Virginia Commonwealth University

VCU Scholars Compass

2017

Biochemical Analysis of Putative Single-Stranded Nucleic Acid Binding Proteins in Porphyromonas gingivalis

Steve H. Kokorelis
Virginia Commonwealth University

Biochemical Analysis of Putative Single-Stranded Nucleic Acid Binding Proteins in *Porphyromonas gingivalis*

A thesis submitted in partial fulfillment for the requirements for the degree of Master of Science in Biochemistry at Virginia Commonwealth University

By

Steve Harry Kokorelis
Bachelor of Science (Biology), James Madison University
Harrisonburg, Virginia
May 2014

Director: **JANINA P. LEWIS**, Ph.D., Professor

The Philips Institute for Oral Health Research
School of Dentistry, Virginia Commonwealth University

Virginia Commonwealth University

Richmond, Virginia

May 2017

Acknowledgments

First, I would like to express my deepest gratitude to my advisor and mentor, Dr. Janina P. Lewis for the opportunity to pursue my graduate study in her lab. Over the past three years, she has inspired me to pursue my goals with hard work and dedication. She has been an excellent teacher, leader and a great inspiration for me. I would like to thank her for guiding me through my project and believing in my capabilities.

I would also like to thank my committee members, Dr. William Barton for supporting me and encouraging me to pursue my goals as well as being an exemplary teacher, and Dr. Renfeng Li for his advice, time and support.

I would also like to extend my thanks and appreciation towards Dr. Faik N. Musayev and J. Neel Scarsdale for their collaborative work and assistance on my project as well as all the inspiring and supportive faculty that I have met at the Philips Institute.

I am truly grateful for my fellow colleagues and friends in the Lewis Lab, both past and present, who have shared this journey with me. I want to thank Ross Belvin, whose guidance and patience has been invaluable to my project. He has been the go-to person, every time I encountered problems and needed an answer. I want to extend a special thanks Ziaullah Haggani who was my partner in crime in perfecting the protein purification process. To the rest of the lab: Nicai Zollar, Kat Sinclair, Chris Pham, and Qin Gui, getting to know and spend time with each and every one of you has made

these past three years a very enjoyable and entertaining experience that I will never forget.

Last but not least, I would like to thank my parents, George and Vicky Kokorelis, as well as my brother and sister, James and Helen Kokorelis. I am forever indebted to them for supporting and believing me throughout this journey. Thank you for your unconditional love and support.

Table of Contents

List of Tables

List of Figures

Abstract

BIOCHEMICAL ANALYSIS OF PUTATIVE SINGLE-STRANDED NUCLEIC ACID
BINDING PROTEINS IN *PORPHYROMONAS GINGIVALIS*

By Steve Harry Kokorelis
Bachelor of Science in Biology, James Madison University, 2014

A thesis submitted in partial fulfillment of the requirements for the degree of Master of
Science in Biochemistry at Virginia Commonwealth University.

Virginia Commonwealth University, 2017

Major Director: Janina P. Lewis, Ph.D., Philips Institute for Oral Health Research

Proteins that bind to both DNA and RNA embody the ability to perform multiple functions by a single gene product. These nucleic acid binding proteins in prokaryotes can play a vital role in many cellular processes, including replication, transcription, gene expression, recombination, and repair, to name a few. Nucleic acid binding proteins have unique functional characteristics that stem from their structural attributes that have evolved in a widely-conserved manner. In *Escherichia coli* (*E. coli*), the highly-conserved histone-like protein, HU, which predominates as a heterodimer of HUα and HUβ, has been found to bind to both dsDNA and ssDNA. Likewise, RNA-binding proteins contain various structural motifs, many of which are also conserved amongst many bacterial species like the RNA recognition motif. However, in *Porphyromonas gingivalis* (*P. gingivalis*), a periodontal pathogen, the histone-like, HU proteins and the RNA-binding protein (RBP) are not well characterized compared to their respective structures in *E. coli*. In our study, we sought to characterize and compare the HU proteins and RBP in order to gain a better understanding of their structure and function

in the cell. We aimed to determine the oligomeric state of the proteins through size exclusion chromatography and comparative analysis. We also sought to determine the binding characteristics to single-stranded DNA. Our data showed the HU proteins predominate as homo-tetramers and RBP as a monomer. We demonstrated single-stranded DNA binding with all three proteins. We found both *P. gingivalis* HU subunits to bind non-specifically to ssDNA but show preferential binding to poly(dG) content, while binding to poly(dA) the weakest. These results show that HUα, HUβ and RBP are novel ssDNA binding proteins in *P. gingivalis,* indicating an expanded role and function within the cell.

Chapter 1 - Background Significance

1.1 - Oral Microbiome

The oral cavity harbors one of the most diverse microbiomes in the human body, with estimates exceeding 700 different bacterial species[1]. This ecosystem of commensal, symbiotic, and pathogenic microorganisms, is arranged in distinct and complex microbial communities, called biofilms, that have adapted to inhabit a variety of niches in the oral cavity. The colonization and characteristics of these bacterial biofilms are highly-regulated by an assortment of environmental factors (e.g. temperature, pH nutrient availability) as well as host factors (e.g. innate and adaptive immunity)[2]. Because of this host-microbe coevolution, a majority of these microorganisms are commensal, and play a vital role in maintaining oral homeostasis. However, even among these commensal flora, there exist some bacteria that are capable of inflicting disease within and beyond the confines of the oral cavity[3].

1.2 – Periodontal Disease

Biofilms that form on the surface of the teeth are known as dental plaque. Poor oral hygiene encourages bacterial growth, allowing for the buildup of plaque on the teeth, as well as under the subgingival tissue. The first bacterial colonizers of the dental plaque are primarily gram-positive, facultative bacteria; but, if the plaque is allowed to buildup and accrue beneath the gingival surface, an oxygen deprived space, then the biofilm will shift toward favoring more gram-negative, anaerobic bacteria[4]. It is this shift, toward favoring anaerobic bacteria with higher virulence attributes, that is linked to the disruption of the normal homeostatic environment. If allowed to fester, enzymes and

toxins produced by the bacteria in the biofilm, will cause a host-immune response resulting in the inflammation and swelling of the gingiva, called gingivitis. This is the hosts natural defense against harmful bacteria and infection. However, if left untreated, gingivitis can progress to periodontitis; a microbial-induced, chronic inflammatory disease which causes the permanent destruction of the tooth-supporting tissues and bone. In the most severe cases of the disease, exfoliation of the teeth and broader systemic complications can occur. Although, there are many bacterial species associated with the onset and progression of periodontal disease[5]; certain, low-abundant, keystone-pathogens are capable of remodeling the structure and composition of the biofilm as it transitions into a dysbiotic state[6]. One among these keystone-pathogens is *Porphyromonas gingivalis*[6].

1.3 – *Porphyromonas gingivalis*

Porphyromonas gingivalis is a non-motile, gram-negative, rod-shaped, asaccharolytic, anaerobic, pathogenic bacterium that forms black colonies on a blood agar plate. *P. gingivalis* is a member of the phylum Bacteroidetes and is predominantly found in the oral cavity. In the oral cavity, it resides almost exclusively in the oxygen-deprived crevices of the subgingival plaque. *P. gingivalis* is theorized to be a keystone pathogen capable of remodeling the microbial community of the biofilm in ways that promote the development and progression of periodontal disease[6]. It produces a number of virulence factors to colonize the host, evade the host defense mechanisms and damage host tissues. It colonizes by adhering and interacting with other microbial species in the biofilm as well as the extracellular matrix and components of the host

cells[7]. These interactions are mediated through the expression of fimbriae and various surface adhesins, which promote the colonization and the maturation of the plaque biofilm[8]. *P. gingivalis* is also capable of producing a variety of enzymes: hydrolytic, proteolytic, and lipolytic, that can cause destruction of the hosts cells and connective tissues[9]. Through a combination of virulent factors and interactions, *P. gingivalis* is able to effectively invade host cells: epithelial, endothelial, fibroblastic and erythrocytic[10-12]. Internalization of *P. gingivalis* into a host cell, allows it to evade the host immunity, survive, replicate and even re-populate back into the extracellular environment[13]. The attachment and colonization of *P. gingivalis* in biofilms as well as its invasion into host cells causes an array of distinct and important changes in gene expression that are crucial for its adaption to the environment and its survival[14].

1.4 – Environmental Stresses of the Mouth

The human mouth is a confined space that is relatively dark and moist, with temperatures around 36°C and a pH around 7[15]. These conditions are ideal for the growth and survival of many micro-organisms. The mouth is a unique part of the body, in that it is comprised of not only, soft mucous membrane consisting of several layers of epithelial cells; but also, a non-shedding rigid surface called teeth. Comparatively, the mucosal surfaces are a habitat to much less bacteria than the teeth due to normal desquamation, which is the natural shedding of the outermost layer of tissue. As a result, the teeth act as an anchor and a more permanent habitat for bacteria to colonize, which leads to the development of diverse and complex microbial biofilms, called dental plaque. Unfortunately for bacteria, the oral cavity is constantly changing throughout the

day, as environment factors and host factors threaten to disrupt the conditions of their habitat. Such environmental factors can include changes in diet, food-consumption, nutrition, hygiene, saliva flow, oxygen levels, as well as; physical factors like chewing, swallowing and brushing, to name a few. Moreover, the host factors, which include adaptive and innate immunity; are more than capable of recognizing and eradicating harmful bacteria under normal homeostatic conditions. As a result, it is imperative for the survival of each individual bacteria in the oral cavity to be able to react, respond and adapt to the changes in its environment. In other words, each bacterium must be able to react to different environmental stimuli with changes in its genotypic expression in order to produce a phenotypic response that preserves or enhances its survival.

1.4.1 – Environmental Stresses on Gene Expression

Microbes such as *Porphyromonas gingivalis* are constantly exposed to a wide range of environmental changes. The cellular ability to constantly sense and adapt to changes in the environment is crucial for maintaining cellular function and homeostasis. The modulation of gene expression plays a central role in cellular adaption. While any step of gene expression may be modulated, most cases of regulation occur at the level of transcription by deciding which genes will be transcribed into an RNA transcript.

The microbial genome is comprised of a variety of genes, some of which are constitutive, meaning they are continuously transcribed, some of which are facultative, meaning they are only transcribed when needed, and others which are inducible, meaning they are transcribed under certain environmental stimuli and regulatory factors. The regulation of transcription, in terms of when and how many copies of RNA are

transcribed is orchestrated by transcription factors. A transcription factor is a protein that has a sequence-specific DNA-binding domain that enables it to bind near a gene of interest and regulate its transcription. Often times, it's a combination of many transcription factors, whether they be activators, which promote transcription, or repressors, which block transcription; that determines whether a gene is expressed or not. In bacteria, the genomic sequence is often organized in such a way that related genes are found in clusters governed by a single promoter. A promoter is a sequence of DNA that allows for the binding of RNA polymerase and other proteins for the initiation of transcription. A cluster of genes under a single promoter is known as an operon, and is a common regulatory feature of prokaryotes. Moreover, each promoter has a sequence-based affinity for RNA polymerase that in the presence of activators or repressors, can determine the degree of gene transcription. On the other hand, the affinity of an RNA polymerase for a promoter can also vary by the binding of sigma factors. Sigma factors are specialized proteins specifically expressed in response to a downstream signaling cascade initiated by biological stimuli of extracellular or intracellular origin. Conversely, anti-sigma factors bind to sigma factors to further regulate or inhibit transcriptional activity in response to changes in the environment. Although, these biological mechanisms are important for cell survival and adaption, they are not the only ones at play.

Post-transcriptional, post-translational and epigenetic modifications in bacteria, while less prevalent, have also been found to play an important role in gene regulation. For instance, the ability of a transcription factor to recruit RNA polymerase can also be modified post-translationally via phosphorylation, acetylation or glycosylation[16].

Furthermore, toxic molecules like some antibiotics can affect the levels of protein expression by inhibiting translation. In prokaryotes, there is also epigenetic effects on gene regulation, whereby architectural modifications to the genome by certain proteins results in changes in gene expression without ever affecting the genetic code directly[17]. Although, bacteria do not have true histones like eukaryotes, they do have several histone-like proteins which are involved in several major pathways. These histone-like proteins have not only been found to play an integral part in bacterial-nucleoid organization but they also have been found to be involved in numerous cellular processes essential for cell survival and adaption to an ever-changing environment.

1.5 – Bacterial Histone-like Proteins

Nearly all prokaryotic cells synthesize abundantly, a set of conserved, small, basic proteins, usually ~90 residues in length, that bind DNA, called histone-like proteins because their biochemical properties resemble eukaryotic histones[18]. While eukaryotic histones have well defined functions in packaging DNA into nucleosomes, the functions of the prokaryotic histone-like proteins are exceedingly diverse in comparison. For example, histone-like proteins have been found to participate in nearly all DNA-dependent functions within the cell from, architectural roles which preserve the structural integrity of the DNA in processes like replication, transcription, translation and recombination, to regulatory roles like controlling gene expression. Currently, proteins considered to be histone-like include: HU (histone-like protein), IHF (integration host factor), FIS (factor for inversion stimulation) and H-NS (histone-like nucleoid structuring). The similarity of these bacterial histone-like proteins to eukaryotic histones

is not based on amino acid sequence relationships but on DNA-binding ability, low molecular mass, copy number and electrostatic charge[19]. Because, these proteins contribute to the organization of the bacterial nucleoid, they are sometimes categorically labeled nucleoid-associated proteins (NAPs). Of the group, the HU protein was the first one described as histone-like and is the most thoroughly studied[20]. The HU protein is also the most ubiquitous of them all, with approximately 98% of all sequenced prokaryotic genomes encoding at least one allele[18]. Most knowledge about HU functions and characteristics has been derived from studies on Escherichia coli (*E. coli*).

1.6 – Histone-like HU Protein in *Escherichia coli*

The histone-like HU protein was first isolated by Josette Rouviere-Yaniv and Francois Gros, in 1975. It was originally called factor U, for the strain of *E. coli* it was first isolated from, U93; but once its resemblance to eukaryotic histone H2B was determined, the letter 'H' was added to create the acronym HU[21]. Relative to the other histone-like proteins in *E. coli*, the HU protein is one of the most abundantly expressed, with purification studies on its intracellular concentration relative to DNA, estimating 60,000 monomers per genome[22]. This highly abundant protein consists of two subunits, HUα and HUβ, with each monomer having a molecular mass of ~9kDa. Sequentially, the HU protein is one of the most conserved DNA binding proteins found in bacteria[21]. In *E. coli*, the two HU subunits share considerable amino acid homology, roughly a 70% identical match, even though they are not genetically linked[20].

Genetically, the two closely related subunits of the HU protein are encoded by the *hupA* and *hupB* genes, which do not form an operon, but instead are located far apart

on the *E. coli* chromosome map at 90 and 10 minutes, respectively[20]. More specifically,

the *hupA* gene has one promoter, while the *hupB* gene has three different promoters[23].

The expression of these two genes in *E. coli* has been shown to vary throughout the

bacterial growth cycle. During the exponential log phase, when the cell is dividing

rapidly, the HU protein predominates as homodimers (HUα_2, HUβ_2), with the alpha

dimer outnumbering the beta dimer, four to one. However, in the stationary phase,

when total expression is highest, 90% of all the HU protein predominates as a

heterodimer (HU$\alpha\beta$), with the two subunits expressed in relatively equal abundance[23].

