An investigation into the functional analysis of BPIFA2 in human saliva

Lin Zhang

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Clinical Dentistry The University of Sheffield

March 2024

Acknowledgments

Firstly, I would like to thank my best supervisor Dr Lynne Bingle for her endless guidance, help and encouragement, as well as her exceptional support during the Covid pandemic. She has always been available to chat about everything not just academic work and she has always known when to cheer me up and how to calm me down. She has helped me become an independent and insightful researcher, and I am truly grateful for her supervision. She is my role model, both academically and in other aspects of life.

I would also like to thank Professor Colin Bingle for his supervision throughout my PhD. I have gained valuable insights from his extensive expertise and considerable experience. I am extremely thankful for his assistance and guidance in exploring the BPIFA2 protein through the application of bioinformatic skills.

Thank you to the technical staffs: Mrs Brenka McCabe, Mr Jason Heath and Mrs Lisa Lijun Chang in dental school who helped me to learn lab techniques and stress relieving chats. And thank you so much to all the lovely colleagues for their kind words and encouragement (Wenyi, Anita, Luna, Basma, Asma, Mariem, Rawan, Cony). I would like to express my gratitude to the Bingle group for their assistance, support, and the enjoyable experiences I had while pursuing my studies (Esra, Fatima, Isabel, Vivian, Miraj, Abdulaziz). A huge thank you to all the generous donors who have provided saliva samples to contribute to my project.

I would like to give a huge thank you to Miss Jiayi Li who encouraged me over the years. She is well-informed about me, and we always share the latest updates with each other. She has always known how to encourage me, like she said, “Just like me running a marathon, I must complete the race.” Thank you for motivating me to finish the lab work and write my thesis during the days when you were enjoying eating, drinking, and having fun.

I would also like to thank Miss Ling Huang and Dr. Yajun Cai for their support throughout my PhD life. Although we operate in diverse fields, we consistently exchange knowledge, inspire, and encourage each other. I deeply appreciate your presence in my life. Expressing gratitude!

And finally, I would deeply be appreciative of my best lovely parents, for their financial support to my PhD degree, unlimited love and encouragement. I could not accomplish this without you! I am sorry to absent all the birthdays, the festivals of family union and my grandma’s funeral over years and have not been back to China in over two years and six months.

Abstract

Introduction

BPIF-containing proteins are members of the BPI-LBP-CETP-PLTP protein family. Previous studies have demonstrated that the most significant diversity exists in the BPIFA branch of the family, and of significance to this project, proteins associated with rodent parotid secretory proteins (PSP) are extremely divergent. However, compelling functional data remains to be elucidated. BPIFA2 orthologues are highly divergent between species, with the human/mouse proteins being 32% identical and some species having evolved a number of BPIFA2 paralogs. BPIFA2, is a glycosylated secretory protein that is highly expressed in mammalian salivary glands and secreted into saliva. As BPIFA2 is structurally related to the LPS binding proteins, BPI and LBP, it has been hypothesised the protein will play an important role in the innate immune defence of the oral cavity (Bingle and Craven, 2004; Bingle *et al.*, 2009). The principal aim of my study, therefore, is to determine the function of BPIFA2 in human saliva through a comprehensive bioinformatic analysis and to compare function with BPIFA2 from multiple mammals across the phylogenetic tree.

Materials and methods

BPIFA2 expression in whole human saliva was determined by Western blot and dot blot. Recombinant human BPIFA2 proteins, with and without N-glycosylation, and five animal species were synthesised by gene cloning, calcium phosphate transfection into HEK293 cells and quantified by BCA, ELISA and dot blot assays. A series of bacterial assays, including pull down, agglutination and biofilm formation assays, were used to determine the interaction between BPIFA2 proteins and the bacterial strains *S. gordonii, S. mutans*, *P. aeruginosa*, *E. coli* and *S. aureus.* Lipid strips were employed to investigate the interaction of BPIFA2 and specific lipids and an opsonisation assay was carried out to determine if bacterial strains coated with BPIFA2 were phagocytosed by MM6 cells. A comprehensive bioinformatics analysis of BPIFA2 across mammalian species by searching online platforms and databases was also undertaken.

Results

BPIFA2 was differentially expressed in whole, human saliva from a number of volunteers. Recombinant proteins were successfully produced by transfection, both the wild type and the mutant variants of recombinant human BPIFA2 demonstrated significant binding capacity with *S. gordonii, S. mutans*, *P. aeruginosa*, *E. coli* and *S. aureus.* Both human and animal recombinant BPIFA2 induce agglutination of *P. aeruginosa* and *E. coli,* but no agglutination was observed with *S. gordonii, S. mutans* or *S. aureus.* Human and animal BPIFA2 demonstrated varying levels of inhibition of biofilm formation with the same bacterial species. Recombinant BPIFA2 from the macaque showed binding to the same lipids as human WT BPIFA2, but no positive binding was detected with armadillo, mouse, squirrel monkey or dog BPIFA2; interestingly mutated forms of human BPIFA2, with glycosylation sites removed, also did not bind lipids. All bacterial strains were opsonised by human and animal BPIFA2 proteins. Bioinformatics results suggest that the function of BPIF-containing proteins may have changed with evolution as humans produce different BPIF-containing proteins and/or different quantities of these proteins in comparison to our nearest relatives (gorilla and chimpanzee). Multiple forms of BPIFA2 have been identified in ruminants with distinct dietary preferences.

Conclusions

This thesis has comprehensively explored BPIFA2, from its evolutionary conservation to its functional roles. BPIFA2 in mammals has diverged with evolutionary development and dietary changes. The host defence role of BPIFA2 in humans is not conclusive. The results indicated that BPIFA2 may be involved in interactions related to sensing and nutrition, rather than host defence. The true function of BPIFA2 requires further investigation.

Contents

List of figures

List of tables

List of abbreviations

AEP	Acquired enamel pellicle
AIDS	Acquired immunodeficiency syndrome
AKI	Acute kidney injury
BASE	Breast cancer and salivary gland expression
BCA	Bicinchoninic acid
BHI	Brain-Heart Infusion
BLAT	BLAST-like alignment tool
BLAST	Basic local alignment search tool
BPI	Bactericidal permeability increasing
BPIFA1	BPI-fold containing family protein A1
BPIFA2	BPI-fold containing family protein A2
BPIFA3	BPI-fold containing family protein A3
BPIFA4	BPI-fold containing family protein A4
BPIFA5	BPI-fold containing family protein A5
BPIFB1	BPI-fold containing family protein B1
BSP	Bovine salivary protein
CACTUS	Comparative algorithm for chaining transcripts using splice sites
CD14	Cluster of differentiation 14
cDNA	Complementary DNA
CETP	Cholesteryl ester-transfer protein
CFUs	Colony-forming units
cGVHD	Chronic graft-versus-host disease
CMV	Cytomegalovirus
COPD	Chronic obstructive pulmonary disease
CRSwNP	Chronic rhinosinusitis with nasal polyps
DMBT1	Deleted in malignant brain tumours 1
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ESTs	Expressed Sequence Tags
FASTA	Fast alignment
FBS	Foetal bovine serum
GFP	Green fluorescent protein

HEK 293	Human embryonic kidney 293 cells
HIV	Human immunodeficiency viruses
HPA	Human Protein Atlas
HCT	Hematopoietic stem cell transplantation
KO	Knock-Out
LBP	Lipopolysaccharide-binding protein
LPS	Lipopolysaccharide
MD-2	Myeloid differential protein 2
MIC	Minimum Inhibitory Concentration
MM6	Mono Mac 6
NCBI	National Center for Biotechnology Information
NEAA	Essential amino acids
NPC	Nasopharyngeal carcinoma
OD	Optical density
OPI	Oxaloacetate pyruvate insulin. Pathogen-
PAMPs	associated molecular patterns
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PLTP	Phospholipid-transfer protein
PG	Peptidoglycan
PLUNC	Palate Lung Nasal Clone
PSP	Parotid secretory protein
qPCR	Quantitative polymerase chain reaction
SDS	Sodium dodecyl sulfate
RPMI	Roswell Park Memorial Institute
sIgA and G	Secretory IgA and IgG
SOB	Super Optimal Broth
SMGB	Submandibular gland protein B
SNPs	Single nucleotide polymorphisms
TAE	Tris-acetate-EDTA buffer
TBS	Tris-buffered saline
TLR4	Toll-like receptor 4
TNF-α	TNF-α Tumour necrosis factor alpha
TOPO cloning	Topoisomerase based cloning
TULIP	Tubular-lipid binding protein
WT	Wild-type

Chapter 1: Literature Review

1.1 Oral cavity, salivary glands and saliva

1.1.1 Oral cavity

The oral cavity, commonly referred to as the mouth, is an important anatomical structure located at the beginning of the digestive system (Atkinson and White, 1992). It serves as the entry point for food intake, facilitating mastication, swallowing and initial digestion. Also, the oral cavity is lined by a specialised mucous membrane, the oral mucosa (Laine and Smoker, 1996), which is richly supplied with blood vessels and nerve endings. The oral cavity contains several organs such as the tongue, teeth, salivary glands and their related ducts (Dewhirst *et al.*, 2010). The tongue is a muscular organ located on the floor of the mouth that plays a crucial role in taste perception, food manipulation, and speech production. Teeth are located in the alveolar processes of the maxilla and mandible and help break down food through chewing. The major salivary glands, the parotid, submandibular and sublingual glands, secrete saliva into the oral cavity (Ten Cate, 1998), which aids in lubrication, enzymatic digestion and antimicrobial defence. Saliva also contains enzymes such as amylase, which initiate the digestion of carbohydrates. Moreover, the oral cavity is equipped with a plethora of lymphoid tissues, including the tonsils and lingual tonsils, which form part of the body's immune system, defending against pathogens entering through the oral route (Laine and Smoker, 1996).

1.1.2 Salivary glands

Salivary glands secrete saliva into the oral cavity of humans and other vertebrates. Salivary glands can be divided into two groups: minor and major (Ten Cate, 1998). There are roughly 600-1000 minor salivary glands located below the epithelial layer, distributed across the entire mouth cavity and found in the labia, buccal mucosa, palate and lingual region (Ghannam *et al*, 2023). The sublingual, submandibular and parotid glands comprise the three pairs of major salivary glands (Figure 1.1). Sublingual glands are the smallest and are located between the tongue and teeth. Sublingual glands are mostly composed of mucous cells producing a mucous saliva. Submandibular glands, the second largest glands are located at the back of jaw, and are composed of a mixture of serous and mucous cells; they secrete seromucous saliva (Tucker, 2007). The saliva from both the sublingual and submandibular glands feeds into the mouth through the major submandibular duct, the Whartons duct. It is thus difficult to collect pure saliva from either gland. The largest of the major glands are the parotid glands, which are located below and forward of the ears. (Tucker, 2007) These glands are comprised solely of serous cells, producing serous saliva secreted into the mouth through the Stenson's duct. The primary function of salivary glands is to produce and secrete saliva, which serves several important roles in oral health and digestion. Disorders affecting salivary gland function can have significant consequences for oral health and overall well-being. Conditions such as xerostomia (dry mouth) (Chaudhury *et al.*, 2015), sialadenitis (inflammation of salivary glands) (Chi *et al.*, 2024), and salivary gland tumours can impair saliva production and lead to complications such as difficulty swallowing, oral infections and dental caries (Pedersen *et al.*,

2018). Diagnosis and management of salivary gland disorders often require a multidisciplinary approach involving dentists, otolaryngologists, and oral surgeons (Ship, 2002).

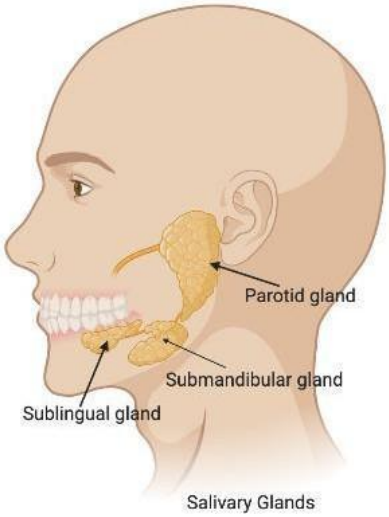


Figure 1.1. The location of three major salivary glands in human oral cavity. The image was generated by BioRender: https://www.biorender.com/

1.1.3 Saliva

Saliva is a colourless and water-like mixture secreted by both major and minor glands (Zhang *et al.*, 2016). Unstimulated whole saliva comprises 65% submandibular saliva, 20% parotid saliva, 7-8% sublingual saliva and the remaining portion from minor glands (Humphrey and Williamson, 2001). When saliva is stimulated, more than 50% comes from the parotid gland (Komine *et al.*, 2007). Saliva contains inorganic molecules such as Na+, K+, Ca2+ and thiocyanate ions (Zhang *et al.*, 2016), and organic molecules such as glycosylated proteins which gives it a sticky consistency (Vidotto *et al.*, 2010; Carneiro *et al.*, 2012). Saliva helps to lubricate the oral mucosa, facilitating speech, mastication, and swallowing. It also prevents

friction between oral surfaces, reducing the risk of injury and irritation (De Almeida *et al*, 2008). Additionally, saliva can provide genetic information through both human and oral microorganism-derived DNA or RNA (Humphrey and Williamson, 2001). With the rapid progress that has been made in salivary proteomics, transcriptomics and genomics, saliva has been recognised as a potential source of biomarkers and, as it is relatively easy to collect in a non-invasive manner, its potential as a prognostic/diagnostic tool continues to be investigated (Vitorino *et al.*, 2004).

Saliva contains many proteins and peptides such as mucins, histatins, statherin, lactoferrin, lactoperoxidase, lysozyme, secretory immunoglobulin A and amylase which contribute to its antimicrobial properties. Salivary proteins are a diverse group of biomolecules, each with its distinct structure and purpose. Salivary mucins and salivary agglutinin are in the first line of defence against oral bacteria as they bind to bacteria by lectin-like-carbohydrate interactions, leading to bacterial aggregation and thus preventing attachment to, colonisation, or infection of oral tissues (Frenkel and Ribbeck, 2015). As a component of the protective pellicle layer, investigations have concentrated on the interactions between oral *streptococci* and salivary mucins (Al-Hashimi and Levine, 1989; Frenkel and Ribbeck, 2015).

Mucin is a highly glycosylated and significant salivary protein complex that plays an important role in digestion and oral processing of food partially through its ability to form gels. Oral food processing is complex but occurs in a step-by-step manner, with salivary mucin being involved in mastication, transportation and swallowing through interactions with teeth, the palate and the tongue. The molecular weight of mucin complexes are exceptionally high with the gel forming mucins being up to 4x 10^4 kDa (Thornton *et al.*, 1999). The main

salivary glycoproteins consist of two types of mucins, notably MUC7 and MUC5B. MUC7 is a monomeric mucin with a low molecular weight, specifically ranging from 130 to 180kD, whereas MUC5B is a high-molecular weight, oligomeric mucin with a total molecular mass ranging from 2 to 4x 10^4kDa (Takehara *et al.*, 2013). MUC5B is produced by the mucous acini of the submandibular and sublingual salivary glands. It has three distinct subtypes and functions to lubricate the mouth epithelium by creating a gel-like substance. The gel-like layer covering the oral epithelium acts as a barrier against microorganisms (Frenkel and Ribbeck, 2015).

Histatins are small, cationic proteins with molecular weights ranging from 3-4.5 kDa mainly produced by the parotid glands. Studies have demonstrated that histatin possesses many indirect antibacterial properties such as metal ion binding, neutralisation of bacterial lipopolysaccharide(LPS), and inhibition of proteinases. Histatins can directly affect such as *Streptococcus mutans* and *Candida albicans* by integrating into their cytoplasmic membrane, enhancing permeability, and leading to growth suppression or death (Nieuw Amerongen and Veerman, 2002).

Statherin is a histatin related protein and is a small, acidic protein with a molecular weight of around 5 kDa that is produced by the parotid and submandibular glands (Raj *et al.*, 1992). Statherin is a phosphoprotein that attaches to hydroxyapatite and helps in the formation of the enamel pellicle (Sabatini *et al.*, 1987). Statherin prevents the initial or natural formation of calcium phosphate on tooth surfaces, which is essential for maintaining tooth structure and it is recognised for its ability to enhance the attachment of *Actinomyces. viscosus* to teeth and has specific binding sites for *P. gingivalis* fimbriae (Amano *et al.*, 1996). Furthermore,

statherin promotes the transition of *C. albicans* hyphae to yeast, a significant transition as the hyphal form of *C. albicans* is more invasive and can lead to oral candidiasis, a fungal infection in the mouth. By promoting the conversion to the yeast form, statherin may help to reduce the pathogenicity of *C. albicans*, thereby aiding in the oral defence against oral candidiasis (Leito *et al.*, 2009). Statherin binding to mucins forms protein complexes which protect the proteins from microbial proteolytic action and encourage the clumping together of bacteria, which can then be eliminated by swallowing (Bruno *et al.*, 2005).

Lactoferrin is a glycoprotein with a molecular weight of around 80 kDa and is produced by the serous acinar cells of both major and minor salivary glands. Lactoferrin belongs to the transferrin family of iron-binding proteins and functions in an antibacterial manner by binding ferric iron (Komine *et al.*, 2007). Lactoferrin binding to iron prevents microbes including bacteria, yeasts, and parasites from accessing this crucial nutrient needed for their growth. Lactoferrin can enhance the attachment and clustering of specific microbes and also aid in the degradation of microbial cell membranes. Apo-lactoferrin (iron-free form of lactoferrin) has been demonstrated to cause agglutination of various oral bacteria such as *S. mutans, S. sobrinus, Streptococcus rattus, S. sanguinis* and *P. gingivalis*, while the iron-saturated form only agglutinates *S. mutans* (Soukka *et al.*, 1993). This suggests that lactoferrin has bacteriolytic properties that are independent of its iron-binding capabilities (Ab, 1993; Arslan, Leung and Wu, 2009). Lactoferrin in whole saliva is derived from neutrophil granulocytes and gingival crevicular fluid and expression levels increase during infections and inflammatory states (Reitamo, Konttinen and Segerberg-Konttinen, 1980; Eberhard *et al.*, 2006). Lactoferrin can neutralise LPS by binding to the lipid A part, competing with LPS

binding protein (LBP), and inhibiting the formation of the LBP-CD14-TLR4 complex, which leads to a decreased inflammatory response (Elass-Rochard *et al.*, 1998; Komine *et al.*, 2007).

Lactoperoxidase is a member of the salivary peroxidase system which includes thiocyanate (SCN^-) ions, hydrogen peroxide and peroxidase. Lactoperoxidase and other proteins are incorporated into oral health products, such as tablets, toothpaste, mouth rinse or gel, to improve the antibacterial properties of saliva. Research indicates that consistent use of lactoferrin and lactoperoxidase-containing products can alter the microbial environment in the mouth, potentially leading to better oral health, including addressing issues such as bad breath and gum conditions (Shin *et al.*, 2011; Morita *et al.*, 2017; Pedersen *et al.*, 2019).

Lysozyme is a 14 kDa protein and contributes to the natural defensive mechanisms found in saliva. Lysozyme in whole saliva is derived mostly from the major and minor salivary glands, with a smaller contribution from gingival crevicular fluid and salivary leukocytes. Lysozyme functions by catalysing the hydrolysis of the β-1,4-glycosidic linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine in the polysaccharide layer of the cell wall of Gram-positive bacteria (Gibbons, de Stoppelaar and Harden, 1966; Balekjian, Hoerman and Berzinskas, 1969). Gram-negative bacteria are less sensitive to lysozyme than Gram-positive bacteria due to the highly protective LPS coating that restricts lysozyme access. Lysozyme may cause oral bacteria, such as *streptococci*, to clump together, reducing their capacity to stick to oral surfaces and helping to remove microorganisms from the oral cavity (O'Sullivan, Jenkinson and Cannon, 2000). Studies have also shown that lysozyme has antifungal and antiviral properties (Yeh *et al.*, 1997; Tsang and Samaranayake, 1999; Samaranayake *et al.*, 2001).

Secretory Immunoglobulin A (sIgA) and Immunoglobulin G are two major antibody (immunoglobulin) classes present in human saliva (Brandtzaeg, 2013). Dimeric IgA is produced by plasma cells in the stroma of the salivary glands which is subsequently carried across the glandular epithelial cells by the polymeric Ig receptor, also known as the membrane secretory component. And then IgA is exocytosed at the apical surface of the epithelial cell following cleavage of the Ig receptor (Brandtzaeg and Prydz, 1984; Brandtzaeg, 2013). The majority of IgG in saliva comes from blood through passive leaking into the gingival crevicular fluid, with only a small portion being produced by the salivary glands. The amounts of salivary IgM, IgD and IgE are very small and come mainly from gingival crevicular leakage. Salivary IgM levels are related to serum IgM levels and periodontal inflammation (Brandtzaeg, 2013). The main role of sIgA as a primary defensive mechanism in the oral cavity appears to be its ability to bind to antigens in saliva, the oral mucosa and the acquired enamel pellicle, a process known as immunological exclusion (Brandtzaeg, 2013). The secretory component of sIgA defends the immunoglobulin from degradation by proteolytic enzymes, however, certain bacteria such as *S. sanguinis*, *S. mitis*, *P. gingivalis*, *Prevotella*, and *Capnocytophaga* species can enzymatically break down parts of the sIgA (Kilian *et al.*, 1996; Russell *et al.*, 1999). The antibacterial capabilities of sIgA include preventing microbial adherence to mucosal and dental surfaces and increasing the removal of microbes such as *S. mutans* from the oral cavity by agglutination. Salivary immunoglobulins work together with the innate defence systems and the production of salivary antibodies in response to *streptococci* which is important in controlling the initial colonisation of oral surfaces (Russell *et al.*, 1999; Borges *et al.*, 2015).

Amylase, is a member of an enzyme family with the most abundant component being α-amylase, and is widely distributed in plants (malt and cornflower), *Bacillus* and *Aspergillus* species, pancreas and oral fluid of human and other mammals. Salivary α-amylase is around 50 kDa in size and is produced by the parotid and submandibular glands (Contreras-Aguilar *et al.*, 2017). Amylase facilitates the adherence of bacteria and the fermentation of bacterial sugars in the oral cavity. Studies have shown that amylase is found in the enamel pellicle, indicating binding to the tooth surface and oral *streptococci* including *S. mitis*, *S. anginosus* and *S. gordonii* but not to *S. oralis*, *S. mutans* or *S. sanguinis* and Gram-negative bacteria which lead to the formation of dental plaque (Aguirre *et al.*, 1987; Scannapieco, Torres and Levine, 1993; Hannig *et al.*, 2004).

These proteins and peptides collaborate to protect the oral cavity, rather than acting alone, to provide multiple defensive actions against microorganisms in the oral cavity.

Another major salivary protein is BPIFA2/SPLUNC2, an abundant small, secreted protein found in mammalian saliva such as rodents (Poulsen *et al.*, 1986; Ballli, 1992), cows (Haigh *et al.*, 2008), pig (Yin *et al.*, 2006), primates and humans (Bingle, Bingle and Craven, 2011). This will be the focus of my thesis and is introduced in detail in the following sections:

1.2 The BPI-LBP-PLTP-CETP protein family

The BPI-LBP-CETP-PLTP protein family consists of Bactericidal Permeability Increasing protein (BPI), LBP, Cholesteryl Ester Transfer protein (CETP) and Phospholipid Transfer protein (PLTP). All members of this family have the ability to bind to and transfer lipids, thus mediating LPS responses (Hailman *et al.*, 1996; Bingle and Craven, 2002) and are members

of the wider tubular-lipid binding protein (TULIP) family (Kopec, Alva and Lupas, 2011; Alva and Lupas, 2016). BPI and LBP are structurally related and have similar functions in mediating signals from LPS (Elsbach *et al*., 1979; Eckert *et al*., 2013).

Lipopolysaccharide

LPS is an important component of the bacterial cell surface and is a phosphorylated glycolipid specifically produced by Gram-negative bacteria (Alexander *et al*., 2001). The pathogenic capacity of Gram-negative bacteria is often related to their cell wall composition, and more specifically to their LPS layer. Gram-positive bacteria, however, only have a thick peptidoglycan layer and no outer lipid membrane. LPS is composed of a lipid A region, an O-polysaccharide and core polysaccharide and is a phosphorylated glucosamine disaccharide combined with several fatty acids. Lipid A is the source of bacterial endotoxin activity and helps the LPS molecule anchor onto the outer cell membrane. O-polysaccharide, also called O- antigen, consists of multiple and repetitive oligosaccharides. Multiple O-chains make LPS smooth, whereas the absence of O-chains make the LPS rough and more hydrophobic (Tsujimoto, Gotoh and Nishino, 1999; Rittig *et al*., 2003). The core polysaccharide contains variable carbohydrate chains and LPS core polysaccharides in some bacteria also contain non-carbohydrate components, such as ethanolamine substituents, phosphates and amino acids. Lipid A attaches directly to the core polysaccharide. LPS is recognized as an endotoxin leading to an inflammatory response in mammals. LPS binds to LPS receptors, including cluster of differentiation 14 (CD14), toll-like receptor 4 (TLR4) and myeloid differential protein 2 (MD-2), on cell membranes to promote the secretion of various cytokines by inflammatory cells.

1.2.1 Lipopolysaccharide-binding protein

LBP is a 60 kDa, soluble, acute phase glycoprotein largely produced by hepatocytes. In the human bloodstream, LPS binds to LBP and the resulting LPS-LBP complex has affinity for the immunity receptor CD14 (Schumann, 1992). The N-terminal of LBP binds to the lipid A region of LPS (Weiss, 2003) and then the catalytic LPS molecules are delivered to both membrane-bound and soluble CD14 to form a specific recognition receptor complex (LPS-CD14). This process increases host cell sensitivity to LPS (Ryu *et al*., 2017). Interestingly, LPS has a high binding capacity for CD14, but only in the presence of LBP, indicating that LBP plays a key role in the immune reaction to Gram-negative bacteria (Hailman *et al*., 1994). Both CD14+ cells (monocytes and macrophages) and CD14-cells (epithelial and endothelial) can be activated by transferring LPS into these cells (Funda *et al.*, 2001). Membrane-bound CD14 (mCD14) remains associated with the activating cells when it is combined with LBP while soluble CD14 (sCD14) separates from the complex when bound to LBP; this is the key difference between mCD14 and sCD14.

1.2.2 Bactericidal Permeability Increasing protein

BPI is a 55 kDa protein, considered to be largely produced by neutrophils and is recognised for its ability to kill Gram-negative bacteria and neutralise LPS (Bülow *et al.*, 2024). BPI contains 456 residues and the structure of crystal BPI is shown in Figure 1.2 (Beamer, Carroll and Eisenberg, 1997). The NH_2-terminal domain is displayed in light blue on the left side, while the COOH-terminal domain is displayed in dark blue on the right side. Additionally, two phosphatidylcholine molecules are represented in red. The linker is shown in a yellow

colour, while the disulphide bond is represented using a ball-and-stick model. Despite BPI having 44% sequence similarity to LBP at the protein level (Eckert *et al.*, 2013), it has been described as a natural antagonist to LBP because its mode of activity, preventing LPS-dependent stimulation of immune cells, is opposite to that of LBP (Weiss *et al*., 2003; Wittmann *et al*., 2008). Like LBP, BPI is a host endotoxin-binding protein that is able to interact with lipid A and inner core sugars to enable host defence recognition of Gram-negative bacteria (Gazzano-Santoro *et al.*, 1995). The N-terminal domains of BPI and LBP have significant structural and functional similarities (Beamer, Carroll and Eisenberg, 1997), however, the C-terminal domain of BPI has no bactericidal activity nor the ability to bind LPS (Ooi *et al*., 1991). The functional differences between BPI and LBP are therefore related to the carboxy terminal domains (Abrahamson *et al*., 1997; Eckert *et al*., 2013). BPI has the ability to bind to lipoteichoic acid (LTA) which is a lipid-like compound from Gram-positive bacteria. LTA is considered a pathogen-associated molecular pattern (PAMP), which means it is recognized by the immune system as a signature of microbial presence. BPI's binding to LTA is functionally significant because it allows BPI to interact with Gram-positive bacteria in a manner similar to how it interacts with Gram-negative bacteria. This interaction is part of the host defence mechanism that enables the recognition and response to bacterial infections. (Akira, Uematsu and Takeuchi, 2006). Lipoproteins and peptidoglycan (PG), found in both Gram-positive and Gram-negative bacteria, are highly effective immunostimulants.

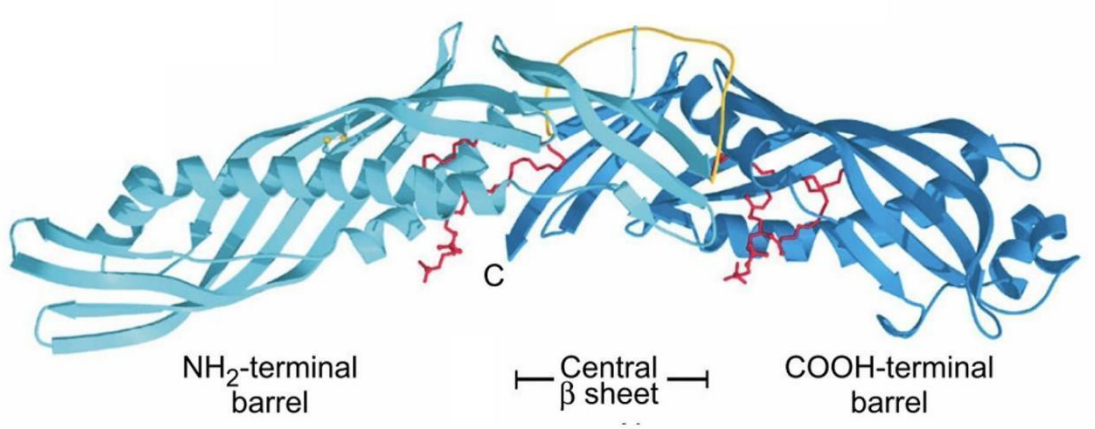


Figure 1.2. The predicted crystal structure of human BPI. BPI has 456 residues demonstrating its boomerang (Beamer, Carroll and Eisenberg, 1997). NH_2-terminal domain is illustrated in a light blue (left), the COOH-terminal domain is presented in a dark blue (right) and β sheet is in the centre. Two molecules of phosphatidylcholine with a yellow linker are highlighted in red. The ball-stick model was used to represent a disulphide bond. The crystal structure of BPI was generated by MOLSCRIPT and RASTER3D.

1.2.3 Cholesteryl Ester Transfer protein and Phospholipid Transfer protein

CETP and PLTP are both lipid-binding molecules. CETP can transfer cholesteryl esters from high density lipoproteins (HDL), leading to lowering of HDL concentrations. In contrast, PLTP can increase HDL levels by removing phospholipids from triglyceride-rich lipoproteins (Masson *et al*., 2009). Although members of the wider BPIF-containing family, these proteins will not be discussed further as they play no significant role in bacterial interactions with the immune system.

1.3 PLUNC proteins

The Palate Lung Nasal Clone (PLUNC) was originally reported in the nasal epithelium of the embryonic mouse and the trachea and bronchi of adult mouse lung (Weston *et al*., 1999) demonstrating a restricted pattern of gene expression. *Plunc* expression increased between

days 13 and 14 of gestation, corresponding with the elevation and fusion of the palatal shelf. *Plunc* expression was observed in the nasal collumella, turbinates, and nasal canal of adult mice, similar to developing mice. Strong expression was observed in the outer epithelial layers of the respiratory passages, extending through the trachea and bronchioles. The expression decreased significantly at the next bronchiole branch and became sporadic, eventually becoming absent in the distal regions of the lung, including the terminal bronchioles, respiratory bronchioles, and alveoli (Weston *et al.*, 1999). PLUNC was also shown to share some amino acid similarity with BPI and LBP.

1.3.1 Murine PLUNC and human PLUNC

The murine PLUNC (*Plunc*) gene encodes a protein of 272 amino acids (Weston *et al.*, 1999). Amino acid sequence analysis showed similarities between PLUNC and two mouse salivary gland proteins: von Ebner minor salivary gland protein and parotid secretory protein (PSP) precursor. These proteins exhibited low sequence similarity, but the signal peptide sequences were more similar. Murine PSP shared 14 identical or conserved amino acids out of the first 15 amino acids of mPLUNC and bovine PSP shared 12 identical or conserved amino acids with mPLUNC (Weston *et al.*, 1999). Murine PSP was demonstrated to bind to membrane proteins of *Listeria monocytogenes, Aggregatibacter actinomycetemcomitans, Escherichia coli* and *S. mutans* with the interactions relying on the presence of Zn^{2+} in saliva, indicating that the protein might serve as a host protein that can regulate the growth or colonisation rates of bacteria (Robinson *et al.*, 1997).

Human *PLUNC* encodes a 256 amino acid protein and like the mouse orthologue it has a high leucine content (Bingle and Bingle, 2000). Subsequent research led to the discovery of a family of PLUNC related proteins encoded by adjacent genes in an approximately 300 kb region of chromosome 20q11. Related loci exist in other species (Bingle, Bingle and Craven, 2011).

1.3.2 PLUNC family of proteins

Soon after the identification of *PLUNC* multiple additional related genes were identified and classified into two groups based on their amino acid composition. The original PLUNC protein contains a single BPI related domain and became Short PLUNC1 (SPLUNC1) as shown in Figure 1.3A (Bingle and Craven, 2002). Two other single BPI-domain containing proteins were also identified: SPLUNC2 (originally identified as Parotid Secretory Protein, PSP) and SPLUNC3. A second group of proteins were designated Long PLUNCs (LPLUNCs) as they contain domains similar to both BPI domains as described in Figure 1.3B. This group originally contained LPLUNC1, LPLUNC2, LPLUNC3 and LPLUNC4 (Bingle and Craven, 2002). The PLUNC family has been expanded as genomic resources have improved.

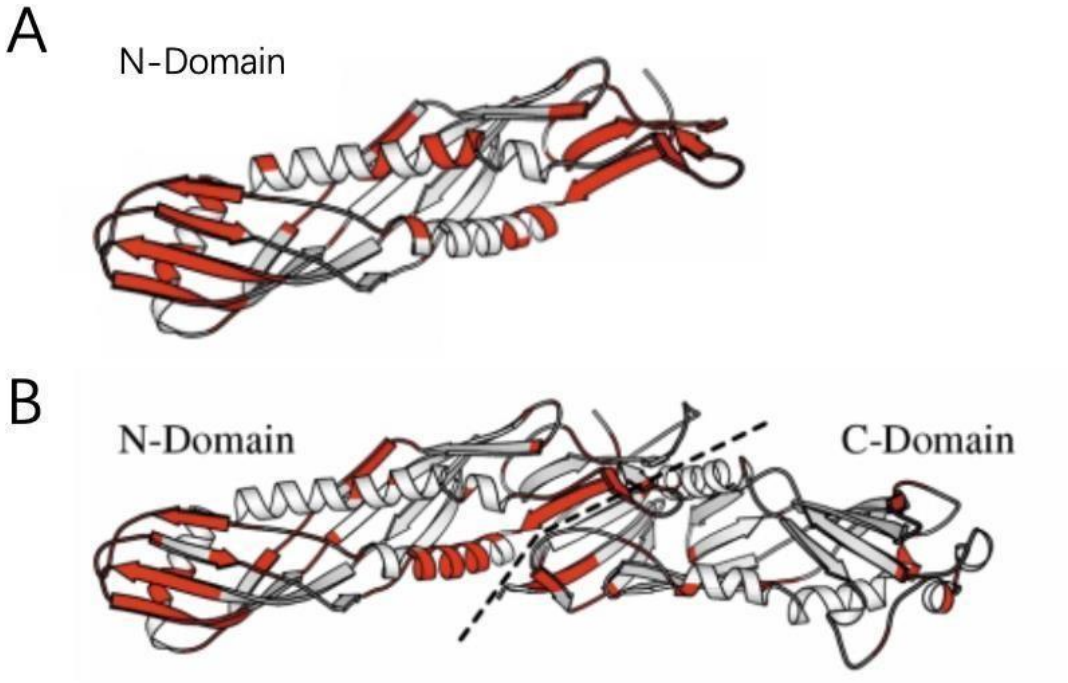


Figure 1.3. The predicted secondary structure of SPLUNC (BPIFA) and LPLUNC (BPIFB) proteins. (A) The structure of SPLUNC (BPIFA) proteins superimposed on the NH_2- terminal domain of BPI. (B) The structure of LPLUNC (BPIFB) proteins superimposed on both the NH_2-terminal domain and COOH-terminal domain of BPI. The less well conserved regions of secondary structures compared to the structure of BPI are highlighted in red. 3DPSSM was used for the structure prediction (Bingle and Craven, 2002).

1.4 PLUNC nomenclature

The expansion of the PLUNC protein family resulted in various alternative names being assigned to its members. For example, SPLUNC1, originally identified as PLUNC, was also referred to as LUNX and SPURT; SPLUNC2 was alternatively named PSP and C20ORF70, and LPLUNC1 was known as Von Ebner minor salivary gland protein and C20ORF114. A standardised naming system was deemed necessary to address the confusion caused by the variety of names used for each family member (Bingle, Seal and Craven, 2011). In 2011 the TULIP (tubular lipid-binding) superfamily including BPI, LBP, CETP, PLTP, and PLUNC was renamed as the BPI-fold containing family. The scheme uses BPIF (BPI fold containing

family) as the root.

Due to the ‘Short’ PLUNC proteins, having homology only to the N-terminal domain of BPI, they were designated BPIFAs while ‘long’ PLUNC proteins with homology to both the N- and C-terminal domains of BPI were designated BPIFBs.

To include all members of the BPI-fold family, such as paralogues from other species, gene names were also extended with letters 'A', 'B', 'C', etc., to distinguish between the two lineages. BSP30 proteins from Cow, previously referred to as BSP30A, BSP30B, BSP30C, and BSP30D, were reclassified as BPIFA2A, BPIFA2B, BPIFA2C, and BPIFA2D. Similarly, mouse and rat PSP were recognised as BPIFA2E, establishing them as part of the extended BPIFA2 sub-family. The established human *BPIF* gene locus and phylogenetic tree of BPIF-containing proteins that underpinned this nomenclature are presented in Figures 1.4 and 1.5.

The term BPIF will be used throughout this thesis.

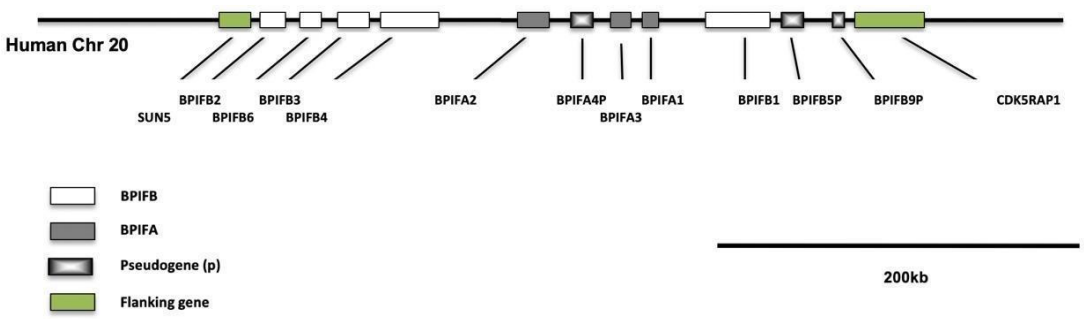


Figure 1.4. Distribution of the human *BPIF* gene loci. The *BPIF* gene locus in humans is located on chromosome 20 and covers an area of around 300kb. It consists of three pseudogenes, marked by shaded boxes, three *BPIFA* genes, represented by grey boxes, and five *BPIFB* genes, symbolised by white boxes. The *BPIF* region is flanked by two genes that are not related, shown in green. Modified from the Genome Reference Consortium. Scale bar is 200kb.

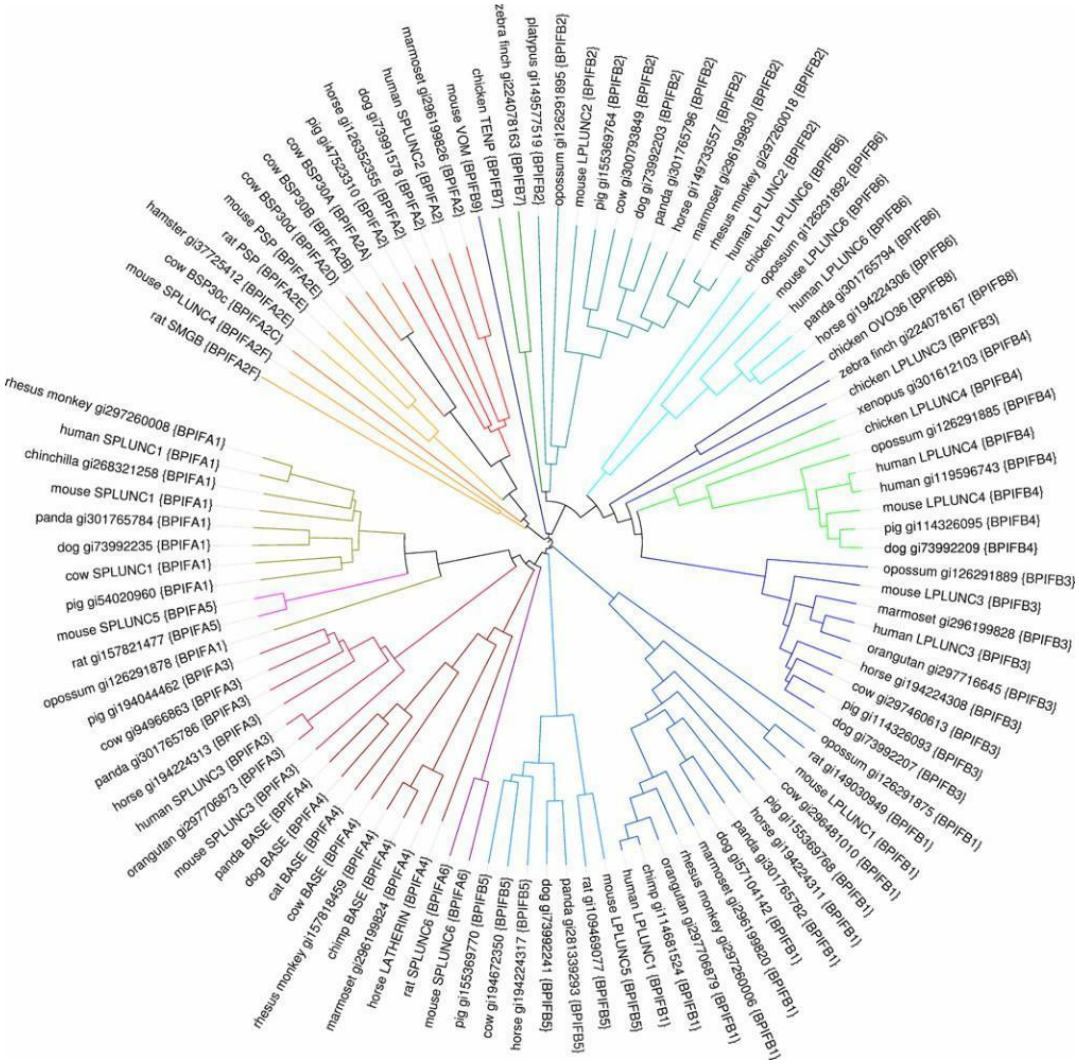


Figure 1.5. Phylogenetic tree of BPIF-containing proteins used to establish the family nomenclature (Bingle, Seal and Craven, 2011). The tree was generated using the Clustalw (https://www.genome.jp/tools-bin/clustalw) and visualised using Interactive Tree Of Life (ITOL:https://itol.embl.de/). The proteins used in this study were identified by the National Centre for Biotechnology Information (NCBI) through BLAST searches.

1.5 BPIF Proteins in other species

During the development of the systematic nomenclature, BPIF proteins were identified in multiple mammalian species including xenopus, chicken, zebra finch and marsupials (opossum). To date it appears that *BPIFA* genes are only found in mammals, with some BPIFB genes being found in birds and amphibia. (Bingle, Seal and Craven, 2011). This

analysis described all unique mammalian BPIF-containing proteins and supported the earlier hypothesis that BPIF proteins are limited to vertebrates. It also confirmed that BPIFA proteins appear to be exclusive to the therian lineage (mammals and marsupials). Research, principally from my host group, has also shown that BPIF proteins exhibit significant divergence in both paralogous and orthologous relatives. These protein families are highly diverse, with individual paralogues usually sharing sequence identities of less than 30% (Bingle and Craven, 2004; Bingle, Seal and Craven, 2011). Therefore, investigating the study of BPIFA proteins across mammalian evolution is valuable for our understanding of their specific functions.

1.5.1 *BPIF* expression

To date most research has focused on BPIFA1, the first member to be reported, with limited studies related to BPIFA2 and almost no studies associated with BPIFA3. Originally BPIFA1 expression was described in the palate, nasal cavity and upper airways of adult mice (Weston *et al.*, 1999). Multiple studies have confirmed that BPIFA1 is a secretory protein and can be detected in saliva, nasal lavage fluid, airway lining fluid and middle-ear effusions in a number of mammals including chinchilla, mice and humans (Vitorino *et al.*, 2004; Wu *et al.*, 2005; McGillivary and Bakaletz, 2010; Preciado *et al.*, 2010). Immunohistochemical staining analysis showed that strong staining was described in the submucosal gland cells and mucous cells of salivary glands (Bingle *et al*., 2005), but not in serous cells (Vargas *et al*., 2008). In humans, BPIFA1 was found on the surface of the ciliated epithelium in the upper airways and also in early fetal oral tissues by using immunohistochemical analysis (Alves *et al.*, 2017). BPIFA1 has been shown to interfere with biofilm formation by *Pseudomonas aeruginosa*

biofilms *in vitro* (Gakhar *et al*., 2010). *P. aeruginosa* is a common opportunistic pathogen that forms biofilms, which are organised groups of bacteria bound in an extracellular matrix. Biofilms prevent bacteria against potential threats such as antibiotics and the host's immune system. By preventing biofilm formation, BPIFA1 helps maintain a mucosal environment less hospitable to bacterial colonization and persistent infections, thereby contributing to mucosal surface defence (Gakhar *et al*., 2010). It has also been reported that BPIFA1 expression is associated with infectious and/or inflammatory disorders. Compared to healthy individuals, reduced BPIFA1 was demonstrated in cigarette smokers, patients suffering from allergic rhinitis and those exposed to air pollutants (De Smet *et al.*, 2018). The differentiation status of the epithelial cell population in the nasal and upper respiratory tract is a decisive factor in the amount of BPIFA1 secreted. Infection and inflammation of the nasal cavity results in reduced BPIFA1 secretion (Ghafouri *et al.*, 2002; Yin *et al.*, 2006). In oral lichen planus patients, the level of BPIFA1 is much lower than that noted in controls by using 2-DE gel analysis, suggesting it might be a useful biomarker in this disease (Yang *et al*., 2006). BPIFA1 can also contribute to the mucociliary clearance of bacteria in the airways and protect the middle ear by regulating mucosal liquid volume (Haverkos, 2003; Garcia-Caballero *et al.*, 2009). Further studies, however, are still needed to conclusively confirm these observations and function of the protein. BPIFB1 is often compared with BPIIFA1 as they have similar expression profiles, however, our understanding of the true role of BPIFB1 remains to be fully elucidated as do the expression patterns and roles of other BPIFB proteins. BPIFB1 is produced by goblet cells in the submucosal glands and epithelium of both larger and smaller airways and the trachea whereas BPIFA1 is not detected here (Bingle *et al*., 2010) and staining of BPIFB1 in minor salivary glands and minor mucosal glands in the nasal cavity is greater than that in major

salivary glands, (Bingle *et al*., 2010). Previous data from our lab showed that BPIFB1 staining was increased in advanced cystic fibrosis (Bingle *et al.*, 2012) and that it is also prominently produced in the bronchiolised epithelium that lines the honeycomb cysts in typical interstitial pneumonia (Bingle *et al.*, 2013). Our group has also demonstrated an association between increased levels of BPIFB1 in sputum and the development of airflow limitation over a 4-year period in chronic obstructive pulmonary disease (COPD). This association is particularly significant among those who are currently smoking. (Gao *et al.*, 2015).

Bpifa5 is a novel, rodent restricted gene that originated from a duplication of the *Bpifa1* gene (Bingle *et al.*, 2004; LeClair *et al.*, 2004). The sequence similarities between BPIFA1 and BPIFA5 are around 60% and the expression of *Bpifa5* is limited to the interpapillary epithelium on the dorsal surface of the tongue (LeClair *et al.*, 2004). The function of BPIFA5 remains unknown.

BPIFA4, expressed in the salivary glands and originally identified as the putative breast cancer marker gene *BASE (breast cancer and salivary gland expression)* (Egland *et al.*, 2003), is a pseudogene in humans and is not present in rodents (Bingle, Bingle and Craven, 2011). *BPIFA4* is an authentic gene in chimpanzees (and other primates) and in horses. It displays significant divergence, with pairwise similarities ranging from 40% to 50% when horse is compared to the chimpanzee sequence (Egland *et al.*, 2003; McDonald *et al.*, 2009); this divergence possibly indicates a functional change in horses. Horse BPIFA4, previously referred to as latherin, was first identified in the 1980s, is found in high quantities in horse sweat and has been shown to act as a surface-active protein without sugars; it is believed to be responsible for the froth observed in horses after intense exercise (Beeley, Eason and Snow,

1986). Latherin is synthesized and stored in granules in the sweat glands of the horse skin and is also localised in the salivary glands (McDonald *et al.*, 2009). It appears that BPIFA4 has been subjected to a number of lineage specific evolutionary influences potentially resulting in a unique function.

BPIFA3 consists of 253 amino acids and is predominantly expressed in the testes and appears to undergo multiple alternative splicing events (Bingle, Bingle and Craven, 2011). There is a shortage of information about BPIFA3, but *Bpifa3* deficient mice do not have a testicular phenotype (Miyata *et al.*, 2016).

1.6 BPIFA2 proteins in human

BPIFA2 is 25 kDa, 249 amino acid protein predominantly expressed in major salivary glands and secreted in saliva (Bingle *et al.*, 2009). Mature BPIFA2 protein contains two N-glycosylation sites located at residues Asn 124 and Asn 132 (Ramachandran *et al.*, 2006) . Previous data in our lab also demonstrated that BPIFA2 from human saliva and analysed by Western blotting appears as several positive bands with a range of sizes, which suggests BPIFA2 undergoes differential glycosylation (Bingle *et al.*, 2009). BPIFA2 is mainly detected in the interlobular ducts and serous acinar cells of the major salivary glands (sublingual, submandibular and parotid gland) with no protein being found in the mucous acinar cells. BPIFA2 has also been observed in several minor salivary glands, and the minor glands of the vallecular region of the tongue where it exhibited a staining pattern similar to that of the parotid gland. The minor glands in the posterior tongue exhibited a similar pattern to the sublingual and submandibular glands. No BPIFA2 was observed in the minor glands located deeper in the respiratory system (the submucosal glands), the respiratory mucosa and tissues

outside the oral cavity (Bingle *et al.*, 2009). Expression of *BPIFA2* is very limited outside of the salivary glands as is clearly illustrated using data from the Human Protein Atlas (HPA; https://www.proteinatlas.org/; Figure 1.6).

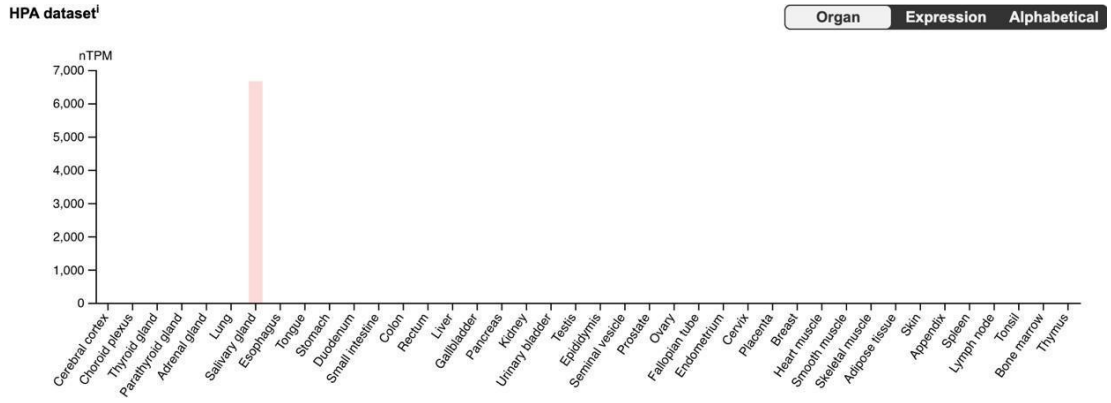


Figure 1.6. Expression of *BPIFA2* in human organ tissues. The Human Protein Atlas database (https://www.proteinatlas.org/) was used to visualize the expression locations of *BPIFA2* in human organs. The HPA RNA-seq tissue data is presented in nTPM (normalised protein coding transcript per million) as shown on the Y-axis, which represents the average values of the individual samples from each tissue.

Consistent with protein localisation data, at the single cell level *BPIFA2* expression is mostly seen in serous cells in salivary glands in the HPA database (Figure 1.7).

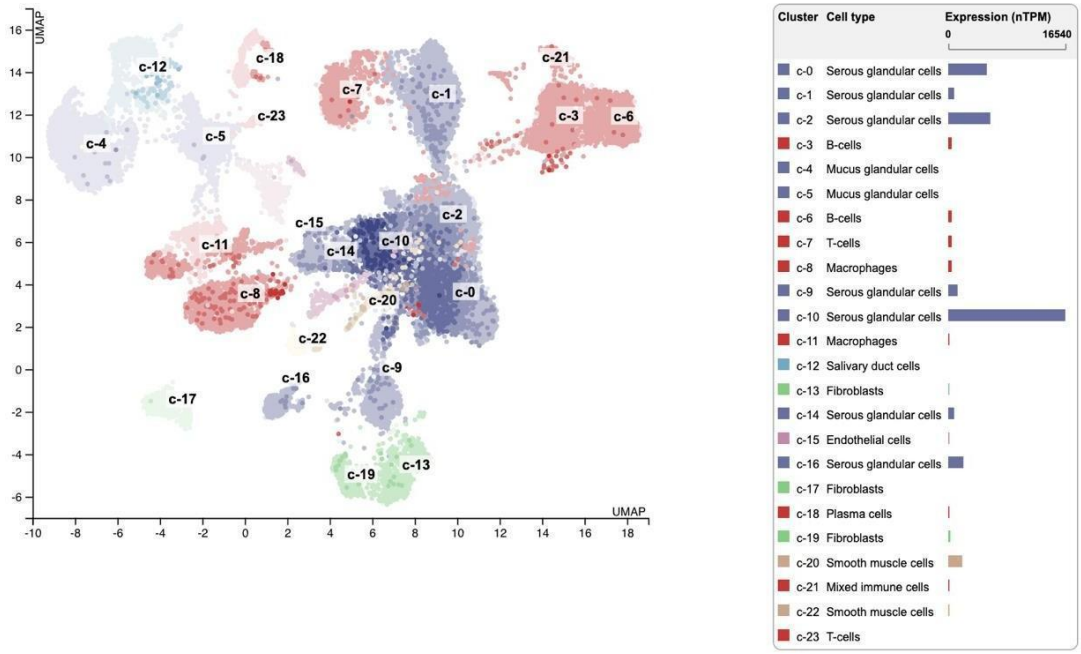


Figure 1.7. Expression of *BPIFA2* at the single cell level in human salivary glands. The Human Protein Atlas database (https://www.proteinatlas.org/) was used to visualize *BPIFA2* expression in single cells of salivary glands. The clusters highlighted in different colours as shown on the left side represent cell types and the numbers correspond to the list of cell types on the right side. The expression levels are represented by nTPM.

The predicted structure of human BPIFA2, generated by threading analysis shows the same β-barrel structure found in other family members (Figure 1.8). The human BPIFA2 protein includes an N-terminal signal sequence (residues 1-16), typical of a secretory protein, it is mostly hydrophobic and consists of a substantial proportion of leucine/isoleucine residues (26%). The predicted isoelectric point (pI) BPIFA2 is 5.35 and as with all BPIF proteins, BPIFA2 includes two cysteine residues that are expected to form a single disulphide bond (Bingle *et al.*, 2004).

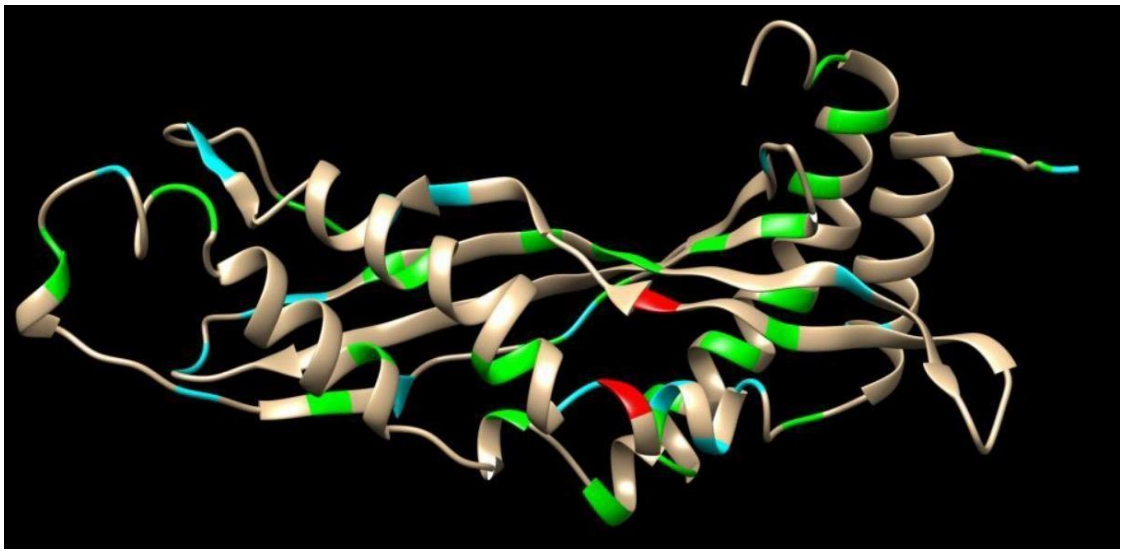


Figure 1.8. The predicted structure of BPIFA2. There are two cysteine residues (red), hydrophobic residues leucine (Green) and Isoleucine (Blue). The human *BPIFA2* protein was modelled using the Phyre server and visualised with Chimaera (http://www.cgl.ucsf.edu/chimera).

Another interesting finding from protein studies was that when two BPIFA2-specific antibodies (BPIFA2A and BPIFA2B; chapter 2 section 2.2.2 for further information regarding antibody generation; figure 1.9) raised to different epitopes of the protein, one to an epitope in the centre of the protein BPIFA2(A*)* and another to an epitope at the C-terminus BPIFA2(B), were used very different patterns of expression on serial tissue sections were seen. The reasons for these differences in expression have not yet been fully elucidated.

