INVESTIGATION OF BACTERIAL INFECTION STRATEGIES THROUGH BACTERIA-HUMAN PROTEIN-PROTEIN INTERACTIONS BY BIOINFORMATICS APPROACHES

by

Büşra GÜRSES

B.S., Chemical Engineering, Yıldız Technical University, 2011

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering

Boğaziçi University

2014

to my precious family

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor, Prof. Kutlu Ülgen, for her support, counsel and kind attitude. Also, I sincerely thank Assist. Prof. Saliha Durmuş Tekir for her everlasting support and willingness to contribute her knowledge to my studies.

Additionally, I would like to thank Prof. Belma Kın Özbek and Assoc. Prof. Elif Özkırımlı Ölmez for devoting their time to read and comment on my thesis.

I gratefully acknowledge, this thesis has been financially supported by Boğaziçi Research Fund through project 7082.

I would also like to thank the former and the current members of Biosystems Engineering Research Group at Boğaziçi University for their support and companionship. I would also like to thank the former and the current members of Catrel team for their friendship and cheerfulness.

I am also deeply grateful to my best friends Özge Ertem and Didem Kabakçı for their never-ending help and friendship. My special thanks are due to Sezin Sezen for her priceless friendship, everlasting support, encouraging and cheerful words, helping me out through countless days and nights, being the most understanding friend during my hardest times.

Last, but certainly not the least, I am forever indepted to my father Nesim Gürses, my mother Leyla Gürses and my brother Buğra Gürses for providing me with peace, support, love, patience, joy and advices throughout my whole life. I feel very lucky for having them. This thesis would not be possible without them.

ABSTRACT

INVESTIGATION OF BACTERIAL INFECTION STRATEGIES THROUGH BACTERIA-HUMAN PROTEIN-PROTEIN INTERACTIONS BY BIOINFORMATICS APPROACHES

Infectious diseases, which have been a serious threat for human life in ancient times, is still leading the causes of death and disability worldwide. Pathogens (viruses, bacteria, protozoa, fungi, etc.) express a wide range of molecules that bind to host cell targets to manipulate human mechanisms for their own advantage, and result in infection in the host organism. Well characterization of these interspecies (pathogen-human) interactions is required for a complete understanding of pathogenesis. The key point of fighting with infectious diseases is the identification and characterization of the strategies used by these pathogens to interact with the host, as these strategies are usually unique to specific pathogens or conserved across several different species. Within the framework of this project, bacterial infection mechanisms are investigated through bacteria-human protein-protein interactions. The targeted human proteins with specific properties different from the non-targeted ones are analyzed to enlighten several infection mechanisms. Gene Ontology enrichments of targeted human proteins are investigated; that is, all three GO terms are examined to identify the terms having significant association with each human protein set studied. In addition to GO terms, Kegg pathways, transcription factor targets and pathway commons of the targeted human proteins are identified. A thorough bioinformatic based analysis of bacteria-human protein interaction network is still missing in the literature. This study is the first attempt to comprehensively investigate bacterial infection strategies through bacteria-human protein-protein interactions. Common and specific infection strategies of different types of bacteria are identified, contributing to the novelty of the present research work.

ÖZET

BAKTERİYEL ENFEKSİYON STRATEJİLERİNİN BAKTERİ-İNSAN PROTEİN-PROTEİN ETKİLEŞİMLERİ KULLANILARAK BİYOİNFORMATİK ARAÇLAR İLE İNCELENMESİ

Eski çağlarda insan yaşamı için ciddi tehdit oluşturan enfeksiyöz hastalıklar, günümüzde de çok sayıda ölüme ve sakatlığa sebep olmaktadırlar. Çeşitli patojen molekülleri insan proteinleri ile etkileşerek, insanın hücresel mekanizmalarını patojenin kendi yararına kullanmasına izin verdirerek enfeksiyona neden olurlar. Türler arası (patojen-insan) bu etkileşimlerin detaylı aydınlatılması patojenezi bütünüyle anlamak için gereklidir. Enfeksiyöz hastalıklarla mücadelede kilit nokta patojenler tarafından konak ile etkileşime geçmek için kullanılan stratejilerin (genellikle belirli patojenlere ortak ya da farklı türlere özgü) tanımlanması ve çözümlenmesidir. Bu çalışma çerçevesinde, bakteri-insan protein-protein etkileşimleri yoluyla bakteriyel enfeksiyon mekanizmalarının ortaya çıkarılması amaçlanmıştır. Hedeflenmeyen proteinlerden farklı belirli özelliklere sahip olan ve hedef alınan insan proteinlerinin özelliklerinin incelenerek yeni araştırma olanaklarına öncülük etmesi beklenmektedir. Çalışılan her insan protein seti ile istatistiksel olarak anlamlı ve önemli ilişkisi bulunan terimleri belirlemek için tüm gen ontolojisi terimleri taranmıştır. Gen ontoloji terimlerinin yanısıra, hedeflenen insan proteinlerinin Kegg izyolları, transkripsiyon faktör hedefleri ve ortak izyolları incelenmiştir. Bakteri-insan protein etkileşim ağyapılarının biyoinformatik tabanlı kapsamlı bir analizi literatürde henüz mevcut değildir. Bu tez, bakteri-insan protein-protein etkileşimleri sayesinde bakteriyel enfeksiyon stratejilerinin ortaya çıkarılmasında en geniş kapsamlı çalışmalardan biri olmuştur. Bakteri-konak protein-protein etkileşim verilerinin toplanarak ve incelenerek değişik bakteri türlerinin ortak ve farklı olan enfeksiyon stratejilerinin araştırılmış olması, bu çalışmanın özgün değerini oluşturmaktadır.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

APCs	Antigen-presenting cells
CTLs	Cytotoxic T-lymphocytes
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
GO	Gene Ontology
KEGG	Kyoto Encyclopedia Of Genes And Genomes
LPS	Lipopolysaccharide
MP	Mononuclear phagocytes
NK	Natural killer
PHI	Pathogen-host Interaction
PHISTO	Pathogen-host Interaction Search Tool
PPI	Protein-Protein Interaction
TF	Transcription factor
TFT	Transcription factor target
TLRs	Toll-like receptors
TNF	Tumor necrosis factor
WebGestalt	Web Based Gene Set Analysis Toolkit
WHO	World Health Organization

1. INTRODUCTION

Since ancient times, even in today's modern world, pathogenic organisms remain to be the source of mortality. According to World Health Organization, more than 20% of total deaths in the world are because of the infectious diseases. However, with the developing technology, the amount of annual infectious disease deaths is aimed to drop over the next 20 years (WHO, 2012). The continual evolution of emerging and reemerging diseases remains a dominant feature of domestic and international public health considerations in the 21st century (Fauci, 2001). Infectious microorganisms, pathogens (viruses, bacteria, fungi, etc.), cause diseases by physical interactions with human proteins. Well characterization of these interspecies (pathogen-human) interactions is required for a complete understanding of pathogenesis. Pathogenic microorganisms interact with human proteins on the surface of the human cell or within the interior of the human cell to manipulate cellular mechanisms in order to use the host cell's functions to their own advantage. This results in diseases in host organisms.

The advances in high-throughput protein interaction detection methods have enabled the collection of large-scale data on pathogen-host protein-protein interactions (PHIs). In recent years virus-based infections (Calderwood *et al*., 2007; de Chassey *et al*., 2008; Shapira *et al*., 2009; Jager *et al.*, 2012; Durmus Tekir *et al.*, 2012) were mainly focused as data for other pathogenic organisms are insufficient. A thorough analysis of bacterial infection mechanisms is still missing. PHI data of bacteria-human systems and the related studies have been very rare up to nowadays, however recent experimental studies have provided enough bacterial PHIs in order to obtain statistically meaningful results, favoring a good opportunity to present the current picture of pathogenesis of bacterial infections. The first high-throughput experimental study producing bacterial PHI data between bacterial pathogens and their human hosts was revealed in 2010 by Dyer and coworkers.

In the scope of this thesis, up-to-date PHI data obtained from PHISTO are studied with a specific focus on bacterial infections of human. The purpose is to provide initial insights on bacterial infection mechanisms. The targeted human proteins are investigated thoroughly in terms of their Gene Ontology annotations (biological process, molecular

function, and cellular component), Kegg pathways, pathway commons and transcription factor targets to be able to figure out mechanisms in human attacked by bacterial pathogen. Special attention is paid to the human proteins that are targeted by 3 bacterial families which are *Enterobacteraceae, Bacillaceae* and *Francisellaceae* as they have large-scale data available.

The second chapter of this thesis, entitled as "Background Aspects" summarizes the fundamental information about bacteria, human immune system and the efforts about host-pathogen interactions. The next chapter gives details about how the analysis were done. The fourth chapter focuses on the results of the analysis on interactions between human proteins and bacteria proteins. The last chapter includes the main conclusions about this study along with the contributions to the research and the recommendations for future work.

2. BACKGROUND ASPECTS

2.1. Bacteria

2.1.1. Evolution of Bacterial Science

Nothing is known about the origin of bacteria, but because of their relative structural simplicity, it seems likely that they preceded the eukaryotic forms in the process of evolution. Anton van Leeuwenhoek was the first person to see bacteria. He first observed bacteria through his single-lens microscope in 1674. During his lifetime he made more than 250 microscopes consisting of home ground lenses mounted in brass and silver plates. His greatest discovery was in 1675 when he saw bacteria, fungi and many protozoa in rain water. He called them "animalcules". Leeuwenhoek sent a report of his sightings of bacteria and algae to the Royal Society in London. For his contributions, he was honoured as Father of Microbiology (Porter, 1976).

In 1850s, Louis Pasteur demonstrated that the fermentation process was caused by the growth of microorganisms or bacteria. Pasteur discovered the process of pasteurization, killing bacteria by heating. He coined the term vaccine. He invented a number of vaccines including the one against rabies. He also studied the bacterium that causes fowl cholera (Barnett, 2003).

In 1876, Robert Koch showed that bacteria can cause disease for the first time. This was followed by Robert Koch's experiments on bacteria as a source of disease, specifically the anthrax bacillus, for which he won the Nobel Prize in 1905. He introduced staining techniques and also methods of obtaining bacteria in pure culture using solid media. He discovered bacillus tuberculosis in 1882 and Vibrio cholera in 1883 (Sakula, 1983). In the 19th century, to identify pathogens that could be isolated with the techniques of the day, Koch's postulates were formulated by Robert Koch and Friedrich Loeffler as follows: the microbe must be present in every case of the disease, the organism must be grown in a pure culture from diseased hosts, the same disease must be produced when a pure culture of the organism is introduced into a susceptible host, and the organism must be recovered from

the experimentally infected host (Walker *et al*., 2006). After Koch discovered *Vibrio Cholera*, the german medical scientist Paul Ehrlich developed the first theory concerning how bacteria cause diseases and how the immune system fights these micro-organisms in 1880s (Ehrlich, 2013).

A young Dutch pathologist Wakker, working on the so-called yellow disease of hyacinth, between the years of 1883-89 proved that it was caused by bacteria. He found bacteria abundantly in the diseased tissues and he was able to incite the disease consistently by direct inoculation (Van Eijk *et al*., 1976). In early 1900s a Russian immunologist Dr. Eli Metchnikoff suggested that a synergistic interaction exists between bacteria and their host, known as phagocytosis later (Cavaillon, 2011). In 1874, a Norwegian physician Hansen described leprosy bacillus and identified bacterium Mycobacterium leprae as the causative agent of leprosy (Irgens, 2002). In 1881, a Scottish surgeon Ogston discovered staphylococcus (Elek, 1959). During 1880s, Loeffler described the diphtheria bacillus and its isolation in pure culture, and proved its relationship to the disease diphtheria (Pappenheimer *et al*., 1984).

A century ago, Hans Christian Gram developed a staining procedure which enabled the separation of most commonly encountered bacteria into two groups: gram positive and gram negative (Buck, 1982). It has been found that some organisms retain the violet stain and others are readily decolorized by the alcohol and take the counterstain. Those that retain the first stain are spoken of as Gram-positive organisms; whereas those that fail to retain the primary stain but take the counter-stain are called Gram-negative organisms (Salle, 1996).

An American microbiologist and biophysicist Carl Woese was interested in genetic code in evolutionary terms. When he first set forth the three-domain classification of life, he proposed the names "Archaebacteria" and "Eubacteria" for the two prokaryotic domains. These were replaced with the names "Archaea" and "Bacteria" in Woese *et al*. (1990). The taxonomic name "Bacteria" refers only to the eubacteria. The informal name "bacteria" is occasionally used loosely in the literature to refer to all of the prokaryotes, and care should be taken to interpret its meaning in any particular context.

2.1.2. Bacterial Cell Structure

Bacterial cells occur in all sorts of different shapes and sizes depending on the kind of organism and on the way in which it has been grown, but for many purposes it is possible to disregard these variations and to consider the common properties of the 'general bacterial cell' (Figure 2.1) (Mann, 1996). When unstained, most bacteria are transparent, colorless and apparently homogeneous bodies (Hiss *et al*., 1928). The hereditary material (DNA) is embedded in the cytoplasm which, surrounded by the cell membrane, is called the protoplast. Within the rigid cell wall of a typical bacterial cell is a protoplast consisting of nuclear body, cytoplasm and cell-membrane and in the cytoplasm probably containing various structures and non-living inclusions. Outside the protoplast lies the cell wall. (Hawker *et al*., 1968). It is of complex chemical composition, consistig of proteins, polysaccharides and sometimes lipids. Many bacterial cells are surrounded by a gummy mass called the capsule which is the extracellular coating. Capsule production is one of the major virulence factors utilised by bacteria to evade clearance from an infectious site. Specifically, the capsule provides bacteria with protection from the host immune response as well as antibiotics. The capsule protects bacteria from phagocytosis by not allowing opsonising antibodies to be recognised by phagocytic host defence cells (Wilson *et al*., 2001).

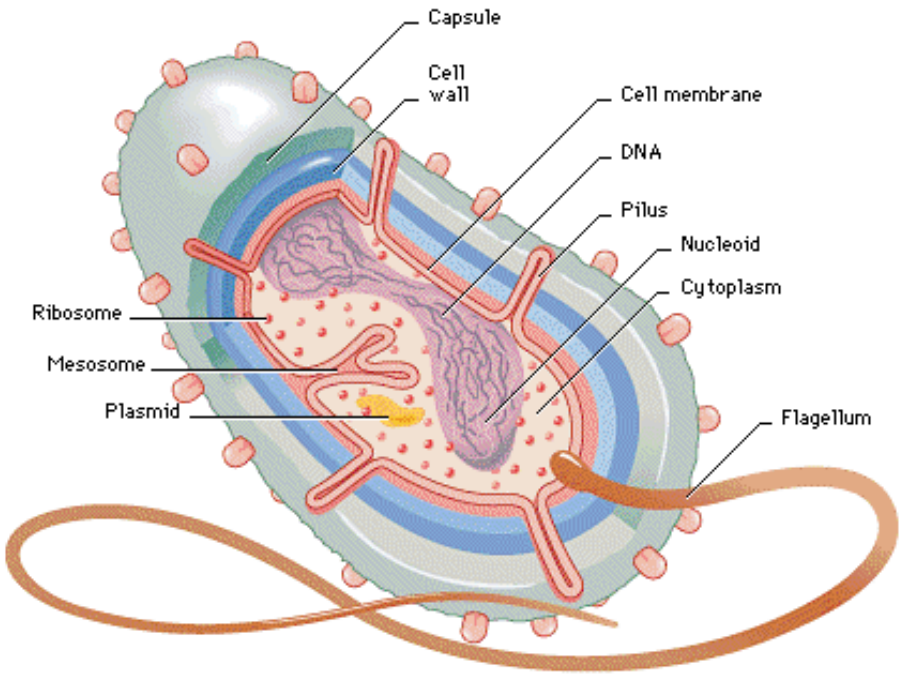


Figure 2.1. Cell Structure of a Bacteria (Mann, 1996).

Bacteria are widely distributed, occuring nearly everywhere. Some species are essential for the maintenance of life through their ability to break down plant and animal remains, others parasitize living plants and animals and in some instances cause serious disease by interference with normal form and function. They are found in all soils and moisturous places. Although they occur in air, it is not their natural home, as under ordinary conditions they can not grow and multiply in it. There are typically 40 million bacterial cells in a gram of soil and a million bacterial cells in a millilitre of fresh water. There are approximately 5×10^{30} bacteria on Earth, forming a biomass which exceeds that of all plants and animals (Goel, 2014). They are found on the surface but not on the inside of undamaged fruit and vegetables. All food, except that recently cooked, contains bacteria, the number and kind varying with the nature and the age of the food. Also they can be expected to be found on the surface of the skin and mucous membranes. When unstained, most bacteria are transparent, colorless and apparently homogeneous bodies. (Hiss *et al*., 1928).

2.1.3. Types of Bacteria

Many different characteristics are used in classifying and identifying bacteria (Table 2.1). These include general tests that are applied for virtually all bacteria, and very specialized tests that are used to identify specific bacteria strains. However there has been some difficulties in classification of bacteria. First of all, the small, simple structure of these microorganisms makes it tough to work out a satisfactory classification. Secondly, even in our modern world man’s knowledge of the characters of the bacteria is limited (Greaves *et al*., 1946).

Bacterial species have long been classified on the basis of their characteristic cell shapes (Figure 2.2). They are described as cocci if the cells are spherical and as rods if the cells are shaped like cylinders. Other shapes include curved rods, spirals and others. The arrangement pattern of the cells is also an important characteristic that distinguishes one type from other. For example, in *Escherichia coli*, the cells occur singly whereas other species have cells that occur as pairs, tetrads, chains or irregular clusters (Atlas, 1995).

Table 2.1. Examples of bacterial classification according to their different characteristics (Atlas, 1995).

Criterion	**Example**
Morphology	Cell shape, cell size, arrangement of flagella, capsule, endospores
Staining reactions	Gram stain, acid-fast stain
Growth and nutritional characteristics	Appearance in liquid culture, colonial morphology, pigmentation, energy sources, carbon sources, fermentation products, modes of metabolism (autotrophic, heterotrophic, fermentative, respiratory)
Biochemical characteristics	Cell wall constituents, pigment biochemicals, storage inclusions, antigens, RNA molecules
Physiological and ecological characteristics	Temperature range and optimum, oxygen relationships, pH tolerance range, osmotic tolerance, salt requirement, antibiotic sensitivity
Genetic characteristics	DNA mole % G + C, DNA hybridization

Staining reactions are used to characterize the bacteria. The gram stain procedure is the most widely used differential staining procedure in today's world. By the specific chemical composition of the cell wall, gram staining reaction is determined. Gram positive bacteria stain blue-purple by the gram staining procedure, whereas gram negative bacteria stain pink-red colored. Examples of gram positive and gram negative bacteria are shown in the Figure 2.3.

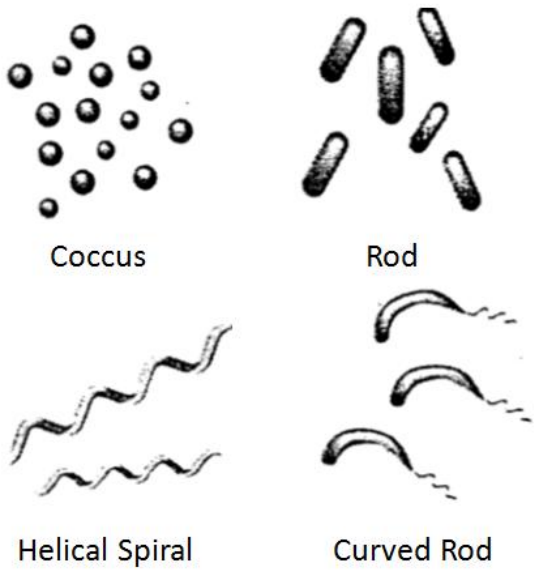


Figure 2.2. Bacterial Shapes.