Even though the HU protein is highly conserved amongst all bacteria, only a few

enterobacteria species, namely *Escherichia coli, Salmonella typhimurium* and *Serratia*

marcescens, predominate in the heterodimer state, HU$\alpha\beta$[20]. Granted, the two distinct

HU subunits in *E. coli* are highly homologous, the reason or advantage for favoring a

heterodimer state only in enterobacteria, while favoring the homodimer or homo-

tetramer state in virtually all other bacteria examined, remains elusive[20]. Furthermore,

the HU protein doesn't necessary just form dimers but is known to also form higher-

ordered structures. For example, cross-linking studies as well as sedimentation

measurements in *Bacillus stearothermophilus*, has revealed the HU protein to

predominate as a homotetramer[21]. Therefore, caution should be exercised when

interrelating biochemical studies on *E. coli* HU with other bacteria.

Knockout studies with *E. coli* HU protein have found that single *hupA* or *hupB* gene

mutations do not significantly impair growth, however; double knockouts, completely

lacking both HU subunits, do disrupt various cellular processes such as DNA

organization, replication and transposition, which result in phenotypic changes like slow

growth and irregular cell division[24,25]. It is worth mentioning that a double knockout of

hupAB genes in *Bacillus subtilis*, a bacterium with no other histone-like proteins, is

lethal[26]. This suggests histone-like proteins share function and are critical for cell

survival.

Structurally and chemically, previous studies have shown that the two *E. coli* HU

subunits vary in terms of their binding affinities and specificities, which suggests they

may have evolved to perform distinct *in vivo* functions[27]. For this reason, as well as the

nature of the *E. coli* protein coexisting in all three dimeric forms (-αβ, -α$_2$, -β$_2$), there has

been extensive studies comparing the states. Crystal structural studies in the absence

of DNA have shown that the HU protein favors the formation of a conserved, compact,

hydrophobic-core consisting of two intertwining α-helical segments and two positively

charged β-ribbon 'arms', that are usually disordered in the absence of DNA, protruding

from the sides[28-30]. Co-crystal structures with DNA have shown that both, HUαβ and

HUα$_2$, bind double-stranded (ds)DNA in stringent conditions and single-stranded

(ss)DNA in low-salt, with low affinity and non-specifically[18,31]. Comparatively, HUβ$_2$

binds with the least affinity[31]. Nevertheless, all three *E. coli* HU structures bind with a

high affinity to cruciform DNA, meaning DNA with junctions, nicks, gaps, forks, and

overhangs[31]. Therefore, although the *E. coli* HU protein does not have sequence-

specific binding, it does appear to show some differential DNA substrate selectivity. For

example, studies on the *E. coli* HU heterodimer show it has a strong preference for G/C

rich sequences, in both dsDNA and ssDNA, forming a higher-ordered duplex with

increased stability and increased heat resistance[18,32]. On the other hand, DNA

sequences rich in A/T content formed unstable complexes[18,32]. In *E. coli*, the HU

heterodimer has also been found to require 9-11 nucleotides (nt) of dsDNA to bind, with a dissociation constant for supercoiled DNA, relaxed DNA and bacterial RNA at 450, 1300, and 2500 nM, respectively[32]. For ssDNA, the HU heterodimer requires 24nt for initial binding and then a subsequent 12nt for each additional dimer, all under low salt conditions (<20mM NaCl)[18]. This ability for global recognition of such diverse nucleic acid structures can be inferred from its diverse function within the cell.

1.6.1 – Functions of HU in *Escherichia coli*

Functionally, the *E. coli* HU protein has been found to be involved in a wide variety of cellular processes. HU was first described as a histone-like protein, with an architectural function in stabilizing, maintaining and modulating the structure of bacterial DNA. However, as more research has revealed, the HU protein also plays an intricate role in many DNA metabolic processes such as recombination, replication, transcription, translation, and repair, as well as gene regulation[33-37].

The histone-like HU protein is often labeled a nucleoid-associated protein for its role in bacterial nucleoid organization. More specifically, it gets its name for its ability to induce negative supercoiling into relaxed, circular DNA in the presence of topoisomerase I, and then condense it into nucleosome-like structures[38,39]. This induction of negative super-helical tension into bacterial DNA is a highly regulated and important process; for the loss in this tension, has shown to cause dramatic disturbances in vital cellular processes, such as DNA replication, recombination and transcription[40]. The preservation and maintenance of this process is linked to homeostatic regulation of the enzymes, gyrase and topoisomerase I, as well as the HU

protein. Gyrase is responsible for introducing negative supercoiling into DNA, while topoisomerase I and HU relax the DNA to prevent excessive supercoiling. A double knockout study on the HU protein demonstrated this state of homeostatic regulation, by showing that topoisomerase I expression and activity increased significantly to compensate for the loss in HU[38].

The HU protein also has the extensive capacity to regulate translation of genes. Knockouts, in one or both of the *hup* genes in *E. coli* has shown to alter the expression of 353 genes, many linked to the anaerobic response, acid stress response, high osmolality and SOS induction[36]. These 353 genes constitute the HU regulon and correspond to 229 operons, comprising 8% of the entire *E. coli* genome[36]. Furthermore, the HU protein is also responsible for negatively regulating the expression of its own genes, *hupA* and *hupB*[35]. Upstream of the promoter region of the *E. coli hupA* gene is a region of inverted repeats which have the potential to form cruciform DNA structures[41]. It is hypothesized that the HU protein can facilitate the formation of cruciform structures by inducing negative super helical tension at the promoter region, and in doing so; block the access of RNA polymerase and inhibit transcription[41]. The *hupB* gene operates under three different promoters, so its expression is less affected by the steric hindrance imposed by cruciform structures[41].

In much the same way the HU protein induces the formation of cruciform structures in certain regions of the DNA to regulate gene expression, it may also be able to regulate cellular processes, like DNA replication[42]. For example, one study found that the *E. coli* HU protein was able to act as a stimulatory factor for the initiation of oriC-dependent DNA replication by increasing replication threefold[34]. However, because HU

is not required to bind to DNA for replication to occur, it is hypothesized that HU is acting as an accessory factor, inducing a conformation change in the dsDNA to open up a bubble at the origin site (oriC)[34].

The architectural role of the *E. coli* HU protein is attributed to its assembly of higher order HU:DNA structures which facilitate the bending and looping of DNA at specific sequences to stimulate DNA recognition by other proteins[43]. Several studies hypothesize, that this bending of DNA is also precise enough to ensure that specific nucleotide sequences are recognized during site-specific recombination[33,44]. A double-knockout study on the two HU subunits support this claim, by showing the mutant strains were indeed deficient in homologous recombination[33,44].

These diverse and complex mechanistic functions found in the *E. coli* HU protein gives insight into the significant role the protein likely plays in gene regulation. Although the precise pathway by which the HU protein is able to react to biological stimuli and respond with appropriate gene regulation, still remains to be explained.

1.6.2 – Biological Significance of HU in *Escherichia coli*

The pleiotropic roles of the *E. coli* HU protein can be inferred from various phenotypic changes exhibited by knockout *hupAB* double mutants, as well as the wide variety of cellular processes the protein partakes in. Inside the oral cavity, bacteria are exposed to many environmental conditions often involving extreme fluctuations in temperature, salt and acidity, to name a few. Some studies have determined that the HU protein is well adapted, not only to handle these environmental stresses, but also to

react to them accordingly and help produce the appropriate phenotypic response needed for the cell adaption.

Even though, cell viability in double knockout HU mutants is usually not compromised in *E. coli*, many physiological changes have been identified. For starters, cellular growth rate is severely restricted, extending the doubling time by as much as three-fold[24,31]. Colony morphology is also affected resulting in the formation of tiny colonies which often lack a nucleoid region during cell division[24,31].

Changes in the pH or acidity of an environment is also major environmental factor that can determine the growth and pathogenicity of a microbe. Double-knockout *E. coli* HU mutants under acid stress were found to have a 14-fold decrease in cell survival compared to the wild type (WT), while the single HU knockout had a similar survival to the WT.[45] This observation implies that the HU protein plays an important role in the regulation of genes responsible for acid resistance and/or growth at low pH.

Another phenotypic change in *E. coli* HU double knockouts is the resistance to gamma irradiation, whereby; cell survival decreased 5-fold compared to the WT cells[46]. Much like the acid stress test, the single knockout HU mutants had a similar survival to that of the WT, suggesting the presence of the HU protein in either of the homodimer states is sufficient enough to restore function[46].

Studies on sudden changes in temperature, both from heat shock and cold shock proved to be lethal for the double knockout *E. coli* HU mutants[47]. Interestingly enough, the WT under the cold shock conditions, survived by downregulating the expression of *hupA* and upregulated the expression of *hupB*, causing the homodimer, $HU\beta_2$ to predominate[47].

Even though the precise mechanistic pathways by which the HU protein is able to produce these phenotypic responses, to changes in its environment, remains to be explained; they certainly support the significant role the protein must play in cell survival, adaptation and even pathogenesis.

1.7 – Histone-like HU Protein in *Porphyromonas gingivalis*

An *in silico* analysis has found that *Porphyromonas gingivalis* possess both subunits: HUα and HUβ encoded by PG1258 and PG0121 genes, respectively[48]. In one study, a connection between the HUβ protein and the regulation of the K-antigen capsule operon in *P. gingivalis* was determined by generating an *erythromycin* insertion-deletion mutant with the PG0121 coding gene[48]. This studied suggested the HU protein modulates the expression of surface polysaccharides in *P. gingivalis*. In another study, the HU PG0121 was shown to bind to double-stranded (ds)DNA with a strong preference for cruciform structures and DNA composed of G/C rich content[49]. However, in the same study they were unable to show binding to ssDNA under low or high salt conditions, despite acknowledging the likelihood the protein forms multiple hairpin structures[49].

Much like how the *E. coli* HU protein was found to play a biologically significant role in gene expression by binding to RpoS mRNA, a stress sigma factor of RNA polymerase; we suspect the PG HU protein also plays a significant role in gene expression by binding to single-stranded DNA[50]. A growing body of evidence has shown that proteins containing the highly conserved small DNA-binding domain found in HU

and proteins containing the highly-conserved RNA recognition motif, reserve the ability

to perform multiple functions such as binding to dsDNA, ssDNA and RNA[51].

1.8 – RNA-Binding Proteins

In eukaryotic cells, a multitude of RNA-binding proteins (RBPs) play important roles

in many metabolic process from RNA splicing and processing to regulation of DNA

transcription and RNA regulation[52]. The characterization of these proteins led to the

identification of several RNA-binding motifs, found to be highly conserved throughout

the kingdom of life, including prokaryotes and even viruses[53]. The RNA recognition motif

(RRM), was first identified in 1988, and the consensus sequence was thought to be only

involved in RNA interaction[54]. Later, however it was shown that this protein domain was

sufficient for a wide range of functions, including binding to ssDNA and proteins[53]. To

date, there are only 85 known proteins containing the RRM domain in bacteria and six

such proteins in viruses, whereas; in eukaryotes, the RNA recognition motif is one of the

most abundant protein domains, found in 3541 different proteins[53]. Nevertheless,

whether it be animal, plant, fungal, or bacterial cells, they practically all have RNA-

binding proteins wherever RNA is present[55]. This suggests that it is an ancient protein

structure with important functions.

The RNA-binding domain is approximately 90 amino acids in length and contains

two conserved sequences, necessary and sufficient for binding, called ribonucleoprotein

(RNP) 1 and 2. RNP-1 is a central sequence about eight residues in length that is highly

conserved, positively charged and defined as Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe/Tyr-

Val/Ile/Leu-X-Phe/Tyr, where X can be any amino acid[53]. RNP-2 is a comparatively

less-conserved sequence, 6 residues in length, found near the N-terminal domain and defined as Ile/Val/Leu-Phe/Tyr-Ile/Val/Leu-X-Asn-Leu[53]. It has been shown that this protein domain can modulate its structure and folds to recognize many RNAs, DNAs and proteins in order to achieve a multitude of biological functions[53].

1.8.1 – Functions of RBPs in Bacteria

In order for bacteria to survive in changing environments, the cell must be able to react to environmental stimuli, alter its gene expression and adjust protein levels accordingly. One way this can be achieved is by regulating transcription initiation with sigma factors and proteins that activate or repress transcription[56]. Another way is post-transcriptionally, by modulating RNA decay, translation initiation efficiency or transcript elongation[56].

Ribonucleases (RNases) can degrade target mRNAs by binding to specific recognition sites and forming endoribonucleolytic cleavages[56]. Functionally, the RBP can prevent the degradation of mRNA transcripts or small (s)RNAs, by directly binding to and shielding these RNase recognition sites[56]. RBPs can also regulate translation, both positively or negatively, by changing the secondary structure of mRNAs to either hide or expose the RNase recognition sites[56].

Some RBPs can also control gene expression by altering the efficiency of translation initiation. Functionally, the RBP can inhibit translation by directly binding to the ribosomal binding site which contains the Shine-Dalgarno sequence responsible for recruiting the ribosome[56]. By directly competing with ribosomes, RBPs can modulate the efficiency of translation initiation. RBPs can also regulate translation, both positively or

negatively, by changing the secondary structure of mRNAs to either hide or expose the ribosomal binding site[56].

RBPs can also regulate translation initiation or RNA stability indirectly by assisting in the recruitment and interaction of other regulatory molecules, like sDNAs and proteins[56]. Intermolecularly, RBPs can assist in the base pairing between regulatory molecules and the mRNA[56]. These chaperon-like RBPs can either positively or negatively affect translation depending on the regulatory molecule and the site at which it binds to.

The last mechanism by which RBPs can post-transcriptionally affect gene expression is through the regulation of transcription elongation. During transcription, primary mRNA transcripts are elongated until a terminator is reached. There are two classes of terminators: intrinsic termination or factor-dependent termination. In terms of intrinsic termination, the RBP can stabilize either the terminating structure or an alternative secondary structure to prevent a termination structure from ever forming[56]. Often the formation of both structures is mutually exclusive. The effect of this type of regulation is either the prevention or assistance of premature termination of mRNA transcript. In terms of factor-dependent termination, the RBP can bind to the mRNA transcript inducing a secondary structural change to either hide or expose a transcription-factor binding site[56].

1.9 – RNA-Binding Protein in *Porphyromonas gingivalis*

There is currently very little research regarding any RNA-binding proteins in *Porphyromonas gingivalis.* However, recently, high-throughput sequencing technologies

have provided the opportunity to relate functional genomics to basic biology, revealing

PG0627 as a likely candidate for an RNA-binding protein in *Porphyromonas gingivalis*[57].

1.10 – Single-stranded Nucleic Acid Binding Proteins

Proteins that bind to both DNA and RNA embody the ability to perform multiple functions by a single gene product. These nucleic acid binding proteins in prokaryotes can play a vital role in many cellular processes, including replication, transcription, gene expression, recombination, and repair, to name a few. Nucleic acid binding proteins have unique functional characteristics that stem from their structural attributes that have evolved in a widely-conserved manner. In *Escherichia coli* (*E. coli*), the highly-conserved histone-like protein, HU, has been found to bind to both dsDNA and ssDNA and have a multi-purpose function within the cell[18]. Likewise, a growing body of evidence has showed that a number of proteins, containing the highly-conserved RNA recognition motif, like the RNA-binding protein (RBP) is capable of binding to both, DNA and RNA[53]. Although, the HU protein does not possess any sequence or structural homology to the RNA recognition motif, it does contain a highly-conserved, small DNA-binding domain formed from two β-ribbon arms and an α-helical core, that in *E. coli*, has been shown to be capable of binding to both DNA and RNA[18]. In this study, we aim to investigate the biochemical characteristics of the histone-like HU protein and the RNA-binding protein (RBP) found in *Porphyromonas gingivalis*; in the hopes of better understanding their function and potential role within the cell.