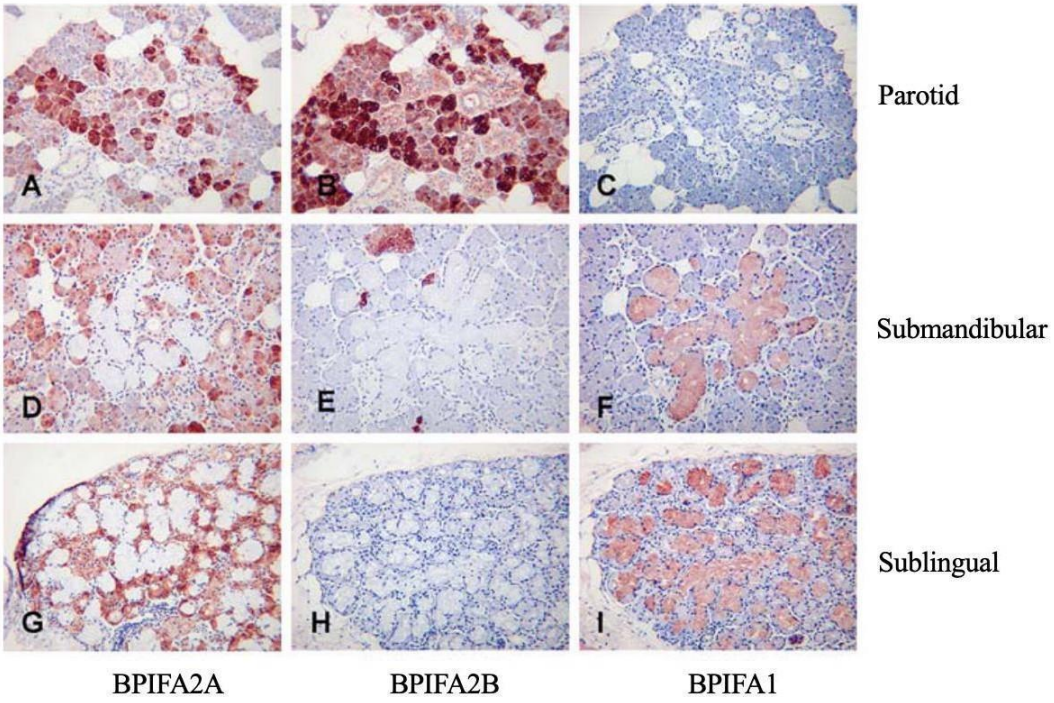


Figure 1.9. Distribution of BPIFA2 in major salivary glands. BPIFA2(A) antibody, raised to an epitope in the centre of the protein (A, D and G) and BPIFA2 (B), raised to an epitope at the C-terminus (B, E and H). Parotid (A-C), submandibular (D-F) and sublingual (G-I) glands were studied. Original magnification x100 (Bingle *et al*, 2009).

1.6.1 Function of BPIFA2

Although BPIF proteins are structurally similar to host defence proteins BPI and LBP, and BPIFA2 is known to be expressed in gingival keratinocytes in response to heat-killed bacteria and inflammatory cytokines such as TNF-α (Tumour necrosis factor-alpha) (Shiba *et al.*, 2005), there is a lack of comprehensive functional research on BPIFA2. The oral cavity serves as a convenient model and saliva offers an abundant supply of BPIFA2 natural protein for functional research purposes. Therefore, previous studies have focused on developing a purification technique for purifying native BPIFA2 from stimulated whole saliva and employed the purified BPIFA2 protein for functional assays. Numerous purification methods were used, including precipitation with ethanol/acetone and ammonium sulphate which have

previously been successfully used to purify BPIFA1 (Campos *et al.*, 2004), size exclusion chromatography and ion exchange chromatography. BPIFA2 was previously reported to be successfully purified by affinity chromatography and identified by trypsin mass fingerprinting from whole saliva (Prokopovic *et al.*, 2014), however, this could not be replicated in our lab using human saliva as a source of the protein (Lunn, 2014). The use of size exclusion chromatography to remove contaminating proteins led to a significant reduction in yield with similar issues occurring when ion exchange chromatography was used. Ammonium sulphate fractionation was not efficient in separating BPIFA2 from other salivary proteins while ethanol purification and the use of temperatures greater than 4°C have been demonstrated to cause protein deformation and denaturation (Lin *et al.*, 2004; Zellner *et al.*, 2005).

Although purification of BPIFA2 from human saliva has not yet been achieved bioactive peptides (GL13NH2 and GL13K) have been designed from a region of the human BPIFA2 sequence (141-153) and used in antibacterial functional assays (Gorr *et al.*, 2008, 2011). The GL13 peptides were shown to inhibit the activity of LPS (Geetha *et al.*, 2005), to reduce biofilm formation of *P. aeruginosa* (Hirt and Gorr, 2013) and induce agglutination in both Gram-negative bacteria (*P. aeruginosa* and *Aggregatibacter actinomycetemcomitans*) and Gram-positive bacteria (*S. gordonii*) (Gorr *et al.*, 2008).

A recent study demonstrated the utilisation of titanium surfaces coated with cysteine-terminated GL13K for dental implants, suggesting BPIFA2 has the antimicrobial ability to decrease bacterial viability with *S. gordonii* and *S. mutans* which are the main bacteria that colonise and cause diseases on dental surfaces. (Mutreja *et al.*, 2023). BPIFA2 has an affinity for hydrophobic silicone, and can enhance the attachment of *Candida albicans* to silicone (Holmes *et al.*, 2014). The adsorbed BPIFA2 layer can serve as a conditioning film, providing adhesion sites for *C. albicans* via hydrophobic and electrostatic interactions. And BPIFA2 is able to aggregate *C. albicans*, potentially promoting their adherence to surfaces. BPIFA2 has the ability to bind to LPS and can induce microbial agglutination (Gorr *et al.*, 2011). BPIFA2 aggregates bacterial cells by cross-linking LPS molecules on their surfaces, leading to clumping and immobilization of bacteria. BPIFA2's ability to bind LPS and agglutinate bacteria is crucial in preventing infections in the oral cavity.

BPIFA2 is a component of acquired enamel pellicle (AEP), (Siqueira et al., 2010) a biofilm that forms on the tooth surface and consists of bioadhesion and bio-adsorption proteins, glycoproteins and biomolecules in the oral cavity. Another form of pellicle, the mucosal pellicle, is composed of other glycoproteins including MUC5B and MUC7 (salivary mucins), MUC1 (and epithelial mucin) and secretory IgA (Gibbins et al., 2014). The AEP protects against acids produced by bacteria in the mouth. The mucosal pellicle plays important roles in lubrication and keeping the mouth healthy as well as against bacterial colonization (Bradway *et al.*, 1989; Ployon *et al*., 2016). Compared to the AEP, there are few studies on the mucosal pellicle, due to the difficulty in experimentally preparing the biological structure (Hannig *et al*., 2017).

It has been hypothesized that members of this family might share related biological functions due to their structural resemblance. In order to elucidate the true function of BPIFA2 it is pertinent for us to consider evidence suggesting roles for other family members. On the basis of the marked hydrophobicity of BPIFA1 and also the expression pattern, Bartlett *et al* (2011) hypothesised that BPIFA1 would act as an airway surfactant and that native BPIFA1 would influence maintenance of surface tension in airway epithelium (Bartlett *et al*., 2011). Schürch and colleagues (1990) suggested that low surface tension in airway secretions helps to immerse foreign particles on the airway surface. Thus, one of the functions of BPIFA1 in the airway may be to reduce the surface tension to a level that allows ingested bacteria and other small molecules to be incorporated into the liquid part of the airway surface and to facilitate their elimination by macrophages, or other components related to innate immunity, rather than direct killing of bacteria. While stimulated saliva obtained from *Bpifa2* knock-out (KO) mice exhibited reduced spreading ability on a hydrophobic surface compared to wild-type saliva and the surface tension of KO saliva was comparable to that of water (Nandula *et al.*, 2020). These findings suggest that BPIFA2 is a salivary surfactant that plays a crucial role in maintaining the low surface tension of mouse saliva. The BPIFA2 derived peptide, GL13NH2 was reported to be more closely related to physiologically active peptides found in surfactant protein and salivary agglutinin which is encoded by *deleted in malignant brain tumours 1* (DMBT1) gene compared to BPI due to 69% sequence similarity between DMBT1 and GL13 (Gorr *et al.*, 2008).

Future work should explore the potential for other BPIF family members to inhibit biofilm formation of other bacterial species and fully explore their potential surfactant activity. This

could be of particular importance when considering the role of BPIFA2 which could be involved in facilitating the organisation of the mucosal and/or enamel pellicle and controlling the growth of commensal versus non-commensal bacteria in the oral cavity.

1.6.2 BPIFA2 in disease

Research into differential expression of BPIFA2 and disease development has not been studied extensively, however, there are a small number of studies describing changes in levels of BPIFA2 in human saliva. For example, reduced levels of BPIFA2 were found in whole unstimulated saliva from patients with invasive periodontitis compared to healthy adults, suggesting BPIFA2 could be used in the development of a salivary-based test to assess the degree of disease (Wu *et al*., 2009). Low concentrations of BPIFA2 were detected in mucositis patients treated with radiotherapy for head and neck cancer (González-Arriagada *et al.*, 2015). Although these cases showed alterations in BPIFA2, the functional significance remains unclear. In contrast, acquired immunodeficiency syndrome (AIDS) patients with cytomegalovirus (CMV) infection and mycobacteriosis had a greater level of BPIFA2 expression compared to HIV (human immunodeficiency virus)-negative controls and HIV patients with chronic non-specific sialadenitis(Da Silva *et al.*, 2011). The staining intensity for BPIFA2 was higher in the staining lesions compared to the surrounding areas. CMV mostly infects ducts and fails to produce a granulomatous response. This could explain the increased staining observed in the outer region. Hence, BPIFA2 was suggested to have an anti- inflammatory function in protecting epithelial cells against infection. A study has suggested that BPIFA2 could be a potential biomarker for the early detection of Acute Kidney

Injury (AKI) as increased BPIFA2 expression was demonstrated in chronic kidney disease in both mice and humans. Previous studies have not shown expression of any BPIF-containing proteins in normal kidneys and so the reasons for this novel increase of BPIFA2 remains to be elucidated (Kota *et al*., 2017; Honore, De Bels and Spapen, 2018). A recent study showed decreased levels of BPIFA2 suggesting alterations in their immune response in patients with hyposalivation and chronic graft-versus-host disease (cGVHD) resulting from inflammation of salivary glands while another study suggested that BPIFA2 could be a potential biomarker for detecting radiation exposure at an early stage, as a notable rise in BPIFA*2* levels in mouse serum was found within the first 12 hours following radiation exposure (He *et al.*, 2022). Thus, investigations into the biological function of BPIFA2 will help to understand the possibility of using the protein as a biomarker for diagnosis of early-stage disorders and diseases.

1.7 BPIFA2 protein in mammals

BPIFA2 is the most complex and diverse member of the BPIF-containing family as multiple homologous and orthologous proteins have been detected in different species, such as cow (BPIFA2A-D), rat and mouse (BPIFA2E and BPIFA2F), hamster (BPIFA2E) pig, chimpanzee, gorillas, rhesus monkey (Gorr *et al.*, 2011). The expression of BPIFA2 (PSP) in rodent salivary glands has been studied extensively (Sivakumar *et al.*, 1998) as has the equivalent protein, BPIFA2 (BSP) in cows (Wheeler *et al.*, 2002). The BPIFA2 group of genes has clearly experienced an important level of divergence throughout mammalian evolution (Figure 1.5).

1.7.1 BPIFA2 proteins in rodent

BPIFA2 was identified as the human orthologue of rat and mouse parotid secretory protein despite their protein sequences sharing only 33% identity (Bingle *et al.*, 2009). Rodent BPIFA2 is a 23 kDa protein that was actually the first BPIF family member identified and cloned (Madsen and Hjorth, 1985; Shaw and Schibler, 1986; Ballli, 1992). With a similar distribution as the human protein BPIFA2, it is highly expressed in the parotid gland (Harrison and Zimmerman, 1984). Rodent BPIFA2 was identified to have potential function in host defence, as it was found to have binding ability to bacterial membranes (Robinson *et al*., 1997) and to have anticandidal activity (Khovidhunkit *et al*., 2005); a biological function has, however, not yet been fully developed. Interestingly, study of the BPIFA2 KO mouse model showed that not only did the saliva contain more amylase than in the wild type mice but also that the KO saliva was less able to spread on hydrophobic surfaces (Nandula *et al.*, 2020). Therefore, there could be a role for BPIF-containing proteins in controlling surface tension of liquids lining the oral cavity as well as the respiratory tract.

Submandibular gland protein B (SMGB), a 25 kDa protein, was the second PSP-related protein identified in rats (Ballli, 1992) and mice (Ball, Mirels and Hand, 2004). SMGB is not a direct orthologue of BPIFA2, but rather it is a rodent-specific member of the BPIF family that has been designated as BPIFA2F (Wheeler *et al.*, 2011).Like *Bpifa2*, *Bpifa2f* is expressed in submandibular glands and in sublingual and parotid glands of new-born rats. However, in adult rats, the expression of these two genes is opposite. *Bpifa2f* is expressed at much higher levels than *Bpifa2e* in sublingual glands, while *Bpifa2e* is more highly expressed in acinar cells of parotid glands (Ballli, 1992; Ball, Mirels and Hand, 2004).

Although the two proteins are the main products of the neonatal submandibular gland, they are not produced in the seromucous acinar cells of the adult parotid gland, with other key secretory proteins.

1.7.2 BPIFA2 proteins in cattle

Bovine salivary protein (BSP), a 30 kDa protein, is another PSP-related protein. BSP30 was originally proposed to be associated with cattle's susceptibility to pasture bloat, a metabolic disorder characterised by the build-up of persistent foam in the rumen. This foam accumulation disrupts the eructation process, causing rumen distension and respiratory issues (Clarke and Reid, 1974). The exact function of BSP30 was unclear, however, it was considered to contribute to the occurrence of bloat (Rajan *et al.*, 1996). BSP30 is now known to be four distinct proteins produced from four related genes originally known as *BSP30A*, *BSP30B*, *BSP30C* and *BSP30D*. They are now known as BPIFA2A to D. *Bpifa2a* and *Bpifa2b* are highly expressed in the parotid gland (Wheeler *et al*., 2002) and have been shown to have bacteriostatic activity against *P. aeruginosa* and *Streptococcus pneumoniae* through their high affinity to ligands derived from these pathogens rather than direct binding to LPS (Haigh *et al*., 2008). Similar antibacterial and anti-inflammatory effects have been described for human BPIFA2 (Geetha *et al.*, 2003, 2005) and it has been shown to agglutinate *P. aeruginosa* and *E.coli* (Gorr *et al*., 2008), suggesting that BPIFA2 and its paralogues have developed distinct functions during the evolution of distinct mammals. BPIFA2B was reported to bind to lipids present on the surface of bacteria in rumen, which suggests it is a lipid binding protein (Zhang *et al.*, 2019). BPIFA2C and BPIFA2D were identified following genomic analysis of

duplication events (Wheeler *et al.*, 2007). Presumably, BPIFA2 in cattle serves a purpose within saliva as it is expressed by the salivary glands and locally in the rumen, as the ruminant lifestyle requires large volumes of saliva to be produced to age digestion. Further research is needed to fully elucidate the involvement of BPIFA2 proteins in the physiology of ruminant digestion and its potential implications for rumination and overall digestive health in cattle.

1.7.3 BPIFA2 proteins in great apes

There are limited studies of BPIFA2 proteins in great apes, which might be considered a model lineage to investigate biological evolution. As shown in Figure 1.10, whole saliva collected from great apes contains considerably more BPIFA2 than from human saliva, even though the primates share a similar sequence to humans. This interesting finding suggests that humans may not require large amounts of the protein in their saliva and this species difference could perhaps suggest further studies aimed at elucidating the true function of the protein. In this context it is also worth highlighting again that *BPIFA4* is a salivary gland expressed pseudogene in man but is fully functional in primates suggesting that its loss may also be associated with lifestyle adaptations that have occurred during recent evolution.

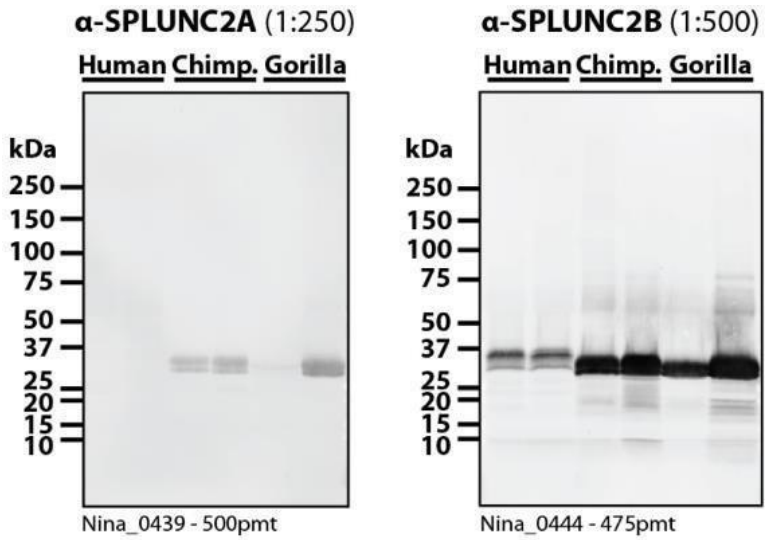


Figure 1.10. Comparison of BPIFA2 levels in humans and great apes analysed by immunoblot. This shows saliva samples from two donors probed with our two different BPIFA2 antibodies (Unpublished data shared by Professor Stefan Ruhl, University at Buffalo).

1.8 Hypothesis and Aims

BPIFA2 is a highly glycosylated protein that is produced from both major and minor salivary glands. In man it is essentially a salivary gland specific protein. The exact function of many BPIF protein family members remains unknown. BPI and LBP are LPS binding proteins, and can control the growth and behavior of pathogenic bacteria including by direct binding, opsonisation to enhance phagocytosis, reduction of the immune response to inhibit levels of inflammation or by directly preventing bacterial growth and adhesion. These are all key functions required within the oral cavity. Due to the similarities between the predicted structure of BPIFA2 and that of BPI and LBP, I hypothesise that BPIFA2 will have similar functions to BPI and LBP.

Hypothesis:

Our central hypothesis is that BPIFA2 is a host defence protein in saliva and plays a role in antimicrobial defence of the mouth.

The principal aim of my project, therefore, is to determine the function of BPIFA2 in human saliva, to be achieved through a number of more specific aims:

1. To provide a comprehensive bioinformatic analysis of BPIFA2 including comparison of multiple sequences across mammals and undertake a human genetic variant analysis (described in chapter 3),

2. To analyse BPIFA2 in the unstimulated saliva of healthy individuals and to synthesise and analyse recombinant BPIFA2 from a number of species (described in chapter 4).

3. To undertake functional analysis of recombinant BPIFA2 with lipids, microbes and macrophages (described in chapter 5).

Chapter 2: Materials and methods

2.1 Cell culture

Human embryonic kidney 293 (HEK 293) cells provide a useful tool for research as they are relatively easy to transfect and allow the production of recombinant proteins that include correct folding and structure of the proteins and also, importantly for this project, post-translational modifications such as glycosylation.

A monocytic cell line, Mono Mac 6 (MM6), was used to determine the interactions between BPIFA2 proteins and immune cells.

The cell lines are shown in Table 2.1.

Table 2.1. Cell lines.

Cell line	Tissue original	Provided by
HEK 293	Established from human embryonic kidney	Dr. Lynne Bingle (Academic Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, UK)
Mono Mac 6	Established from peripheral blood of a patient with monoblastic leukaemia	Dr. Lynne Bingle (Academic Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, UK)

2.1.1. Culture media and conditions

HEK 293 cells were cultured in low glucose DMEM (Sigma-Aldrich) supplemented with 10% (v/v) foetal bovine serum (FBS; Sigma), 2mM L-glutamine (Sigma-Aldrich), 100 μg/ml

Penicillin and 100 U/ml Streptomycin (Sigma- Aldrich) and incubated at 37°C in a humidified 5% CO_2 incubator.

MM6 cells are non-adherent and were cultured in RPMI 1640 L-Glutamine (Sigma-Aldrich) supplemented with 200 U/ml Penicillin, 200 µg/ml Streptomycin (Sigma- Aldrich), 1% (v/v) 100X non-Essential amino acids (NEAA; Sigma- Aldrich), 1% (v/v) OPI-supplement (Sigma-Aldrich) and 10% (v/v) foetal bovine serum (FBS; Sigma). Mono Mac cells were incubated at 37°C in a humidified 5% CO_2 incubator.

2.1.2. Trypsinization and subculturing of cells

HEK 293 cells were cultured in T75 flasks, the media changed every two days and cells subcultured by trypsinisation when they had reached 80% confluency. Cells were washed twice with Ca^{2+}/Mg^{2+}-free PBS and incubated in 0.05% trypsin/EDTA at 37°C for 5 minutes to allow detachment from the flask surface. 10ml of DMEM media containing 10% (v/v) foetal bovine serum (Sigma) was added in order to neutralize the trypsin. (Sigma-Aldrich). Cells were pelleted by centrifugation at 1,000x *g* for 5 minutes.

MM6 cells were cultured in T75 flasks, the media changed every week and cells subcultured when they had reached 80% confluency. As MM6 cells are non-adherent they were subcultured by taking 1ml of cell suspension and transferring to a new flask containing appropriate growth medium.

2.1.3. Counting cells

4.5 x 10^6 cells of HEK 293 cells were plated into a T75 flask to ensure the correct confluency at the correct time, for successful transfection. Cells were trypsinised and pelleted according to the protocol given in section 2.1.2. The cell pellet was resuspended with 10ml of DMEM and 10 µl of this was added to a haemocytometer to count the cells.

The following formula was used to calculate the number of cells per millilitre.

Concentration (cells/ml) = (Total number of cells x Dilution Factor 10^4)/4 (square)

Total number of cells= Concentration (cells /ml) x Volume of the sample (ml)

2.1.4. Freezing cells

After trypsinisation (see section 2.1.2), the cells were counted (section 2.1.3) to acquire a cell density of 1x 10^6 cells/ml. Cells were mixed with cryopreservation medium, normal growth medium (as prepared in section 2.1.1) with the addition of 10% Dimethyl sulfoxide (DMSO; Thermo Fisher Scientific) and aliquoted into 1ml cryovials, prior to storage in a Mr. Frosty overnight at -80°C. All cryovials were then transferred into liquid nitrogen for long-term storage.

2.2 Protein analysis

2.2.1 Human samples

Whole saliva samples were collected from 18 healthy volunteers; the distribution of age and

gender is shown in Table 2.2. The saliva was collected according to University of Sheffield ethics approved project 040965. Detailed information regarding the protocol and consent forms used in this study is given in the appendix as described the participant consent form). All volunteers were recruited by emails and gave informed consent to participate in the research study. All volunteers were healthy and free of oral diseases on the day of collection. Eating and drinking was not permitted for 2 hours prior to collecting samples and the collection time was always between 9am and 9.30am. Unstimulated saliva was collected and purified by centrifuging at 13,000rpm for 10 minutes to remove cells and debris. Equal volumes of purified saliva samples were mixed with SDS lysis buffer (2X) and stored at -20°C for further experiments.

Table 2.2. Human saliva collection.

Volunteers Sample Number	Gender	Age
1	Female	30
2	Male	23
3	Female	32
4	Female	38
5	Male	50
6	Female	36
7	Female	60
8	Female	33
9	Female	50
10	Male	55
11	Female	23
12	Female	26
13	Female	48
14	Female	40
15	Female	40
16	Male	27
17	Female	60
18	Male	28

2.2.2 Antibody generation

Previously, BPIFA2 antibodies were generated (Eurogentec) by injecting two synthetic peptides 156-168: (VTIETDPQTHQPV) (designated BPIFA2A) and 236–249 (VDNPQHKTQLQTLI) (designated BPIFA2B) into rabbits. The pooled final serum was used for purification against each single peptide, resulting in specific antibodies against each unique peptide (Bingle *et al.*, 2009b). The antibody was validated by transcription/translation *in vitro.*

The antibodies used for Western blotting, dot blotting and enzyme-linked immunosorbent assay (ELISA) are summarised in Table 2.3.

Table 2.3. Antibodies used for Western blotting and dot blotting. Working dilution and secondary antibody are listed.

Primary antibody	Manufacturer	Description	Working dilution	Secondary antibody
BPIFA2A	Eurogentec	Rabbit monoclonal to human BPIFA2	1:250	Anti-rabbit- HRP
BPIFA2B	Eurogentec	Rabbit monoclonal to human BPIFA2	1:500	Anti-rabbit- HRP
Anti-FLAG	Sigma	Mouse monoclonal to FLAG-tagged proteins	1:1000	Anti-mouse- HRP

2.2.3 Western blotting

All protein samples analysed by Western Blotting were lysed in 2X SDS lysis buffer (Table 2.4) and heated at 95°C for 5 minutes prior to loading. A 12% or 15% acrylamide gel was used depending upon the level of protein separation required. The recipes for making gels are outlined in Tables 2.5 to 2.8. Gels were subject to 100V for 15 minutes, 150V for 45 minutes and then 200V to ensure separation of proteins. Proteins were transferred to a nitrocellulose membrane under the following transfer conditions: 1.0A and 25V for 30 minutes, using a Bio-Rad, Trans-Blot Turbo. Ponceau S (Sigma Aldrich) was used to stain the membranes to confirm full transfer of proteins. 5% skimmed milk in Tris-buffered saline (1X TBS) was used to block the membrane overnight at 4°C on a shaking platform. Primary antibody (Table 2.3) was diluted in 5% skimmed milk in 1X TBS with 0.1% Tween 20 (1X TBS-Tween) and incubated with the membrane at room temperature on a shaking platform for one hour. The membrane was washed three times for 12 minutes in 1X TBS-Tween and then incubated with a horseradish peroxidase (HRP) conjugated secondary antibody (Table 2.3) for 15 minutes on a rolling platform at room temperature. The membrane was washed 3 times for 12 minutes in 1X TBS-Tween. Enhanced Chemiluminescence (ECL) Western Blotting substrate (Thermo Fisher Scientific) and a Li-Cor C-Digit Western Blot Scanner was used to visualise proteins. A Compact X4 automatic X-ray film processor (Xograph) was also used to visualise signals following exposure to X-ray film (Thermo Fisher Scientific). ImageJ was used for band quantification and densitometry analysis was performed as described in the section 2.2.5.

Table 2.4. The constituents of SDS lysis buffer.

Reagent	Weight/ volume
Bromophenol Blue	0.05g
Protease Inhibitor tablets	1 tablet
1M DTT	1ml
0.5M Tris-HCl pH6.8	1.25ml
10% SDS	2ml
Glycerol	2ml
Distilled water	3.55ml

Resolving gel (ml)

Table 2.5. Formula for solutions to make resolving gel of 2x 1.00mm SDS PAGE gels using 30% Acrylamide.

%Gel	Acrylamide (30%)	Lower Tris	H_2O
12 ml	3.75ml	2.5ml	3.55ml
15 ml	4.687ml	2.5ml	2.613ml

Table 2.6. Formula for solutions to make resolving gel of 2x 1.00mm SDS PAGE gels using 40% Acrylamide.

%Gel	Acrylamide (40%)	Lower Tris	H_2O
12ml	3ml	2.5ml	4.3ml
15ml	3.75ml	2.5ml	3.55ml

5µl Tetramethylethylenediamine (TMED, Thermo Scientific Pierce) and 350µl 10% ammonium persulfate (APS) were added to the acrylamide/tris/H_2O mixture. Gels were poured immediately. Acrylamide solution was poured up to 3/4 of the total height of the glass plates and overlayed with isopropanol until the gel polymerized. The overlay was washed twice with distilled water before pouring the stacking gel.

Stacking gel (ml)

Table 2.7. Formula for solutions to make stacking gel of 2x 1.00mm SDS PAGE gels using 30% Acrylamide.

Acrylamide (30%)	Upper Tris	H_2O
1.219ml	2.1ml	4.48ml

Table 2.8. Formula for solutions to make stacking gel of 2x 1.00mm SDS PAGE gels using 40% Acrylamide.

Acrylamide (30%)	Upper Tris	H_2O
0.975ml	2.1ml	4.725ml

– 17µl TMED and 100µl 10% APS were added to the acrylamide/tris/H_2O mixture. The gel was poured just short of the top of the glass plates and the sample well comb inserted; the gel was then left to polymerise.

2.2.4 Dot blotting

1µl and 2µl of recombinant proteins were dotted separately on PVDF membranes (Millipore) and air-dried for 5 minutes to ensure samples were entirely dry. 5% skimmed milk in 1X TBS was used to block the membranes for 1 hour on a shaking platform. Primary antibody (Table 2.3) was diluted in 5% skimmed milk in 1X TBS-Tween and incubated with the membranes at room temperature on a shaking platform for one hour. The membranes were washed three times for 12 minutes in 1X TBS-Tween and then incubated with a HRP conjugated secondary antibody (Table 2.3) for 15 minutes on a shaking platform at room temperature. The membranes were washed 3 times for 12 minutes in 1X TBS-Tween. Enhanced Chemiluminescence (ECL) Western Blotting substrate (Thermo Fisher Scientific) and a Li-Cor C-Digit Western Blot Scanner were used to visualise dots. ImageJ was used for dot quantification and densitometry analysis was performed as described in section 2.2.5.

2.2.5 Densitometry

Western blots and dot blots were developed using the Li-Cor C-Digit Western Blot Scanner and analysed by densitometry using ImageJ software. Bands and dots were manually highlighted and the signal within each area of focus was measured. GraphPad Prism 9 software was used to determine statistical significance.

2.2.6 SDS-PAGE analysis

A 12% acrylamide gel was used to visualise the successful purification of recombinant protein preparations using Coomassie blue staining.

2.2.6.1 Coomassie blue staining

0.1% Coomassie R-250 was prepared in 40% ethanol, 10% acetic acid. After gel electrophoresis, the SDS-PAGE gels were rinsed 3 times for 5 minutes with 50ml deionized distilled water to remove SDS and buffer salts. Gels were then fully immersed in Coomassie blue stain solution for 2 hours at room temperature with gentle shaking. A destain solution containing 10% ethanol and 7.5% acetic acid was then used to wash the stained SDS-PAGE gels at room temperature with gently shaking, for 1-3 hours or until the desired background was visualised using the InGenius3 gel documentation system (Syngene).

2.2.7 BCA assay

Bicinchoninic acid (BCA) protein assay (Thermo Scientific, TM Pierce) was used according to the manufacturer's instructions to quantify recombinant protein. Bovine Serum Albumin (BSA)was used to construct a standard curve. Absorbance was detected at 570nm using a Tecan infinite M200 microplate reader.

2.2.8 Enzyme-linked immunosorbent assay

The concentration of recombinant BPIFA2 proteins was quantified using a DYKDDDDK-Tag Protein ELISA Kit (Abcam, ab285234). The ELISA was carried out following the manufacturer's protocol. DYKDDDDK Standard provided in the kit was used to construct a standard curve. Absorbance was detected at 450nm using a Tecan infinite M200 microplate reader.

2.3 Synthesis of recombinant BPIFA2 protein

Constructs for the production of recombinant BPIFA2 proteins (WT, 124 and 132) were previously established and verified by the Bingle group and stored at -80°C as glycerol stocks. These constructs were made in the mammalian expression vector pVR1255 and each had an in- frame FLAG-Tag at the C-terminus. BPIFA2 is N-glycosylated at asparagine (N)-124 and asparagine-132 and so to determine if glycosylation is related to function these sites were mutated to alanine (A) to produce proteins without the individual, or both glycosylation sites for comparison with the wide type protein. A new construct for the double mutant was generated alongside a further five constructs from the mouse, armadillo, macaque, squirrel

monkey and dog were generated to study the function of BPIFA2 across the mammalian phylogeny.

2.3.1 Plasmid Midi prep kit isolation

A ZymoPURE II Plasmid Midiprep kit was utilised according to the manufacturer's instructions to purify DNA from an overnight bacterial culture (as described in section 2.5.2) which was initially centrifuged at 6,000x g for 15 minutes at 4°C (Ohaus). 8ml ZymoPURE™ P1 was added to the bacterial pellet which was resuspended completely by pipetting. 8ml of ZymoPURE™ P2 was added to the mixture and immediately mixed by gently inverting the tube 6 times before incubation at room temperature for 3 minutes. 8ml of ZymoPURE™ P3 was then added and mixed gently by inverting the tubes until the sample was completely yellow. 10ml of this mixture was added to the 15ml Reservoir-X/Zymo-Spin™ V-PS Column and centrifuged at 500x *g* for 2 minutes. This step was repeated until the entire sample had passed through the column and then 5ml of ZymoPURE™ Wash 1 was added to the Zymo-Spin™ V-PS Column and centrifuged at 500x *g* for 2 minutes. The flow through was discarded, 5 ml of ZymoPURE™ Wash 2 was added to the Zymo-Spin™ V-PS Column and centrifuged at 500x *g* for 2 minutes. The flow through was discarded and the 15ml Reservoir-X removed from the Zymo-Spin™ V-PS Column, which was then placed in a collection Tube, centrifuged at 16,000x *g* for 1 minute to remove residual wash buffer and this step repeated with a clean 1.5ml Eppendorf tube. The Zymo-Spin™ V-PS Column was transferred to a clean 1.5ml tube and 200µl of ZymoPURE™ Elution Buffer was added to the column membrane and incubated at room temperature for 2 minutes to dissolve the DNA. The column

was centrifuged at 16,000x *g* for 1 minute at 4°C and DNA yield was quantified using a Thermo Scientific NanoDrop™ 1000 Spectrometer (Thermo Fisher Scientific).

2.3.2 TOPO TA subcloning

Topoisomerase-based cloning (TOPO cloning) technique was used to subclone amplified BPIFA2 products (Biomatik constructs for BPIFA2 glycosylation double mutant (124/132), armadillo, dog, macaque, squirrel monkey and mouse) directly into a PCR®II- TOPO® 4.0 kb vector (Thermo Fisher Scientific). Cloning was carried out according to the manufacturer's instructions. Plasmids were introduced into an *E. coli* strain (Invitrogen) containing the cloned topA gene (TOPO TA cloning) by transformation, and the transformed cells incorporating the plasmids were selected by growing on LB-agar plates supplemented with ampicillin (50 μg/ml). The transformed colonies were selected and cultured in 2ml of LB-broth supplemented with ampicillin (50 μg/ml) at 37°C and 220rpm overnight. The plasmid DNA was extracted from the bacterial culture using Monarch® Plasmid Mini Kit, and the amount of plasmid DNA obtained was quantified using a Thermo Scientific NanoDrop™ 1000 Spectrometer (Thermo Fisher Scientific).

2.3.3 Plasmid Mini prep kit isolation

The overnight culture described in section 2.3.2 was used for plasmid DNA isolation using a Monarch® Plasmid Mini Kit (NEB) based on the manufacturer's instructions. DNA yield was quantified using a Thermo Scientific NanoDrop™ 1000 Spectrometer (Thermo Fisher Scientific) and stored at -20°C.

2.3.4 Plasmid Maxiprep kit isolation

To obtain higher DNA yields, 100 ml of LB-broth containing kanamycin (50 μg/ml) was cultured with 0.5ml of transformant culture from stocks stored at 4°C. The bacterial culture was incubated at 37°C and 220rpm overnight and this overnight culture used for plasmid DNA isolation using a Monarch® Plasmid Maxi Kit (NEB) based on the manufacturer's instructions. DNA yield was quantified using a Thermo Scientific NanoDrop™ 1000 Spectrometer (Thermo Fisher Scientific) and aliquoted and stored at -20°C.

2.3.5 Isolation of insert DNA and vector preparation

2.3.5.1 Human BPIFA2 double mutant construct

The target vector was the pVR1255 described below (Figure 2.1). The pVR1255 vector, BPIFA2 insert from the wild-type BPIFA2 plasmid and Bluescript plasmid were double digested with BamHI and NotI restriction enzymes (NEB) as showed in Table 2.9 and incubated at 37°C overnight. A 1% agarose gel was used to detect the fragment size of human BPIFA2 (~750 bp) and empty VR1255 (~5-6 kb). The target vector and insert were isolated using a Wizard® SV Gel and PCR Clean-Up Kit (Promega). DNA yield was determined using a Thermo Scientific NanoDrop™ 1000 Spectrometer (Thermo Fisher Scientific).

2.3.5.2 Animal BPIFA2 constructs

The pVR1255 vector was obtained from the wild-type BPIFA2 plasmid double digested with BamHI and NotI restriction enzymes. The armadillo and dog BPIFA2 inserts were double digested with BgIII and NotI restriction enzymes and macaque, mouse and squirrel monkey BPIFA2 inserts were double digested with BamHI and NotI restriction enzymes (Table 2.9)

and incubated at 37°C overnight. A 1% agarose gel was used to detect the ~772 bp fragment of mouse BPIFA2 and the 814 bp armadillo, dog, macaque and squirrel monkey inserts. The target inserts were isolated using a Wizard® SV Gel and PCR Clean-Up Kit (Promega). DNA yield was determined using a Thermo Scientific NanoDrop™ 1000 Spectrometer (Thermo Fisher Scientific).

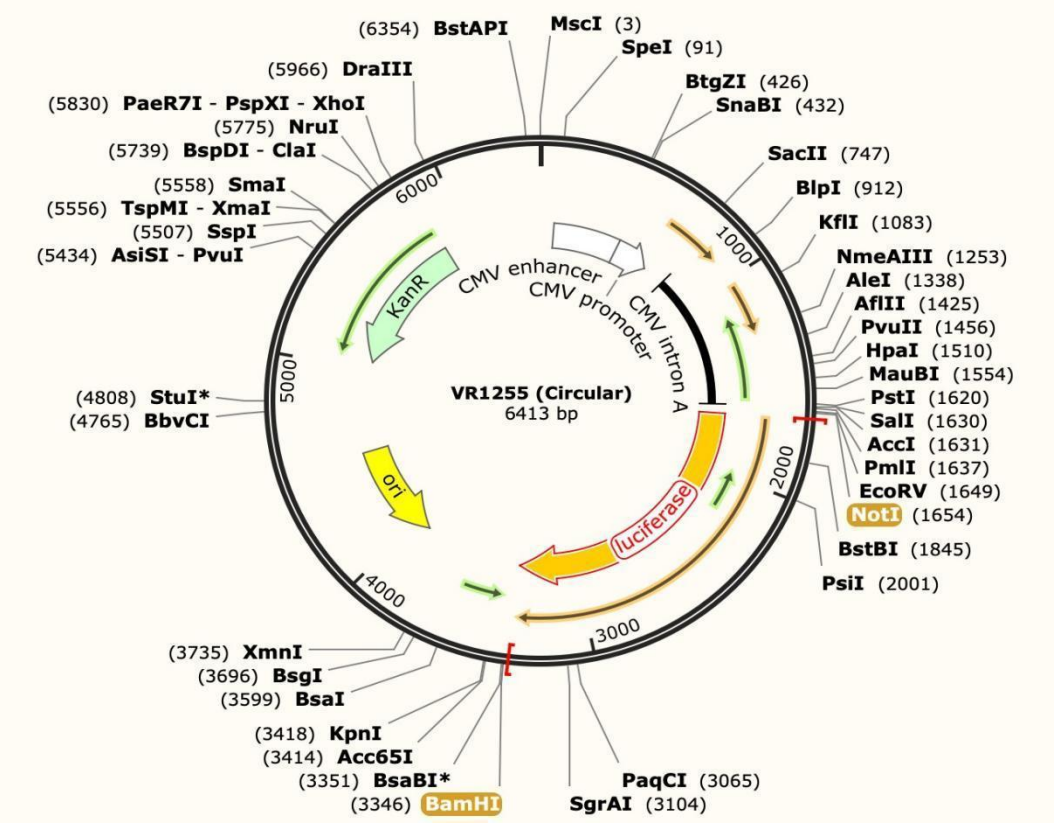


Figure 2.1. Plasmid map of target vector pVR1255. The restriction sites and kanamycin resistance gene (Kan) are shown. BamHI and NotI were used to produce the VR1255 vector, and the luciferase gene (LUX) was removed and replaced with compatible sticky terminals for BPIFA2 insertion. The size of the plasmid vector is 6413 bp.

Table 2.9 Restriction enzyme sites utilised for Biomatik construct digestion.

Construct name	Restriction enzyme site 1	Restriction enzyme site 2
Armadillo	BgI II	Not I
Dog	BgI II	Not I
Macaque	BamH I	Not I
Mouse	BamH I	Not I
Squirrel monkey	BamH I	Not I
Double mutant hBPIFA2	BamH I	Not I

2.3.6 Ligation

A portion of the BPIFA2 insert obtained from gel extraction, as outlined in section 2.3.5, was loaded onto a 1% agarose gel. A PCR Ranger 100 bp DNA ladder (Norgen) was used to compare reference bands. The BPIFA2 insert was ligated into the double cut VR1255 using T4 DNA ligase (Promega) and Ligase 10X Buffer (Promega) following the manufacturer's protocol in a total volume of 10µl.

2.3.7 Transformation

12.5µl *E. coli* (Invitrogen) was mixed with 5 ml ligation mixture and incubated on ice for 10 minutes. The mixture was heat-shocked at 42°C in a water bath for 45 seconds. 100µl Super Optimal Broth (SOB) medium was added and incubated at 37°C for 30 minutes before plating onto LB-agar plates containing kanamycin (50 µg/ml) and overnight incubation at 37°C. Individual colonies were selected and incubated in 3 ml LB-broth with kanamycin (50 µg/ml)

at 37°C and 220rpm overnight. Glycerol stocks were prepared by adding 0.5ml of bacterial culture to 0.5ml of 50% glycerol and stored at -80°C.

2.3.8 Confirmation of target insertion

The plasmid DNA was digested by restriction enzymes and the expected sizes of the fragments were detected using agarose gel electrophoresis to ensure the insert had been successfully ligated with the vector as described in Figure 2.1.

2.3.9 Sequencing of clones

DNA sequencing was used to confirm the accuracy of the nucleotide sequence of the BPIFA2 insert in the VR1255 vector (Research Core Facility, Medical School, University of Sheffield, UK). Submitted the prepared sample to the sequencing facility and perform the Sanger sequencing and verified the obtained sequence with the reference sequence or expected template using alignment software (BLAST). The sequence of primers is described in Table 2.10.

Table 2.10. The sequence of primers used for confirmation of BPIFA2 insert.

Primer	Sequence (5'-3')
VR1255Rnew	TGAGTGGGGCTGGAACTAAG
VR1255FSeq	AATAGCTGACAGACTAACAGACTG

2.3.10 Calcium phosphate transfection

Transfections were carried out in mammalian cells to produce recombinant protein, as this should allow for correct protein folding and important post-translational modifications, such as glycosylation, of significance to our proposed downstream experiments. 4.5 x 10^6 cells were seeded into a T75 flask and incubated at 37°C, 5% CO_2 overnight. A media change, to low glucose Dulbecco's Modified Eagle Medium (DMEM; Sigma-Aldrich) containing 10% (v/v) foetal bovine serum (FBS; Sigma), 2mM L-glutamine, 100 µg/ml of Penicillin and 100 U/ml of was carried out two hours prior to transfection. Nuclease-free H_2O, 2X HEPES buffered saline (HBS, Table 2.11) and 2.5 M $CaCl_2$ (Table 2.12) were sterile-filtered using 0.22µm syringe filters (Sartorius) before transfection. HBS is used to maintain the pH of the cell culture and mixed with a $CaCl_2$ solution containing the DNA. DNA and calcium phosphate co- precipitate on the surface of the cell, quickly and accurately, and are taken up by the cell by endocytosis. BPIFA2 (15µg) and pEGFP-N1 (3.75µg) were mixed with 140.6µl of $CaCl_2$ and the total volume made up to 1.125ml with filtered nuclease-free H_2O. 1.125ml of 2X HBS was added to the plasmid-$CaCl_2$ and after leaving to stand for 15 minutes at room temperature the mixture was added dropwise to the cells and incubated overnight at 37°C, 5% CO_2. After 24 hours, the media was replaced with fresh DMEM with 10% (v/v) FBS. After a further 24 hours, cells were washed three times with PBS and phenol red-free, serum-free DMEM added prior to a further 24 hours incubation. The media was then collected (conditioned medium) and replaced with phenol red-free, serum-free DMEM; this process was repeated every 24 hours for a further three days for further collections of conditioned medium, which was kept at -20°C for use within 1 month or used immediately after

concentration for Anti-FLAG resin purification.

Table 2.11 Recipe for 2X HBS solution.

Reagent	Weight(gram)
Sodium chloride (NaCl)	8g
Sodium phosphate dibasic heptahydrate (Na_2HPO_4-$7H_2O$)	0.2g
4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid (HEPES)	6.5g
500ml of distilled water was added, and the pH was adjusted to 7.0.	
10ml aliquoted into 15ml sterile tubes and stored at -20°C.	

Table 2.12 Recipe for 2.5 M $CaCl_2$ solution.

Reagent	Weight and Volume
Anhydrous calcium chloride ($CaCl_2$, Sigma)	27.27g
Distilled water	100ml
5ml aliquoted into 7ml sterile bijou tubes and stored at -20°C.	

2.3.11 Conditioned media concentration

A centrifugal concentrator column (Sartorius) with a molecular weight cut off 10 kDa was used to concentrate the conditioned media at 6,000rpm. Concentration continued until the

original volume of 20ml was reduced to 6ml.

2.3.12 Anti-FLAG M2 affinity gel purification

The recombinant BPIFA2 proteins have a FLAG-tag sequence (DYKDDDDK) enabling the use of anti- FLAG M2 affinity gel (Sigma) for purification. The cell remnants from conditioned media were discarded by centrifugation at 15,000×*g* for 15 minutes followed by filtration through 0.22µm filters. The resin was washed with 0.1M glycine-HCl (pH3.5) and Tris buffered saline (TBS; pH7.5) prior to use. 100µl of affinity gel was added to 6ml phenol-red free conditioned media and mixed at 4°C overnight on a roller. The resin was washed to elute proteins which had not bound but importantly all washes were retained to assess loss of recombinant protein. Washing continued until absorbance at 280nm was less than 0.05 compared to the blank solution. Bound protein was eluted from the affinity gel with 150µl of FLAG peptide (150 ng/µl; Sigma) by incubating at 4°C overnight with gentle rolling. The resin was centrifuged, and the supernatant was transferred into a fresh 1.5ml Eppendorf tube. Eluted protein was aliquoted and stored at -20°C. Western blotting was used to confirm the presence of proteins.

2.4 Glycosylation analysis

2.4.1 PNGase F enzyme treatment

PNGase F enzyme was used to remove carbohydrate residues from recombinant proteins. 9µl glycoprotein sample was combined with 1µl of glycoprotein denaturing buffer (10X) and nuclease-free H_2O (Biolabs) was added to a final volume of 10µl. Glycoproteins were

denatured by heating to 100°C for 10 minutes, chilled on ice and centrifuged for 10 seconds. The reaction volume was increased to 20µl by adding 2µl PNGase F enzyme, 2µl 10% NP-40, 2µl 10X glycobuffer 3 and nuclease-free H_2O (Biolabs) and incubated at 37°C for 1 hour. Untreated samples were compared to PNGase F treated samples by Western blotting.

2.5 Bacteria assays

2.5.1 Bacterial strains

Table 2.13 Bacterial strains.

Bacteria strain	Growth Conditions	Bacterial infection category
Streptococcus mutans (Ingbritt)	Anaerobic/CO2	Gram-positive
Streptococcus gordonii (Challis)	Anaerobic/CO2	Gram-positive
Pseudomonas aeruginosa (Clinical strain)	Aerobic	Gram-negative
Staphylococcus aureus (Oxford)	Aerobic	Gram-positive
Escherichia coli (U125643)	Aerobic	Gram-negative

2.5.2 Bacterial culture

7.2g Columbia agar powder was added to 200ml distilled water and sterilised by autoclaving. The agar was cooled to 55°C in a water bath and 5-7% (15ml) of room temperature defibrillated horse blood added. 20ml of agar was aliquoted into each Petri dish and the plates left to set and dry at 50°C for 10-15 minutes. Plates were then stored at 4°C.

14.8g of Brain-Heart Infusion (BHI) powder (Sigma-Aldrich) was added to 400ml of distilled water. 20ml aliquots were prepared for autoclaving.

LB broth was prepared by adding 7.5g LB broth powder (Fisher Chemical) to 300ml of distilled water and sterilised by autoclaving. 25ml aliquots were added to 50ml tubes and 25µl kanamycin (50 mg/ml stock concentration) added. A disposable loop was used to transfer bacterial samples into the broth. All liquid cultures of bacteria were incubated in a shaking (200rpm) incubator overnight at 37°C.

2.5.3 Pull down assay

A pull-down assay was used to determine whether BPIFA2 protein binds specifically to bacterial strains. BHI broth cultures of *S. gordonii, P. aeruginosa, S. aureus, E. coli,* and *S. mutans* were cultured overnight and absorbance (600nm) adjusted to an optical density (OD) of 1 by diluting with BHI media. 1ml of bacterial culture was centrifuged at 13,000 rpm for 5 minutes, the aqueous layer discarded, and the pellet washed twice with 1ml PBS by centrifugation at 13,000rpm for 5 minutes. 20µl of each BPIFA2 recombinant protein, saliva (previously clarified through centrifugation; positive control) or PBS (negative control) was

added to the bacterial pellet and the samples incubated for 1 hour at 37°C. The samples were twice washed with 1ml PBS by centrifugation at 13,000rpm for 5 minutes before transfer into a glass vial, as BPIFA2 has been shown to bind to plastic Eppendorf tubes and washed twice with 1ml PBS by centrifugation at 13,000 rpm for 5 minutes. 20µl PBS was used to resuspend the pellet before the addition of 20µl of 2X SDS lysis buffer. The samples were boiled at 95°C for 5 minutes and subjected to Western blotting. The processes of pull-down assay are described in Figure 2.2.

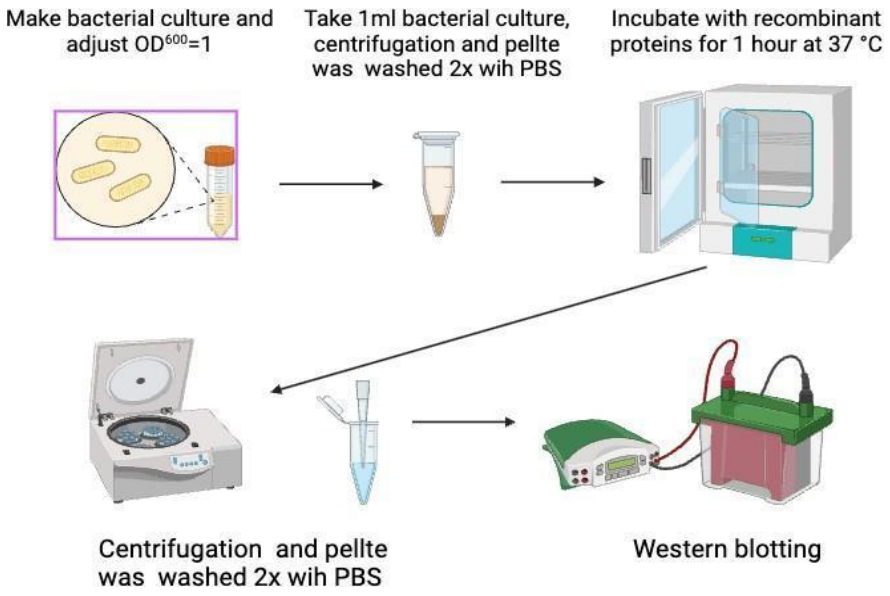


Figure 2.2. Diagram of pull-down assay protocol.

2.5.4 Agglutination assay

S. gordonii, P. aeruginosa, S. aureus, E. coli and S. mutans strains were cultured in BHI broth overnight and the optical density adjusted to an OD^{600} of 1 with BHI media. Whole saliva centrifuged at 13,000rpm to remove debris was used as an agglutination positive control and sterile PBS was used as negative control. 1mL of bacterial suspension was washed in PBS by

centrifugation at 13,000rpm for 5 minutes and then resuspended in 1 ml of PBS. Each 1ml aliquot of bacterial suspension was treated with 1µl of fluorescein isothiocyanate (FITC; 1mg/ml; Sigma) and kept in the dark for 15 minutes at 4°C with agitation. The bacteria were resuspended in 1.5mL PBS and washed four times with 1.5ml PBS to remove excess FITC label and finally resuspended in 1 ml PBS. 100µl of FITC-stained bacterial suspension was centrifuged at 13,000rpm for 5 minutes and the supernatant discarded. The pellet was resuspended in 100µl of purified saliva, purified BPIFA2 recombinant proteins or PBS. Each bacterial suspension was added to a U-bottomed 96-well plate (Greiner) and plates were kept at 4°C in the dark overnight. Images were captured using a Gel Documentation and Analysis System (Syngene). GelDoc Go Imaging System (Bio-Rad) were used to visualize the agglutination patterns of all bacterial strains were performed to the same processes without FITC staining. The processes of the agglutination assay are described in Figure 2.3.

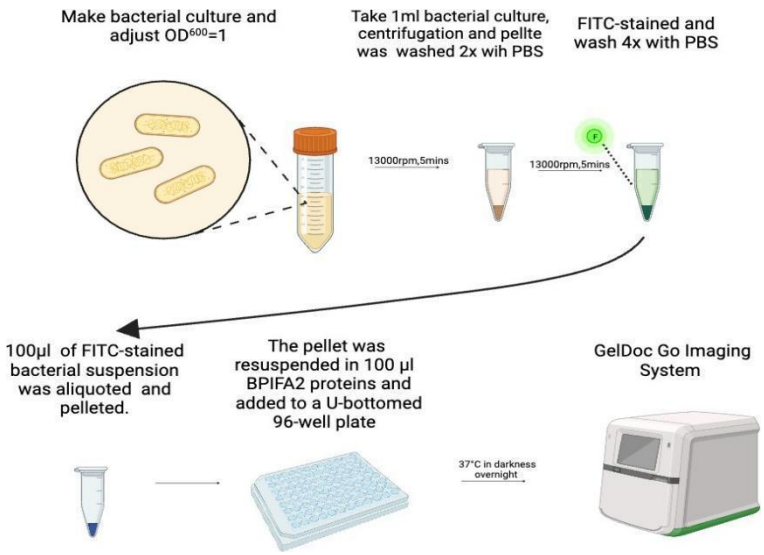


Figure 2.3. Diagram of agglutination assay protocol.

2.5.5 Biofilm formation assay

The optical density of an overnight broth culture was set to an OD^{600} of 1.0. 100µl of the bacterial culture was pipetted into individual wells of a 96-well plate and 100µl purified recombinant BPIFA2 proteins or an equal volume of sterile PBS (negative control) or whole saliva centrifuged at 13,000rpm to remove cells and debris (positive control) were immediately added to the bacteria. The sterile BHI broth was incubated alone to determine background staining and was removed from all absorbance measurements. Plates were placed in an incubator at 37°C for 48 hours to facilitate the formation of biofilms. The plate was washed four times with PBS and 200µl of 0.1% crystal violet was added into each well to stain the remaining biofilm for 10 minutes at room temperature. The wells were washed with PBS to remove dye until no colour residue remained and plates left at room temperature to air dry. 200µl of 33% acetic acid was added to all wells to solubilise bacteria-bound crystal violet and mixed by pipetting up and down. A microplate reader (Tecan) was used to measure the presence of biofilm at OD^{570}. The processes of the biofilm formation assay are described in Figure 2.4.

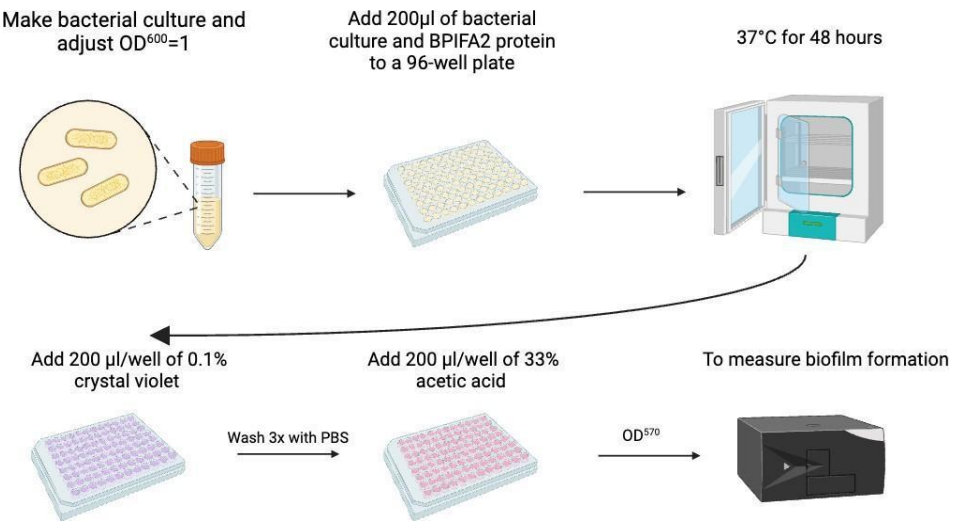


Figure 2.4. Diagram of biofilm formation assay protocol.

2.6 Protein-lipid interaction assay

To evaluate the binding between proteins and lipids, membrane lipid strips (Echelon Biosciences Inc; Figure 2.5) were used according to the manufacturer's instructions. 1µl of PI (4,5) P2 Grip and 1µl of secondary antibody were pipetted directly onto open areas of the dry membrane to confirm activity of the secondary HRP antibody and detection reagent prior to adding any samples. The lipid strips were blocked with 5mL of blocking buffer (TBS-Tween 20, 0.25% Tween-20) containing 1% (w/v) non-fat dry milk overnight at 4°C with gentle agitation. The lipid strips were incubated with purified saliva, purified recombinant proteins and PI(4,5)P2 Grip^{TM} (0.5µg ml^{-1}) as positive control, at room temperature for 1 hour with gentle shaking. The lipid strips were washed 3 times for 10 minutes each with 1X TBS-Tween and incubated with primary antibody (Anti-BPIFA2B (for human proteins) (1:500), Anti-FLAG (for animal proteins) (1:1000) or Anti-GST (for positive control, Sigma, UK) (1:2000))

at room temperature for 1 hour with gentle shaking. The lipid strips were washed 3 times for 10 minutes each with 1X TBS-Tween and incubated with secondary antibody (Anti- Rabbit HRP (1:3000) for BPIFA2B detection and Anti-mouse HRP (1:3000) for GST and FLAG detection) at room temperature for 30 minutes with gentle shaking. The lipid strips were washed 2 times in 1X TBS-Tween followed by a final wash with 1X TBS. Enhanced Chemiluminescence (ECL) Western Blotting substrate (Thermo Fisher Scientific) and a Li-Cor C-Digit Western Blot Scanner were used to determine the binding of BPIFA2 to the membrane lipids. The key processes of the protein-lipid binding assay are described in Figure 2.6.

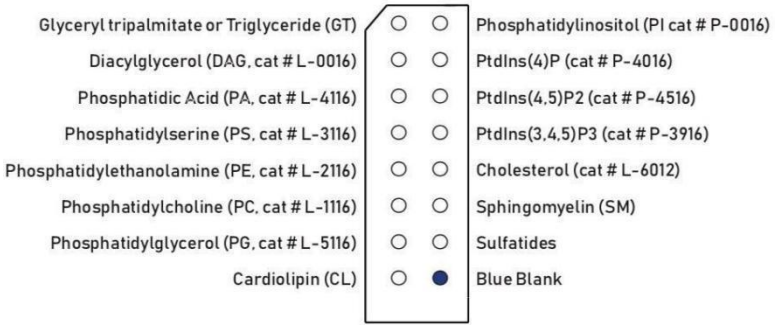


Figure 2.5. Membrane lipid strip template from Echelon Biosciences Inc.

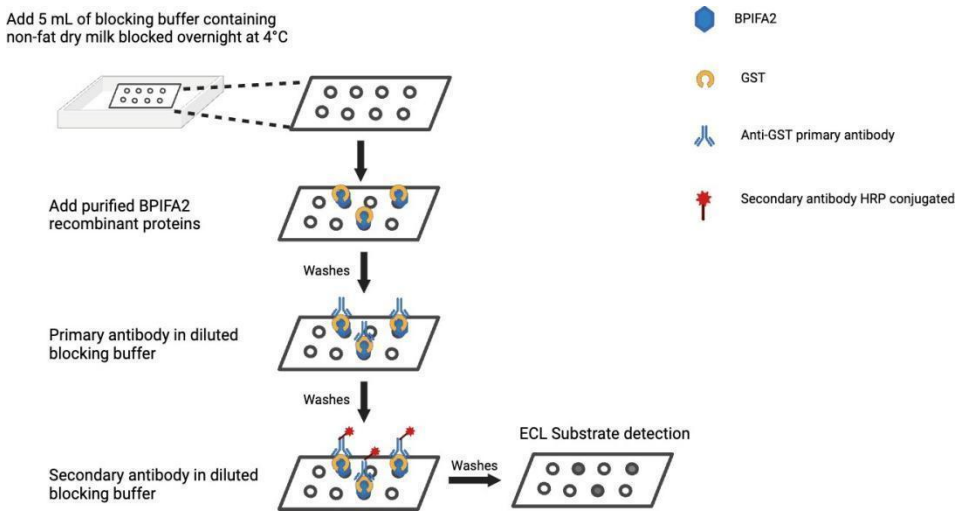


Figure 2.6. Diagram of protein-lipid binding assay protocol.