Additionally, bacteria can be classified according to their relationship with oxygen. These are specified as aerobic bacteria and anaerobic bacteria. Aerobics are able to use oxygen, whereas anaerobic bacteria can sustain itself without the presence of oxygen. Moreover, genetic characteristics are employed in modern classification systems. In some clinical tests, species are identified using DNA hybridization. (Schaechter *et al*., 2006; Atlas, 1995).

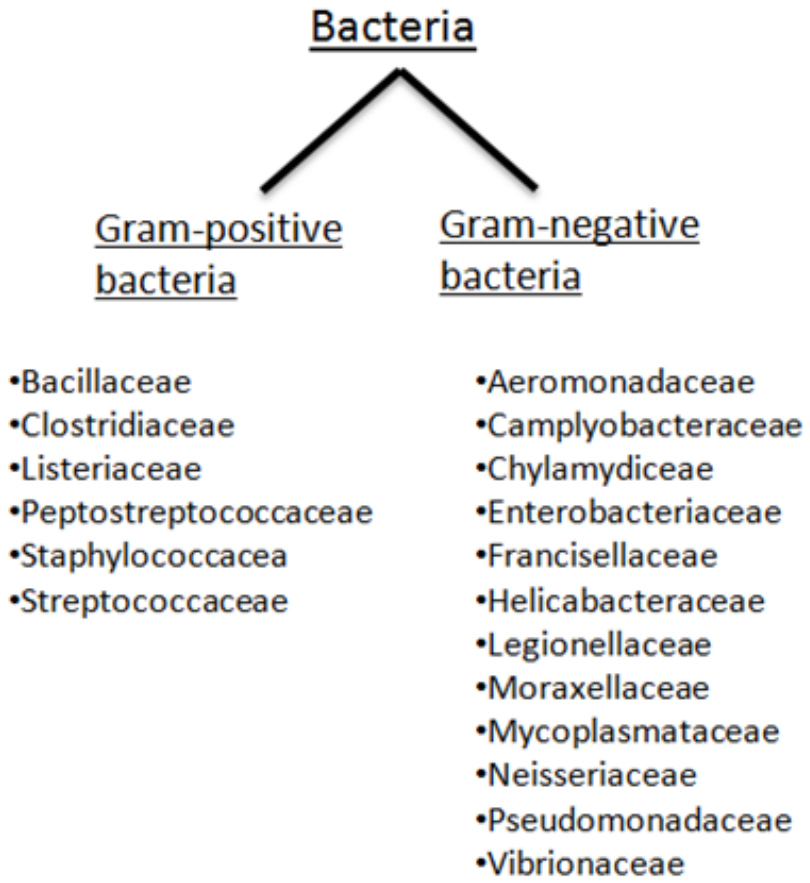


Figure 2.3. Types of Bacterial Families according to their Gram Staining Characteristics.

2.1.4. Bacterial Infections

The human body has an extensive population of microogranisms on the skin and the mucous membranes lining the mouth, gut, excretory, and reproductive system. These microorganisms are often beneficial and sometimes necessary to maintain good health. However, another group of microorganisms, such as bacteria, use direct and indirect means

to colonize, invade, and damage the human body, leading to infectious diseases (Madigan *et al*., 2006).

Infectious diseases which cause about 14 million human deaths annually, are caused by six types of pathogens: extracellular bacteria, intracellular bacteria, viruses, parasites, fungi and prions. Infection occurs when an organism successfully avoids innate defense and colonizes a niche in the body. What follows is a biological "horse race" in which the pathogen tries to replicate and expand its niche, while the immune system tries to eliminate the pathogen. Only if the replication of the pathogen causes detectable clinical damage, then the host experiences "disease". Microbial toxins released by pathogen can cause disease even in the absence of widespread colonization (Saunders *et al*., 2011). Infections and their bacterial pathogens with their mortality rates are shown in the Table 2.2.

Table 2.2. The infections caused by bacterial pathogens.

Bacterial Families	**Infectious diseases**	**Overall Mortality Rates**	**Reference**
Bacillaceae	Anthrax, food poisoning	Anthrax: 20%	Valcheva-Komitska *et al.*, 2007; WHO, 2012
Clostridiaceae	Gastro-intestinal disease, botulism, tetanus	Botulism: 5-10% Tetanus: 11%	DeFranco *et al.*, 2007; CDC, 1991
Listeriaceae	Food poiosoning, Listeriosis	Listeriosis: 20 %	DeFranco *et al*., 2007; Fitzpatrick *et al*., 2008.
Peptostreptococcaceae	Respiratory tract, intra-abdominal and subcutaneous infections	Respiratory tract:6.9 %	Finegold, 1977; Brook, 2007; Beaglehole *et al*., 2004; WHO, 2004.
Staphylococcacea	Severe skin, wound infection	Wound infection: 30%	DeFranco *et al*., 2007; Wyllie *et al*., 2006.
Streptococcaceae	Pneumonia, scarlet fever, pharyngitis	Pneumonia: 18% Scarlet fever: 1 %	DeFranco *et al.*, 2007; American Academy of Pediatrics, 2009.

Table 2.2. The infections caused by bacterial pathogens (cont.)

Bacterial Families	**Infectious diseases**	**Overall Mortality Rates**	**Reference**
Aeromonadaceae	Gastroenteritis and bacterial septicemia	Gastroenteritis: 1.4 million	Abbott *et al*., 2010; Elliott, 2007.
Campylobacteraceae	Gastroenteritis, diarrhea and periodontitis	Diarrhea: 15% (under the age 5)	Humphrey *et al*., 2007; Jassim *et al.*, 2011; WHO, 2009
Chylamydiceae	Blindness, Trachoma, sterility	Not available	Ryan *et al*., 2004.
Enterobacteriaceae	Bacteremia, lower respiratory tract infections, skin and soft tissue infections, urinary tract infections, endocarditis, intra-abdominal infections, septic arthritis, osteomyelitis, and ophthalmic infections	Enterobacteriaceae related diseases: 40% to 50%	CDC, 2013; Fraser *et al*., 2014.
Francisellaceae	Tularemia	15%	Penn, 2005; CFSPH, 2009.
Helicabacteraceae	Chronic gastritis, gastric cancer and gastric ulcers	Gastric cancer: 8.8%	Shiotani *et al*., 2002; Ferlay *et al*., 2013.
Legionellaceae	Legionellosis	5%-30%	Palusińska-Szysz *et al*., 2009; CDC, 2013.

Table 2.2. The infections caused by bacterial pathogens (cont.)

Bacterial Families	**Infectious diseases**	**Overall Mortality Rates**	**Reference**
Moraxellaceae	Otitis, sinusitis, or bronchitis	Sinusitis: 7%	Brook, 2007; Gallagher *et al*., 1998
Mycoplasmataceae	Atypical pneumonia, tuberculosis, leprosy	Pneumonia: 7%	William *et al*., 2006.
Neisseriaceae	Gonorrhea and meningitis	Gonorrhea: rare	DeFranco *et al*., 2007; Ohneck *et al*., 2011.
Pseudomonadaceae	Respiratory tract infections, soft tissue infections, urinary tract infections and pneumonia	7%	Conti *et al*., 1996
Vibrionaceae	Gastroenteritis, acute and fatal septicaemia, cholera	Gastroentreritis: 60-80% Cholera:58,000–130,000 deaths/year	Horseman *et al*., 2013; Manjunath, 2007; Lozano *et al*., 2012.

2.2. Human Immune System

Human beings and other mammals are continuously exposed to substances that are capable of causing them harm. As a consequence, they also have an amazing complex system of responses to attacks from outside the body and that is called the immune system. The immune system is compromised from the cells and molecules responsible for immunity, and the collective and coordinated response to the introduction of unfamiliar substances by these cells and molecules is called the immune response (Abbas, 2011). The term immunity is derived from the Latin word immunitas which referred to the protection from legal prosecution offered to Roman senators during their regimes. Therefore, historically, immunity mean protection from disease (Abbas *et al*., 2010).

The outcome of our understanding of the immune system and its functions has been remarkable since 1960s. Developments in biochemistry and genetics have changed immunology from largely descriptive science into one in which diverse immune phenomena can be explained by structural and biochemical terms (Abbas *et al*., 2010).

2.2.1. Innate and Adaptive Immunity

Immunity requires the recognition and elimination or containment of infectious organisms. This is obtained by two systems broadly classified as innate immunity and adaptive immunity (DeFranco *et al*., 2007).

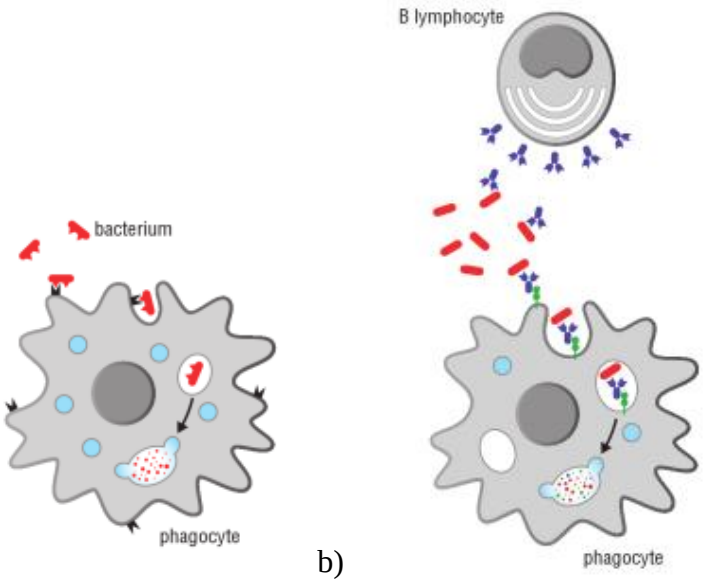
Innate immunity (also called natural or native immunity) provides the early line of defense against microbes which refers to the fact that this type of host defense is always present in healthy individuals. It is prepared to block the entry of microbes and to rapidly eliminate microbes that succeed in entering host tissues (Abbas, 2011). This type of immune system consists of cellular and biochemical defense mechanisms that distinguish host cells from those of infectious agents. Innate immunity specifically targets microbes and is a powerful early defense mechanism capable of controlling and even eradicating infections, thus it is essential for host defense and is responsible for early detection and containment of microbes (Abbas *et al*., 2010; Forst, 2006). The first line of its defense is provided by epithelial barriers and specialized cells and natural antibiotics present in epithelia. All of these function to block the entry of pathogen. The main components of innate immunity are physical and chemical barriers, phagocytic and natural killer (NK) cells, blood proteins and cytokines which are proteins that regulate and coordinate many activities of the innate immunity cells (Abbas *et al*., 2010). The defense begins in the skin and the epithelia of the respiratory, intestinal, urinary and reproductive tracts with antimicrobial peptides that are thought to be important for protection against bacterial infection. Innate immunity in the blood and tissues is provided principally by phagocytic cells that recognize surface components of bacteria and engulf them (DeFranco *et al*., 2007).

In contrast to the innate immunity, adaptive immunity develops as a response to infection and adapts to it. Unlike innate immune responses, the adaptive responses are

highly specific to the particular pathogen that induced them. They can also provide long-lasting protection (Alberts *et al*., 2002). The cells of adaptive immunity, which are called lymphocytes, express receptors that specifically recognize different substances, antigens, produced by microbes as well as noninfectious molecules. In addition it has an extraordinary capacity to distinguish between different microbes and molecules and for this reason it is also called specific immunity. Innate and adaptive immune responses are components of an integrated system of host defense in which numerous cells and molecules function cooperatively (Table 2.3).

Table 2.3. Characteristics and components of innate and adaptive immunity (Abbas, 2011).

	Innate	**Adaptive**
Characterisctics		
Specificity	For structures shared by groups of related microbes	For antigens of microbes and for nonmicrobial antigens
Diversity	Limited; germline-encoded	Very large;receptors are produced by somatic recombination of gene segments
Memory	None	Yes
Nonreactivity to self	Yes	Yes
Components		
Cellular and chemical barriers	Skin, mucosal epithelia; antimicrobial chemicals	Lymphocytes in epithelia; antibodies secreted at epithelial surfaces
Blood proteins	Complement, others	Antibodies
Cells	Phagocytes (macrophages, neutrophills), NK cells	Lymphocytes



a) b)

Figure 2.4. (a) Example of Innate Immune System (b) Example of Adaptive Immune System (DeFranco *et al*., 2007).

Examples of innate and adaptive immunity are shown in the Figure 2.4. A phagocyte of the innate immune system recognizes a conserved surface component of a bacterium, and ingests and destroys it (Figure 2.4-a). A lymphocyte of the adaptive immune system produces antibodies that recognize a variable component of the surface of a bacterium by means of a binding site that is itself highly variable. A non-variable portion of the antibody is then recognized by a receptor on the phagocyte, which is thereby activated to ingest the bacterium and destroy it. In this way the adaptive immune system enables microorganisms that have masked the conserved components recognized by innate immune cells to be destroyed by these cells (Figure 2.4-b) (DeFranco *et al*., 2007).

2.2.2. Cells of the Immune System

The cells of the immune system consist of lymphocytes, effector cells and antigen-presenting cells (APCs) (Table 2.4). Lymphocytes are the only cells in the body capable of specifically recognizing and distinguishing different antigenic determinants and are responsible for the two defining characteristics of the adaptive immune response, specificity and memory. All lymphocytes are morphogically similar and rather unremarkable in appearence, however they are extremely heterogenous in lineage, function and phenotype. They are also capable of complex biological responses and activities.

Protective immunity to microbes can be adoptively transferred from immunized to healthy individuals only by lymphocytes or their secreted products (Abbas *et al.,* 2010). Lymphocytes fall into two major classes; B lymphocytes and T lymphocytes. B lymphocytes are the only cells capable of producing antibodies. They also express membrane forms of antibodies that serve as the receptors which recognize antigens and initiate the process of activation of the cells. T lymphocytes, the mediators of cellular immunity, were named after their precursors which arise in the bone marrow. These precursors migrate to and mature in the thymus; which "T" refers to. The sites in which mature lymphocytes are produced are called the generative lymphoid organs and the mature B and T cells are called the naive lymphocytes. When naive lymphocytes recognize microbial antigens and also receive additional signals induced by microbes, the antigen-specific lymphocytes proliferate and differentiate into effector cells and memory cells, i.e. the differentiation of naive lymphocytes into effector cells and memory cells is initiated by antigen recognition (Alberts *et al*., 2002).

Effector cells of the immune system itself are the cells that eliminate microbes and they consist of lymphocytes and other leukocytes. The elimination of microbes often requires the participation of other nonlymphoid leukocytes, such as granulocytes and macrophages. These may function as effector cells in both adaptive and innate immunity (Abbas, 2011).

The common portals (skin, gastrointestinal tract, respiratory tract) of entry for microbes contain specialized antigen-presenting cells (APCs) located in the epithelium. These cells capture the microbes, display them to lymphocytes and provide signals that stimulate the proliferation and differentiation of the lymphocytes. By convention, APC usually refers to a cell that displays antigens to T lymphocytes. The major type of these cells that is involved in initiating T cell responses is the dendritic cell (formed or marked like dentrite of a branching form).

Table 2.4. Cell types of immune system and their functions (Abbas, 2011).

Cell Type	**Principal Function(s)**
Lymphocytes: B lymphocytes; T lymphocytes; natural killer cells	Specific recognition of antigens: **B lymphocytes**: mediators of humoral immunity **T lymphocytes**: mediators of cell-mediated immunity **Natural killer cells**: cells of innate imunity
Antigen-presenting cells: dentritic cells; macrophages;follicular dendritic cells	Capture of antigen for display to lymphocytes: **Dendritic cells**: initiation of T cell responses **Macrophages:** initiation and effector phase of cell-mediated immunity **Follicular dendritic cells**: display of antigens to B lymphocytesin humoral immune responses
Effector cells: T lymphocytes; macrophages; granulocytes	Elimination of antigens: **T lymphocytes**: helper T cells and cytotoxic T lymphocytes Macrophages and monocytes. cells of the mononuclear-phagocytes system Granulocytes: neutrophils, eosinophills

2.2.3. Tissues of the Immune System

The tissues of the immune system consist of the generative lymphoid organs, in which T and B lymphocytes mature and become competent to respond to antigens and the secondary lymphoid organs in which adaptive immune responses to microbes are initiated. Included in the generative lymphoid organs of adult mammals are the bone marrow, where all the lymphocytes arise and B cells mature, and the thymus, where T cells mature and reach a stage of fuctional competence (DeFranco *et al*., 2007).

The peripheral lymphoid organs which consist of the lymp nodes, the spleen, the mucosal and cutaneous immune systems, are organized to optimize interactions of antigens, APCs, and lymphocytes in a way that promotes the development of adaptive immune responses. The anatomic organization of peripheral lymphoid organs enables

APCs to concentrate antigens in these organs and lymphocytes to locate and respond to the antigens (Abbas, 2011).

2.2.4. Immunity to Extracellular Bacteria

Extracellular bacteria (*Clostridiaceae, Staphylococcacea, Streptococcaceae, Camplyobacteraceae, Helicabacteraceae, Mycoplasmataceae, Neisseriaceae, Pseudomonadaceae, Vibrionaceae*) (Table 2.9) are capable of replicating outside host cells, for example, in the circulation, in connective tissues and in tissue spaces such as the lumens of the airways and gastrointestinal tract (Table 2.5). The disease caused by pathogenic extracellular bacteria follow two principle mechanisms. First, these bacteria induce inflammation, which results in tissue loss at the site of infection. Second, toxins, which have diverse pathologic effects, are produced by extracellular bacteria. Such toxins may be endotoxins, which are components of bacterial cell walls, or exotoxins, which are actively secreted by bacteria. The endotoxin in gram- negative bacteria are called lipopolysaccharide (LPS) which increases the negative charge of the cell membrane and helps stabilize the overall membrane structure (Abbas *et al*., 2010). On the other hand, exotoxins are highly potent and can cause major damage to the host by destroying cells or disrupting normal cellular metabolism (Ryan *et al*., 2010).

Complement activation, phagocytosis and the inflammatory response are the principal mechanisms of innate immunity to extracellular bacteria. Gram positive bacteria have cell walls containing a thick layer of peptidoglycan which stays purple colored after Gram staining. On the other hand, LPS in the cell walls of gram-negative bacteria stays red colored after Gram staining. Peptidoglycan and LPS activate the alternative pathway of complement activation in the absence of antibody. One result of complement activation is opsonization (rendering of bacteria and other cells subject to phagocytosis) and enhanced phagocytosis (the process by which certain cells of the innate immune system, including macrophages and neutrophils, engulf large particles such as intact microbes) of bacteria. Phagocytes use various surface receptors to recognize appropriately opsonized bacteria.

Table 2.5. Examples of extracellular bacteria and their infections (Abbas, 2011).

Pathogen	**Examples of human diseases**	**Mechanism of pathogenicity**
Staphylococcus aureus	Skin and soft tissue infections, lung abscess; systemic: toxic shock syndrome, food poisoning	Skin infections; acute inflammation induced by toxins; cell death caused by pore-forming toxins Systematic: enterotoxin ("superantigen")-induced cytokine production by T cells causing skin necrosis,shock,diarrhea
Streptococcus pyogenes (group A)	Pharyngitis, Skin infections: impetigo, erysipelas; cellulitis, Systemic: scaret fever	Acute inflammation induced by various toxins,e.g., streptolysin O damages cel membranes (antiphagocytic action of capsular polysaccharides).
Streptococcus pyogenes (pneumococcus)	Pneumonia, meningitis	Acute inflammation induced by cell wall constituents; pneumolysin is similar to streptoysin O.
Escherichia coli	Urinary tract infections, gastroenteritis, septic shock	Toxins act on intestinal epithelium and cause increased chloride and water secretion; endotoxin (LPS) stimulates cytokine secretion by macrophages.
Vibrio cholerae	Diarrhea (cholera)	Cholera toxin ADP ribosylates G protein subunit,which leads to increased cyclic AMP in intestinal epithelial cells and results in chloride secretion and water loss.
Clostridium tetani	Tetanus	Tetanus toxin binds to the motor end-plate at neuromuscular junctions and causes irreversible muscle contraction.
Neisseria meningitidis (meningococcus)	Meningitis	Acute inflammation and systemic disease caused by potent endotoxin
Corynebacterium diphtheriae	Diphteria	Diptheria toxin ADP ribosylates elongation factor-2 and inhibits protein synthesis

Opsonized bacteria is the foreign pathogen in the human body, and a floating antigen is attached to the protein on the bacteria so that a phagocyte can easily recognize the bacteria in order to engulf and destroy it. Toll-like receptors (TLRs) of phagocytes participate in activation of phagocytes as a result of encounter with microbes. These TLRs are an evolutionarily conserved family of pattern recognition receptors expressed on many cell types, which play essential roles in innate immune responses to microbes. Some of the microbial products that stimulate TLR signals include gram negative bacterial LPS, gram positive bacterial peptidoglycan, bacterial lipoproteins, lipoteichoic acid, lipoarabianomannan, zymosan, the bacterial flagellar protein flagellin and respiratory syncytial virus fusion protein (Abbas, 2011).