Chapter 2 – Hypothesis and Aims

2.1 – Hypothesis

We hypothesize that *Porphyromonas gingivalis* genes PG0121, PG1258, and PG0627 code for single-stranded nucleic acid binding proteins, HUβ, HUα and RBP respectively. Furthermore, we predict the HU proteins will bind non-specifically to single-stranded DNA but will have stronger preference for poly(dG), while binding to poly(dA) the weakest. We suspect the HU proteins will form multimeric states but predominate as homo-tetramers while RBP will predominate as a monomer.

2.2 – Aims

There is currently very little research regarding any single-stranded nucleic acid binding proteins in *P. gingivalis* or the regulatory mechanisms by which they operate gene expression, therefore; the main purpose of this project is to investigate and better characterize these proteins.

Aim 1: To conduct a bioinformatic analysis on the *P. gingivalis* genes: PG0121, PG1258 and PG0627 and their respective gene products to determine their functional homology to the extensively researched *E. coli* homologs.

Aim 2: To express and isolate each protein using histidine-tag column purification and size exclusion chromatography.

Aim 3: To determine the oligomeric state of each protein using size exclusion chromatography comparative analysis, because it is currently unknown and there are conflicting reports on homologous protein studies.

Aim 4: To determine the binding characteristics and conditions for each protein to single-stranded DNA, using electrophoretic mobility shift assays (EMSAs). The *P. gingivalis* HU protein is known to bind dsDNA but current studies have failed to determine whether it can bind ssDNA. Understanding the binding characteristics of these proteins will give us a better understanding of the potential roles it may place in cellular processes.

Chapter 3 – Materials and Methods

3.1 – Bioinformatic Analysis

For the bioinformatic analysis, we used fast, scalable generation of high-quality protein multiple sequence alignments with Clustal Omega[58]. For the generation of protein homology models, one to one threading from the Phyre 2 web portal for protein modeling, prediction and analysis was used[59]. Molecular graphics and analysis were performed with the UCSF Chimera package[60]. The theoretical protein molecular weights were calculated by the addition of average isotopic masses of amino acids using the analysis tools on the ExPASy Server[61].

3.2 - Cloning and Expression of Recombinant *Porphyromonas gingivalis* HU

The coding regions, PG0121 and PG1258 from *P. gingivalis* strain W83, were originally sub-cloned into the pCR2.1-TOPO Vector. For cloning and expression of PG0121 and PG1258 a modified (m-)pET21d vector was used (Table 2). Primers were used to PCR amplify PG0121 and PG1258 from *P. gingivalis* strain W83 genomic DNA (Table 1). The forward primer was designed to have a restriction site for BamH1. The reverse primer was designed to have a restriction site for Xho1. The m-pET21d vector contains a 6x Histidine-Tag on the amino-terminus for protein purifications.

A double digestion was performed on the respective PCR amplified genes with the m-pET21d vector. The digestions were run on a 1% agarose gel and gel extracted (QIAGEN) after positive verification. The gene inserts and vector were T4 DNA ligated (NEB) to generate m-pET21d-*gene* (Figure 1) and transformed into One Shot TOP10 Chemically Competent *Escherichia coli.* Lastly, the transformed *E. coli* cells were

screened for successful ligation using Carbenicillin and then transformed into

BL21(DE3) Chemically Competent *Escherichia coli.*

3.3 - Cloning and Expression of Recombinant *Porphyromonas gingivalis* RBP

The coding regions, PG0627 from *P. gingivalis* strain W83, were originally sub-cloned into the pCR2.1-TOPO Vector. For cloning and expression of PG0627, pET-30a vector was used (Table 2). PG0627 from *P. gingivalis* strain W83 genomic DNA was inserted between the restriction sites Nde1 and Xho1 on the pET-30a vector to produce a 6x Histidine-Tag on the Carboxy-terminus for protein purifications (Figure 2).

A double digestion was performed on the respective PCR amplified genes with the pET-30a vector. The digestions were run on a 1% agarose gel and gel extracted (QIAGEN) after positive verification. The gene inserts and vector were T4 DNA ligated (NEB) to generate pET30a-*gene* (Figure 2) and transformed into One Shot TOP10 Chemically Competent *Escherichia coli.* Lastly, the transformed *E. coli* cells were screened for successful ligation using Kanamycin and then transformed into BL21(DE3) Chemically Competent *Escherichia coli.*

Table 1 - Primers used for cloning recombinant *P. gingivalis* HU

Name	BamH1 Sequence	Description
PG0121-Forward	CTTCCAG**GGATCC**ATGAACAAGACAGATTTTATTGC	Forward primer for *PG0121*
PG0121-Reverse	CAGCGCA**CTCGAG**TTACTTAAGTTCCAAAGTAGAGCCC	Reverse primer for *PG0121*
	Xho1	

Name	BamH1 Sequence	Description
PG1258-Forward	CTTCCAG**GGATCC**ATGACGAAAGCTGACGTAGTGAAC	Forward primer for *PG1258*
PG1258-Reverse	CAGCGCA**CTCGAG**TTAGTCTTGTTTCATCTGACTCATAAAG	Reverse primer for *PG1258*
	Xho1	

Table 2 – Expression Vectors used in this study

P. gingivalis HU	Description
m-pET21d	Modified pET21d vector with *P. gingivalis* HU insertion Contains a 6x His-tag on amino-terminal

P. gingivalis RBP	Description
pET30a	pET30a vector with *P. gingivalis* RBP insertion Contains His-tag on carboxy-terminal

Figure 1 – Cloning strategy to insert *P. gingivalis* HU genes into m-pET21d vector.

A double digestion was performed using BamH1 and Xho1 restriction enzymes to

cleave the vector at the specified restriction sites. T4 DNA ligation was used to insert

the respective *P. gingivalis* HU genes into the m-pET21d vector. The HU insert was

PCR amplified using primers designed to contain restriction sites for BamH1 and Xho1.

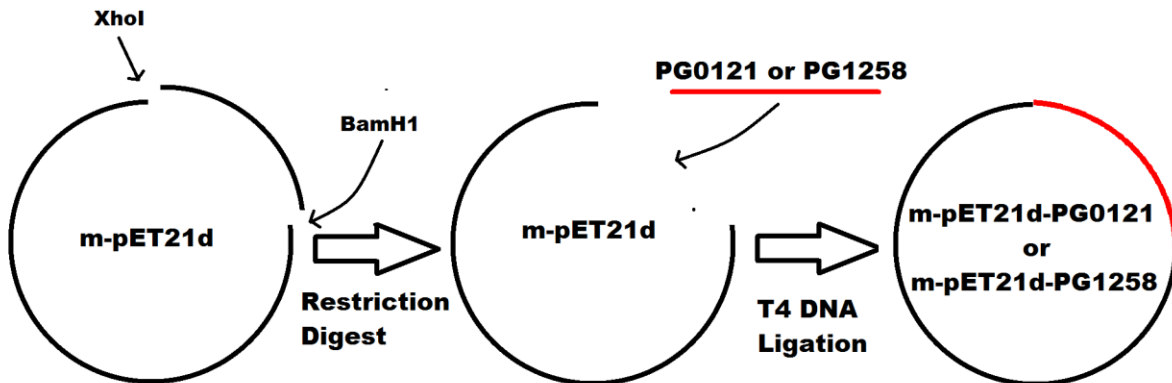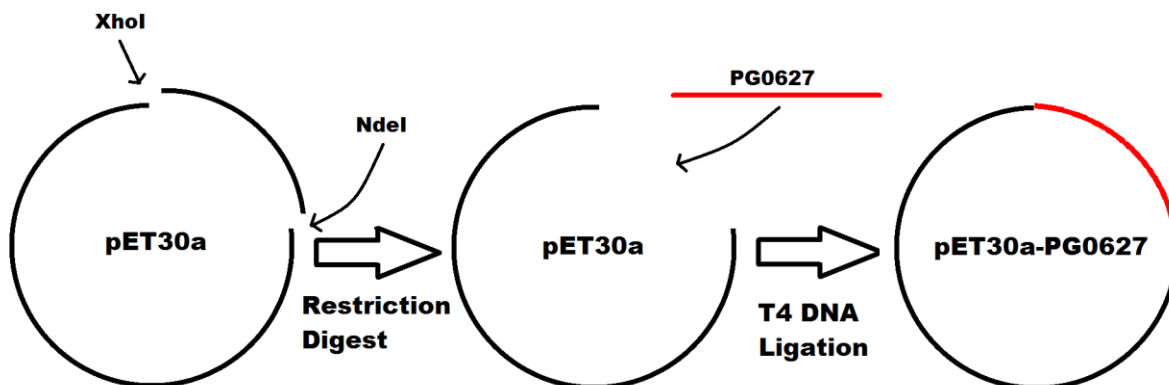**Figure 2 – Cloning strategy to insert *P. gingivalis* RBP into pET30a vector.**

A double digestion was performed using Nde1 and Xho1 restriction enzymes were used

to cleave the vector at the specified restriction sites. T4 DNA ligation was used to insert

P. gingivalis RBP gene into the pET30a vector. The RBP insert was PCR amplified

using primers designed to contain restriction sites for Nde1 and Xho1.

3.4 Purification of Recombinant *P. gingivalis* HU and RBP

3.4.1 Preparing Starter Cultures and Auto-Induction Media

The recombinant *P. gingivalis* HU and RBP strains were used to inoculate separate media containing Luria-Bertani (LB) and antibiotic: carbenicillin (50ug/mL) for HU and kanamycin (50ug/mL) for RBP. These cultures were left overnight to grow at 37°C at 220 RPM.

The following day, the starter cultures were used to inoculate auto-induction (AI) media, containing 200g/L of Glycerol, 20g/L of glucose, 80g/L of lactose, 1mM MgSO4, and antibiotic: carbenicillin (50ug/mL) for HU and kanamycin (50ug/mL) for RBP, respectively. The AI media were left overnight to grow at 37°C at 220 RPM.

3.4.2 - His-Tag Purification of Recombinant *P. gingivalis* HU and RBP

After overnight growth, the cell cultures were centrifuged at 8,000 RPM for 10 minutes. The cells were washed with PBS buffer and then resuspended in 20uL of His Binding Buffer (50mM Na_2HPO_4, 300mM NaCl, and 20mM imidazole adjusted to pH 8). To facilitate lysis of the cells, Lysozyme (10mg/mL) and 10X CelLytic™ B cell Lysis Reagent (Sigma) were added to the His Binding Buffer. Additionally, benzonase (Sigma) was added at 25units per 20mL of buffer to degrade genomic DNA. The cells were vortexed and incubated for 30 minutes at room temperature. After lysis, the cells were centrifuged at 15,000 RPM for 20 minutes and the lysate (supernatant) was collected. The cell lysates were passed through a Ni-NTA Resin (Qiagen) flow column which has been equilibrated with His Binding Buffer. The column was washed with His Wash Buffer (50mM NaH_2PO_4, 300mM NaCl, and 30mM imidazole adjusted to pH 8).

The column was washed again with a second His Wash Buffer (50mM NaH_2PO_4, 300mM NaCl, and 60mM imidazole adjusted to pH 8). The His-tagged protein was eluted from the Ni-NTA Resin using His Elution Buffer (50mM NaH_2PO_4, 300mM NaCl, and 250mM imidazole adjusted to pH 8). The elutions were run on a 12% Bis-Tris denaturing gel to assess purity of protein.

3.4.3 – Ammonium Sulfate Precipitation of HU and RBP

In order to concentrate the proteins, elution collections were salted out by adding ammonium sulfate to reach 60% saturation at 20°C. The proteins were centrifuged at 15,000 RPM for 10 minutes. The cell lysates were resuspended in a Tris-Base Buffer solution (20mM Tris-Base, 200mM NaCl, 1mM Tris(2-carboxyethyl)phosphine hydrochloride adjusted to pH 8).

3.4.4 – Dialysis of HU and RBP

In order to reduce residual salts, elution collections were dialyzed with 7,000 MWCO SnakeSkin™ Dialysis Tubing (Thermo Scientific) in a Tris-Base Buffer solution (20mM Tris-Base, 200mM NaCl, 1mM Tris(2-carboxyethyl)phosphine hydrochloride adjusted to pH 8) overnight at 4°C.

3.4.5 – Size Exclusion Chromatography

Dialyzed cell lysates were run on an ÄKTA™ pure Fast Protein Liquid Chromatography machine using a Superdex 75 (10/300 GL) column (GE Healthcare) or a HiLoad Superdex 75 (16/600) column (GE Healthcare). The columns were

equilibrated using a Tris-Base Buffer solution (20mM Tris-Base, 200mM NaCl, 1mM

Tris(2-carboxyethyl)phosphine hydrochloride adjusted to pH 8). The proteins were

collected from the column using the generated elution fraction profile based on UV

absorption at 220nm. The collected fractions were loaded with LDS(4X) Sample Buffer

(Invitrogen) onto a 12% Bis-Tris denaturing gel run set at 155V for 55minutes using

MES SDS Running Buffer (novex). After electrophoresis, the gel was stained using a

Pierce 6x His Protein Tag Stain Reagent kit (ThermoFisher) to detect the presence of

the His-tagged proteins. The fraction collections were stored at 4°C for later studies.

Molecular weight markers were run on both Superdex 75 columns using the

same Tris-Base Buffer solution to generate a calibration curve based on volume at

which the markers and proteins eluted off the column. Using the curve, an equation was

derived and used to determine the experimental molecular weight of eluted proteins.

The following molecular markers were used: Blue Dextran (2000 kDa), Albumin (66

kDa), Carbonic Anhydrase (29 kDa) and Cytochrome C (12 kDa).

3.5 – Electrophoretic Mobility Shift Assay with Recombinant *P. gingivalis* HU and RBP

For all electrophoretic mobility shift assays (EMSA), a target-specific, ssDNA

oligonucleotide with a length of 48nt was designed for both HU and RBP (adapted from

Kamashev, et al. 2007). This sequence was comprised of G/C rich content and was

modified with a 5' IRDye® 700 (IDT) fluorophore (Table 3). For inhibition studies, an

identical sequence without the fluorescent tag was designed (Table 3). For non-specific

binding studies, a random sequence, ssDNA oligonucleotide with a length of 48nt, without a fluorescent tag was designed (Table 3).

Both, recombinant *P. gingivalis* HU and RBP proteins were purified via His-tag affinity chromatography and size exclusion chromatography before being used for this assay. HcpR from *P. gingivalis* was purified using the His-tag purification system from a pET30 vector to run as a negative control. Two types of shifts were performed: uninhibited shifts and competitive inhibition shifts.

3.5.1 - Uninhibited Shift Assays

Increasing concentrations of purified recombinant *P. gingivalis* HU and RBP ranging in concentrations from 0nM – 7nM were incubated for at least 1hour, in the dark, at room temperature, with fluorescently labeled ssDNA (10nM) under the following binding conditions: 20mM Tris-HCl (pH 8), 10mM NaCl, 0.1mM EDTA, 0.05 mg/mL BSA, and 7% Glycerol (adapted from Kamashev, et al. 2007). All fluorescently labeled DNA was heated to 100 degrees centigrade for 8 minutes and kept on ice prior to mixing to prevent the single-stranded DNA from annealing to itself. The samples were loaded onto a 10% TBE Native Gel (Invitrogen) buffered with 27mM Tris-Borate, 0.1mM EDTA and run at 100 volts for 90 minutes. The gel shifts were captured on an Odyssey® CLx imaging system (Li-COR).