2.7 Interaction between BPIFA2 proteins and MM6 cells

2.7.1 Protein extraction from MM6 cell line

Prior to investigating any interaction between BPIFA2 proteins and MM6 cells it was necessary to determine whether MM6 cells produce BPIFA2 and so lysate from MMC 6 cells was used for BPIFA2 Western blot analysis.

1 x10^6 MM6 cell suspension was centrifuged at 2,500x *g* for 10 minutes and the supernatant discarded. The cell pellet was washed once by resuspending the cell pellet in ice-cold PBS and centrifuged at 2,500x *g* for 10 minutes. The cell pellet was mixed with ice-cold Radioimmunoprecipitation assay buffer (RIPA) lysis buffer (~1mL per 100mg or ~100µL of wet cell pellet, ChemCruz®, Biotechnology, USA) to which EDTA free protease inhibitor cocktail (Roche, Basel, Switzerland) had previously been added. The cell lysis mixture was incubated at 4°C for 30 minutes with constant agitation, and the cell debris was removed by centrifugation at 14,000x *g* for 15 minutes. The supernatant was transferred to a new

microcentrifuge tube and the protein concentration determined by BCA protein assay. Absorbance was detected at 570nm using a Tecan infinite M200 microplate reader. 10µl of lysis sample and purified saliva as positive control were mixed with 10µl of 2X SDS lysis buffer, boiled at 95°C for 5 minutes and loaded onto a 12% SDS-PAGE gel for Western blotting.

2.7.2 Opsonisation assay

An opsonisation assay was used to determine whether BPIFA2 coated bacterial strains are differentially phagocytosed by MM6 cells. Overnight cultures of *S. gordonii, P. aeruginosa, S. aureus, E. coli* and *S. mutans* were cultured in BHI broth, and absorbance (600nm) was adjusted to an optical optical density of 1 by diluting with BHI media.

10µL of bacterial suspension was mixed with 10µL purified saliva (positive control), BPIFA2 recombinant proteins or PBS (negative control) and incubated for 1 hour at 37°C. Coated and uncoated bacteria were seeded into separate wells of a 24-well plate that had previously been seeded with 1 x10^6 MM6 cells/ well and incubated for 30 minutes to allow infection/interaction. The cells/ bacteria samples were collected into clean microcentrifuge tubes and the samples were centrifuged at 400x *g* for 5 minutes before washing once with complete medium and resuspending in 0.5% saponin (Sigma) for 5 minutes in 5% CO^2 at 37°C. MM6 lysates were serially diluted and plated onto LB-agar plates to count bacteria which had either attached to the MM6 cells or had been phagocytosed by the cells. The processes of the opsonisation assay are described in Figure 2.7.

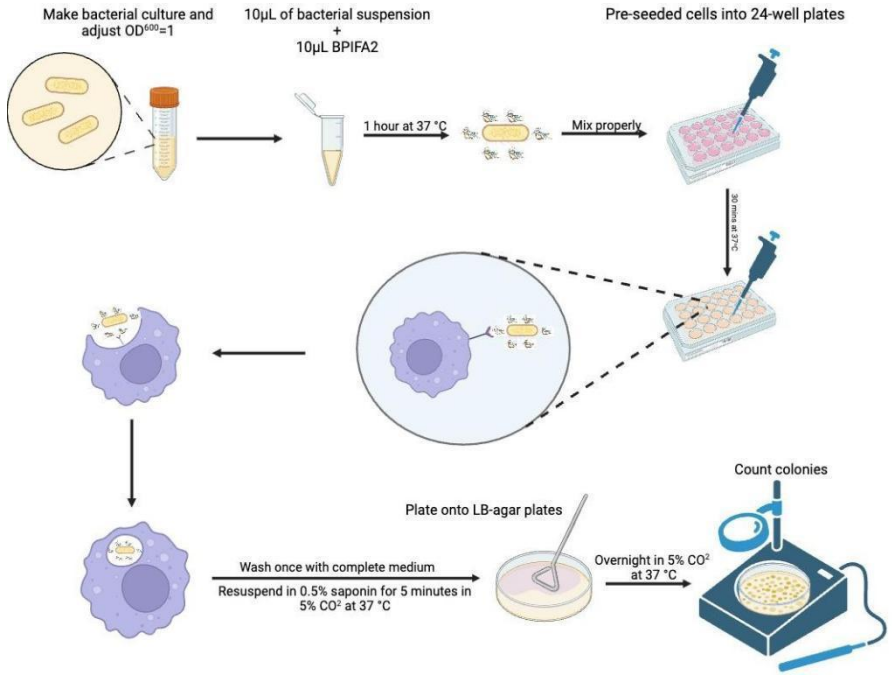


Figure 2.7. Diagram of opsonisation assay protocol.

2.8 Online Resources

All online software tools and platforms used in this study are summarised in Table 2.14.

Table 2.14. The online analysis tools and platforms used in the study.

Tools/Platforms	Purpose	URL Link
BioRender	Image generation	https://www.biorender.com/
Ensembl database	Genomic information	https://www.ensembl.org/index.html
Genome Data Viewer (GDV)	Genomes viewer	https://www.ncbi.nlm.nih.gov/genome/gdv/
GnomAD v4.0.0	Genomes viewer	https://gnomad.broadinstitute.org/
Human Protein Atlas (HPA)	Gene expression	https://www.proteinatlas.org/
Job Dispatcher (Clustal Omega)	Multiple sequence alignment	https://www.ebi.ac.uk/jdispatcher/
NCBI blastp	Protein BLAST: search protein databases using a protein query	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
NCBI VecScreen	Sequences for vector contamination	https://www.ncbi.nlm.nih.gov/tools/vecscreen/
NEBcutter V3.0	Restriction enzyme maps	https://nc3.neb.com/NEBcutter/

2.9 Statistics

Statistical significance was determined using the GraphPad Prism 10 programme. A one-sample t-test was employed to compare two groups, an analysis of variance (ANOVA) analysis for multiple groups and a P-value below 0.05 was deemed statistically significant. ANOVA assumes that the differences between observed and predicted values are normally distributed. Before applying ANOVA, a check for normality was carried out using the Shapiro-Wilk Test; a widely used test for small to medium sample sizes ($n < 50$) and iff $p > 0.05$, the data do not significantly deviate from normality. If the ANOVA indicates a significant effect ($p < 0.05$), this means that at least one group differs from the others, however, ANOVA doesn't specify which groups differ. Post-hoc tests are selected based on variance equality and research design to identify specific group differences. When discussing group-specific sample sizes, the symbol 'n' denotes technical repetitions, which refer to the number of samples, replicates, or individuals in a specific group. The purpose of technical repetitions is to control for variability introduced by experimental procedures or equipment. *In vitro* experiments (e.g., Western blot, ELISA): ≥3 technical replicates per sample are common. The purpose of biological repetition is to account for natural variation between individual biological units and ensure the results are representative. Biological repetitions are shown as 'N' refers to the total number of samples, subjects, or observations across all groups in an experiment. For example, cell culture experiments: ≥3 biological replicates are common while human saliva studies usually consist of ≥20 due to greater variability. Asterisks indicate the level of statistical significance (*P <0.05; **P <0.01; ***P <0.001; ****P <0.0001), whereas "NS" denotes non-significance (P >0.05).

Chapter 3: Bioinformatics analysis of BPIFA2 gene

3.1 Introduction

BPIFA2 is the most diverse member of the BPIF-containing family as multiple homologous and orthologous proteins have been detected in many species, including mice (Poulsen *et al.*, 1986), rats (Ballli, 1992), cows (Haigh *et al.*, 2008), pigs (Yin *et al.*, 2006), chimpanzee, gorillas, rhesus monkey and human (Bingle and Bingle, 2011). The sequence similarity between BPIFA2 in humans and mice is just over 30% which make it one of the most divergent orthologous protein pairs (Emes *et al.*, 2003). Such a low level of similarly made the assignment of orthologous relationship difficult to confirm (Bingle, Bingle and Craven, 2011). As highlighted in the introduction *BPIFA2* is highly expressed in the salivary glands of both species and the protein has been assumed to function in the saliva and oral cavity. Rats contain two *Bpifa2*-related genes, both of which are expressed in the salivary glands. The second of these, originally known as *Smgb*, (now *Bpifa2f*) evolved from a duplication event and in mice this second gene has been shown to be a pseudogene (Ball, Mirels and Hand, 2004).

Although the majority of studies on BPIFA2 have been undertaken in rodents and man, BPIFA2 related proteins have been identified in cow saliva and multiple *BPIFA2* related genes have been identified, suggesting a lineage specific expansion of the gene repertoire (Wheeler *et al.*, 2011). Cows are ruminants, a group of mammals that possess a unique digestive system characterized by a multi-compartment stomach, including the rumen,

reticulum, omasum, and abomasum. The presence of BSP30 (a BPIFA2 orthologue) was suggested to be linked to cattle's vulnerability to pasture bloat (Rajan *et al.*, 1996), a metabolic condition marked by the accumulation of stable foam in the rumen, leading to a malfunction in the eructation process, resulting in rumen distension and breathing difficulties (Clarke and Reid, 1974). The multiple *Bpifa2* related genes in the cow genome are also expressed in different tissue of the rumen as well as in the salivary glands. Previous sequence analysis of a preliminary sheep genome assembly also suggested the presence of numerous BPIFA2/BSP30-like proteins in sheep (Wheeler *et al.*, 2011). However, the specific role of BPIFA2 proteins in ruminant animals, in the context of rumination and digestion, is not well-documented. The expansion of the *BPIFA2* related genes in certain lineages suggests that some of these may have evolved new functions and this in turn suggests that detailed comparative analysis may provide useful functional information. As well as gene expansion within the BPIFA portion of the family, it is also well established that the locus contains many species and lineage-specific pseudogenes. An example of such a gene is *BPIFA4P(BASE)*, which is a salivary gland enriched human specific pseudogene that contains a single base deletion compared to other primate genes (Bingle, Bingle and Craven, 2011). *BASE* (breast cancer and salivary gland expression) encodes a secreted protein expressed in breast cancer cells, suggesting *BASE* might be a potential biomarker for breast cancer diagnosis (Egland *et al.*, 2003). Understanding the pattern of gene loss in the BPIF family may also provide important clues to the function of the proteins. Previous research has shown that cetaceans lack salivary glands and do not require saliva for digestion (Huelsmann *et al.*, 2019) and therefore it is of interest to study the conservation of *BPIFA2* in whales and dolphins compared with the gene in terrestrial mammals.

As highlighted in the introduction, human *BPIFA2* is located in the center of the *BPIF* locus on chromosome 20 (Bingle *et al.*, 2009) and this portion of the locus that is most variable between species. The underlying assumption is that these genes are involved in saliva or digestion, so genes in animals with different digestive systems may have evolved different functions. In humans, genetic variants may impact functions differently and are often specific to various ethnic groups. In humans, it is possible that the *BPIFA2* gene may have evolved due dietary changes and processes compared to other primates.

There has been very little new analysis of *BPIFA2* and related genes generated over the last 15 years during which time there has been a very significant increase in the amount of accessible genomic and expression data that can be studied. As BPIFA2 is a salivary gland-enriched protein, understanding the conservation of the gene and its variation across species may be important for the understanding of its salivary function. In this chapter, I conducted a comprehensive analysis of BPIFA2 across mammalian species by searching online platforms and databases.

3.2 Aims

This chapter focuses on a comparative analysis of *BPIFA2* and *BPIFA2*-related genes between human and animal orthologues using multiple online platforms (see section 2.8 for the tools used in this chapter.) in conjunction with previous published studies which only described paralogs within the same organism due to a duplication gene. All the platforms and tools used in this chapter are free and clearly describe the outputs. The Ensembl database was chosen for its comprehensive genomic data, enabling exploration of BPIFA2 gene structure,

variants, and orthologs. Basic Local Alignment Search Tool (BLAST) is essential for identifying homologous sequences, uncovering evolutionary relationships, and characterising conserved domains. Clustal Omega tool ensures accurate multiple sequence alignment for evolutionary and functional comparisons. BLAST-like alignment tool (BLAT) was chosen for its speed and accuracy in mapping BPIFA2 sequences to a genome, enabling efficient localisation and analysis of gene regions. UCSF Chimera is used for its advanced molecular visualization capabilities, allowing detailed exploration of BPIFA2's 3D structure and interactions. Genome Data Viewer (GDV) provides an interactive platform to visualise BPIFA2 within its genomic context, integrating annotations and comparative genomics. GnomAD v4.0.0 database is ideal for population-level variant analysis of BPIFA2, providing insights into genetic diversity and pathogenic variants. These platforms and tools are reliable, user-friendly, and tailored to BPIFA2's genomic, proteomic, and functional analysis. A comprehensive analysis of this nature can be used to highlight potentially important differences between species that may be related to function of the proteins and provide a rational background for the wider studies presented in this thesis.

3.3 Results

3.3.1 Comparative analysis of BPIFA2 across multiple genomes

Initially, a phylogenetic tree of BPIFA2 was extracted from Ensembl online database (https://www.ensembl.org/index.html) using the human BPIFA2 protein sequence as the query and covering all mammalian branches (Figure 3.2). It is important to note that the Ensembl database does not contain all genomes and is therefore not a complete representation

of mammalian phylogeny. On this tree some branches are compressed for simplicity. The tree shows that all classes of mammals contain BPIFA2 orthologs with the majority of species containing a single gene. As noted previously, cows contain multiple BPIFA2 proteins and the red highlighted section of Figure 3.1 (enlarged in Figure 3.2) shows that there are multiple *BPIFA2* related genes in cows. This is consistent with the previous identification of BSP30A, BSP30B, BSP30C and BSP30D, all of which are found in bovine saliva (Wheeler *et al.*, 2011). Multiple *BPIFA2* related genes are also seen in other ruminant animals including sheep, goats, sheep, bison, yak and deer. Three BPIFA2-like proteins are also identified in the horse (a non- ruminant species).

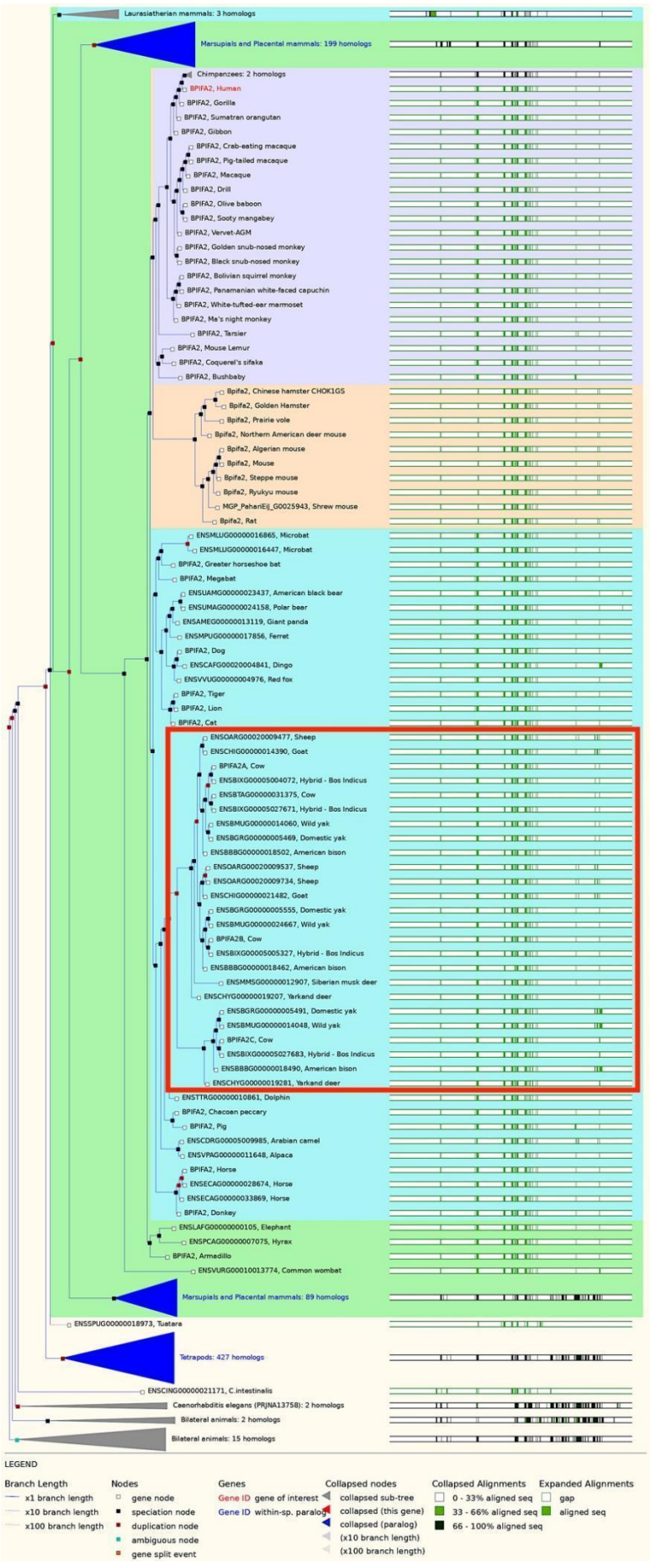

Laurasiatherian mammals: 3 homologs
Marsupials and Placental mammals: 199 homologs
Chimpanzees: 2 homologs
BPIFA2, Human
BPIFA2, Gorilla
BPIFA2, Sumatran orangutan
BPIFA2, Gibbon
BPIFA2, Crab-eating macaque
BPIFA2, Pig-tailed macaque
BPIFA2, Macaque
BPIFA2, Drill
BPIFA2, Olive baboon
BPIFA2, Sooty mangabey
BPIFA2, Vervet-AGM
BPIFA2, Golden snub-nosed monkey
BPIFA2, Black snub-nosed monkey
BPIFA2, Bolivian squirrel monkey
BPIFA2, Panamanian white-faced capuchin
BPIFA2, White-tufted-ear marmoset
BPIFA2, Me's night monkey
BPIFA2, Tarsier
BPIFA2, Mouse Lemur
BPIFA2, Coquerel's sifaka
BPIFA2, Bushbaby
Bpifa2, Chinese hamster CHOK1GS
Bpifa2, Golden Hamster
Bpifa2, Prairie vole
Bpifa2, Northern American deer mouse
Bpifa2, Algerian mouse
Bpifa2, Mouse
Bpifa2, Steppe mouse
Bpifa2, Ryukyu mouse
MGP_PahariEij_G0025943, Shrew mouse
Bpifa2, Rat
ENSMLUG00000016865, Microbat
ENSMLUG00000016447, Microbat
BPIFA2, Greater horseshoe bat
BPIFA2, Megabat
ENSUAMG00000023437, American black bear
ENSUMAG00000024158, Polar bear
ENSAMEG00000013119, Giant panda
ENSMPUG00000017856, Ferret
BPIFA2, Dog
ENSCAFG00020004841, Dingo
ENSVVUG00000004976, Red fox
BPIFA2, Tiger
BPIFA2, Lion
BPIFA2, Cat
ENSOARG00020009477, Sheep
ENSCHIG00000014390, Goat
BPIFA2A, Cow
ENSBIXG00005004072, Hybrid - Bos Indicus
ENSBTAG00000031375, Cow
ENSBIXG00005027671, Hybrid - Bos Indicus
ENSBMUG00000014060, Wild yak
ENSBGRG00000005469, Domestic yak
ENSBBBG00000018502, American bison
ENSOARG00020009537, Sheep
ENSOARG00020009734, Sheep
ENSCHIG00000021482, Goat
ENSBGRG00000005555, Domestic yak
ENSBMUG00000024667, Wild yak
BPIFA2B, Cow
ENSBIXG00005005327, Hybrid - Bos Indicus
ENSBBBG00000018462, American bison
ENSMMSG00000012907, Siberian musk deer
ENSCHYG00000019207, Yarkand deer
ENSBGRG00000005491, Domestic yak
ENSBMUG00000014048, Wild yak
BPIFA2C, Cow
ENSBIXG00005027683, Hybrid - Bos Indicus
ENSBBBG00000018490, American bison
ENSCHYG00000019281, Yarkand deer
ENSTTRG00000010861, Dolphin
BPIFA2, Chacoan peccary
BPIFA2, Pig
ENSCDRG00005009985, Arabian camel
ENSVPAG00000011648, Alpaca
BPIFA2, Horse
ENSECAG00000028674, Horse
ENSECAG00000033869, Horse
BPIFA2, Donkey
ENSLAFG00000000105, Elephant
ENSPCAG00000007075, Hyrax
BPIFA2, Armadillo
ENSVURG00010013774, Common wombat
Marsupials and Placental mammals: 89 homologs
ENSSPUG00000018973, Tuatara
Tetrapods: 427 homologs
ENSCING00000021171, C.intestinalis
Caenorhabditis elegans (PRJNA13758): 2 homologs
Bilateral animals: 2 homologs
Bilateral animals: 15 homologs
LEGEND
Branch Length
x1 branch length
x10 branch length
x100 branch length
Nodes
gene node
speciation node
duplication node
ambiguous node
gene split event
Genes
Gene ID gene of interest
Gene ID within-sp. paralog
Collapsed nodes
collapsed sub-tree
collapsed (this gene)
collapsed (paralog)
(x10 branch length)
(x100 branch length)
Collapsed Alignments
0 - 33% aligned seq
33 - 66% aligned seq
66 - 100% aligned seq
Expanded Alignments
gap
aligned seq

Figure 3.1. Phylogenetic tree of BPIFA2 proteins. A phylogenetic tree of BPIFA2 proteins was derived from the ENSEMBL database (ENSGT01100000263546) using the human BPIFA2 protein sequence as query (shown in red). BPIFA2 proteins from ruminant animals are highlighted by the red box.

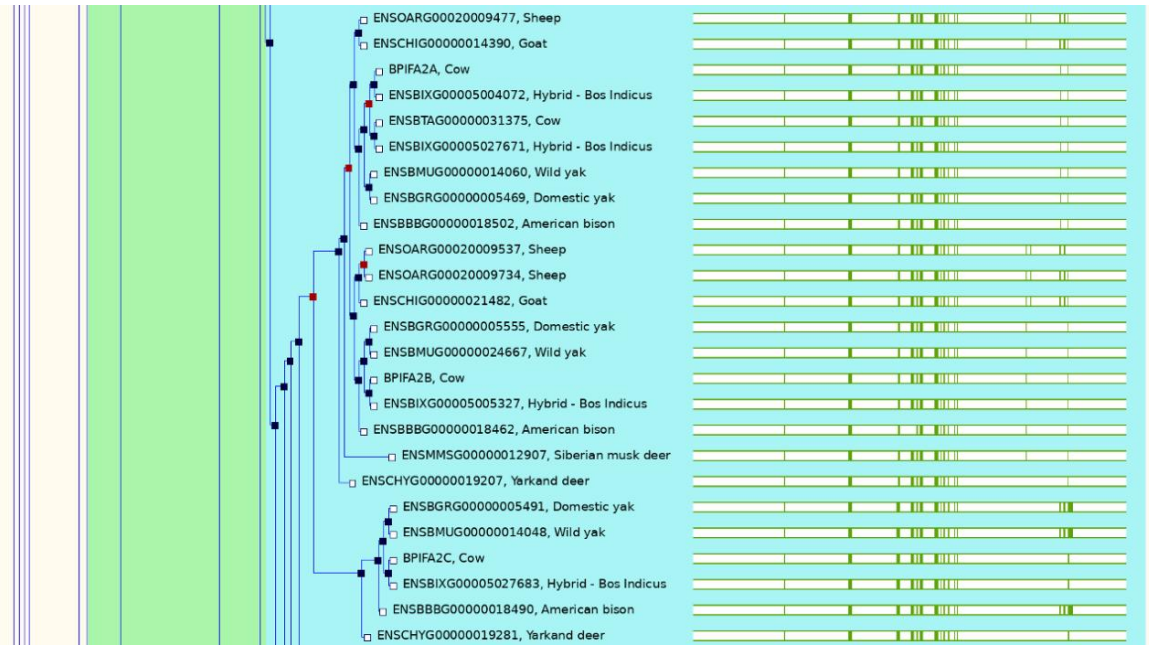


Figure 3.2. Expansion of the Phylogenetic tree of ruminant BPIFA2 proteins. The ruminant proteins of the BPIFA2 tree from the Figure 3.1 above have been expanded for clarity. The red nodes show duplication events.

To investigate the level of BPIFA2 sequence conservation across species, the human BPIFA2 sequence was analysed by a BLASTP search of the NCBI reference sequence (RefSeq) protein database. The results of this multiple sequence alignment (MSA) are presented in Figure 3.3.

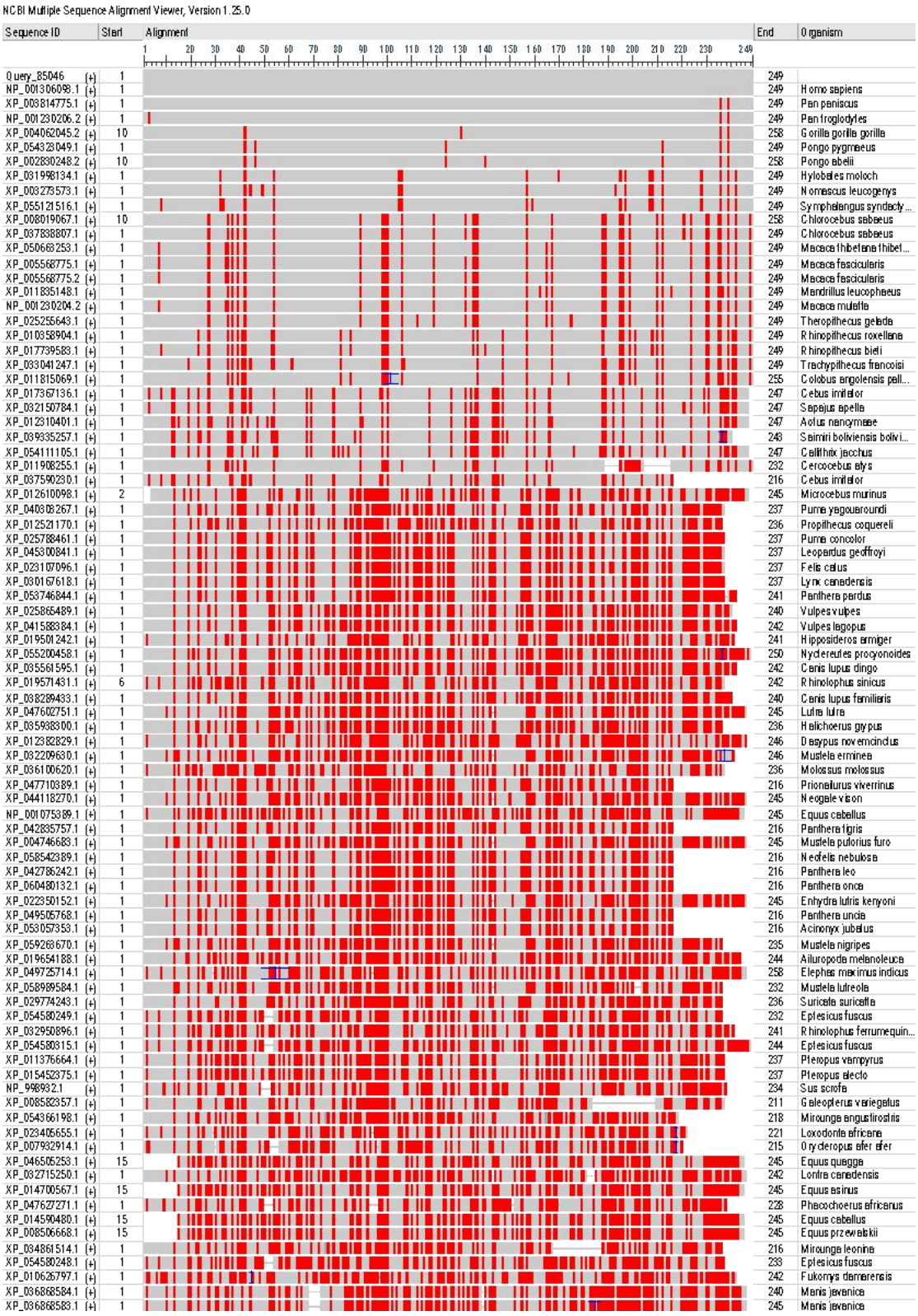


Figure 3.3. Multiple sequence alignments of BPIFA2. FASTA sequences of BPIFA2 were aligned using BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The top 86 proteins are shown (unexpected extreme long sequences were removed). Conserved regions are highlighted in grey and amino acid differences are shown in red.

The most notable observation from this alignment is the dissimilarity seen between the primate BPIFA2 sequences and those obtained from other species. For clarity the grey colour shows protein residues that are completely conserved and the red colour shows the amino acid differences. The presence of the highest homology within the primate group is evident, as is a notable distinction in the protein lengths. This shows non-primate proteins are shorter in length. Although this analysis shows a low level of conservation across species, in this format this type of image does not show this to be seen at the individual amino acid level and to do this, a comparative analysis of BPIFA2 across 18 representative mammalian species using the Clustal Omega alignment tool (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) (Figure 3.4). Sequences that were used for this analysis were taken from across the phylogenetic tree displayed above (Figure 3.2) and were chosen to represent as much as possible every branch of mammalian phylogeny. The sequences were obtained from Genbank as RefSeq sequences and were visually inspected to confirm that they were complete predictions as shown in appendix.

Consistent with what is shown in Figure 3.3, the most striking observation is the lack of overall similarity across the proteins with only 13 of 249 amino acids completely conserved accross these sequences. These are highlighted in dark grey in Figure 3.4. Seven of these amino acids are located in the signal peptide region (position: 1-20) at the beginning of sequences, which suggests evolutionary conservation of the protein targeting and secretion mechanisms. It implies that these sequences have been maintained throughout evolution due to their functional importance. There are some insertions in the sequences in some species around the N-terminal region, a potential deletion in the hedgehog (*Erinaceus europeaus*) as

highlighted by yellow box and a C-terminal extension in the dolphin (*Tursiops truncatus*) sequence as highlighted by green box.

Glycosylation sites in BPIFA2 are critical for its structural stability, secretion, and functional roles in immunity. The MSA also shows a lack of conservation of the predicted N-glycosylation sites (NXS or NXT) highlighted by red colour in Figure 3.4. The glycosylation sites are varied and not conserved across all species. Fourteen species have N-glycosylation sites located in the sequences as shown in Figure 3.4. The macaque (*Macaca mulatta*) has three N-glycosylation sites at positions 124, 165 and 188 which has the most sites across all the species. Humans (*Homo sapiens*), chimpanzees (*Pan troglodytes*) and capped leaf monkeys (*Trachypithecus francoisi*) have two N-glycosylation sites at positions 124 and 132. Gorilla (*Gorilla gorilla gorilla*) has two N-glycosylation sites at positions 133 and 141. Mouse (*Mus musculus*) has one N-glycosylation site at position 215. Dog (*Canis lupus dingo*), Bottlenose dolphin *(Tursiops_truncatus),* pig (*Sus scrofa*), and Sumatran orangutan (*Pongo abeli*i) have one N- glycosylation site at positions 132 which seems to be in one of the same positions of human. Common marmoset (*Callithrix_jacchus*) and brown capuchin (*Sapajus_apella*) are new world monkeys that have the same N-glycosylation site at position 147, and the site position is different from all of the species aligned in the figure. Armadillo (*Dasypus novemcinctus*) has one N-glycosylation site at position 120 that is different from all of the species.

Notably, N-glycosylation sites at positions 124 and 132 in chimpanzees are in the same positions as in humans, which suggests the function of BPIFA2 may be the same between these two species.

There are no N-glycosylation sites in five of the sequences aligned here, rat (*Rattus norvegicus*), hedgehog (*Erinaceus europaeus*), squirrel monkey (*saimiri boliviensis boliviensis*) gray mouse lemur (*Microcebus murinus*) and elephant (*Elephas_maximus_indicus*). Functional analysis (see chapter 5) is necessary to understand the impact of N-glycosylation sites on BPIFA2 proteins across species by comparing the function of wild-type BPIFA2 and mutant forms (both single and double N-glycosylation site mutants) in humans. Additionally, BPIFA2 from armadillo, mouse, macaque, squirrel monkey, and dog, were also compared with human BPIFA2 in functional assays. The choice of these species was made as they covered a wide phylogenetic range.

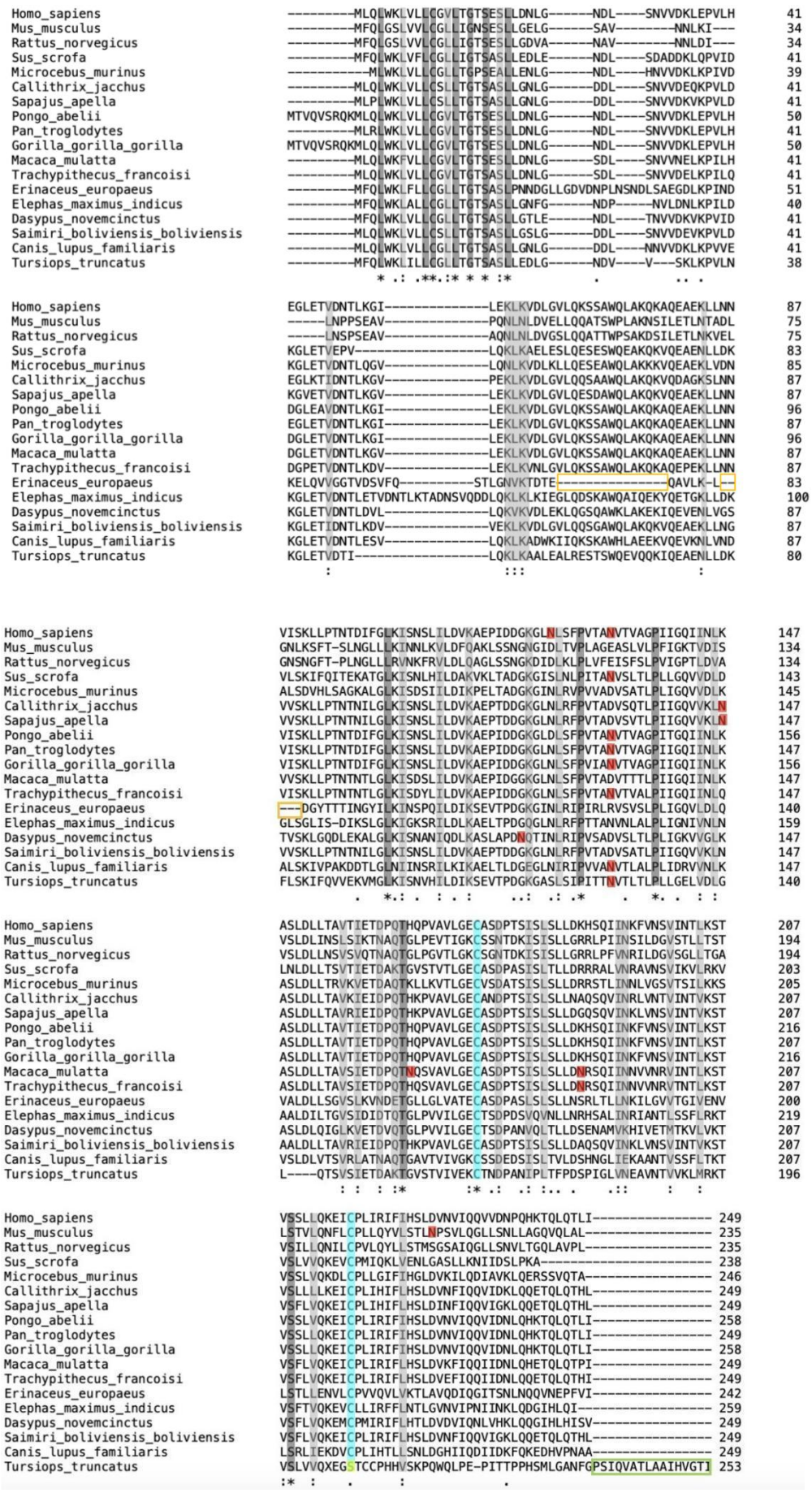

Homo_sapiens
Mus_musculus
Rattus_norvegicus
Sus_scrofa
Microcebus_murinus
Callithrix_jacchus
Sapajus_apella
Pongo_abelii
Pan_troglodytes
Gorilla_gorilla_gorilla
Macaca_mulatta
Trachypithecus_francoisi
Erinaceus_europaeus
Elephas_maximus_indicus
Dasypus_novemcinctus
Saimiri_boliviensis_boliviensis
Canis_lupus_familiaris
Tursiops_truncatus

Figure 3.4. Multiple sequence alignments of BPIFA2 protein across 18 mammalian species. ClustalOmega alignment tool (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) was used to compare multiple sequence alignments. A star (*) represented completely conserved amino acid residues, a colon (:) represented semi conserved residues and a period (.) represented weakly similar residues. Highly conserved residues are highlighted in dark grey and semi- conserved residues are highlighted in light grey. N-glycosylation sites are highlighted in red and cysteine residues formed disulfide bonds in protein are highlighted in blue. The deletion residues are highlighted in the yellow box and extension residues are highlighted in the green box.

Previous studies have demonstrated that all BPIF proteins have two cysteine residues that form a disulfide bond (Bingle *et al.*, 2004). The cysteines at positions 174 and 217 are highlighted in blue colour in the alignment across all 18 species in the MSA although it should be noted that the second cysteine residue as highlighted in yellow colour in the dolphin (*Tursiops truncates*) sequence is shifted a little compared to the other species. This bond is likely to have a significant role in the function of BPIFA2, as it is also conserved among the wider LBP/BPI family and related members (Beamer, Carroll and Eisenberg, 1997, 2008; Krasity *et al.*, 2011).

3.3.2 Exploration of BPIFA2 in human and five animal species

Due to the low level of sequence conservation and N-glycosylation sites I reasoned that functional studies performed with a number of different proteins might be valuable. In addition to humans, I chose five representative animal species, mouse, squirrel monkey, macaque, armadillo and dog to cover as great a phylogenetic range between species as I could. The MSA of these BPIFA2 proteins is shown in Figure 3.5. The total amino acid residues were 249 in dog, armadillo, squirrel monkey, macaque and human, but at 235 amino acids in the mouse, it is shorter than other species. There is a clear deletion at the beginning of the N-

terminal portion immediately after the signal peptide (highlighted by the red box).

In these six species, the alignment shows that 42 of 249 amino acids are completely conserved (highlighted in green in Figure 3.5) and ten of these are located in the signal peptide region (position 1-20). The N-glycosylation sites are highlighted in yellow. The human protein has two at positions 124 and 132. There are 3 sites at positions 124, 165 and 188 of macaque one of which 124 is at the same position as in humans. The other 2 sites (position 165 and 188) are different from sites in other species. Dogs have one site at position 132 which is at the same position as is in humans. Mouse has one site at position 215 which is different from others. There are no N-glycosylation sites in armadillo and squirrel monkeys. The cysteine residues forming the disulfide bonds are highlighted in blue at positions 174 and 217.

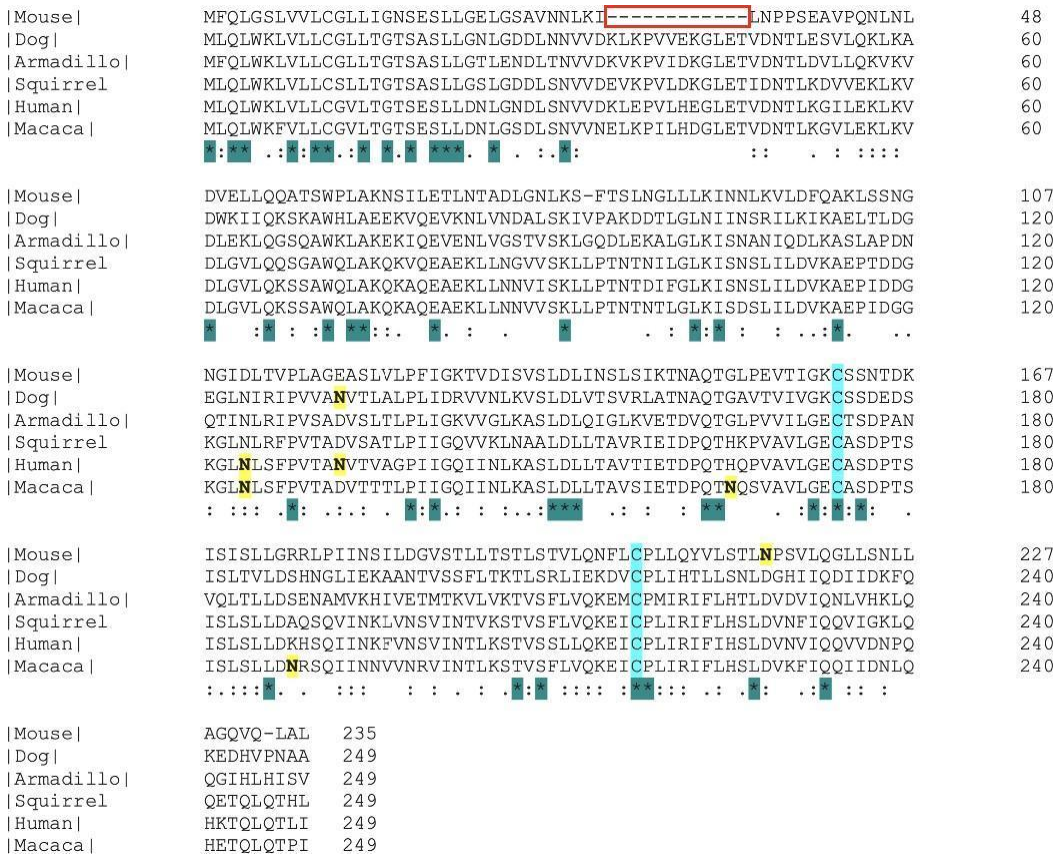
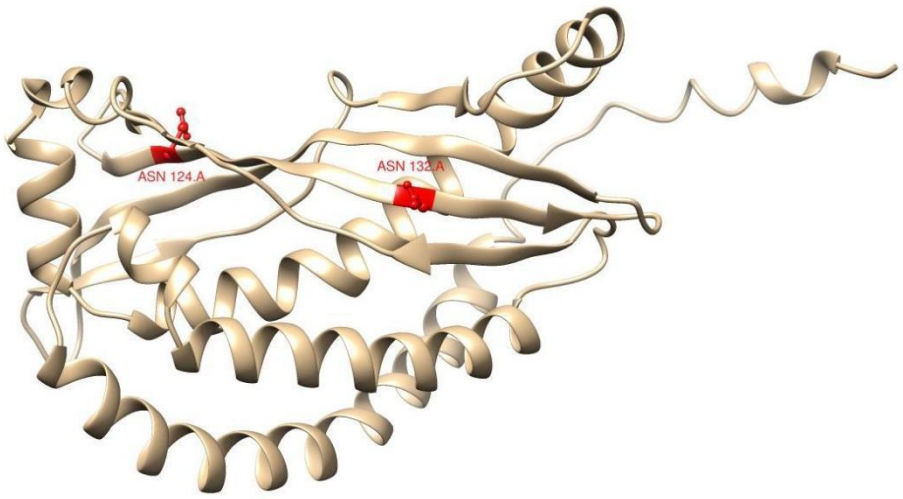


Figure 3.5. Multiple sequence alignments of BPIFA2 protein in human and 5 animal species. ClustalOmega alignment tool (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) was used to compare multiple sequence alignments. A star (*) represented completely conserved amino acid residues highlighted in green colour, a colon (:) represented semi conserved residues and a period (.) represented weakly similar residues. The deletion residues are highlighted in red box. N-glycosylation sites are highlighted in yellow and cysteine residues formed disulfide bonds in protein are highlighted in blue.

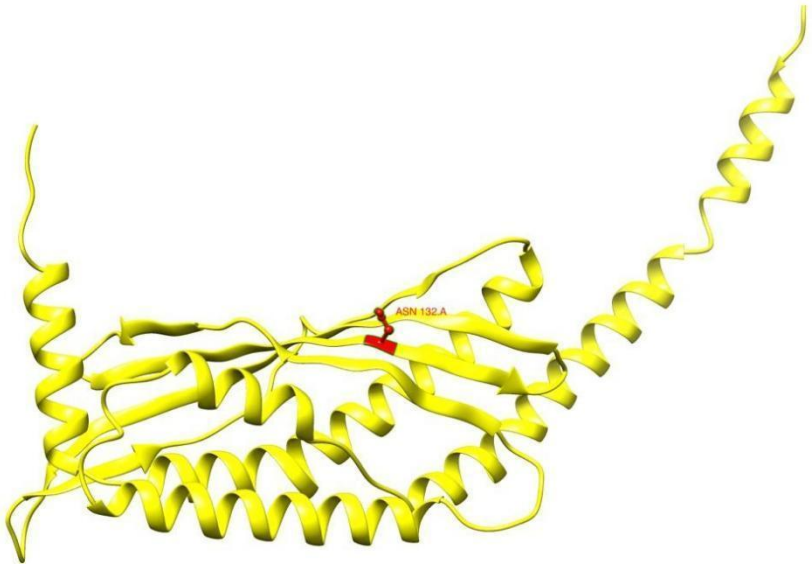
To investigate the relationship between N-glycosylation sites and protein structure, the UCSF Chimera online tool 1.17.3 was used to model and visualize the predicted secondary structures BPIFA2 (Figure 3.6). To date there is no published BPIFA2 structure. The structures of human (Figure 3.6A), dog (Figure 3.6B), macaque (Figure 3.6C) and mouse (Figure 3.6D) were predicted by Alpha Fold Protein Structure Database (https://alphafold.ebi.ac.uk/) which is an artificial intelligence (AI) system developed to predict a three-dimensional structure of protein from its amino acid sequence (Jumper *et al.*,

2021). The rationale for generating these models was that I wished to see if the glycosylation sites were physically close and in the same structural region of the protein in spite of their lack of conservation in the alignment. In these models the sites are shown in red. For the N-glycosylation sites of positions124 and 132 on the beta sheet of the protein structure in human, dog and macaque proteins and for the N-glycosylation sites of positions at 165, 188 and 215 were closer to alpha helical regions of the protein structure in macaque and mouse proteins. These results suggest that the position of N-glycosylation sites is probably not important.

A



B



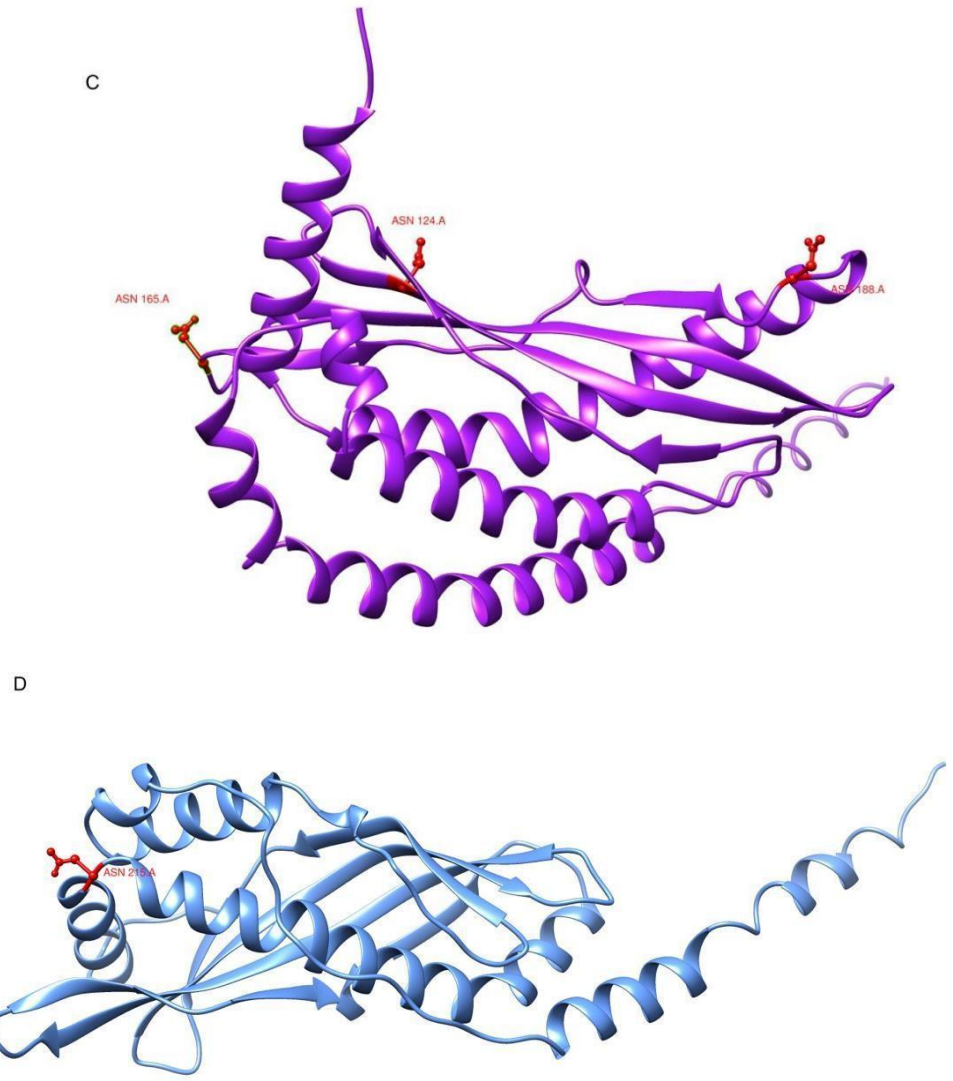


Figure 3.6. N-glycosylation sites of BPIFA2 in 4 animal species. N-glycosylation sites were highlighted in red colour using UCSF Chimera 1.17.3 (https://www.cgl.ucsf.edu/chimera/). Position 124 and 132 were highlighted in human (A), position 132 was highlighted in dog (B), position 124, 165 and 188 were highlighted in macaque (C) and position 215 was highlighted in mouse (D).

3.3.3 *BPIFA2 gene* loss in cetaceans

In the MSA in Figure 3.4 the predicted BPIFA2 sequence from the dolphin (*Tursiops truncatus*) displayed some differences from the other species at the end of C-terminal in that the sequence length is longer than seen in the other species. It is known that dolphins and other cetaceans have residual (or no) salivary glands as their function is presumably not

required as they have transitioned to an aquatic lifestyle (Huelsmann *et al.*, 2019). It was therefore interesting to look at the *BPIFA2* gene in greater detail in cetaceans. To do this I employed the CACTUS (Comparative Algorithm for Chaining Transcripts Using Splice sites) multiple genome alignment tool on the BLAT (BLAST-like alignment tool) server (Armstrong *et al.*, 2020). The Cactus track on BLAT displays a multiple sequence alignment for up to 241 mammalian genomes from the Zoonomia project (Vignieri, 2023), which aims to understand mammalian genome evolution.

To investigate this, I used BPIFA2 on the human BLAT track as the query and in the Cactus track, selected all 16 cetacean genomes along with the hippopotamus genome, as this represents the closest living land dwelling ancestor to cetaceans. The alignment is shown in Figure 3.7

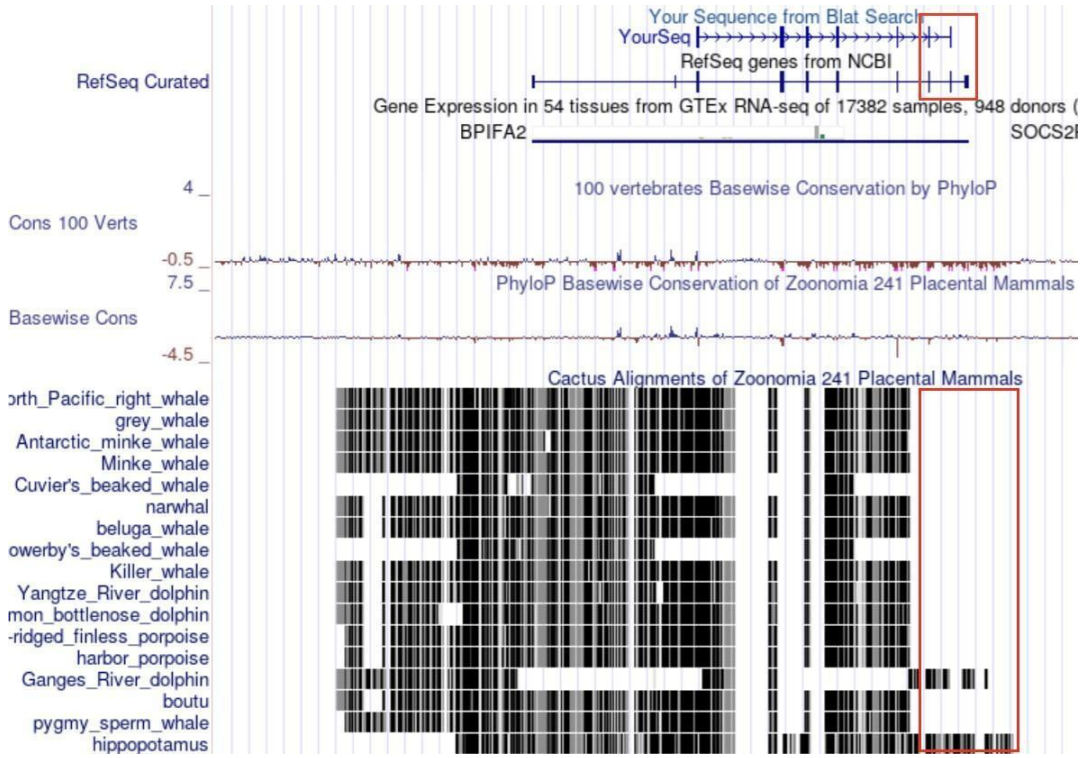


Figure 3.7. *BPIFA2* locus shows gene loss in cetaceans. Human BPIFA2 was used as a query sequence and the genomes of 16 cetaceans and the hippopotamus were visualized on BLAT: https://genome.ucsc.edu/cgi-bin/hgBlat. The loss of the 3' exons is shown in the red boxes.

The black and white blocks at the bottom reflect alignment patterns of genomic sequences between various species. The blocks represent aligned (similar) portions of the genome, while gaps may reflect insertions, deletions, or highly divergent regions. The red box highlights a region where the alignment blocks (representing exon regions) are absent in most cetacean species compared to hippopotamus. The absence of these blocks suggests that these species do not have these particular exons and non-coding exons for *BPIFA2*, due to evolutionary loss (Huelsmann *et al.*, 2019).

The expanded region including *BPIFA4P* is shown in Figure 3.8. As highlighted previously the *BPIFA4P* gene is a pseudogene in humans since a single nucleotide loss leads to a premature

stop codon (Bingle, Bingle and Craven, 2011). In Figure 3.8, BPIFA4 is not found in human saliva even though it is highly expressed in human salivary glands at the RNA level (Egland *et al.*, 2003). The gene is however functional in great apes. The alignment shows that the *BPIFA4* gene is also lost from the locus in cetaceans compared to chimpanzees. This analysis clearly shows that *BPIFA2* and the related salivary gland enriched gene *BPIFA4* are lost from the genomes of cetaceans.

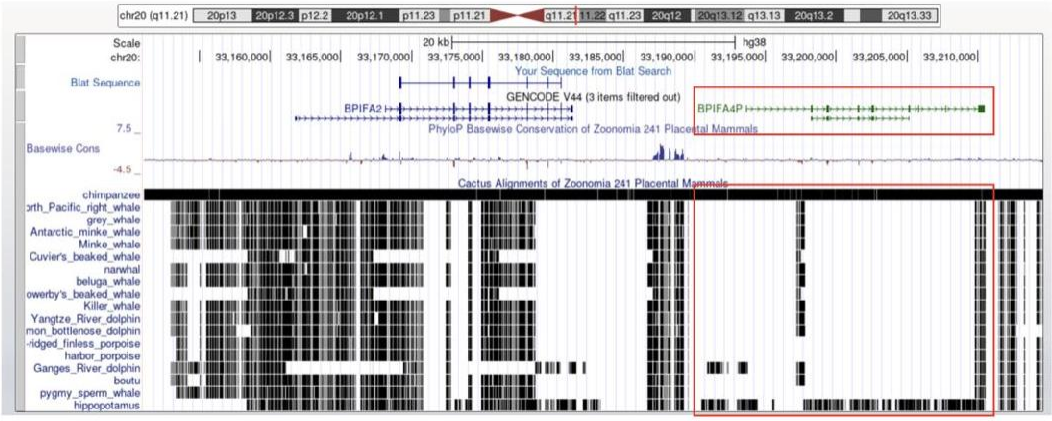


Figure 3.8. The *BPIFA2* locus shows gene loss in cetaceans. Human BPIFA2 was used as a query sequence and 16 cetacean genomes, hippopotamus and chimpanzee were visualized by Human BLAT search: https://genome.ucsc.edu/cgi-bin/hgBlat. *BPIFA4P* was highlighted in red boxes.

3.3.4 *BPIFA2* gene expansion in ruminant mammals

As highlighted above BPIFA2 and related proteins, have undergone an expansion in the bovine lineage and it has been suggested that this may relate to the function of the proteins in saliva and digestion in this specialized mammalian lineage. The biological function of BPIFA2 proteins in ruminant animals, particularly in rumination and digestion needs extensive research and validation; however, *BPIFA2* may be used as an example for comprehending aspects of the molecular evolution of gene families.

3.3.4.1 *BPIFA2* related genes in cattle

National Center for Biotechnology Information (NCBI) Genome Data Viewer (GDV) was used to investigate and visualize the *BPIFA2* genes in cattle (*Bos taurus)* as described in Figure 3.9. This analysis used the ARS-UCD2.0 assembly released on October 3rd 2023. LOC numbers (e.g., "LOC112441508") are identified for specific loci within the genome which shows the loci as highlighted in red boxes. The horizontal scale marks the genomic coordinates in kilobases (Kb), indicating the location on the chromosome.

From the data shown in Figure 3.9, between *BPIFB4* and *Bpifa4* there are six *BPIFA2* related genes in cattle, including four true genes (*Bpifa2a, Bpifa2b, Bpifa2c* and *LOC112441508*) that are highlighted in red boxes and two pseudogenes (*Bpifa2D* and *LOC112449237*) that are highlighted in black boxes in the region of the genome. *LOC132346950* appears to be a pseudogene related to *b4*. Each box corresponds to a gene, and the direction of the arrows indicates the direction of transcription from the 5' end to the 3' end. All of the complete genes are supported by RNAseq data from bovine tissues displayed underneath the locus region.