Humoral immunity is the principal protective immune response against extracellular bacteria and it functions to block infection, eliminate the microbes and neutralize their toxins. Because extracellular bacteria can not routinely hide within host cells, antibodies are generally highly effective to these species. Antibody responses against extracellular bacteria are directed against cell wall antigens, and secreted and cell-associated toxins, which may be polysaccharides or proteins. The polysaccharides are prototypic thymus-independent antigens and major function of humoral immunity is defense against polysaccharide-rich encapsulated bacteria. They are also perfect for Ti antigens for B cell activation while other bacterial components supplying Td antigens induce primarily a Th2 response that provides T cell help for antibacterial B cells. Neutralization is mediated by high-affinity immunoglobulin G (lgG) and lgA isotypes, opsonization by some subclasses of lgG, and complement activation of lgM and subclasses of lgG. Smaller lgG antibodies, which protect the tissues, neutralize bacteria by physically preventing them from attaching host cell surfaces. The protein antigens also activate CD4+ helper T cells, which produce cytokines that stimulate antibody production, induce local inflammation and enhance the phagocytic and microbial activities of macrophages and neutrophils (Figure 2.5) (Abbas, 2011; Saunders, 2011). Interferon-α (IFN-α) is the T cell cytokine responsible for macrophage activation, and tumor necrosis factor (TNF) and lymphotoxin trigger inflammmation (Schroder *et al.*, 2004).

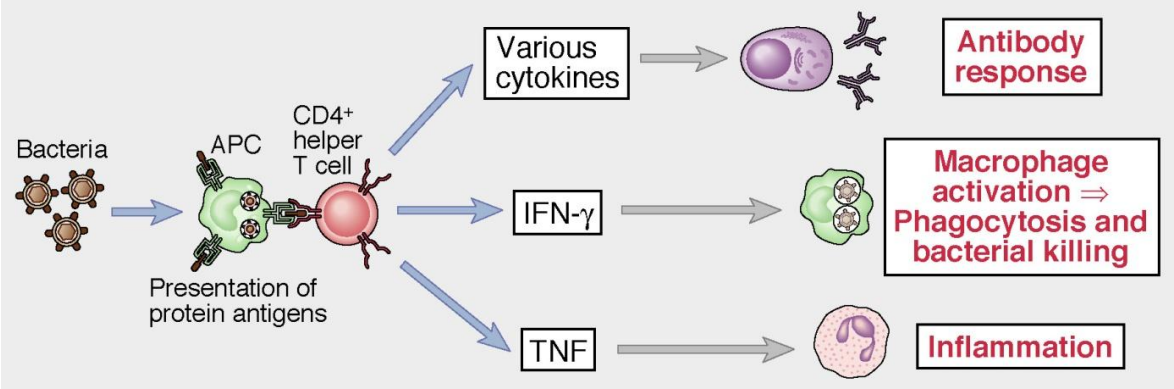


Figure 2.5. The Extracellular Immunity Mechanism (Abbas, 2011).

The principal injurious effects of host responses to extracellular bacteria are inflammation and septic shock. Septic shock is a medical condition as a result of severe infection and sepsis, though the microbe may be systemic or localized to a particular site. It can cause multiple organ disfunction syndrome (formerly known as multiple organ failure) and death. Its most common victims are children, immunocompromised individuals, and the elderly, as their immune systems cannot deal with the infection as effectively as those of healthy adults. Frequently, patients suffering from septic shock are cared for in intensive care units. The mortality rate from septic shock is approximately 25-50% (Kumar *et al*., 2007). Bacterial sepsis and septic shock result from the overproduction of inflammatory mediators as a consequence of the interaction of the immune system with bacteria and bacterial wall constituents in the body (Berkel *et al*., 2003). It is believed that Gram negative and Gram positive bacteria may activate a common pathway of events that lead to septic shock (Verhoef *et al*., 2003). The early phase of septic shock is caused by cytokines produced by macrophages that are activated by microbial components, particularly LPS which is a major pathogen-associated molecular pattern that is recognized by host Toll-like receptor 4 (TLR4), inducing innate immune responses including inflammation (Rhen *et al*., 2003). Recent studies suggest that TLRs, inflammatory cytokines, free radicals, macrophage migration inhibitory factor, signal protein kinases, and transcription factors, all play an important role in the pathobiology of gram negative mediated septic shock (Das, 2000; Guha *et al*., 2001). Gram positive bacteria can also induce septic shock and are increasingly recognised as major contributers to nosocomial sepsis. Although they do not have endotoxin, the presence of these bacteria in tissues provokes an inflammatory response that is similar to that triggered by Gram negative LPS.

Bacteria frequently implicated in septic shock include Gram negative bacteria such as *Escherichia coli, P aeruginosa,* and *Meningococci,* and Gram positive bacteria such as *Staphylococcus aureus, Staphylococcus epidermidis,* and *Streptococci* (Walker, 1998).

The virulence of extracellular bacteria has been associated with some mechanisms that resist to innate immunity (Table 2.6) (Abbas, 2011). The major mechanisms used by these bacteria to evade humoral immunity is genetic variation of surface antigens. Some surface antigens of bacteria such as gonococci are contained in their pili, which are the structures responsible for bacterial adhesion to host cells. The ability to alter antigens helps bacteria to evade the attack of pili specific antibodies.

Table 2.6. The mechanism of immune evasion of extracellular bacteria (Abbas, 2011).

Mechanism of immune evasion	**Examples**
Antigenic variation	*Neisseria gonorrhoeae, Escherichia coli, Salmonella typhimurium*
Inhibition of complement activation	Many bacteria
Resistance to phagocytosis	*Pneumococcus*
Scavenging of reactive oxygen intermediates	Catalase-positive *staphylococci*

2.2.5. Immunity to Intracellular Bacteria

A characteristic of intracellular bacteria (*Listeriaceae, Chylamydiceae, Enterobacteriaceae, Francisellaceae, Legionellaceae*) (Table 2.9) is their ability to survive and even replicate within phagocytes. The pathological consequences of infections by many intracellular bacteria are due to the host response to these microbes (Table 2.7). Similar to extracellular bacteria, most intracellular bacteria enter the host via breaches in the mucosae and skin, but some are introduced directly into the bloodstream by the bites of vectors. Once inside the host, intracellular bacteria elude phagocytes, complement activation and antibodies by moving right inside host cells to reproduce (Saunders *et al*., 2011). Intracellular pathogens often reside within specialized compartments and might evade or minimize extracellular innate immune signaling, raising the questions of whether infected host cells can distinguish intracellular bacteria from extracellular bacteria, and if such a distinction would have immunological consequences (O'Riordan *et al*., 2002). More

evidence for intracellular bacterial recognition was provided by Brumell and co-workers (Perrin *et al*., 2004) who showed that, upon entry of bacteria into the cytoplasm, the bacterial surface becomes decorated by ubiquitin. To date, it is unclear which bacterial ligands or host proteins trigger the process of ubiquitylation, which proteins are targeted for modification or, at the systemic level, what function ubiquitylation of bacteria has in the innate immune response.

Table 2.7. Examples of intracellular bacteria and their infections (Abbas, 2011).

Pathogen	**Examples of human diseases**	**Mechanisms of pathogenicity**
Mycobacteria	Tuberculosis, leprosy	Macrophage activation resulting in granulomatous inflammation and tissue destruction
Listeria monocytogenes	Listeriosis	Listeriolysin damages cell membranes
Legionella pneumophila	Legionnaires' diseases	Cytotoxin lyses cells and causes lung injury and inflammation

The innate immune response to intracellular bacteria is mediated mainly by phagocytes and natural killer (NK) cells. Phagocytes, initially neutrophils and later macrophages, ingest and attempt to destroy these microbes, but pathogenic intracellular bacteria are resistant to degradation within phagocytes. In addition to phagocytosis, macrophages can carry out TLR-mediated endocytosis of intracellular bacteria. Once activated by TLR engagement, the macrophages produce pro-inflammatory cytokines to promote NK cell activation and Th1 differentiation. Intracellular bacteria activate NK cell by inducing expression of NK cell-activating ligands on infected cells or by stimulating dendritic cells. The NK cells secrete IFN-γ which promotes macrophage activation directly (Th1 activation indirectly) and promotes killing of the phagocytosed bacteria. In other words, NK cells provide an early defense against these pathogens before the development of adaptive immunity (Saunders, 2011; Abbas, 2011).

The major protective immune response aginst intracellular bacteria is T-cell mediated immunity which is the effector function of T lymphocytes and serves as the defense mechanism against pathogens that survive and replicate within phagocytes and nonphagocytic cells. Cytotoxic T-lymphocytes (CTLs) are critial for resolving many intracellular bacterial infections. If the bacterium replicates in the cytosol of the infected cell, some of its component proteins enter the endogenous antigen processing pathway. CD4+ T cells make a significant contribution to defense against intracellular bacteria. They respond to protein antigens of phagocytosed microbes with CD8+ T cells. Phagocytosed bacteria stimulate CD8+ T cell responses if bacterial antigens are transported from phagosomes into the cytosol or if the bacteria escape from phagosomes and enter the cytoplasm of infected cells.

Antibodies can make an important contribution to host defense against at least some intracellular bacteria. The antibody-bound bacteria are unable to enter host cells and are eliminated by opsonized phagocytosis or classical complement-mediated lysis. The macrophage activation that occurs in response to intracellular microbes is also capable of causing tissue damage. This damage may be the result of delayed type hypersensitivity (DTH) reactions to microbial protein antigens (Saunders, 2011; Abbas, 2011).

Intracellular bacteria have the capacity to survive and replicate inside mononuclear phagocytes (MP) and, sometimes, within certain other host cells. MP are potent effector cells that are able to engulf and kill many bacterial invaders. Therefore, intracellular bacteria had to exploit potent evasion mechanisms that allow their survival in this hostile environment (Kaufmann, 1993).

Table 2.8. The mechanism of immune evasion of intracellular bacteria (Abbas *et al*., 2010).

Mechanism of immune evasion	**Examples**
Inhibition of phagolysosome formation	*Mycobacterium tuberculosis, Legionella pneumophila*
Inactivation of reactive oxygen and nitrogen species	*Mycobacterium leprae (*phenolic glycolipid*)*
Disruption of phagosome membrane, escape into cytoplasm	*Listeria monocytogenes (*hemolysin protein*)*

Different intracellular bacteria have developed various strategies to resist destruction by phagocytosis (Table 2.8). These include avoiding phagosomal destruction and antibodies, escaping into the cytosol, directly scavenging or inactivating microbicidal substances.

Table 2.9. Bacterial families according to their Gram positive/Gram negative and extracellular/intracellular properties.

Bacterial Families	**Gram positive/Gram negative**	**Extracellular/Intracellular**
Bacillaceae	Gram positive	Not Available
Clostridiaceae	Gram positive	Extracellular
Listeriaceae	Gram positive	Intracellular
Peptostreptococcaceae	Gram positive	Not Available
Staphylococcacea	Gram positive	Extracellular
Streptococcaceae	Gram positive	Extracellular
Aeromonadaceae	Gram negative	Not Available
Camplyobacteraceae	Gram negative	Extracellular
Chylamydiceae	Gram negative	Intracellular
Enterobacteriaceae	Gram negative	Intracellular
Francisellaceae	Gram negative	Intracellular
Helicabacteraceae	Gram negative	Extracellular
Legionellaceae	Gram negative	Intracellular
Moraxellaceae	Gram negative	Not Available
Mycoplasmataceae	Gram negative	Extracellular
Neisseriaceae	Gram negative	Extracellular
Pseudomonadaceae	Gram negative	Extracellular
Vibrionaceae	Gram negative	Extracellular

2.2.6. Vaccine Development

Infectious diseases have been a serious threat for human life since the ancient times. The function of immune system and immunology is the defense against infectious pathogens. In today's modern era, immunology is an experimental science, in which explanations of immunologic phenomena are based on experimental observations and

conclusions drawn from them. The evolution of immunology as an experimental discipline has depended on man's ability to manage the function of the immune system under controlled conditions (Abbas, 2011). It is mostly believed among historians that some kind of inoculation was developed in India or China before the 16th century (Lombard *et al*., 2007). Chinese customs suggest that using powders made from the skin lesions of patients who are infected with smallpox, make children resistant to the disease (Lund *et al*., 2005). The most outstanding manipulation of scientists is the successful vaccination of Edward Jenner which was against smallpox in 1796 (Henderson *et al*., 2003). He recognized that milkmaids who had recovered from cowpox, never caught more serious smallpox. Hence, he injected the material from cowpox lesion into a child's arm. After this child was inoculated with smallpox, he did not develop the disease (Abbas, 2011). The history of vaccination began with Louis Pasteur. In 1881, he framed the hypothesis that pathogens could be attenuated by the exposure to environmental factors such as high temperature, oxygen and chemicals (Plotkin, 2005).

Vaccination is the most effective method for protecting individuals against infections in order to stimulate immune responses against microbes. A succesful vaccine needs to mimic as close as possible the real biological entity from which it is derived in order to be recognized by the host immune system as real danger (Buonaguro *et al*., 2011). If the infectious agent does not establish latency, if it does not undergo much or any antigenic variation and if it does not interfere with the host immune response response, then vaccines are effective (Abbas, 2011).

In the 20th century, the outcome of vaccinology has been outstanding. Surely, vaccines have been the most considerable tools for preventing infection, dissability and death. Although there are enormous number of successes of vaccines worldwide, continual frustration has resulted from the fact that there are still millions of diseases uncontrolled by vaccination (Fauci, 2001). Fortunately, advanced technologies and new tools allow for the development of new vaccines against microbes.

2.3. Host- Pathogen Interaction

Despite extended technological advances in medicine, pathogenic organisms, which cause infectious diseases worldwide, remain the source of mortality. While traditional biological research, which isolates and studies small sets of components, may provide some insights, these approaches are not well suited to address interaction mechanisms on a larger and more general scale. To this end, a systems biology approach is an emerging strategy to better comprehend the underlying mechanisms that occur during host pathogen interactions (Wang *et al*., 2013).

Most of the terminology used to define the host-pathogen interactions, which are enormously complex processes, has been in use for nearly a century. In the earliest studies, pathogens were thought to be primary aggressors that governed the host-pathogen interaction, resulting in disease. Later, new information about the behaviors of pathogen and their hosts resulted in the understanding that the pathogen-host interaction does not always result in disease. The knowledge on host-pathogen interactions and the relationship between them thoroughly, led to the introduction of terms to explain states in which microbes exist within hosts without causing disease and why some microbes only cause disease in certain hosts. (Casadevall *et al*., 2000). Each component of this relationship is in constant interaction with the other. Once established onto or into a particular body site of the host, microorganisms develop a particular relationship with that host. The relationship between host-pathogen may be one of the symbiosis, commensalism, or parasitism. Symbiosis is defined as the long-term relationship between two species which are called symbionts. An association in which both organisms apparently benefit is described as mutualism whereas a state of infection that results in either no damage or clinically inapparent damage to the host, though it can elicit an immune response is commensalism. In parasitism, one organism-the parasite-benefits, and the host is unfavorably affected (Mahon *et al*., 2007).

The outcome of host-pathogen interactions is effected by numerous factors. The status of the host's immune system and ability of the host to defend itself from microbial invasion, combined with microbial factors inherent to the invading organism, often determine whether disease will occur. Knowledge and understanding of the pathogen-host

relationship is important to appreciate and understand the concepts involved in the pathogenesis of infectious diseases (Mahon *et al*., 2007).

The mechanism by which a pathogen is able to invade a host cell is the main aspect of any pathogen-host system. Within these complex systems, protein-protein interactions (PPIs) between surface proteins form the foundation of communication between a host and a pathogen and play a vital role in initiating infection. These interactions allow the pathogen to enter the host cell, manipulate important cellular processes, multiply, and invade other cells (Dyer *et al*., 2008; Durmus Tekir *et al.,* 2012).

The very first prokaryotic PPI map was built for Helicobacter pylori (Rain *et al*., 2001). Other large-scale prokaryotic networks eventually were developed for *Campylobacter jejuni* (Parrish *et al*., 2007), *Treponema pallidum* (Titz *et al*., 2008), *Mycobacterium tuberculosis* (Wang *et al*., 2010) and *Bacillus subtilis* (Marchadier *et al*., 2011). Genome-scale analysis of interacting proteins that assemble into protein complexes were performed for *E. coli* (Butland *et al*., 2005; Arifuzzaman *et al*., 2006) and *Mycoplasma pneumoniae* (Kühner *et al*., 2009).

In recent years, large-scale protein-protein interaction (PPI) networks, both interspecies and intraspecies, have been determined (Table 2.9) (Durmuş Tekir *et al.,* 2011). The published PPI networks of humans and some model organisms provide valuable references for investigating protein interaction networks between pathogens and their hosts (LaCount *et al*., 2005; Li *et al*., 2004; Parrish *et al*., 2007; Rual *et al.,* 2005; Stelzl *et al.,* 2005). However, up to very recent years bacterial PHI data have been very scarce. Nearly all of the 10.477 human-pathogen protein-protein interactions are for viral systems (98.3%), with the majority belonging to the human-HIV systems (77.9%). These human-pathogen PPIs involve 1.233 unique human proteins, of which 1.109 are known to interact with at least one other human protein. Of these 1.233 human proteins, 221 interact with at least two pathogen groups (182 with more than one viral pathogen and 20 with more than one bacterial pathogen) (Dyer *et al*., 2008).

Table 2.10. Large-scale PPI networks (Durmuş Tekir *et al*., 2011).

Pathogen name	**Pathogen type**	**# of PPIs**	**References**
H. Pylori	Gram- bacteria	1280	Rain *et al.*, 2001
E. coli	Gram- bacteria	716	Butland *et al.*, 2005
E. coli	Gram- bacteria	11511	Arifuzzaman *et al.*, 2006
C. jejuni	Gram- bacteria	11687	Parrish *et al.*, 2007
T. pallidum	Gram- bacteria	3649	Titz *et al.*, 2008
M. pneumoniae	Bacteria without cell wall	178	Kühner *et al.*, 2009
M. tuberculosis	Bacteria without cell wall	8042	Wang *et al.*, 2010
B. subtilis	Gram+ bacteria	793	Marchadier *et al.*, 2011

Dyer *et al*. (2010) revealed the first high-throughput experimental study to produce bacterial PHI data which constitute the first extensive protein interaction networks constructed for bacterial pathogens (*Bacillus anthracis, Francisella tularensis,* and *Yersinia pestis*) and their human hosts (Dyer *et al*., 2010). Typically, data detailing pathogen-host interactions is ascertained from small-scaled experiments that are designed to target specific proteins, complexes, or pathways of interest. This is evident from the number of interactions between host and bacterial pathogens currently available in seven public resources. Then another high-throughput experimental study generating PHI data of Yersinia pestis was reported (Yang *et al*., 2011). Recently Durmus Tekir *et al*. (2012) analyzed 23.435 interactions between 3419 proteins of viral, bacterial, protozoan and fungal pathogens (totally 257 strains) and 5210 proteins of human obtained from the database PHISTO.