3.5.2 - Competitive Inhibition Shift Assays

Purified recombinant HU proteins were incubated with fluorescent (labeled) and non-fluorescent (unlabeled) probes at varying concentrations (Table 3). 2uM *P.*

gingivalis HU proteins were incubated separately with both labeled and unlabeled probes under the following conditions: 0.5nM labeled probe with increasing concentrations of identical unlabeled probe at 0.5nM, 5nM, 25nM. Likewise, 0.5nM labeled probe was added with increasing concentration of random unlabeled probe at 0.5nM, 5nM, and 25nM. All probes were heated to 100 degrees centigrade for 8 minutes and kept on ice prior to mixing to prevent the single-stranded DNA from annealing to itself. All reactions were incubated for at least 1hour, in the dark, at room temperature, under the following binding conditions: 20mM Tris-HCl (pH 8), 10mM NaCl, 0.1mM EDTA, 0.05 mg/mL BSA, and 7% Glycerol (adapted from Kamashev, et al. 2007). The samples were loaded onto a 10% TBE Native Gel (Invitrogen) buffered with 27mM Tris-Borate, 0.1mM EDTA and run at 120 volts for 100 minutes. The gel shifts were captured on an Odyssey® CLx imaging system (Li-COR).

Table 3 – EMSA Primers for *P. gingivalis* HU and RBP

Name	Sequence	Description
Labeled Probe	5'- /5IRD700/ AGTCTAGAGTGCAGTTGAGT CCTTGCTACGACGGATCCCTTAGGTCAG -3'	Fluorescently labeled single-stranded DNA
Unlabeled Probe	5'- AGTCTAGAGTGCAGTTGAGTCCTTGCTA CGACGGATCCCTTAGGTCAG -3'	Unlabeled duplicate single-stranded DNA
Random Probe	5'- ATTCTAGATTACATTTTAATCCTTACTA CGACTTATCCCTTAAATCAA -3'	Unlabeled random seq. single-stranded DNA

Chapter 4 – Results

4.1 – Bioinformatic Analysis

4.1.1 – *P. gingivalis* HUα (PG1258) and HUβ (PG0121) are homologous to *E. coli* HU subunits.

An amino acid sequence comparison shows that PG0121 has a 47.73% identity to the β subunit of the *E. coli* HU (Figure 3A). Likewise, PG1258 has a 37.78% identity to the α subunit of the *E. coli* HU (Figure 3B). An amino acid sequence comparison between the *P. gingivalis* HU subunits shows a 29.55% identity were as the *E. coli* HU subunits shared a 68.89% identity (Figure 3C and 3D). From the multi-sequence alignment, we could indicate which regions of each sequence are homologous. One such area of significant conservation is the histone-like DNA binding site found between the 44th and 65th residue in all four proteins (Figure 4).

The theoretical molecular weight of *P. gingivalis* HUα and HUβ was calculated to be 10,298.96 Da and 9,650.23 Da, respectively. The theoretical molecular weight of *P. gingivalis* RBP was calculated to be 11,500.83 Da. After the additions of their respective histidine tags, the theoretical molecular weight for *P. gingivalis* HUα, HUβ and RBP was calculated to be 12,320.13 Da, 11,671.39 Da, and 12,323.68 Da, respectively.

A superposition of the backbone resulted in a root mean square deviations (RMSDs) for the models and their respective *E. coli* HU crystal structure templates that was very low (between 69 atom pairs <0.1 Å for HU PG0121 and between 68 atoms pairs 0.163 Å for HU PG1258) indicating a high degree of structural similarity, as expected from their high degree of sequence identity (Figure 5A and 5B). Both, *P. gingivalis* HU subunit homology models indicated structurally conserved elements of the

HU 'body' and b-ribbon 'arms' that protrude to the side (Figure 5A and 5B). The conserved geometry and positive electrostatic potential of these individual arms suggests that the *P. gingivalis* HU will also be capable of binding DNA in a similar manner as the *E. coli* HU homolog.

For the *P. gingivalis* RNA-binding protein, which is 97 amino acids in length we were able to solve a three-dimensional crystal structure of the protein to 1.9 Å (Figure 6). RBP contains the highly-conserved RNA recognition motif (RRM) consisting of the RNP-1 consensus sequence, necessary and sufficient for binding and found in eukaryotes, prokaryotes, and viruses. The RRM is found between the 2nd and 92nd residue, comprising nearly the entire protein's structure. The RRM consists of four anti-parallel beta-strands and two alpha-helices arranged in a β_1-α_1-β_2-β_3-α_2-β_4 fold with side chains that stack with RNA bases (Figure 6). From the crystal structure, we can also deduce that the active state of the protein predominates in the monomeric form. It has been shown in previous literature, that this protein domain can modulate its structure and folds to recognize many RNAs, DNAs, and proteins in order to achieve a multitude of biological functions.

Three-dimensional (3D) homology models of the *P. gingivalis* HU homodimers were modeled after the *E. coli* HU heterodimer crystal structure. The β-ribbon 'arms' for the *P. gingivalis* HU homology models were removed to more accurately compare to the surface area of the *E. coli* HU heterodimer crystal structure. It is worth mentioning that the β-ribbon 'arms' had no impact on the structural alignment of HU 'body' or the interface between the dimers. The *E. coli* HU subunits had a total solvent excluded surface area of 4223.0 Å^2 and 4295 Å^2, respectively; while the dimer structure had an

area of 6525.0 $Å^2$ (Figure 7A). The total solvent excluded surface area of the interface

between the dimers was calculated to be 996.5 $Å^2$ (Figure 7A). The homology model of

the PG1258 HU homodimer had a total surface area of 4368.8 $Å^2$ for each subunit and

an area of 7425.2 $Å^2$ for the complete structure (Figure 7B). The total solvent excluded

area of the interface between the dimers was calculated to be 656.2 $Å^2$ (Figure 7B).

The homology model of the PG0121 HU homodimer had a total solvent excluded

surface area of 4563.0 $Å^2$ for each subunit and an area of 6685.9 $Å^2$ for the complete

structure (Figure 7C). The total solvent excluded surface area of the interface between

the dimers was calculated to be 1220.1 $Å^2$ (Figure 7C).

A. Clustal Omega Sequence Alignment between HU PG0121 and *E. coli* HUβ

```
                1: Ecoli_HUβ    100.00    47.73
                2: PG_0121       47.73   100.00
```

```
Ecoli HUβ     MNKSQLIDKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFGTFAVKERAARTG
PG0121        MNKTDFIAAVAEKANLTKADAQRAVNAFAEVVTEQMNAGEKIALIGFGTFSVSERAARKG
              ***::::*   :*   *:::** * **::*:   ***.:: *:.:**:*****:*.*****.*

Ecoli HUβ     RNPQTGKEITIAAAKVPSFRAGKALKDAVN
PG0121        INPKTKKSISIPARKVVRFKPGSTLELK--
               **:* *.*:* * **   *: *.:*:
```

B. Clustal Omega Sequence Alignment between HU PG1258 and *E. coli* HUα

```
                1: Ecoli_HUα    100.00    37.78
                2: PG_1258       37.78   100.00
```

```
Ecoli HUα     MNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFGTFKVNHRAERTG
PG 1258       MTKADVVNAIAKSTGIDKETTLKVVESFMDTIKDSLSEGDNVYLRGFGSFIVKERAEKTA
              *.*:::::.**:.: :.*   :   .:** : :*.:**.*** * * ***:* *:.***:*.

Ecoli HUα     RNPQTGKEIKIAAANVPAFVSGKALKDAVK--
PG 1258       RNISKQTTIIIPKRNIPAFKPSKIFMSQMKQD
              **  .. .  * *   *:***  .* : . :*
```

C. Clustal Omega Sequence Alignment between *E. coli* HUβ and *E. coli* HUα

```
                1: Ecoli_HUβ    100.00    68.89
                2: Ecoli_HUα     68.89   100.00
```

```
Ecoli HUβ     MNKSQLIDKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFGTFAVKERAARTG
Ecoli HUα     MNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFGTFKVNHRAERTG
              ***:**** **   *::**: *   **:: :*::******** * ******* *:.** ***

Ecoli HUβ     RNPQTGKEITIAAAKVPSFRAGKALKDAVN
Ecoli HUα     RNPQTGKEIKIAAANVPAFVSGKALKDAVK
              *********.****:**:* :*********:
```

D. Clustal Omega Sequence Alignment between HU PG0121 and HU PG1258

```
                1: PG_0121     100.00    29.55
                2: PG_1258      29.55   100.00
```

```
PG 0121       MNKTDFIAAVAEKANLTKADAQRAVNAFAEVVTEQMNAGEKIALIGFGTFSVSERAARKG
PG 1258       MTKADVVNAIAKSTGIDKETTLKVVESFMDTIKDSLSEGDNVYLRGFGSFIVKERAEKTA
              *.*:*.: *:*:.: : *   : :.*::* :.:.:.:. *::: * ***:* *.*** :..

PG 0121       INPKTKKSISIPARKVVRFKPGSTLELK----
PG 1258       RNISKQTTIIIPKRNIPAFKPSKIFMSQMKQD
              *  ..:.:* ** *::  ***.. :   :
```

Figure 3 – Clustal Omega Pairwise Sequence Alignment between HU PG0121, HU PG1258 and *E. coli* HU subunits.

Pairwise sequence alignments of HU PG0121, HU PG1258, *E. coli* HUα and *E. coli* HUβ, using Clustal Omega at EMBL-EBI. A) Pairwise sequence alignment between HU PG0121 and E. coli HUβ. B) Pairwise sequence alignment between HU PG1258 and E. coli HUα. C) Pairwise sequence alignment between E. coli HUβ and E. coli HUα. D) Pairwise sequence alignment between HU PG0121 and HU PG1258. An * (asterisk) indicates positions which have a single, fully conserved residue. A : (Colon) indicates conservation between groups of strongly similar properties – scoring > 0.5 in the Gonnet PAM 250 matrix. A . (period) indicates conservation between groups of weakly similar properties – scoring ≤ 0.5 in the Gonnet PAM 250 matrix.

```
PG1258        MTKADVVNAIAKSTGIDKETTLKVVESFMDTIKDSLSEGDNVYLRGFGSFIVKERAEKTA
PG0121        MNKTDFIAAVAEKANLTKADAQRAVNAFAEVVTEQMNAGEKIALIGFGTFSVSERAARKG
Ecoli_HUα     MNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFGTFKVNHRAERTG
Ecoli_HUβ     MNKSQLIDKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFGTFAVKERAARTG
              *.*::.:  :*   : : *   :   .::    :.:.:. *: : * ***:* *..** :..

PG1258        RNISKQTTIIIPKRNIPAFKPSKIFMSQMKQD
PG0121        INPKTKKSISIPARKVVRFKPGSTLELK----
Ecoli_HUα     RNPQTGKEIKIAAANVPAFVSGKALKDAVK--
Ecoli_HUβ     RNPQTGKEITIAAAKVPSFRAGKALKDAVN--
               *  .. . * *   ::   *   .. :
```

Figure 4 – Clustal Omega Multiple Sequence Alignment Between HU PG0121, HU PG1258 and *E. coli* HU subunits.

Multiple sequence alignment of HU PG0121, HU PG1258, E. coli HUα and E. coli HUβ, using Clustal Omega at EMBL-EBI. An * (asterisk) indicates positions which have a single, fully conserved residue. A : (Colon) indicates conservation between groups of strongly similar properties – scoring > 0.5 in the Gonnet PAM 250 matrix. A . (period) indicates conservation between groups of weakly similar properties – scoring ≤ 0.5 in the Gonnet PAM 250 matrix.
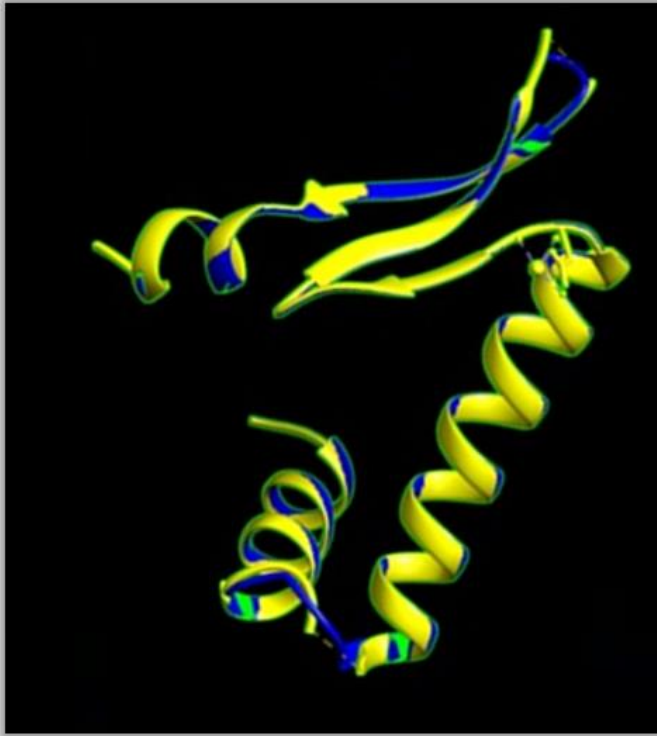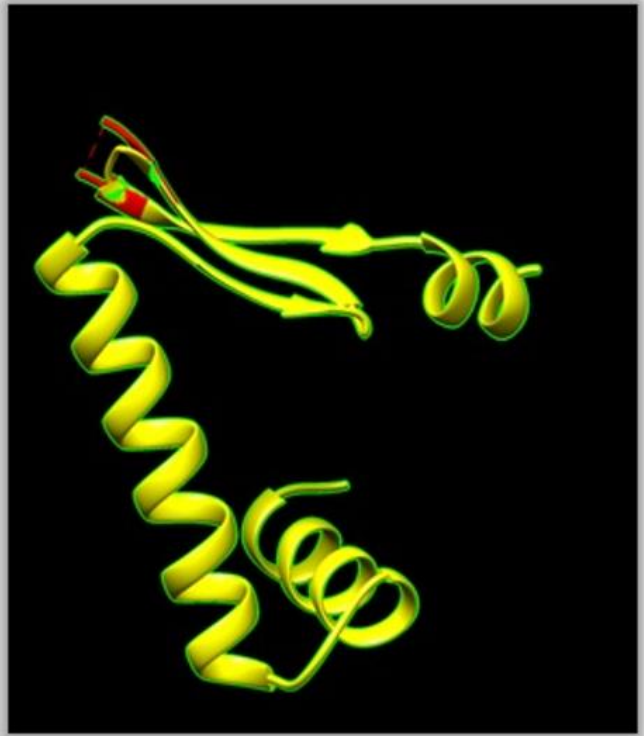
Figure 5 – Homology model of *P. gingivalis* HU proteins based on *E. coli* HU crystal structures.