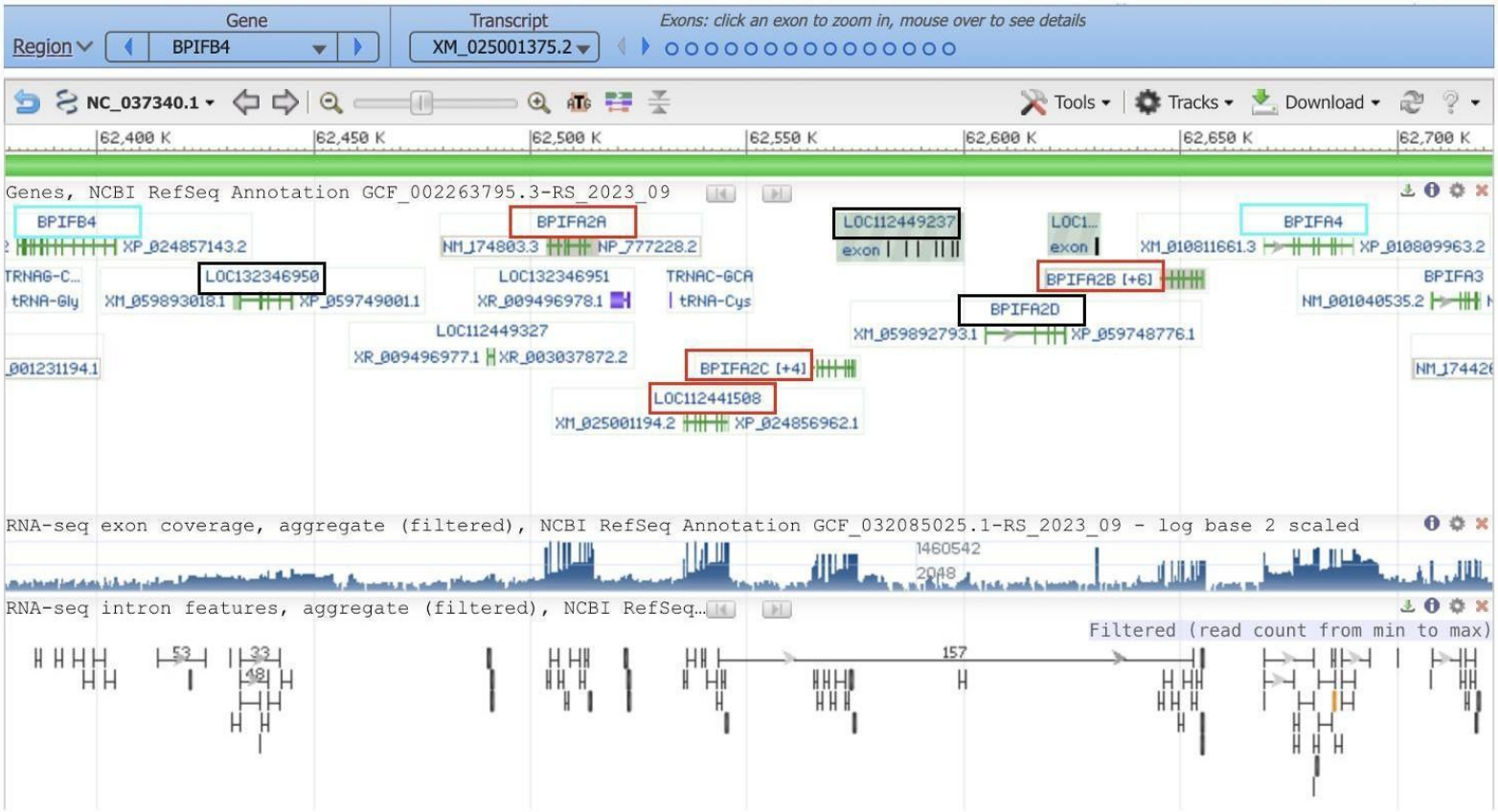


Figure. 3.9. The genetic expression of *BPIFA2 (BSP30)* in cattle. BSP30 was used as query and the genome was visualized on the National Center for Biotechnology Information (NCBI) Genome Data Viewer (GDV): https://www.ncbi.nlm.nih.gov/genome/gdv/. *Bpifa2a*, *Bpifa2b, Bpifa2c* and *LOC112441508* were highlighted in red boxes. And *Bpifa2d*, *LOC132346950* and *LOC112449237* were highlighted in black boxes. *BPIFA4* and *BPIFB4* were highlighted in blue box. The arrows on the boxes show the direction of the gene transcription. The RNAseq data provides evidence for all the complete genes within the locus regions.

Next, I extracted protein sequences from NCBI database through the GDV site and generated a MSA of bovine BPIFA2-related proteins with ClustalOmega (https://www.ebi.ac.uk/jdispatcher/msa/clustalo). Five BPIFA2-related sequences BPIFA2A, LOC112441508 (an uncharacterized protein), BPIFA2C, BPIFA2D and BPIFA2D and BPIFA4 were compared in Figure 3.10.

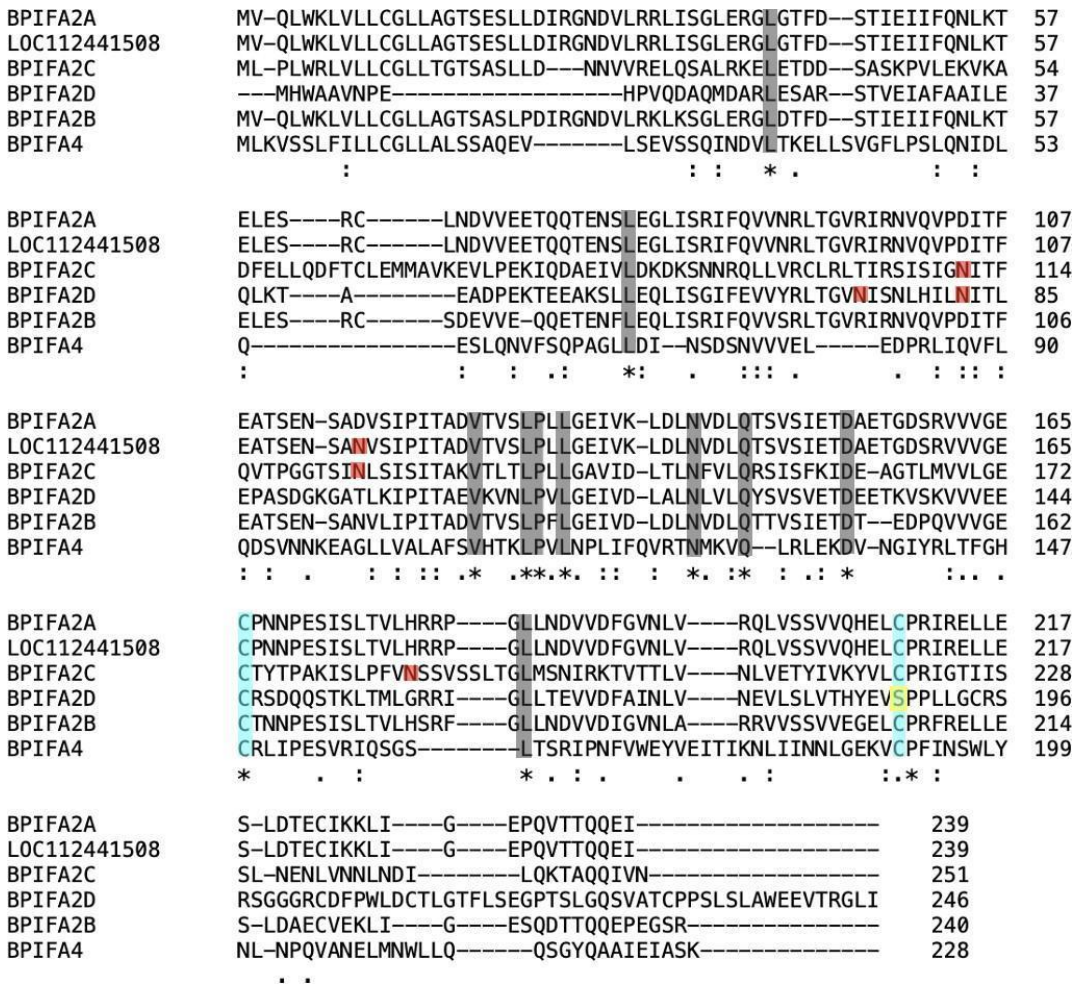


Figure 3.10. Multiple sequence alignments of BPIFA2-related proteins in cattle (Bos taurus). ClustalOmega alignment tool (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) was used to compare multiple sequence alignments. A star (*) represented completely conserved amino acid residues, a colon (:) represented semi conserved residues and a period (.) represented weakly similar residues. Highly conserved residues are highlighted in dark grey and N- glycosylation sites are highlighted in red. Cysteine residues formed disulfide bonds in protein are highlighted in blue.

From the output, it can be seen there were only 9 completely conserved amino acids as highlighted in dark grey. The cysteine residues forming the disulfide bonds are conserved in all variants as highlighted in blue at positions 173 and 200. The addition of the predicted BPIFA2D sequence (which is missing many of the coding exons required to make a complete BPIFA2 related protein), cysteine was replaced by serine at the second position (as highlighted in blue at position 200). In these six sequences, both of the cysteine sites are separate from the locations of the cysteines as shown in Figure 3.4.

The MSA also clearly shows a lack of conservation of the N-glycosylation sites highlighted by red colour in Figure 3.10. LOC112441508 has one N-glycosylation site at position 116, BPIFA2D has two N-glycosylation sites at position 74 and 82 and BPIFA2C has three N-glycosylation sites at position 111, 124 and 186. Each of these places is distinct from the individual sequence and also the N-glycosylation sites are different from the sites in other mammals shown in Figure 3.4. There is no N-glycosylation site shown in BPIFA2A, BPIFA2B and BPIFA4.

I next compared exon sizes between the bovine *Bpifa2*-related genes and human *BPIFA2*, to provide insights into the evolutionary conservation and divergence of genes. The exon sizes and protein lengths of humans were used as direct comparators. Exon sequences and predicted protein lengths were collected from NCBI GDV and their lengths verified using SnapGene Viewer: https://www.snapgene.com/snapgene-viewer. These are shown in Table 3.1.

Table 3.1 The comparison of exon sizes between human and cattle. The sizes of individual exons (in bp) for the cattle genes are compared with humans. The sizes of exon were collected from NCBI Genome Data Viewer database and were verified by SnapGene Viewer: https://www.snapgene.com/snapgene-viewer.

Gene	Protein Length	**Exon 1**	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	*Exon 7*	Exon 8	**Exon 9**
hBPIFA2	249	55	168	145	108	153	82	*64*	78	188
Bpifa2a	236	15	161	131	105	153	82	*64*	75	182
LOC112441508	239	24	161	133	105	153	82	*64*	75	182
Bpifa2c	251	231	145	163	108	150	94	*64*	79	183
Bpifa2d	246		82	133	108	153	272			
Bpifa2b	240	105	161	130	105	147	82	*64*	74	182

The protein length of human BPIFA2 is 249. The bovine proteins have different predicted lengths, BPIFA2A is 236, LOC112441508 is 239, BPIFA2C is 251, BPIFA2D is 246 and BPIFA2B is 240. Most of the genes have nine exons with the start codon (in exon 2) and the stop codon (in exon 8). Only five exons (Exon 2-5) appear in *Bpifa2d* of which only exons 3-5 are of the appropriate size. This observation suggests that *Bpifa2d* is a pseudogene. Any resultant protein generated from this would lack functional domains leading to a loss or alteration of protein function. Each exon size of the other *Bpifa2*-related genes is similar to the exon sizes in the human gene, particularly, the size of exon 7 (shown in *italics)* which is 64bp. This exon is completely conserved in all *BPIF*-related genes and contains the second cysteine residue forming the disulfide bonds as described in Figure 3.4.

3.3.4.2 *BPIFA2* related genes in sheep

Having clarified the complete *BPIFA2*-related gene repertoire in cattle, I next investigated the potential expansion of *BPIFA2* related genes in the sheep using a similar approach and the locus is shown in Figure 3.11. LOC numbers (e.g., "LOC101114720") are displayed for specific uncharacterized genes within the genome. The output identifies nine “complete” genes (highlighted in red boxes) and three pseudogenes (highlighted black boxes). RNAseq data from sheep tissues displayed at the bottom of the figure, provides transcriptional evidence for all the complete genes within the locus. *LOC101116500 (BPIFA4*) and *BPIFB4* flank the *BPIFA2* related gene containing region.

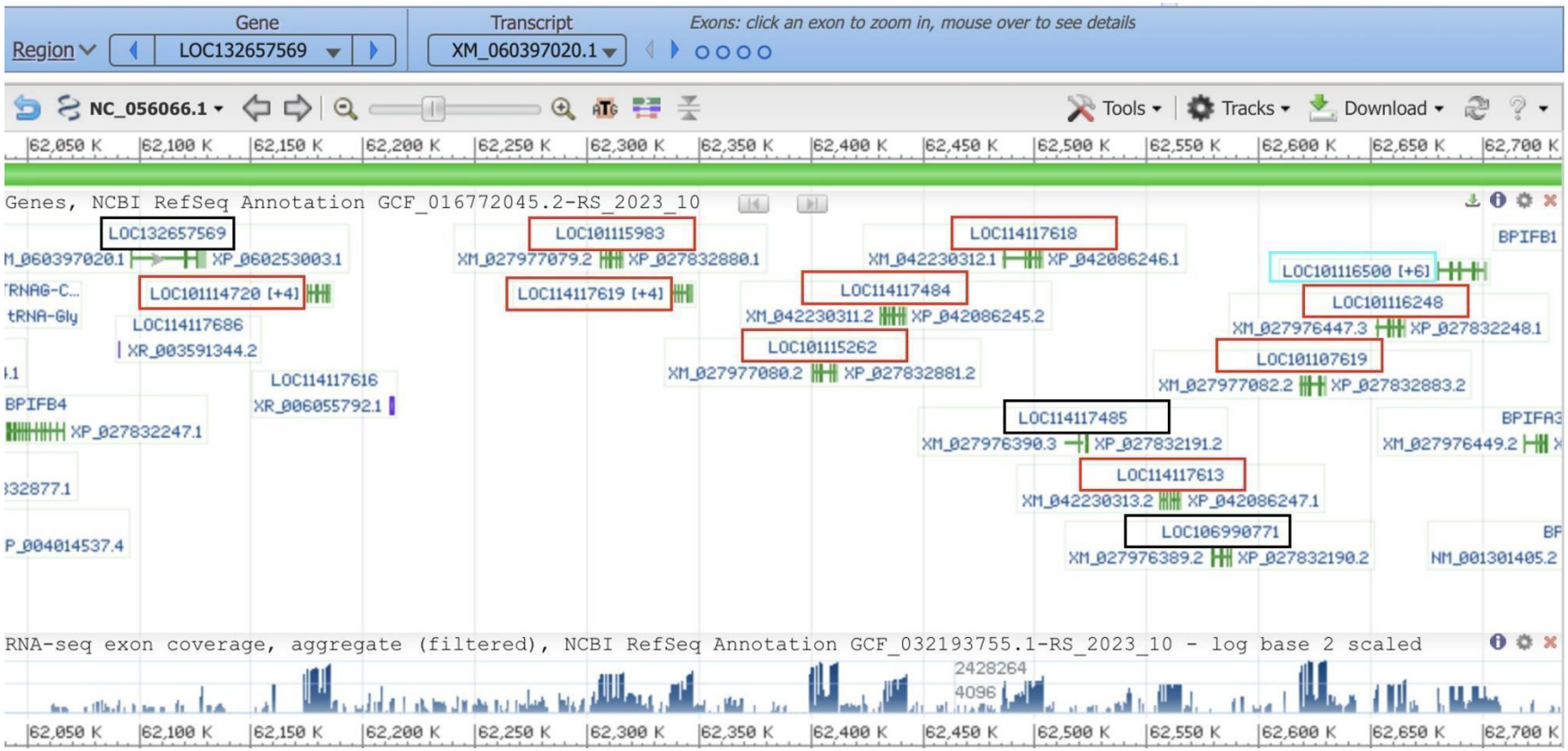


Figure 3.11. The gene expression of *BPIFA2* in sheep. BPIFB1 was used as query genome was visualized by National Center for Biotechnology Information (NCBI) Genome Data Viewer (GDV): https://www.ncbi.nlm.nih.gov/genome/gdv/. Multiple true genes in sheep are highlighted in red boxes and pseudogenes are highlighted in black boxes. *LOC101116500 (BPIFA4)* was highlighted in blue box to confirm the proper genomic location. The arrows on the boxes show the direction of the gene transcription. The RNAseq data provides evidence for all the complete genes within the locus regions.

A further comparison of sheep proteins (*ovis aries*) was performed using ClustalOmega. There are nine uncharacterized BPIFA2-like protein sequences including LOC101114720, LOC101115983, LOC114117619, LOC101115262, LOC114117484, LOC114117618, LOC114117613, LOC101167619 and LOC101116248 were compared in Figure 3.12.

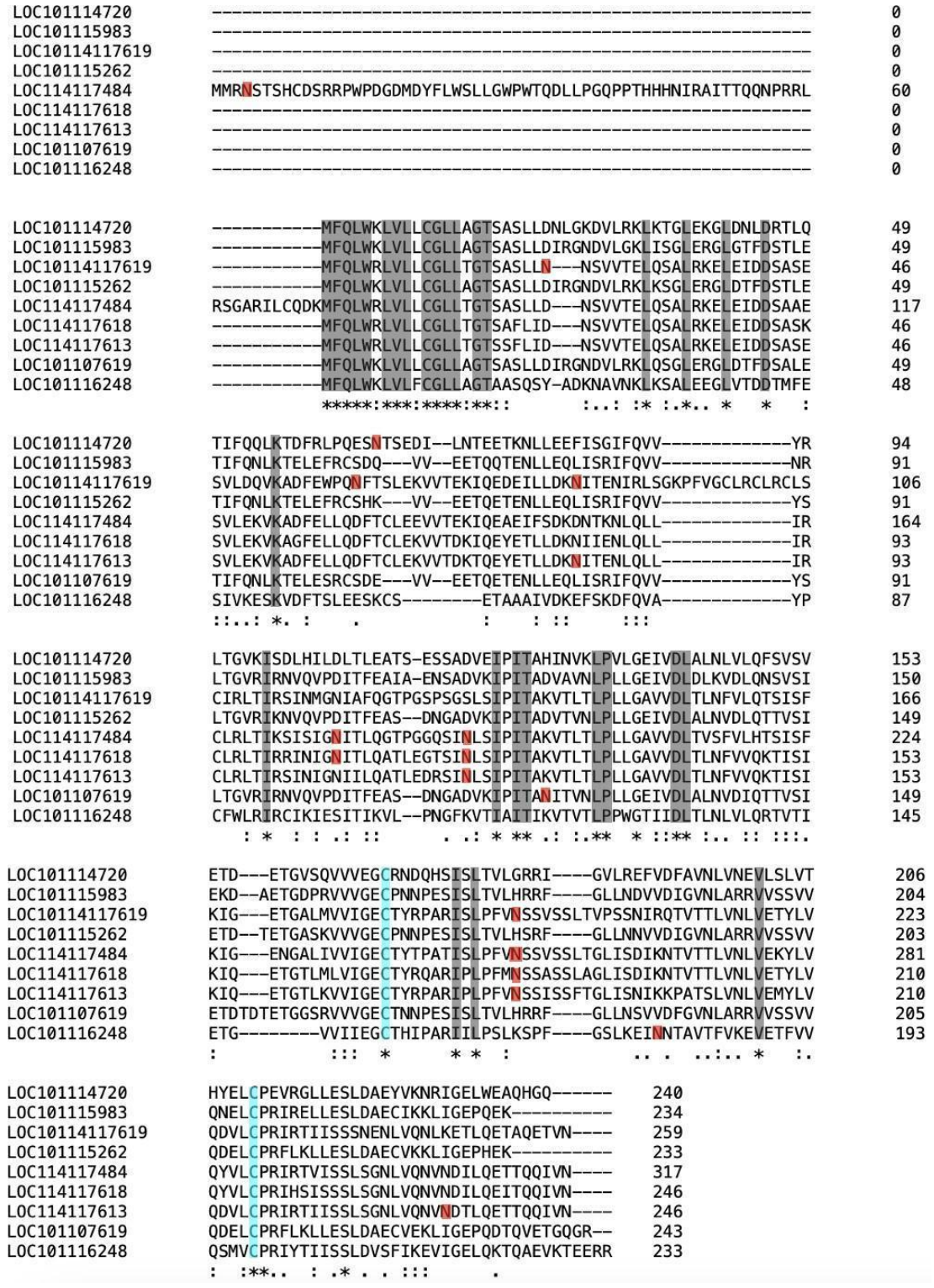


Figure 3.12. Multiple sequence alignments of BPIFA2 proteins in sheep. ClustalOmega alignment tool (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) was used to compare multiple sequence alignments. A star (*) represented completely conserved amino acid residues, a colon (:) represented semi conserved residues and a period (.) represented weakly similar residues. Highly conserved residues are highlighted in dark grey and N-glycosylation sites are highlighted in red. Cysteine residues formed disulfide bonds in protein are highlighted in blue.

There are 29 amino acids completely conserved as highlighted in dark grey and 14 of 29 are located in the signal peptide region (position: 1-20) at the beginning of sequences. The two cysteine residues are conserved in all proteins as highlighted in blue at positions 239 and 285. LOC114117484 has the longest amino acid length (317) and there is sequence extension at the beginning of sequence. This arises from an "extra" 5' exon that is present in this transcript. It is worth noting that a Met residue is also found in the position conserved in all other proteins. It has N-glycosylation sites at positions 4, 177, 190 and 252. LOC114117613 also has four N-glycosylation sites at positions 83, 119, 181 and 234 and the former two sites are at the same positions as in LOC114117484 at position 177 and 190. There are another four N-glycosylation sites in LOC114117619 at positions 23, 61, 83 and 194. The last 194 site is at the same last position in LOC114117484, LOC11411763 and LOC114117618 which has another two sites at position 106 and 119. The length of amino acid sequences in LOC114117613 and LOC114117618 are 246 and LOC114117619 is 259. Both LOC101114720 and LOC101116248 have one N-glycosylation site at position 66 and 178, respectively. The amino acid sequence lengths of LOC101114720, LOC101116248 and LOC101167619 are 240, 233 and 243. There is no N-glycosylation site in LOC101115983 with 234 amino acids and LOC101115262 with 233 amino acids.

Given that I identified a large number of uncharacterised *Bpifa2*-like genes in sheep, I sought to compare exon sizes between sheep and humans. Exon sizes and protein lengths were collected from the NCBI Genome Data Viewer database and were verified by SnapGene Viewer. These are shown in Table 3.2.

Table 3.2 Comparison of exon sizes between human and sheep. The sizes of individual exons (in bp) for sheep genes are compared with humans. The sizes of exon were collected from NCBI Genome Data Viewer database and were verified by SnapGene Viewer: https://www.snapgene.com/snapgene-viewer.

Gene	Protein Length		**Exon 1**	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	*Exon 7*	Exon 8	**Exon 9**
hBPIFA2	249		55	168	145	108	153	82	*64*	78	188
LOC101114720	240			160	142	105	150	82	*64*	72	174
LOC101115983	234		29	161	133	105	153	82	*64*	60	183
LOC114117619	259		29	152	187	108	150	94	*64*	67	182
LOC101115262	233		31	161	133	102	153	82	*64*	60	183
LOC114117484	317	70	145	152	148	108	150	94	*64*	79	183
LOC114117618	246		29	152	148	108	150	94	*64*	79	183
LOC114117613	246			152	148	108	150	94	*64*	79	188
LOC101167619	243		29	161	133	102	159	82	*64*	71	179
LOC101116248	233		29	163	124	102	135	82	*64*	78	138

Table 3.2 shows nine complete *BPIFA2*-like genes identified in sheep. These are described as LOC numbers as they have not yet been characterized and named. The human *BPIFA2* gene was used as a root to compare each exon size with sheep variants. The protein sequence length of the sheep proteins is generally comparable to that of humans, with the exception of LOC114117484, which contains one additional exon compared to other genes. The other genes have nine exons and there is good conservation of each exon size is also comparable to those in humans, particularly with exon 7 consistently measuring 64 in size. Again, the exon 7 is conserved in most complete genes as it contains a cysteine residue that forms disulfide bonds in the protein.

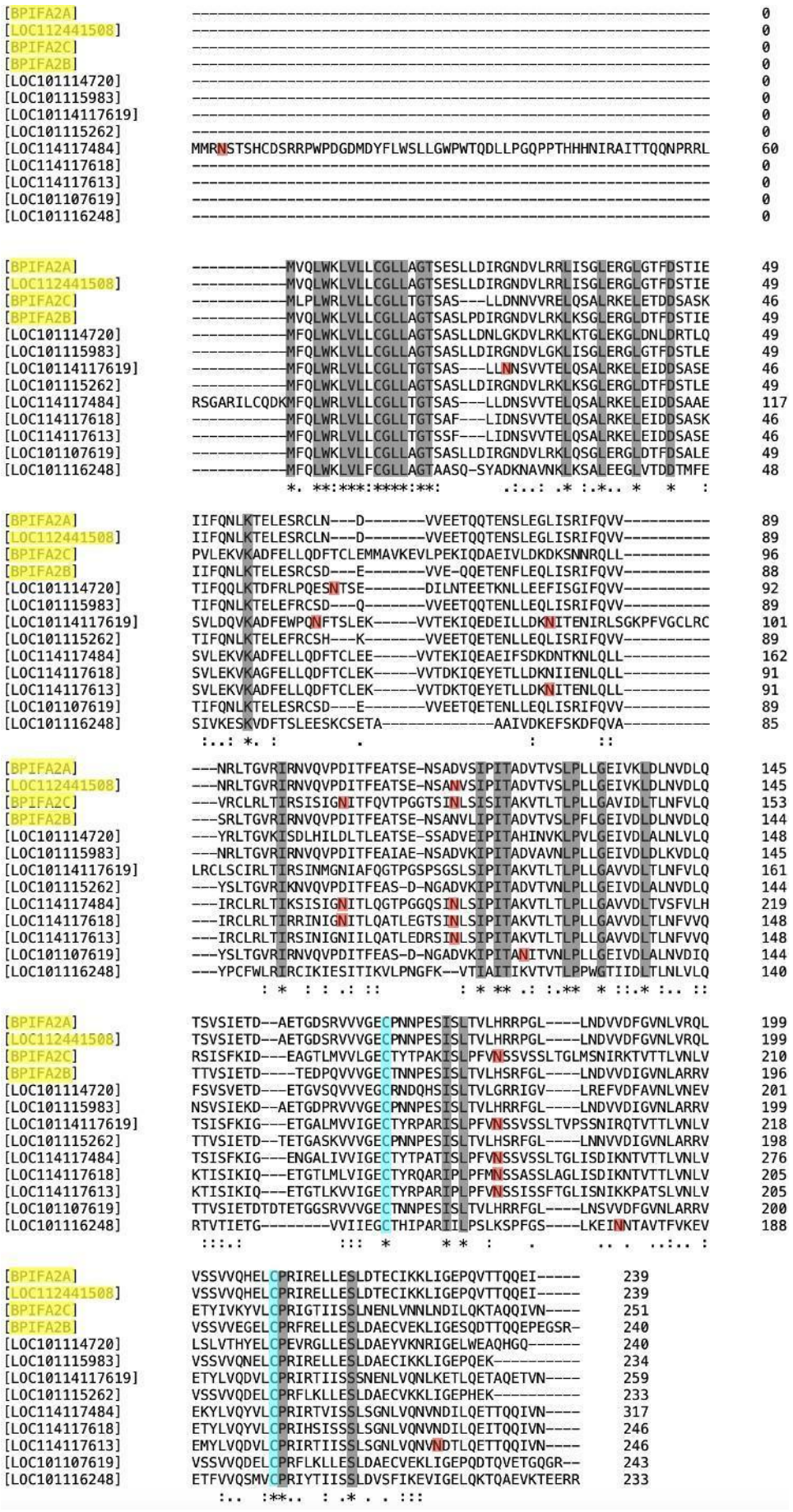

[BPIFA2A]
[LOC112441508]
[BPIFA2C]
[BPIFA2B]
[LOC101114720]
[LOC101115983]
[LOC10114117619]
[LOC101115262]
[LOC114117484]
[LOC114117618]
[LOC114117613]
[LOC101107619]
[LOC101116248]

Figure 3.13. Comparison of multiple sequence alignments of BPIFA2-like protein between cattle and sheep. ClustalOmega was used to compare multiple sequence alignments. A star (*) represented completely conserved amino acid residues, a colon (:) represented semi conserved residues and a period (.) represented weakly similar residues. Highly conserved residues are highlighted in dark grey and N-glycosylation sites are highlighted in red. Cattle sequences are highlighted in yellow. Cysteine residues formed disulfide bonds in protein are highlighted in blue.

To look at the evolutionary relationship between the BPIFA2-related protein from these two species I generated a MSA with cattle and sheep sequences using ClustalOmega. Figure 3.13 shows that the similarity across the amino acids with only 29 of residues completely conserved as highlighted in dark grey. Twelve of these are present in the signal peptide region which suggest evolutionary conservation across the sequences of protein. As expected, the two cysteine residues are conserved. It is clear that the N-glycosylation sites are varied in these two species.

To investigate the evolutionary relationships of BPIFA2-like proteins in cattle and sheep a phylogenetic tree was generated by the BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Human BPIFA2 was used as the root to the query sequence as highlighted in yellow, and the cow proteins BPIFA2A, BPIFA2B, BPIFA2C and LOC112441508 proteins are highlighted in red. The tree represents a single lineage that diverges into two ruminant species. The blue nodes represent common ancestors or significant evolutionary events, with the branching distance reflecting the evolutionary relationships among the proteins. The lengths of the horizontal lines indicate genetic distance or evolutionary time.

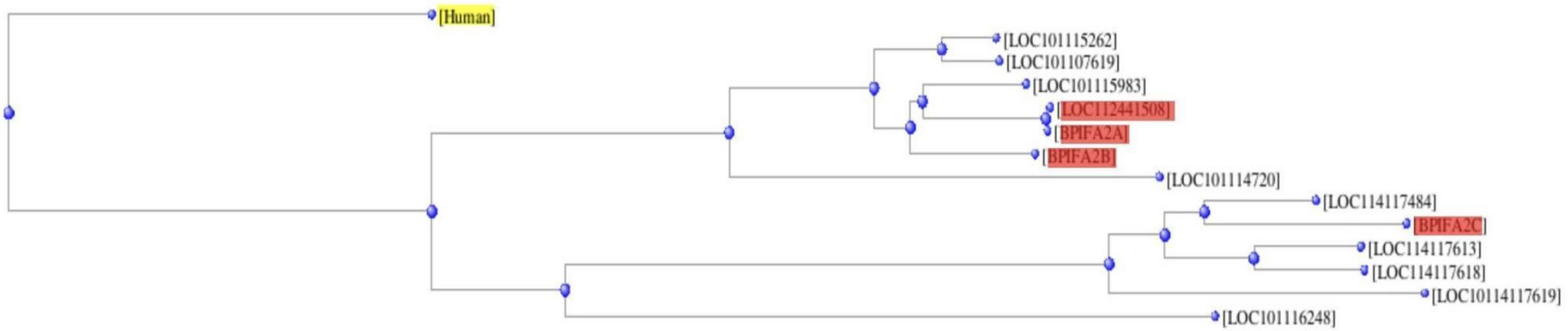


Figure 3.14. A phylogenetic tree of evolutionary distance in cattle and sheep. A phylogenetic tree was generated by BLAST ((https://blast.ncbi.nlm.nih.gov/Blast.cgi)) to compare the evolutionary distance between cattle and sheep. Human BPIFA2 protein sequence as the query as highlighted in yellow and covering cattle and sheep branches. BPIFA2A, BPIFA2B, BPIFA2C and LOC112441508 proteins expressed in cattle were highlighted in red.

Figure 3.14 shows that LOC101115262 is closest to LOC101117619 and less close to LOC101115983, indicating a close evolutionary relationship between them. LOC112441508 seems to be a BPIFA2A-like protein rather than the longer distance between BPIFA2A and BPIFA2B, distinct from LOC101114720. It is interesting that BPIFA2C, in cattle, has a closer relationship with sheep LOC114117484 than other orthologues in cattle. LOC114117613 and LOC114117618 also have a closer distance, but these two are distinct from LOC10114117619 and then further related to LOC101116248.

3.3.5 Genetic variations of human BPIFA2

The comparative analysis that I have performed above has confirmed that BPIFA2 and its' paralogues are the most rapidly evolving members of the wider BPIF-protein family and suggest that this variation may in some way be associated with different salivary function across species. As my study is really focused on functional analysis of human BPIFA2 in saliva, I considered it to be important to investigate the genetic variation of human *BPIFA2*. The most important consideration in studying genetic modification or protein mutations is the variations that might involve changes in the structure or modify the function of the protein.

In order to identify the genetic variations and their locations in the human *BPIFA2* coding region, the Ensembl database (https://www.ensembl.org/index.html) was used to visualise the cDNA sequence of BPIFA2. The output in Figure 3.15 shows the comparison between the transcript gene sequence of *BPIFA2* on the top and their translated amino acid sequences below to identify variations and predict their impact on the encoded protein. This transcript is a product of gene *BPIFA2* (ENSG0000013150.11) and is made up of 9 exons which are 2

non-coding exons and 7 coding exons. Only the coding region is displayed in this image. A large number of variations (as highlighted in different colours on the top section) in the transcript sequence can be seen in the figure. Many of these variants have the potential to alter the amino acid sequence and can result in changes to the protein's structure and function. For example, a non-synonymous mutation, one specific type of single nucleotide polymorphism (SNPs) changes the amino acid sequence, possibly impacting the protein's function.

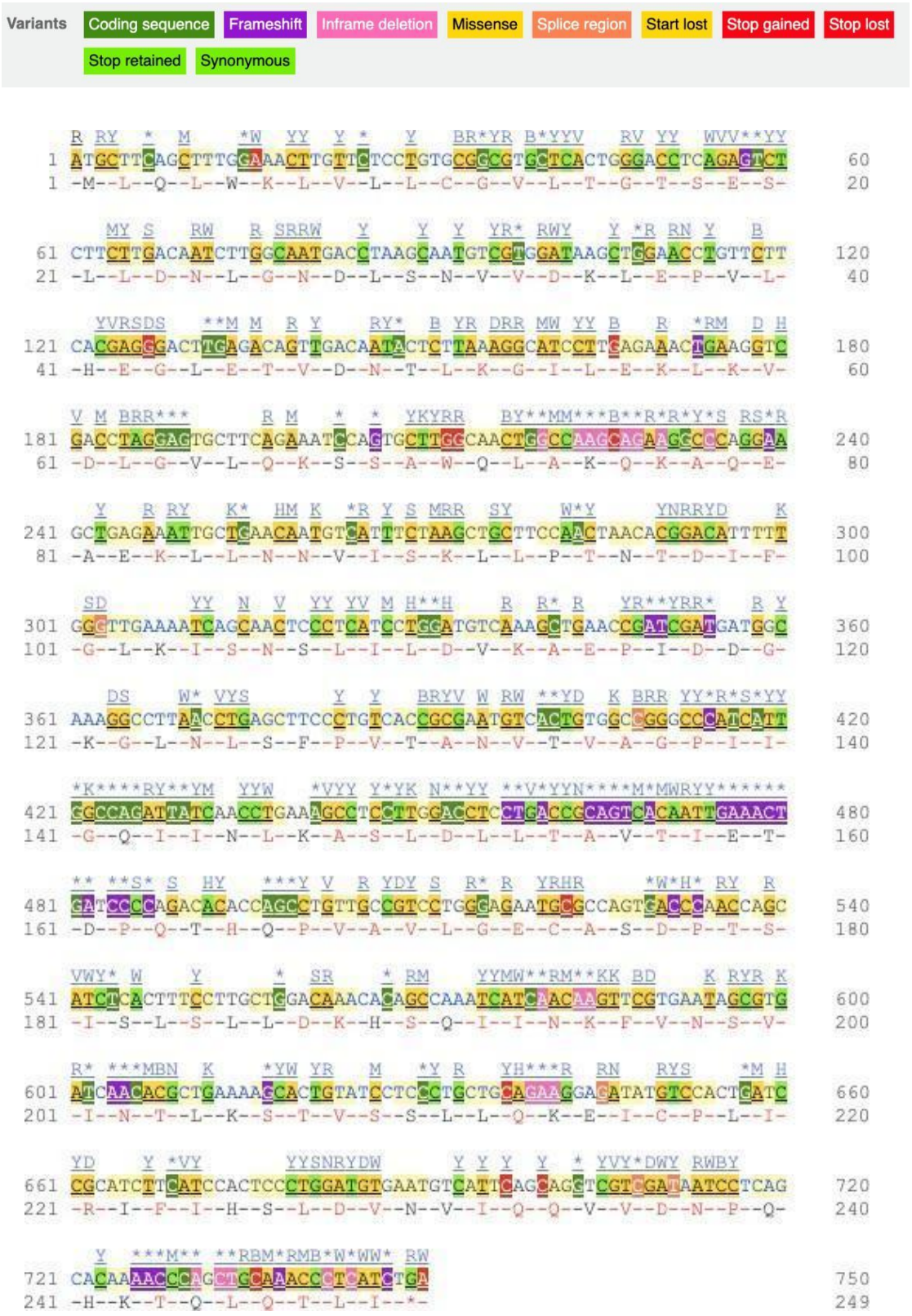


Figure 3.15. The locations and genetic variations in *BPIFA2*. The transcript was derived from the ENSEMBL database (ENST00000354932.6) using the human BPIFA2 as a query. The annotation of variants is displayed in the top section.

To more fully investigate the location of variations within the coding region of *BPIFA2*, the gnomAD (Genome Aggregation Database) v4.0.0 brower (https://gnomad.broadinstitute.org/) was used to visualize the gene as shown in Figure 3.16. gnomAD aggregate exome and genome sequencing data from a variety of large-scale sequencing projects and enables interrogation of a large amount of genomic information in a web based format (Chen *et al.*, 2024).

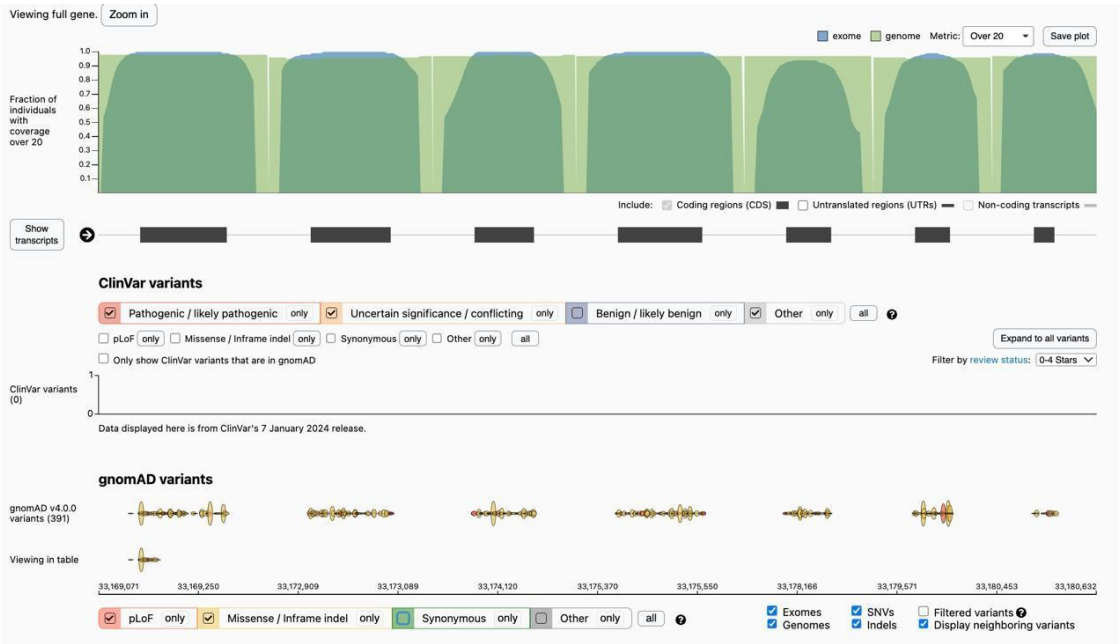


Figure 3.16. The position of non-synonymous variants on the coding region of *BPIFA2.* The non-synonymous variations of BPIFA2 were extracted from the gnomAD v4.0.0 database available at: https://gnomad.broadinstitute.org/.

The top panel displays the fraction of individuals with sequence coverage at different positions within the *BPIFA2* gene in the coding region (made of seven coding exons). The lower part of the figure shows variants from the gnomAD database mapped onto the gene structure. Each symbol represents a variant, and they are coloured according to the type: PLoF (Predicted Loss of Function), missense or in frame deletions or insertions, synonymous mutation, or but not used in my analysis, other types of variants including a splice or intronic

modifications. In Figure 3.16, there are 391 variation events related to non-synonymous variants on the coding region of the gene (displayed on the site as an interactive list). Non-synonymous mutations are commonly influenced by natural selection. A nonsense mutation is a specific form of non-synonymous mutation that results in a replacement of a codon that typically codes for an amino acid with a stop codon in the genetic sequence.

Manual interrogation of this list allowed a detailed exploration of five most prevalent variations and information related to these were extracted from the Ensembl and GnomAD databases based on allele count (frequency of mutation) and number of homozygotes (homozygote penetration) to analyse the SNPs in various ethnic groups as shown in Figure 3.17.

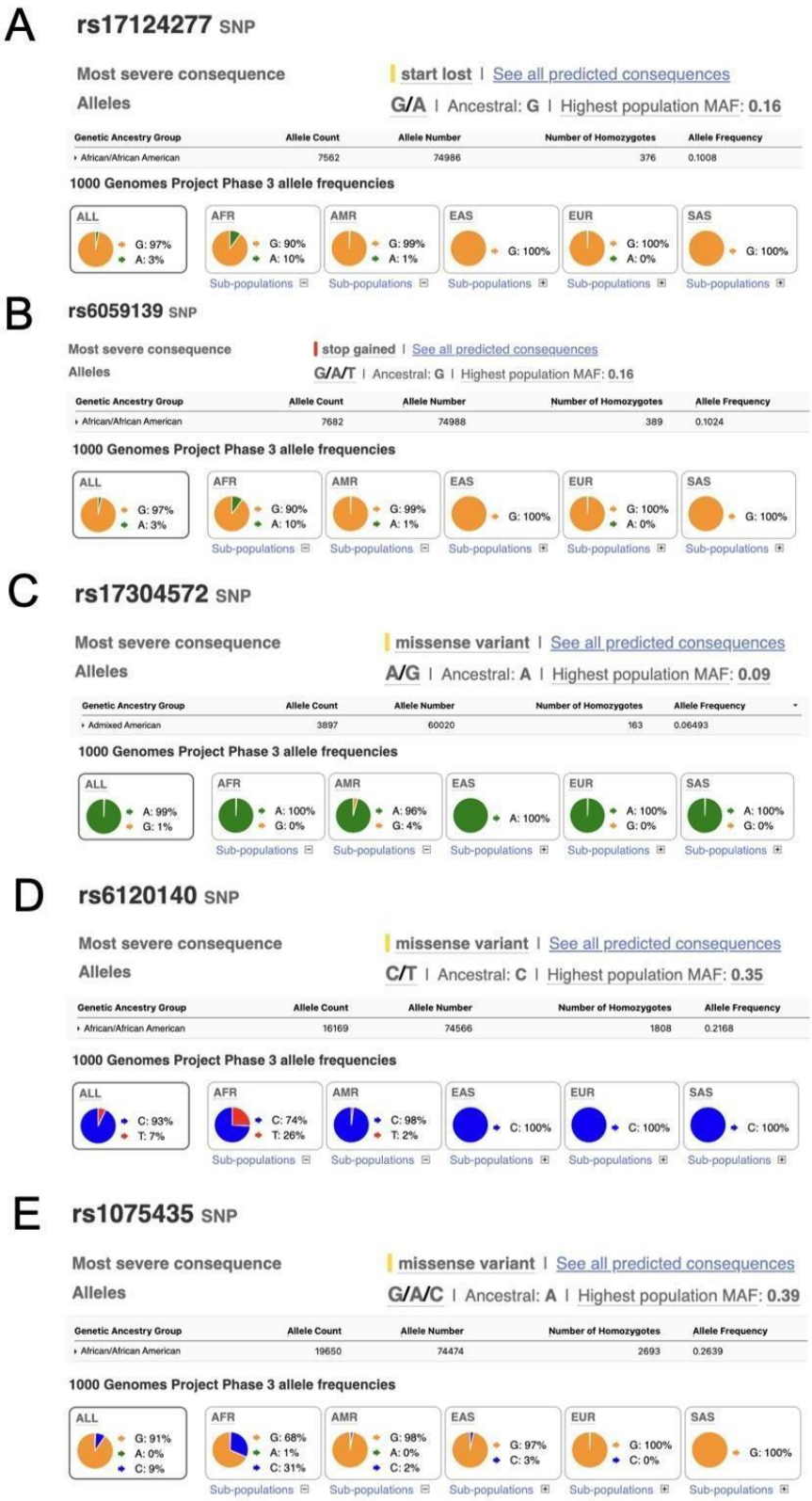


Figure 3.17. The five most common variants in human BPIFA2. The distribution of ethnic groups were collected from the gnomAD v4.0.0 database available at: https://gnomad.broadinstitute.org/. Each panel (A to E) presents data on a difference indicated by its reference SNP ID (rsID) at the left top positions.

The output showed that the variant rs17124277 in panel A generated the loss of the start codon which indicated a mutation at the start codon leading to no protein translated. The figure shows 'stop gained' in panel B (rs6059139), which suggests a mutation leading to a premature stop codon resulting in the absence of production of a complete protein, and 'missense variant' in panels C (rs17304572), D (rs6120140), and E (rs1075435), indicating non-synonymous mutations that result in a change to the amino acid sequence of the encoded protein. For example, the amino acid of K (Lysine) was changed to E (Glutamate) in rs17304572, R (Arginine) was replaced by C (Cysteine) in rs6120140 and Valine (V) was changed to Leucine

(L) in rs1075435. The alleles involved in the SNP are listed with the ancestral allele mentioned first, followed by the variant allele, such as "G/A" for rs17124277 in panel A, "G/A/T" for rs6059139, "A/G" for rs17304572, "C/T" for rs6120140 and "G/A/C" for rs1075435. The population frequencies are represented by pie charts for each panel, showing five major population groups: African/ African American (AFR), Admixed American (AMR), East Asian (EAS), European (EUR), and South Asian (SAS). These charts provided valuable insights into the distribution of these variants among various human populations and the overall population. Potentially, the most significant result is the mutation leading to the absence of the start codon resulting in the absence of protein production that occurs most frequently (and almost exclusively) among individuals of African/ African American (AFR) origin. The observation of missense variants mostly happens in Admixed American (AMR). These interesting results indicate that some individuals of African descent will be unable to produce BPIFA2 in their saliva due to a genetic change, even though they have functional salivary glands. That might suggest BPIFA2 does not play any biological role in that

population due to a range of biological and evolutionary processes, health conditions, and selected pressures or adaptations to environmental factors.

3.4 Discussion

BPIFA2 is the most divergent protein among BPIF family members expressed in multiple species and it has been hard to identify the orthologous or paralogous relationship based on low sequence similarity (Bingle, Seal and Craven, 2011). There has been a lack of investigations into BPIFA2 since the advent of more extensive annotated genomes. In this chapter, I set out to update the knowledge around BPIFA2. Multiple alignment tools and databases were used to investigate the similarity of the BPIFA2 amino acid sequences in mammals which show there is a low level of sequence similarity across mammalian phylogeny. My analysis also suggests that N-glycosylation is not seen on all BPIFA2 proteins and where it is seen sites are distributed at various locations across the sequence. The structural modeling work shows that it is not the case that these sites are located in one place within the structure of the protein. Glycosylation might be important to the function of BPIFA2, and this remains to be defined. As was expected, the two cysteine residues are always conserved in the protein. The second cystine is always encoded by residues 4, 5 and 6 of exon 7 of the genes. This exon is completely conserved across all *BPIFA2* and *BPIFA2*-related genes. The exception to this is in the cetacean lineage where incomplete BPIFA2 proteins are seen due to lineage specific exon loss. The cysteine residues have the ability to create a disulfide bond, which is also consistently preserved in LBP/BPI family proteins. (Beamer, Carroll and Eisenberg, 1997, 2008). This observation indicates that these cysteines have important functions in the three-dimensional structure and in the biological function.

The preservation of cysteine positions in the protein is crucial for understanding its structural-functional relationships and predicting the effects of genetic variations across species.

The differential localisation of N-glycosylation sites in primates suggests that conducting additional investigation to examine the function of BPIFA2 between primates and humans might be interesting. A published paper reported that there was less BPIFA2 expressed in human saliva compared to great apes (chimpanzee and gorilla) (Thamadilok *et al.*, 2020), which suggests humans may need less of BPIFA2 protein. The difference in the levels of BPIFA2 between humans and primates may be due to human adaptation to dietary changes and evolutionary processes.

One of the most striking observations from the comparative genomic point of view is that there is only a *Bpifa2* pseudogene seen in cetaceans such as dolphins and whales due to a deletion of a genomic region that causes the loss of the 3'exons of the gene loss. Loss of exons in this region could decrease mRNA stability as it can be affected by elements in the 3' untranslated regions (3' UTRs). And the loss of the last exons introduces a premature stop codon, this could trigger nonsense-mediated decay, leading to the reduction of mRNA and therefore less protein being produced. Proteins may lose important functional domains encoded by the missing exons, potentially rendering the proteins inactive or less functional.

In addition to the loss of *Bpifa2* 3' exons there is also loss of the whole region of the genome in cetaceans that encodes *Bpifa4*, another salivary gland enriched member of the BPIFA subfamily. A single nucleotide loss has been identified in *BPIFA4P* (Bingle, Bingle and Craven, 2011), resulting in it becoming a pseudogene in humans, although it is present as an

authentic gene in primates and horses.

The loss of the *BPIFA4* gene in both humans and cetaceans when compared to chimpanzee might be a case of convergent evolution, where unrelated lineages (humans and cetaceans) have independently lost a gene due to similar selective pressures or environments that make the gene unnecessary or disadvantageous. The *BPIFA4* gene may have been lost as part of an evolutionary trade-off, where the loss of one function provides some other advantage or is linked to the loss of a no longer necessary function. For example, an olfactory receptor gene associated with the sense of smell in whales which rely less on their sense of smell (Kishida *et al.*, 2015). In this case, the absence of the *Bpifa4* gene and the loss of exons of *Bpifa2* in cetaceans can be understood as an evolutionary adaptation related to their lifestyle, diet, and the way they ingest food. Cetaceans feed by swallowing their prey whole rather than chewing it. Saliva is less important when food is not chewed because its primary functions in terrestrial animals include lubricating and beginning the digestion of food in the mouth. It should be noted that while cetaceans may not have fully functional salivary glands similar to those of terrestrial mammals, this does not necessarily mean they lack them entirely (Huelsmann *et al.*, 2019).

In terms of *BPIFA4* loss in human compared to chimpanzee which results in no BPIFA4 being secreted into saliva, it has been shown that there are lower levels of BPIFA2 in human saliva compared to great apes like chimpanzees and gorillas (Thamadilok *et al.*, 2020), possibly due to humans requiring less BPIFA2 protein as they adjust to dietary changes. The less BPIFA2 might reflect differences in diet, environment, or behavior between human and chimpanzee.

Humans eat cooked food which is much easier to digest and extract nutrients from when compared to the eating of raw food by primates. Similarly, the loss of *BPIFA4* might have provided some selective advantage during human evolution.

The presence of six *BPIFA2* related genes in cattle confirm and extend data that suggests a gene family expansion. Here, I was able to better define the complement *Bpifa*2-related genes in cows due to there now being a much better quality annotated genome. In this analysis, the gene previously identified as *Bpifa2d* (Wheeler *et al.*, 2007) has been shown to be pseudogene. In the original bovine *BPIF* gene study, it was only a gene that was the result of a prediction performed by the Genescan gene prediction program. It is notable that there is one new complete gene in the locus, *LOC112441508*. This gene contains a full complement of exons and has been shown to be transcribed in full by analysis by RNAseq which is able to reveal sequences that are being actively transcribed in the cells or tissues being studied.

The most likely reason for having multiple *Bpifa2*-related genes in cattle is gene duplication The multiple copies of the gene may specialize for various functions or in acquiring a new function which is different from the one it originally had. This is termed neofunctionalization and can lead to an expansion of the functional distribution of the gene family. *BpifaA2*-like genes are expressed in the bovine salivary glands and part of the rumen, and they may well play a role in digestion, oral health, or immune function in the oral cavity. Although examination of the expression patterns of the potential new gene across different tissues, developmental stages, or in response to various stimuli are still required and experimental validation such as gene cloning and expression in an effective system to verify that it is able to produce functional proteins *in vitro* and quantitative PCR (qPCR) can be used to validate

and quantify expression levels.

Multiple variants of *Bpifa2*-related genes in sheep, which is another classical ruminant, were also identified. This builds on the observation generated using partial ESTs (Expressed Sequence Tags) (Wheeler *et al*., 2011). My study has identified a total of nine functional genes and three non-functional pseudogenes. Additional investigation is required to completely clarify the role of BPIFA2 proteins in the functioning of ruminant digestion and its possible consequences for rumination and general digestive well-being in these animals as the dietary habits of ruminant are distinct from other land-dwelling mammals.

By inspection of the sequence-based phylogenetic tree in cattle and sheep proteins were assigned to the appropriate position according to the evolutionary distance. Their positioning on the tree could provide valuable insights into the evolutionary patterns of protein activities. BPIFA2C was identified in cattle and found to share a common evolutionary ancestor with LOC114117484 in sheep. As the multiple sequence alignments shown in Figure 3.13, it has been observed that BPIFA2C contains three N-glycosylation sites that are located exactly at the same place as the last three of four N-glycosylation sites in LOC114117484 which suggests the glycosylation sites has been maintained through evolution and the function may be conserved between the two proteins. This finding indicates a promising opportunity to investigate this unique BPIFA2-like proteins in cattle and the glycosylation sites have remained unchanged throughout evolution suggesting that the function of these sites may be preserved in both proteins in two ruminants. Further research could expand to a bigger group of ruminant mammals including goats, deer, and giraffe to give a better understanding of the expansion of this group of proteins in ruminants.

Given that I have shown how variable BPIFA2 and BPIFA2-related proteins are in mammals and the loss of the gene in some species (as well as the loss of *BPIFA4* in man) it became important to look at genetic variation in the human gene. There is a lot of genetic data that can be investigated to look at such variation and by doing this, I was able to identify some abundant non-synonymous mutations within *BPIFA2*. This investigation of the human mutations identified a start site mutation and a premature stop codon mutation that would result in variable levels of expression or lack of expression of the *BPIFA2* gene in specific ethnic groups such as people of African origin. This suggests that genetic variants in humans may impact the function of *BPIFA2* and are specific to distinct ethnic groupings. As I previously mentioned, the disparity in the level of BPIFA2 between humans and great apes may be attributed to the evolutionary and dietary alterations. Therefore, it can be hypothesised that humans may have required less BPIFA2 protein function as we have evolved.

Although the developments and methodologies significantly contribute to the ongoing expansion of genetic databases such as NCBI and our understanding of genomic complexity of *BPIFA2*, extensive experimental validation still is necessary to support my bioinformatic analysis. Such an analysis may yield relationships between *BPIFA2* variants and specific traits and or diseases.

In summary, this chapter has provided a detailed bioinformatic background that has confirmed the diversity of the BPIFA2 branch of the wider BPIF family. It has identified gene loss and gene expansion in lineages that have different digestive traits and requirements for saliva as well as identified mutations in the human gene that may affect protein secretion and

function. It has allowed me to choose a number of protein sequences from different species to include in the functional work described in this thesis.

Chapter 4: BPIFA2 expression in human saliva and synthesis and analysis of recombinant BPIFA2.

4.1 Introduction

The composition of saliva can be influenced by food and beverages (Bozorgi, Holleufer and Wendin, 2020). For example, consuming food can trigger the creation of saliva and alter its consistency, while specific meals or beverages may introduce extra components into the saliva, such as sugars and acids (Bozorgi, Holleufer and Wendin, 2020; Simões *et al.*, 2021). In order to eliminate the influence of these variables, it is standard to request that volunteers abstain from eating for 2 hours prior to providing a saliva sample. This facilitates the collection of sample that accurately reflects the initial condition of an individual's saliva, uncontaminated by recent dietary factors (Al Habobe *et al.*, 2024). Stimulated saliva produced in response to sensory stimuli such as taste, smell, or chewing, tends to have a higher flow rate and contains higher concentrations of certain ions, proteins, and enzymes in comparison to unstimulated saliva (Al Habobe *et al*., 2024).

Human BPIFA2 is predominantly expressed in major salivary glands and secreted in saliva and although the predicted molecular weight is 25 kDa, as an N-glycosylated protein, the true molecular weight is often greater. Glycosylation is a post-translational modification in which carbohydrate molecules (glycans) are covalently attached to specific amino acid residues on proteins. There are two primary types of glycosylation, N-linked and O-linked glycosylation. Glycans attach to the nitrogen atom of asparagine (Asn) residues within a specific consensus sequence (Asn-X-Ser/Thr, where X is any amino acid except proline) to form N-linked

glycosylation and glycans are attached to the oxygen atom of serine (Ser) or threonine (Thr) residues, often without a defined consensus sequence, to form O-linked glycosylation. Glycosylation plays a crucial role in protein folding, stability, localization, and function and glycosylated regions on proteins play key roles in immune modulation, including pathogen recognition.

Previous data has shown that expression of human BPIFA2 varies within a population and that there are multiple isoforms, probably due to differential glycosylation (Bingle *et al.*, 2009). Human BPIFA2 contains 249 amino acids with two N- glycosylation sites on residues 124 and 132 (Gorr *et al.*, 2011). To investigate the relationship between glycosylation sites and the, as yet, unknown biological function of BPIFA2, four FLAG-tagged expression constructs of BPIFA2 were previously developed in our laboratory; wild-type, single glycosylation mutants and a double glycosylation mutant in which the N residues were replaced by A. As discussed in chapter 3 BPIFA2 orthologues are highly divergent between species, with the human/mouse proteins being 32% identical and some species having evolved a number of BPIFA2 paralogs, such as bovine salivary proteins which are encoded by four genes *BSP30A*, *BSP30B*, *BSP30C* and *BSP30D* and are found in bovine saliva. Also as discussed in chapter 3 five constructs from the mouse, armadillo, macaque, squirrel monkey and dog were generated in this study in order to determine the function of BPIFA2 across the mammalian phylogeny. Transient calcium phosphate transfections, centrifugal concentration and Anti-FLAG M2 affinity gel purification was carried out in order to produce recombinant protein from the above outlined BPIFA2 constructs.

4.2 Aims

- To enhance understanding of the expression and variability of BPIFA2 in human saliva, contributing to insights into its biological and functional roles.
- To investigate the glycosylation sites of BPIFA2 in human saliva and understand their variability, contributing to insights into its structural and functional roles.

To develop and purify recombinant BPIFA2 proteins using mammalian expression systems for subsequent functional and structural analyses.

4.3 Materials and methods

The relevant materials and methods utilised in this chapter are detailed in chapter 2, please refer to the following:

- To mitigate these variations in saliva constituents due to diurnal and dietary influences and to enhance comparability between individuals, unstimulated saliva samples were collected in a manner that ensured consistency in the time of collection. Human saliva sample pretreatment was carried out as described in section 2.2.1. Western blot and dot blots were carried out as described in sections 2.2.3 and 2.2.4.
- Deglycosylation analysis of human saliva was carried out using PNGase F enzyme

treatment as described in section 2.4.1.

- Gene cloning of human, armadillo, mouse, macaque, squirrel monkey and dog BPIFA2 was carried out as described in section 2.3.2.
- Recombinant protein was produced using the calcium phosphate transfection method as described in section 2.3.10 and purified from conditioned media using anti-FLAG affinity resin as described in section 2.3.12. The purified proteins were detected by Coomassie blue staining as described in section 2.2.6.1 with specific protein identification by Western blotting as described in section 2.2.3. Protein concentration was initially quantified by the BCA assay and ELISA assay as described in section 2.2.7 and 2.2.8, and then quantified by dot blots and densitometry as described in section 2.2.4.

4.4 BPIFA2 expression in human saliva

4.4.1 Expression levels of BPIFA2 in whole human saliva

Whole human saliva was collected from eighteen healthy, free of oral disease, volunteers and centrifuged to remove cells and debris. In Figure 4.1, 1µl of clarified saliva was dotted onto PVDF membranes and the native expression levels of BPIFA2 detected using specific antibodies originally generated by Eurogentec and previously validated in our lab (Bingle *et al.*, 2009a). The resultant intensity of each dot was determined using densitometry analysis (Figure 4.2). Variations in expression levels were observed among different individuals. For example, greater levels of expression of BPIFA2 were seen in samples from volunteers 1, 7, 8, 9 and 18 in comparison with other volunteers. No BPIFA2 expression was seen in the sample from volunteer 10. As shown in Table 4.1, 27% of the volunteers enrolled in this study were

male and all were aged between 23 and 60 years. Volunteers 1, 7 ,8 and 9 were female and aged 30, 60,33 and 50, respectively and volunteer 18 was a 28-year-old male. Volunteer 10 who did not express BPIFA2 in their saliva (Figure 4.3) was a 55-year-old male. Variations in the *BPIFA2* gene might result in non-functional or absent protein expression. Genetic mutations, such as promoter deletions or nonsense mutations, could downregulate or silence BPIFA2 expression. Similar cases have been observed in a western African ethnic group as described in chapter 3 section 3.3.5, where BPIFA2 expression is potentially lost. These interesting results suggest there is no real gender or age difference in BPIFA2 expression in human saliva and it would be of great interest for further study.