3. MATERIALS AND METHODS

3.1. Bacteria- Human PHI Data

The bacteria-human PHIs were downloaded from PHISTO in January 2014. PHISTO is a pathogen-human interaction search tool which serves as an up-to-date and functionally-enhanced source of comprehensive PHI data through a user-friendly interface. In PHISTO, the pathogens, for which experimental PHI data are available, are presented as: Pathogen type → Family → Species → Strain (Durmuş Tekir *et al*., 2013). The downloaded family-based bacteria-human PHIs include 18 bacterial families. 6 of these families are Gram positive bacteria while 12 of them are Gram negative bacteria. PHI data belonging to these Gram negative and Gram positive bacterial families (Tables 3.1 and 3.2) were analyzed comparatively.

Table 3.1. Contents of Gram negative bacteria-human PHI data.

Gram-negative bacteria family	**# of strains**	**# of PHIs**	**# of pathogen proteins**	**# of human proteins**
Aeromonadaceae	1	1	2	2
Camplyobacteraceae	1	4	1	3
Chylamydiceae	2	20	3	21
Enterobacteriaceae	15	4455	1304	2244
Francisellaceae	2	1341	346	988
Helicabacteraceae	3	5	4	5
Legionellaceae	1	1	1	1
Moraxellaceae	1	1	1	1
Mycoplasmataceae	1	1	1	1
Neisseriaceae	1	17	2	17
Pseudomonadaceae	1	14	3	10
Vibrionaceae	1	2	2	2
TOTAL	30	5858	1669	2822

Table 3.2. Contents of Gram positive bacteria-human PHI data.

Gram positive bacteria family	**# of strains**	**# of PHIs**	**# of pathogen proteins**	**# of human proteins**
Bacillaceae	2	3182	943	1751
Clostridiaceae	2	45	10	6
Listeriaceae	2	7	6	5
Peptostreptococcaceae	2	7	2	5
Staphylococcacea	4	16	13	14
Streptococcaceae	4	12	11	7
TOTAL	16	3270	985	1778

3.2. Human Protein Sets

A total of 16 sets of human proteins interacting with bacteria proteins were constructed from PHI data downloaded from PHISTO to analyze the properties of targeted human proteins in comparison to non-targeted ones (Table 3.3) as follows: the sets targeted by Gram positive bacteria (Gram positive bacteria–targeted set), Gram negative bacteria (Gram negative bacteria-targeted set), only Gram positive bacteria, not targeted by any Gram negative bacteria (only Gram positive bacteria-targeted set) and only Gram negative bacteria, not targeted by any Gram positive bacteria (only Gram negative bacteria-targeted set). These sets were constructed to observe the attacking characteristics specific to Gram negative and Gram positive bacteria. For a deeper comparison, human proteins interacting with two Gram positive bacteria families (two Gram positive bacteria-targeted set) and three Gram positive bacteria families (three Gram positive bacteria-targeted set) and additionally, human proteins interacting with two Gram negative bacteria families (two Gram negative bacteria-targeted set) and three Gram negative bacteria families (three Gram negative bacteria-targeted set) were analyzed. The sets of human proteins interacting with both Gram positive and Gram negative bacteria together (Gram positive-Gram negative bacteria–targeted set) was used to obtain the common infection strategies of bacterial pathogens. Furthermore, the sets of human proteins targeted by bacterial families with large-scale PHI data were analyzed to investigate the human mechanism attacked by each bacterial family in the PHI data.

Table 3.3. Number of total human proteins targeted by bacterial families.

Human protein sets	**Number of Total Human Proteins targeted**
All Bacteria-targeted set	3698
Gram Negative bacteria-targeted set	2822
Gram Positive bacteria-targeted set	1778
Only Gram Negative bacteria-targeted set	1920
Only Gram Positive bacteria-targeted set	876
Two Gram Negative bacteria-targeted set	452
Three Gram Negative bacteria-targeted set	9
Two Gram Positive bacteria-targeted set	6
Three Gram Positive bacteria-targeted set	2
Both Gram Positive and Gram Negative bacteria – targeted set	902
Six Gram Positive bacteria families-targeted set	1778
Five Gram Negative bacteria families-targeted set	2809
Enterobacteriacea-Bacillaceae-Francisellaceae bacteria-targeted set	3645
Enterobacteriaceae-targeted set	2244
Bacillaceae–targeted set	1751
Francisellaceae-targeted set	988

3.3. GO Enrichment Analysis

Enriched GO (Ashburner *et al*., 2000) terms of all 16 human protein sets were found using BiNGO plugin (ver. 2.44) of Cytoscape (ver. 2.8.1) (Maere *et al*., 2005). Significance level was set to 0.05 which means that only terms enriched with a p-value of at most 0.05 were considered. The Gene Ontology (GO) is used to address the consistent descriptions of gene products in different databases. It consists of three hierarchically structured terms that describe gene products with respect to their associated biological processes, molecular functions and cellular components.

3.4. KEGG Pathway Analysis

KEGG (Ogata *et al*., 1999) Pathway Analysis of all 16 human protein sets were conducted using WebGestalt (Zhang *et al*., 2005) software. EntrezGene was selected as the reference set which is needed to perform statistical analysis to identify enriched gene sets. KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database resource that integrates genomic, chemical, and systemic functional information. It is widely used as a reference knowledge base for integration and interpretation of large-scale datasets generated by genome sequencing and other high-throughput experimental data.

3.5. Pathway Commons Analysis

Coupled with KEGG Pathway Analysis, Pathway Commons Analysis of 15 sets of human proteins targeted by bacterial pathogens were performed using WebGestalt (Zhang *et al*., 2005). Pathway Commons is a database which has the collection of publicly available pathway information from multiple organisms. It currently includes over 1400 pathways and 687 000 interactions from sources as follows: BioGRID (Breitkreutz *et al*., 2008), HPRD (Keshava *et al*., 2009), HumanCyc (Romero *et al*., 2005), IntAct (Aranda *et al*., 2010), MINT (Ceol *et al*., 2009), NCI/Nature PID (Schaefer *et al*., 2009) and Reactome (Matthews *et al*., 2009). These information can be usefully combined with high-throughput genomic data and clinical phenotype data to investigate the network properties of specific disease types and to build classifiers for disease subtypes (Cerami *et al*., 2011).

3.6. Transcription Factor Targets Analysis

Eukaryotes have developed a complex way of controlling expression of their existing genes via a group of proteins known as transcription factors (TFs) (Phillips *et al*., 2008). This system controls which genes are turned on or off in the genome by binding to DNA and other proteins. So a transcription factor is a protein that binds to specific DNA sequences; however transcription factor targets (TFTs) are these types of DNA sequences. TFTs are specific sets of genes that share a common TF-binding site defined in the TRANSFAC database (Wingender, 2008). TFT analysis is conducted to find out whether the genes encoding the targeted human proteins have the same DNA sequences. If they do

contain the same DNA sequences, it means that input proteins may be regulated by the same TFs. Using WebGestalt (Zhang *et al*., 2005), the web-based gene set analysis toolkit, TFTs of targeted human protein sets are estimated. This tool incorporates information from different public resources such as NCBI Gene, GO, KEGG and MsigDB. All available TFTs are collected in the Molecular signature Database (MsigDB) (Liberzon *et al*., 2011) and are restored in WebGestalt. EntrezGene was again selected as the reference set to perform statistical analysis. The results were evaluated by the hypergeometric test using the seven most enriched terms with maximum significance level or p-value of 0.05.

4. RESULTS AND DISCUSSIONS

4.1. All Gram Positive Bacteria versus All Gram Negative Bacteria

Bacteria-human PHI data downloaded from PHISTO reveal that Gram negative bacteria target 1.5 times more human proteins than Gram positive bacteria, resulting in nearly two times greater number of PHIs (Table 4.1).

Table 4.1. Comparison of bacterial PHI contents: Gram positive bacteria versus Gram negative bacteria.

Bacteria Type	**# of strains**	**# of PHIs**	**# of pathogen proteins**	**# of human proteins**
Gram Positive Bacteria	16	3270	985	1778
Gram Negative Bacteria	30	5858	1669	2822
TOTAL	46	9035	2654	3698

Regarding to the ratio of the strain numbers of Gram negative to Gram positive bacteria, the number of the PHIs and targeted human proteins seems to be linear. In addition, two (*Enterobacteriaceae* and *Francisellaceae*) of the three bacterial families, which have large-scale data, belong to Gram negative bacteria families. These families have more than 98% of PHIs and targeted human proteins.

4.1.1. Targeted Human Proteins

The distribution of 3698 bacteria-targeted human proteins according to their attacking bacteria types is shown in the Venn diagram (Figure 4.1). About 25% of these human proteins (902) are commonly targeted by both Gram negative and Gram positive bacteria. In the PHI data analyzed, one of the important observations about these common 902 human proteins is that none of them is targeted by more than one Gram positive and more than one Gram negative bacteria.

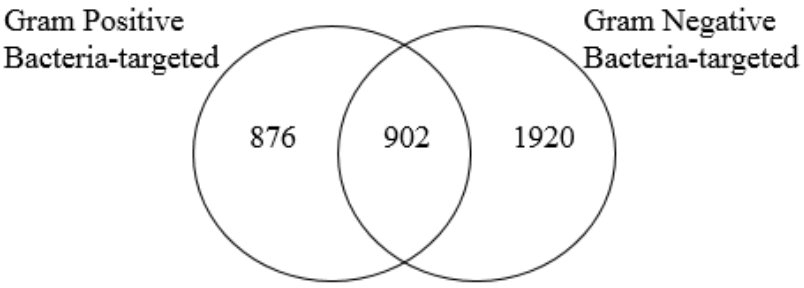


Figure 4.1. The Number of Human Proteins Targeted by Gram Negative Bacteria and/or Gram Positive Bacteria.

Considering 3698 human proteins, the most targeted ones are listed in the Table 4.2 according to their attacking gram positive and gram negative bacterial families. Especially TRAF6 (TNF receptor-associated factor 6), TLR4 (Toll-like receptor 4), IRAK2 (Interleukin-1 receptor-associated kinase-like 2), IRAK1 (Interleukin-1 receptor-associated kinase 1), MAP3K1 (Mitogen-activated protein kinase kinase kinase 1, E3 ubiquitin protein ligase), TLR1 (Toll-like receptor 1), PELI1 (E3 ubiquitin-protein ligase pellino homolog 1), MYD88 (Myeloid differentiation primary response protein), PELI2 (E3 ubiquitin-protein ligase pellino homolog 2), IRAK4 (Interleukin-1 receptor-associated kinase 4), PLMN (Plasminogen) and other following human proteins may have the potential to give important insights about infection strategies.

Table 4.2. Human proteins targeted by two Gram positive bacteria families and three Gram negative bacteria families.

Protein	**Attacking Gram Positive bacterial families**	**Attacking Gram Negative Bacterial families**
TRAF6- TNF receptor-associated factor 6	-	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae, Helicobacteriacea*
IRAK2- Interleukin-1 receptor-associated kinase-like 2	-	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*

Table 4.2. Human proteins targeted by two Gram positive bacteria families and three Gram negative bacteria families (cont.)

Protein	**Attacking Gram Positive bacterial families**	**Attacking Gram Negative Bacterial families**
TLR4- Toll-like receptor 4	-	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
IRAK1-Interleukin-1 receptor-associated kinase 1	-	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
MAP3K1- Mitogen-activated protein kinase kinase kinase 1, E3 ubiquitin protein ligase	-	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
TLR1-Toll-like receptor 1	-	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
PELI1- E3 ubiquitin-protein ligase pellino homolog 1	-	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
MYD88- Myeloid differentiation primary response protein	-	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
PELI2- E3 ubiquitin-protein ligase pellino homolog 2	-	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
IRAK4- Interleukin-1 receptor-associated kinase 4	-	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
COL1A2-Collagen alpha-2(I) chain	*Bacillaceae, Streptococcaceae*	*Franciselaceae*
1GHG1-Ig gamma-1 chain C region	*Bacillaceae, Streptococcaceae, Staphylococcacea*	*Enterobacteriaceae, Francisellaceae*
PLMN- Plasminogen	*Bacillaceae, Streptococcaceae, Staphylococcacea*	*Enterobacteriaceae*

Table 4.2. Human proteins targeted by two Gram positive bacteria families and three Gram negative bacteria families (cont.)

Protein	**Attacking Gram Positive bacterial families**	**Attacking Gram Negative Bacterial families**
ENPL-Endoplasmin	*Bacillaceae, Listeriaceae*	*Enterobacteriaceae, Francisellaceae*
EGFR-Epidermal growth factor receptor	*Bacillaceae, Staphylococcacea*	*Enterobacteriaceae, Francisellaceae*
COL1A1-Collagen alpha-1(I) chain	*Bacillaceae, Streptococcaceae*	*Enterobacteriaceae, Francisellaceae*
RAC1-Ras-related C3 botulinum toxin substrate 1	*Bacillaceae, Peptostreptococcaceae*	*Enterobacteriaceae, Pseudomonadaceae*
A4D2P1- Ras-related C3 botulinum toxin substrate 1	*Bacillaceae, Peptostreptococcaceae*	*Enterobacteriaceae, Pseudomonadaceae*

One of the interesting points of these highly targeted human proteins is the similar behaviour of Gram positive and Gram negative bacterial families as they attack the same human proteins. Gram negative bacterial families which are observed to behave similarly are *Enterobacteriaceae, Chylamydiaceae, Neisseriaceae* whereas Gram positive bacterial families include mostly *Bacillaceae.* It is expected that *Enterobacteriaceae, Bacillaceae and Francisellaceae* will be among the families attacking highly targeted human proteins as they have more than 98% of the high-throughput large-scale PHI data. However, *Francisellaceae* having 23% of PHIs is not found among those Gram negative families.

As a matter of fact, these bacterial families behaving similarly in terms of attacked human proteins, cause very serious diseases. Gram positive bacterial family *Bacillaceae* cause anthrax (Dimitrova *et al*., 2007) which has a 20% mortality rate worldwide (WHO, 2012) and food poisoning (Vos *et al*., 1984). Gram negative bacterium *Enterobacteriaceae* cause lower respiratory tract infections, skin and soft-tissue infections, urinary tract infections (UTIs), endocarditis, intra-abdominal infections, septic arthritis, osteomyelitis, CNS infections, and ophthalmic infections (Fraser *et al*., 2014). These diseases caused by *Enterobacteriaceae* have 40% to 50% mortality rates (CDC, 2013). Gram negative

bacterium *Chylamydiaceae* cause blindness (Ryan *et al*., 2004) while *Neisseriaceae* cause gonorrhea and meningitis (Abbas, 2011) which has a 30% mortality rate (Beek, 2012). *Streptococcaceae* cause pneumonia, which has 18% mortality rate worldwide, scarlet fever and pharyngitis (DeFranco *et al*., 2007).

Table 4.3. Human proteins targeted by two and three Gram positive bacterial families.

UniProt ID	**Human Protein Name**	**Targeting Gram positive family**
P01857	IGHG1- Ig gamma-1 chain C region	*Bacillaceae, Streptococcaceae, Staphylcoccaceae*
P00747	PLMN- Plasminogen	*Bacillaceae, Streptococcaceae, Staphylcoccaceae*
P08123	COL1A2-Collagen alpha-2(I) chain	*Bacillaceae, Streptococcaceae*
P02452	COL1A1-Collagen alpha-1(I) chain	*Bacillaceae, Streptococcaceae*
P63000	RAC1-Ras-related C3 botulinum toxin substrate 1	*Bacillaceae, Peptostreptococcaceae*
P14625	ENPL-Endoplasmin	*Bacillaceae, Listeriaceae*
P00533	EGFR-Epidermal growth factor receptor	*Bacillaceae, Staphylococcacea*
A4D2P1	A4D2P1- Ras-related C3 botulinum toxin substrate 1	*Bacillaceae, Peptostreptococcaceae*

Attacking bacterial families for highly targeted human proteins such as two and three Gram positive-targeted set and three Gram negative-targeted set are also given in Tables 4.3 and 4.4, respectively. The human proteins targeted by three Gram positive bacterial families are IGHG1 (Ig gamma-1 chain C region) and PLMN (Plasminogen). They are targeted by *Bacillaceae*, *Streptococcaceae* and *Staphylococcacea.* IGHG1 plays important roles in immune system, antigen and protein binding. On the other hand, PLMN plays important roles in dissolving the fibrin of blood clots and acting as a proteolytic factor in a variety of other processes including embryonic development, tissue remodeling, tumor invasion, and inflammation.

Table 4.4. Human proteins targeted by three Gram negative bacterial families.

UniProt ID	Human Protein Name	Targeting Gram positive family
O00206	TLR4- Toll-like receptor 4	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
O43187	IRAK2- Interleukin-1 receptor-associated kinase-like 2	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
P51617	IRAK1-Interleukin-1 receptor-associated kinase 1	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
Q13233	MAP3K1- Mitogen-activated protein kinase kinase kinase 1, E3 ubiquitin protein ligase	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
Q5FWG5	TLR1-Toll-like receptor 1	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
Q96FA3	PELI1- E3 ubiquitin-protein ligase pellino homolog 1	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
Q99836	MYD88- Myeloid differentiation primary response protein	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
Q9HAT8	PELI2- E3 ubiquitin-protein ligase pellino homolog 2	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*
Q9NWZ3	IRAK4- Interleukin-1 receptor-associated kinase 4	*Enterobacteriaceae, Chylamydiaceae, Neisseriaceae*

4.1.2. GO Enrichment Analysis

The GO process terms of targeted human proteins can be used to indicate the human biological processes that are attacked by Gram positive and Gram negative bacteria. GO (Ashburner *et al.*, 2000) enrichment analysis of all human protein sets were performed using BiNGO plugin (ver. 2.44) of Cytoscape (ver. 2.8.1) (Maere *et al*., 2005) as explained before in Chapter 3. Significance level was set to 0.05 which means that only terms enriched with a p-value of at most 0.05 were considered. GO enrichment terms of human proteins attacked by all Gram negative and Gram positive bacteria are given in the Table 4.5. Special attention should be paid to the results of sets of human proteins interacting with only Gram positive bacteria proteins (Table 4.6), with only Gram negative bacteria proteins (Table 4.7) and commonly with both Gram positive and Gram negative bacteria

proteins (Table 4.8) to understand specific and common attack strategies of different bacteria types.

Table 4.5. First 20 enriched GO Process terms of human proteins (3698) targeted by all bacterial families.

GO Process Term	**p-value**
regulation of biological process	7.84E-08
negative regulation of biological process	8.48E-08
biological regulation	1.36E-07
response to cytokine stimulus	6.00E-07
cellular component organization	3.04E-06
cytoskeleton organization	4.05E-06
cell death	6.17E-06
regulation of cellular component organization	6.39E-06
multi-organism process	6.62E-06
death	7.03E-06
cellular process	8.24E-06
cellular macromolecule localization	9.56E-06
mRNA metabolic process	1.19E-05
organelle organization	1.62E-05
regulation of cellular process	1.80E-05
negative regulation of cellular process	1.90E-05
RNA metabolic process	1.96E-05
regulation of gene-specific transcription from RNA polymerase II promoter	2.10E-05
regulation of immune system process	2.23E-05
intracellular transport	2.26E-05

From the enriched GO process terms of human proteins targeted by all Gram positive and Gram negative bacterial families (Table 4.5), it is seen that bacteria have tendency to target human proteins which are enriched highly in the regulation of metabolic processes and cellular processes in addition to immune system processes. Both Gram negative and Gram positive bacterial families (especially Gram positive bacterial families) target human proteins involved in regulation of metabolic processes in order to use human metabolism

for their own advantage (Tables 4.5-4.8). Bacterial pathogens use common regulatory mechanisms, to control the expression of their virulence genes in response to environmental conditions encountered during infection of the human host, including changes in temperature, pH, osmotic strength, oxygen availability, and nutrient conditions (Wilson *et al*., 2002). Destroying the regulation of metabolic processes may enable bacteria to easily benefit from human metabolism as well as its nutrition and energy resources (Durmuş Tekir, 2013).