Three-dimensional (3D) homology models of the *P. gingivalis* HU protein subunits generated using one to one threading from the Phyre 2 web portal. Visualizations were generated using the UCSF Chimera package. A) Homology model of the HU PG0121, in yellow, based on the *E. coli* HUβ crystal structure in blue. RMSD value between 69 atom pairs is <0.1A. B) Homology model of the HU PG1258, in yellow, based on the *E. coli* HUα crystal structure in blue. RMSD value between 68 atom pairs is 0.163A.
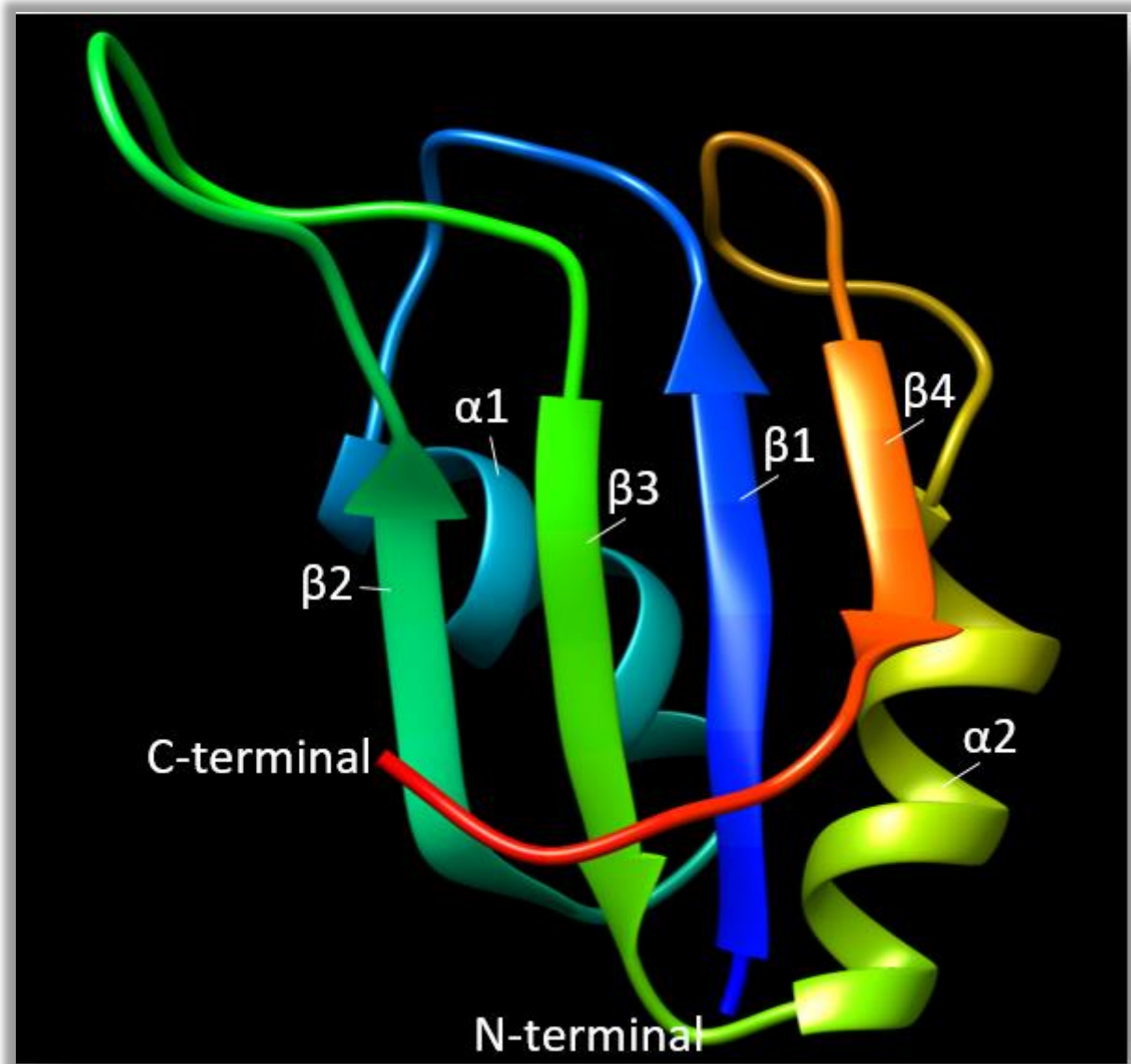
Figure 6 – The three-dimensional Crystal Structure of *P. gingivalis* RNA-binding protein has been solved to 1.9Å.

The three-dimensional crystal structure of *P. gingivalis* RBP has been solved to 1.9Å.

The RRM consists of four anti-parallel beta-strands and two alpha-helices arranged in a β_1-α_1-β_2-β_3-α_2-β_4 topological pattern with flexible side chains that stack with RNA bases.

The protein predominates in a monomeric state.
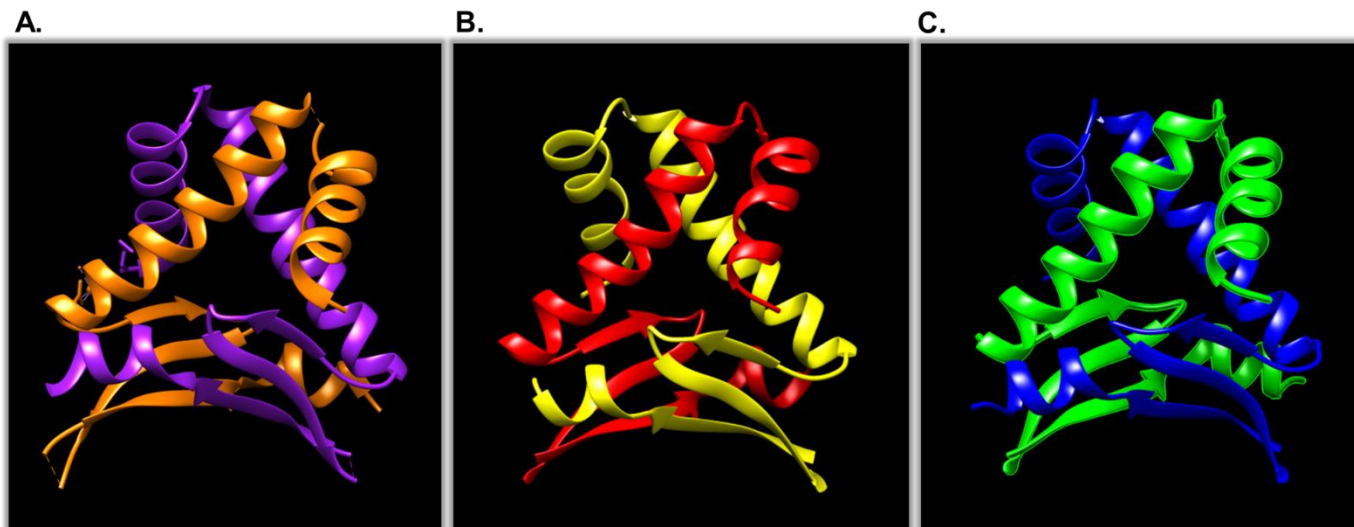
Figure 7 – Homology models of the *P. gingivalis* HU homodimers and the crystal structure of the *E. coli* HU heterodimer.

Three-dimensional homology models of the *P. gingivalis* HU homodimers were modeled after the *E. coli* HU heterodimer crystal structure. Visualizations were generated using the UCSF Chimera package. The β-ribbon 'arms' for the *P. gingivalis* HU homology models were removed to more accurately compare to the surface area of the *E. coli* HU heterodimer crystal structure. A) Crystal structure of the *E. coli* HU heterodimer: the α-subunit in purple and the β-subunit in orange. The total solvent excluded surface area of the interface between the dimers was calculated to be 996.5 Å^2. B) Homology model of the PG1258 HU homodimer: the β-subunits are labeled in red and yellow. The total solvent excluded surface area of the interface between the dimers was calculated to be 656.2 Å^2. C) Homology model of the PG0121 HU homodimer: the α-subunits are labeled in green and blue. The total solvent excluded surface area of the interface between the dimers was calculated to be 1,220.1 Å^2.

4.2 – Purification of Recombinant *P. gingivalis* HU and RBP

4.2.1 – His-Tag Purification of Recombinant *P. gingivalis* HU and RBP

The gene products from PG0121, PG1258, and PG0627, were purified using Ni-NTA resin via His-tag purification system. The purified proteins were run on a 12% Bis-Tris denaturing gel with MES running buffer. The purified RBP with the His-tag appears approximately at 12 kDa (Figure 8A). The purified PG0121 and the purified PG1258 with 6x His-tag appear at approximately 12 kDa (Figure 8B and 8C). Because the purification of PG1258 resulted in two bands: a less expressed band at approximately 25 kDa and a more expressed band at approximately 12 kDa; a 6x His Protein Tag Stain kit was used to confirm the size and location of the protein. The His-tag stain detected the presence of the His-tag only in the more expressed band at approximately 12 kDa (Figure 8D).
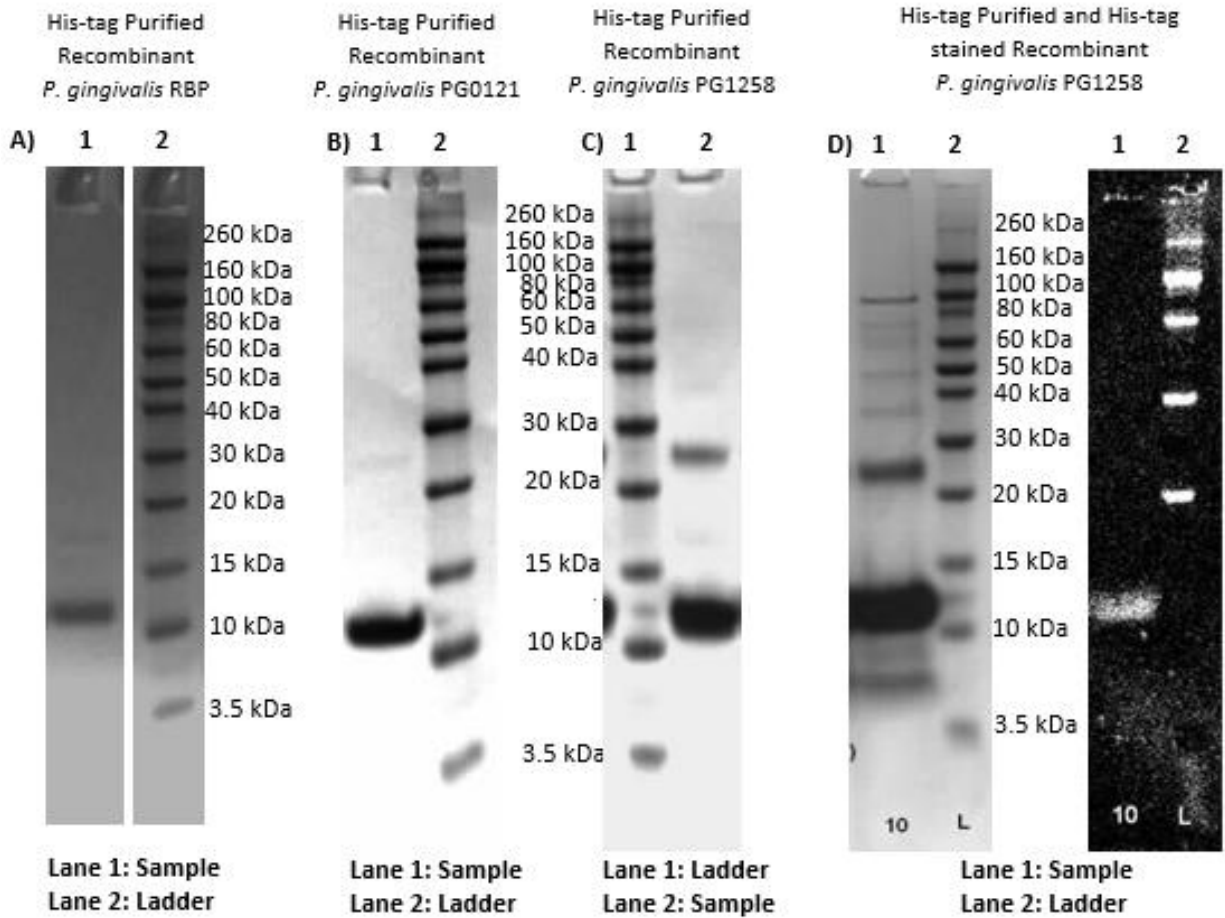
Figure 8 – Purification of Recombinant *P. gingivalis* HU and RBP

The elution for each of the His-tag purified proteins: RBP, PG0121 and PG1258 were

run on 12% Bis-tris denaturing gels with MES running buffer set at 155 Volts for 55

minutes. Each sample was loaded with 4X LDS Buffer (NuPAGE). A) Lane 1 contains

the His-tag purified Recombinant *P. gingivalis* RBP. Lane 2 contains pre-stained protein

ladder (Novex Sharp). B) Lane 1 contains the His-tag purified Recombinant *P. gingivalis*

PG0121. Lane 2 contains pre-stained protein ladder (Novex Sharp). C) Lane 1 contains

pre-stained protein ladder (Novex Sharp). Lane 2 contains the His-tag purified

Recombinant *P. gingivalis* PG1258 with two bands, a major and a minor at

approximately 12 and 25 kDa. D) Is an image of the same gel, before and after His-tag

stain containing the His-tag purified PG01258. The His-tag appears only in the major 12

kDa band.

4.3 – Size Exclusion Chromatography of Recombinant *P. gingivalis* HU and RBP

The purified recombinant *P. gingivalis* HU and RBP were concentrated using ammonium sulfate precipitation and then dialyzed in a Tris-Base buffer (27mM Tris-Base, 200mM NaCl, 1mM at pH=8) before running on an ÄKTA™ pure Fast Protein Liquid Chromatography machine using a Superdex 75 (10/300 GL). Molecular weight markers were also run on the column using the same buffer. An elution profile was generated (Figure 9) and used to create a standard calibration curve (Figure 10) based on the volume at which the markers eluted off the column (Table 4). The molecular weight of PG0121, PG1258 and RBP was calculated using both the standard calibration curve equation and the volume at which the proteins eluted off the column.

4.3.1 – His-Tag Purified Recombinant *P. gingivalis* RBP on Superdex 75

The purified His-tagged recombinant *P. gingivalis* RBP elution was dialyzed in a Tris-Base buffer and then run on the Superdex 75 column to generate an elution profile (Figure 9). At approximately 13.6mL the protein eluted. Using the standard calibration curve, the molecular weight of the protein was calculated to be approximately 5kDa in size (Table 4), suggesting that RBP in its native state is a monomer. The fraction collections from the column were run on a 12% Bis-Tris denaturing gel to confirm sizes and the protein was found to run at approximately 12 kDa (Figure 11).

4.3.2 – His-Tag Purified Recombinant HU PG0121 on Superdex 75

The purified His-tagged recombinant PG0121 elution was dialyzed in a Tris-Base buffer and then run on the Superdex 75 column to generate an elution profile (Figure 9).

Based on an average of five separate runs, the protein eluted at approximately 10.6mL (Table 4). Using the standard calibration curve, the molecular weight of the protein was calculated to be approximately 48.5kDa in size (Table 4), suggesting that the protein in its native state is a tetramer. The fraction collections from the column were run on a 12% Bis-Tris denaturing gel to confirm sizes and the protein was found to run at approximately 12 kDa (Figure 11).

4.3.3 – His-Tag Purified Recombinant HU PG1258 on Superdex 75

The purified His-tagged recombinant PG1258 elution was dialyzed in a Tris-Base buffer and then run on the Superdex 75 column to generate an elution profile (Figure 9). Based on the average of five separate runs, the protein eluted at approximately 10.7mL (Table 4). Using the standard calibration curve, the molecular weight of the protein was calculated to be approximately 46.2kDa in size (Table 4), suggesting that the protein in its native state is a tetramer. The fraction collections from the column were run on a 12% Bis-Tris denaturing gel to confirm sizes and the protein was found to run at approximately 12 kDa (Figure 11).