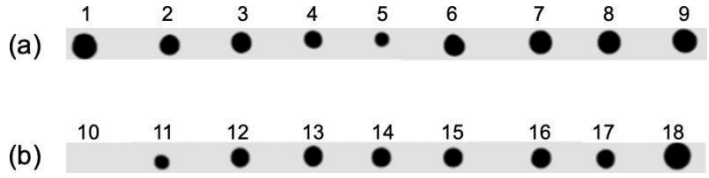


Figure 4.1. Dot blot analysis of BPIFA2 expression levels in whole human saliva with BPIFA2(B) primary antibody. Dot blot analysis was used to detect the native expression levels of BPIFA2 in whole saliva derived from 1-9 (a) and 10-18 (b) healthy volunteers. (Representative image from n=3).

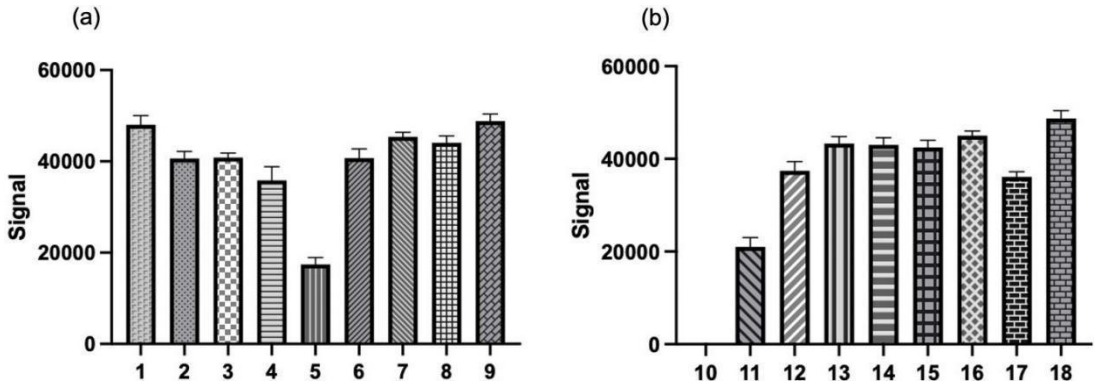


Figure 4.2. Densitometry of dot blot for BPIFA2 expression in whole human saliva. Densitometry analysis derived from dot blots comparing BPIFA2 expression in whole human saliva from 1-9 (a) and 10-18 (b) healthy volunteers. BPIFA2(B) antibody was used for detection of the protein. The X-axis represents the sample number. The Y-axis represents the relative intensity of protein bands detected. Data presented as the mean ± SD (N=3).

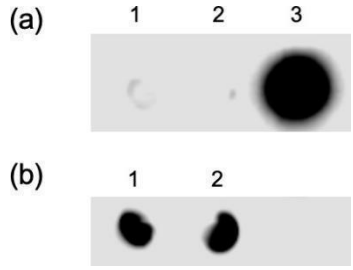


Figure 4.3. Validation of dot blot for BPIFA2 expression in whole human saliva BPIFA2(B) primary antibody. Dot blot analysis was used to validate the native expression levels of BPIFA2 in old (1) and new (2) whole saliva samples from volunteer 10 and old sample (3) from volunteer 17 (a). Expression of BPIFA2 in new saliva samples form volunteers 17 (1) and 11(2) (b).

Table 4.1. The gender and age distribution of saliva samples.

Volunteers Sample Number	Gender	Age
1	Female	30
2	Male	23
3	Female	32
4	Female	38
5	Male	50
6	Female	36
7	Female	60
8	Female	33
9	Female	50
10	Male	55
11	Female	23
12	Female	26
13	Female	48
14	Female	40
15	Female	40
16	Male	27
17	Female	60
18	Male	28

The native expression of BPIFA2 from the same human whole saliva samples was also detected by Western blotting using the same BPIFA2(B) antibody. In comparison to the dot blot analysis, Western blotting allowed us to see multiple bands; similar results have been seen in previous studies (Bingle *et al.*, 2009b). The multiple bands indicate that there are several protein isoforms of BPIFA2 present in saliva; these variations are probably due to differential glycosylation at N-linked and O-linked glycosylation sites.

The Western blot results demonstrated similar differences in expression levels of BPIFA2 in the eighteen healthy volunteers (Figure 4.4) to that seen with dot blots, and densitometry was again employed to quantify the intensity of bands (Figure 4.5). There were clearly multiple bands in volunteers 1, 2, 7, 8, 9, 11, 12, 13, 17 and 18 compared to other volunteers and, consistent with Figure 4.1, there was no BPIFA2 expression in volunteer 10. Less intensive bands expressed by volunteers 3, 4, 5, 6, 14, 15 and 16 again demonstrated consistent results

between the Western blot and dot blot analysis.

Densitometry analysis indicated that there were significant individual differences in native BPIFA2 expression in human whole saliva. The Western blot analysis revealed the presence of several isoforms of BPIFA2 in saliva, whereas the dot blot analysis provides a more direct assessment of the total expression of BPIFA2 in human saliva.

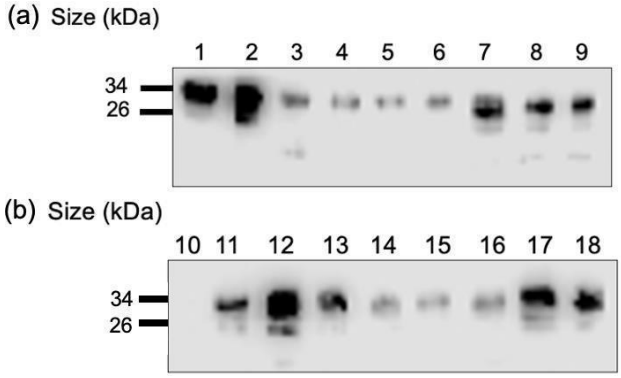


Figure 4.4. Western blot analysis of BPIFA2 expression levels in whole human saliva. Western blot analysis was used to indicate the expression levels of BPIFA2 in whole saliva derived from 1-9 (a) and 10-18 (b) healthy volunteers. BPIFA2(B) antibody used for detection of specific protein (Representative image from n=3).

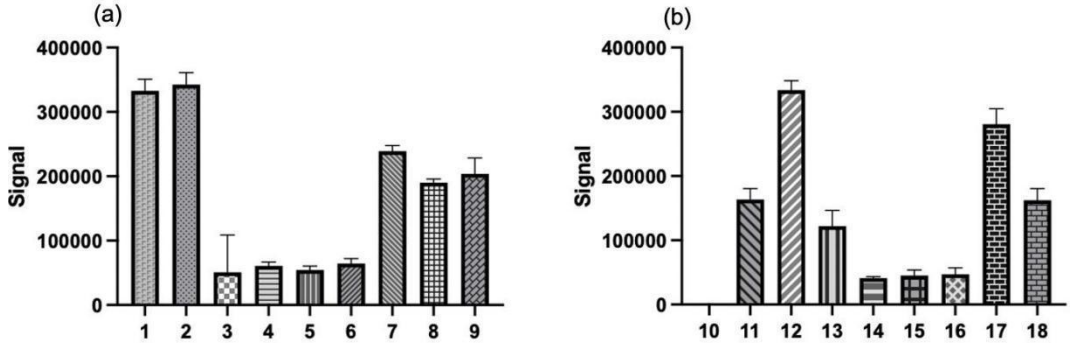


Figure 4.5. Densitometry of Western blot for BPIFA2 expression in whole human saliva. Densitometry analysis of Western blots comparing BPIFA2 expression in whole human saliva from 1-9 (a) and 10-18 (b) healthy volunteers. The X-axis represents the sample number. The Y-axis represents the relative intensity of protein bands. BPIFA2(B) antibody used for detection of samples. Data presented as the mean ± SD (N=3).

Two BPIFA2-specific antibodies BPIFA2(A) and BPIFA2(B) (see chapter 2 for further information regarding antibody generation and validation; Bingle *et al.*, 2009) raised to different epitopes of the protein, one to an epitope in the centre of the protein (BPIF2A) and another to an epitope at the C-terminus (BPIF2B), demonstrated different expression patterns of BPIFA2 in human saliva samples (Bingle *et al.*, 2009). The BPIFA2(A) antibody indicated clear expression of a 34 kDa (potentially glycosylated) protein in the saliva of volunteers 1, 2, 9 and 12 and a less intensive band in volunteers 6 and 11 in comparison to the same samples analysed with the BPIFA2(B) antibody (Figures 4.6 and 4.7). Quantification of the differences in expression levels is shown in the densitometry analysis (Figures 4.5 and 4.8). Overall, the Western blots treated with BPIFA2(A) showed expression of the protein in fewer volunteers with less intense bands compared to those samples analysed with BPIFA2(B). The reasons for these differences in expression have not been fully elucidated. To investigate why BPIFA2 expression in human saliva might be missing or might vary between individuals. For example, ELISA could be used to quantify BPIFA2 levels and compare across individuals to confirm the absence of BPIFA2 protein. Whole Genome Sequencing (WGS) of the BPIFA2 locus would allow detection of mutations, single nucleotide polymorphisms, insertions, or deletions.

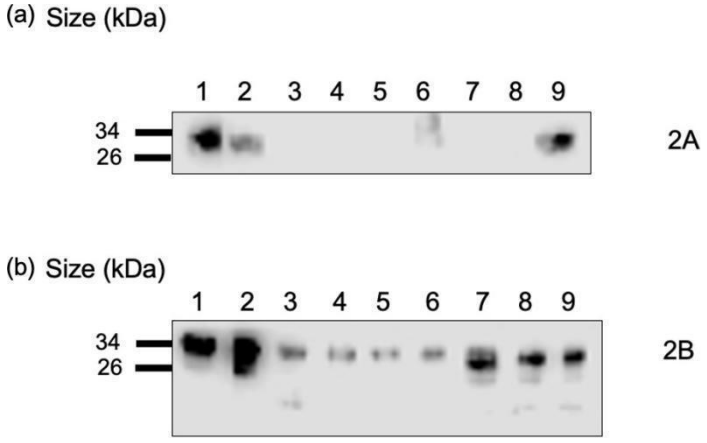


Figure 4.6. Western blot analysis of BPIFA2 expression levels in whole human saliva comparing BPIFA2(A) and BPIFA2(B) primary antibodies. Western blot analysis was used to indicate the expression levels of BPIFA2 in whole saliva derived from healthy volunteers as labelled (Lanes 1-9). BPIFA2(A) (a) and BPIFA2(B) (b) antibody used for detection of specific protein (Representative image from n=3).

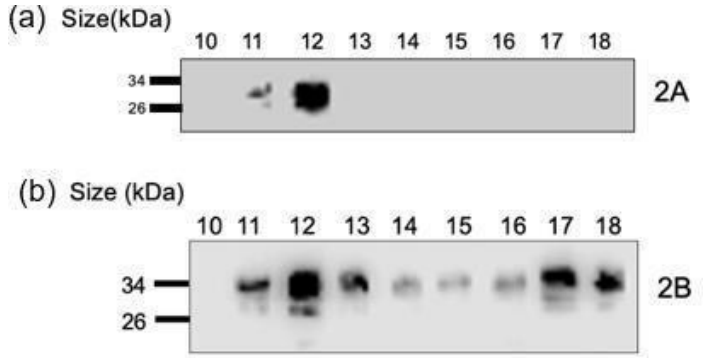


Figure 4.7. Western blot analysis of BPIFA2 expression levels in whole human saliva comparing BPIFA2(A) and BPIFA2(B) primary antibodies. Western blot analysis was used to indicate the expression levels of BPIFA2 in whole saliva derived from healthy volunteers as labelled (Lanes 10-18). BPIFA2(A) (a) and BPIFA2(B) (b) antibody used for detection of specific protein (Representative image from n=3).

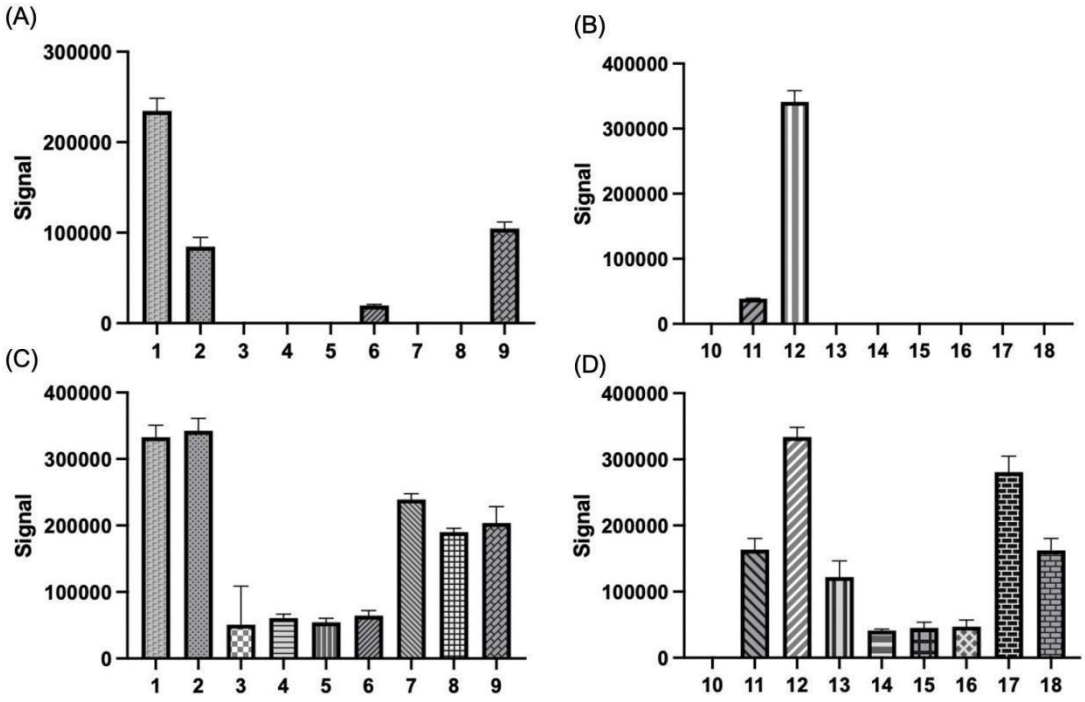


Figure 4.8. Densitometry of Western blot for BPIFA2 expression. Densitometry analysis derived from analysis of Western blots of whole human saliva from 1-9 (A/C) and 10-18 (B/D) healthy volunteers. BPIFA2(A) antibody used for detection of samples in Figure A/B and BPIFA2(B) antibody used for detection of samples in Figure C/D. The X-axis represents the sample number. The Y-axis represents the relative intensity of protein bands detected by densitometry analysis and the signal values indicate protein expression levels. Data presented as the mean ± SD (N=3).

4.4.2 BPIFA2 glycosylation analysis in human saliva

To investigate potential N-glycosylation of BPIFA2, saliva samples from volunteers were incubated with PNGase, which removes N-linked carbohydrate residues, and then subjected to Western blot analysis. Results show a shift in the molecular weight of the specific BPIFA2 bands suggesting carbohydrates have been removed (Figure 4.9). The multiple bands present on the undigested samples were reduced to a single band of lower molecular mass following PNGase treatment in all saliva samples. The molecular weight of the BPIFA2 in the enzyme-treated samples was around 25 kDa which is the molecular weight of the protein as predicted by its amino acid sequence. The Western blot results also demonstrated a difference in the expression of BPIFA2 isoforms between the volunteers. Apart from volunteer 12, the molecular weight of the BPIFA2 isoforms was significantly reduced after enzyme treatment in all saliva samples. There appears to be little impact of the PNGase treatment on saliva from volunteer 12 such that not all of the glycosylation may have been removed, which suggests that more enzyme is needed to fully deglycosylate this sample.

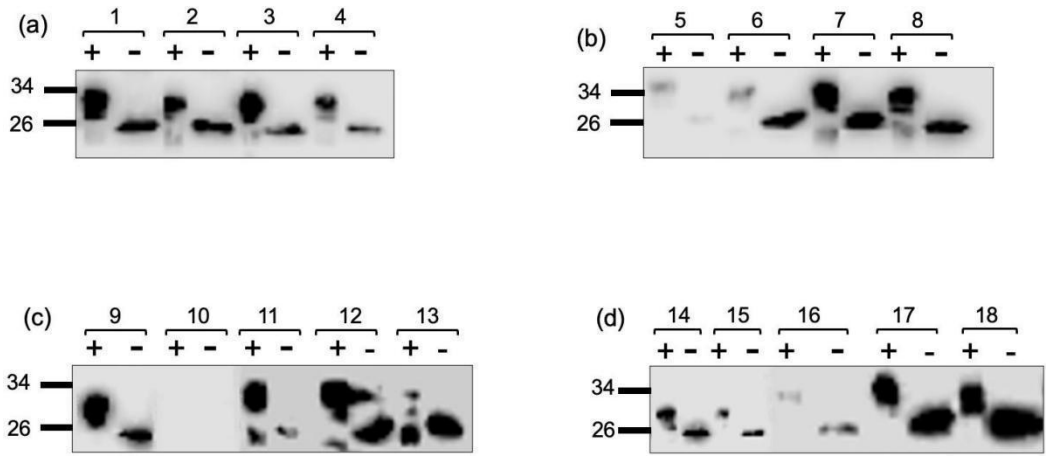


Figure 4.9. Western blot analysis of BPIFA2 expression levels in whole human saliva following PNGase F enzyme treatment. Saliva samples from donor 1-4 (a), 5-8 (b), 9-13 (c) and 14-18 (d) were digested with PNGase F, analysed by Western blotting using the BPIFA2(B) antibody and compared to untreated saliva samples. "+" represents untreated saliva and "-" represents saliva digested with PNGase. (Representative image from n=3)

Densitometry analysis was used to quantify the intensity of the bands following treatment of the samples with PNGase F enzyme (Figure 4.10).

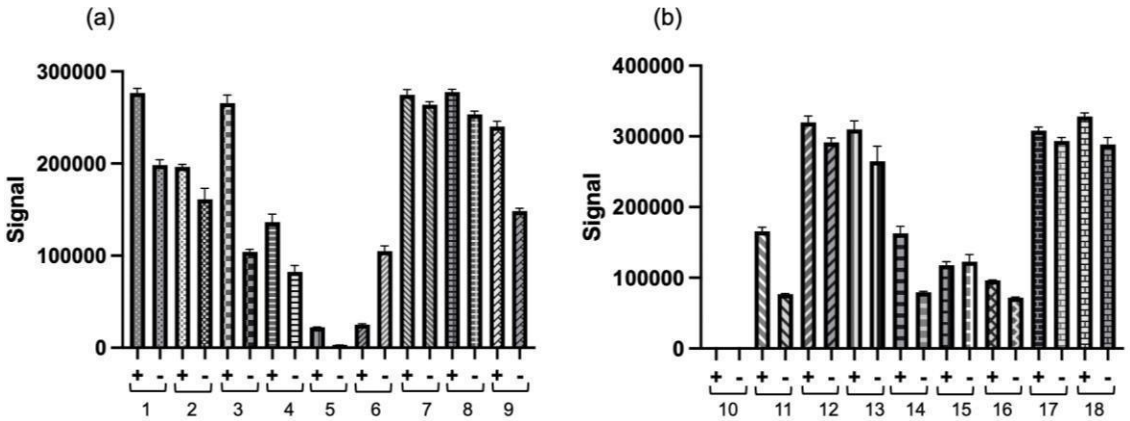


Figure 4.10. Densitometry analysis of Western blot analysis of BPIFA2 expression levels in whole human saliva following treatment with PNGase F enzyme. Saliva samples from donors 1-9 (a) and 10-18 (b) digested with PNGase F were analysed by Western blotting using the BPIFA2(B) antibody and compared to untreated saliva controls. "+" represents untreated saliva samples and "-" represents saliva samples digested with PNGase. The X-axis represents the sample number. The Y-axis (signal) represents the relative intensity of protein (expression level) detected by densitometry analysis. Data presented as the mean ± SD (N=3).

4.5 Construct development.

Constructs for the production of recombinant human BPIFA2 proteins were previously established and verified by the Bingle group and stored at -80 °C as glycerol stocks. These constructs were made in the mammalian expression vector pVR1255 and each one had an in-frame FLAG-Tag at the C-terminus. BPIFA2 is N-glycosylated at asparagine-124 and asparagine-132 and so to determine if glycosylation is related to function these sites were mutated to alanine to produce proteins without the individual glycosylation sites for comparison with wide-type sample. A new construct for the 124/132 double glycosylation mutant in which the 2 N residues were replaced by A was required as although a construct of this type had previously been developed in the laboratory it did not produce recombinant protein in my initial transfection experiments and when sequenced it was discovered that it was no longer the correct sequence for this construct.

Similar FLAG-tagged expression constructs of five animal species, mouse, armadillo, macaque, squirrel monkey and dog, were generated as part of this thesis in order to study the function of BPIFA2, *in vitro*, spanning the mammalian phylogeny.

4.5.1 TOPO TA subcloning

Empty VR1255 vector and the BPIFA2 insert were used for cloning and the expected fragment sizes were detected using a 1% agarose gel (Figure 4.11). BamH1 and Not1 were selected as target sites as the insert was under these cutting sites and the enzymes will not cut elsewhere in the plasmid.

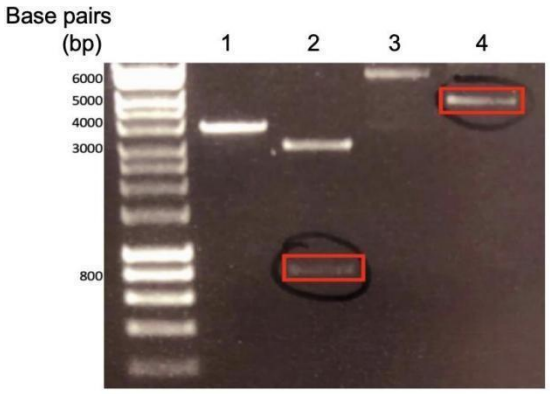


Figure 4.11. 1% agarose gel was used to detect the fragment size of human BPIFA2 (~750 bp) and human empty VR1255 (~5-6 kb). Lane 1- uncut Bluescript plasmid; lane 2- double digestion of Bluescript plasmid with red box representing target BPIFA2 insert; lane 3- uncut VR1255; lane 4- double digestion of VR1255. BamH1 and Not1 were used for double digest.

The 124/132 mutant BPIFA2 insert (~750 bp) was ligated into the double-cut VR1255 vector (~5-6 kb) using T4 DNA ligase (Promega) and Ligase 10X Buffer (Promega). The target vector was then transformed into *E. coli* via heat-shock and after overnight culture colonies were picked and incubated overnight in LB-broth with kanamycin (50 μg/mg). A ZymoPURE II Plasmid Midiprep kit was used to isolate plasmid DNA from the bacterial culture as described in section 2.3.1. Prior to sequencing the target plasmid was again digested with BamH1 and Not1 restriction enzymes and a 1% agarose gel confirmed the fragments were the correct size: human BPIFA2 ~750 bp and human empty VR1255 ~5-6 kb as shown in Figure 4.12.

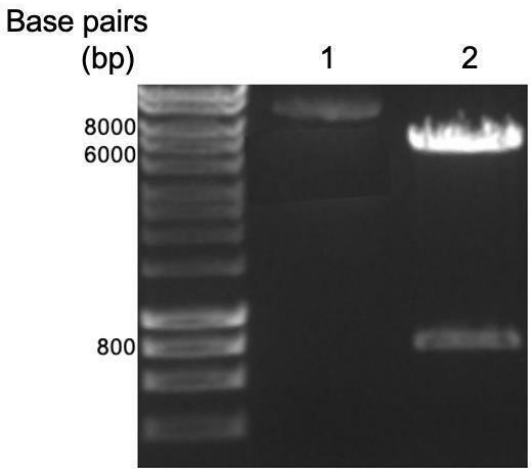


Figure 4.12. 1% agarose gel was used to detect the fragment size of human BPIFA2 (~750 bp) and empty VR1255 (~5-6 kb) after Plasmid Mini Kit (ZymoPURE) purification. Lane 1- uncut double mutant BPIFA2 plasmid; lane 2- double digest of double mutant BPIFA2 plasmid. Upper band is empty VR1255 (~5-6 kb) and the lower band is BPIFA2 (~750 bp). BamH1 and Not1 restriction enzymes were used for the double digest.

Expected double mutation plasmids were sequenced (Figure 4.13) and the Basic Local Alignment Search Tool (BLAST) used to find, and align, regions of local similarity between sequences. Expert, identities and gaps are the three important parameters to determine the results from BLAST analysis. Expert indicates the possibility of random matching, the greater the E value, the greater the possibility of random matching. When the E value is close to zero or zero, it is essentially a perfect match. Identity or similarity indicates the number of matched bases as a percentage of the total sequence length. Gaps representing insertion or deletion; "-" was used to indicate these in the sequence. From this information we can confirm that we have obtained the target plasmid for further experiments.

TTGTCATCGTCGTCCTTGTAGTCGATGAGGGTTTGCAGCTGGGTTTTGTGCTGAGGATTATCGACGACCTGCTGAA
TGACATTCACATCCAGGGAGTGGATGAAGATGCGGATCAGTGGACATATCTCCTTCTGCAGCAGGGAGGATACAG
TGCTTTTCAGCGTGTTGATCACGCTATTCACGAACTTGTTGATGATTTGGCTGTGTTTGTCCAGCAAGGAAAGTGA
GATGCTGGTTGGGTCACTGGCGCATTCTCCCAGGACGGCAACAGGCTGGTGTGTCTGGGGATCAGTTTCAATTGT
GACTGCGGTCAGGAGGTCCAAGGAGGCTTTCAGGTTGATAATCTGGCCAATGATGGGCCCGGCCACAGTGACCT
GCGCGGTGACAGGGAAGCTCAGCTGAAGGCCTTTGCCATCATCGATCGGTTCAGCTTTGACATCCAGGATGAGGG
AGTTGCTGATTTTCAACCCAAAAATGTCCGTGTTAGTTGGAAGCAGCTTAGAAATGACATTGTTCAGCAATTTCTCA
GCTTCCTGGGCCTTCTGCTTGGCCAGTTGCCAAGCACTGGATTTCTGAAGCACTCCTA

Score	Expect	Identities	Gaps	Strand
1018 bits(551)	0.0	559/563(99%)	0/563(0%)	Plus/Minus

```
Query  24   GATGAGGGTTTGCAGCTGGGTTTTGTGCTGAGGATTATCGACGACCTGCTGAATGACATT  83
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  891  GATGAGGGTTTGCAGCTGGGTTTTGTGCTGAGGATTATCGACGACCTGCTGAATGACATT  832

Query  84   CACATCCAGGGAGTGGATGAAGATGCGGATCAGTGGACATATCTCCTTCTGCAGCAGGGA  143
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  831  CACATCCAGGGAGTGGATGAAGATGCGGATCAGTGGACATATCTCCTTCTGCAGCAGGGA  772

Query  144  GGATACAGTGCTTTTCAGCGTGTTGATCACGCTATTCACGAACTTGTTGATGATTTGGCT  203
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  771  GGATACAGTGCTTTTCAGCGTGTTGATCACGCTATTCACGAACTTGTTGATGATTTGGCT  712

Query  204  GTGTTTGTCCAGCAAGGAAAGTGAGATGCTGGTTGGGTCACTGGCGCATTCTCCCAGGAC  263
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  711  GTGTTTGTCCAGCAAGGAAAGTGAGATGCTGGTTGGGTCACTGGCGCATTCTCCCAGGAC  652

Query  264  GGCAACAGGCTGGTGTGTCTGGGGATCAGTTTCAATTGTGACTGCGGTCAGGAGGTCCAA  323
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  651  GGCAACAGGCTGGTGTGTCTGGGGATCAGTTTCAATTGTGACTGCGGTCAGGAGGTCCAA  592

Query  324  GGAGGCTTTCAGGTTGATAATCTGGCCAATGATGGGCCCGGCCACAGTGACCTGCGCGGT  383
            |||||||||||||||||||||||||||||||||||||||||||||||||| | ||||||
Sbjct  591  GGAGGCTTTCAGGTTGATAATCTGGCCAATGATGGGCCCGGCCACAGTGACATTCGCGGT  532

Query  384  GACAGGGAAGCTCAGCTGAAGGCCTTTGCCATCATCGATCGGTTCAGCTTTGACATCCAG  443
            |||||||||||||| | ||||||||||||||||||||||||||||||||||||||||||
Sbjct  531  GACAGGGAAGCTCAGGTTAAGGCCTTTGCCATCATCGATCGGTTCAGCTTTGACATCCAG  472

Query  444  GATGAGGGAGTTGCTGATTTTCAACCCAAAAATGTCCGTGTTAGTTGGAAGCAGCTTAGA  503
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  471  GATGAGGGAGTTGCTGATTTTCAACCCAAAAATGTCCGTGTTAGTTGGAAGCAGCTTAGA  412

Query  504  AATGACATTGTTCAGCAATTTCTCAGCTTCCTGGGCCTTCTGCTTGGCCAGTTGCCAAGC  563
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  411  AATGACATTGTTCAGCAATTTCTCAGCTTCCTGGGCCTTCTGCTTGGCCAGTTGCCAAGC  352

Query  564  ACTGGATTTCTGAAGCACTCCTA  586
            |||||||||||||||||||||||
Sbjct  351  ACTGGATTTCTGAAGCACTCCTA  329
```

Figure 4.13. Sequencing of double mutant plasmids. Sequencing of double mutant plasmids by The Genomics Core Facility, The Medical School, The University of Sheffield, UK shows successful cloning.

In order to produce the animal constructs the empty VR1255 vector was again obtained from the wild-type BPIFA2 plasmid double digested with BamHI and NotI restriction enzymes. The armadillo and dog BPIFA2 inserts were double digested with BgIII and NotI restriction enzymes and the macaque, mouse and squirrel monkey BPIFA2 inserts were double digested with BamHI and NotI restriction enzymes; all animal constructs were from Biomatik (Figure

4.14). BamH1, BgIII and Not1 were selected for the target sites as the insert was under these cutting sites and these enzymes would not cut elsewhere in the plasmid.

In Figure 4.13, the size of armadillo (lane 2), dog (lane 3), macaque (lane 5) and squirrel monkey (lane 6) BPIFA2 inserts were the expected size of 814 bp and the mouse BPIFA2 insert (lane 4) was the expected size of 772 bp. The animal inserts were ligated into the double-cut VR1255 vector (~5-6 kb; lane 1) using T4 DNA ligase and Ligase 10X Buffer. The target vectors were then transformed into *E.coli* via heat-shock and after overnight culture colonies were picked and incubated overnight in LB-broth with kanamycin (50 μg/mg). A ZymoPURE II Plasmid Midiprep kit was used to isolate plasmid DNA from the overnight bacterial culture.

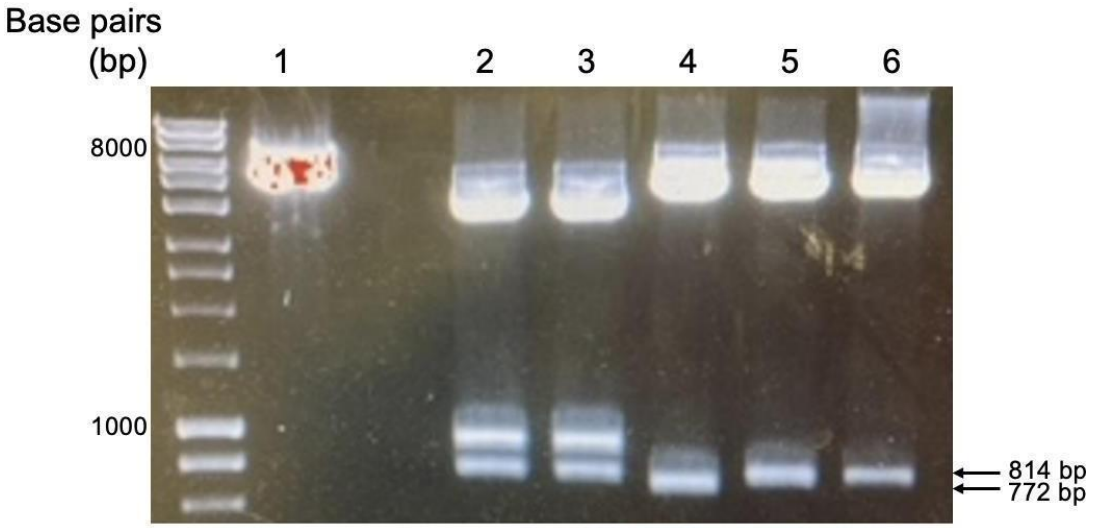


Figure 4.14. 1% agarose gel was used to detect the fragment size of animal BPIFA2 inserts (814 bp and 772 bp) and empty human VR1255 (~5-6 kb) following digestion. Lane 1- VR1255; lane 2- armadillo, lane 3- dog, lane 4- mouse, lane 5- macaque and lane 6- squirrel monkey. BamH1 and Not1 were used for the double digest.

Successful incorporation of animal BPIFA2 constructs was confirmed by the presence of bands of the expected size when analysed by 1% gel electrophoresis of digested fragments

(Figure 4.15). A HindIII digestion was performed on the squirrel monkey plasmid, a Pst1 digestion on the macaque plasmid, an Xba1 digestion on the mouse plasmid, and both the dog and armadillo plasmids were digested with BamH1. The agarose gels show that multiple plasmid preparations contained successfully incorporated inserts of the expected size (Figure 4.16). The expected size of the squirrel monkey and macaque plasmid vectors was 5512 bp, the expected size of the dog and armadillo plasmid vectors was 5506 bp and the expected size of the mouse plasmid vector was 5470 bp. Specific restriction enzymes were used to make the five animal constructs and so DNA sequencing was not required for further confirmation. The plasmid maps of the animal BPIFA2 constructs are described in Figure 4.16.

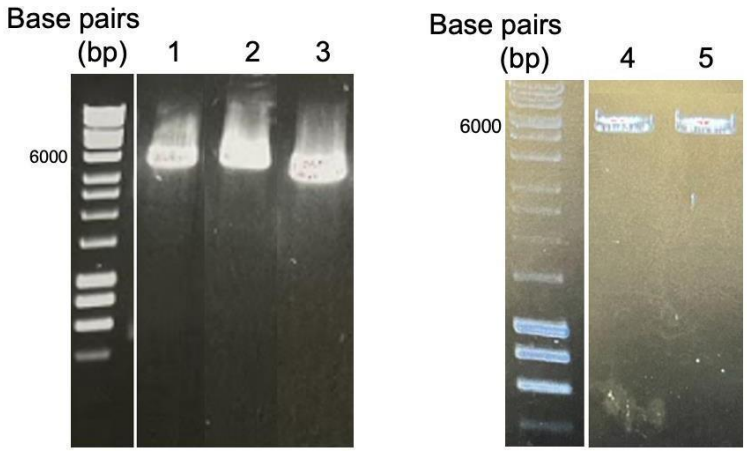


Figure 4.15. 1% agarose gel was used to detect the expected fragment size of animal BPIFA2 constructs for validation of gene insertion. The resultant plasmid fragments were resolved on a 1% agarose gel to confirm successful incorporation of the BPIFA2 insert. Lane 1- Hind III digest of squirrel monkey plasmid, expected size 5512 bp, lane 2- digest of macaque plasmid, expected size 5512 bp, lane 3- Xba1 digest of mouse plasmid, expected size 5470 bp, lane 4- BamH1 digest of armadillo plasmid, expected size 5506 bp and lane 5- BamH1 digest of dog plasmid, expected size 5506 bp.

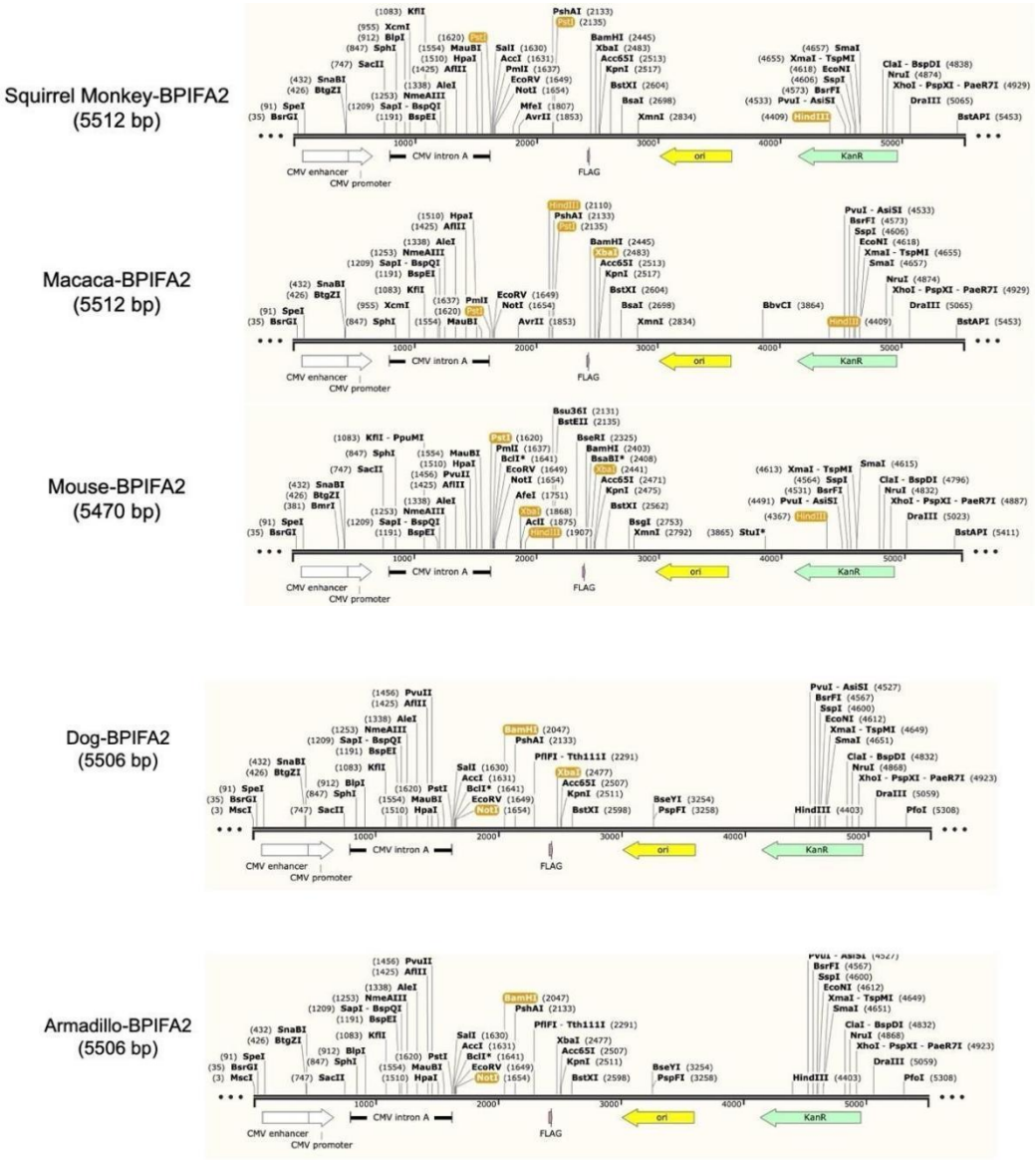


Figure 4.16. Plasmid maps of animal BPIFA2 constructs. The restriction sites and kanamycin resistance gene (Kan) are shown. BamHI, NotI and BgIII were the digest sites used to produce VR1255 vector and the luciferase gene (LUX) was removed and replaced with compatible sticky terminals for the BPIFA2 insertion. The size of squirrel monkey and macaque plasmids was 5512 bp, the dog and armadillo plasmids were 5506 bp and the mouse plasmid was 5470 bp.

4.6 Protein purification

Transient transfections were carried out in order to produce recombinant protein from four different human BPIFA2 constructs and five animal BPIFA2 constructs; this would allow collection of WT human protein and also human proteins with one or both glycosylation sites

removed: 124, 132 and 124/132 and BPIFA2 proteins from animal species. Calcium phosphate was used to transfect the novel DNA into HEK293 cells in T75 flasks and as a mean of assessing transfection efficiency we simultaneously transfected the cells with the vector enhanced green fluorescent protein (pEGFP-N1; 3.75 μg). Recombinant protein was purified from culture media, beginning 48 hours post-transfection by transferring the cultures into phenol red-free growth medium. Conditioned media containing BPIFA2 proteins was collected 24 hours after transfer to phenol red-free media and every 24 hours for the subsequent 3 days. A representative image of human BPIFA2/pEGFP-N1 transfection efficiency in HEK293 cells is shown in Figure 4.17 and successful transfection of animal BPIFA2/ pEGFP-N1 is shown in Figure 4.18.

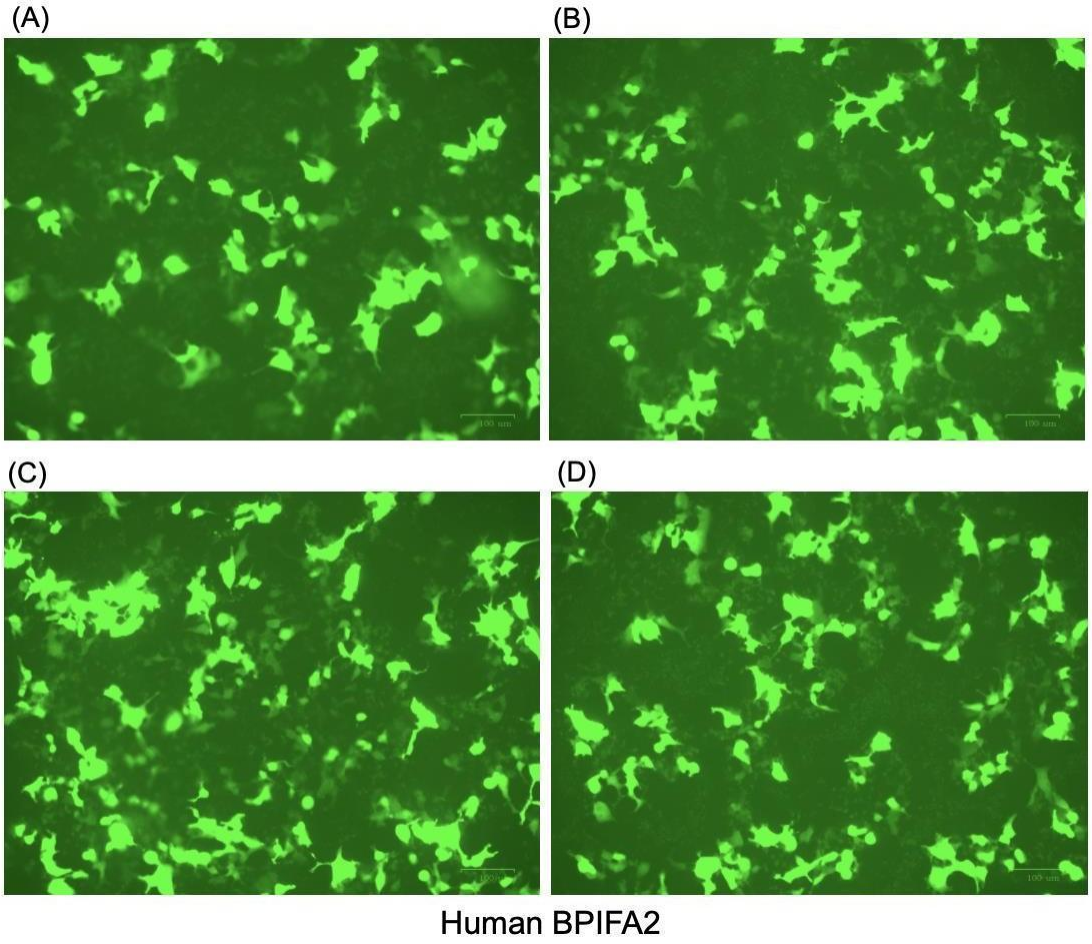


Figure 4.17. HEK293 cells transfected with human BPIFA2 and eGFP-N1 using calcium phosphate. Representative images of transfected HEK293 cells 48 hours post-transfection. (A) WT BPIFA2, (B) 124 single mutant, (C) 132 single mutant and (D) 124/132 double mutant. Transfection images were detected by the ZOE Fluorescent Cell Imager. Original magnification of the images was X10. (Representative image from n=5)

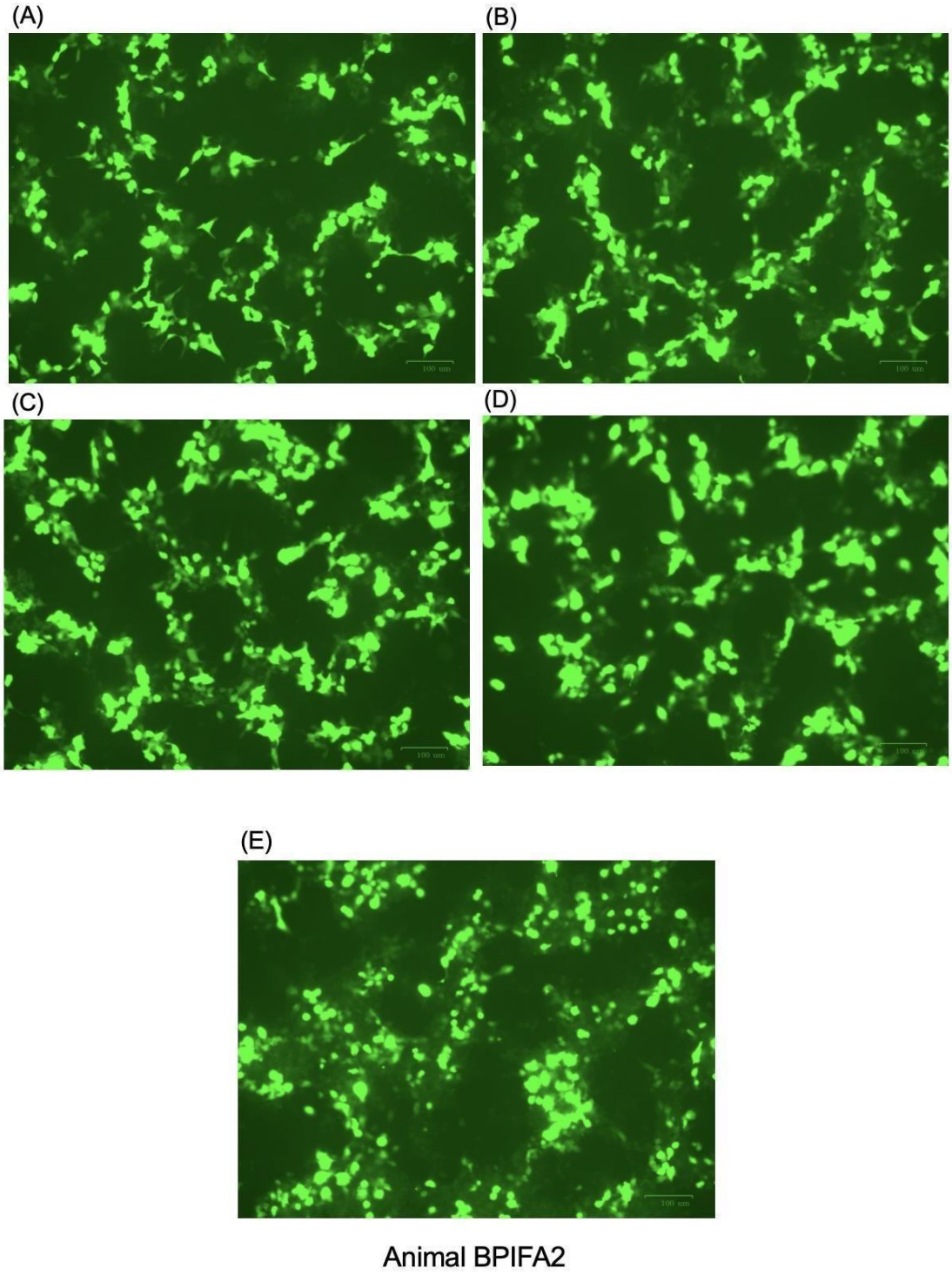


Figure 4.18. HEK293 cells transfected with animal BPIFA2 and eGFP-N1 using calcium phosphate. Representative images of transfected HEK293 cells 48 hours post-transfection. (A) dog, (B) armadillo, (C) mouse, (D) squirrel monkey and (e) macaque BPIFA2. Transfection images were detected by the ZOE Fluorescent Cell Imager. Original magnification of the images was X10. (Representative image from n=5)

Following transfection, a centrifugal concentrator column with a molecular weight cut off of 10 kDa was used to concentrate the conditioned media collected from the transfected cells. The initial volume of 20 ml was reduced to a final volume of 3 ml after concentration. The recombinant protein (which was FLAG-tagged) was purified from concentrated conditioned media using FLAG affinity resin. 100µl of resin was added to 6 ml conditioned media for 2 hours at 4^{o}C on a roller followed by the addition of 150 µl FLAG peptide (150 ng/µl) to elute the FLAG-tagged recombinant protein. Coomassie staining shows that the samples contained predominantly purified BPIFA2 recombinant proteins with very limited amounts of contaminating proteins in the human samples and no contaminating proteins in the animal samples (Figure 4.19). Protein dimers, complexes formed by two protein molecules, were seen in the animal BPIFA2 samples. The interaction can be temporary or long-lasting, based on a number of factors including the affinity of the surfaces involved. Similar findings were also observed on Western blot analysis as shown in Figure 4.20.

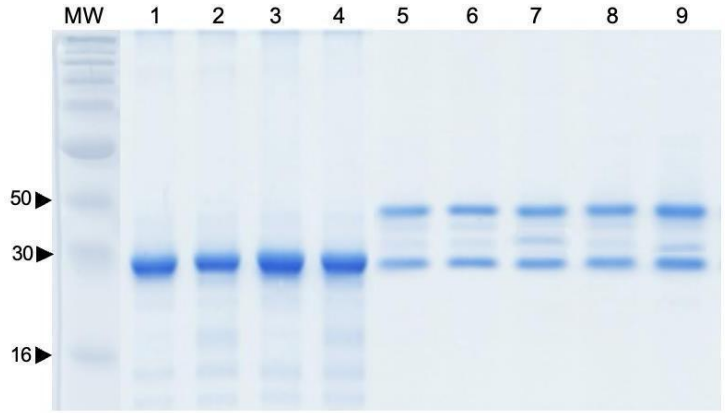


Figure 4.19. Coomassie blue staining of purified recombinant proteins from HEK 293 cells transfected with BPIFA2 constructs. Representative images show human and animal FLAG-tagged recombinant protein was purified from conditioned media using Anti-FLAG M2 affinity gel and FLAG peptide. Lane numbers represent the following: 1 - WT BPIFA2, 2 -124 single mutant, 3 - 132 single mutant and 4 - 124/132 double mutant. 5 - armadillo 6 - mouse, 7 - macaque, 8 - squirrel monkey and 9 - dog BPIFA2. (Representative image from n=3).

Western blotting was used to analyse the success of recombinant protein purification and concentration (Figure 4.20). Human BPIFA2 is present as multiple bands representing the different isoforms of the protein (124 in lane 2 and 132 in lane 3, Figure 4.20A). Western blotting was also used to analyse the success of animal recombinant protein concentration and purification (Figure 4.20B). Interesting novel bands were seen at around 52 kDa, but to date we do not know the true identity of these bands; it is possible they are protein dimers. Most of the glycosylation sites in the animals studied in this project are different to those in humans apart from position124 and 132 in the human, which are the same as two of the glycosylation sites in the macaque and position 132 human, which is the same as the only glycosylation site in dog (Figure 4.21). Further experiments need to be carried out to investigate the role for the animal glycosylation sites.

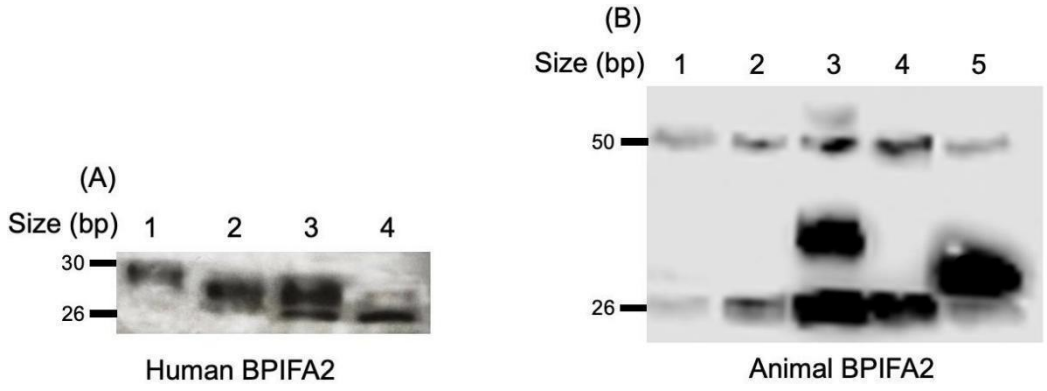


Figure 4.20. Western blot analysis of purified recombinant proteins from HEK 293 cells transfected with BPIFA2 constructs. Representative images show human (A) and animal (B) FLAG-tagged recombinant protein purified from conditioned media using Anti-FLAG M2 affinity gel and FLAG peptide. A specific BPIFA2 antibody was used for Western blot analysis of human proteins and an Anti-flag M2 antibody for animal species. Lane numbers represent the following: (A) 1 - WT BPIFA2, 2 -124 single mutant, 3 - 132 single mutant and 4 - 124/132 double mutant. (B) 1- armadillo 2 - mouse, 3 - macaque, 4 - squirrel monkey and 5 - dog BPIFA2 (Representative image from n=4).

Figure 4.21. Multiple protein sequences of BPIFA2 in mouse, dog, armadillo, squirrel monkey human and macaque. A blue star (*) indicates completely conserved residues, a colon (:) indicates semi conserved residues and period (.) indicates weakly similar residues. Cysteine residues involved in the formation of a disulphide bond are highlighted in green. N-glycosylation sites are highlighted in yellow and the same sites as the glycosylation site in humans are highlighted with a red box. The sequences of BPIFA2 were aligned using the online tool, BLAST: https://blast.ncbi.nlm.nih.gov/Blast.cgi.

After purification, BCA and ELISA assays were used to determine total recombinant protein concentration in each sample as depicted in Figure 4.22. The ELISA assay (Figure 4.22B) showed a much lower concentration compared to the BCA assay (Figure 4.22A) when identical purified recombinant protein samples were analysed simultaneously. A potential reason for the discrepancies between the two assays is the folding or conformation of the protein. The recombinant protein used in this study was attached to a FLAG-tag at the end of the C-terminal domain which might cause the proteins to be folded in a way that embeds the FLAG-tag inside the protein structures and less readily recognized by the Anti-FLAG

antibody, leading to difficulties in concentration determination.

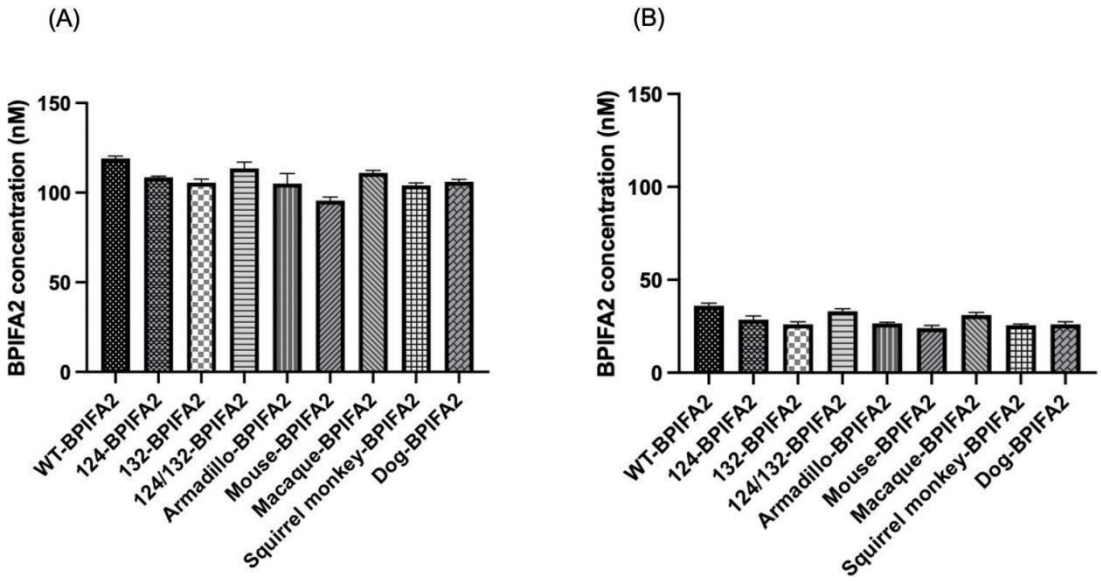


Figure 4.22. Comparison of BCA assay and ELISA analysis. BCA assay (A) and ELISA analysis (B) were used to quantify the concentration of recombinant BPIFA2 proteins after purification. The Y-axis represents the concentration of recombinant proteins. Data presented as the mean ± SD (N=3; n=3).

Additional dot blot analysis was, therefore, used to quantify the concentration of purified recombinant proteins in humans and animals (Figure 4.23). The densitometry results for dot blotting show that the purification efficiency of producing recombinant proteins with four human constructs and five animal constructs is at similar levels, suggesting that the protein concentration is comparable between the constructs (Figure 4.24).

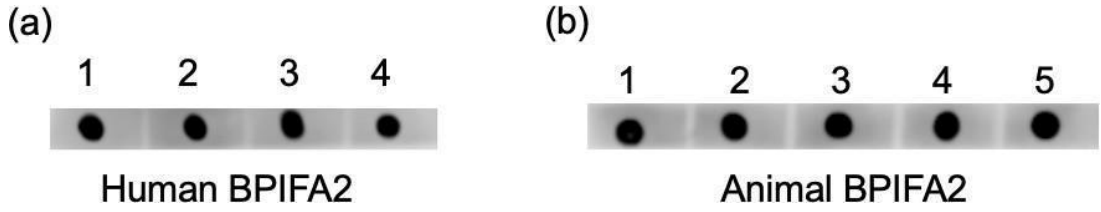


Figure 4.23. Dot blot analysis of purified recombinant proteins from HEK 293 cells transfected with BPIFA2 constructs. Representative images show human (a) and animal (b) FLAG-tagged recombinant protein purified from conditioned media using Anti-FLAG M2 affinity gel and FLAG peptide. A specific BPIFA2 antibody was used for analysis of human proteins and an Anti-flag M2 antibody for animal proteins. Lane numbers represent the following: (a)1 - WT BPIFA2, 2 -124 single mutant, 3 - 132 single mutant and 4 - 124/132 double mutant. (b)1- armadillo, 2 – mouse, 3 – macaque, 4 – squirrel monkey and 5 – dog BPIFA2. (Representative image from n=3).