Table 4.6. First 20 enriched GO Process terms of human proteins (876) targeted by only Gram positive bacterial families.

GO Process Term	**p-value**
biological regulation	1.61E-10
positive regulation of biological process	4.74E-10
regulation of biological process	8.29E-10
regulation of cellular process	1.10E-09
positive regulation of cellular process	2.09E-08
negative regulation of biological process	6.49E-08
cellular process	2.46E-07
cellular component organization	2.59E-07
cell death	3.58E-07
response to wounding	4.02E-07
death	4.55E-07
response to stress	6.20E-07
response to organic substance	8.81E-07
negative regulation of cellular process	9.14E-07
membrane invagination	1.40E-06
endocytosis	1.40E-06
response to chemical stimulus	1.74E-06
regulation of cellular component biogenesis	1.78E-06
regulation of cell communication	1.90E-06
biological regulation	1.61E-10

Infection mechanisms can be understood more thoroughly by analyzing the human proteins targeted by multiple bacterial families. Therefore, the results for three and two

Gram positive bacteria-targeted set (Table 4.9) and for three Gram negative bacteria-targeted set (Table 4.10) are discussed below. GO enrichment maps of human proteins attacked by all Gram negative and Gram positive bacteria, by only Gram negative bacteria proteins, by only Gram positive bacteria proteins, by both Gram positive and Gram negative bacteria proteins, by two Gram positive bacteria families and by three Gram negative bacteria families are shown in the Appendix A.1-A.6.

Table 4.7. First 20 enriched GO Process terms of human proteins (1920) targeted by only Gram negative bacterial families.

GO Process Term	**p-value**
biological regulation	4.12E-19
regulation of biological process	9.10E-19
regulation of cellular process	7.57E-17
intracellular transport	9.22E-15
cellular process	2.58E-14
cellular macromolecule metabolic process	3.92E-14
establishment of protein localization	8.74E-14
protein transport	1.36E-13
protein localization	3.72E-13
positive regulation of biological process	4.05E-13
positive regulation of cellular process	6.80E-13
establishment of localization in cell	7.18E-13
macromolecule localization	1.21E-12
macromolecule metabolic process	5.52E-12
cellular localization	1.28E-11
RNA splicing	5.27E-11
RNA metabolic process	1.01E-10
cellular protein localization	1.65E-10
cellular macromolecule localization	2.35E-10
biological regulation	4.12E-19

From the results of GO enrichment terms of three Gram negative bacteria-targeted set (Table 4.10), it can be deduced that gram negative bacteria prefers to attack proteins involved generally in human immunity system. Hence, the most specific bacterial infection

strategy is through evading or suppressing human immune responses as also concluded previously (Lai *et al*., 2001; Park *et al*., 2002; Zhang *et al*., 2005; Dyer *et al*., 2010). By attacking human proteins functioning in innate and adaptive immunity (i.e. TLR4 and TLR7), inflammation (i.e. NF-κB and BCL6), activation of T cells (i.e. CXCR4 and LCK) (Zhang and Ghosh, 2000; Alonso *et al*., 2004; Oda and Kitano, 2006; Dyer *et al*., 2010), bacteria try to manipulate the human immune system

Table 4.8. First 20 enriched GO Process terms of human proteins (902) targeted by both Gram positive and Gram negative bacterial families.

GO Process Term	**p-value**
biological regulation	3.16E-16
regulation of biological process	1.17E-15
regulation of cellular process	3.55E-14
cellular component organization	2.78E-12
mRNA metabolic process	1.17E-11
RNA splicing	1.61E-10
RNA metabolic process	1.70E-10
negative regulation of biological process	1.22E-09
interspecies interaction between organisms	2.89E-09
cellular component assembly	3.11E-09
cytoskeleton organization	3.78E-09
actin cytoskeleton organization	4.96E-09
cellular process	7.27E-09
actin filament-based process	2.40E-08
protein complex biogenesis	3.43E-08
protein complex assembly	3.43E-08
nucleic acid metabolic process	3.80E-08
gene expression	4.35E-08
mRNA processing	5.00E-08
biological regulation	3.16E-16

It is seen from Table 4.9 that positive regulation of lamellipodium (a cytoskeletal protein actin projection on the mobile edge of the cell) assembly, regulation of lamellipodium assembly and lamellipodium assembly attracts attention. Small GTPases of

the Rho family, the human protein RAC, which has been reported in colorectal, pancreatic, breast, and testicular cancers and in various leukemias (Fritz *et al*., 1999; Schnelzer *et al.*, 2000; Wang *et al.*, 2009) is the key regulator of actin assembly and control the formation of lamellipodia (Etienne-Manneville *et al*., 2002).

Table 4.9. First 20 enriched GO Process terms of human proteins (6) targeted by two Gram positive bacterial families (two Gram positive bacteria-targeted set).

GO Process Term	**p-value**
localization within membrane	3.26E-04
skin morphogenesis	2.05E-02
positive regulation of lamellipodium assembly	2.05E-02
anatomical structure morphogenesis	2.46E-02
positive regulation of Rho protein signal transduction	3.08E-02
regulation of lamellipodium assembly	3.08E-02
negative regulation of receptor-mediated endocytosis	4.30E-02
actin filament organization	5.37E-02
regulation of respiratory burst	5.74E-02
regulation of hydrogen peroxide metabolic process	5.74E-02
cellular component organization	5.90E-02
ruffle organization	9.22E-02
extracellular matrix organization	1.24E-01
actin filament polymerization	1.35E-01
regulation of apoptosis	1.41E-01
regulation of programmed cell death	1.48E-01
regulation of cell death	1.54E-01
response to stress	2.26E-01
negative regulation of cell-substrate adhesion	2.45E-01
lamellipodium assembly	2.45E-01

Table 4.10. First 20 enriched GO Process terms of human proteins (9) targeted by three Gram negative bacterial families (three Gram negative bacteria-targeted set).

GO Process Term	**p-value**
I-kappaB kinase/NF-kappaB cascade	3.61E-10
interleukin-1-mediated signaling pathway	7.17E-06
signal transmission via phosphorylation event	1.79E-05
intracellular protein kinase cascade	1.79E-05
positive regulation of NF-kappaB transcription factor activity	6.78E-05
immune response-activating signal transduction	6.78E-05
immune response-regulating signaling pathway	9.69E-05
cytokine-mediated signaling pathway	1.82E-04
initiation of signal transduction	1.93E-04
signal initiation by diffusible mediator	1.93E-04
signal initiation by protein/peptide mediator	1.93E-04
toll-like receptor signaling pathway	3.26E-04
positive regulation of transcription regulator activity	3.46E-04
positive regulation of transcription factor activity	3.46E-04
positive regulation of immune system process	4.28E-04
cellular response to lipopolysaccharide	4.86E-04
positive regulation of DNA binding	5.77E-04
pattern recognition receptor signaling pathway	5.84E-04
activation of immune response	6.56E-04

4.1.3. KEGG Pathway Analysis

One of the most important tasks in the high-throughput studies is to identify the pathways that are involved in the biological processes. KEGG (Ogata *et al*., 1999) Pathway Analysis of all human protein sets were conducted using the web-based gene set analysis toolkit WebGestalt (Zhang *et al*., 2005) software. EntrezGene was selected as the reference set, which is needed to perform statistical analysis to identify enriched gene/protein sets as explained before in Chapter 3. It is NCBI's database for gene-specific information and focuses on the genomes that have been completely sequenced, that have an active research

community to contribute gene-specific information, or that are scheduled for intense sequence analysis (Maglott *et al.*, 2005). 10 pathways which may have importance in explaining the behaviour of targeted human proteins were tabulated for each human protein set. KEGG Pathways enriched with human proteins attacked by all Gram negative and Gram positive bacteria are given in the Table 4.11. Special attention should be paid to the results of human proteins interacting with only Gram positive bacteria proteins (Table 4.12), with only Gram negative bacteria proteins (Table 4.13) and both Gram positive and Gram negative bacteria proteins (Table 4.14) to understand specific and common attack strategies of different bacteria types. Pathways of two Gram positive bacteria-targeted set (Table 4.15) and three Gram negative bacteria-targeted set (Table 4.16) are also given.

Table 4.11. First 10 enriched KEGG pathways in human protein set targeted by all bacterial families.

KEGG Pathways	**adjP-value**	**p-value**
Osteoclast differentiation	1.16E-14	5.61E-17
B cell receptor signaling pathway	2.86E-13	2.76E-15
Spliceosome	4.78E-12	6.93E-14
Leishmaniasis	3.73E-11	7.21E-13
Pathways in cancer	4.45E-10	1.08E-11
Regulation of actin cytoskeleton	4.45E-10	1.29E-11
Endocytosis	4.35E-09	1.47E-10
Renal cell carcinoma	5.82E-09	2.25E-10
Prostate cancer	1.21E-08	5.24E-10
Toll-like receptor signaling pathway	2.48E-08	1.20E-09

KEGG pathways enriched in human proteins targeted by Gram positive and Gram negative bacteria highlight a number of important pathways in immune response. B cell receptor signaling pathway, endocytosis, Toll-like receptor signaling (TLR) pathway and mitogen-activated protein kinase (MAPK) signaling pathway were especially found to be significant, in human proteins targeted by Gram negative bacterial groups. TLRs and MAPK cascade play pivotal roles in host defense in response to microbial infections (Yang *et al.*, 2011). TLR signaling is an important signaling network for both innate and adaptive immunity and is the front-line subsystem against invasive pathogens (Iwasaki and Medzhitov, 2004). In mammalians, MAPK cascade is involved in all aspects of immune

responses, from the initiation phase of innate immunity, to activation of adaptive immunity and to cell death when immune function is complete (Dong *et al*., 2002).

Table 4.12. First 10 enriched KEGG pathways in human protein set targeted by only Gram positive bacterial families.

KEGG Pathways	**adjP-value**	**p-value**
Regulation of actin cytoskeleton	2.07E-08	1.30E-10
Staphylococcus aureus infection	6.52E-06	1.07E-07
B cell receptor signaling pathway	6.52E-06	1.23E-07
Hematopoietic cell lineage	7.51E-06	1.89E-07
Complement and coagulation cascades	8.52E-06	2.68E-07
Glioma	2.25E-05	8.50E-07
Dorso-ventral axis formation	4.16E-05	1.83E-06
Pathways in cancer	6.02E-05	3.03E-06
Melanoma	2.00E-04	1.40E-05
Long-term potentiation	2.00E-04	1.20E-05

Table 4.13. First 10 enriched KEGG pathways in human protein set targeted by only Gram negative bacterial families.

KEGG Pathways	**adjP-value**	**p-value**
Bacterial invasion of epithelial cells	2.11E-05	1.38E-07
Proteasome	2.11E-05	2.22E-07
Spliceosome	2.50E-05	4.63E-07
Endocytosis	2.50E-05	5.27E-07
B cell receptor signaling pathway	8.25E-05	2.17E-06
Osteoclast differentiation	2.00E-04	5.70E-06
Lysosome	5.00E-04	1.82E-05
Shigellosis	6.00E-04	2.77E-05
Pathogenic Escherichia coli infection	6.00E-04	2.97E-05
Antigen processing and presentation	8.00E-04	4.00E-05

One of the other important human protein targeted by bacteria is epidermal growth factor receptor (EGFR), which is targeted by *Bacillaceae* (Gram positive), *Staphylococcaceae* (Gram positive, extracellular), *Enterobacteriaceae* (Gram negative,

intracellular), *Francisellaceae* (Gram negative, intracellular) (Table 4.2). EGFR plays a fundamental role in the morphogenesis of organisms and is also involved in the development and growth of many types of human tumour cells (Yarden, 2001). The interactions between bacteria-human proteins should be investigated thoroughly in order to understand the immune system in human as a defense mechanism completely

Table 4.14. First 10 enriched KEGG pathways in human protein set targeted by both Gram positive and Gram negative bacterial families.

KEGG Pathways	**adjP-value**	**p-value**
Focal adhesion	6.75E-07	7.94E-09
Endocytosis	6.75E-07	8.88E-09
Adherens junction	4.99E-05	9.84E-07
Osteoclast differentiation	4.00E-04	1.39E-05
Regulation of actin cytoskeleton	4.00E-04	1.45E-05
Pancreatic cancer	5.00E-04	1.94E-05
Pathways in cancer	1.20E-03	5.30E-05
Phagosome	1.40E-03	1.00E-05
Bacterial invasion of epithelial cells	1.40E-03	9.78E-05
Viral myocarditis	1.40E-03	9.78E-05

It is observed that the human protein MAP3K1 takes place in Neurotrophin signaling pathway, RIG-I-like receptor signaling pathway, Ubiquitin mediated proteolysis and MAPK signaling pathway. Neurotrophin signaling which plays an important role for neural development and additional higher-order activities, is regulated by connecting a variety of intracellular signaling cascades, which include MAPK pathway. RIG-I-like receptor signaling pathway receptors are responsible for detecting viral pathogens and generating innate immune responses. The ubiquitin mediated proteolysis, similar to MAPK pathway, plays an important role in a broad array of basic cellular processes which include regulation of cell cycle and immune and inflammatory responses (Ciechanover *et al*., 2000).

Table 4.15. First 10 enriched KEGG pathways in human protein set targeted by two Gram positive bacterial families (two Gram positive bacteria-targeted set).

KEGG Pathways	adjP-value	p-value	Human Proteins
Focal adhesion	1.50E-06	1.36E-07	COL1A2, COL1A1, RAC1, EGFR
Pancreatic cancer	0.04E-02	0.02E-02	RAC1, EGFR
Pathways in cancer	0.04E-02	7.94E-05	RAC1, HSP90B1, EGFR
Prostate cancer	0.04E-02	0.03E-02	HSP90B1, EGFR
ECM-receptor interaction	0.04E-02	0.03E-02	COL1A2, COL1A1
Epithelial cell signaling in Helicobacter pylori infection	0.04E-02	0.02E-02	RAC1, EGFR
Protein digestion and absorption	0.04E-02	0.02E-02	COL1A2, COL1A1
Adherens junction	0.04E-02	0.02E-02	RAC1, EGFR
Amoebiasis	0.05E-02	0.04E-02	COL1A2, COL1A1
Regulation of actin cytoskeleton	0.18E-02	0.16E-02	RAC1, EGFR

TLR signaling pathway, Leismaniasis, Chagas disease and Toxoplasmosis are controlled by the human proteins TRAF6, TLR4, IRAK1, MyD88, IRAK4, which are found in significant KEGG pathways (Table 4.16). Infection of Leishmania which is called Leismaniasis is achieved by alteration of signaling events in the host cell, resulting abnormalities in immune responses. During Chagas disease, a tropical parasitic disease, the pathogen manipulates the host innate immunity. Toxoplasmosis, a chronic infection, causes resistant damage in the host's immune system changing the host's immune processes. TRAF6, the connection between the adaptive and innate immune responses (Choi, 2005) and MyD88, the core protein of the bow-tie structured TLR network (Beutler, 2004), are found to be the out-degree hubs which means that any mutations in these highly connected signaling proteins may cause immunological disorder (Özbabacan *et al*., 2011). IRAK1 and IRAK4 are active kinases dissociating from the receptor-adapter complex upon phoshorylation which plays a significant role in cellular processes and activating TRAF6 (Picard, 2003). They are involved in signaling innate immune responses from TLRs and play a critical role in initiating innate immune response against foreign

pathogens. IRAK4 shares the domain structure of the other IRAKs and can activate similar signal transduction pathways such as MAPK signaling pathway (Li, 2002).

Table 4.16. First 10 enriched KEGG pathways in human protein set targeted by three Gram negative bacterial families (three Gram negative bacteria-targeted set).

KEGG Pathways	**adjP-value**	**p-value**	**Human Proteins**
Leishmaniasis	6.10E-10	6.10E-11	TRAF6, TLR4, IRAK1, MYD88, IRAK4
Chagas disease (American trypanosomiasis)	1.33E-09	3.99E-10	TRAF6, TLR4, IRAK1, MYD88, IRAK4
Toll-like receptor signaling pathway	1.33E-09	3.61E-10	TRAF6, TLR4, IRAK1, MYD88, IRAK4
Toxoplasmosis	2.66E-09	1.33E-09	TRAF6, TLR4, IRAK1, MYD88, IRAK4
Neurotrophin signaling pathway	2.66E-09	1.10E-09	TRAF6, IRAK2, IRAK1, MAP3K1, IRAK4
Apoptosis	6.52E-08	3.91E-08	IRAK2, IRAK1, MYD88, IRAK4
Malaria	0.03E-02	0.02E-02	TLR4, MYD88
RIG-I-like receptor signaling pathway	0.05E-02	0.04E-02	TRAF6, MAP3K1
Ubiquitin mediated proteolysis	0.17E-02	0.15E-02	TRAF6, MAP3K1
MAPK signaling pathway	0.59E-02	0.59E-02	TRAF6, MAP3K1

As seen from the results of enriched KEGG Pathways of two Gram positive bacteria-targeted set and all bacteria-targeted set, focal adhesion, pacreatic cancer, pathways in cancer, prostate cancer and epithelial cell signaling in *Helicobacter pylori* infection seem to be significant. Literature studies show that bacterial invasion is linked to cancer (Jones, 2000; Egi *et al*., 2007; Beglinge *et al*., 2007; Kocazeybek, 2003; Ning *et al*., 2004; Namiki *et al*., 2009; Borriello *et al*., 1983; Murray *et al*., 1980). RAC1 human protein, taking place

in focal adhesion and pathways in cancer (Table 4.15) (Chapter 4.1.2), has an important regulatory role specifically in cell motility and cell growth (Parri *et al*., 2010). RAC1 human protein is targeted by *Bacillaceae* (Gram positive), *Peptostreptococcaceae* (Gram positive), *Enterobacteriaceae* (Gram negative, intracellular), *Pseudomonadaceae* (Gram positive, extracellular) (Table 4.2). Deregulation of cell motility is one of the distinct issues in cancer cell invasion and metastasis (Hanahan *et al.,* 2011). The focal adhesion is a prominent determinant in cancer initiation, progression and metastasis (Luo and Guan, 2010).

4.1.4. Pathway Commons Analysis

Along with KEGG Pathway Analysis, Pathway Commons Analysis of all human protein sets were conducted using WebGestalt (Zhang *et al*., 2005) software (Tables 4.17-4.22). EntrezGene was again selected as the reference set which is needed to perform statistical analysis to identify enriched protein sets as explained in Chapter 3.

Table 4.17. First 10 enriched Pathway commons in human protein set targeted by all bacterial families.

Pathway Commons	**adjP-value**	**p-value**
TRAIL signaling pathway	2.98E-61	2.88E-64
Proteoglycan syndecan-mediated signaling events	4.48E-61	8.65E-64
PAR1-mediated thrombin signaling events	3.13E-60	9.33E-63
Thrombin/protease-activated receptor (PAR) pathway	3.13E-60	1.21E-62
Syndecan-1-mediated signaling events	7.50E-60	3.62E-62
IFN-gamma pathway	1.99E-59	1.15E-61
LKB1 signaling events	2.99E-59	2.81E-61
Glypican pathway	2.99E-59	2.89E-61
Nectin adhesion pathway	2.99E-59	2.65E-61

Similar to the results of the KEGG Pathways analysis, the pathways related to immunity such as Toll Like Receptor 4 (TLR4) Cascade, TRAF6 mediated induction of NFkB and MAP kinases upon TLR7/8 or 9 activation, Toll Like Receptor 5 (TLR5) Cascade, MyD88:Mal cascade initiated on plasma membrane, appeared as the most significant ones attacked by three Gram negative bacteria-targeted set. The analysis of

three Gram negative bacteria-targeted set similarly revealed the human proteins of TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1, TLR4 in pathway commons. Moreover, human proteins PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1 are included in the pathways, Alpha9 beta1 integrin signaling events, IFN-gamma pathway, Signaling events mediated by focal adhesion kinase, IL3-mediated signaling events and EGF receptor (ErbB1) signaling pathway attacked by two Gram positive bacteria.