Blue Dextrane (2,000 kDa)

Albumin (66.5 kDa)

Carbonic Anhydrase (30 kDa)

Cytochrome C (12 kDa)

PG0121

PG1258

RBP

48
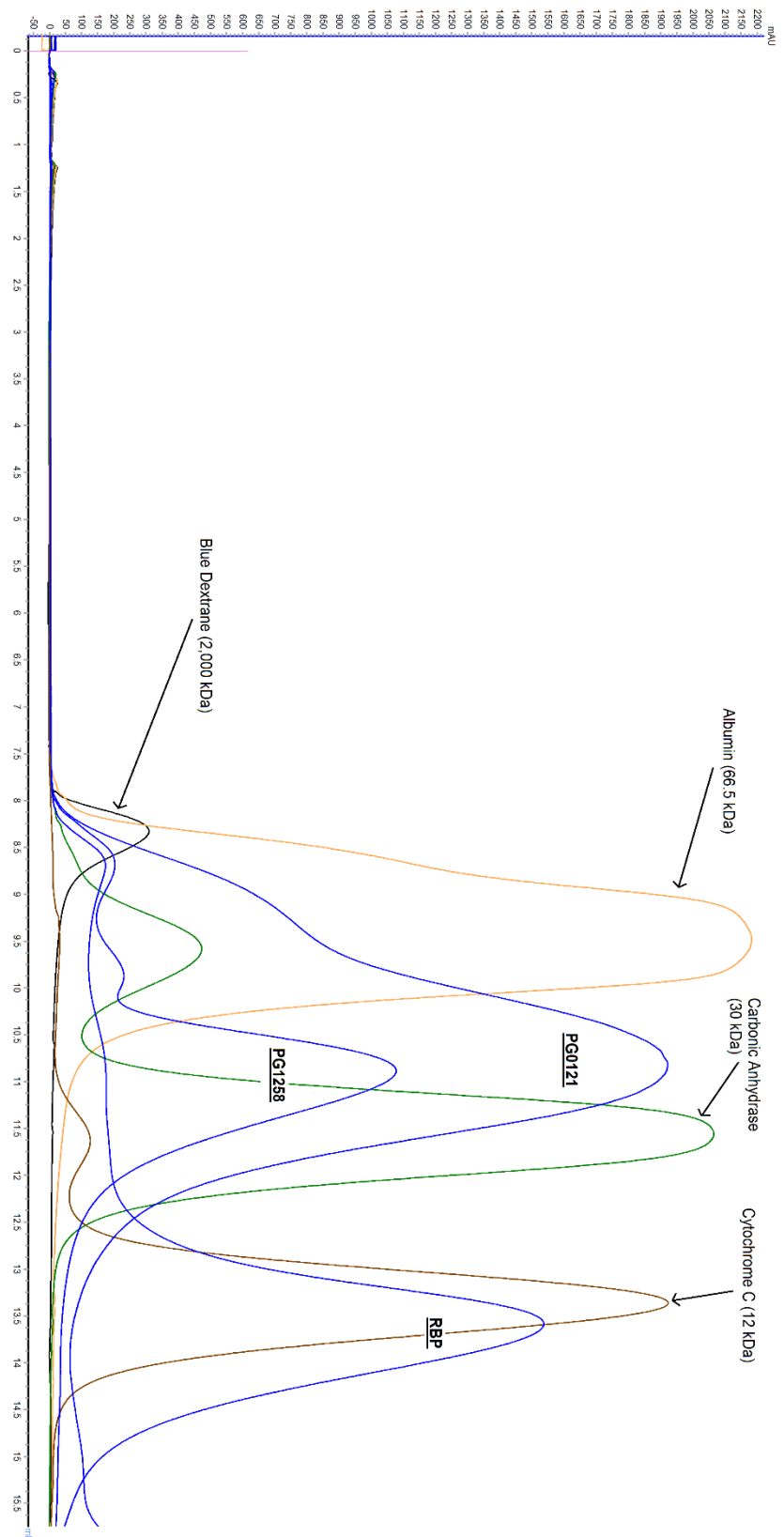
Figure 9 - Superdex 75 Elution Profiles for Recombinant *P. gingivalis* HU & RBP and Molecular Weight Standards

The elution profiles generated from the His-tag purified recombinant *P. gingivalis* HU and RBP run on a Superdex 75 column. The elution profiles for each of these purifications were overlaid for direct comparison. For the HU and RBP proteins, five separate elution profiles were generated (only one shown in graph) to determine an average elution volume for the respective subunits. HU PG0121 eluted on average at 10.6mL while HU PG1258 eluted on average at 10.8mL from the column. RBP PG0627 eluted from the column 13.24mL. The elution profiles were overlaid to determine the approximate molecular weights using the markers as reference points. Blue Dextran (2,000 kDa) was used to determine the void volume, shown in black. Albumin (66.5 kDa) shown in green and Cytochrome C shown brown (12 kDa), were used as the last two reference weights.
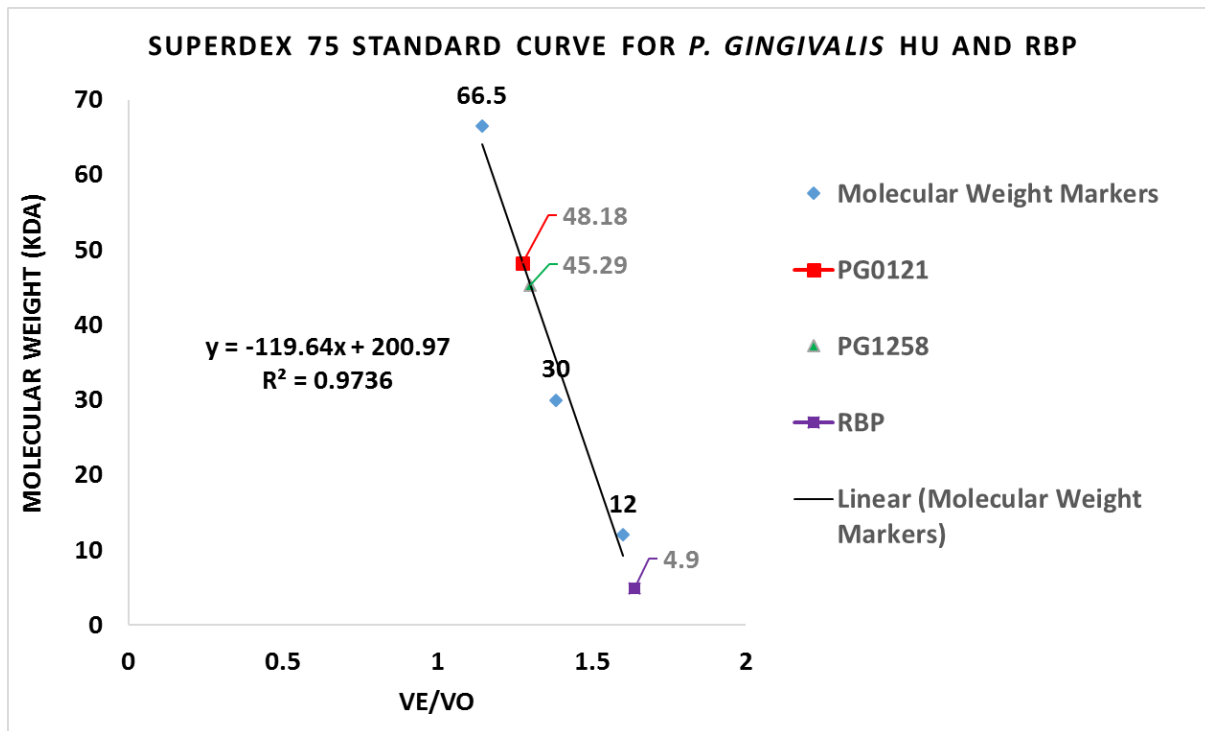
Figure 10 - Superdex 75 Standards Curve for Recombinant *P. gingivalis* HU & RBP

The line of best fit was generated from the molecular weight of each marker (y-axis) compared to the ratio of their respective elution volume (Ve) and void volume (Vo). The derived equation from the line of best fit was used to calculate the molecular weights of the recombinant *P. gingivalis* HU and RBP samples eluted from the Superdex 75 column. The representative HU subunit values, appear to form similar and higher molecular weight species compared to the representative RBP value, which appears to reside closer to a monomeric state.

Table 4 - Elution Volumes, Calculated Weights and Oligomeric Predictions for *P. gingivalis* HU and RBP

Standards	Elution Volume	Ve/Vo Value	kDa
Blue Dextran	8.3		2000
Albumin	9.5	1.145	66.5
Carbonic Anhydrase	11.5	1.386	30
Cytochrome C	13.3	1.602	12

Name	Elution Volume					Average Elution Volume	Average Ve/Vo	Calculated kDa	Oligomeric Prediction
	Run 1	Run 2	Run 3	Run 4	Run 5				
His-Tag Purified HUβ PG0121	10.2	9.9	10.7	10.9	11.2	10.58	1.275	48.465	4
His-Tag Purified HUα PG1258	11.2	10.4	10.8	11.3	10	10.74	1.294	46.158	4
His-Tag Purified RBP PG0627	13.4	13.6	13.5	13.1	12.6	13.24	1.595	10.113	1

51

Elution Fractions of
Recombinant
P. gingivalis HU and RBP

Lane 1: Ladder
Lane 2: PG0121
Lane 3: PG1258
Lane 4: RBP

Figure 11 - Gel Fractions of Recombinant *P. gingivalis* HU and RBP from

Superdex 75 Column

The elution fractions for each of the purified proteins off the Superdex 75 column were

run on 12% Bis-tris denaturing gel with MES running buffer set at 155 Volts for 55

minutes. Each sample was loaded with 4X LDS Buffer (NuPAGE). Lane 1 contains pre-

stained protein ladder (Novex Sharp). Lane 2 contains the elution fraction of

recombinant PG0121. Lane 3 contains the elution fraction of recombinant PG1258.

Lane 4 contains the elution fraction of the recombinant *P. gingivalis* RBP. All three

denatured proteins are approximately 12 kDa in size.

4.4 - Electrophoretic Mobility Shift Assay with Recombinant *P. gingivalis* HU and RBP

4.4.1 - Shift Assay with Recombinant *P. gingivalis* HU and RBP

For the uninhibited shift assays, increasing concentrations of purified

recombinant *P. gingivalis* HU and RBP (0-7nM) were incubated with just the specific,

fluorescently-tagged ssDNA of 48 nucleotides in length (adapted from Kamashev, et al

2007). A visible shift band appears for HU PG0121 at 0.5nM and increases in

fluorescence intensity as protein concentration increases (Figure 12A). Similarly, a

visible shift band appears for HU PG1258 at 2nM and increases in fluorescence

intensity as protein concentration increase (Figure 12B). For *P. gingivalis* RBP the shift

band appears at 0.5nM and increases in fluorescence intensity as protein concentration

increases (Figure 12C). On the other hand, the control lanes with no protein and only

labeled probe showed no shifting (Figure 12A-C). It is worth noting the small indication

on for *p. gingivalis* RBP was due to overflow (Figure 12C).

For the competitive inhibition studies, the protein concentrations were kept at a

consistent concentration of 2μm and the fluorescently labeled target sequence probe at

0.5nM. For the competitive inhibitors, increasing concentrations of both unlabeled target

sequence probe and the unlabeled random sequence probe were administered at

0.5nM, 5nM, and 25nM. For HU PG0121, there is no visible shift for the control lanes,

containing either no protein or the negative control HcpR (Figure 13). A prominent shift

is visible in the lane with only the labeled probe and protein (Figure 13). There is a

visible decrease in shift intensity observed for the three lanes containing increase

concentrations of unlabeled target sequence probe, demonstrating that the labeled

probe is being outcompeted (Figure 13). Likewise, there is a visible decrease in shift

intensity observed for the three lanes containing the unlabeled random sequence probe

(Figure 13).

For HU PG1258, there is no visible shift for the control lanes, containing either no

protein or the negative control HcpR (Figure 14). A prominent shift is visible in the lane

with only the labeled probe and protein (Figure 14). There is a visible decrease in shift

intensity observed for the three lanes containing increase concentrations of unlabeled

target sequence probe, demonstrating that the labeled probe is being outcompeted

(Figure 14). Likewise, there is a visible decrease in shift intensity observed for the three

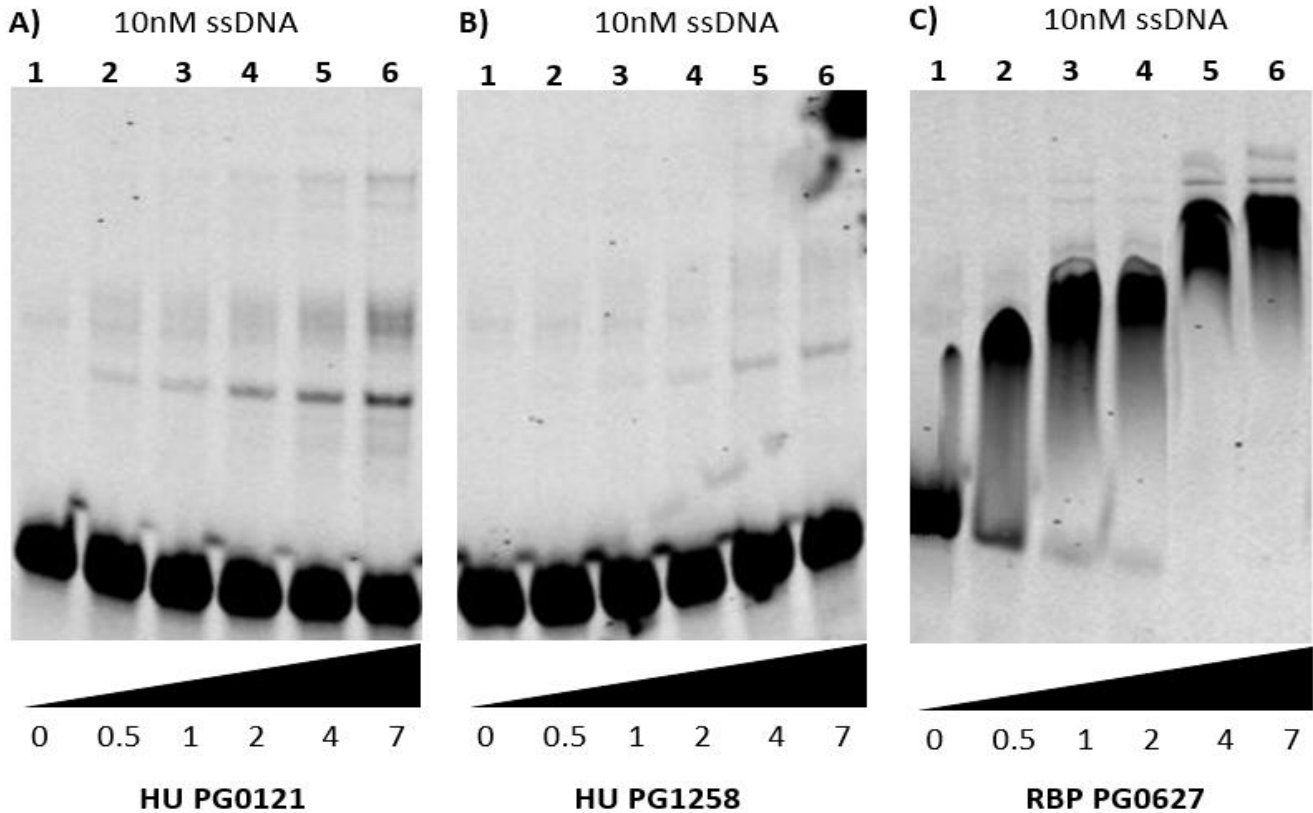lanes containing the unlabeled random sequence probe (Figure 14).