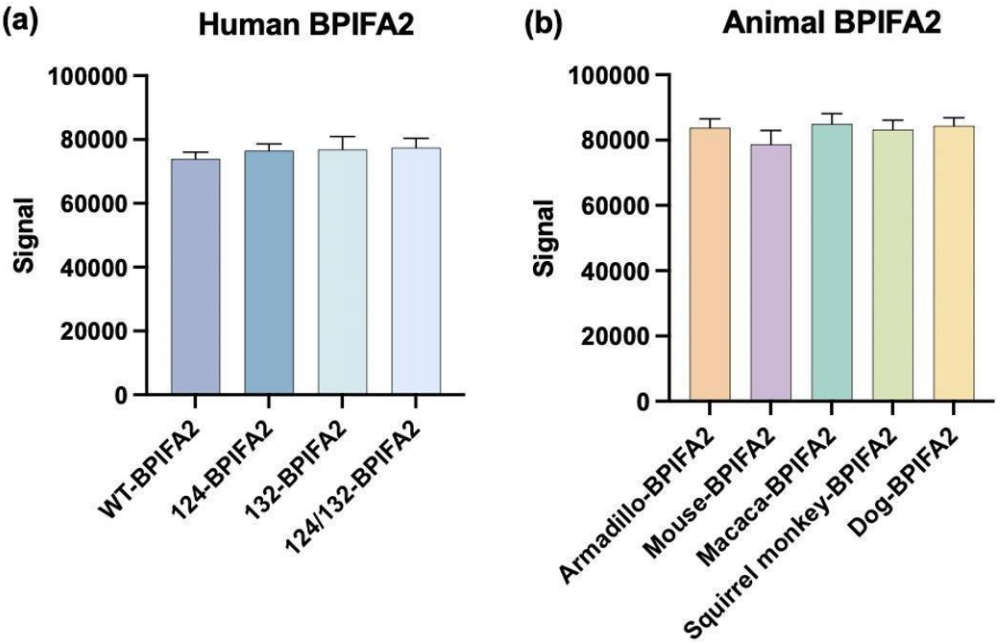


Figure 4.24. Densitometry of dot blot analysis of purified recombinant proteins from HEK 293 cells transfected with BPIFA2 constructs. Lane numbers represent the following: (a) WT-BPIFA2, 124 single mutant, 132 single mutant and 124/132 double mutant. (b) armadillo, mouse, macaque, squirrel monkey and dog BPIFA2. The Y-axis (signal) represents the relative intensity of protein expression. Data presented as the mean ± SD (N=3).

4.7 Discussion

This chapter identifies the variability in BPIFA2 levels in saliva across individuals and highlights the importance of methods like Western blotting and dot blotting for quantifying and identifying BPIFA2 expression. Identifying and characterizing glycosylation patterns through enzymatic treatments (e.g., PNGase F) can elucidate their impact on BPIFA2 function in saliva. Producing recombinant BPIFA2 in mammalian expression cell lines allows for analysis of the protein *in vitro*. Recombinant BPIFA2 was ideal for preserving post-translational modifications like glycosylation, ensuring that the recombinant proteins are as physiologically relevant as possible.

In this study, the use of unstimulated saliva is based on BPIFA2 is produced by the serous acini in the main glands and subsequently released into saliva through the ducts (Bingle *et al.*, 2009). The composition of saliva can fluctuate based on circadian rhythms (Dawes, 1972), which are influenced by the time of day. Saliva can exhibit diurnal fluctuations in BPIFA2 proteins and other indicators such as hormones or other salivary proteins (Kobayashi *et al.*, 2017). Additional investigation is required to evaluate the expression levels of BPIFA2 in stimulated and unstimulated saliva in humans. Data published from PSP Knock-out and Wild-type mouse models, developed by Professor Sven-Ulrik Gorr, University of Minnesota, USA, has been shared with us. This data suggested that the mice do not develop a significant phenotype following knock out of the BPIFA2 gene. However, it was noted that whilst there was no difference in the amount of saliva produced upon pilocarpine stimulation between WT and KO mice, there was a significant difference (60% more) in the amount of amylase present in KO mice (Nandula *et al.*, 2020) . Exploring the relationship between BPIFA2 and other

secretions in the oral cavity, including salivary amylase, will allow us to investigate the continued importance of this protein in human saliva.

Unstimulated whole saliva from eighteen volunteers was collected at the same time of day (9am-9.30am) and samples were subjected to dot blotting and Western blotting using antibodies BPIFA2(A) and BPIFA2(B) (see chapter 2; section 2.2.2 for BPIFA2 antibody generation and validation). The resultant dot blots with BPIFA2(B) antibody show that significant variations in expression levels were observed among different volunteers, suggesting that BPIFA2 expression is related to individual differences but these are not influenced by gender or age. It is interesting to note that there is no BPIFA2 expression in saliva of volunteer 10, a 55-year-old male. A study of genetics has also shown that eight SNPs were found in all exons, including exon 2 which is a start codon leading to protein translation, intron borders and the 2 kb promoter region of the BPIF gene showing a notable link to disease susceptibility of nasopharyngeal carcinoma (NPC) in Chinese population (He et al., 2005). BPIFA1 and BPIFA2 belong to the same protein family that implies potential similarities in both structure and function. If single nucleotide polymorphisms (SNPs) in BPIFA1 are recognised to influence its function or structure, it is possible that similar SNPs in BPIFA2 could have comparable consequences. Our results in chapter 3 section 3.3.5 showed that SNPs expression across wider populations of people has found that a population from Western Africa may not produce any BPIFA2 protein. Although the association between BPIFA2 protein loss in that population and possible diseases remains unknown. We have not been able to study the Western African or Chinese populations, but further experiments could be carried out on our volunteer 10, who does not appear to express BPIFA2 in their saliva.

And further parafilm chew-stimulated saliva could be collected from this volunteer compared to unstimulated sample in BPIFA2 expression as both BPIFA2 and stimulated saliva are mainly expressed by parotid glands (Aps and Martens, 2005; Bingle *et al.*, 2009). Although the study revealed that the levels of biomarkers such as ammonium, enzymes and proteins in unstimulated saliva were similar to those in chew-stimulated saliva (Al Habobe *et al.*, 2024).

All of the saliva samples used in this study were collected from healthy volunteers at the same time in the morning, without stimulation and without eating or drinking for 2 hours prior to collection. There is limited research on BPIFA2 levels in saliva from patients with diseases, however, some cases of differential production have been observed. Previous results showed that alterations in BPIFA2 levels in patients undergoing autologous hematopoietic stem cell transplantation (auto-HCT), allogeneic hematopoietic stem cell transplantation (allo-HCT), and those with oral chronic graft-versus-host disease (cGVHD) may be associated with modifications in saliva composition, which suggesting that BPIFA2 could serve as a promising biomarker for the oral side effects of HCT. Reduced levels of BPIFA2 seen during both hyposalivation and cGVHD which is a side effect of allo-HCT (Innocentini *et al.*, 2022).

BPIFA2 is mostly found in the serous cells of the major salivary glands and in the sero-mucous tubules of the smaller glands in the oral mucosa, posterior tongue, and tonsil (Laine and Smoker, 1996). Saliva, both serous and mucous cell-derived contains many host defence proteins, and it is well-established that saliva plays a number of important roles in maintaining oral health. It has been hypothesised that due to the expression sites of many BPIF family members, and the fact that they are secreted proteins, these proteins will also place a host- defence role. Thus, it has been proposed that BPIFA2, secreted into saliva, will

play a role in the innate immunological defence of the oral cavity. My studies were designed to test this hypothesis and the data we have presented, showing variability of expression within a population leads us to suggest that additional research should be conducted to compare levels of BPIFA2 between healthy volunteers and patients with salivary gland disease, particularly infectious diseases such as sialadenitis. This could help us understand the potential for BPIFA2 as a biomarker of salivary gland disease.

Our Western blot analysis with the BPIFA2(B) antibody clearly shows that multiple protein isoforms of BPIFA2 exist in saliva, which are likely a result of distinct glycosylation patterns. This result is corroborated by a proteomic analysis of glycosylated proteins in whole saliva in which BPIFA2 was identified as containing N-linked glycoproteins (Ramachandran *et al.*, 2006) The exact function of this differential glycosylation of BPIFA2, however, is still uncertain. It is well-known that numerous proteins in saliva undergo complex modifications after they are synthesised, such as glycosylation. It is presumed that these modifications significantly impact their function. For example, mucin is highly glycosylated with extremely adhesive characteristics and expressed in the mucous glands, and is commonly found in association with other salivary proteins providing antibacterial activity (Nieuw Amerongen and Veerman, 2002).

BPIFA2 proteins have been described as the most diverse group of paralogous proteins within the family with theBPIFA2 genes havingexperienced an important degree of divergence during mammalian evolution. For example, PSP, as discussed in chapter 1, has significant homology to BPIFA2 expressed in mice with the human/mouse proteins being 32% identical, even though sequence similarity of PSP with other BPIF family members is very low (Weston

et al., 1999). SMGB was the second PSP-related protein identified in rats and whilst not an orthologue of BPIFA2 it is a rodent-specific member of the BPIF family (Ball, Mirels and Hand, 2003; Bingle, Seal and Craven, 2011). BSP are encoded by four genes BSP30A, BSP30B, BSP30C and BSP30D and are found in bovine saliva. Both PSP and SMGB in rodents and BSP30A and B in bovine have been shown to be N-glycosylated. Studies have demonstrated that Rodent PSP exhibits anti-*candidal* properties (Robinson *et al.*, 1997) and has the capacity to attach to bacterial membranes (Khovidhunkit *et al.*, 2005). BPIFA2 also exhibits a strong attraction to hydrophobic silicone and may improve the adhesion of *Candida albicans* to silicone surfaces (Holmes *et al.*, 2014). BSP 30A and B have the ability to inhibit growth of *P. aeruginosa* and *Streptococcus pneumoniae* (Wheeler *et al.*, 2002). Current information on BSP function related to glycosylation presents numerous unresolved challenges. BSP30B has a higher level of glycosylation than BSP30A and the glycosylation status of BSP30B could partially explain the growth inhibitory activity of BSP30B towards *S. pneumoniae* (Haigh *et al.*, 2008). The data indicates that the antibacterial activity corresponds to the high level of N- glycosylation. Electrophoresis suggests that BSP 30B has an apparent molecular weight 5kDa higher than BSP30A, which could be a result of post-translational modifications, such as glycosylation (Wheeler *et al.*, 2011). There is little information regarding BSP30C and BSP30D relative to BSP30A and BSP30B. N-glycosylation is significant in immunity and glycoproteins on the surface of cells can be recognized by the immune system as self or non-self, influencing immune responses (Pandey *et al.*, 2022). Glycosylation is not limited to immune proteins exclusively. It provides the purpose of stabilising proteins. In the case of immune proteins, glycosylation facilitates interaction between host cells and proteins through lectins (Frenkel and Ribbeck, 2015). Discovering the

specific interactions between the BPIFA2 protein and macrophage cells, particularly immune cells, could provide more evidence of the BPIFA2 protein's immune-related role. An interesting research prospect involves investigating the functional assays of both the wild type and mutant variants of human BPIFA2 to determine the impact of N-glycosylation sites in chapter 5.

Although the function of BPIFA2 is not fully known, it's diverse expression could be related to biological evolution. It has been shown that saliva collected from great apes contains considerably more BPIFA2 than that from human saliva (Thamadilok *et al.*, 2020). This interesting finding suggests that humans no longer require large amounts of the protein in their saliva and this species difference could perhaps suggest further studies aimed at elucidating the true function of the protein. We still need to determine the reasons for these differences in expression in further research.

Previous data have shown that there appear to be significant differences in expression of BPIFA2 in human saliva depending on whether antibody BPIFA2(A) or BPIFA2(B) is used for detection (Bingle et al., 2009b). To study this in greater detail, whole saliva collected from eighteen individuals was used for Western blotting with the BPIFA2A antibodies, and our results showed clear differences in the banding pattern. The Western blots treated with BPIFA2(A) showed expression of the protein in fewer volunteers with less intense bands compared to those samples analysed with BPIFA2(B). Previous data from Western blot analysis and staining of glands (chapter 1; figure 1.9) indicated smaller bands were detected by BPIFA2(A) compared to BPIFA2(B) in unstimulated whole human saliva (Bingle *et al.*, 2009b).

The two BPIFA2-specific antibodies are raised to different epitopes of the protein, one to an epitope in the centre of the protein (BPIF2A) and another to an epitope at the C-terminus (BPIF2B). The reasons for the differences in expression patterns seen with these antibodies have not yet been fully elucidated but there are potentially two explanations. Firstly, alternative splicing of the BPIFA2 protein could potentially result in the production of different isoforms, which may not contain the epitopes needed to interact with the antibody. Previous research indicates that BPIFA2 can undergo alternative splicing and is likely regulated by two separate regulatory regions: a novel 5' exon located approximately 7 kb upstream of exon 2 and either exon 5 or exon 3 reaches into the intron (Bingle *et al.*, 2009).

To address the issue of possible differential glycosylation of BPIFA2, the human saliva samples were treated with N-glycanase, PNGase F, to remove the glycosylation, and were then subjected to Western blotting with antibody BPIFA2(B). The results showed a reduction of the distinct bands recognised by the BPIFA2 antibody to a single band of approximately 25 kDa, the predicted molecular weight based on amino acid sequence, suggesting that the larger molecular weight bands represented differentially glycosylated isoforms of the same protein. In the saliva samples of volunteers 17 and 18, a smaller shift in molecular weight and lower reduction in intensity suggests the protein still has carbohydrates attached, perhaps through O-linked sites. It is not currently possible to predict the presence or location of O-linked glycosylation sites and thus it is not possible to eliminate these sites to remove all traces of carbohydrate. The saliva sample of volunteer 12 indicated that additional digestion enzymes may be required to completely deglycosylate the carbohydrates attached.

Recombinant proteins are an effective resource and extensively used for protein functional analysis. To accomplish optimal production of target proteins, it is crucial to carefully design cloning and expression systems to meet the specific requirements of the final product. In this study, constructs lacking one or both glycosylation sites were generated in order to compare the function of "wild-type" BPIFA2 with non-glycosylated proteins. A further five constructs from the mouse, armadillo, macaque, squirrel monkey and dog were generated using the same method to study the function of BPIFA2 across the mammalian phylogeny.

The VR1255 vector has been used for designing target BPIFA2 plasmids as it is stable with high expression. A FLAG tag was integrated into the reverse primer utilised for the amplification of BPIFA2, so that the tag was inserted directly prior to the stop codon. The inclusion of a FLAG tag, consisting of a peptide sequence of eight amino acids (DYKDDDDK), was crucial for the subsequent isolation of BPIFA2. This is because monoclonal antibodies specifically designed to recognise the FLAG epitope can be employed for the purpose of affinity purification, utilising readily accessible resins that are commercially available.

After successfully inserting multiple BPIFA2 fragments into vectors, the plasmids were used for transfection experiments. BPIFA2 was initially transiently transfected into mammalian cells (HEK293) in a T75 flask using calcium phosphate. The calcium phosphate method (see chapter 2; section 2.3.10 for detailed information) facilitates the production of BPIFA2 protein in an effective and cost-effective way enabling large amounts of protein to be produced for downstream functional assays.

HEK293 cell line was used as the target cell line for transfection as mammalian cells are able to synthesise proteins with similar molecular structures to endogenous human and animal proteins. This cell line has the ability to produce complex glycosylated proteins, which is particularly significant in the case of BPIFA2, since it has been reported to have a complex glycosylation profile. Glycosylation can also impact various parameters such as protein production, elimination rate, durability, as well as bioreactivity and immunogenicity. HEK293 cells are commercially used for the synthesis of biotherapeutic proteins, demonstrating its aptness for the production of recombinant proteins (Dumont *et al.*, 2016).

The recombinant BPIFA2 protein was purified from conditioned media from transfected HEK293 cells using an affinity resin that has a monoclonal antibody immobilised on it. This antibody specifically binds to FLAG-tagged fusion proteins. The Coomassie-stained gels show bands indicating the presence solely of BPIFA2 proteins after purification and confirm that the purification process was successful in isolating pure BPIFA2 proteins from conditioned media. By comparing the intensity of protein bands, it is possible to estimate the quantity of protein in the samples and that the BPIFA2 protein samples are pure and of sufficient quality and quantity for the downstream use. The purified products were also analysed using Western blotting, employing a specific BPIFA2 antibody for the four human constructs and an anti- FLAG monoclonal antibody for the five animal constructs. The positive band of an identical molecular weight to that detected by specific BPIFA2 antibodies and anti-FLAG antibodies was visualised by Western blotting, supporting the identity of the recombinant protein of constructs.

The combination of the methods used in this chapter provided a comprehensive understanding

of BPIFA2's role in saliva. The investigation of BPIFA2 expression in natural variability reveals insights into BPIFA2 levels and susceptibility to oral health. The findings also revealed the relationship of glycosylation variability and the role of BPIFA2 in immune defence. Variations in glycosylation may impact the ability of BPIFA2 to bind to pathogens or microbes, which could potentially alter its antimicrobial efficacy. This will be further investigated through the functional assays described in chapter 5. The results of dot blot and densitometry analysis showed that the production efficiency of nine constructs was at the same level, suggesting we had used a reliable and reproducible method and had the correct tools for our downstream functional assays.

Chapter 5: Investigating the functional analysis of recombinant BPIFA2 with lipids, microbes and macrophage cells.

5.1 Introduction

BPI and LBP are innate immune proteins with well-established host defence roles due to their interactions with and upon LPS. Due to structural similarities, it was hypothesised that BPIFA1 (PLUNC), the first BPIF-containing protein identified in the mouse palate, lung and nasal tissues would play a similar host defence role and in 2010 a study by Gakhar et al demonstrated binding to LPS (Gakhar *et al.*, 2010), antibiofilm properties of the protein and also that it had surfactant activity. BPIFA2, a salivary member of the family, is mainly expressed in serous acinar cells and interlobular ducts of major salivary glands including human parotid, sublingual and submandibular glands. Similar to other BPIF-containing proteins, it has been hypothesised that due to the structural similarity to BPI and LBP, BPIFA2 will also play an important role in host defence (Bingle and Craven, 2002). As discussed in chapter 3, due to glycosylation of the protein there are multiple isoforms of BPIFA2 in saliva and these could be important for any innate immune function (Bingle *et al.*, 2009). Functional studies of human BPIFA2 are still very limited and the physiological function of BPIFA2 in saliva has yet to be established.

In this study, recombinant protein has been successfully produced and purified from four human and five animal constructs, providing an effective resource for functional studies. Our aims were to compare the function of "wild-type" human BPIFA2 with mutant forms, where

single N-glycosylation sites have been mutated and a further mutant form with both N-glycosylation sites mutated, thus producing BPIFA2 with normal and reduced glycosylation. We also wanted to compare human protein with various animal forms of BPIFA2 in functional assays to determine if the role of BPIFA2 is conserved. The animal species used in this study were armadillo, macaque, mouse, squirrel monkey and dog, as they span the evolutionary history of mammalian species. Armadillo is the most divergent in the phylogenetic tree of BPIFA2 and the mouse and dog are less divergent. The macaque is an old world monkey and squirrel monkey a new world monkey. New world monkeys diverged from old world monkeys approximately 40 million years ago which are more closely related to humans and share a more recent common ancestor.

It is well-known that the microenvironment of the oral cavity is extremely complicated due to diversity of surfaces such as teeth, gums and oral mucosa with a wide range of commensal and non-commensal microorganisms, including bacteria, fungi, and viruses (Elias and Banin, 2012). The interaction between commensal oral bacteria and the immune system requires them to work together to maintain oral health and to prevent oral diseases (Shang *et al.*, 2018). Commensal bacteria usually have an intimate relationship with their host through the bacteria themselves and also multiple components (e.g. peptidoglycans, bacterial DNA, LPS). In this study *in vitro* assays such as pull-down, biofilm and agglutination assays were performed to investigate the interaction of BPIFA2 with: *S. gordonii* (Gram-positive, oral commensal*)* and *S. mutans* (Gram-positive, pathogen); non-commensal: *P. aeruginosa* (Gram-negative) and *E. coli* (Gram-negative) and *S. aureus* (Gram-positive, commensal and pathogen).

The pull-down assay was used to determine if BPIFA2 protein binds specifically to bacterial strains while the agglutination assay was used to determine if BPIFA2 induces bacterial agglutination or clumping, as it has been shown to exhibit microbial agglutination and LPS binding properties (Gorr *et al.*, 2011). The survival of oral bacteria depends significantly on their capacity to attach to surfaces and form biofilms and it has previously been shown that synthetic peptides, GL13 (BPIFA2 141-153) have antibacterial activity through inhibition of LPS (Abdolhosseini *et al.*, 2012). The peptides did not kill the bacteria directly, rather they decreased bacterial adherence (biofilm) of *P. aeruginosa* and *E. coli* and promoted agglutination (Gorr *et al.*, 2008, 2011). BPIFA1 was also reported to inhibit biofilm formation of *P. aeruginosa in vitro* suggesting a role in the prevention of chronic infection of mucosal surfaces (Gakhar *et al.*, 2010). Studying the impact of BPIFA2 on biofilm formation would allow us to evaluate any functional activity linked to the BPIFA2 and further understand its biological function or significant role in host defence. BPI and LBP have influence on the growth and activity of pathogenic bacteria through a number of mechanisms, such as direct binding, suppression of the immune response to regulate inflammation, direct inhibition of bacterial growth and adherence and opsonisation to enhance phagocytosis. In order to determine if BPIFA2 plays a role in opsonisation and/or internalisation of bacteria we used Monomac 6 cells, an immortalised macrophage cell line, to study interactions between BPIFA2 proteins and phagocytic cells.

5.2 Aims

- To compare the functional properties of BPIFA2 proteins from specific animal

species with human variants, providing insights into evolutionary conservation and species-specific adaptations in antimicrobial and biofilm-modulating activities.

- To elucidate the potential role of BPIFA2 in lipid interactions and the relevance to glycosylation sites and species.

- To explore the interactions between BPIFA2 and oral bacteria and elucidate potential roles in modulating microbial behavior and inflammatory responses.

5.3 Materials and methods

The relevant materials and methods utilised in this chapter are detailed in chapter 2, please refer to sections:

- The pull-down assay was used to determine whether BPIFA2 protein binds specifically to the bacterial strains *S. gordonii, P. aeruginosa, S. aureus, E. coli, S. mutans* as described in section 2.5.3.

- The agglutination assay was used to determine the ability of BPIFA2 to induce bacterial agglutination as described in section 2.5.4.

- The biofilm formation assay was used to determine whether BPIFA2 protein has the ability to inhibit oral commensal and non-commensal bacterial biofilm formation as described in section 2.5.5.

- The lipid-binding assay described in this study was used to investigate the binding of

BPIFA2 to various lipids as described in section 2.6.

– The opsonisation assay was used to determine whether BPIFA2-protein coated bacterial strains are differentially phagocytosed by MM6 cells as described in section 2.7.

5.4 Functional assays of BPIFA2

5.4.1 Pull-down assay

A pull-down assay was used to determine whether recombinant BPIFA2 binds specifically to bacteria; both commensal and non-commensal, found in the oral cavity. The $A_{(600)}$ of an overnight culture of the bacterial stains *E. coli*, *S. aureus, P. aeruginosa, S. gordonii and S. mutans* was measured and 1ml of bacteria with OD1 was used per treatment. The bacteria were washed twice in 1ml PBS buffer by centrifugation (13,000rpm for 5 minutes) and incubated at 37 °C for 1 hour with either saliva, as a positive control, PBS buffer, as a negative control, or recombinant human BPIFA2. After three washing steps to remove unbound protein from the bacteria both the bacterial pellet and the supernatant containing unbound protein were analysed by Western blotting using a specific BPIFA2 antibody (Figure 5.1).

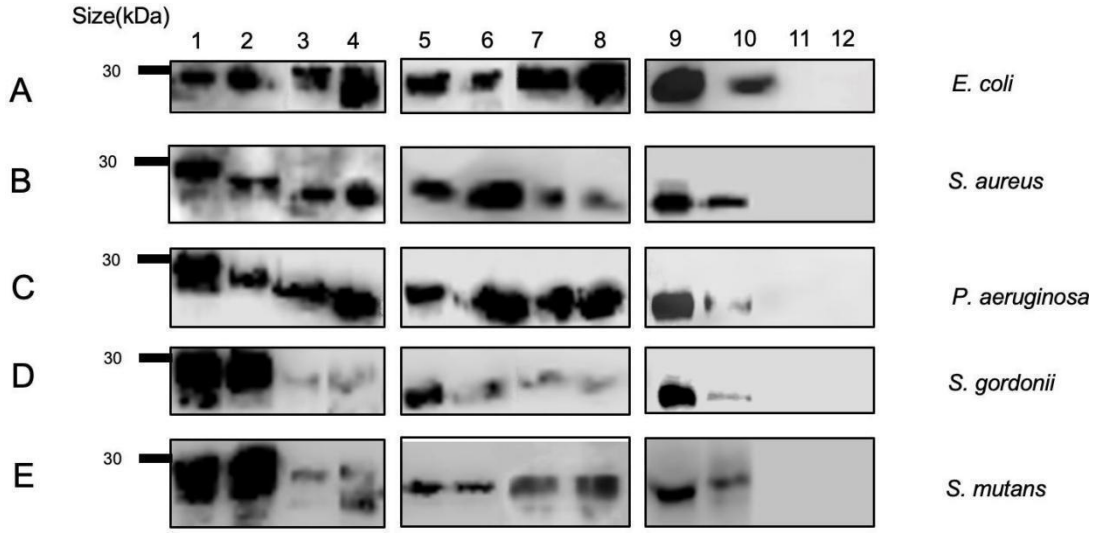
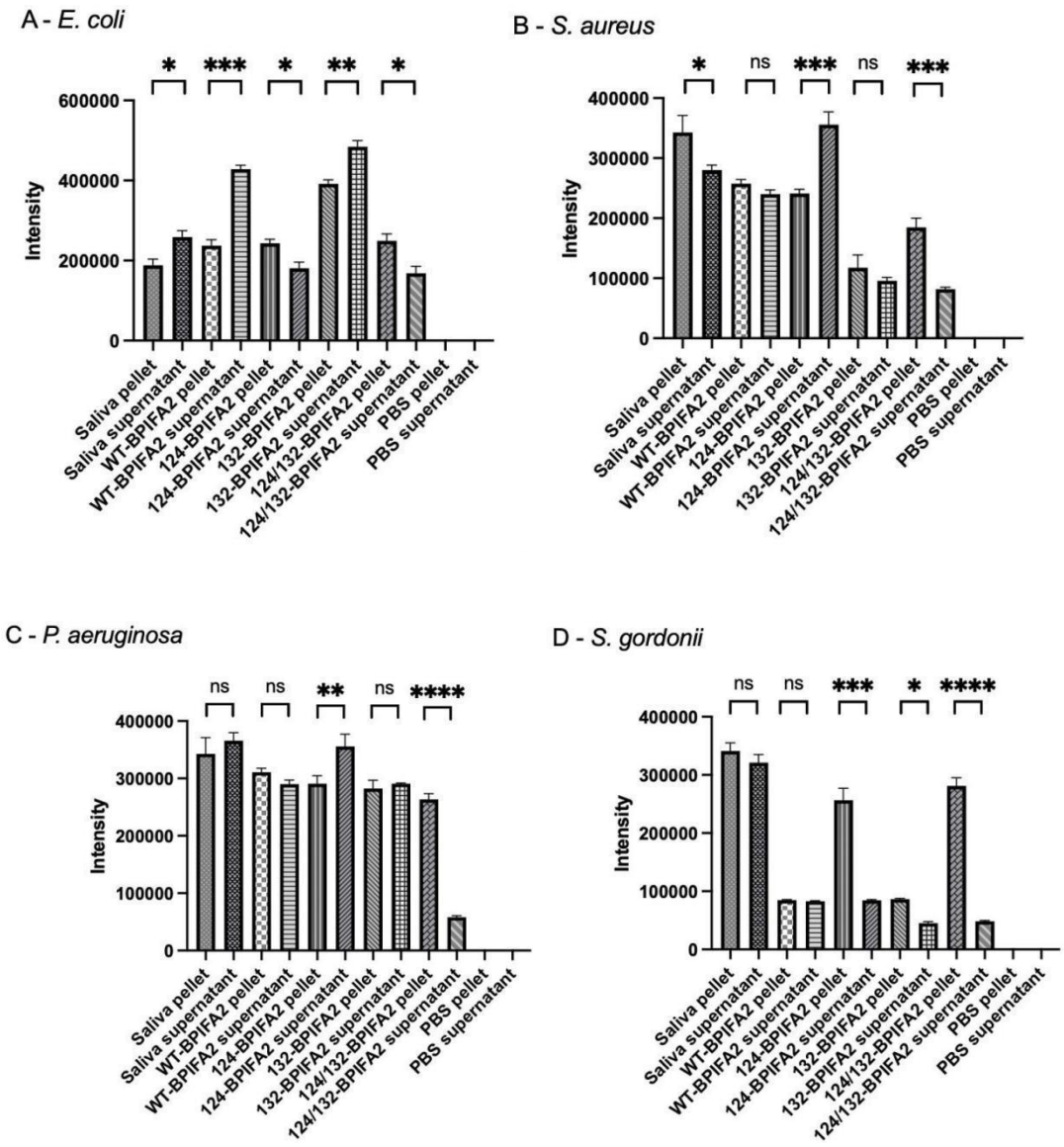


Figure 5.1. Binding capacity of human recombinant BPIFA2 to bacteria. The capacity of BPIFA2 to bind to *E. coli* (A), *S. aureus* (B), *P. aeruginosa* (C)*, S. gordonii* (D) and *S. mutans* (E) was assessed using a pull-down assay and subsequent Western blotting with BPIFA2(B) antibody. Human saliva served as a positive control and PBS as a negative control. Lane numbers in each blot represent the following: 1 – bacterial pellet with bound human saliva, 2 – unbound human saliva, 3– bacterial pellet with bound WT BPIFA2, 4 – unbound WT BPIFA2, 5 – bacterial pellet with bound 124 BPIFA2, 6 – unbound 124 BPIFA2, 7–bacterial pellet with bound 132 BPIFA2, 8 – unbound 132 BPIFA2, 9 –bacterial pellet with bound 124/132 BPIFA2, 10 – unbound 124/132 BPIFA2, 11 - PBS treated bacterial pellet and 12 –unbound PBS buffer. Representative image shown (n=3).


A - E. coli
B - S. aureus
C - P. aeruginosa
D - S. gordonii
Intensity
Saliva pellet
Saliva supernatant
WT-BPIFA2 pellet
WT-BPIFA2 supernatant
124-BPIFA2 pellet
124-BPIFA2 supernatant
132-BPIFA2 pellet
132-BPIFA2 supernatant
124/132-BPIFA2 pellet
124/132-BPIFA2 supernatant
PBS pellet
PBS supernatant

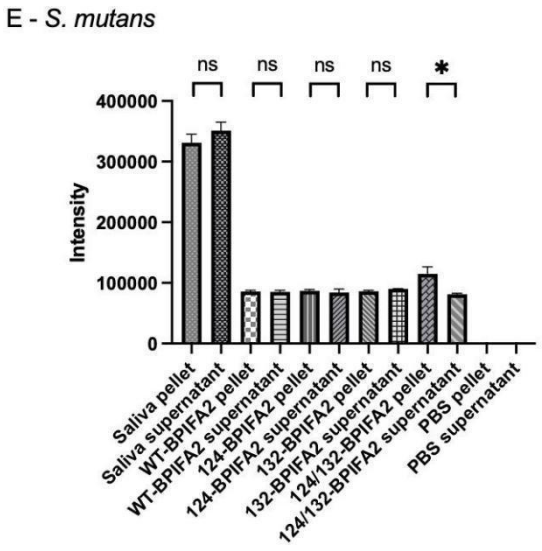


Figure 5.2. Densitometry of pull-down assay for the interaction between recombinant BPIFA2 and bacteria. Densitometry analysis derived from analysis of Western blots comparing BPIFA2 interaction with *E. coli* (A), *S. aureus* (B), *P. aeruginosa* (C), *S. gordonii* (D) and *S. mutans* (E). BPIFA2(B) antibody used for detection of samples. Data presented as the mean ± SD (N=3); * P ≤ 0.05; ** P ≤0.01; *** ≤ 0.001; **** P ≤ 0.0001; ns = not significant.

The Western blots showed that all bacterial strains bound to the BPIFA2 present in whole saliva (lane 1) but no protein was detected in the PBS negative controls (lanes 11 and lane 12). This suggests that the BPIFA2 found in whole saliva had the ability to bind to the bacteria. Unbound BPIFA2 was also detected in the saliva samples (lane 2) following incubation with all bacterial strains suggesting a significant amount of protein had not bound to the bacteria; it is possible there was significantly more BPIFA2 in the saliva than needed to bind to these bacteria.

Densitometry analysis was used to quantify the intensity of the Western blot bands and thus the relative amount of BPIFA2 bound to the bacteria (Figure 5.2).

The Western blots demonstrated significant binding of all human BIFA2 proteins to *E. coli* but

also that not all of the protein in our samples attached to the bacteria as protein was detected in the supernatants as well as the pellet; binding appeared comparable to that seen with the saliva positive control (Figure 5.1A). The densitometry analysis allows a more quantitative comparison of relative binding of protein and bacteria, where, for example, lower levels of WT (lane 3) and 132 BPIFA2 (lane 7) were noted in *E. coli* pellets in comparison to that seen in the supernatants (lanes 4 and 8 respectively). Inversely, and interestingly, more BPIFA2 was found in the bacterial pellets than supernatants when the interactions between *E. coli* and the 124 and double mutant proteins were analysed. While Figure 5.1B shows all recombinant human proteins bound to *S. aureus* and the densitometry analysis showed similar levels of both bound and unbound WT and 132 proteins, more of the double mutant bound to the *S. aureus* than was detected in the supernatant but inversely, less of the 124 BPIFA2 protein bound to the bacteria. Human BPIFA2 proteins bound to *P. aeruginosa* (Figure 5.1C) in a similar manner to that seen with *S. aureus;* both are opportunistic pathogens commonly associated with multiple human infections not limited mainly to the oral cavity, meaning they can cause disease primarily in people with weakened immune systems. Further research is required to investigate the specificity of the interaction between the 124 BPIFA2 and *S. aureus* and *P. aeruginosa*.

Overall, there seemed to be less protein detected in the Western blots of interactions between BPIFA2 and *S. gordonii* (Figure 5.1D) this was most apparent, and reflected in the densitometry data, with the WT and 132 proteins. The 124 and double mutant proteins both indicated significant binding to the bacteria as more protein was detected in the pellet than in the supernatant; this is similar to the results seen with *E. Coli*. Interestingly, the binding of all forms of BPIFA2 to *S. mutans* (Figure 5.1E) was lower overall and there was very little difference

between bound and unbound samples. This suggests the proteins did not adhere to *S. mutans* in the same way as they did to the other strains tested. With the exception of the 124 and double mutant proteins the binding of BPIFA2 to *S. gordonii* was comparable to that of *S. mutans* and so we might hypothesise that as these are both oral commensal Gram-positive bacteria the lack of an outer lipid membrane results in reduced binding capacity. However, there is significant variation across the strains of bacteria tested and the forms of BPIFA2 and so a larger panel of bacteria would need to be tested to confirm such a hypothesis.

Unfortunately, due to time limitations pull-down assays with animal BPIFA2 proteins and bacterial strains were not undertaken.

5.4.2 Agglutination assay

Successful agglutination was observed following incubation of saliva (positive control) and BPIFA2 recombinant proteins and bacterial strains confirming the ability of BPIFA2 to induce bacterial agglutination. Absence of agglutination was identified by observing a tight cluster of bacteria at the bottom of the well (as in the negative controls) whereas the presence of agglutination was identified by the appearance of a flat or smaller cluster, as in the saliva samples.

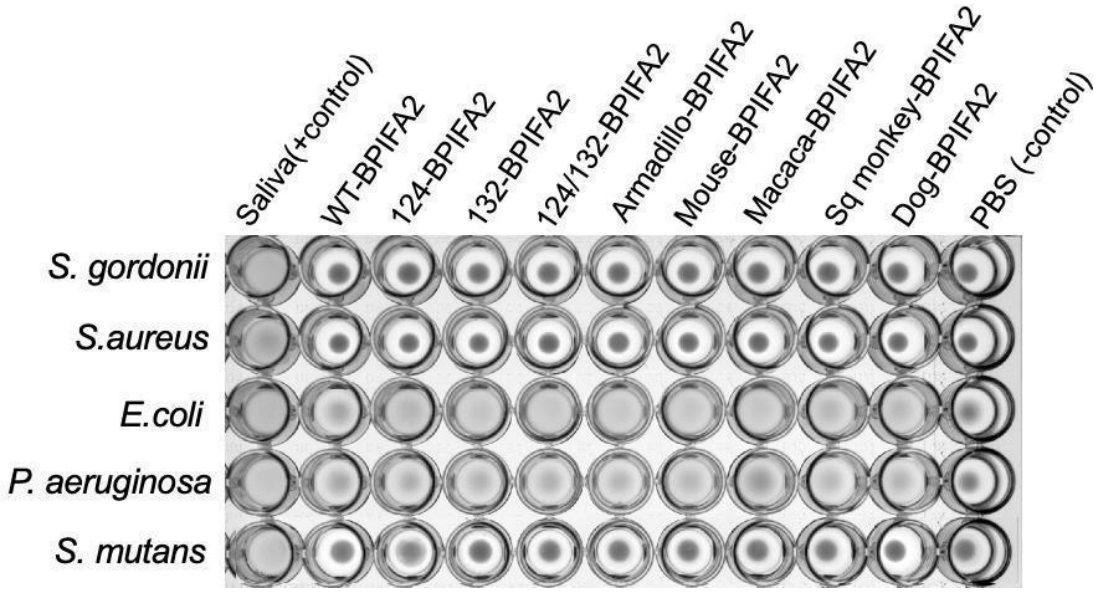
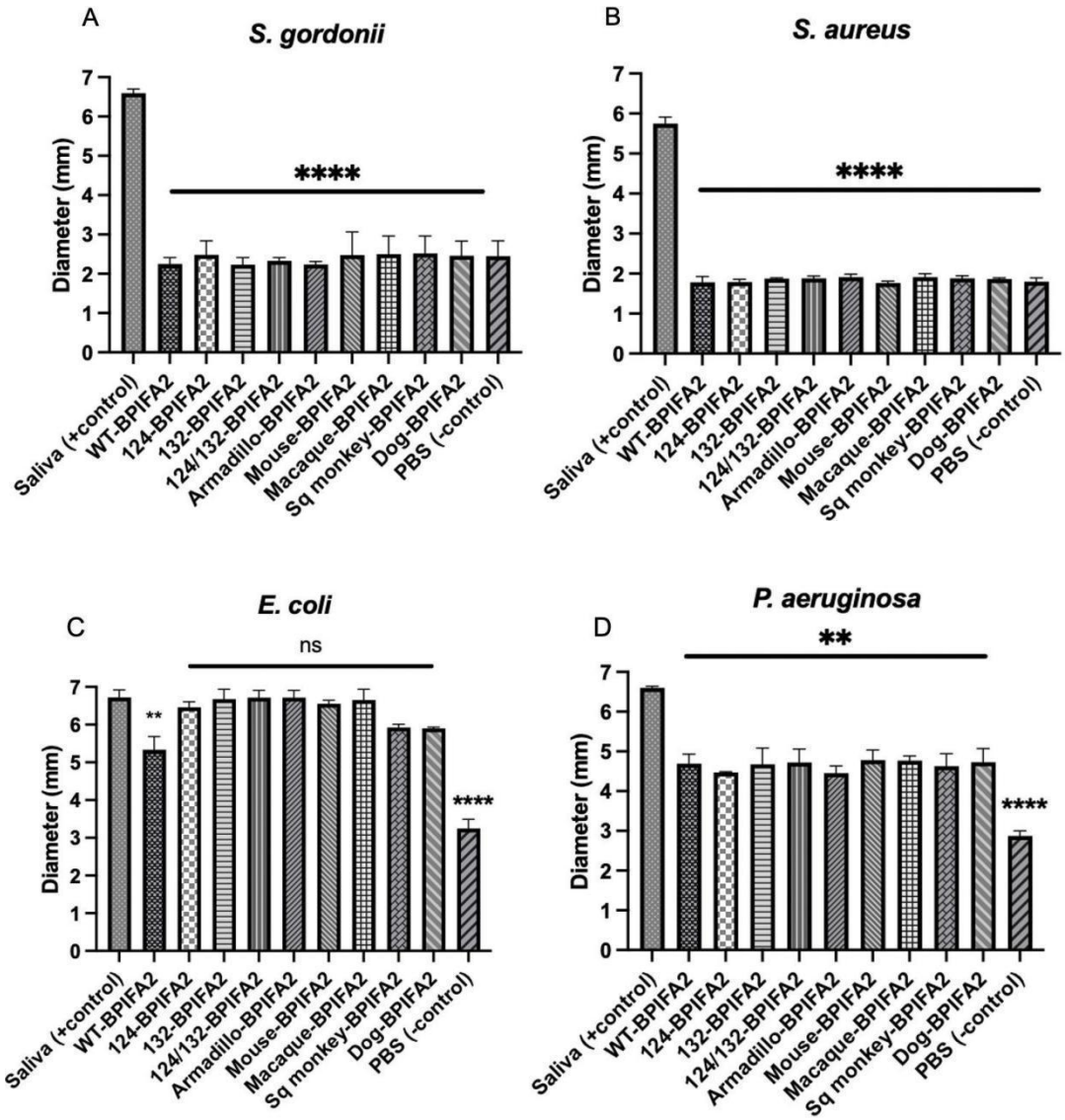


Figure 5.3. The agglutination capacity of recombinant BPIFA2. Bacterial strains *S. gordonii, S. aureus, E. coli, P. aeruginosa* and *S. mutans* were incubated with human and animal BPIFA2 overnight to determine the capacity of BPIFA2 to induce agglutination. Saliva was used as a positive control and PBS as a negative control (n=3; representative image shown).


A
S. gordonii
Diameter (mm)
****
B
S. aureus
Diameter (mm)
****
C
E. coli
ns
**
****
Diameter (mm)
D
P. aeruginosa
**
****
Diameter (mm)
Saliva (+control)
WT-BPIFA2
124-BPIFA2
132-BPIFA2
124/132-BPIFA2
Armadillo-BPIFA2
Mouse-BPIFA2
Macaque-BPIFA2
Sq monkey-BPIFA2
Dog-BPIFA2
PBS (-control)

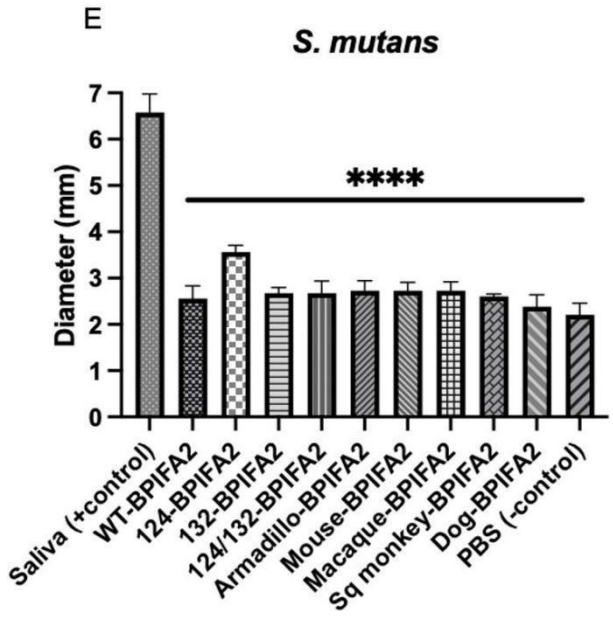


Figure 5.4. Quantitative analysis of agglutination as a result of the incubation of recombinant BPIFA2 proteins and bacterial strains. *S. gordonii,* (A), *S. aureus* (B), *E. coli* (C), *P. aeruginosa* (D) and *S. mutans* (E) were incubated with human and animal BPIFA2 for 24 hours. Human saliva was used as a positive control and PBS as a negative control. Data presented as the mean ± SD (N=3); * $P \leq 0.05$; ** $P \leq 0.01$; *** ≤ 0.001; **** $P \leq 0.0001$; ns = not significant.

Human saliva caused agglutination of all bacterial strains tested while no bacterial agglutination was seen with the negative control (PBS). Incubation of *S. gordonii, S. aureus* and *S. mutans* with all human and animal BPIFA2 proteins resulted in a tight cluster of bacteria at the bottom of the wells indicating the proteins did not induce agglutination of these particular bacterial strains. *P. aeruginosa* and *E. coli,* appeared as flat buttons, similar to that seen with saliva, suggesting positive agglutination induced by all BPIFA2 recombinant proteins tested (Figure 5.3). A relatively subjective and visual measurement to quantify the diameter of agglutinated patterns using Image J software and statistical analysis is described in Figure 5.4. The incubation with *S. gordonii* (A) and*S. aureus* (B) resulted in small diameters with both human and animal BPIFA2 that were the same as negative controls,

suggesting a significant difference from saliva positive controls. Incubation with BPIFA2 with *E. coli* (C) and *P. aeruginosa* (D) bacterial strains caused more agglutination, as a flat pattern, suggesting significant agglutination had occurred with these samplesthan the other bacteria tested.

Initially, the level of agglutination is subjectively assessed based on the pattern, density, and uniformity of the clumps formed. This qualitative assessment provides a rough, manual and visual estimation of the extent of agglutination.

5.4.3 Biofilm formation assay

It has been demonstrated that BPIFA2 peptide (GL13K) and BPIFA1 have the ability to prevent *P. aeruginosa* from forming biofilm (Gorr *et al.*, 2008; Gakhar *et al.*, 2010). A biofilm assay would enable us to look at any similar functional activity associated with BPIFA2. Recombinant BPIFA2 was incubated with bacterial cultures for 48 hours in a 96-well plate and biofilms were stained with 0.1% crystal violet which was subsequently dissolved in acetic acid to allow for the quantification of the biofilms using a Tecan microplate reader. Wells containing sterile broth were also incubated and subsequently stained to enable background readings to be subtracted from all other values (Figure 5.5).

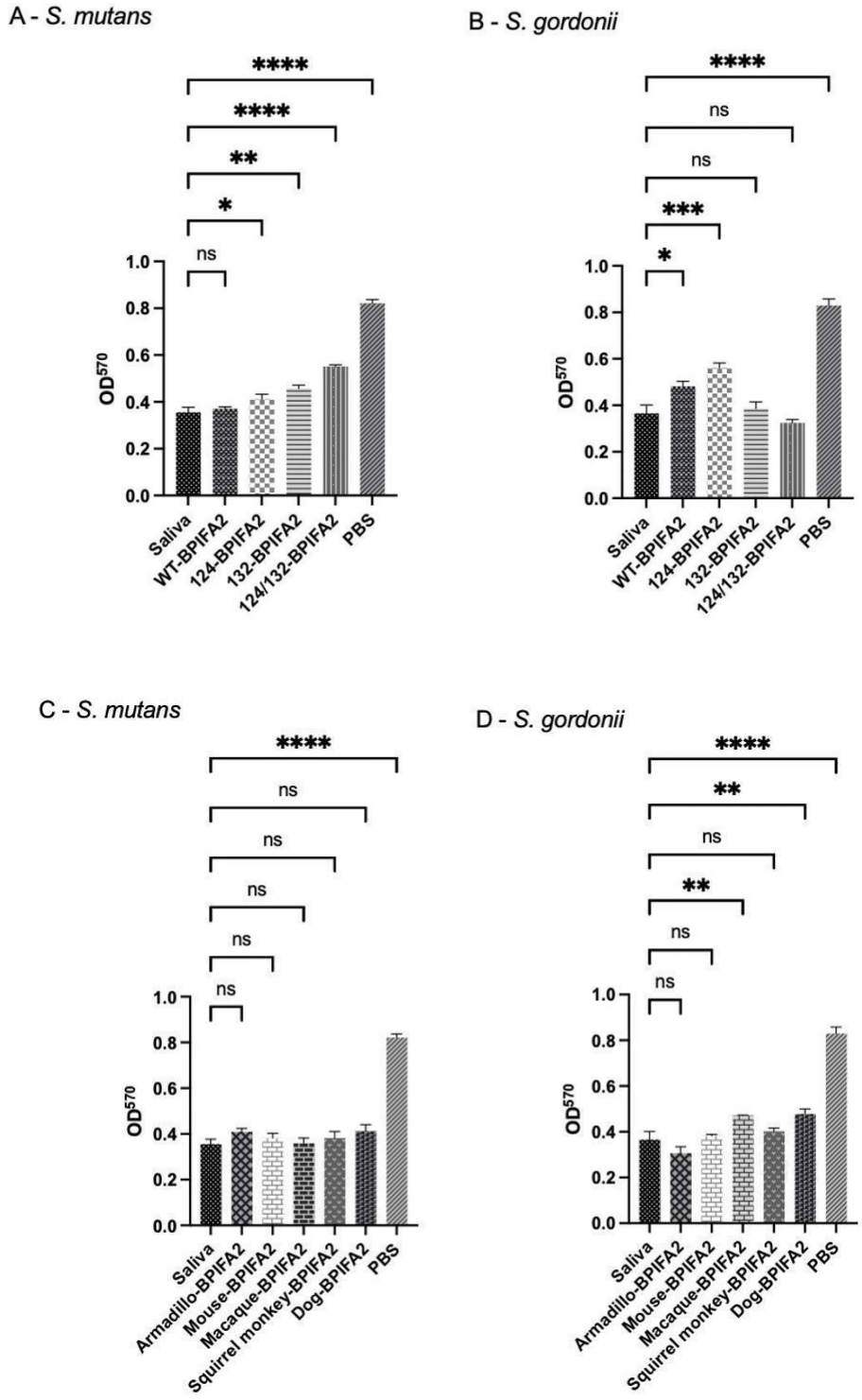


Figure 5.5. The effect of recombinant BPIFA2 on oral commensal bacterial biofilm formation. *S. mutans* (A and C) and *S. gordonii* (B and D) were incubated with human and animal BPIFA2 for 48 hours. Human saliva was used as a positive control and PBS as a negative control. The resultant biofilms that formed in the presence of BPIFA2 were stained and quantified. Data presented as the mean ± SD (N=3; n=3); * $P \leq 0.05$; ** $P \leq 0.01$; *** ≤ 0.001;**** $P \leq 0.0001$; ns = not significant.

S. mutans and *S. gordonii* were selected as oral commensal bacterial strains and human saliva was used as a positive control and PBS as a negative control. Human WT and 124 BPIFA2 inhibited *S. mutans* biofilm formation in a similar manner to saliva but 132 BPIFA2 and the double mutant (124/132) did not (Figure 5.5A and 5.5B). Inversely, and interestingly, 132 and 124/132 BPIFA2 inhibited biofilm formation of *S. gordonii* but the WT and 124 BPIFA2 did not. Formation of biofilm by both *S. mutans* and *S. gordonii* was inhibited by all animal BPIFA2 proteins apart from when *S. gordonii* was incubated with macaque and dog BPIFA2 (Figure 5.5C and 5.5D).

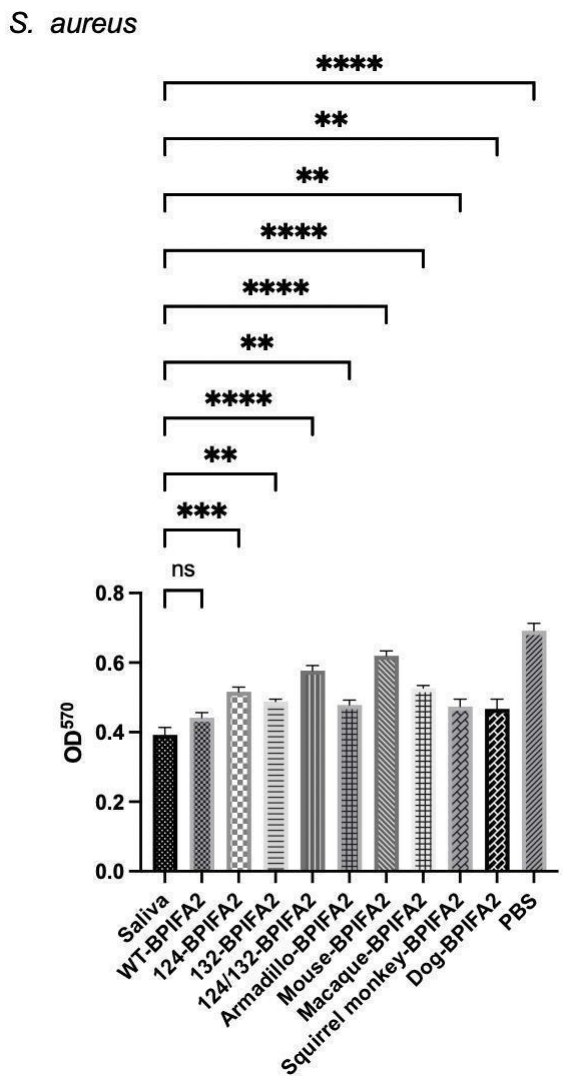


Figure 5.6. The effect of recombinant BPIFA2 on a respiratory commensal bacterial biofilm formation. *S. aureus* was incubated respectively with human BPIFA2 and animal BPIFA2 for 48 hours. Human saliva served as a positive control and PBS as a negative control. The resultant biofilms that formed in the presence of BPIFA2 were stained and quantified. Data presented as the mean ± SD (N=3; n=3); * $P \leq 0.05$; ** $P \leq 0.01$; *** ≤ 0.001; **** $P \leq 0.0001$; ns = not significant.

S. aureus was selected as a respiratory commensal bacterial strain and human saliva was used as a positive control and PBS as a negative control (Figure 5.6). Neither human 124, 132 nor 124/132 BPIFA2 inhibited *S. aureus* biofilm formation but WT BPIFA2 inhibited biofilm formation in a similar manner to that seen with saliva. None of the animal proteins were able to inhibit the formation of *S. aureus* biofilm.

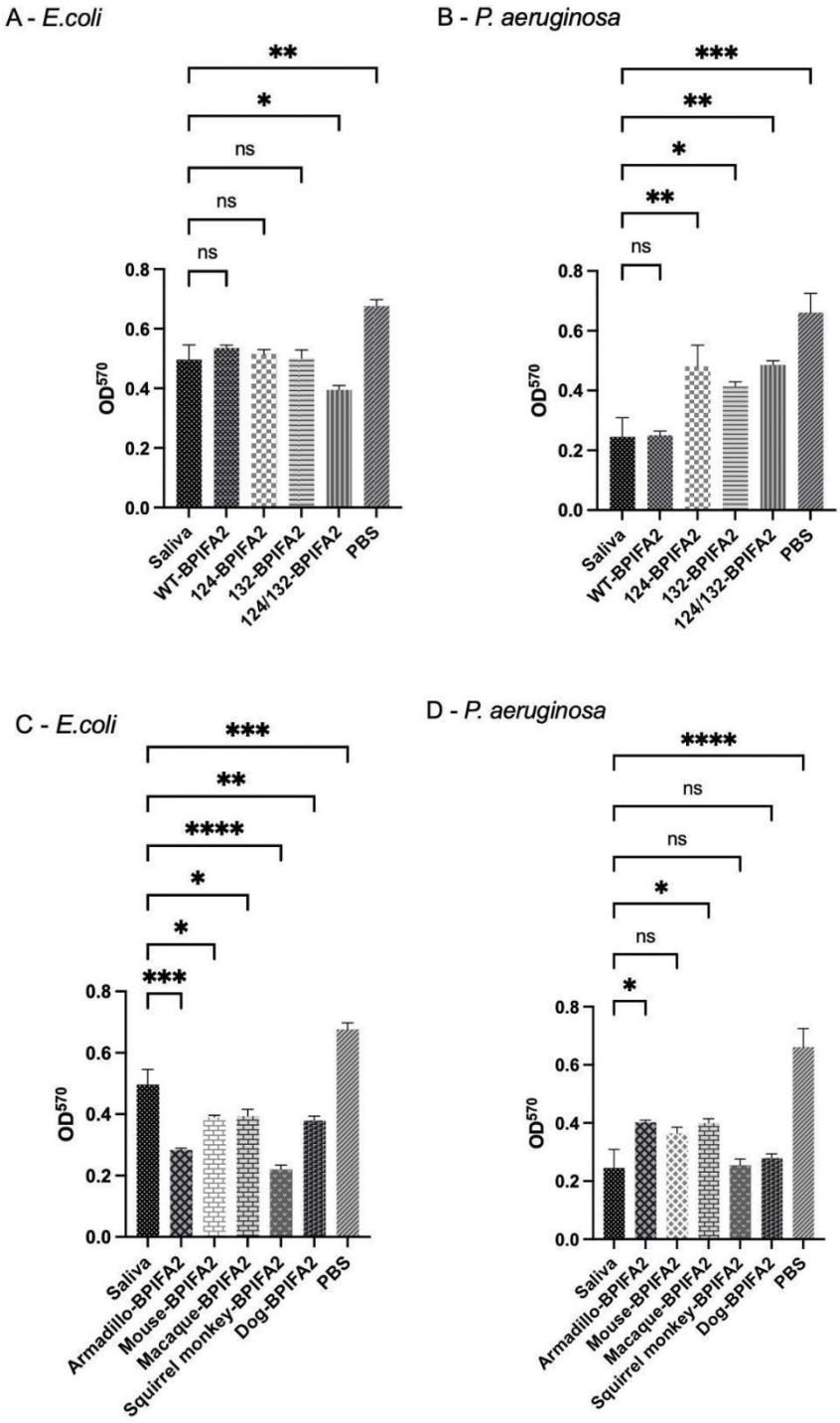


Figure 5.7. The effect of recombinant BPIFA2 on oral non-commensal bacteria biofilm formation. *E. coli* and *P. aeruginosa* were incubated with human (A and B) and animal (C and D) BPIFA2 for 48 hours. Human saliva served as a positive control and PBS as a negative control. The resultant biofilms that formed in the presence of BPIFA2 were stained and quantified. Data presented as the mean ± SD (N=3; n=3); * $P \leq 0.05$; ** $P \leq 0.01$; *** ≤ 0.001; **** $P \leq 0.0001$; ns = not significant.

E. coli and *P. aeruginosa* were selected as oral non-commensal bacterial strains and the results demonstrate that WT, 124 and 132 BPIFA2 inhibit *E. coli* biofilm formation in a similar manner to saliva but the 124/132 BPIFA2 had no effect (Figure 5.7A). While the results of biofilm formation with *P. aeruginosa* suggested that the WT and 132 BPIFA2 inhibited biofilm formation the 124 and 124/132 BPIFA2 did not (Figure 5.7B). Both *S. aureus* and *P. aeruginosa* are respiratory pathogens but clearly there are differences in their ability to form biofilm in the presence of some of the human BPIFA2 proteins (Figure 5.6) as human 124, 132 and 124/132 BPIFA2 did not inhibit *S. aureus* biofilm formation but WT BPIFA2 did. *E. coli* biofilm formation was inhibited by saliva but, interestingly, BPIFA2 from all of the animal species tested inhibited biofilm to a greater extent (Figure 5.7C). The incubation of mouse, squirrel monkey and dog BPIFA2 inhibited the formation of *P. aeruginosa* biofilm but armadillo and macaque did not (Figure 5.7D). Again, there was a contrast between the response of *S. aureus* and *P. aeruginosa* as none of the animal forms of BPIFA2 inhibited biofilm formation of *S. aureus*.

S. aureus helps to provide the structural support for biofilm formation that adhere to surfaces and are encased in a protective matrix (Lynge Pedersen and Belstrøm, 2019). *P. aeruginosa* was known for biofilm formation and resistance to antibiotics and it is a significant concern in chronic infections. However, *S. aureus* is a Gram-positive bacterium and *P. aeruginosa* is a Gram-negative bacterium. The two strains have different morphologies, and also differ in their ability to adhere to surfaces and to resist host immune responses. These differences help explain how they interact with BPIFA2 proteins in distinct ways.

In this study our bacterial biofilm research focused on mono-species cultures, which could be considered a limitation as natural biofilm communities consist of a diverse range of bacteria. The composition of a mixed biofilm and the interactions among its microorganisms play a crucial role in determining the growth and form of the community (Elias and Banin, 2012). Therefore, conducting experiments on multiple species of bacteria in both biofilm formation assays and biofilm disruption assays would be helpful. These factors involve the facilitation of biofilm development by co-aggregation, potentially resulting in enhanced resistance to BPIFA2 when compared to biofilms that contain a single species.