Table 4.18. First 10 enriched Pathways commons in human protein set targeted by only Gram positive bacterial families.

Pathway Commons	**adjP-value**	**p-value**
TRAIL signaling pathway	1.49E-20	2.21E-23
Proteoglycan syndecan-mediated signaling events	4.19E-20	2.09E-22
Beta1 integrin cell surface interactions	4.19E-20	3.10E-22
Integrin family cell surface interactions	4.19E-20	2.01E-22
Syndecan-1-mediated signaling events	4.19E-20	2.92E-22
Plasma membrane estrogen receptor signaling	7.96E-20	9.42E-22
PAR1-mediated thrombin signaling events	7.96E-20	8.26E-22
Thrombin/protease-activated receptor (PAR) pathway	7.96E-20	8.82E-22
IL5-mediated signaling events	1.17E-19	1.56E-21

Table 4.19. First 10 enriched Pathways commons in human protein set targeted by only Gram negative bacterial families.

Pathway Commons	**adjP-value**	**p-value**
Proteoglycan syndecan-mediated signaling events	7.88E-32	9.31E-35
Thrombin/protease-activated receptor (PAR) pathway	3.10E-31	1.10E-33
PAR1-mediated thrombin signaling events	3.10E-31	9.69E-34
Plasma membrane estrogen receptor signaling	4.94E-31	9.27E-33
PDGFR-beta signaling pathway	4.94E-31	1.38E-32
Class I PI3K signaling events	4.94E-31	1.38E-32
Syndecan-1-mediated signaling events	4.94E-31	3.01E-33
ErbB1 downstream signaling	4.94E-31	1.38E-32
Proteoglycan syndecan-mediated signaling events	7.88E-32	9.31E-35

Table 4.20. First 10 enriched Pathways commons in human protein set targeted by both Gram positive and Gram negative bacterial families.

Pathway Commons	**adjP-value**	**p-value**
Thrombin/protease-activated receptor (PAR) pathway	1.17E-27	6.88E-30
Beta1 integrin cell surface interactions	1.17E-27	3.16E-30
Syndecan-1-mediated signaling events	1.17E-27	6.88E-30
PAR1-mediated thrombin signaling events	1.17E-27	6.34E-30
Proteoglycan syndecan-mediated signaling events	2.63E-27	2.24E-29
Integrin family cell surface interactions	2.63E-27	2.70E-29
Plasma membrane estrogen receptor signaling	2.63E-27	2.56E-29
TRAIL signaling pathway	5.60E-27	6.58E-29
IGF1 pathway	8.79E-27	1.31E-28

Table 4.21. First 10 enriched Pathway commons in human protein set targeted by two Gram positive bacterial families (two Gram positive bacteria-targeted set).

Pathway Commons	**adjP-value**	**p-value**	**Human Proteins**
Alpha9 beta1 integrin signaling events	1.59E-07	7.02E-08	PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1
IFN-gamma pathway	1.59E-07	6.73E-08	PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1
Signaling events mediated by focal adhesion kinase	1.59E-07	6.49E-08	PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1
Internalization of ErbB1	1.59E-07	6.49E-08	PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1
IL3-mediated signaling events	1.59E-07	6.70E-08	PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1
EGFR-dependent Endothelin signaling events	1.59E-07	6.52E-08	PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1
Insulin Pathway	1.59E-07	6.49E-08	PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1
Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met)	1.59E-07	6.64E-08	PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1

Table 4.21. First 10 enriched Pathway commons in human protein set targeted by two Gram positive bacterial families (two Gram positive bacteria-targeted set) (cont.).

Pathway Commons	**adjP-value**	**p-value**	**Human Proteins**
Nectin adhesion pathway	1.59E-07	6.70E-08	PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1
EGF receptor (ErbB1) signaling pathway	1.59E-07	6.49E-08	PLG, HSP90B1, COL1A2, COL1A1, EGFR, RAC1

Table 4.22. First 10 enriched Pathway commons in human protein set targeted by three Gram negative bacterial families (three Gram negative bacteria-targeted set).

Pathway Commons	**adjP-value**	**p-value**	**Human Proteins**
Activated TLR4 signalling	4.28E-21	5.33E-23	TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1, TLR4
Toll Like Receptor 4 (TLR4) Cascade	4.28E-21	8.91E-23	TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1, TLR4
Toll Receptor Cascades	1.41E-20	4.42E-22	TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1, TLR4
TRAF6 mediated induction of NFkB and MAP kinases upon TLR7/8 or 9 activation	3.89E-19	2.81E-20	TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1
Toll Like Receptor 5 (TLR5) Cascade	3.89E-19	3.24E-20	TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1
MyD88 cascade initiated on plasma membrane	3.89E-19	2.81E-20	TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1
MyD88 dependent cascade initiated on endosome	3.89E-19	3.24E-20	TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1
Toll Like Receptor 10 (TLR10) Cascade	3.89E-19	3.24E-20	TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1

Table 4.22. First 10 enriched Pathway commons in human protein set targeted by three Gram negative bacterial families (three Gram negative bacteria-targeted set) (cont.).

Pathway Commons	adjP-value	p-value	Human Proteins
Toll Like Receptor 7/8 (TLR7/8) Cascade	4.11E-19	4.28E-20	TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1
MyD88:Mal cascade initiated on plasma membrane	4.11E-19	4.28E-20	TRAF6, IRAK1, PELI2, IRAK4, IRAK2, PELI1, MYD88, MAP3K1

4.1.5. Transcription Factor Target Analysis

The transcription factor target analysis of human protein sets (All bacteria set, only Gram positive bacteria-targeted set, only Gram negative bacteria-targeted set, both Gram positive and Gram negative bacteria–targeted set, two Gram positive bacteria-targeted set, three Gram negative bacteria-targeted set) were conducted by using the web-based gene set analysis toolkit WebGestalt (Zhang *et al*., 2005). EntrezGene was again selected as the reference set to perform statistical analysis. The human proteins attacked by bacteria are explored whether the genes encoding these sets of proteins are significantly enriched with potential target sites, TF targets (TFTs). The top targets of transcription factors of human proteins attacked by bacteria are listed in Tables 4.23 to 4.28.

Table 4.23. First 8 enriched TFTs in human proteins targeted by all bacterial families.

TFT	adjP-value	p-value
hsa_GGGCGGR_V$SP1_Q6	2.22E-56	3.61E-59
hsa_RYTTCCTG_V$ETS2_B	7.32E-26	2.38E-28
hsa_SCGGAAGY_V$ELK1_02	1.90E-24	9.27E-27
hsa_RCGCANGCGY_V$NRF1_Q6	5.77E-23	4.69E-25
hsa_GGGAGGRR_V$MAZ_Q6	5.77E-23	3.89E-25
hsa_MGGAAGTG_V$GABP_B	9.00E-20	8.78E-22
hsa_GCCATNTTG_V$YY1_Q6	1.97E-19	2.24E-21
hsa_CTTTGT_V$LEF1_Q2	1.55E-15	2.01E-17

Table 4.24. First 8 enriched TFTs in human proteins targeted by only Gram positive bacterial families.

TFT	adjP-value	p-value
hsa_GGGCGGR_V$SP1_Q6	1.32E-12	2.21E-15
hsa_GGGAGGRR_V$MAZ_Q6	4.79E-08	1.60E-10
hsa_CTTTGT_V$LEF1_Q2	1.52E-06	7.63E-09
hsa_V$EGR1_01	4.30E-06	2.87E-08
hsa_MGGAAGTG_V$GABP_B	9.22E-06	7.70E-08
hsa_V$NGFIC_01	1.04E-05	1.04E-07
hsa_TGCGCANK_UNKNOWN	1.75E-05	2.05E-07
hsa_V$MYCMAX_02	1.98E-05	2.64E-07

Table 4.25. First 8 enriched TFTs in human proteins targeted by only Gram negative bacterial families.

TFT	adjP-value	p-value
hsa_GGGCGGR_V$SP1_Q6	4.15E-26	6.77E-29
hsa_GCCATNTTG_V$YY1_Q6	1.49E-11	4.87E-14
hsa_RYTTCCTG_V$ETS2_B	4.90E-11	2.40E-13
hsa_V$YY1_02	1.78E-10	1.16E-12
hsa_SCGGAAGY_V$ELK1_02	2.44E-10	1.99E-12
hsa_RCGCANGCGY_V$NRF1_Q6	4.73E-10	4.63E-12
hsa_V$YY1_Q6	5.45E-08	6.22E-10
hsa_MGGAAGTG_V$GABP_B	6.06E-08	7.91E-10

It is observed that the most significant transcription factor may be SP1 which may regulate the only Gram positive, only Gram negative and both Gram negative and Gram positive bacteria-targeted human proteins by binding to the target sequence GGGCGGR. The TFT namely hsa_GGGCGGR_V$SP1_Q6 matches annotation with the TF SP1. SP1 is involved in many cellular processes, including cell differentiation, cell growth, apoptosis, immune responses, response to DNA damage, and chromatin remodeling. In case of the three Gram negative and two Gram positive bacteria targeted human proteins, the TFTs, which have the highest p-value, are hsa_CTGYNNCTYTAA_UNKNOWN and hsa_V$CP2_01 respectively. hsa_CTGYNNCTYTAA_UNKNOWN has the binding target sequence CTGYNNCTYTAA, however it does not match any known TF. PELI2, IRAK4

and TLR4 are mostly found to be enriched with that particular TFT. hsa_V$CP2_01 matches annotation with TF CP2. COL1A2, COL1A1 and RAC1 are found to be enriched with this TFT. CP2 regulates erythroid gene expression, plays a role in the transcriptional switch of globin gene promoters, and it activates many other cellular and viral gene promoters.

Table 4.26. First 8 enriched TFTs in human proteins targeted by both Gram positive and Gram negative bacterial families.

TFT	adjP-value	p-value
hsa_GGGCGGR_V$SP1_Q6	1.75E-20	2.91E-23
hsa_GGGAGGRR_V$MAZ_Q6	7.77E-13	2.58E-15
hsa_CTTTGT_V$LEF1_Q2	3.63E-11	1.81E-13
hsa_RCGCANGCGY_V$NRF1_Q6	5.39E-08	3.58E-10
hsa_GCCATNTTG_V$YY1_Q6	1.77E-07	1.47E-09
hsa_GGGTGGRR_V$PAX4_03	1.84E-07	1.83E-09
hsa_TAANNYSGCG_UNKNOWN	9.72E-07	1.13E-08
hsa_CACGTG_V$MYC_Q2	1.25E-06	1.66E-08

Table 4.27. First 8 enriched TFTs in human proteins targeted by two Gram positive bacterial families (two Gram positive bacteria-targeted set).

TFT	adjP-value	p-value	Human Proteins
hsa_V$CP2_01	0.92E-02	0.23E-02	COL1A2, COL1A1
hsa_V$SRF_Q6	0.92E-02	0.20 E-02	COL1A2, COL1A1
hsa_TATAAA_V$TATA_01	9.86E-02	4.89E-02	COL1A2, COL1A1
hsa_GGGTGGRR_V$PAX4_03	9.67E-02	4.93E-02	COL1A1, RAC1
hsa_TTGTTT_V$FOXO4_01	1.77E-02	1.13E-02	COL1A2, COL1A1
hsa_CAGGTG_V$E12_Q6	1.77E-02	1.54E-02	COL1A2, COL1A1
hsa_GGGAGGRR_V$MAZ_Q6	1.77E-02	1.34E-02	COL1A2, COL1A1
hsa_GGGCGGR_V$SP1_Q6	2.03E-02	2.03E-02	COL1A2, COL1A1

Table 4.28. First 8 enriched TFTs in human proteins targeted by three Gram negative bacterial families (three Gram negative bacteria-targeted set).

TFT	adjP-value	p-value	Human Proteins
hsa_CTGYNNCTYTAA_UNKNOWN	0.42E-02	0.06E-02	PELI2, IRAK4
hsa_V$PU1_Q6	0.73E-02	0.42E-02	TLR4, IRAK4
hsa_V$NKX62_Q2	0.73E-02	0.45E-02	IRAK1, PELI2
hsa_TTAYRTAA_V$E4BP4_01	0.73E-02	0.52E-02	IRAK1, PELI2
hsa_V$CREBP1_01	0.73E-02	0.25E-02	IRAK1, PELI2
hsa_SMTTTTGT_UNKNOWN	1.46E-02	1.25E-02	IRAK1, PELI2
hsa_RGAGGAARY_V$PU1_Q6	1.87E-02	1.87E-02	TLR4, IRAK4

4.2. Bacterial Families

For a family-based analysis of PHI data, five families (*Enterobacteriaceae, Francisellaceae, Chylamydiaceae, Neisseriaceae, Pseudomonadaceae*) of Gram Negative bacteria and six families (*Bacillaceae, Clostridiaceae, Listeriaceae, Peptostreptococcaceae, Staphylococcacea, Streptococcaceae*) of Gram positive bacteria were selected as shown in the Figures 4.2 and 4.3. To observe the family-based interaction characteristics of bacterial proteins with human proteins, the targeted human proteins are presented with respect to their interactions with the families. The common human proteins that all Gram negative bacteria target may reflect the infection strategies of these bacterial families.

4.2.1. Gram Negative Bacterial Families

The distribution of the Gram negative bacteria-targeted human proteins are shown in the Figure 4.2. The black, red, blue, green and yellow colored lines denote *Pseudomonadaceae* (Extracellular), *Francisellaceae* (Intracellular), *Chylamydiceae* (Intracellular), *Enterobacteriaeae* (Intracellular), and *Neisseriaceae* (Extracellular) Gram negative bacterial families, respectively.

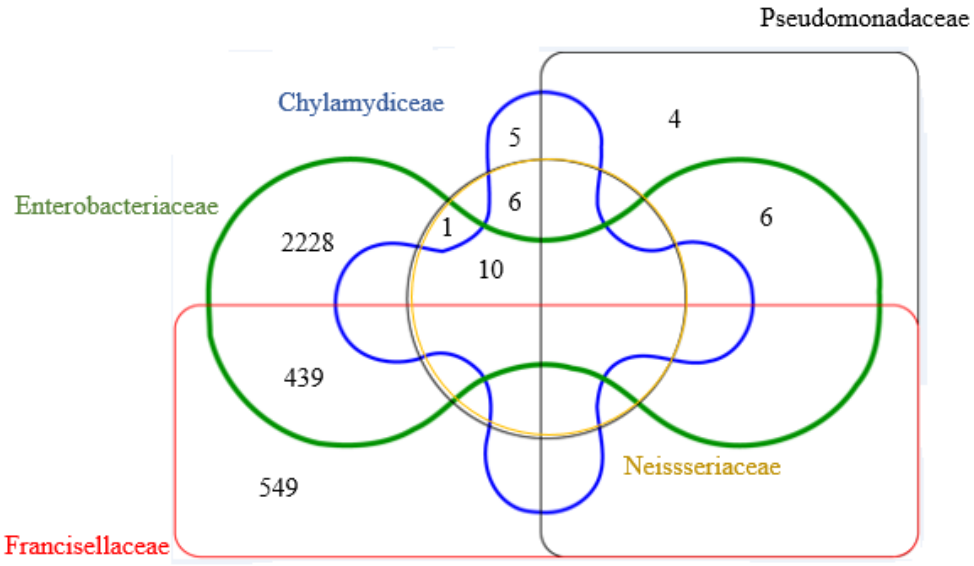


Figure 4.2. The Number of Human Proteins Targeted by Gram Negative Bacteria.

Table 4.29. Commonly targeted human proteins by Gram negative bacteria families.

Gram Negative Bacteria sets	**Commonly Targeted Human proteins**
Enterobacteriaceae-Chylamydiceae	TLR4, IRAK2, IRAK1, MAP3K1, TLR1, PELI1, MYD88, PELI2, IRAK4, TRAF6
Enterobacteriaceae-Neisseriaceae	ITGB1, TLR4, IRAK2, IRAK1, MAP3K1, TLR1, PELI1, MYD88, PELI2, IRAK4, TRAF6
Enterobacteriaceae-Pseudomonadaceae	YWHAE, YWHAZ, YWHAH, RAC1, A4D2P1, D0PNI1
Chylamydiceae- Neisseriaceae	TLR2, CD14, TIRAP, BTK, TLR6, LY96, TLR4, IRAK2, IRAK1, MAP3K1, TLR1, PELI1, MYD88, PELI2, IRAK4, TRAF6
Enterobacteriaceae-Chylamydiceae-Neisseriaceae	TLR4, IRAK2, IRAK1, MAP3K1, TLR1, PELI1, MYD88, PELI2, IRAK4, TRAF6

It is observed from the Venn diagram that there is no common human protein that all Gram negative bacteria target together. The common human proteins targeted by two and three bacterial families are shown in the Table 4.29. *Chylamydiceae, Enterobacteriaeae* and *Neisseriaceae* together target the human proteins namely TLR4, IRAK2, IRAK1, MAP3K1, TLR1, PELI1, MYD88, PELI2, IRAK4 and TRAF6. These human proteins are

previously discussed in Chapter 4.1.3 and Chapter 4.1.4 in the light of human immune system.

The human proteins, TRAF6, IRAK 2, IRAK 1, MAP3K1, IRAK4 which are targeted by three Gram negative bacterial families are enriched with neurotrophin signaling pathway as discussed in Chapter 4.1.3. Neurotrophin signaling pathway plays an important role for neural development and additional higher-order activities. It is regulated by connecting a variety of intracellular signaling cascades, which include MAPK pathway.

Table 4.30. Specific human proteins targeted by *Chylamydiceae* and *Pseudomonadaceae.*

Gram Negative Bacteria Families	**Specific Human Proteins**
Chylamydiceae	CASB, TR10B, TNR6, TR10A, TNR1A.
Pseudomonadaceae	YWHAQ, YWHAG, SFN, YWHAB.

From the Table 4.30, it is seen that specific human proteins targeted by only *Chylamydiceae*, which causes blindness and trachoma, are CASB, TR10B, TNR6, TR10A, TNR1A. The specific human proteins targeted by only *Pseudomonadaceae*, which causes respiratory tract infections, soft tissue infections, urinary tract infections and pneumonia, are YWHAQ, YWHAG, SFN, YWHAB.

4.2.2. Gram Positive Bacterial Families

The distribution of the Gram positive bacteria-targeted human proteins is shown in the Figure 4.3. The black, red, blue, green, yellow and pink colored lines denote Gram positive bacteria families *Bacillaceae*, *Streptococcaceae* (Extracellular), *Clostridiaceae*, *Staphylococcaceae* (Extracellular), *Peptostreptococcaceae* and *Listeriaceae* (Intracellular) respectively.

It is observed from the Venn diagram that there is no common human protein that all Gram positive bacteria target together. The common human proteins targeted by two and three bacterial families are shown in the Table 4.31 *Bacillaceae*, *Staphylococcaceae* and *Streptococcaceae* together target the human proteins, IGHG1 and PLMN. IGHG1 functions in complement activation and innate immune response and plays an important

role in Reactome (Croft *et al.*, 2011) pathway immune system, organism-specific biosystem, whereas PLMN functions in cellular protein metabolic process and negative regulation of cell proliferation. Human proteins RAC1 (commonly targeted by *Bacillaceae* and *Peptostreptococcaceae)* and EGFR (commonly targeted by *Bacillaceae* and *Staphylococcaceae)* are found in KEGG pathways of pancreatic cancer, pathways in cancer and prostate cancer (Table 4.15 and 4.33).