Figure 12 - Electrophoretic Mobility Shift Assays with Recombinant *P. gingivalis* HU and RBP

Purified recombinant *P. gingivalis* HU and RBP were tested for binding to a ssDNA oligonucleotide with a length of 48nt (adapted from Kamashev, et al 2007). The sequence was designed with a fluorescent tag. Each reaction was incubated for one hour in the dark and analyzed using electrophoresis on a 10% TBE Native Gel (Invitrogen) buffered with 27mM Tris-Borate, 0.1mM EDTA. The gel shifts were captured on an Odyssey® CLx imaging system (Li-COR). A) Increasing concentrations (0–7 nM) of PG0121 binding to 10nM ssDNA. B) Increasing concentrations (0–7 nM) of PG1258 binding to 10nM ssDNA. C) Increasing concentrations (0–7 nM) of RBP binding to 10nM ssDNA. The control lanes had 10nM ssDNA with no protein.

Figure 13 - Electrophoretic Mobility Shift Assays with Recombinant HU PG0121 and Competitive Inhibition.

Purified recombinant *P. gingivalis* HU PG0121 at a concentration of 2µM was tested for binding to a fluorescently tagged, target-sequence, ssDNA oligonucleotide at 2nM concentration and a length of 48nt. Two competitive inhibitors, an unlabeled target

sequence probe and an unlabeled random sequence probe, were added at increasing

concentrations of 0.5nM, 5nM, and 25nM. Two control lanes were run, one containing

no protein and the other, a negative control HcpR protein at a concentration of 2µM.

Each reaction was incubated for one hour in the dark and analyzed using

electrophoresis on a 10% TBE Native Gel (Invitrogen) buffered with 27mM Tris-Borate,

0.1mM EDTA. The gel shifts were captured on an Odyssey® CLx imaging system (Li-
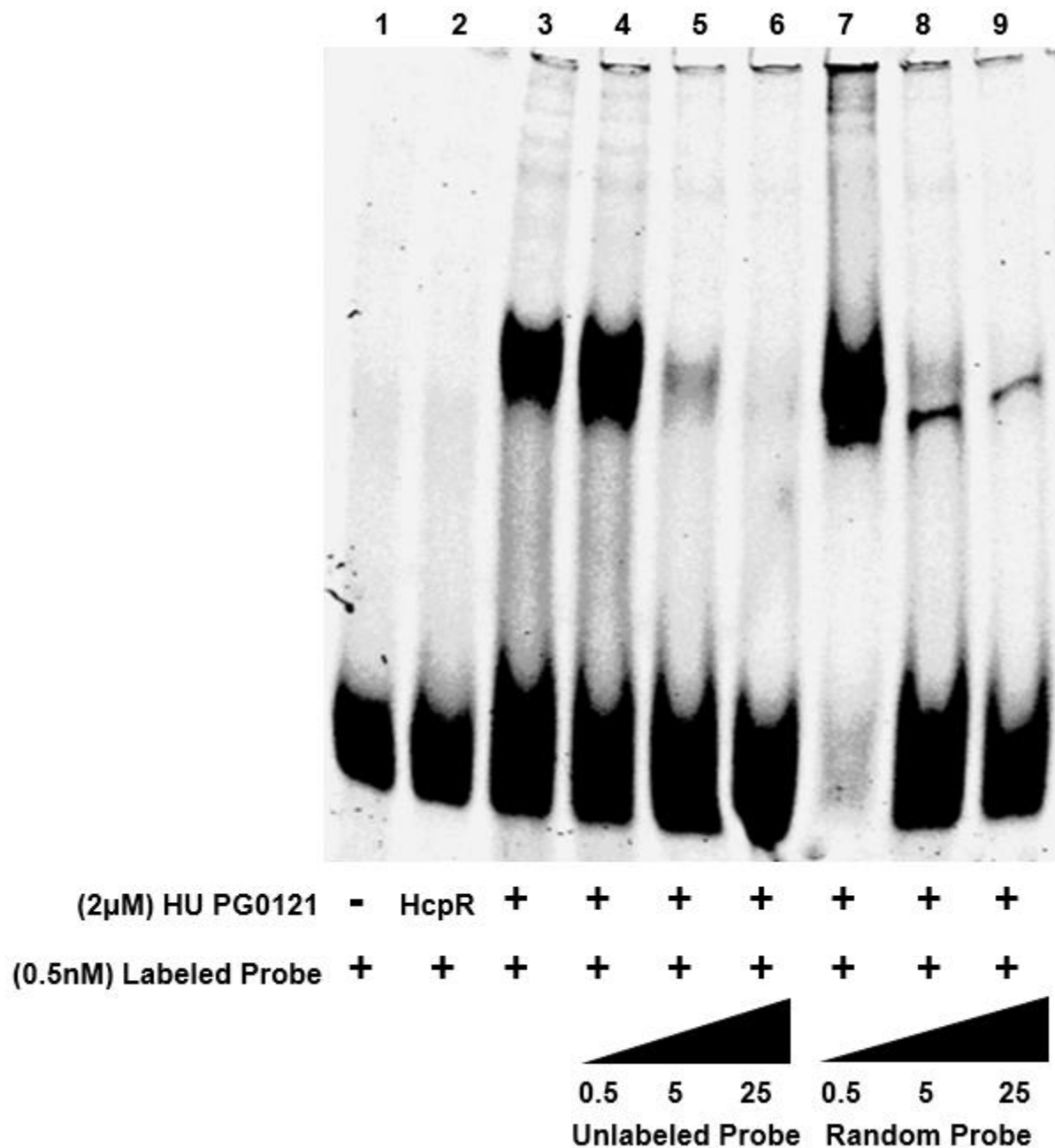
COR).

Figure 14 - Electrophoretic Mobility Shift Assays with Recombinant HU PG1258 and Competitive Inhibition.

Purified recombinant *P. gingivalis* HU PG1258 at a concentration of 2µM was tested for binding to a fluorescently tagged, target-sequence, ssDNA oligonucleotide at 2nM concentration and a length of 48nt. Two competitive inhibitors, an unlabeled target

sequence probe and an unlabeled random sequence probe, were added at increasing

concentrations of 0.5nM, 5nM, and 25nM. Two control lanes were run, one containing

no protein and the other, a negative control HcpR protein at a concentration of 2μM.

Each reaction was incubated for one hour in the dark and analyzed using

electrophoresis on a 10% TBE Native Gel (Invitrogen) buffered with 27mM Tris-Borate,

0.1mM EDTA. The gel shifts were captured on an Odyssey® CLx imaging system (Li-
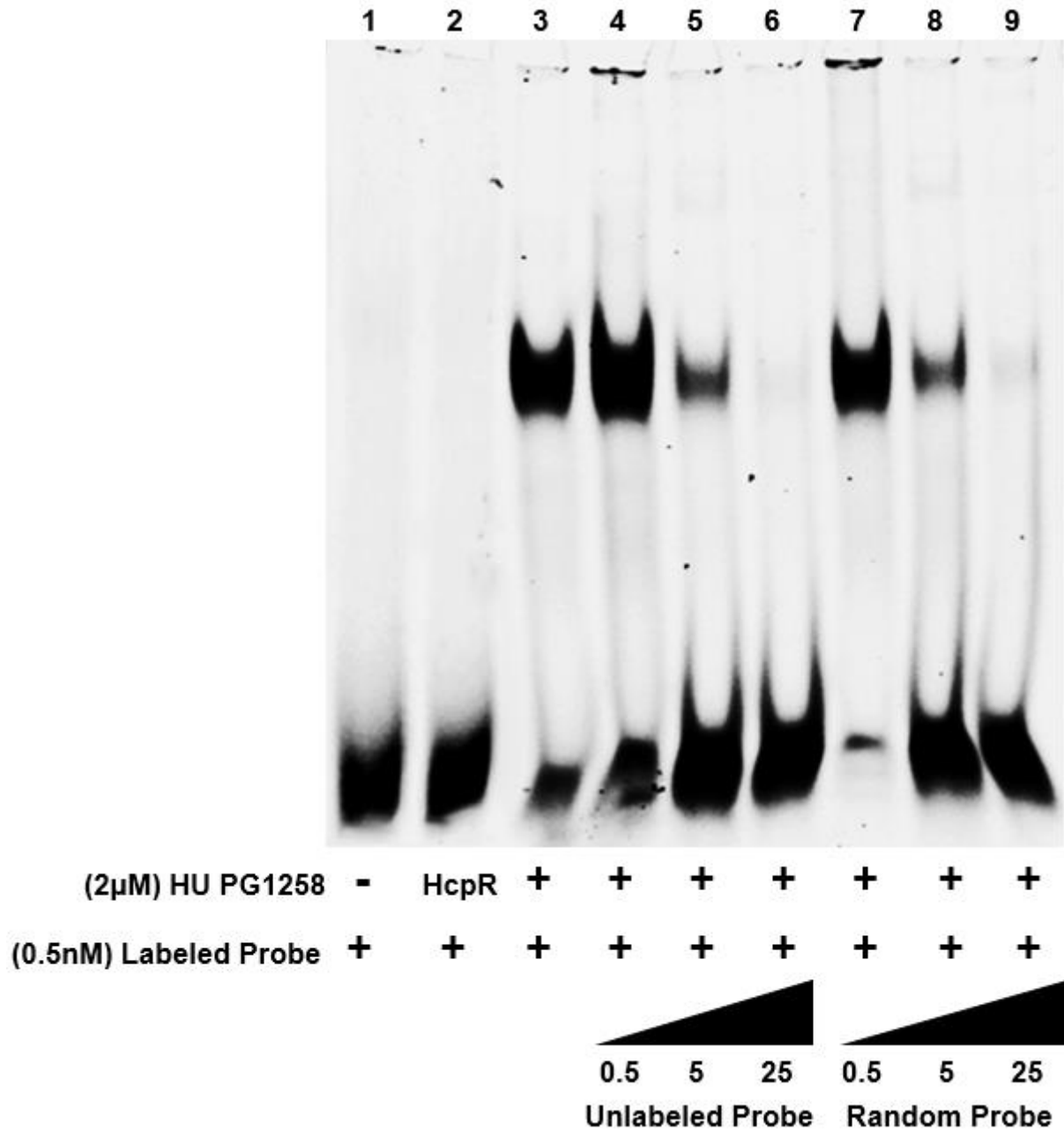
COR).

Chapter 5 – Discussion

The bioinformatic analysis showed that *Porphyromonas gingivalis* possess both subunits: HUα and HUβ encoded by PG1258 and PG0121 genes, respectively. Interestingly, the amino acid sequence comparison between the *P. gingivalis* HU subunits showed only a 29.55% identity were as the *E. coli* HU subunits shared a 68.89% identity (Figure 3C and 3D). This roughly 40% difference in sequence identity between *P. gingivalis* HU subunits from the *E. coli* HU subunits, supports the notion that the HU proteins in *P. gingivalis* favor a homo-tetramer state rather than the heterodimer state found in *E. coli.* This is also supported by literature, which has shown that the HU protein predominates as a homo-dimer or homo-tetramer during the stationary phase of cell growth in virtually all bacteria except enterobacteria, like *E. coli* which predominates (90%) as a heterodimer[20]. The reason or advantage for favoring a heterodimer state only in enterobacteria, while favoring the homodimer or homo-tetramer state in virtually all other bacteria examined, remains elusive[20].

Nevertheless, the *P. gingivalis* HU subunit homology models, derived from the *E. coli* HU crystal structures, confirms a high degree of structural conservation, both in terms of the HU 'body' and β-ribbon 'arms' that protrude to the side (Figure 5A and 5B). This conservation in geometry and positive electrostatic potential of these individual arms suggests that the *P. gingivalis* HU is capable of binding to ssDNA in a similar manner as that already been shown in *E. coli* HU. Moreover, the multi-sequence alignments in Figure 4, indicate that the histone-like DNA binding site found between the 44th and 65th residue for all four proteins is mutually conserved alluding to the potential for similar physiochemical properties.

For the *P. gingivalis* RNA-binding protein, which is 97 amino acids in length we were able to produce a three-dimensional crystal structure of the protein (Figure 6). The structure confirmed that RBP predominates as a monomer and is comprised of a highly-conserved RNA recognition motif (RRM). This protein consists of four anti-parallel beta-strands and two alpha-helices arranged in a β_1-α_1-β_2-β_3-α_2-β_4 manner which is consistent with the RRM found throughout the entire kingdom of life, including viruses (Figure 6)[53]. It has been shown in previous literature, that this conserved protein domain is capable of modulating its structure and folds to recognize many RNAs, DNAs, and proteins in order to achieve a multitude of biological functions[53].

In order to determine if there were any aberrant structural differences between the *E. coli* HU heterodimer and the *P. gingivalis* HU homodimers, the total solvent excluded surface area of the interface between each dimer was compared. For the *E. coli* HU heterodimer, the area was determined to be 996.5 Å^2 while the area for the PG1258 HUβ homodimer and PG0121 HUα homodimer was determined to be 656.2 Å^2 and 1,220.1 Å^2, respectively. This indicates that proteins structurally align differently to each other but that the PG0121 HUα homodimer is most similar to *E. coli* HU heterodimer structure.

For the His-tag purification, the gene products from PG0121, PG1258, and PG0627, were purified using Ni-NTA resin via the His-tag purification system. The purified HU PG0121 and the purified HU PG1258 with their 6x His-tag appear at approximately 12 kDa, as expected (Figure 8B and 8C). However, the purification of HU PG1258 resulted in the purification of two separate bands: a less expressed band at approximately 25 kDa and a more expressed band at approximately 12 kDa. The 6x His

Protein Tag Stain kit was used to confirm the size and location of the protein and although the His-tag stain only detected the presence of His-tag at the more expressed band at approximately 12 kDa, there remains a certain degree of uncertainty as to whether or not the stain was sensitive enough to detect the less expressed band at 25 kDa (Figure 8D). Another reason for uncertainty, is the less expressed band at 25 kDa, is approximately double the size of the 12,320.13 Da monomer, suggesting that the second band is not a contaminant, but instead a dimer or aggregate of the purified protein at quantities too low to be detected by the His-tag stain.

The elution profiles for each protein shows that the HU subunits are very similar in size, both predominating as a homo-tetramers as indicated by their largest peaks and their average elution volumes (Figure 9, 10 and Table 4). However, it is also worth noting that the elution profile for both HU subunits also had a smaller, second peak, suggesting the presence of a less prominent homodimer, as well. Likely, the proteins are in equilibrium between the homo-tetramer and homodimer form. Further studies, including small-angle X-ray scattering (SAXS) analysis or analytical ultracentrifugation can be used to validate the oligomeric state of these proteins.

For the uninhibited shift assays, increasing concentrations of recombinant *P. gingivalis* HU and RBP (0-7nM) demonstrated visible shifting of the ssDNA oligonucleotide with increasing intensity. For HU PG0121, the first visible shift was observable at a concentration of 0.5nM, were as for HU PG1258 the first visible shift was observable at a concentration of 2nM, suggesting that HU PG0121 has a greater affinity for the ssDNA oligonucleotide than HU PG1258 (Figure 12A and 12B). For *P. gingivalis* RBP the shift band was first observable at a concentration of 0.5nM and was

much stronger in intensity, resulting in the complete shifting of all 10μM fluorescent probe, suggesting the RBP has the greatest affinity for ssDNA binding of the three proteins (Figure 12C). It is worth noting that the small indentation of shifting occurring in the no protein lane was due to overflow (Figure 12C).