5.4.4 The interaction between recombinant BPIFA2 and lipids

Similarities between the predicted structure of BPIF proteins and BPI, led to the hypothesis that BPIF proteins, including BPIFA2, would possess a hydrophobic crevice appropriate for the binding of lipid-like molecules (Hailman *et al.*, 1994; Bingle and Craven, 2002). It was hypothesised further that this lipid binding would be related to the innate immune function of the proteins as shown with BPI (Gazzano-Santoro *et al.*, 1995). Commercial lipid strips, containing samples of 15 physiologically active lipids that are found in cell membranes, were used to evaluate which, if any, lipids BPIFA2 proteins bind to and whether glycosylation alters this interaction. PI (4,5) P2 Grip™ and Anti-GST HRP were provided by the manufacturer for use as primary and secondary antibodies to ensure the lipid binding assay was successful. Enhanced Chemiluminescence Western Blotting substrate and a Li-Cor C-Digit Western Blot Scanner were used to determine the binding of BPIFA2 to the membrane lipids. As expected, there was clear binding with the PtdIns (4,5) P2 lipid using PI (4,5) P2 Grip™ and Anti-GST HRP antibodies (Figure 5.8A). Whole human saliva was used as a

positive control and showed clear binding to the lipids Phosphatidic acid (PA), PtdIns (4)P, PtdIns (4,5)P2 and PtdIns (3,4,5)P3 (Figure 5.8B). We were able to demonstrate that WT BPIFA2 also bound to (PtdIns(4)P, PtdIns (4,5)P2 and PtdIns (3,4,5)P3) (Figure 5.8C) but none of the other human proteins bound to any of the lipids suggesting deglycosylation prevented the lipid-protein interaction (Figure 5.8DEF).

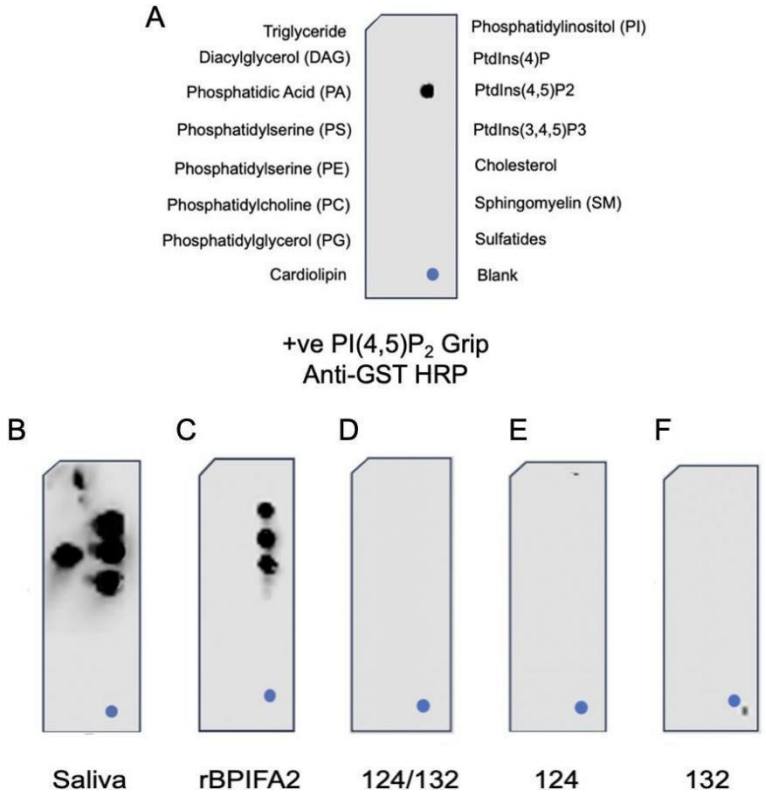


Figure 5.8. Interaction between human recombinant BPIFA2 and membrane lipids. Membrane lipids were used to determine which lipids the proteins bind to and if glycosylation alters this interaction. A. Example of positive binding of primary antibody PI (4,5) P2 Grip™ and secondary antibody Anti-GST HRP. B. Whole human saliva. C. WT BPIFA2. D. 124/132 BPIFA2. E. 124 BPIFA2. F. 132 BPIFA2. BPIFA2(B) antibody was used for specific detection of samples (n=3; representative image shown).

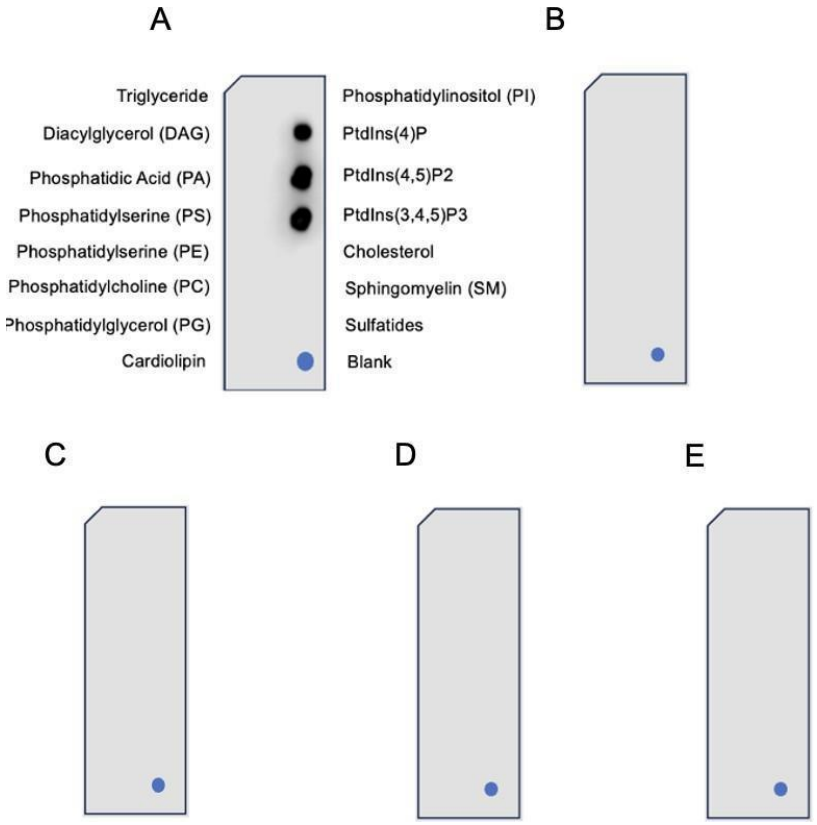


Figure 5.9. Interaction between animal BPIFA2 and membrane lipids. Membrane lipids were used to determine lipid-protein interaction between animal BPIFA2 protein and lipids. A. Macaque BPIFA2. B. Armadillo BPIFA2. C. Mouse BPIFA2. D. Squirrel monkey BPIFA2. E. Dog BPIFA2. Anti-flag M2 antibody was used for detection of samples (n=3; representative image shown).

For the interaction between animal BPIFA2 proteins and membrane lipids, recombinant BPIFA2 from the macaque (Figure 5.9A) showed binding to the same lipids as human WT BPIFA2, namely PtdIns(4)P, PtdIns (4,5)P2 and PtdIns (3,4,5)P3 but no positive binding was detected with armadillo, mouse, squirrel monkey or dog BPIFA2 (Figure 5.9BCDE). To date there is very limited information available regarding the interaction between BPIFA2 and lipids, however, this study emphasizes that the interaction appears to depend on the origin and glycosylation status of the protein.

5.4.5 Opsonisation assay

A preliminary western blot analysis (described in chapter 2; section 2.7.1) was used to determine whether BPIFA2 is produced by MM6 cells (Figure 5.10). Saliva was a positive control and PBS a negative control. The same bacterial strains as used in previous assays (sections 5.4.1-5.4.3)*)* were coated with BPIFA2, added to tissue culture wells containing MM6 cells and co-incubated for 1 hour at 37°C. MM6 lysates were seeded onto LB-agar plates to quantify the bacteria colonies that had either adhered to the MM6 cells or had been ingested/phagocytosed by the cells; non-attached bacteria were removed. However, counting colony-forming units (CFUs) did not reveal whether the host cells were surface-bound or phagocytosed. Human saliva (lane 1) was used as a positive control and as expected there was no BPIFA2 expression by MM6 cells (lane 2).

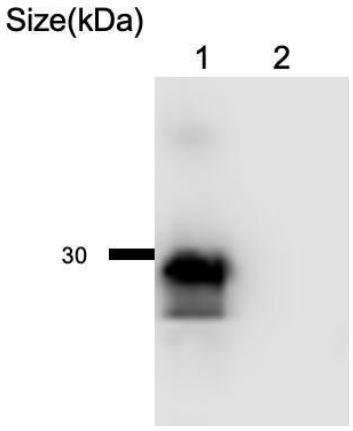


Figure 5.10. Western blot analysis of MM6 cell expression. Western blot analysis was used to determine whether there is BPIFA2 expression in MM6 cells. Lane 1-human saliva, lane 2-MM6 cells. BPIFA2(B) antibody was used for detection of specific protein (Representative image from n=3).

The opsonisation assay was used to determine whether BPIFA2 protein-coated bacteria would be differentially phagocytosed by MM6 cells as this could give information regarding a mechanism of action related to host defence. However, the results only described phagocytosis and adherence to the MM6 cells as a whole section and compared it to non-attached macrophages in this chapter.

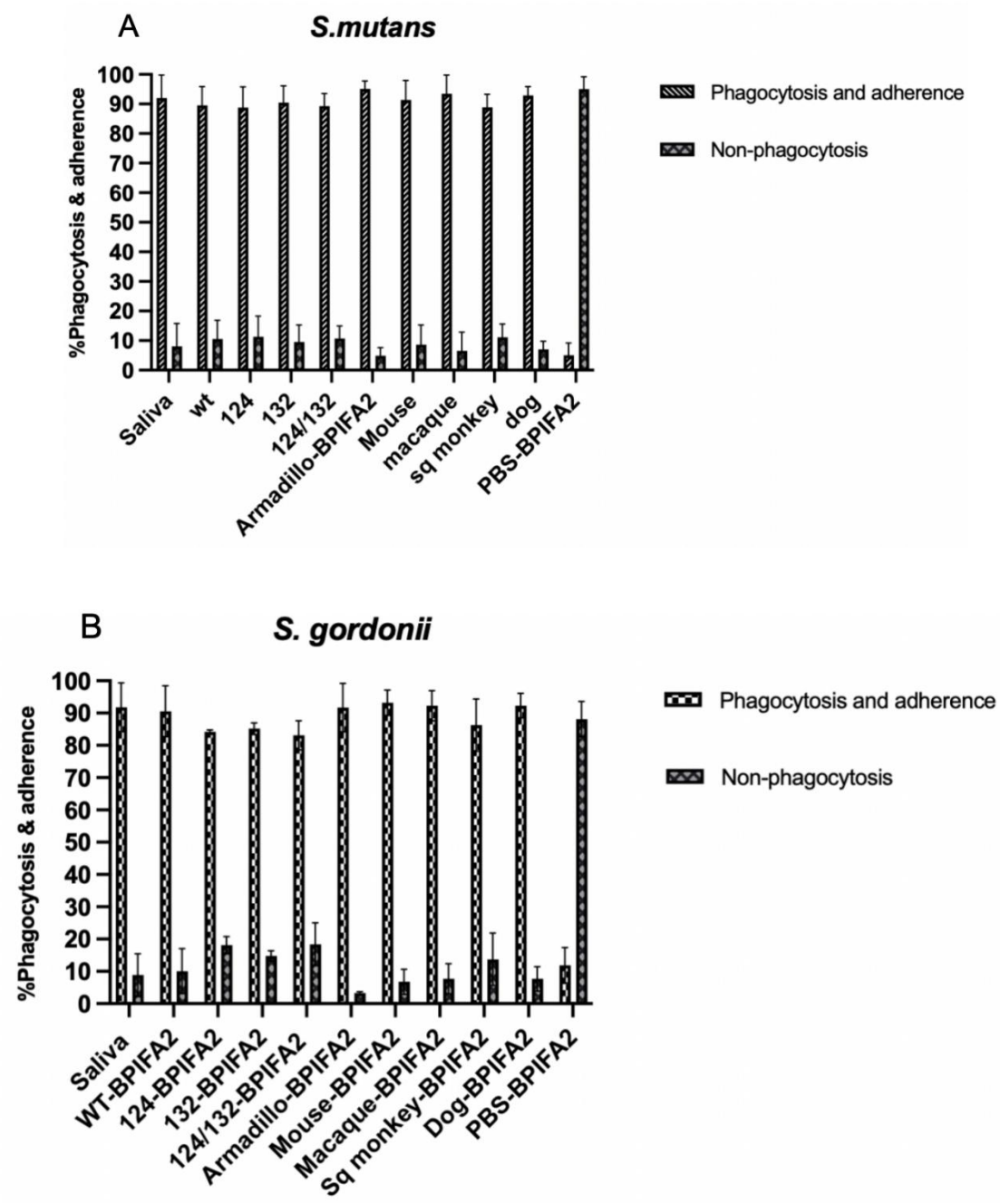


Figure 5.11. Phagocytosis and/or adherence of oral commensal bacteria strains following opsonisation with BPIFA2. Saliva was used as a positive control and PBS was a negative control. *S. mutans* (A) and *S. gordonii* (B) were coated with human BPIFA2 and animal BPIFA2 as described in the X-axis. The Y-axis represents the percentage of bacteria that were phagocytosed and/or attached to the surface of MM6 cells. Data presented as mean ± SEM (N=3).

The results show that *S. mutans* was opsonized by all human and all animal species in a similar manner to the positive control, saliva (Figure 5.11A).

Similar results were seen with *S. gordonii* (Figure 5.11B) as all human and animal BPIFA2 proteins opsonised the bacteria in a similar manner to saliva.

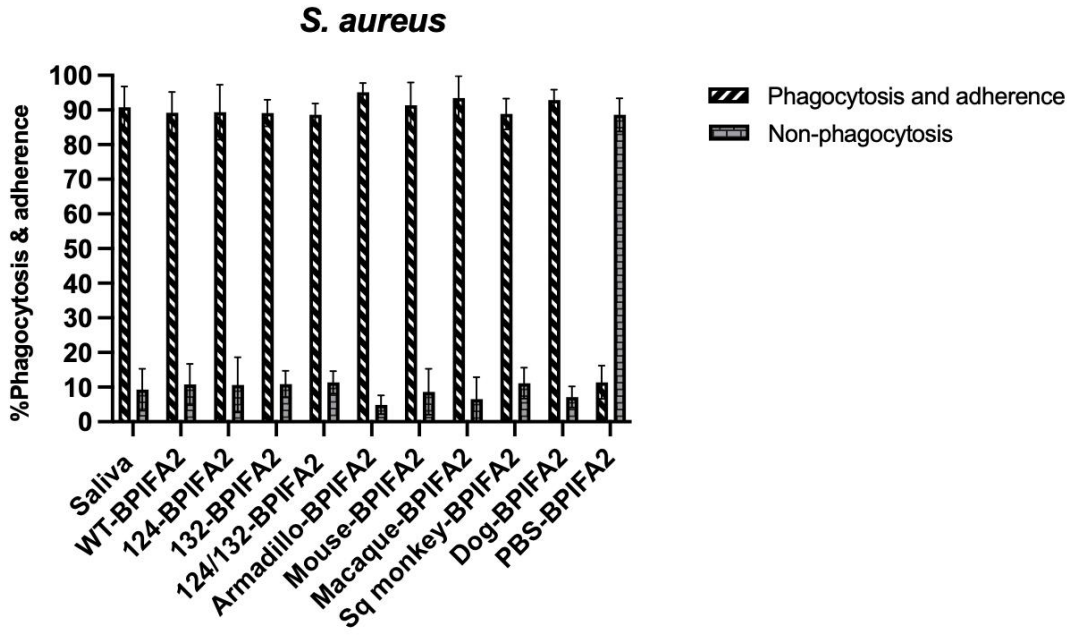


Figure 5.12. Phagocytosis and/or adherence of a respiratory commensal bacterial strain following opsonisation with BPIFA2. Saliva was used as a positive control and PBS was a negative control. *S. aureus* was coated with human BPIFA2 and animal BPIFA2 as described in the X-axis. The Y-axis represents the percentage of bacteria that were phagocytosed and/or attached to the surface of MM6 cells. Data presented as mean ± SEM (N=3).

S. aureus was selected as a respiratory commensal bacterial strain and the results demonstrate that, as with *S. mutans* and *S. gordonii* (Figure 5.11) all human and animal BPIFA2 proteins opsonised *S. aureus* (Figure 5.12) in a similar manner to saliva.

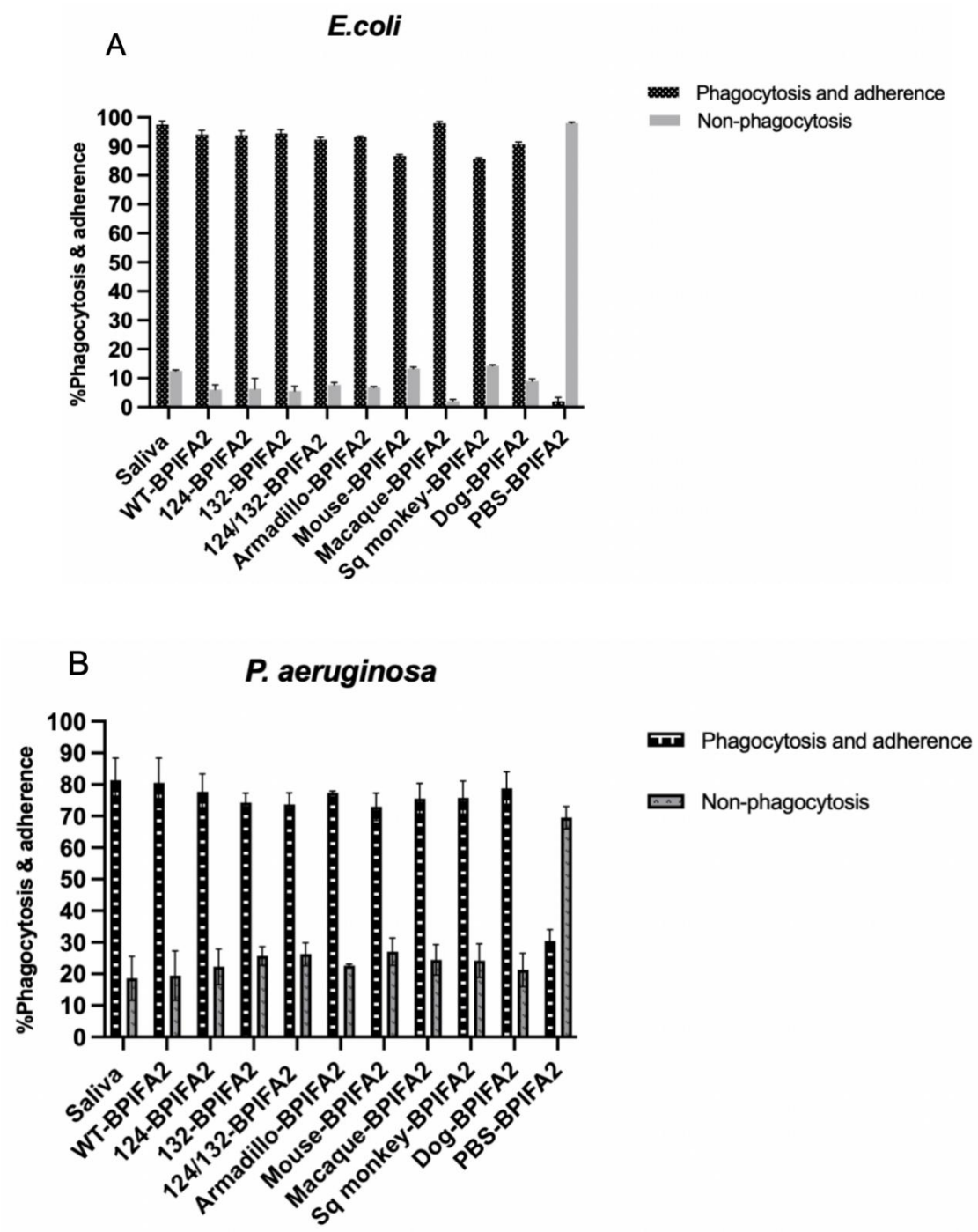


Figure 5.13. Phagocytosis and/or adherence of non-commensal oral bacteria strains following opsonisation by BPIFA2. Saliva was used as a positive control and PBS was a negative control. *E. coli* (A) and *P.aeruginosa* (B) were coated with human BPIFA2 and animal BPIFA2 as described in the X-axis. The Y-axis represents the percentage of bacteria that were phagocytosed and/or attached to the surface of MM6 cells. Data presented as mean ± SEM (N=3).

The results of Figure 5.13 A and B show that both *E. coli* and *P. aeruginosa* were opsonised by all human and animal BPIFA2 proteins. In Figure 5.12, it was demonstrated that all human and animal BPIFA2 proteins increased the adherence/phagocytosis of *S. aureus,* a respiratory pathogen, in a similar manner to saliva while in comparison BPIFA2 coating increased adherence/phagocytosis of another respiratory pathogen, *P. aeruginosa,* to a relatively lesser extent and with a greater number of non-attached bacteria being found in the washes. It is possible that *P. aeruginosa* exhibits swarming motility, which means they can move across the agar surface and form a thin layer of cells rather than distinct colonies. This spreading can make it hard to count individual colonies since they merge into a film or a "swarm" rather than remaining isolated.

5.5 Discussion

The purpose of the investigations described in this chapter was to fill in the research gaps in the existing literature concerning the antibacterial capabilities and potential role of BPIFA2 in host defence. Alongside the pull-down binding assay, agglutination, biofilm formation, protein-lipid interaction and opsonisation assays have been carried out, with diverse responses being observed. Bacterial pull-down is an effective tool for investigating interactions between bacterial and host proteins. It has been suggested that BPIFA2 functions as an antibacterial agent based on its predicted structure and similarity to known LPS binding proteins, BPI and LBP (Bingle and Craven, 2002). The synthetic BPIFA2 GL13 peptide aligns with the active regions of the BPI protein which, along with BPIFA1 has been shown to function in host defence by causing agglutination and inhibiting biofilm formation (Gorr *et al.*, 2008; Gakhar *et al.*, 2010). Agglutination is triggered by glycoproteins such as salivary agglutinin and

mucins, which can lead to bacterial clumping together of microbial cells and thus inhibiting biofilm formation (Frenkel and Ribbeck, 2015). Furthermore, biofilm formation plays an important role in the initial stages of periodontal disease by triggering an innate immune inflammatory response (Frenkel and Ribbeck, 2015). A lipid binding assay was also selected for this study as the predicted structural similarities between BPI and BPIF proteins led to the hypothesis that BPIFA2 would have the ability to attach to bind to lipid-like compounds in order to function as an innate immune protein (Bingle and Craven, 2002). Opsonisation through BPIFA2 was investigated for the first time in this study, as phagocytosis is essential for the innate immune response and this assay is able to provide direct observations of macrophage cells showing phagocytic activity towards the protein-coated bacterial strains. Therefore, the production of recombinant BPIFA2 protein from both human and animal sources as explained in chapter 4, allowed for the assessment of the biological functions of BPIFA2 using *in vitro* functional assays.

To date the literature has mainly focused on investigating the function of BPIF proteins by examining their impact on non-oral commensal bacteria such as *E. coli* and *P. aeruginosa,* including studies on the function of BPIFA2. For instance, previous studies have demonstrated that BPIF proteins exhibit surfactant-like characteristics, suggesting they are involved in reducing surface tension in the upper respiratory tract. Moreover, it has been demonstrated that BPIF inhibits the growth of *P. aeruginosa* biofilms, indicating that one potential role of the BPIF protein is to impede persistent infections on mucosal surfaces (Gakhar *et al.*, 2010). A limitation of this study is that it focused on the BPIFA1 protein and has not considered the function of other family members.

The peptide GL13K obtained from the BPIFA2 sequence (141-153) has been shown to have antibacterial properties towards *P. aeruginosa* and *E. coli* (Gorr *et al.*, 2008, 2011). As this is a small peptide of human BPIFA2, it could be argued that the functional results are inclusive as the folding and conformation of the protein could greatly affect its functionality.

In the human oral cavity, the normal microflora comprises multiple bacteria strains, which can be found on any surfaces including those covered by epithelial cells; it has been proposed that a minimum of 1000 species exist. Oral bacteria are a diverse population and include Gram- negative *cocci* and *bacilli*, such as *Neisseria* (Kilian *et al.*, 1996) and *Fusobacterium*; Gram- positive Streptococcus, such as *S. mitis*, *S. mutans* and *S. sanguis* and oral pathogens such as *P. gingivalis* and *T. forsythia* (Aas *et al.*, 2005; Verma, Garg and Dubey, 2018).

Previously, BPIFA2 has been shown to bind lipopolysaccharide, an important component of the bacterial cell surface as a phosphorylated glycolipid specifically produced by Gram-negative bacteria (Abdolhosseini *et al.*, 2012). In this study, both WT and deglycosylated forms of human BPIFA2 showed positive binding to the LPS of *E. coli* which is in agreement with a previously published study that demonstrated that the BPIFA2 peptide (GL13K) was able to bind to *E. coli* and had bactericidal activity (Gorr *et al.*, 2011). The agglutination results further confirmed the hypothesis that BPIF proteins, due to their shared structural characteristics, exhibit similar antibacterial activity to that of BPI and LBP, by inducing agglutination of *E. coli*. In contrast, previously investigated data from our lab showed no binding to *E. coli* but this may have been due to the fact that ethanol precipitation was used as a purification method and the BPIFA2 was contaminated by an antibacterial protein, cystatin

(Lunn, 2014).

Previous studies investigating the interactions with, and effects on, *S. aureus* by BPIF proteins are limited. BPIFA1 is a close family member of BPIFA2 and has been to demonstrated to play an important role in the regulation of the flow of fluids on airway surfaces and also to provide defence against bacterial infections such as *P. aeruginosa* and *S. aureus* (Schicht *et al.*, 2015).Tsou *et al* (2014) found that patients with chronic rhinosinusitis with nasal polyps (CRSwNP) who needed several sinus operations due to recurrent or persistent sinusitis exhibited significantly reduced expression of BPIFA1 compared to those who did not require repeated sinus surgery. The study also showed that when BPIFA1 was decreased, CRSwNP patients were more vulnerable to bacterial infections in the mucosal epithelia and submucosal glands through bacterial colonization, particularly *S. aureus* and *P. aeruginosa* due to a reduced immune response. This suggests that decreased expression of BPIFA1 may contribute to the infection of *S. aureus* and *P. aeruginosa* infections in individuals with CRSwNP (Tsou *et al.*, 2014). One possible explanation is that BPIFA2 and BPIFA1 are part of the BPIF protein family, suggesting they may exhibit functional similarities.

In this study, while the pull-down assay showed positive binding of human BPIFA2 to *S. aureus* only WT BPIFA2 was able to inhibit biofilm formation and none of the human or animal BPIFA2 proteins analysed were able to induce bacterial agglutination. It is possible that BPIFA2 interacted with lipid-like compounds such as lipoteichoic acid (LTA) from Gram- positive bacteria in the pull down assay because as with LPS of Gram-negative bacteria, LTAs from Gram-positive bacteria are pathogen-associated molecular patterns (PAMPs) recognized by the immune system (Akira, Uematsu and Takeuchi, 2006). As

described in previous study, BPI has binding ability for various LPS preparations derived from different bacteria, all of which have negatively charged anionic groups (Gazzano-Santoro *et al.*, 1995). BPI was also identified to have affinity to LTA due to an N-terminal lipid-binding pocket that has the capacity to incorporate hydrophobic acyl-chains (Beamer, Carroll and Eisenberg, 1997; Bülow *et al.*, 2018). Therefore, the positive binding of BPIFA2 with Gram-negative bacterial strains could be explained by sharing an N-terminal domain with BPI which indicates functional similarity.

P. aeruginosa is also a respiratory pathogen that, in this study, showed similar results to those found with *S. aureus*; positive binding of BPIFA2 and agglutination with both human and animal BPIFA2 proteins. These positive results are supported by a previous study in which a peptide (GL13K) derived from the functional sequence of human BPIFA2 was able to cause agglutination of *S.gordonii and P. aeruginosa* resulting in an inhibition of the interaction between LPS and LBP (Gorr *et al.*, 2011) and the peptide inhibited the spread of *P. aeruginosa* infection in a lettuce leaf model, indicating that GL13K exhibits *in vivo* activity (Gorr *et al.*, 2008).

Human BPIFA2 showed positive binding with *S. gordonii* and *S. mutans* but no agglutination activity was demonstrated with either bacterial strain. Previously, similar data from our lab indicated that in the presence of BPIFA2, the biofilm produced by *S. mutans* did not spread but noticeable clusters of bacteria developed suggesting that BPIFA2 was unable to induce agglutination (Lunn, 2014). As mentioned previously the active GL13K peptide from human BPIFA2 sequence (141-153) was able to induce agglutination against *S.gordonii* (Gorr *et al.*, 2008, 2011), however, a major difference between my study and those of Gorr et al (2008,

2011) was that I used Anti-FLAG purified recombinant BPIFA2 protein for functional assays rather than His-tag peptides from the active human BPIFA2 sequence and thus the folding and conformation of the recombinant protein may lead BPIFA2 to behave differently. Additionally, both the amount of protein or peptide used and the time allowed for agglutination to take place could have an impact on the process. It would be necessary to determine the time-dependent progression of bacterial agglutination with overnight bacterial cultures.

In this study, agglutination was assessed qualitatively rather than quantitatively, however, a more quantitative method should be considered as this would allow more accurate determination of the ability of the whole BPIFA2 protein to cause bacterial agglutination. For example, previous studies have shown that bacteria subjected to crystal violet staining after overnight incubation in a 96-well plate, followed by the careful removal of the supernatant, results in an accumulation of aggregated bacteria at the bottom of the plate. Images of the aggregates were then captured using a digital camera (Gorr *et al.*, 2008) and agglutination assessed by quantification of thecrystal violet staining instead of the “flat” or “tight patterns” we used. Recently, agglutination assays involving the use of particles such as latex beads that have been coated with proteins have been described (Kobayashi *et al.*, 2023). Following the assay, the quantity of particles within the agglutinated clumps can be determined using techniques such as flow cytometry or microscopy. It should be noted, however, that the latex beads could themselves induce bacterial agglutination. Also, photometric measurement was tried many times in my project without positive results. Optimisation of my method, through the use of a spectrophotometer would allow determination of the absorbance or turbidity of

the samples before and after agglutination as the degree of agglutination is inversely proportional to the absorbance or turbidity of the samples. Thus, the quantification of the extent of agglutination could be determined by comparing the absorbance or turbidity values of the samples before and after treatment.

This study demonstrated significant inhibition of *S. mutans* biofilm formation by all animal BPFIA2 proteins, and a significant number of the animal proteins also disrupted *S.gordonii* biofilm formation. Our preliminary data therefore suggests that BPIFA2 selectively interacts with, and controls activity of, bacteria but further study is required to determine any true relationship between the animal BPIFA2 proteins we chose to study and bacterial strains.

WT and deglycosylated forms of human BPIFA2 bound to all bacterial strains tested and thus it appears that glycosylation does not affect direct interaction and binding of BPIFA2 to bacteria. The results from my bacterial binding assays also showed that BPIFA2 proteins remained in the supernatant samples suggesting that a significant amount of protein had not bound to the bacteria. After the purification of recombinant protein, it was attempted to quantify the concentration of purified BPIFA2 by ELISA and BCA assays as described in chapter 4; section 4.6; figure 4.22. The ELISA assay was unable to determine the precise concentration of purified BPIFA2, however, quantification via the BCA assay was possible when the same amount of each recombinant protein was used for concentration detection. To date the exact physiological concentration of BPIFA2 has not been determined and it is entirely possible that significantly more recombinant BPIFA2 was used in this study than is found in the normal oral cavity.

Further experiments could determine and quantify the binding capacity of BPIFA2 with specific bacterial strains. Agglutination of *P. aeruginosa* and *E. coli*. by armadillo, mouse, macaque, squirrel monkey and dog BPIFA2 was similar to that seen with human BPIFA2 but the results seen in the biofilm formation experiments were less consistent. The results of bacterial binding assays showed that human BPIFA2 binds to oral commensal, non-commensal and respiratory bacteria suggesting that BPIFA2 might be a "sticky" protein that is adhering to bacteria through a non-specific interaction. In the future, it would be beneficial to investigate similar pull down assays using the animal BPIFA2 proteins where we would expect to see some positive binding in the incubation with macaque, squirrel monkey and dog BPIFA2, particularly since macaque and squirrel monkey are primates that are evolutionarily closer to human BPIFA2 than mouse and armadillo (chapter 3; figure 3.1). No or reduced binding might be expected in the interaction of bacterial species with mouse BPIFA2 as this has a shorter amino acid sequence than other species including humans (chapter 3; figure 3.5). Furthermore, there is a noticeable deletion at the initial sequence of the N-terminal, indicating a potential misfolding of the protein and consequent loss of function in the mouse. We would also expect to see reduced binding between bacteria and armadillo BPIFA2 due to the extremely divergent evolutionary relationship (chapter 3; figure 3.1) compared to humans and other mammals used in this study.

Further investigations, in which BPIFA2 proteins are incubated overnight in 96-well plates prior to the addition of the bacteria would allow us to examine the ability of BPIFA2 to prevent biofilm development rather than disrupting formed biofilms; determining whether BPIFA2, as a salivary protein, has a true host defence function. In the oral cavity, most

biofilms are composed of multiple species of microorganisms rather than being dominated by a single species (Elias and Banin, 2012). These multispecies biofilms are commonly found on various surfaces in the mouth, including teeth, gums, tongue, and oral mucosa (Shang *et al.*, 2018). Thus it would be of value to carry out multispecies bacterial biofilm formation and disruption assays with *S. gordonii* and *S.mutans* as oral commensal strains, *S. mutans* as an oral commensal strain or *S.aureus* as a respiratory commensal.

As mentioned in chapter 1 (section 1.2) BPIFA2 is a member of the TULIP superfamily, which also contains BPI, LBP and other PLUNC proteins. These proteins play a role in either the host defence mechanism against bacteria by binding to LPS, or in the transfer of lipids between lipoprotein particles. BPIFA2 was described as an N-glycosylation protein (Ramachandran *et al.*, 2006), and our saliva analysis (chapter 4; section 4.4) also demonstrated the presence of multiple BPIFA2 isoforms in the oral cavity. Previous data also showed that BPIFA2 may be regulated differentially in different salivary gland tissues, where it may play a role in the innate immune response (Bingle *et al.*, 2009). The function of this differential glycosylation of BPIFA2, however, is still uncertain. LPS consists of a lipid A region, which is responsible for the bacterial endotoxin action. Glycosylation can alter the affinity and specificity of a protein for the bacterial lipids due to the addition of carbohydrate groups that can change the physical and chemical properties of the protein surface, which plays a significant role in biological functions. Currently, there is little information available regarding the interaction between BPIFA2 and lipids. BPIFA2B (BSP30B) is a BPIFA2 paralogue expressed in bovine saliva that was discovered to bind to lipids present on the surface of bacteria, as determined by the crystal structure and its ability to bind to oleic acid

(Zhang *et al.*, 2019).

This study found that WT BPIFA2 bound to PIP lipids (PtdIns(4)P, PtdIns (4,5)P2 and PtdIns (3,4,5)P3) but none of the deglycosylated human proteins, 124, 132 nor 124/132 BPIFA2 bound to any of the lipids; this suggests that there are no lipid-protein interactions following deglycosylation and that the lipid-protein interaction is dependent upon glycosylation sites in the wild-type recombinant BPIFA2 protein. However, it is interesting to observe that removal of only one glycosylation site disrupted lipid binding of the protein potentially indicating that the BPIFA2-lipid interactions need more than one N-glycosylation site to be *in situ*.

Recombinant BPIFA2 from macaques interacted with the same lipids as the WT human protein (PtdIns(4)P, PtdIns (4,5)P2, PtdIns (3,4,5)P3) but no binding was observed with any of the other animal species. A potential reason for this could be that macaque BPIFA2 has three N- glycosylation sites (124, 165 and 188; chapter 3; figure 3.5) the most glycosylation sites in the six species, including humans, studied in my project and interestingly the macaque, a classical type of primate, is evolutionary closer to humans and one of the glycosylation sites is shared with the human protein. It would be of interest, in future, to expand our studies to include chimpanzees and gorillas. Chimpanzees have two N-glycosylation sites at exactly the same positions, 124 and 132, as those in humans and gorillas have two N-glycosylation sites but at slightly different sites, 133 and 141. Previous data has shown that both chimpanzees and gorillas have higher BPIFA2 expression in saliva compared to humans (Thamadilok *et al.*, 2020) and whilst we might anticipate similar lipid binding of chimpanzee BPIFA2 it would be of great interest to mutate the glycosylation sites as used in this study (chapter 2; section 2.3).

No lipid binding was noted with mouse, dog, armadillo, squirrel monkey BPIFA2. There is no N-glycosylation site in the protein sequence of the squirrel monkey armadillo or dog, all of which are relatively divergent mammals from humans (chapter 3; figure 3.1). While there is a site located at position 215 in the mouse, the level of sequence similarity is significantly lower compared to other mammals. Additionally, there is a deletion of an amino acid at the beginning of the sequence (chapter 3; figure 3.5.), which results in the loss or alteration of function. It is possible that the position of the N-glycosylation site is not significant, especially when the protein structures are comparable to each other, as shown by the predicted protein structures generated by Alpha Fold (chapter 3; figure 3.6).

The PIP lipids, including PtdIns(4)P, PtdIns (4,5)P2 and PtdIns (3,4,5)P3, are located in plasma membranes. A previous study in rats suggested that binding to PIP lipids could assist in localizing the BPIFA2 protein to granules of the acinar cells for release into the salivary ducts (Venkatesh *et al.*, 2011); this suggests the functional importance of binding to membrane lipids is negligible.

Very few previous studies have investigated the role of BPIF proteins in the opsonisation of bacteria and presentation to macrophages. A previous study indicated that BPIFA2-like (BSP 30A and B) proteins in cattle suppress the activity of bovine saliva by a mechanism that is not dependent on LPS opsonisation (Haigh *et al.*, 2008). Opsonisation is important as a first line of defence against bacterial infections. Previously published papers have shown that the N-terminal domain of BPI, with a positive charge, will bind to the negatively charged lipid A part of LPS and the C-terminal domain of BPI facilitates the binding of bacteria to neutrophils (or macrophages), which demonstrates both its bactericidal and LPS-neutralizing actions

(Iovine, Elsbach and Weiss, 1997). The salivary protein, agglutinin, is a highly glycosylated protein expressed by serous acini and it has been suggested that it enhances phagocytosis by neutrophils and macrophages through recognising and binding mannose molecules from lectin (Leito *et a.*, 2011).

In this study, we investigated the same oral bacteria and pathogens as in our previous functional assays, namely *P. aeruginosa, E. coli*, *S. gordonii, S. mutans* and *S. aureus*. MonoMac 6 cells are immortalised human macrophage cells commonly used in immunological research and were cultured under appropriate conditions (chapter 2; section 2.1.1). The assay undertaken allowed us to investigate the adherence of bacteria to the MM6 cells and also allowed us to assume that some of these bacteria would have been phagocytosed by the cells. It was not able to say definitively that all bacteria interacting with the MM6 cells were in fact phagocytosed but in future studies antibiotic treatment would remove adherent, but non-ingested cells would provide this data. Nevertheless, it was able to demonstrate differential adherence/phagocytosis of BPIFA2-coated bacteria by MM6 cells; the differential adherence could reflect variations in proteins expressed on the bacterial surface or modifications that influence their recognition and uptake by phagocytic cells.

The opsonisation method used in this study is novel in the BPIFA2 protein field and the results give an informative direction for functional analysis as they indicate that BPIFA2 may play an important role in modulating host-pathogen interactions and immune responses. However, we did not determine whether the BPIFA2 caused the bacteria to simply attach to the MM6 cells or had been fully phagocytosed by the cells. Further investigations including detailed characterisation of the protein coatings and their effects on bacterial adherence,

internalisation, and intracellular fate within MM6 cells, are warranted to elucidate the underlying mechanisms and biological significance of these observations. Further investigations could also include studies which would allow BPIFA2 pre-opsonisation of MM6 whereby the MM6 cells are coated with BPIFA2, the bacteria are introduced and opsonisation is further investigated.

The findings in this chapter highlight the functional roles of BPIFA2, including its antimicrobial, biofilm-modulating, and immunomodulatory properties. Comparative studies of BPIFA2 proteins from various animal species with human variants revealed both evolutionary conservation and species-specific adaptations. Investigations of BPIFA2's lipid-binding ability helped us understand its association with glycosylation sites, as well as its structural and functional complexity. Furthermore, the exploration of BPIFA2’s interactions between macrophages and oral bacteria has provided valuable insights into its ability to modulate microbial behavior and inflammatory responses, emphasizing its relevance in oral health.

Chapter 6: Finial discussion

The overarching aim of the work described in this thesis was to determine the function of BPIFA2. Comparative bioinformatics analyses were performed to investigate BPIFA2's evolutionary conservation and divergence across multiple mammalian species. Variations in expression levels and glycosylation patterns were analyzed, contributing to an understanding of its structural and functional diversity. Simultaneously, recombinant BPIFA2 from multiple species was synthesized using a mammalian expression cell line, enabling downstream experimental studies. Functional studies were conducted to elucidate BPIFA2's interactions with lipids, microbes, and macrophages. Lipid-binding assays evaluated its potential roles in lipid transport or signaling. Microbial interaction studies revealed species-specific effects on bacterial agglutination, biofilm formation, and antimicrobial activity. Interactions with macrophages provided insights into BPIFA2's immunomodulatory roles, including potential impacts on inflammatory responses. These thorough methodologies combined evolutionary, biochemical, and functional investigations, enhancing the comprehension of BPIFA2's activities across species and its implications for antibacterial properties and inflammation within the human immune system.

BPIF-containing (PLUNC) proteins display notable sequence diversity and distinct biological roles have been suggested for various members of the family (Abdolhosseini *et al.*, 2012). BPIFA2, a protein found in both major and minor salivary glands and secreted into saliva, has been hypothesised to play a role in innate immune defence. One of the objectives in this study was to determine the nature of expression of BPIFA2 in human saliva and to determine genetic variants using bioinformatics databases and tools. Sequence alignments, phylogenetic

trees, and conserved domain analyses highlighted structural similarities and species-specific adaptations. Additionally, human genetic variant analysis identified polymorphisms in BPIFA2, providing insights into potential variations in its function or expression among individuals.

BPIFA2 expression was studied using unstimulated saliva samples collected from eighteen healthy individuals in which variables were controlled by standardising the conditions, such as always collecting samples at the same time of day, ensuring volunteers had not eaten or drunk anything other than water for at least one hour prior to collection. This would allow us to ensure reliability and validity of samples and results. The results we obtained indicate that the expression of BPIFA2 is not affected by gender or age, however, it was notable that volunteer 10 (55-year-old, male) showed no BPIFA2 expression (chapter 4; figure 4.1 and figure 4.4), even after conducting a second validation by dot blotting (chapter 4; figure 4.3). Further investigations are required to evaluate the level of BPIFA2 expression in both stimulated and unstimulated saliva in volunteers, particularly volunteer 10.

Bioinformatic investigations (section 3.3.5) into human genetic variations have revealed that certain ethnic groups, such as the African population, show different levels of expression and/or a complete lack of expression of the BPIFA2 gene, leading to the absence of BPIFA2 expression. A possible reason for this could be the presence of genetic variations in the start codon that affect the functioning of proteins and are unique to specific ethnic groups.

The difference in the expression level of BPIFA2 between humans and great apes can be attributed to evolutionary factors as well as alterations in diet (see chapter 1; section 1.7.3;

figure 1.10). Similar observations in the *BPIFA4* gene which seems to have undergone several evolutionary changes unusual to its lineage, and this may have led to the development of a distinct function. *BPIFA4* is a functional gene in great apes and horses, however, it has been lost in human and aquatic mammals (chapter 3; section 3.3.3). The *BPIFA4* gene may have undergone loss as a result of an evolutionary trade-off, wherein the loss of one function provides some other advantageous trait or is associated with the loss of a function that is no longer required. In ruminants, such as cattle and sheep, there are multiple *Bpifa2*-associated genes indicating a proliferation of the gene family. An important discovery is the identification of three pseudogenes due to the absence of exons in cattle gene expression: *LOC112449237*, *LOC132346950* (formerly known as *BPIFA2A*), and *BPIFA2D* (chapter 3; section 3.3.4.1). Another significant finding in cattle, from a transcriptomic study using RNAseq, is the identification of a newly characterised gene, referred to as *LOC112441508*.The gene could have developed specialisation or acquired new functions, distinct from its initial function, through the presence of numerous duplicates. This process is known as neofunctionalization and can result in an increase in the functional diversity of the gene family. The sheep, a traditional ruminant, was discovered to have multiple variants of *Bpifa2*, nine functional genes and three non-functional pseudogenes. Currently, there is a shortage of understanding regarding the sheep genes and additional investigation is required to completely clarify the role of BPIFA2 proteins in the functioning of ruminant digestion and its possible consequences for rumination and general digestive well-being in these animals as the dietary habits of ruminant are so distinct from normal mammals or cetaceans. Cetaceans are a specific species of animal that lacks salivary glands and does not rely on saliva for digesting (Huelsmann *et al.*, 2019) . The cetaceans lack the expression of BPIFA2 due to exon

deletion.

While the lack of research into the absence of BPIFA2 in African populations, and the presence of numerous BPIFA2-related proteins in ruminants, it is apparent that the functional significance of gene differences between species remains unknown.

Recombinant proteins are highly valuable and extensively employed for investigating protein function. In our laboratory, we previously purified recombinant WFDC (Armes, 2019) and used similar techniques in this study to synthesise and analyse recombinant BPIFA2 from a number of animal species including armadillo, mouse, macaque, squirrel monkey, dog alongside human wild type and mutated forms of the protein. Purified recombinant proteins facilitated the exploration of conserved and unique features across species. In order to accomplish the best possible synthesis of targeted proteins, it is important to carefully plan and design cloning and expression systems that have been modified to match the unique requirements of the final products. This study aimed to produce constructs that lack one or both glycosylation sites in order to compare the functionality of the "wild-type" BPIFA2 protein with non-glycosylated proteins. The purified recombinant protein from human and animal constructs was then utilised to determine the true function of BPIFA2.

The oral cavity is a complicated environment that produces around 1500mL of saliva each day to maintain health (Pittman *et al*., 2023). Also, the oral cavity hosts a diverse microbiome, consisting of bacteria, fungi, viruses, and other microorganisms (Elias and Banin, 2012). This microbiome plays a vital role in oral health, aiding in digestion and protecting against pathogens (Dong *et al.*, 2022). The oral mucosa and saliva contain immune cells and

immunological defence components that help in fighting infections (Hannig *et al.*, 2017). The antimicrobial action of purified BPIFA2 was evaluated against a range of bacterial strains including *P. aeruginosa, E. coli*, *S. gordonii, S. mutans* and *S. aureus* through the use of multiple assays, including bacterial binding, biofilm formation and agglutination. These methods demonstrate the binding of BPIFA2 to various bacteria, but it does not establish whether BPIFA2 binds directly to the bacteria or if it binds indirectly through other salivary constituents such as amylase or mucin, which then bind directly to the bacteria in an oral complex microenvironment. Additionally, my results suggest that salivary proteins, including BPIFA2, bind to bacteria, but it does not yet confirm absolutely whether the N-glycosylation sites of BPIFA2 are related to the bacterial binding ability. We were also not able to show conclusively whether BPIFA2 binds specifically or non-specifically to commensal and/or non-commensal bacteria.

A previous study demonstrated the successful purification of BPIFA2 through affinity chromatography and demonstrated that their pure BPIFA2 inhibited the growth of *P. aeruginosa* within the microgram concentration range, as demonstrated by the minimum inhibitory concentration (MIC) assay (Prokopovic *et al.*, 2014). Techniques such as size exclusion chromatography, ion exchange chromatography, ethanol/acetone and ammonium sulphate precipitation have proven ineffective in isolating functional, pure BPIFA2 protein from human saliva, probably due to continuing contamination and loss of function caused by the isolation procedure (Lunn, 2014). Nasal lavage fluid contains a wide range of proteins, including antibacterial proteins such as lysozyme, lactoferrin, and mucins (Ali, Wilson and Pearson, 2002), which have the potential to disrupt the binding of BPIFA1 to bacteria. The

research presented here provides evidence that the binding of BPIF-containing protein from complex secretions may occur through the involvement of other accessory proteins. BPIFA2 in whole saliva exhibited binding affinity towards all of the examined microorganisms. It would be interesting to examine the interaction between purified mucins, obtained from mucus using denaturation solvents (Fass and Thornton, 2023), and purified BPIFA2. Additionally, in our previous studies it was found that the most effective technique for purifying BPIFA2 from saliva was through the use of native gel electrophoresis combined with protein elution, but this methods leading to amylase contamination, which indicated a strong correlation between amylase and BPIFA2 (Lunn, 2014). Amylase is a salivary protein that has evolved in some species to play a role in nutrition. Previous studies have suggested that mice do not exhibit any noticeable immune characteristics following BPIFA2 gene deletion. However, the amount of saliva produced upon pilocarpine stimulation was the same in both WT and KO mice and there was an important distinction in the amount of amylase present in KO mice, which was 60% higher than WT mice (Nandula *et al.*, 2020). Further investigations would allow us to determine if there is a true relationship between amylase and BPIFA2 production and also if the regulation of one gene is related to the other. It would also allow us to determine any close functional ties between BPIFA2 and amylase.

A further consideration might be that due to the anaerobic nature of bacteria such as *Porphyromonas gingivalis* (Gram-negative), it might not come into direct contact with salivary proteins such as BPIFA2; we have not yet determined any interaction between BPIFA2 and *P. gingivalis*. The colonisation of the oral cavity by *C. albican*s, a commensal yeast, can cause oral diseases such as candidiasis and stomatitis, particularly in older patients

with dentures . Studies have shown that BPIFA2 exhibits a strong attraction to hydrophobic silicone and may in fact improve the adherence of *Candida albicans* to silicone surfaces. (Holmes *et al.*, 2014). Expanding research to include the interactions of BPIFA2 with *C. albicans* and several bacterial species could provide significant insights into its wider biological functions.

In this study, we have demonstrated that BPIFA2 has the ability to bind to a range of bacteria, and to reduce biofilm formation and prevent agglutination of some of these species. We have not yet been able to demonstrate bacterial killing, which is a significant aspect of human host defence. Further investigation should be carried out to examine whether BPIFA2 has the ability to inhibit bacterial growth or induce bacterial death. It would be advantageous to specifically focus on both the interactions and the act of killing. Investigating the activity of BPIFA2, both in isolation and in conjunction with mucins and amylase, is essential for achieving any real host defence functions and to finally determine the true role of this protein in maintaining oral health.

From an evolutionary standpoint, humans and rodents have undergone an extremely long and divergent evolutionary process while divergence between humans and chimpanzees and gorillas occurred relatively recently in evolutionary terms. The complexity of our genes may be attributed to the complicated nature of evolution, as shown in the potential relationship with amylase. Therefore, we speculate that BPIFA2 functions through interactions with other salivary constituents such as mucin and amylase and potentially plays a role in sensing and nutrition, rather than host defence. The future plan will be optimizing recombinant production of BPIFA2 and mucin or amylase in the lab and developing scalable methods for purifying

these recombinant proteins without losing bioactivity. Investigate how recombinant BPIFA2 and mucin/amylase work together to inhibit biofilms and agglutinate bacteria.

In conclusion, this thesis comprehensively investigated the functional analysis of BPIFA2, providing insights into its structural, evolutionary and functional roles. The bioinformatic analysis in chapter 3 identified conserved and species-specific features across mammalian BPIFA2 sequences and revealed human genetic variants that may influence its expression or function. In chapter 4, BPIFA2 was identified in human saliva and examined for its varying expression and glycosylation patterns. And recombinant BPIFA2 was produced and studied across six species within humans, enabling a systematic comparative investigation. Further exploration of BPIFA2's functional properties was described in chapter 5. This chapter mainly revealed and highlighted BPIFA2's interactions with lipids, microbes, and macrophages in antimicrobial, biofilm-modulating, and immunomodulatory activities. Overall, these chapters achieve the initial aims of the thesis by combining bioinformatics, biochemical and functional methodologies to further the comprehension of BPIFA2's biological significance across various species and its prospective implications for oral and systemic health.

Bibliography

Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I., and Dewhirst, F. E. (2005) 'Defining the Normal Bacterial Flora of the Oral Cavity', *Journal of Clinical Microbiology*, 43(11), 5721–5732. doi: 10.1128/JCM.43.11.5721.

Abdolhosseini, M., Sotsky, J. B., Shelar, A. P., Joyce, P. B., and Gorr, S. U. (2012) 'Human parotid secretory protein is a lipopolysaccharide-binding protein: Identification of an anti-inflammatory peptide domain', *Molecular and Cellular Biochemistry*, 359(1–2), 1–8. doi: 10.1007/s11010-011-0991-2.

Abrahamson, S. L., Wu, H. M., Williams, R. E., Der, K., Ottah, N., Little, R., Gazzano-Santoro, H., Theofan, G., Bauer, R., Leigh., *et al* (1997) 'Biochemical characterization of recombinant fusions of lipopolysaccharide binding protein and bactericidal/permeability-increasing protein. Implications in biological activity', *Journal of Biological Chemistry*, 272(4), 2149–2155. doi: 10.1074/jbc.272.4.2149.

Aguirre, A., Levine, M. J., Cohen, R. E., and Tabak, L. A. (1987) 'Immunochemical quantitation of α-amylase and secretory IgA in parotid saliva from people of various ages', *Archives of Oral Biology*, 32(4), 297–301. doi: 10.1016/0003-9969(87)90024-0.

Akira, S., Uematsu, S., and Takeuchi, O. (2006) 'Pathogen recognition and innate immunity', *Cell*, 124(4), 783–801. doi: 10.1016/j.cell.2006.02.015.

Alexander, C., and Rietschel, E. T. (2001). *Bacterial lipopolysaccharides and innate immunity*. Journal of endotoxin research, 7(3), 167–202.

Alva, V., and Lupas, A. N. (2016) 'The TULIP superfamily of eukaryotic lipid-binding proteins as a mediator of lipid sensing and transport', *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*. Elsevier B.V., 1861(8), 913–923. doi: 10.1016/j.bbalip.2016.01.016.

Alves, D. B. M., Bingle, L., Bingle, C. D., Lourenco, S. V., Silva, A. A., Pereira, D. L., and Vargas, P. (2017) 'BPI-fold (BPIF) containing/plunc protein expression in human fetal major and minor salivary glands', *Brazilian Oral Research*, 31(0), 1–9. doi: 10.1590/1807-3107bor-2017.vol31.0006.

Al Habobe, H., Haverkort, E. B., Nazmi, K., Van Splunter, A. P., Pieters, R. H. H., and Bikker, F. J. (2024) 'The impact of saliva collection methods on measured salivary biomarker levels', *Clinica Chimica Acta*. Elsevier B.V., 552(November 2023), 117628. doi: 10.1016/j.cca.2023.117628.

Al-Hashimi I, Levine MJ (1989). 'Characterization of in vivo salivary-derived enamel pellicle', *Arch Oral Biology,* 34(4):289-295. doi:10.1016/0003-9969(89)90070-8.

Amano, A., Kataoka, K., Raj, P. A., Genco, R. J., and Shizukuishi, S. (1996) 'Binding sites of salivary statherin for Porphyromonas gingivalis recombinant fimbrillin', *Infection and Immunity*, 64(10), 4249–4254. doi: 10.1128/iai.64.10.4249-4254.1996.

Aps, J. K. M., and Martens, L. C. (2005) 'Review: The physiology of saliva and transfer of drugs into saliva', *Forensic Science International,* 150(2–3), 119–131. doi: 10.1016/j.forsciint.2004.10.026.

Armes, H. (2019) "An investigation into the function of WFDC", Available at : https://etheses.whiterose.ac.uk/

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., and *et al*. (2020) 'Progressive Cactus is a multiple-genome aligner for the thousand-genome era', *Nature*. Springer US, 587(7833), 246–251. doi: 10.1038/s41586- 020-2871-y.

Arslan, S. Y., Leung, K. P., and Wu, C. D. (2009) 'The effect of lactoferrin on oral bacterial attachment', *Oral Microbiology and Immunology*, 24(5), 411–416. doi: 10.1111/j.1399-302X.2009.00537.x.

Atkinson, M. E., and White, F. H. 1992. *Principles of anatomy and oral anatomy for dental students*, Edinburgh, Churchill Livingstone.

Balekjian, A. Y., Hoerman, K. C., and Berzinskas, V. J. (1969) 'Lysozyme of the human parotid gland secretion: Its purification and physicochemical properties', *Biochemical and Biophysical Research Communications*, 35(6), 887–894. doi: 10.1016/0006-291X(69)90707-4.

Ball, W. D., Mirels, L., and Hand, A. R. (2003) 'Psp and Smgb: a model for developmental and functional regulation in the rat major salivary glands', *Biochemical Society Transactions*, 31(4), 777–780. doi: 10.1042/bst0310777.

Bartlett, J. A., Gakhar, L., Penterman, J., Singh, P. K., Mallampalli, R. K., Porter, E., and McCray, P. B., Jr. (2011) 'PLUNC: a multifunctional surfactant of the airways.', *Biochemical Society transactions*, 39(4), 1012–1016. doi: 10.1042/BST0391012.

Beeley, J. G, Eason, R., and Snow, D. H. (1986) 'Isolation and characterization of latherin, a surface-active protein from horse sweat', *Biochemical Journal*, 235(3), 645–650. doi: 10.1042/bj2350645.

Beamer, L. J., Carroll, S. F., and Eisenberg, D. (1997) 'Crystal structure of human BPI and two bound phospholipids at 2.4 Angstrom resolution', *Science*. American Association for the Advancement of Science, 276(5320), 1857–1860. doi: 10.1126/science.276.5320.1857.

Beamer, L. J., Carroll, S. F., and Eisenberg, D. (2008) 'The BPI/LBP family of proteins: A structural analysis of conserved regions', *Protein Science*, 7(4), pp. 906–914. doi: 10.1002/pro.5560070408.

Bingle, C. D., and Bingle, L. (2000) 'Characterisation of the human plunc gene , a gene product with an upper airways and nasopharyngeal restricted expression pattern', *Biochimica et biophysica acta*, 1493(3), 363–367. doi: 10.1016/s0167-4781(00)00196-2.

Bingle, C. D., Bingle, L., and Craven, C. J. (2011) 'Distant cousins: genomic and sequence diversity within the BPI fold-containing (BPIF)/PLUNC protein family', *Biochemical Society Transactions*. Portland Press Ltd., 39(4), 961–965. doi: 10.1042/bst0390961.

Bingle, C. D., and Craven, C. J. (2002) 'PLUNC: A novel family of candidate host defence proteins expressed in the upper airways and nasopharynx', *Human Molecular Genetics*, 11(8), 937–943. doi: 10.1093/hmg/11.8.937.

Bingle, C. D., and Craven, C. J. (2004) 'Meet the relatives: a family of BPI- and LBP-related proteins.', *Trends in immunology*. Elsevier, 25(2), 53–5. doi: 10.1016/J.IT.2003.11.007.

Bingle, C. D., and Gorr, S. U. (2004) 'Host defense in oral and airway epithelia: Chromosome 20 contributes a new protein family', *International Journal of Biochemistry and Cell Biology*, 36(11), 2144–2152. doi: 10.1016/j.biocel.2004.05.002.

Bingle, C. D., LeClair, E. E., Havard, S., Bingle, L., Gillingham, P., and Craven, C. J. (2004) 'Phylogenetic and evolutionary analysis of the PLUNC gene family', *Protein Science*, 13(2), 422–430. doi: 10.1110/ps.03332704.

Bingle, C. D., Seal, R. L., and Craven, C. J. (2011) 'Systematic nomenclature for the PLUNC/PSP/BSP30/SMGB proteins as a subfamily of the BPI fold-containing superfamily', *Biochemical Society Transactions*, 39(4), 977–983. doi: 10.1042/BST0390977.

Bingle, C. D., Wilson, K., Lunn, H., Barnes, F. A., High, A. S., Wallace, W. A., Rassl, D., Campos, M. A., Ribeiro, M., and Bingle, L. (2010) 'Human LPLUNC1 is a secreted product of goblet cells and minor glands of the respiratory and upper aerodigestive tracts', *Histochemistry and Cell Biology*, 133(5), 505–515. doi: 10.1007/s00418-010-0683-0.

Bingle, L., Barnes, F. A., Lunn, H., Musa, M., Webster, S., Douglas, C. W., Cross, S. S., High, A. S., and Bingle, C. D. (2009) 'Characterisation and expression of SPLUNC2, the human orthologue of rodent parotid secretory protein', *Histochemistry and Cell Biology*, 132(3), 339– 349. doi: 10.1007/s00418-009-0610-4.