Table 4.31. Specific human proteins targeted by *Streptococcaceae, Staphylococcaceae, Clostridiaceae, Peptostreptococcaceae* and *Listeriaceae.*

Gram Positive Bacterial Families	**Specific Human Proteins**
Streptococcaceae	CD59, C4BPA
Staphylococcaceae	THRB, Q8WLQ7, TVB1, CO3, CO5, CRP, FINC, D7RIH8, DRA, TNR1A, CFAB.
Clostridiaceae	SNP25, VAMP2, SYT1, VAMP1, STX1A, SYT2
Peptostreptococcaceae	RASH, RASK, RASN.
Listeriaceae	MET (I), CADH1, A8K1U7, MET (II)

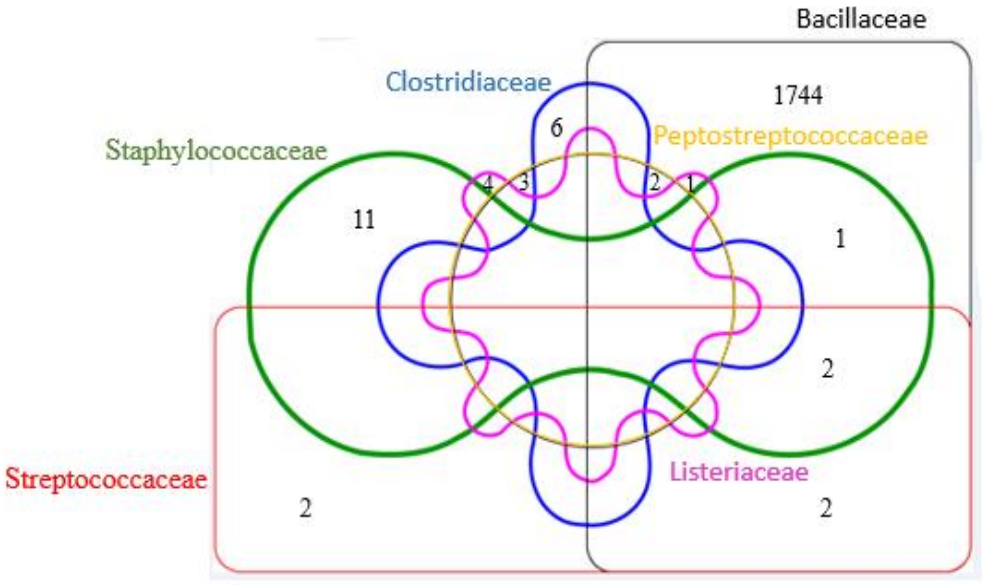


Figure 4.3. The Number of Human Proteins Targeted by Gram Positive Bacteria.

The specific human proteins (Table 4.32) targeted by *Streptococcaceae,* causing pneumonia, scarlet fever and pharyngitis, are CD59 and C4BPA. THRB, Q8WLQ7, TVB1, CO3, CO5, CRP, FINC, D7RIH8, DRA, TNR1A and CFAB are the human proteins tsrrgeted by *Staphylococcaceae* which causes severe skin, wound infections. Only *Peptostreptococcaceae,* resulting in respiratory tract, intra-abdominal and subcutaneous infections, attack RASH, RASK, RASN. Only *Listeriaceae, causing* food poisoning, and listeriosis,attack human proteins MET (I), CADH1, A8K1U7, MET (II).

The human proteins targeted by only one bacterial family are important mainly to reveal the infection strategies specific to that bacterium. The human proteins SNP25, VAMP2, SYT1, VAMP1, STX1A and SYT2 are targeted only by *Clostridiaceae*. Among these, VAMP1 and VAMP2 are involved in the targeting and/or fusion of transport vesicles to their target membrane. They also play important roles in neurotransmitter secretion, regulation of exocytosis, synaptic transmission and vesicle fusion. These two human proteins cause tetanus in human body (Galli *et al*., 1998).

Table 4.32. Commonly targeted human proteins by Gram positive bacteria families.

Gram Positive Bacteria sets	**Commonly Targeted Human proteins**
Staphylococcaceae-Streptococcaceae	IGHG1, PLMN
Bacillaceae-Streptococcaceae	IGHG1, PLMN, CO1A2, CO1A1
Bacillaceae- Staphylococcaceae	IGHG1, PLMN, EGFR
Bacillaceae-Listeriaceae	ENPL
Bacillaceae-Peptostreptococcaceae	RAC1, A4D2P1
Bacillaceae-Streptococcaceae- Staphylococcaceae	IGHG1, PLMN

4.3. Bacterial Strains

To understand the bacterial families' infection strategies, strains of each family were investigated intensely (Table 4.34). As *Aeromonadaceae* (Gram negative), *Camplyobacteraceae* (Gram negative, Extracellular), *Legionellaceae* (Gram negative, intracellular), *Moraxellaceae* (Gram negative), *Mycoplasmataceae* (Gram negative, extracellular), *Neisseriaceae* (Gram negative, extracellular), *Pseudomonadaceae* (Gram Gram negative, extracellular), and *Vibrionaceae* (Gram negative, extracellular) only have

one strain, it is not possible to analyze the intraspecies interactions with the human proteins. On the other hand, *Chylamydiceae* (Gram negative, intracellular) , Enterobacteriaceae (Gram negative, extracellular), *Francisellaceae* (Gram negative, extracellular), *Helicabacteraceae* (Gram negative, extracellular), *Clostridiaceae (*Gram positive, extracellular), *Bacillaceaae* (Gram positive), *Listeriaceae* (Gram positive, intracellular), *Peptostreptococcaceae* (Gram positive), *Staphylococcacea* (Gram positive, extracellular), and *Streptococcaceae* (Gram positive, extracellular) have more than one strain, and this makes it possible to study the interactions between the human proteins and the strains of each family (Figures 4.4- 4.13).

As seen from the Figures 4.4-4.13, only the strains of *Peptostreptococcaceae, Listeriaceae* and *Clostridiaceae* have common human proteins. Due to the lack of data, it is not possible to get statistically meaningful analyses. However, these first attempts with limited results may reflect the initial insights about the infection strategies.

Table 4.33. PHI data belonging to all bacterial families and strains.

Families	**Bacteria strains**	**# of PHIs**	**# of pathogen proteins**	**# of human proteins**
Aeromonadaceae	*Aeromonas hydrophila*	1	1	1
Camplyobacteraceae	*Campylobacter jejuni*	2	2	2
Legionellaceae	*Legionella pneumophila SUBSPECIES PNEUMOPHILA STRAIN PHILADEPHIA*	2	2	2
Moraxellaceae	*Moraxella catarrhalis*	4	4	4
Mycoplasmataceae	*Mycoplasma arthritidis*	1	1	1
Neisseriaceae	*Neisseria meningitidis SEROGROUP B STRAIN MC58*	3	3	3
Pseudomonadaceae	*Pseudomonas aeruginosa*	1	1	1

Table 4.34. Contents of PHI data belonging to all bacteria families and strains (cont.).

Families	Bacteria strains	# of PHIs	# of pathogen proteins	# of human proteins
Vibrionaceae	*Vibrio cholerae*	1	1	1
Chylamydiceae	*Chlamydia trachomatis STRAIN A / HAR-13*	1	1	1
	Chlamydia trachomatis STRAIN D / UW-3 / CX	20	2	20
	TOTAL	20	3	21
Enterobacteriaceae	*Yersinia Pestis*	4203	1227	2196
	Yersinia Pestis KIM 10+	217	47	193
	Shigella Flexneri	20	9	13
	TOTAL	4395	1267	2208
Francisellaceae	*Francisella tularensis SUBSPECIES TULARENSIS SCHU S4*	1317	342	974
	Francisella tularensis SUBSPECIES TULARENSIS	24	4	23
	TOTAL	1341	346	988
Helicabacteraceae	*Helicobacter pylori*	2	2	2
	Helicobacter pylori 26695	1	1	1
	Helicobacter pylori G27	2	1	2
	TOTAL	5	4	5

Table 4.34. Contents of PHI data belonging to all bacteria families and strains (cont.).

Families	Bacteria strains	# of PHIs	# of pathogen proteins	# of human proteins
Clostridiaceae	*Clostridium botulinum*	6	9	6
	Clostridium botulinum F STRAIN LANGELAND	1	1	1
	TOTAL	45	10	6
Bacillaceaae	*Bacillus anthracis*	3181	942	1750
	Bacillus subtilis	1	1	1
	TOTAL	3182	943	1751
Listeriaceae	*Listeria monocytogenes*	5	5	5
	Listeria monocytogenes STRAIN EGD-e	2	2	2
	TOTAL	7	6	5
Peptostreptococcaceae	*Clostridium difficile*	2	1	2
	Clostridium sordellii	5	1	5
	TOTAL	7	2	5

Table 4.34. Contents of PHI data belonging to all bacteria families and strains (cont.).

Families	Bacteria strains	# of PHIs	# of pathogen proteins	# of human proteins
Staphylococcacea	*Staphylococcus aureus*	12	10	11
	Staphylococcus aureus SUBSPECIES AUREUS / STRAIN NCTC 8325	3	2	3
	Staphylococcus aureus STRAIN Mu50 / ATCC 700699	1	1	1
	Staphylococcus aureus SUBSPECIES AUREUS / STRAIN N315	1	1	1
	TOTAL	16	13	14
Streptococcaceae	*Streptococcus dysgalactiae SUBSPECIES EQUISSIMILIS*	1	1	1
	Streptococcus intermedius	1	1	1
	Streptococcus pyogenes	7	7	3
	Streptococcus GROUP G	3	2	3
	TOTAL	12	11	7

Commonly targeted human proteins (Table 4.32) may give us some important clues about the bacterial infection mechanisms. The common human proteins targeted by all two strains of *Peptostreptococcaceae* are RASN and RAC1 (Figure 4.7). RASN, GTPase NRas, binds GDP/GTP and possesses intrinsic GTPase activity. RAC1, Ras-related C3

botulinum toxin substrate 1 (Rho family, small GTP binding protein Rac1), is involved in small GTPase mediated signal transduction.

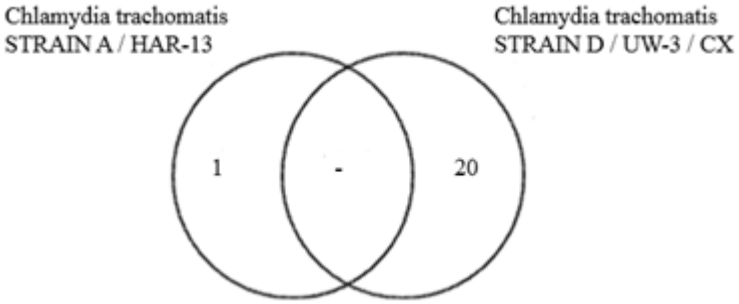


Figure 4.4. The Number of Chylamydiceae Strains-Targeted Human Proteins that are Grouped based on their Interactions with Strains; *Chlamydia Trachomatis STRAIN A / HAR-13* and *Chlamydia Trachomatis STRAIN D / UW-3 / CX.*

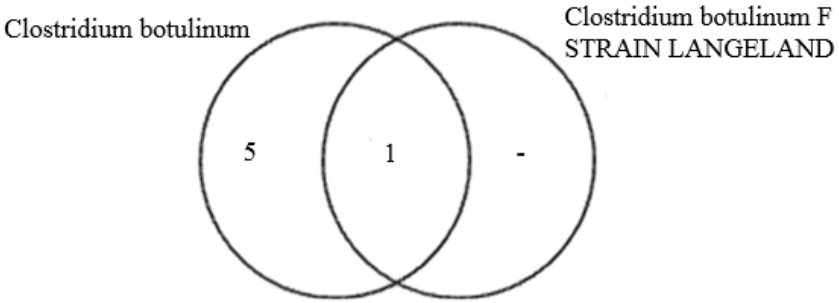


Figure 4.5. The Number of *Clostridiaceae* Strains-Targeted Human Proteins that are Grouped based on their Interactions with Strains; *Clostridium Botulinum* and *Clostridium Botulinum F STRAIN LANGELAND.*

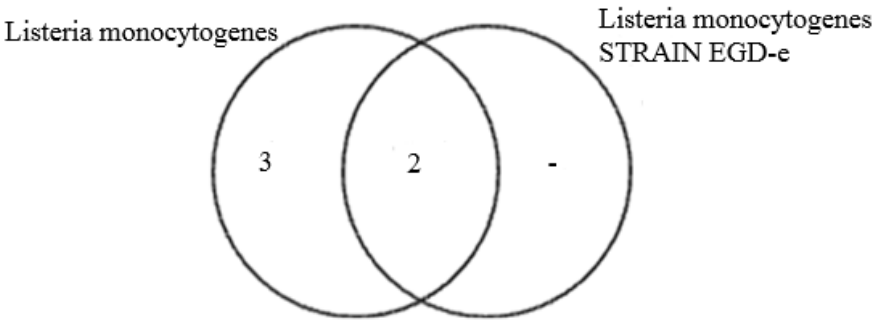


Figure 4.6. The Number of *Listeriaceae* Strains-Targeted Human Proteins that are Grouped based on their Interactions with Strains; *Listeria Monocytogenes* and *Listeria Monocytogenes STRAIN EGD-E.*

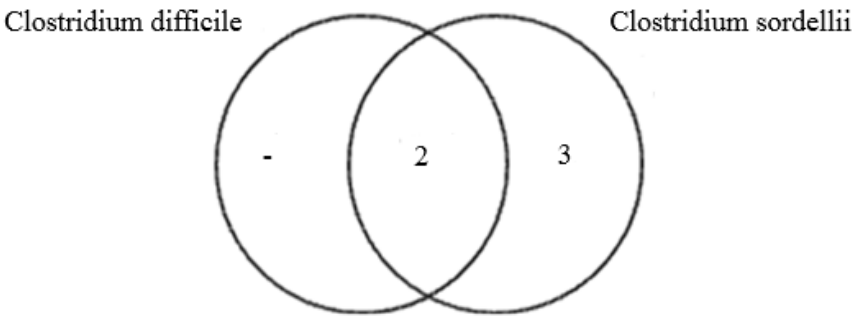


Figure 4.7. The Number of *Peptostreptococcaceae* Strains-Targeted Human Proteins that are Grouped based on their Interactions with Strains; *Clostridium Difficile* and *Clostridium Sordellii.*

The common human protein targeted by all two strains of *Clostridiaceae* is VAMP2 (Figure 4.5), vesicle-associated membrane protein 2, which is involved in the targeting and/or fusion of transport vesicles to their target membrane.

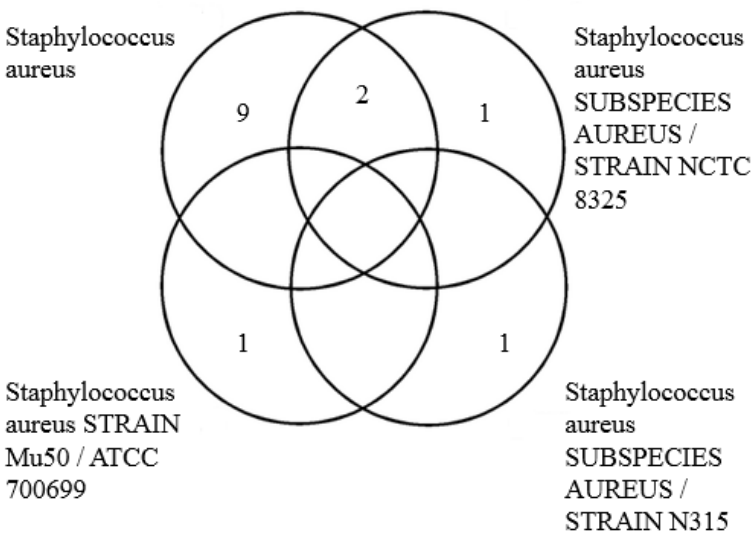


Figure 4.8. The Number of *Staphylococcacea* Strains-Targeted Human Proteins that are Grouped based on their Interactions with Strains; *Staphylococcus Aureus, Staphylococcus Aureus SUBSPECIES AUREUS / STRAIN NCTC 8325, Staphylococcus Aureus STRAIN Mu50/ATCC 700699* and *Staphylococcus Aureus SUBSPECIES AUREUS / STRAIN N315.*

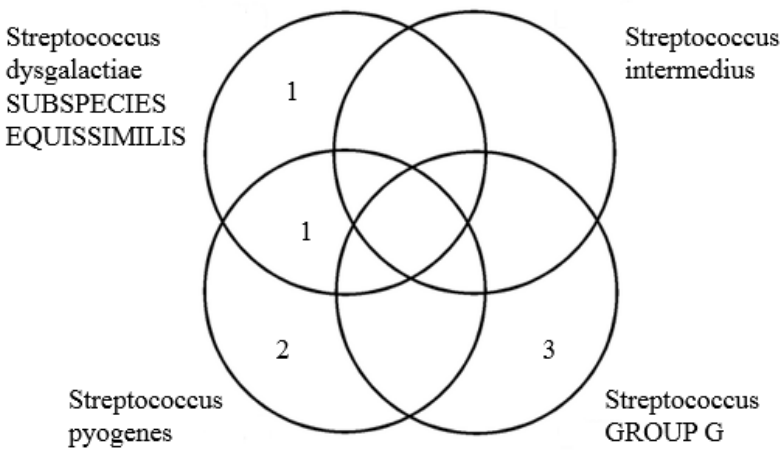


Figure 4.9. The Number of *Streptococcaceae* Strains-Targeted Human Proteins that are Grouped based on their Interactions with Strains; *Streptococcus Dysgalactiae SUBSPECIES EQUISSIMILIS, Streptococcus Intermedius, Streptococcus Pyogenes* and *Streptococcus GROUP G*.

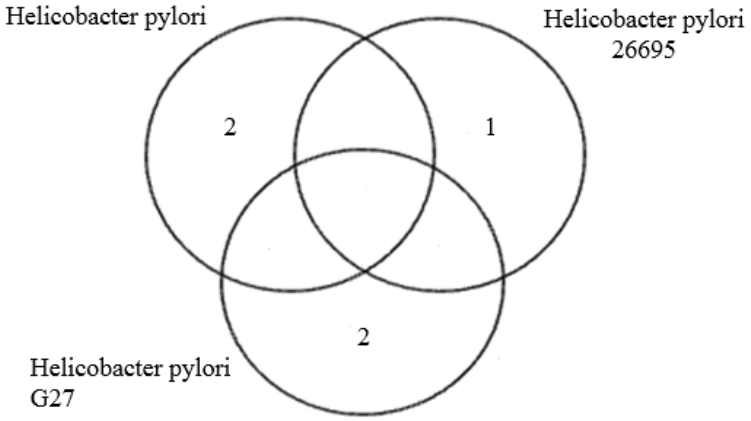


Figure 4.10. The Number of *Helicobacteriaceae* Strains-Targeted Human Proteins that are Grouped based on their Interactions with Strains; *Helicobacter Pylori, Helicobacter Pylori 26695* and *Helicobacter Pylori G27*.

The common human proteins targeted by all two strains of *Listeriaceae* are MET and CADH1 (Figure 4.6). MET, Hepatocyte growth factor receptor, regulates many physiological processes including proliferation, scattering, morphogenesis and survival. CADH1, cadherin-1, is involved in mechanisms regulating cell-cell adhesions, mobility and proliferation of epithelial cells.

