science*, *307* (5714), 1459-1461.

Voss, H., Wiemann, S., Grothues, D., Sensen, C., Zimmermann, J., Schwager, C.*, et al.* (1993). Automated low-redundancy large-scale DNA sequencing by primer walking. *BioTechniques*, *15* (4), 714-721.

Welch, T. J. and Bartlett, D. H. (1997). Cloning, sequencing and overexpression of the gene encoding malate dehydrogenase from the deep-sea bacterium Photobacterium species strain SS9. *Biochimica et biophysica acta*, *1350* (1), 41-46.

Welch, T. J. and Bartlett, D. H. (1998). Identification of a regulatory protein required for pressure-responsive gene expression in the deep-sea bacterium Photobacterium species strain SS9. *Molecular microbiology*, *27* (5), 977-985.

Welch, T. J., Farewell, A., Neidhardt, F. C. and Bartlett, D. H. (1993). Stress response of Escherichia coli to elevated hydrostatic pressure. *Journal of bacteriology*, *175* (22), 7170-7177.

Whitman, W. B., Coleman, D. C. and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, *95* (12), 6578-6583.

Wool, I. G. (1979). The structure and function of eukaryotic ribosomes. *Annual review of biochemistry*, *48*, 719-754.

Xu, Y., Nogi, Y., Kato, C., Liang, Z., Ruger, H.-J., De Kegel, D.*, et al.* (2003). Moritella profunda sp. nov. and Moritella abyssi sp. nov., two psychropiezophilic organisms isolated from deep Atlantic sediments. *International journal of systematic and evolutionary microbiology*, *53* (Pt 2), 533-538.

Yano, Y. (1996). Ph.D. Thesis. *Hokkaido University* .

Yano, Y., Nakayama, A. and Yoshida, K. (1995). Population Sizes and Growth Pressure Responses of Intestinal Microfloras of Deep-Sea Fish Retrieved from the Abyssal Zone. *Applied and environmental microbiology*, *61* (12), 4480-4483.

Yayanos, A. A. (1986). Evolutional and ecological implications of the properties of deep-sea barophilic bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *83* (24), 9542-9546.

Yayanos, A. A. and Pollard, E. C. (1969). A study of the effects of hydrostatic pressure on macromolecular synthesis in Escherichia coli. *Biophysical journal*, *9* (12), 1464-1482.

Yayanos, A. A., Dietz, A. S. and Van Boxtel, R. (1981). Obligately barophilic bacterium from the Mariana trench. *Proceedings of the National Academy of Sciences of the United States of America*, *78* (8), 5212-5215.

Yayanos, A. (1995). Microbiology To 10,500 Meters in the Deep Sea. *Annual Review of Microbiology*, *49* (1), 777-805.

Yayanos, A., Dietz, A. and van Boxtel, R. (1979). Isolation of a Deep-Sea Barophilic Bacterium and Some of Its Growth Characteristics. *Science*, *205* (4408), 808-810.

Zipp, A. and Kauzmann, W. (1973). Pressure denaturation of metmyoglobin. *Biochemistry*, *12* (21), 4217-4228.

Zobell, C. E. and Morita, R. Y. (1957). Barophilic bacteria in some deep sea sediments. *Journal of bacteriology*, *73* (4), 563-568.

ZoBell, C. E. and Morita, R. Y. (1954). Effects of high hydrostatic pressure on physiological activities of marine microorganisms. *O.N.R. Progress Report*, No. 7.