Bingle, L., and Bingle, C. D. (2011) 'Distribution of human PLUNC/BPI fold-containing (BPIF) proteins', *Biochemical Society Transactions*. 39(4), 1023–1027. doi: 10.1042/bst0391023.

Bingle, L., Cross, S. S., High, A. S., Wallace, W. A., Devine, D. A., Havard, S., Campos, M. A., and Bingle, C. D. (2005) 'SPLUNC1 (PLUNC) is expressed in glandular tissues of the respiratory tract and in lung tumours with a glandular phenotype', 205(4), 491–497. doi: 10.1002/path.1726.

Borges, M. C., Sesso, M. L., Roberti, L. R., de Menezes Oliveira, M. A., Nogueira, R. D., Geraldo-Martins, V. R., and Ferriani, V. P. (2015) 'Salivary antibody response to streptococci in preterm and full term children: A prospective study', *Archives of Oral Biology*, 60(1), 116–125. doi: 10.1016/j.archoralbio.2014.08.003.

Bozorgi, C., Holleufer, C., and Wendin, K. (2020) 'Saliva secretion and swallowing—The impact of different types of food and drink on subsequent intake', *Nutrients*, 12(1), 256. doi: 10.3390/nu12010256

Bradway, S. D., Bergey, E. J., Jones, P. C., and Levine, M. J. (1989) 'Oral mucosal pellicle. Adsorption and transpeptidation of salivary components to buccal epithelial cells', *Biochemical Journal*, 261(3), 887–896. doi: 10.1042/bj2610887.

Brandtzaeg, P. (2013) 'Secretory immunity with special reference to the oral cavity', *Journal of Oral Microbiology*, 5(2013), 1–25. doi: 10.3402/jom.v5i0.20401.

Brandtzaeg, P., and Prydz, H. (1984) 'Direct evidence for an integrated function of J chain and secretory component in epithelial transport of immunoglobulins', *Nature*, 311(5981), 71–73. doi: 10.1038/311071a0.

Bruno, L. S., Li, X., Wang, L., Soares, R. V., Siqueira, C. C., Oppenheim, F. G., Troxler, R. F., and Offner, G. D. (2005) 'Two-hybrid analysis of human salivary mucin MUC7 interactions', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1746(1), 65–72. doi: 10.1016/j.bbamcr.2005.08.007.

Bülow, S., Ederer, K. U., Holzinger, J. M., Zeller, L., Werner, M., Toelge, M., Pfab, C., Hirsch, S., Göpferich, F., Hiergeist, A., *et al* (2024) 'Bactericidal/permeability-increasing protein instructs dendritic cells to elicit Th22 cell response', *Cell Reports*, 43(3). doi: 10.1016/j.celrep.2024.113929.

Bülow, S., Zeller, L., Werner, M., Toelge, M., Holzinger, J., Entzian, C., Schubert, T., Waldow, F., Gisch, N., Hammerschmidt, S., *et al* (2018) 'Bactericidal/Permeability-Increasing Protein Is an Enhancer of Bacterial Lipoprotein Recognition', *Frontiers in immunology*, 9, 2768. doi: 10.3389/fimmu.2018.02768.

Campos, M. A., Abreu, A. R., Nlend, M. C., Cobas, M. A., Conner, G. E., and Whitney, P. L. (2004) 'Purification and Characterization of PLUNC from Human Tracheobronchial Secretions', *American Journal of Respiratory Cell and Molecular Biology*, 30(2), 184–192. doi: 10.1165/rcmb.2003-0142OC.

Carneiro, L. G., Venuleo, C., Oppenheim, F. G., and Salih, E. (2012) 'Proteome data set of human gingival crevicular fluid from healthy periodontium sites by multidimensional protein separation and mass spectrometry', *Journal of Periodontal Research*, 47(2), 248–262. doi: 10.1111/j.1600-0765.2011.01429.x.

Chaudhury, N. M., Shirlaw, P., Pramanik, R., Carpenter, G. H., and Proctor, G. B. (2015) 'Changes in saliva rheological properties and mucin glycosylation in dry mouth', *Journal of Dental Research*, 94(12),1660–1667. doi: 10.1177/0022034515609070.

Chen, S., Francioli, L. C., Goodrich, J. K., Collins, R. L., Kanai, M., Wang, Q., Alföldi, J., Watts, N. A., Vittal, C., Gauthier, L. D., *et al* (2024) 'A genomic mutational constraint map using variation in 76,156 human genomes', *Nature*, 625(7993), 625(7993), 92–100.doi: 10.1038/s41586-023-06045-0.

Chi, Y., Zhang, Q., Qin, Z., Bai, J., Yan, J., Liu, C., and Li, B. (2024) 'Molecular pathology assists the diagnosis of lymphoepithelial sialadenitis, Sjögren's syndrome and extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue', *Journal of Dental Sciences*. Association for Dental Sciences of the Republic of China, 19(1), 130–138. doi: 10.1016/j.jds.2023.05.018.

Chu, H. W., Thaikoottathil, J., Rino, J. G., Zhang, G., Wu, Q., Moss, T., Refaeli, Y., Bowler, R., Wenzel, S. E., Chen, Z., *et al* (2007) 'Function and Regulation of SPLUNC1 Protein in Mycoplasma Infection and Allergic Inflammation', *The Journal of Immunology*, 79(6), 3995–4002. doi: 10.4049/jimmunol.179.6.3995.

Clarke, R. T. J., and Reid, C. S. W. (1974) 'Foamy Bloat of Cattle. A Review', *Journal of Dairy Science*. Elsevier, 57(7), 753–785. doi: 10.3168/jds.S0022-0302(74)84964-7.

Contreras-Aguilar, M. D., Tecles, F., Martínez-Subiela, S., Escribano, D., Bernal, L. J., and Cerón, J. J. (2017). *Detection and measurement of alpha-amylase in canine saliva and changes after an experimentally induced sympathetic activation*. BMC veterinary research, 13(1), 266. doi: 10.1186/s12917-017-1191-4

Cooper, A., Vance, S. J., Smith, B. O., and Kennedy, M. W. (2017) 'Frog foams and natural protein surfactants', *Colloids and Surfaces A: Physicochemical and Engineering Aspects*. Elsevier B.V., 534, 120–129. doi: 10.1016/j.colsurfa.2017.01.049.

Dawes, C. (1972) 'Circadian rhythms in human salivary flow rate and composition', *The Journal of Physiology*, 220(3), 529–545. doi: 10.1113/jphysiol.1972.sp009721.

Da Silva, A. A., Bingle, L., Speight, P. M., Bingle, C. D., Mauad, T., da Silva, L. F., and Vargas, P. A. (2011) 'PLUNC protein expression in major salivary glands of HIV-infected patients', *Oral Diseases*, 17(3), 258–264. doi: 10.1111/j.1601-0825.2010.01733.x.

De Almeida, P.delV., Grégio, A. M., Machado, M. A., de Lima, A. A., and Azevedo, L. R. (2008) '*Saliva Composition and Functions: A Comprehensive Review*', The journal of contemporary dental practice, 9(3), 9(3), 72–80.

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W. H., Lakshmanan, A., and Wade, W. G. (2010) 'The human oral microbiome', *Journal of Bacteriology*, 192(19), 5002–5017. doi: 10.1128/JB.00542-10.

De Smet, E. G., Seys, L. J., Verhamme, F. M., Vanaudenaerde, B. M., Brusselle, G. G., Bingle, C. D., and Bracke, K. R. (2018) 'Association of innate defense proteins BPIFA1 and BPIFB1 with disease severity in COPD', I*nternational Journal of COPD*, 13, 11–27. doi: 10.2147/COPD.S144136.

Dong, J., Li, W., Wang, Q., Chen, J., Zu, Y., Zhou, X., and Guo, Q. (2022). *Relationships Between Oral Microecosystem and Respiratory Diseases*. Frontiers in molecular biosciences, 8, 718222. doi: 10.3389/fmolb.2021.718222.

Dumont, J., Euwart, D., Mei, B., Estes, S., and Kshirsagar, R. (2016). Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical reviews in biotechnology*, 36(6), 1110–1122. https://doi.org/10.3109/07388551.2015.1084266

Eberhard, J., Drosos, Z., Tiemann, M., Jepsen, S., and Schröder, J. M. (2006) 'Immunolocalization of Lactoferrin in Healthy and Inflamed Gingival Tissues', *Journal of Periodontology*, 77(3), 472–478. doi: 10.1902/jop.2006.050186.

Eckert, J. K., Kim, Y. J., Kim, J. I., Gürtler, K., Oh, D. Y., Sur, S., Lundvall, L., Hamann, L., van der Ploeg, A., Pickkers, P., *et al*. (2013) 'The crystal structure of lipopolysaccharide binding protein reveals the location of a frequent mutation that impairs innate immunity', *Immunity*, 39(4), 647–660. doi: 10.1016/j.immuni.2013.09.005.

Egland, K. A., Vincent, J. J., Strausberg, R., Lee, B., and Pastan, I. (2003) 'Discovery of the breast cancer gene BASE using a molecular approach to enrich for genes encoding membrane and secreted proteins', *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), 1099–1104. doi: 10.1073/pnas.0337425100.

Elass-Rochard, E., Legrand, D., Salmon, V., Roseanu, A., Trif, M., Tobias, P. S., Mazurier, J., and Spik, G. (1998) 'Lactoferrin inhibits the endotoxin interaction with CD14 by competition with the lipopolysaccharide-binding protein', *Infection and Immunity*, 66(2), 486–491. doi: 10.1128/iai.66.2.486-491.1998.

Elsbach, P., Weiss, J., Franson, R. C., Beckerdite-Quagliata, S., Schneider, A., and Harris, L. (1979) 'Separation and purification of a potent bactericidal/permeability-increasing protein

and a closely associated phospholipase A 2 from rabbit polymorphonuclear leukocytes: Observations on their relationship', *Journal of Biological Chemistry*, 254(21), 11000–11009.

Elias, S., and Banin, E. (2012) 'Multi-species biofilms: Living with friendly neighbors', *FEMS Microbiology Reviews*, 36(5), 990–1004. doi: 10.1111/j.1574-6976.2012.00325.x.

Emes, R. D., Goodstadt, L., Winter, E. E., and Ponting, C. P. (2003) 'Comparison of the genomes of human and mouse lays the foundation of genome zoology', *Human Molecular Genetics*. 12(7), 701–709. doi: 10.1093/hmg/ddg078.

Frenkel, E. S., and Ribbeck, K. (2015) 'Salivary mucins in host defense and disease prevention', *Journal of Oral Microbiology*, 7, 29759. doi: 10.3402/jom.v7.29759.

Funda, D. P., Tucková, L., Farré, M. A., Iwase, T., Moro, I., and Tlaskalová-Hogenová, H. (2001). CD14 is expressed and released as soluble CD14 by human intestinal epithelial cells in vitro: lipopolysaccharide activation of epithelial cells revisited. *Infection and immunity*, 69(6), 3772–3781. doi: 10.1128/IAI.69.6.3772-3781.2001.

Gakhar, L., Bartlett, J. A., Penterman, J., Mizrachi, D., Singh, P. K., Mallampalli, R. K., Ramaswamy, S., and McCray, P. B., Jr (2010) 'PLUNC Is a Novel Airway Surfactant Protein with Anti-Biofilm Activity', *PLoS ONE*, 5(2), e9098. doi: 10.1371/journal.pone.0009098.

Gao, J., Ohlmeier, S., Nieminen, P., Toljamo, T., Tiitinen, S., Kanerva, T., Bingle, L., Araujo, B., Rönty, M., Höyhtyä, M., *et al* (2015) 'Elevated sputum BPIFB1 levels in smokers with chronic obstructive pulmonary disease: A longitudinal study', *American Journal of Physiology*
- *Lung Cellular and Molecular Physiology*, 309(1), L17–L26. doi: 10.1152/ajplung.00082.2015.

Garcia-Caballero, A., Rasmussen, J. E., Gaillard, E., Watson, M. J., Olsen, J. C., Donaldson, S. H., Stutts, M. J., and Tarran, R. (2009) 'SPLUNC1 regulates airway surface liquid volume by protecting ENaC from proteolytic cleavage', *Proceedings of the National Academy of Sciences of the United States of America*, 106(27), 11412–11417. doi: 10.1073/pnas.0903609106.

Gazzano-Santoro, H., Parent, J. B., Conlon, P. J., Kasler, H. G., Tsai, C. M., Lill-Elghanian, D. A., and Hollingsworth, R. I. (1995) 'Characterization of the structural elements in lipid A required for binding of a recombinant fragment of bactericidal/permeability-increasing protein rBPI23', *Infection and Immunity*, 63(6), 2201–2205. doi: 10.1128/iai.63.6.2201-2205.1995.

Geetha, C., Venkatesh, S. G., Bingle, L., Bingle, C. D., and Gorr, S. U. (2005). Design and validation of anti-inflammatory peptides from human parotid secretory protein. *Journal of dental research,* 84(2), 149–153.

Geetha, C., Venkatesh, S. G., Dunn, B. H., and Gorr, S. U. (2003). Expression and anti-

bacterial activity of human parotid secretory protein (PSP). *Biochemical Society transactions*, 31(Pt 4), 815–818. doi: 10.1042/bst0310815.

Ghafouri, B., Ståhlbom, B., Tagesson, C., and Lindahl, M. (2002) 'Newly identified proteins in human nasal lavage fluid from non-smokers and smokers using two-dimensional gel electrophoresis and peptide mass fingerprinting', *Proteomics*, 2(1), 112–120. doi: 10.1002/1615-9861(200201)2:1<112::AID-PROT112>3.0.CO;2-N.

Ghafouri, B., Kihlström, E., Tagesson, C., and Lindahl, M. (2004) 'PLUNC in human nasal lavage fluid: Multiple isoforms that bind to lipopolysaccharide', *Biochimica et Biophysica Acta - Proteins and Proteomics*. 1699(1-2), 57–63. doi: 10.1016/j.bbapap.2004.01.001.

Ghannam MG, Singh P. Anatomy, Head and Neck, Salivary Glands. [Updated 2023 May 29]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK538325/

Gibbins, H. L., Proctor, G. B., Yakubov, G. E., Wilson, S., and Carpenter, G. H. (2014) 'Concentration of salivary protective proteins within the bound oral mucosal pellicle', *Oral Diseases*, 20(7), 707–713. doi: 10.1111/odi.12194.

Gibbons, R. J., de Stoppelaar, J. D., and Harden, L. (1966) 'Lysozyme Insensitivity of Bacteria Indigenous to the Oral Cavity of Man', *Journal of Dental Research*, 45(3), 877–881. doi: 10.1177/00220345660450036201.

González-Arriagada, W. A., Ramos, L. M., Silva, A. A., Vargas, P. A., Coletta, R. D., Bingle, L., and Lopes, M. A. (2015) 'Salivary BPIFA1 (SPLUNC1) and BPIFA2 (SPLUNC2 A) are modified by head and neck cancer radiotherapy', *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*. 119(1), 48–58. doi: 10.1016/j.oooo.2014.09.026.

Gorr, S. U., Abdolhosseini, M., Shelar, A., and Sotsky, J. (2011) 'Dual host-defence functions of SPLUNC2/PSP and synthetic peptides derived from the protein', *Biochemical Society Transactions*, 39(4), 1028–1032. doi: 10.1042/BST0391028.

Gorr, S. U., Sotsky, J. B., Shelar, A. P., and Demuth, D. R. (2008) 'Design of bacteria-agglutinating peptides derived from parotid secretory protein, a member of the bactericidal/permeability increasing-like protein family', *Peptides*, 29(12), 2118–2127. doi: 10.1016/j.peptides.2008.09.019.

Haigh, B., Hood, K., Broadhurst, M., Medele, S., Callaghan, M., Smolenski, G., Dines, M., and Wheeler, T. (2008) 'The bovine salivary proteins BSP30a and BSP30b are independently expressed BPI-like proteins with anti-Pseudomonas activity', *Molecular Immunology,* 45(7), 1944–1951. doi: 10.1016/j.molimm.2007.10.032.

Hailman, E., Lichenstein, H. S., Wurfel, M. M., Miller, D. S., Johnson, D. A., Kelley, M., Busse, L. A., Zukowski, M. M., and Wright, S. D. (1994) 'Lipopolysaccharide (LPS)-binding protein accelerates the binding of LPS to CD14', *Journal of Experimental Medicine*, 179(1), 269–277. doi: 10.1084/jem.179.1.269.

Hailman, E., Albers, J. J., Wolfbauer, G., Tu, A. Y., and Wright, S. D. (1996) 'Neutralization and transfer of lipopolysaccharide by phospholipid transfer protein', *Journal of Biological Chemistry*, 271(21), 12172–12178. doi: 10.1074/jbc.271.21.12172.

Hannig, C., Attin, T., Hannig, M., Henze, E., Brinkmann, K., and Zech, R. (2004) 'Immobilisation and activity of human α-amylase in the acquired enamel pellicle', *Archives of Oral Biology*, 49(6), 469–475. doi: 10.1016/j.archoralbio.2004.01.005.

Hannig, C., Hannig, M., Kensche, A., and Carpenter, G. (2017) 'The mucosal pellicle – An underestimated factor in oral physiology', *Archives of Oral Biology*, 80, 144–152. doi: 10.1016/j.archoralbio.2017.04.001.

Harrison, B., and Zimmerman, S. B. (1984) 'Nucleic Acids Research Nucleic Acids Research', *Methods*, 12(21), 8235–8251. doi: 10.1007/978-3-642-04898-2_214.

Haverkos, H. W. (2003). Polymicrobial Diseases. *Emerging Infectious Diseases,* 9(1), 141. https://doi.org/10.3201/eid0901.020487.

He, Y., Zhou, G., Zhai, Y., Dong, X., Lv, L., He, F., and Yao, K. (2005) 'Association of PLUNC gene polymorphisms with susceptibility to nasopharyngeal carcinoma in a Chinese population', *Journal of Medical Genetics*, 42(2), 172–176. doi: 10.1136/jmg.2004.022616.

He L, Zhou S, Li W, Wang Q, Qi Z, Zhou P, Wang Z, Chen J, Li Y., and Lin Z (2022) 'BPIFA2 as a Novel Early Biomarker to Identify Fatal Radiation Injury After Radiation Exposure', *Dose- Response*, 20(1), 1–10. doi: 10.1177/15593258221086478.

Hirt, H., and Gorr, S. U. (2013) 'Antimicrobial peptide GL13K is effective in reducing biofilms of Pseudomonas aeruginosa', *Antimicrobial Agents and Chemotherapy*, 57(10), 4903–4910. doi: 10.1128/AAC.00311-13.

Holmes, A. R., Rodrigues, E., van der Wielen, P., Lyons, K. M., Haigh, B. J., Wheeler, T. T., Dawes, P. J., and Cannon, R. D. (2014) 'Adherence of Candida albicans to silicone is promoted by the human salivary protein SPLUNC2/PSP/BPIFA2', *Molecular Oral Microbiology*, 29(2), 90–98. doi: 10.1111/omi.12048.

Honore, P. M., De Bels, D., and Spapen, H. D. (2018) 'BPI fold-containing family a member 2 as a biomarker of acute kidney injury—close but no (clinical) cigar?', *Annals of Translational Medicine*, 6(10), 191–191. doi: 10.21037/atm.2018.03.13.

Huelsmann, M., Hecker, N., Springer, M. S., Gatesy, J., Sharma, V., and Hiller, M. (2019) 'Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations', *Science Advances*, 5(9), 1–12. doi: 10.1126/sciadv.aaw6671.

Humphrey, S. P., and Williamson, R. T. (2001) 'A review of saliva: Normal composition, flow, and function.pdf', *J Prosthet Dent.*, 85(2), 162–169.

Innocentini, L. M. A. R., Silva, A. A., Carvalho, M. A., Coletta, R. D., Corrêa, M. E. P., Bingle, L., Bingle, C. D., Vargas, P. A., and Lopes, M. A. (2022) 'Salivary BPIFA proteins are altered in patients undergoing hematopoietic cell transplantation', *Oral Diseases*, 28(4), 1279–1288. doi: 10.1111/odi.13832.

Iovine, N. M., Elsbach, P. and Weiss, J. (1997) 'An opsonic function of the neutrophil bactericidal/permeability-increasing protein depends on both its N- and C-terminal domains', *Proceedings of the National Academy of Sciences of the United States of America*, 94(20), 10973–10978. doi: 10.1073/pnas.94.20.10973.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., *et al* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), 583–589. doi: 10.1038/s41586-021-03819-2.

Khovidhunkit, W., Hachem, J. P., Medzihradszky, K. F., Duchateau, P. N., Shigenaga, J. K., Moser, A. H., Movsesyan, I., Naya-Vigne, J., Kane, J. P., Feingold, K. R., and Grunfeld, C. (2005) 'Parotid secretory protein is an HDL-associated protein with anticandidal activity', *American Journal of Physiology - Regulatory Integrative and Comparative Physiology*. 288(5), R1306–R1315. doi: 10.1152/ajpregu.00007.2004.

Kilian, M., Reinholdt, J., Lomholt, H., Poulsen, K., and Frandsen, E. V. (1996) 'Biological significance of IgA1 proteases in bacterial colonization and pathogenesis: Critical evaluation of experimental evidence', *Apmis*, 104(5), 21–338. doi: 10.1111/j.1699-0463.1996.tb00724.x.

Kishida, T., Thewissen, J., Hayakawa, T., Imai, H., and Agata, K. (2015) 'Aquatic adaptation and the evolution of smell and taste in whales', *Zoological Letters*, 1(1), 1–10. doi: 10.1186/s40851-014-0002-z.

Kobayashi, J., Matsuyama, S., Shirakura, M., Arita, T., Suzuki, Y., Asanuma, H., Watanabe, S., Hasegawa, H., and Nakamura, K. (2023). Use of the particle agglutination/particle agglutination inhibition test for antigenic analysis of SARS-CoV-2. Influenza and other respiratory viruses, 17(2), e13093. doi: 10.1111/irv.13093.

Kobayashi, H., Song, C., Ikei, H., Park, B. J., Kagawa, T., and Miyazaki, Y. (2017) 'Diurnal changes in distribution characteristics of salivary cortisol and immunoglobulin a concentrations', *International Journal of Environmental Research and Public Health*, 14(9). doi: 10.3390/ijerph14090987.

Komine, K., Kuroishi, T., Ozawa, A., Komine, Y., Minami, T., Shimauchi, H., and Sugawara, S. (2007) 'Cleaved inflammatory lactoferrin peptides in parotid saliva of periodontitis patients', *Molecular Immunology*, 44(7), 1498–1508. doi: 10.1016/j.molimm.2006.09.003.

Kopec, K. O., Alva, V., and Lupas, A. N. (2011) 'Bioinformatics of the TULIP domain superfamily', *Biochemical Society Transactions*, 39(4), 1033–1038. doi: 10.1042/BST0391033.

Kota, S. K., Pernicone, E., Leaf, D. E., Stillman, I. E., Waikar, S. S., and Kota, S. B. (2017) 'BPI Fold-Containing Family A Member 2/Parotid Secretory Protein Is an Early Biomarker of AKI', *Journal of the American Society of Nephrology*, 28(12), 3473–3478. doi: 10.1681/asn.2016121265.

Krasity, B. C., Troll, J. V., Weiss, J. P., and McFall-Ngai, M. J. (2011) 'LBP/BPI proteins and their relatives: Conservation over evolution and roles in mutualism', *Biochemical Society Transactions*, 39(4), 1039–1044. doi: 10.1042/BST0391039.

Laine, F. J., and Smoker, W. R. K. (1996) 'Oral cavity anatomy and pathology', *Radiologist*, 3(1), 37–46. doi: 10.1016/s0887-2171(06)80024-7

LeClair, E. E., Nomellini, V., Bahena, M., Singleton, V., Bingle, L., Craven, C. J., and Bingle, C. D. (2004) 'Cloning and expression of a mouse member of the PLUNC protein family exclusively expressed in tongue epithelium', *Genomics*, 83(4), 658–666. doi: 10.1016/j.ygeno.2003.09.015.

Leito, J. T., Ligtenberg, A. J., Nazmi, K., and Veerman, E. C. (2009) 'Identification of salivary components that induce transition of hyphae to yeast in Candida albicans', *FEMS Yeast Research*, 9(7), 1102–1110. doi: 10.1111/j.1567-1364.2009.00575.x.

Leito, J. T., Ligtenberg, A. J., van Houdt, M., van den Berg, T. K., and Wouters, D. (2011) 'The bacteria binding glycoprotein salivary agglutinin (SAG/gp340) activates complement via the lectin pathway', *Molecular Immunology*. Elsevier Ltd, 49(1–2), 185–190. doi: 10.1016/j.molimm.2011.08.010.

Lin, S. Y., Wei, Y. S., Li, M. J., and Wang, S. L. (2004) 'Effect of ethanol or/and captopril on the secondary structure of human serum albumin before and after protein binding', *European Journal of Pharmaceutics and Biopharmaceutics*, 57(3), 457–464. doi: 10.1016/j.ejpb.2004.02.005.

Lunn, H. L. (2014) 'Purification and functional analysis of BPIFA2', *PQDT - UK and Ireland*. Available at: https://search.proquest.com/docview/1779546964?accountid=13607%0Ahttp://e-tidsskrifter.kb.dk/resolve??url_ver=Z39.882004andrft_val_fmt=info:ofi/fmt:kev:mtx:dissertationandgenre=dissertations+%26+thesesandsid=ProQ:ProQuest+Dissertations+%26+Theses+Globalandatitle.

Lynge Pedersen, A. M., and Belstrøm, D. (2019) 'The role of natural salivary defences in maintaining a healthy oral microbiota', *Journal of Dentistry*. Elsevier, 80(August 2018), S3–S12. doi: 10.1016/j.jdent.2018.08.010.

Madsen, H. O., and Hjorth, J. P. (1985). Molecular cloning of mouse PSP mRNA. *Nucleic acids research*, 13(1), 1–13. https://doi.org/10.1093/nar/13.1.1

Masson, D., Jiang, X. C., Lagrost, L., and Tall, A. R. ((2009) 'The role of plasma lipid transfer proteins in lipoprotein metabolism and atherogenesis', *Journal of Lipid Research*. doi: 10.1194/jlr.R800061-JLR200. 50, S201-S206.

McDonald, R. E., Fleming, R. I., Beeley, J. G., Bovell, D. L., Lu, J. R., Zhao, X., Cooper, A., and Kennedy, M. W. (2009) 'Latherin: A surfactant protein of horse sweat and saliva', *PLoS ONE*, 4(5), 1–12. doi: 10.1371/journal.pone.0005726.

McGillivary, G., and Bakaletz, L. O. (2010) 'The multifunctional host defense peptide SPLUNC1 is critical for homeostasis of the mammalian upper airway', *PLoS ONE*, 5(10). doi: 10.1371/journal.pone.0013224.

Miyata, H., Castaneda, J. M., Fujihara, Y., Yu, Z., Archambeault, D. R., Isotani, A., Kiyozumi, D., Kriseman, M. L., Mashiko, D., Matsumura, T., Matzuk, R. M., *et al.* (2016) 'Genome engineering uncovers 54 evolutionarily conserved and testis-enriched genes that are not required for male fertility in mice', *Proceedings of the National Academy of Sciences of the United States of America*, 113(28), 7704–7710. doi: 10.1073/pnas.1608458113.

Mirels, L., and Ball, W. D. (1992) 'Neonatal Rat Submandibular Gland Protein SMG-A and Parotid Secretory Protein Are Alternatively Regulated Members of a Salivary Protein Multigene Family ', *The Journal of biological chemistry*, 267(4), 2679–2687.

Morita, Y., Ishikawa, K., Nakano, M., Wakabayashi, H., Yamauchi, K., Abe, F., Ooka, T., and Hironaka, S. (2017) 'Effects of lactoferrin and lactoperoxidase-containing food on the oral microbiota of older individuals', *Microbiology and Immunology*, 61(10), 416–426. doi: 10.1111/1348-0421.12537.

Mutreja, I., Lan, C., Li, Q., and Aparicio, C. (2023) 'Chemoselective Coatings of GL13K Antimicrobial Peptides for Dental Implants', *Pharmaceutics*, 15(10), 1–13. doi: 10.3390/pharmaceutics15102418.

Nandula, S. R., Huxford, I., Wheeler, T. T., Aparicio, C., and Gorr, S. U. (2020) 'The parotid secretory protein BPIFA2 is a salivary surfactant that affects lipopolysaccharide action', *Experimental Physiology*, 105(8), 1280–1292. doi: 10.1113/EP088567.

Nieuw Amerongen, A. V., and Veerman, E. C. I. (2002) 'Saliva - The defender of the oral cavity', *Oral Diseases*, 8(1), 12–22. doi: 10.1034/j.1601-0825.2002.1o816.x.

Ooi, C. E., Weiss, J., Doerfler, M. E., and Elsbach, P. (1991) 'Endotoxin-neutralizing properties of the 25 kD N-terminal fragment and a newly isolated 30 kD C-Terminal fragment of the 55-
60 kD bactericidal/permeability-increasing protein of human neutrophils', *Journal of Experimental Medicine*,174(3), 649–655. doi: 10.1084/jem.174.3.649.

O'Sullivan, J. M., Jenkinson, H. F., and Cannon, R. D. (2000) 'Adhesion of Candida albicans to oral streptococci is promoted by selective adsorption of salivary proteins to the streptococcal cell surface', *Microbiology*, 146(1), 41–48. doi: 10.1099/00221287-146-1-41.

Pandey, V. K., Sharma, R., Prajapati, G. K., Mohanta, T. K., and Mishra, A. K. (2022) 'N-glycosylation, a leading role in viral infection and immunity development', *Molecular Biology Reports*. Springer Netherlands, 49(8), pp. 8109–8120. doi: 10.1007/s11033-022-07359-4.

Pedersen, A. M. L., Darwish, M., Nicholson, J., Edwards, M. I., Gupta, A. K., and Belstrøm, D. (2019) 'Gingival health status in individuals using different types of toothpaste', *Journal of Dentistry*. Elsevier, 80(July 2018), S13–S18. doi: 10.1016/j.jdent.2018.08.008.

Pedersen, A. M. L., Sørensen, C. E., Proctor, G. B., Carpenter, G. H., and Ekström, J. (2018) Salivary secretion in health and disease, *Journal of Oral Rehabilitation.* 45(9), 730–746. doi: 10.1111/joor.12664.

Pittman, T. W., Decsi, D. B., Punyadeera, C., and Henry, C. S. (2023). *Saliva-based microfluidic point-of-care diagnostic.* Theranostics, 13(3), 1091–1108. doi: 10.7150/thno.78872

Ployon, S., Belloir, C., Bonnotte, A., Lherminier, J., Canon, F., and Morzel, M. (2016) 'The membrane-associated MUC1 improves adhesion of salivary MUC5B on buccal cells. Application to development of an in vitro cellular model of oral epithelium', *Archives of Oral Biology*. 61, 149–155. doi: 10.1016/j.archoralbio.2015.11.002.

Poulsen, K., Jakobsen, B. K., Mikkelsen, B. M., Harmark, K., Nielsen, J. T., and Hjorth, J. P. 1986) 'Coordination of murine parotid secretory protein and salivary amylase expression.',

The EMBO Journal, 5(8), 1891–1896. doi: 10.1002/j.1460-2075.1986.tb04441.x.

Preciado, D., Goyal, S., Rahimi, M., Watson, A. M., Brown, K. J., Hathout, Y., and Rose, M. C. (2010) 'MUC5B is the predominant mucin glycoprotein in chronic otitis media fluid', *Pediatric Research*, 68(3), 231–236. doi: 10.1203/PDR.0b013e3181eb2ecc.

Prokopovic, V., Popovic, M., Andjelkovic, U., Marsavelski, A., Raskovic, B., Gavrovic-Jankulovic, M., and Polovic, N. (2014) 'Isolation, biochemical characterization and anti-bacterial activity of BPIFA2 protein', *Archives of Oral Biology*. Elsevier Ltd, 59(3), 302–309. doi: 10.1016/j.archoralbio.2013.12.005.

Rajan, G. H., Morris, C. A., Carruthers, V. R., Wilkins, R. J., and Wheeler, T. T. (1996) 'The relative abundance of a salivary protein, bSP30, is correlated with susceptibility to bloat in cattle herds selected for high or low bloat susceptibility', *Animal Genetics*, 27(6), 407–414. doi: 10.1111/j.1365-2052.1996.tb00507.x.

Raj, P. A., Johnsson, M., Levine, M. J., and Nancollas, G. H. (1992) 'Salivary statherin: Dependence on sequence, charge, hydrogen bonding potency, and helical conformation for adsorption to hydroxyapatite and inhibition of mineralization', *Journal of Biological Chemistry*, 267(9), 5968–5976. doi: 10.1016/s0021-9258(18)42650-6.

Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Loo, J. A. (2006) 'Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry', *Journal of Proteome Research*, 5(6), 1493–1503. doi: 10.1021/pr050492k.

Reitamo, S., Konttinen, Y. T., and Segerberg-Konttinen, M. (1980) 'Distribution of lactoferrin in human salivary glands', *Histochemistry*, 66(3), 285–291. doi: 10.1007/BF00495741.

Rittig, M. G., Kaufmann, A., Robins, A., Shaw, B., Sprenger, H., Gemsa, D., Foulongne, V., Rouot, B., and Dornand, J. (2003) ' Smooth and rough lipopolysaccharide phenotypes of Brucella induce different intracellular trafficking and cytokine/chemokine release in human monocytes ', *Journal of Leukocyte Biology*, 74(6), 1045–1055. doi: 10.1189/jlb.0103015.

Robinson, C. P., Bounous, D. I., Alford, C. E., Nguyen, K. H., Nanni, J. M., Peck, A. B., and Humphreys-Beher, M. G. (1997) 'PSP expression in murine lacrimal glands and function as a bacteria binding protein in exocrine secretions', *American Journal of Physiology - Gastrointestinal and Liver Physiology*.272(4Pt 1), G863–G871.
doi: 10.1152/ajpgi.1997.272.4.g863.

Russell, M. W., Hajishengallis, G., Childers, N. K., and Michalek, S. M. (1999) 'Secretory Immunity in Defense against Cariogenic Mutans Streptococci', *Caries Research*, 33(1), 4–15. doi: 10.1159/000016490.

Ryu, J. K., Kim, S. J., Rah, S. H., Kang, J. I., Jung, H. E., Lee, D., Lee, H. K., Lee, J. O., Park, B. S., Yoon, T. Y., and Kim, H. M. (2017) 'Reconstruction of LPS Transfer Cascade Reveals Structural Determinants within LBP, CD14, and TLR4-MD2 for Efficient LPS Recognition and Transfer', *Immunity*. Elsevier Inc., 46(1), 38–50. doi: 10.1016/j.immuni.2016.11.007.

Sabatini, L. M., Carlock, L. R., Johnson, G. W., and Azen, E. A. (1987). 'cDNA cloning and chromosomal localization (4q11-13) of a gene for statherin, a regulator of calcium in saliva'. *American journal of human genetics*, 41(6), 1048–1060.

Samaranayake, Y. H., Samaranayake, L. P., Pow, E. H., Beena, V. T., and Yeung, K. W. (2001) 'Antifungal effects of lysozyme and lactoferrin against genetically similar, sequential candida albicans isolates from a human immunodeficiency virus-infected southern chinese cohort', *Journal of Clinical Microbiology*, 39(9), 3296–3302. doi: 10.1128/JCM.39.9.3296-3302.2001.

Scannapieco, F. A., Torres, G., and Levine, M. J. (1993) 'Salivary α-amylase: Role in dental plaque and caries formation', in *Critical Reviews in Oral Biology and Medicine*. 4(3-4), 301–307. doi: 10.1177/10454411930040030701.

Schicht, M., Rausch, F., Beron, M., Jacobi, C., Garreis, F., Hartjen, N., Beileke, S., Kruse, F., Bräuer, L., and Paulsen, F. (2015) 'Palate lung nasal clone (PLUNC), a novel protein of the tear film: Three-dimensional structure, immune activation, and Involvement in dry eye disease (DED)', *Investigative Ophthalmology and Visual Science*, 56(12), 7312–7323. doi: 10.1167/iovs.15-17560.

Schumann, R. R. (1992) 'Function of lipopolysaccharide (LPS)-binding protein (LBP) and CD14, the receptor for LPS/LBP complexes: a short review', *Research in Immunology*, 143(1), 11–15. doi: 10.1016/0923-2494(92)80074-U.

Shang, L., Deng, D., Buskermolen, J. K., Janus, M. M., Krom, B. P., Roffel, S., Waaijman, T., van Loveren, C., Crielaard, W., and Gibbs, S. (2018) 'Multi-species oral biofilm promotes reconstructed human gingiva epithelial barrier function', *Scientific Reports*, 8(1), 1–12. doi: 10.1038/s41598-018-34390-y.

Shaw, P., and Schibler, U. (1986) 'Structure and expression of the parotid secretory protein gene of mouse', *Journal of Molecular Biology*, 192(3), 567–576. doi: 10.1016/0022-2836(86)90277-9.

Shiba, H., Venkatesh, S. G., Gorr, S. U., Barbieri, G., Kurihara, H., and Kinane, D. F. ((2005) 'Parotid secretory protein is expressed and inducible in human gingival keratinocytes', *Journal of Periodontal Research*. 40(2), 153–157. doi: 10.1111/j.1600-0765.2005.00781.x.

Shin, K., Yaegaki, K., Murata, T., Ii, H., Tanaka, T., Aoyama, I., Yamauchi, K., Toida, T., and Iwatsuki, K. (2011) 'Effects of a composition containing lactoferrin and lactoperoxidase on oral malodor and salivary bacteria: A randomized, double-blind, crossover, placebo-controlled clinical trial', Clinical Oral Investigations, 15(4), 485–493. doi: 10.1007/s00784-010-0422-x.

Schürch, S., Gehr, P., Im Hof, V., Geiser, M., and Green, F. (1990). Surfactant displaces particles toward the epithelium in airways and alveoli. *Respiration physiology*, 80(1), 17–32. https://doi.org/10.1016/0034-5687(90)90003-h

Simões, C., Caeiro, I., Carreira, L., Silva, F. C. E., and Lamy, E. (2021) 'How different snacks produce a distinct effect in salivary protein composition', *Molecules*, 26(9). doi: 10.3390/molecules26092403.

Ship, J. A. (2002) 'Diagnosing, managing, and preventing salivary gland disorders', *Oral Diseases*, 8(2), 77–89. doi: 10.1034/j.1601-0825.2002.2o837.x.

Siqueira, W. L., Margolis, H. C., Helmerhorst, E. J., Mendes, F. M., and Oppenheim, F. G. (2010) 'Evidence of intact histatins in the in vivo acquired enamel pellicle', *Journal of Dental Research*, 89(6), 626–630. doi: 10.1177/0022034510363384.

Sivakumar, S., Mirels, L., Miranda, A. J., and Hand, A. R. (1998) 'Secretory protein expression patterns during rat parotid gland development', *Anatomical Record*, 252(3), 485–497. doi: 10.1002/(SICI)1097-0185(199811)252:3<485::AID-AR17>3.0.CO;2-J.

Soukka, T., Tenovuo, J., and Rundegren, J. (1993). Agglutination of Streptococcus mutans serotype C cells but inhibition of Porphyromonas gingivalis autoaggregation by human lactoferrin. *Archives of oral biology*, 38(3), 227–232. doi: 10.1016/0003-9969(93)90032-h.

Takehara, S., Yanagishita, M., Podyma-Inoue, K. A., and Kawaguchi, Y. (2013) 'Degradation of MUC7 and MUC5B in Human Saliva', *PLoS ONE*, 8(7), 1–9. doi: 10.1371/journal.pone.0069059.

TEN CATE, A. R. 1998. Oral histology : development, structure, and function, St. Louis, Mo. ; London, Mosby.

Thamadilok, S., Choi, K. S., Ruhl, L., Schulte, F., Kazim, A. L., Hardt, M., Gokcumen, O., and Ruhl, S. (2020) 'Human and Nonhuman Primate Lineage-Specific Footprints in the Salivary Proteome', *Molecular Biology and Evolution*, 37(2), pp. 395–405. doi: 10.1093/molbev/msz223.

Thornton, D. J., Khan, N., Mehrotra, R., Howard, M., Veerman, E., Packer, N. H., and Sheehan, J. K. (1999). Salivary mucin MG1 is comprised almost entirely of different glycosylated forms of the MUC5B gene product. *Glycobiology*, 9(3), 293–302. doi: 10.1093/glycob/9.3.293.

Tsang, C. S. P., and Samaranayake, L. P. (1999) 'Salivary lysozyme and related parameters of a predominantly Chinese, HIV-infected cohort in Hong Kong', *Oral Diseases*, 5(3), 241–246. doi: 10.1111/j.1601-0825.1999.tb00308.x.

Tsou, Y. A., Peng, M. T., Wu, Y. F., Lai, C. H., Lin, C. D., Tai, C. J., Tsai, M. H., Chen, C. M., and Chen, H. C. (2014) 'Decreased PLUNC expression in nasal polyps is associated with multibacterial colonization in chronic rhinosinusitis patients', *European Archives of Oto-Rhino-Laryngology*, 271(2), 299–304. doi: 10.1007/s00405-013-2535-8.

Tsujimoto, H., Gotoh, N., and Nishino, T. (1999) 'Diffusion of macrolide antibiotics through the outer membrane of Moraxella catarrhalis', *Journal of Infection and Chemotherapy*, 5(4), 196–200. doi: 10.1007/s101560050034.

Tucker, A. S. (2007) 'Salivary gland development', *Seminars in Cell and Developmental Biology*, 18(2), 237–244. doi: 10.1016/j.semcdb.2007.01.006.

Vargas, P. A., Speight, P. M., Bingle, C. D., Barrett, A. W., and Bingle, L. (2008) 'Expression of PLUNC family members in benign and malignant salivary gland tumours', *Oral Diseases*. doi: 10.1111/j.1601-0825.2007.01429.x.

Veerman, E. C., Ligtenberg, A. J., Schenkels, L. C., Walgreen-Weterings, E., and Nieuw Amerongen, A. V. (1995) 'Binding of Human High-molecular-weight Salivary Mucins (MG1) to Hemophilus parainfluenzae', *Journal of Dental Research*, 74(1), 351–357. doi: 10.1177/00220345950740011101.

Venkatesh, S. G., Goyal, D., Carenbauer, A. L., and Darling, D. S. (2011). Parotid secretory protein binds phosphatidylinositol (3,4) bisphosphate. *Journal of dental research*, 90(9), 1085– 1090. doi: 10.1177/0022034511410699.

Verma, D., Garg, P. K., and Dubey, A. K. (2018). Insights into the human oral microbiome. *Archives of microbiology*, 200(4), 525–540. doi: 10.1007/s00203-018-1505-3.

Vidotto, A., Henrique, T., Raposo, L. S., Maniglia, J. V., and Tajara, E. H. (2010) 'Salivary and serum proteomics in head and neck carcinomas: Before and after surgery and radiotherapy', *Cancer Biomarkers*, 8(2), 95–107. doi: 10.3233/CBM-2011-0205.

Vignieri, S. (2023) 'Zoonomia', *Science*, 380(6643), pp. 356–357. doi: 10.1126/science.adi1599.

Vitorino, R., Lobo, M. J., Ferrer-Correira, A. J., Dubin, J. R., Tomer, K. B., Domingues, P. M., and Amado, F. M. (2004) 'Identification of human whole saliva protein components using proteomics', *Proteomics*, 4(4), 1109–1115. doi: 10.1002/pmic.200300638.

Weiss, J., Elsbach, P., Shu, C., Castillo, J., Grinna, L., Horwitz, A., and Theofan, G. (2003) 'Bactericidal/permeability-increasing protein (BPI) and lipopolysaccharide-binding protein (LBP): Structure, function and regulation in host defence against Gram-negative bacteria', *Biochemical Society Transactions*, 31(4), 785–790. doi: 10.1042/BST0310785.

Weston, W. M., LeClair, E. E., Trzyna, W., McHugh, K. M., Nugent, P., Lafferty, C. M., Ma, L., Tuan, R. S., and Greene, R. M. (1999). Differential display identification of plunc, a novel gene expressed in embryonic palate, nasal epithelium, and adult lung. *The Journal of biological chemistry*, 274(19), 13698–13703. doi: 10.1074/jbc.274.19.13698.

Wheeler, T. T., Haigh, B. J., Broadhurst, M. K., Hood, K. A., and Maqbool, N. J. (2011) 'The BPI-like/PLUNC family proteins in cattle', *Biochemical Society Transactions*, 39(4), 1006–1011. doi: 10.1042/BST0391006.

Wheeler, T. T., Haigh, B. J., McCracken, J. Y., Wilkins, R. J., Morris, C. A., and Grigor, M. R. (2002) 'The BSP30 salivary proteins from cattle, LUNX/PLUNC and von Ebner's minor salivary gland protein are members of the PSP/LBP superfamily of proteins', *Biochimica et Biophysica Acta - Gene Structure and Expression*. 1579(2-3), 92–100. doi: 10.1016/S0167-4781(02)00508-0.

Wheeler, T. T., Hood, K. A., Maqbool, N. J., McEwan, J. C., Bingle, C. D., and Zhao, S. (2007) 'Expansion of the Bactericidal/Permeability Increasing-like (BPI-like) protein locus in cattle', *BMC Genomics*, 8, 1–15. doi: 10.1186/1471-2164-8-75.

Wittmann, I., Schönefeld, M., Aichele, D., Groer, G., Gessner, A., and Schnare, M. (2008) 'Murine Bactericidal/Permeability-Increasing Protein Inhibits the Endotoxic Activity of Lipopolysaccharide and Gram-Negative Bacteria', *The Journal of Immunology*, 180(11), 7546–7552. doi: 10.4049/jimmunol.180.11.7546.

Wu, J., Kobayashi, M., Sousa, E. A., Liu, W., Cai, J., Goldman, S. J., Dorner, A. J., Projan, S. J., Kavuru, M. S., Qiu, Y., and Thomassen, M. J. (2005) 'Differential proteomic analysis of bronchoalveolar lavage fluid in asthmatics following segmental antigen challenge', *Molecular and Cellular Proteomics*, 4(9), 1251–1264. doi: 10.1074/mcp.M500041-MCP200.

Wu, Y., Shu, R., Luo, L. J., Ge, L. H., and Xie, Y. F. (2009) 'Initial comparison of proteomic profiles of whole unstimulated saliva obtained from generalized aggressive periodontitis patients and healthy control subjects', *Journal of Periodontal Research*, 44(5), 636–644. doi: 10.1111/j.1600-0765.2008.01172.x.

Yang, L. L., Liu, X. Q., Liu, W., Cheng, B., and Li, M. T. (2006) 'Comparative analysis of whole saliva proteomes for the screening of biomarkers for oral lichen planus', *Inflammation Research*, 55(10), 405–407. doi: 10.1007/s00011-006-5145-8.

Yeh, C. K., Dodds, M. W., Zuo, P., and Johnson, D. A. (1997) 'A population-based study of salivary lysozyme concentrations and candidal counts', *Archives of Oral Biology*, 42(1), 25–31. doi: 10.1016/S0003-9969(96)00104-5.

Yin, H. F., Fan, B. L., Yang, B., Liu, Y. F., Luo, J., Tian, X. H., and Li, N. (2006) 'Cloning of pig parotid secretory protein gene upstream promoter and the establishment of a transgenic mouse model expressing bacterial phytase for agricultural phosphorus pollution control', *Journal of Animal Science*, 84(3), 513–519. doi: 10.2527/2006.843513x.

Zellner, M., Winkler, W., Hayden, H., Diestinger, M., Eliasen, M., Gesslbauer, B., Miller, I., Chang, M., Kungl, A., Roth, E., and Oehler, R. (2005) 'Quantitative validation of different protein precipitation methods in proteome analysis of blood platelets', *Electrophoresis*, 26(12), 2481–2489. doi: 10.1002/elps.200410262.

Zhang, C. Z., Cheng, X. Q., Li, J. Y., Zhang, P., Yi, P., Xu, X., and Zhou, X. D. (2016). Saliva in the diagnosis of diseases. *International journal of oral science*, 8(3), 133–137. doi: 10.1038/ijos.2016.38.

Zhang, H., Burrows, J., Card, G. L., Attwood, G., Wheeler, T. T., and Arcus, V. L. (2019). The three dimensional structure of Bovine Salivary Protein 30b (BSP30b) and its interaction with specific rumen bacteria. *PloS one*, 14(4), e0206709. https://doi.org/10.1371/journal.pone.0206709

Appendix

Multiple sequence alignments using in chapter 3; figure 3.4

>Homo_sapiens
MLQLWKLVLLCGVLTGTSESLLDNLGNDLSNVVDKLEPVLHEGLETVDNTLKGILE
KLKVDLGVLQKSSAWQLAKQKAQEAEKLLNNVISKLLPTNTDIFGLKISNSLILDVKA
EPIDDGKGLNLSFPVTANVTVAGPIIGQIINLKASLDLLTAVTIETDPQTHQPVAVLGE
CASDPTSISLSLLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDVNVIQQV
VDNPQHKTQLQTLI

>Mus_musculus
MFQLGSLVVLCGLLIGNSESLLGELGSAVNNLKILNPPSEAVPQNLNLDVELLQQATS
WPLAKNSILETLNTADLGNLKSFTSLNGLLLKINNLKVLDFQAKLSSNGNGIDLTVPL
AGEASLVLPFIGKTVDISVSLDLINSLSIKTNAQTGLPEVTIGKCSSNTDKISISLLGRRL
PIINSILDGVSTLLTSTLSTVLQNFLCPLLQYVLSTLNPSVLQGLLSNLLAGQVQLAL

> Rattus_norvegicus
MFQLGSLVVLCGLLIGTSESLLGDVANAVNNLDILNSPSEAVAQNLNLDVGSLQQAT
TWPSAKDSILETLNKVELGNSNGFTPLNGLLLRVNKFRVLDLQAGLSSNGKDIDLKLP
LVFEISFSLPVIGPTLDVAVSLDLLNSVSVQTNAQTGLPGVTLGKCSGNTDKISISLLGR
RLPFVNRILDGVSGLLTGAVSILLQNILCPVLQYLLSTMSGSAIQGLLSNVLTGQLAVP
L

>Sus_scrofa
MFQLWKLVFLCGLLIGTSASLLEDLENDLSDADDKLQPVIDKGLETVEPVLQKLKAE
LESLQESESWQEAKQKVQEAENLLDKVLSKIFQITEKATGLKISNLHILDAKVKLTAD
GKGISLNLPITANVSLTLPLLGQVVDLDLNLDLLTSVTIETDAKTGVSTVTLGECASDP
ASISLTLLDRRRALVNRAVNSVIKVLRKVVSLVVQKEVCPMIQKLVENLGASLLKNII
DSLPKA

>Microcebus_murinus
MLWKLVLLCGLLTGPSEALLENLGNDLHNVVDKLKPIVDKGLETVDNTLQGVLQNL
KVDLKLLQESEAWQLAKKKVQEAEKLVDNALSDVHLSAGKALGLKISDSIILDIKPE
LTADGKGINLRVPVVADVSATLPLIGQVVDLKASLDLLTRVKVETDAQTKLLKVTLG
ECVSDATSISLSLLDRRSTLINNLVGSVTSILKKSVSSLVQKDLCPLLGIFIHGLDVKILQ
DIAVKLQERSSVQTA

> Callithrix_jacchus
MLQLWKLVLLCSLLTGTSASLLGNLGDDLSNVVDEQKPVLDEGLKTIDNTLKGVPE
KLKVDLGVLQQSAAWQLAKQKVQDAGKSLNNVVSKLLPTNTNILGLKISNSLILDV
KAEPTDDGKGLNLRFPVTADVSQTLPIIGQVVKLNASLDLLTAVKIEIDPQTHKPVAV
LGECANDPTSISLSLLNAQSQVINRLVNTVINTVKSTVSLLLLKEICPLIHIFLHSLDVNF
IQQVIDKLQQETQLQTHL

>Sapajus_apella
MLPLWKLVLLCSLLTGTSASLLGNLGDDLSNVVDKVKPVLDKGVETVDNTLKGVLE
KLKVDLGVLQESDAWQLAKQKVQEAEKLLNNVVSKLLPTNTNILGLKISNSLILDVK
AEPTDDGKGLNLRFPVTADVSVTLPIIGQVVKLNASLDLLTAVRIEIDPQTHKPVAVL
GECASDPTSISLSLLDGQSQVINKLVNSVINTVKSTVSFLVQKEICPLIHIFLHSLDINFI
QQVIGKLQQETQLQTHL

>Pongo_abelii
MTVQVSRQKMLQLWKLVLLCGVLTGTSESLLDNLGNDLSNVVDKLEPVLHDGLEA
VDNTLKGILEKLKVDLGVLQKSSAWQLAKQKAQEAEKLLNNVISKLLPTNTDIFGLK
ISNSLILDVKAEPIDDGKGLDLSFPVTANVTVAGPITGQIINLKASLDLLTAVTIETDPQ
THQPVAVLGECASDPTSISLSLLDKHSQIINKFVNSVINTLKSTVSSLVQKEICPLIRIFIH
SLDVNVIQQVIDNLQHKTQLQTLI

>Pan_troglodytes
MLRLWKLVLLCGVLTGTSESLLDNLGNDLSNVVDKLEPVLHEGLETVDNTLKGILEK
LKVDLGVLQKSSAWQLAKQKAQEAEKLLNNVISKLLPTNTDIFGLKISNSLILDVKAE
PIDDGKGLNLSFPVTANVTVAGPIIGQIINLKASLDLLTAVTIETDPQTHQPVAVLGEC
ASDPTSISLSLLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDVNVIQQVI
DNLQHKTQLQTLI

>Gorilla_gorilla_gorilla
MTVQVSRQKMLQLWKLVLLCGVLTGTSESLLDNLGNDLSNVVDKLEPVLHDGLET
VDNTLKGILEKLKVDLGVLQKSSAWQLAKQKAQEAEKLLNNVISKLLPTNTDIFGLK
ISNSLILDVKAEPIDDGKGLNLSFPVIANVTVAGPIIGQIINLKASLDLLTAVTIETDPQT
HQPVAVLGECASDPTSISLSLLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHS
LDVNVIQQVIDNLQHKTQLQTLI

>Macaca_mulatta
MLQLWKFVLLCGVLTGTSESLLDNLGSDLSNVVNELKPILHDGLETVDNTLKGVLEK
LKVDLGVLQKSSAWQLAKQKAQEAEKLLNNVVSKLLPTNTNTLGLKISDSLILDVKA
EPIDGGKGLNLSFPVTADVTTTLPIIGQIINLKASLDLLTAVSIETDPQTNQSVAVLGEC
ASDPTSISLSLLDNRSQIINNVVNRVINTLKSTVSFLVQKEICPLIRIFLHSLDVKFIQQII
DNLQHETQLQTPI

>Trachypithecus_francoisi
MLQLWKLVLLCGVLTGTSASLLDNLGSDLSNVVDELKPILQDGPETVDNTLKDVLE
KLKVNLGVLQKSSAWQLAKQKAQEPEKLLNNVISKLLPTNTNTLGLKISDYLILDVK
AEPIDDGKGLNLSFPVTANVTVALPIIGQIINLQASLDLLTAVSIETDPQTHQSVAVLGE
CASDPTSISLSLLDNRSQIINNVVNRVTNTLKSTVSFLVQKEICPLIRIFLHSLDVEFIQQI
IDNLQQETQLQTHI

>Erinaceus_europaeus
MFQLWKLFLLCGLLTGTSASLPNNDGLLGDVDNPLNSNDLSAEGDLKPINDKELQVV
GGTVDSVFQSTLGNVKTDTEQAVLKLDGYTTTINGYILKINSPQILDIKSEVTPDGKGI
NLRIPIRLRVSVSLPLIGQVLDLQVALDLLSGVSLKVNDETGLLGLVATECASDPASLS
LSLLNSRLTLLNKILGVVTGIVENVLSTLLENVLCPVVQVLVKTLAVQDIQGITSNLN
QQVNEPFVI

>Elephas_maximus_indicus
MFQLWKLALLCGLLTGTSASLLGNFGNDPNVLDNLKPILDKGLETVDNTLETVDNTL
KTADNSVQDDLQKLKLKIEGLQDSKAWQAIQEKYQETGKLLDKGLSGLISDIKSLGL
KIGKSRILDLKAELTPDGQGLNLRFPTTANVNLALPLIGNIVNLNAALDILTGVSIDIDT
QTGLPVVILGECTSDPDSVQVNLLNRHSALINRIANTLSSFLRKTVSFTVQKEVCLLIR
FFLNTLGVNVIPNIINKLQDGIHLQI

>Dasypus_novemcinctus
MFQLWKLVLLCGLLTGTSASLLGTLENDLTNVVDKVKPVIDKGLETVDNTLDVLLQ
KVKVDLEKLQGSQAWKLAKEKIQEVENLVGSTVSKLGQDLEKALGLKISNANIQDL
KASLAPDNQTINLRIPVSADVSLTLPLIGKVVGLKASLDLQIGLKVETDVQTGLPVVIL
GECTSDPANVQLTLLDSENAMVKHIVETMTKVLVKTVSFLVQKEMCPMIRIFLHTLD
VDVIQNLVHKLQQGIHLHISV

> Saimiri_boliviensis_boliviensis

MLQLWKLVLLCSLLTGTSASLLGSLGDDLSNVVDEVKPVLDKGLETIDNTLKDVVE
KLKVDLGVLQQSGAWQLAKQKVQEAEKLLNGVVSKLLPTNTNILGLKISNSLILDVK
AEPTDDGKGLNLRFPVTADVSATLPIIGQVVKLNAALDLLTAVRIEIDPQTHKPVAVL
GECASDPTSISLSLLDAQSQVINKLVNSVINTVKSTVSFLVQKEICPLIRIFLHSLDVNFI
QQVIGKLQQETQLQTHL

> Canis_lupus_familiaris
MLQLWKLVLLCGLLTGTSASLLGNLGDDLNNVVDKLKPVVEKGLETVDNTLESVLQ
KLKADWKIIQKSKAWHLAEEKVQEVKNLVNDALSKIVPAKDDTLGLNIINSRILKIKA
ELTLDGEGLNIRIPVVANVTLALPLIDRVVNLKVSLDLVTSVRLATNAQTGAVTVIVG
KCSSDEDSISLTVLDSHNGLIEKAANTVSSFLTKTLSRLIEKDVCPLIHTLLSNLDGHIIQ
DIIDKFQKEDHVPNAA

> Tursiops_truncatus
MFQLWKLILLCGLLTGTSASLLEDLGNDVVSKLKPVLNKGLETVDTILQKLKAALEA
LRESTSWQEVQQKIQEAENLLDKFLSKIFQVVEKVMGLKISNVHILDIKSEVTPDGKG
ASLSIPITTNVTLTLPLLGELVDLGLQTSVSIETDAKTGVSTVIVEKCTNDPANIPLTFP
DSPIGLVNEAVNTVVKLMRKTVSLVVQXEGSTCCPHHVSKPQWQLPEPITTPPHSML
GANFGPSIQVATLAAIHVGTI

Consent form used in chapter 4; section 4.4

Participant Consent Form

Title of Research Project: Function Analysis of BPIFA2 in human saliva

Name of Researcher: Lin Zhang supervised by Dr Lynne Bingle

Participant Identification Number for this project: **Please initial box**

1. I confirm that I have read and understand the information sheet dated [24th May 2022] explaining the above research project and I have had the opportunity to ask questions about the project. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline. (Contact details of researchers LZhang89@sheffield.ac.uk or l.bingle@sheffield.ac.uk ☐

3. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research. ☐

4. I agree for the data collected from me to be used in future research ☐

5. I agree to take part in the above research project. ☐

Name of Participant	Date	Signature
Name of person taking consent (*if different from lead researcher*)	Date	Signature

To be signed and dated in presence of the participant

Lead Researcher	Date	Signature

To be signed and dated in presence of the participant

Copies:
Once this has been signed by all parties the participant should receive a copy of the signed and dated participant consent form, the information sheet and any other written information provided to the participants. A copy of the signed and dated consent form should be placed in the project's main record (e.g. a site file), which must be kept in a secure location.