For the competitive inhibition studies, HU PG0121 demonstrated clear and prominent shifts that also decreased in intensity as the unlabeled probe was added, further confirming the protein indeed binds to ssDNA (Figure 13). Additionally, a random sequence probe was used in increasing concentrations, which out-competed the labeled probe and decreased shift-band intensity, suggesting HU PG0121 is non-specific binding (Figure 13). However, when comparing the band intensity between the unlabeled and random probe at 5nM and 25nM concentrations, the unlabeled appears to outcompete more effectively and produce a less intense band, suggesting that while HU PG0121 is a non-specific binding protein it does show preferential binding to the labeled-target sequence (Figure 13). Likewise, HU PG1258 also demonstrated clear and prominent shifts that decreased in intensity as the unlabeled probe was added, further confirming the protein binds to ssDNA (Figure 14). Additionally, a random sequence probe was used in increasing concentrations which out-competed the labeled probe, decreasing shift-band intensity, suggesting HU PG1258 is non-specific binding (Figure 14). However, when comparing the band intensity between the unlabeled and random probe at 25nM concentrations, the unlabeled appears to outcompete more effectively and produce a less intense band, suggesting that while HU PG1258 is a non-specific binding protein it does show preferential binding to the labeled-target sequence. Furthermore, when comparing binding characteristics of HU PG0121 and HU PG1258

in terms of the shift-band intensity for the random sequence probe at 5nM and 25nM

concentrations, HU PG0121 is clearly more intense, suggesting once more that HU

PG0121 has a greater preference for the target sequence ssDNA than PG1258. This is

also substantiated in literature, as several studies on the *E. coli* HU protein have also

found that while it binds non-specifically, it does show strong preferential binding to G/C

rich sequences, in both dsDNA and ssDNA[18,32].

Chapter 6 – Conclusion

We were able to determine the oligomeric states of the *Porphyromonas gingivalis* HUα subunits: HUα and HUβ, which are encoded by PG1258 and PG0121 genes respectively, as well as the oligomeric state of the *P. gingivalis* RNA-binding protein encoded by PG0627. Our studies show that the *P. gingivalis* HU proteins are multimeric but predominate as a homo-tetramer and that the *P. gingivalis* RBP exists exclusively as a monomer. We also showed that the *P. gingivalis* HU proteins bind non-specifically to single-stranded DNA with a strong preference for poly(dG) content, while binding to poly(dA) the weakest, which has not been shown before. Also, our results suggest that HU PG0121 has a greater affinity for single-stranded DNA than HU PG1258. These results show that *P. gingivalis* HU and RBP are novel ssDNA binding proteins in *P. gingivalis,* indicating an expanded role and function within the cell.

Bibliography

1. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol*. 2005;43(11):5721-5732. doi: 43/11/5721 [pii].

2. Samaranayake L. Normal oral flora, the oral ecosystem and plaque biofilm. In: *Essential microbiology for dentistry.* Philadelphia: Elsevier; 2006:255-266.

3. Parahitiyawa NB, Jin LJ, Leung WK, Yam WC, Samaranayake LP. Microbiology of odontogenic bacteremia: Beyond endocarditis. *Clin Microbiol Rev*. 2009;22(1):46-64, Table of Contents. doi: 10.1128/CMR.00028-08 [doi].

4. O'Toole G, Kaplan HB, Kolter R. Biofilm formation as microbial development. *Annu Rev Microbiol*. 2000;54:49-79. doi: 10.1146/annurev.micro.54.1.49 [doi].

5. Socransky SS, Haffajee AD. The bacterial etiology of destructive periodontal disease: Current concepts. *J Periodontol*. 1992;63(4 Suppl):322-331. doi: 10.1902/jop.1992.63.4s.322 [doi].

6. Hajishengallis G, Darveau RP, Curtis MA. The keystone-pathogen hypothesis. *Nat Rev Microbiol*. 2012;10(10):717-725. doi: 10.1038/nrmicro2873 [doi].

7. Lamont RJ, Jenkinson HF. Subgingival colonization by porphyromonas gingivalis. *Oral Microbiol Immunol*. 2000;15(6):341-349. doi: omi150601 [pii].

8. Duncan MJ, Nakao S, Skobe Z, Xie H. Interactions of porphyromonas gingivalis with epithelial cells. *Infect Immun*. 1993;61(5):2260-2265.

9. Holt SC, Kesavalu L, Walker S, Genco CA. Virulence factors of porphyromonas gingivalis. *Periodontol 2000*. 1999;20:168-238.

10. Lamont RJ, Yilmaz O. In or out: The invasiveness of oral bacteria. *Periodontol 2000*. 2002;30:61-69. doi: prd3006 [pii].

11. Li Y, Guo H, Wang X, Lu Y, Yang C, Yang P. Coinfection with fusobacterium nucleatum can enhance the attachment and invasion of porphyromonas gingivalis or aggregatibacter actinomycetemcomitans to human gingival epithelial cells. *Arch Oral Biol.* 2015;60(9):1387-1393. doi: 10.1016/j.archoralbio.2015.06.017 [doi].

12. Njoroge T, Genco RJ, Sojar HT, Hamada N, Genco CA. A role for fimbriae in porphyromonas gingivalis invasion of oral epithelial cells. *Infect Immun*. 1997;65(5):1980-1984.

13. Madianos PN, Papapanou PN, Nannmark U, Dahlen G, Sandros J. Porphyromonas gingivalis FDC381 multiplies and persists within human oral epithelial cells in vitro. *Infect Immun*. 1996;64(2):660-664.

14. Hosogi Y, Duncan MJ. Gene expression in porphyromonas gingivalis after contact with human epithelial cells. *Infect Immun*. 2005;73(4):2327-2335. doi: 73/4/2327 [pii].

15. Tang G, Yip HK, Cutress TW, Samaranayake LP. Artificial mouth model systems and their contribution to caries research: A review. *J Dent*. 2003;31(3):161-171. doi: S0300571203000095 [pii].

16. Walsh CT, Garneau-Tsodikova S, Gatto GJ,Jr. Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew Chem Int Ed Engl*. 2005;44(45):7342-7372. doi: 10.1002/anie.200501023 [doi].

17. Jablonka E, Raz G. Transgenerational epigenetic inheritance: Prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev Biol*. 2009;84(2):131-176.

18. Kamashev D, Balandina A, Mazur AK, Arimondo PB, Rouviere-Yaniv J. HU binds and folds single-stranded DNA. *Nucleic Acids Res*. 2008;36(3):1026-1036. doi: gkm667 [pii].

19. Dorman CJ, Deighan P. Regulation of gene expression by histone-like proteins in bacteria. *Curr Opin Genet Dev*. 2003;13(2):179-184. doi: S0959437X0300025X [pii].

20. Oberto J, Drlica K, Rouviere-Yaniv J. Histones, HMG, HU, IHF: Meme combat. *Biochimie*. 1994;76(10-11):901-908. doi: 0300-9084(94)90014-0 [pii].

21. Drlica K, Rouviere-Yaniv J. Histonelike proteins of bacteria. *Microbiol Rev*. 1987;51(3):301-319.

22. Pettijohn DE. Histone-like proteins and bacterial chromosome structure. *J Biol Chem*. 1988;263(26):12793-12796.

23. Claret L, Rouviere-Yaniv J. Regulation of HU alpha and HU beta by CRP and FIS in escherichia coli. *J Mol Biol*. 1996;263(2):126-139. doi: S0022-2836(96)90564-1 [pii].

24. Huisman O, Faelen M, Girard D, Jaffe A, Toussaint A, Rouviere-Yaniv J. Multiple defects in escherichia coli mutants lacking HU protein. *J Bacteriol*. 1989;171(7):3704-3712.

25. Dri AM, Rouviere-Yaniv J, Moreau PL. Inhibition of cell division in hupA hupB mutant bacteria lacking HU protein. *J Bacteriol*. 1991;173(9):2852-2863.

26. Fernandez S, Rojo F, Alonso JC. The bacillus subtilis chromatin-associated protein hbsu is involved in DNA repair and recombination. *Mol Microbiol*. 1997;23(6):1169-1179.

27. Grove A. Functional evolution of bacterial histone-like HU proteins. *Curr Issues Mol Biol*. 2011;13(1):1-12. doi: v13/1 [pii].

28. Tanaka I, Appelt K, Dijk J, White SW, Wilson KS. 3-A resolution structure of a protein with histone-like properties in prokaryotes. *Nature*. 1984;310(5976):376-381.

29. Christodoulou E, Vorgias CE. Cloning, overproduction, purification and crystallization of the DNA binding protein HU from the hyperthermophilic eubacterium thermotoga maritima. *Acta Crystallogr D Biol Crystallogr*. 1998;54(Pt 5):1043-1045.

30. White SW, Appelt K, Wilson KS, Tanaka I. A protein structural motif that bends DNA. *Proteins*. 1989;5(4):281-288. doi: 10.1002/prot.340050405 [doi].

31. Pinson V, Takahashi M, Rouviere-Yaniv J. Differential binding of the escherichia coli HU, homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA. *J Mol Biol*. 1999;287(3):485-497. doi: S0022-2836(99)92631-1 [pii].

32. Krylov AS, Zasedateleva OA, Prokopenko DV, Rouviere-Yaniv J, Mirzabekov AD. Massive parallel analysis of the binding specificity of histone-like protein HU to single- and double-stranded DNA with generic oligodeoxyribonucleotide microchips. *Nucleic Acids Res*. 2001;29(12):2654-2660.

33. Dri AM, Moreau PL, Rouviere-Yaniv J. Role of the histone-like proteins OsmZ and HU in homologous recombination. *Gene*. 1992;120(1):11-16.

34. Hwang DS, Kornberg A. Opening of the replication origin of escherichia coli by DnaA protein with protein HU or IHF. *J Biol Chem*. 1992;267(32):23083-23086.

35. Kohno K, Wada M, Kano Y, Imamoto F. Promoters and autogenous control of the escherichia coli hupA and hupB genes. *J Mol Biol*. 1990;213(1):27-36. doi: S0022-2836(05)80119-6 [pii].

36. Oberto J, Nabti S, Jooste V, Mignot H, Rouviere-Yaniv J. The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction. *PLoS One*. 2009;4(2):e4367. doi: 10.1371/journal.pone.0004367 [doi].

37. Kamashev D, Rouviere-Yaniv J. The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *EMBO J*. 2000;19(23):6527-6535. doi: 10.1093/emboj/19.23.6527 [doi].

38. Bensaid A, Almeida A, Drlica K, Rouviere-Yaniv J. Cross-talk between topoisomerase I and HU in escherichia coli. *J Mol Biol*. 1996;256(2):292-300. doi: S0022-2836(96)90086-8 [pii].

39. Rouviere-Yaniv J, Yaniv M, Germond JE. E. coli DNA binding protein HU forms nucleosomelike structure with circular double-stranded DNA. *Cell.* 1979;17(2):265-274. doi: 0092-8674(79)90152-1 [pii].

40. Gellert M, Fisher LM, Ohmori H, O'Dea MH, Mizuuchi K. DNA gyrase: Site-specific interactions and transient double-strand breakage of DNA. *Cold Spring Harb Symp Quant Biol.* 1981;45 Pt 1:391-398.

41. Kohno K, Yasuzawa K, Hirose M, et al. Autoregulation of transcription of the hupA gene in escherichia coli: Evidence for steric hindrance of the functional promoter domains induced by HU. *J Biochem.* 1994;115(6):1113-1118.

42. Rice PA. Making DNA do a U-turn: IHF and related proteins. *Curr Opin Struct Biol.* 1997;7(1):86-93. doi: S0959-440X(97)80011-5 [pii].

43. Flashner Y, Gralla JD. DNA dynamic flexibility and protein recognition: Differential stimulation by bacterial histone-like protein HU. *Cell.* 1988;54(5):713-721. doi: S0092-8674(88)80016-3 [pii].

44. Kano Y, Imamoto F. Requirement of integration host factor (IHF) for growth of escherichia coli deficient in HU protein. *Gene.* 1990;89(1):133-137.

45. Bi H, Sun L, Fukamachi T, Saito H, Kobayashi H. HU participates in expression of a specific set of genes required for growth and survival at acidic pH in escherichia coli. *Curr Microbiol.* 2009;58(5):443-448. doi: 10.1007/s00284-008-9340-4 [doi].

46. Boubrik F, Rouviere-Yaniv J. Increased sensitivity to gamma irradiation in bacteria lacking protein HU. *Proc Natl Acad Sci U S A.* 1995;92(9):3958-3962.

47. Giangrossi M, Giuliodori AM, Gualerzi CO, Pon CL. Selective expression of the beta-subunit of nucleoid-associated protein HU during cold shock in escherichia coli. *Mol Microbiol.* 2002;44(1):205-216. doi: 2868 [pii].

48. Alberti-Segui C, Arndt A, Cugini C, Priyadarshini R, Davey ME. HU protein affects transcription of surface polysaccharide synthesis genes in porphyromonas gingivalis. *J Bacteriol.* 2010;192(23):6217-6229. doi: 10.1128/JB.00106-10 [doi].

49. Tjokro NO, Rocco CJ, Priyadarshini R, Davey ME, Goodman SD. A biochemical analysis of the interaction of porphyromonas gingivalis HU PG0121 protein with DNA. *PLoS One.* 2014;9(3):e93266. doi: 10.1371/journal.pone.0093266 [doi].

50. Balandina A, Claret L, Hengge-Aronis R, Rouviere-Yaniv J. The escherichia coli histone-like protein HU regulates rpoS translation. *Mol Microbiol.* 2001;39(4):1069-1079. doi: mmi2305 [pii].

51. Balandina A, Kamashev D, Rouviere-Yaniv J. The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids. *J Biol Chem.* 2002;277(31):27622-27628. doi: 10.1074/jbc.M201978200 [doi].

52. Murina VN, Nikulin AD. RNA-binding sm-like proteins of bacteria and archaea. similarity and difference in structure and function. *Biochemistry (Mosc).* 2011;76(13):1434-1449. doi: 10.1134/S0006297911130050 [doi].

53. Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* 2005;272(9):2118-2131. doi: EJB4653 [pii].

54. Dreyfuss G, Swanson MS, Pinol-Roma S. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci.* 1988;13(3):86-91.

55. Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. *Science.* 1994;265(5172):615-621.

56. Van Assche E, Van Puyvelde S, Vanderleyden J, Steenackers HP. RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Front Microbiol.* 2015;6:141. doi: 10.3389/fmicb.2015.00141 [doi].

57. McKenzie RM, Johnson NA, Aruni W, Dou Y, Masinde G, Fletcher HM. Differential response of porphyromonas gingivalis to varying levels and duration of hydrogen peroxide-induced oxidative stress. *Microbiology.* 2012;158(Pt 10):2465-2479. doi: mic.0.056416-0 [pii].

58. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol.* 2011;7:539. doi: 10.1038/msb.2011.75 [doi].

59. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10(6):845-858. doi: 10.1038/nprot.2015.053 [doi].

60. Pettersen EF, Goddard TD, Huang CC, et al. UCSF chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612. doi: 10.1002/jcc.20084 [doi].

61. Wilkins MR, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*. 1999;112:531-552.

Vita

Steve Harry Kokorelis was born on November 11, 1991, in Mineola, New York. He graduated from Midlothian High School in 2010. Steve then attended James Madison University where he graduated with a Bachelor's of Science in Biology in May 2014. Following his graduation from this program, he will attend the University of Maryland School of Dentistry as part of the class of 2021.