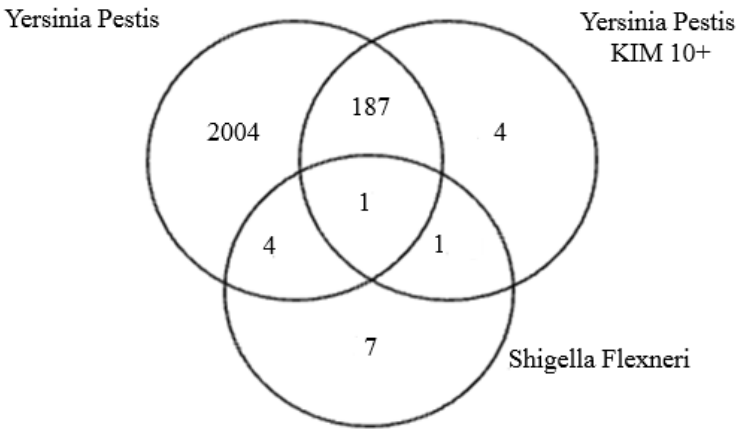


Figure 4.11. The Number of *Enterobacteriaceae* Strains-Targeted Human Proteins that are Grouped based on their Interactions with Strains; *Yersinia Pestis, Yersinia Pestis KIM 10+* and *Shigella Flexneri.*

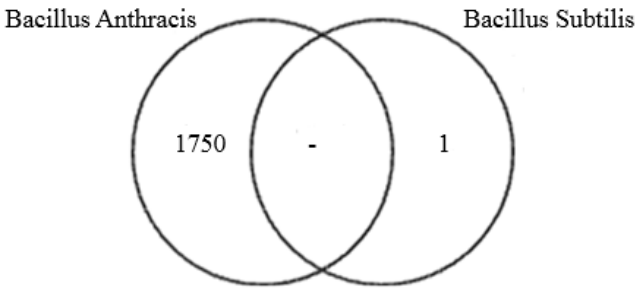


Figure 4.12. The Number of *Bacillaceae* Strains-Targeted Human Proteins that are Grouped based on their Interactions with Strains; *Bacillus Anthracis* and *Bacillus Subtilis.*

The commonly targeted human protein (by the two strains of *Listeriaceae*) MET activating mutations may be involved in the development of a highly malignant, metastatic syndrome known as cancer of unknown primary origin. Defects in MET may be associated with gastric cancer. CADH1 is also involved in some cancer types such as; hereditary diffuse gastric cancer, endometrial cancer, ovarian cancer and breast cancer (EMBL-EBI, 2014). As explained in Chapter 4.1.2, human protein RAC has been reported in colorectal, pancreatic, breast, and testicular cancers and in various leukemias (Fritz *et al.*, 1999; Schnelzer *et al.*, 2000; Wang *et al.*, 2009).

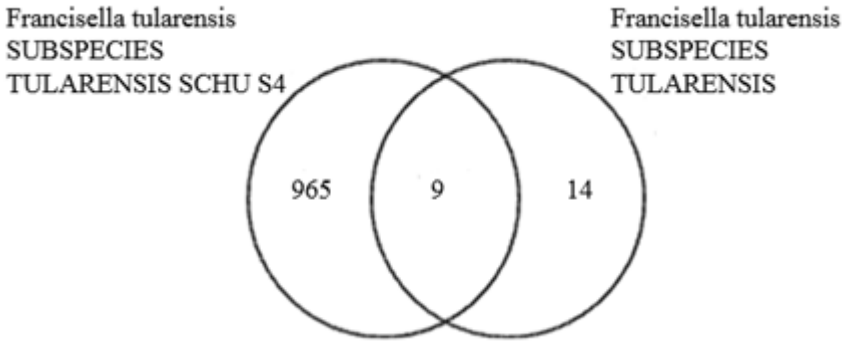


Figure 4.13. The Number of *Francisellaceae* Strains-Targeted Human Proteins That are Grouped Based On Their Interactions with Strains; *Francisella Tularensis SUBSPECIES TULARENSIS SCHU S4* and *Francisella Tularensis SUBSPECIES TULARENSIS*.

The commonly targeted human proteins (Table 4.35) by all three strains of *Enterobacteriacea* is RhoA, Ras homolog gene family member A and transforming protein RhoA (Figure 4.12); it is a small GTPase protein, serving as a target for the yopT cysteine peptidase from Yersinia pestis (causes gastrointestinal disorders). RhoA is enriched with the GO terms as follows; GTPase activity, transforming growth factor beta receptor signaling pathway, small GTPase mediated signal transduction, Rho protein signal transduction, response to mechanical stimulus, cell junction, platelet activation, positive regulation of cell growth, positive regulation of cytokinesis, response to drug, positive regulation of I-kappaB kinase/NF-kappaB cascade, negative regulation of I-kappaB kinase/NF-kappaB cascade, negative regulation of neuron differentiation, positive regulation of translation, and positive regulation of cell adhesion.

Table 4.34. Commonly targeted human proteins by different bacteria families.

Bacterial Families	**Commonly Targeted Human Proteins**
Francisellaceae	TCTP, IGHA2, 1B42, TEF, NFKB1, POGZ, HBA, GDIR2, EMIL1
Enterobacteriaceae	RhoA
Peptostreptococcaceae	RASN, RAC1
Listeriaceae	MET, CADH1
Clostridiaceae	VAMP2

Rho GTPases control multiple cellular processes, including actin and microtubule dynamics, gene expression, the cell cycle, cell polarity regulating cell shape, polarity and locomotion through their effects on actin polymerization, actomyosin contractility, cell adhesion, microtubule dynamics and membrane transport, through their ability to bind to numerous downstream effectors, which lead to diverse parallel downstream signaling pathways (Schwartz, 2004; Buchsbaum, 2007). RhoA plays a central role in the KEGG pathways of bacterial invasion of epithelial cells, Ras signalling pathway, endocytosis, T cell receptor signaling pathway, regulation of actin cytoskeleton, pathways in cancer and colorectal cancer. Several types of human cancers (breast, ovarian, renal, lung and colon) have been analyzed for RhoA mutations (Ridley, 2013). *Bacillus anthracis* and *Bacillus subtilis* seem to have their own infection strategies as they do not target any common human proteins (Figure 4.12).

The strains of Francisellaceae target nine common human proteins (Figure 4.11), namely; TCTP (Translationally-controlled tumor protein), IGHA2 (Ig alpha-2 chain C region), 1B42 (HLA class I histocompatibility antigen, B-42 alpha chain), TEF (Thyrotroph embryonic factor), NFKB1 (Nuclear factor NF-kappa-B p105 subunit), POGZ (Pogo transposable element with ZNF domain), HBA (Hemoglobin subunit alpha), GDIR2 (Rho GDP-dissociation inhibitor 2) and EMIL1 (EMILIN-1). These commonly targeted human proteins by the strains of *Francisellaceae* are enriched with the KEGG pathway, namely neurotrophin signaling pathway, which is regulated by connecting a variety of intracellular signaling cascades, that include MAPK pathway, PI-3 kinase pathway, and PLC pathway, transmitting positive signals like enhanced survival and growth.

Moreover, PAX4, as a transcription factor, may regulate the common human proteins targeted by Francisellaceae strains, containing the motif GGGTGGRR (the TFT is GGGTGGRR_V$PAX4_03). The transcription factor PAX4 is expressed in the developing pancreas (Smith *et al*., 1999) and it has been identified only as a regulator of endocrine development (Sosa-Pineda *et al*., 1997).

4.4. Comparison of Gram positive and Gram negative Bacterial Families and Strains with the Largest PHI Data

4.4.1. Enterobacteriaceae-Bacillaceae-Francisellaceae Families

Enterobacteriaceae (Gram negative, extracellular), *Bacillaceae* (Gram positive) and *Francisellaceae* (Gram negative, extracellular) are bacterial families which have the large-scale PHI data obtained by high-throughput experiments. These PHI data cover 99% of the total data (Table 3.1) and are therefore expected to reflect the behaviour of the bacteria types. The distribution of the human proteins attacked by these bacteria and their contents are given in the Table 4.36 and Figure 4.14.

These three families may have common infection strategies as they share 275 human proteins as target. Although *Enterobacteriaceae* and *Francisellaceae* families belong to Gram negative bacteria and *Bacillaceae* family belongs to Gram positive bacteria, it is surprisingly seen that *Enterobacteriaceae* and *Bacillaceae* commonly target 772 human proteins which is more than the amount of human proteins targeted by two Gram negative families *Enterobacteriaceae* and *Francisellaceae*.

Table 4.35. Contents of PHI data belonging to bacterial families with the largest data.

Bacteria family	**# of strains**	**# of PHIs**	**# of pathogen proteins**	**# of human proteins**
Enterobacteriaceae	15	4455	1304	2244
Francisellaceae	2	1341	346	988
Bacillaceae	2	3182	943	1751
TOTAL	19	8978	2593	3645

The significantly enriched GO terms of human proteins commonly targeted by *Enterobacteriaceae, Bacillaceae* and *Francisellaceae* are negative regulation of biological process, regulation of biological process, biological regulation and intracellular transport (Table 4.34). As *Enterobacteriaceae* and *Francisellaceae* are intracellular Gram negative bacteria, intracellular transport stands out among other terms. Negative regulation of

biological process, regulation of biological process and biological regulation are seen in almost every GO enrichment analysis in Chapter 4.1.2.

KEGG pathways enriched in human proteins commonly targeted by *Enterobacteriaceae*, *Bacillaceae* and *Francisellaceae* reveal that focal adhesion, pathways in cancer, prostate cancer, pancreatic cancer, colorectal cancer and small cell lung cancer are significant (Table 4.38). Especially, human proteins RHOA, HSP90B1 and EGFR are again found to have important roles in these pathways (Table 4.15 and Table 4.37).

Table 4.36. First 20 enriched GO Process terms of human proteins (275) targeted commonly by *Enterobacteriaceae, Bacillaceae* and *Francisellaceae.*

GO Process Term	**p-value**
negative regulation of biological process	3.35E-04
regulation of biological process	2.04E-03
biological regulation	3.18E-03
cytoskeleton organization	7.34E-03
cellular component organization	1.23E-02
cellular macromolecule localization	1.81E-02
cellular process	5.22E-02
gene expression	5.99E-02
cellular protein localization	6.06E-02
mRNA metabolic process	8.62E-02
organelle organization	8.97E-02
negative regulation of cellular process	9.61E-02
negative regulation of apoptosis	1.11E-01
response to inorganic substance	1.13E-01
cell death	1.18E-01
RNA metabolic process	1.20E-01
negative regulation of programmed cell death	1.33E-01
death	1.34E-01
regulation of cellular component organization	1.41E-01
intracellular transport	1.51E-01

It is observed that the most significant transcription factor may be SP1 (Table 4.23-4.26 and 4.39), which may regulate *Enterobacteriaceae, Bacillaceae* and *Francisellaceae* bacteria-targeted common human proteins by binding to the target sequence GGGCGGR. The TFT namely hsa_GGGCGGR_V$SP1_Q6 matches annotation with the TF SP1. SP1 is involved in many cellular processes, including cell differentiation, cell growth, apoptosis, immune responses, response to DNA damage, and chromatin remodeling as mentioned before in Chapter 4.1.5.

Table 4.37. First 10 enriched KEGG pathways in human protein set (275) targeted commonly by *Enterobacteriaceae, Bacillaceae* and *Francisellaceae.*

KEGG Pathways	**adjP-value**	**p-value**
Focal adhesion	2.20E-05	2.32E-07
Pathways in cancer	2.69E-05	7.93E-07
Amoebiasis	2.69E-05	8.51E-07
Viral myocarditis	4.00E-04	2.65E-05
Prostate cancer	4.00E-04	1.55E-05
Leishmaniasis	4.00E-04	3.19E-05
Pancreatic cancer	4.00E-04	2.65E-05
Colorectal cancer	4.00E-04	1.00E-04
Small cell lung cancer	1.10E-04	9.33E-05
Adherens junction	2.60E-04	3.00E-04

Table 4.38. First 8 enriched TFTs in human proteins targeted commonly by *Enterobacteriaceae, Bacillaceae* and *Francisellaceae.*

TFT	**adjP-value**	**p-value**
hsa_GGGCGGR_V$SP1_Q6	4.20E-07	7.71E-10
hsa_CTTTGT_V$LEF1_Q2	1.52E-05	5.56E-08
hsa_V$SMAD_Q6	9.25E-05	5.09E-07
hsa_GGGTGGRR_V$PAX4_03	2.00E-04	1.16E-06
hsa_V$AHRARNT_01	7.00E-04	7.68E-06
hsa_MGGAAGTG_V$GABP_B	7.00E-04	6.82E-06
hsa_TTGTTT_V$FOXO4_01	9.00E-04	1.79E-05
hsa_V$STAT6_02	9.00E-04	1.73E-05

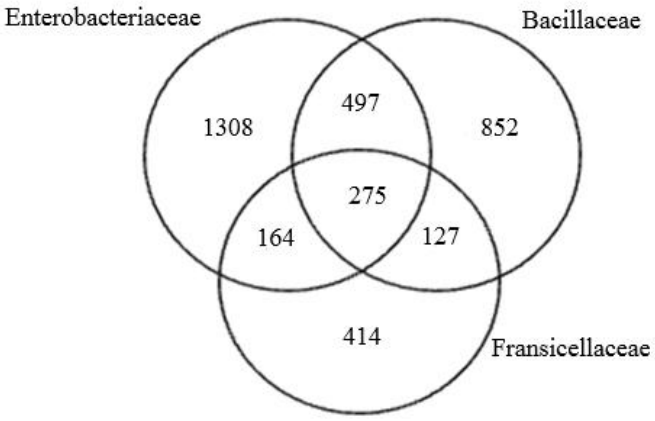


Figure 4.14. The Number of Bacteria Families-Targeted Human Proteins that are Grouped based on their Interactions with Families; *Enterobacteriaceae*, *Bacillaceae* and *Francisellaceae.*

4.4.2. Bacterial Strains with the Largest PHI Data

For a deeper strain-based analysis of PHI data, three bacterial strains of *Enterobacteriacea* with the largest amount of data were selected among its 15 strains. As *Bacillaceae* and *Francisellaceae* have only two strains with limited amount of data, only two strains of these bacterial families were analyzed (Table 4.40).

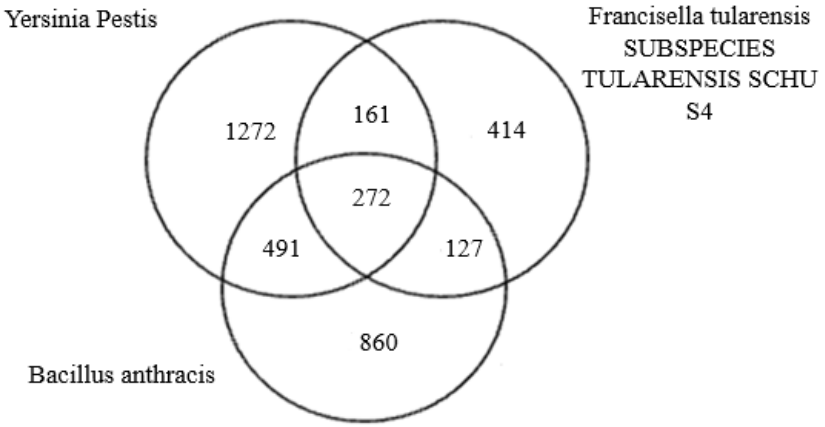


Figure 4.15. The Number of Bacterial Strains-Targeted Human Proteins that are Grouped based on their Interactions with Families; *Yersinia Pestis*, *Bacillus Anthracis* and *Francisella Tularensis SUBSPECIES TULARENSIS SCHU S4.*

Table 4.39. Contents of PHI data belonging to bacteria strains with the largest data.

Families	**Bacteria strains**	**# of PHIs**	**# of pathogen proteins**	**# of human proteins**
Enterobacteriacea	*Yersinia Pestis*	4203	1227	2196
	Yersinia Pestis KIM 10+	217	47	193
	Shigella Flexneri	20	9	13
	TOTAL	4395	1267	2208
Bacillaceae	*Bacillus anthracis*	3181	942	1750
	Bacillus subtilis	1	1	1
	TOTAL	3182	943	1751
Francisellaceae	*Francisella tularensis SUBSPECIES TULARENSIS SCHU S4*	1317	342	974
	Francisella tularensis SUBSPECIES TULARENSIS	24	4	23
	TOTAL	1341	346	988

The distribution of the human proteins targeted by the strains having largest-scale data of *Enterobacteriaceae, Bacillaceae* and *Francisellaceae* are shown in the Figure 4.15.

It is predicted that the bacreial strains Yersinia pestis, *Bacillus anthracis* and *Francisella tularensis SUBSPECIES TULARENSIS SCHU S4* behave similarly with their families *Enterobacteriaceae, Bacillaceae* and *Francisellaceae* as they contain about 99 % of their family based PHIs and hence attack similar numbers of human proteins.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

The following conclusions can be drawn on the basis of the results presented:

- Gram negative bacteria target 2822 human proteins while Gram positive bacteria target 1778 human proteins which is nearly 63% of the former.

- Human proteins (452) targeted by two Gram negative bacteria have five common proteins with those targeted by two Gram positive bacteria. However, the highly targeted human proteins of Gram negative bacteria (three Gram negative bacteria-targeted set) do not have any common human proteins with those targeted by two and three Gram positive bacteria. Thus, one can conclude that Gram positive and Gram negative bacteria may have different infection strategies.

- GO enrichment analysis of the human proteins targeted by pathogens denote important information about the infection mechanisms. As the main infection strategy, human proteins targeted by all Gram positive and Gram negative bacteria are enriched in the regulation of biological processes. Compared to Gram positive bacteria, Gram negative bacteria interact with the human proteins that function in immune system in order to block the human defense mechanism.

- From the analysis of enriched KEGG pathways, B cell receptor signaling pathway, endocytosis, Toll-like receptor signaling (TLR) pathway, cancer pathways and mitogen-activated protein kinase (MAPK) signaling pathway were found to be significant. Gram negative bacteria-targeted human proteins are enriched in pathways related to immunity, on the other hand, Gram positive bacteria-targeted human proteins are associated with pathways in cancer and found to disrupt human defense mechanism.

- From the analysis of pathway commons, toll like receptor 4 (TLR4) cascade, TRAF6 mediated induction of NFkB and MAP kinases upon TLR7/8 or 9 activation, toll like receptor 5 (TLR5) cascade, MyD88:Mal cascade initiated on plasma membrane are found to be significant for Gram negative bacteria targeted human proteins, while Alpha9 beta1 integrin signaling events, IFN-gamma pathway, signaling events mediated by focal adhesion kinase, IL3-mediated signaling events and EGF receptor (ErbB1) signaling pathway are found to be significant for those targeted by Gram positive bacteria.

- The highly encountered target sequence is GGGCGGR and the corresponding transcription factor is SP1 which contains the binding site of the human proteins COL1A2 and COL1A1 targeted by Gram positive bacteria. These proteins have crucial functions in the defense against cancer.

- The Gram negative bacteria, *Enterobacteriaceae*, *Chylamydiaceae* and *Neisseriaceae* may behave similarly as they share 10 targeted human proteins. *Fransicellaceae*'s attacking behavior seems to be specific to itself as 549 human proteins are targeted only by *Fransicellaceae*. It can be observed from the interactions between the proteins of Gram positive bacteria and the targeted human proteins that *Bacillaceae* having more than 97% of PHIs belonging to Gram positive bacteria reflects their infection strategies. It also shares targeted human proteins with other Gram positive bacteria.

- The common human proteins (RASN and RAC1) targeted by two strains of *Peptostreptococcaceae* and the human proteins (MET and CADH1) targeted by two strains of *Listeriaceae* indicate the relationship between cancer diseases and Gram positive bacteria. The commonly targeted human proteins by all three strains of Gram negative bacteria family *Enterobacteriacea* is RhoA which function in immune system. The KEGG pathways of the commonly targeted human proteins by the strains of *Francisellaceae* show that *Francisellaceae* mainly aims to attack human immune system.

- It is shown that *Enterobacteriaceae* (Gram negative, extracellular), *Bacillaceae* (Gram positive) and *Francisellaceae* (Gram negative, extracellular) may have common infection strategies as they share 275 common human proteins. It is surprisingly seen that *Enterobacteriaceae* and *Bacillaceae* commonly target 772 human proteins which is more than the amount of human proteins targeted by two Gram negative families of *Enterobacteriaceae* and *Francisellaceae.*

5.2. Recommendations

This study aims to provide initial insights on bacterial infection mechanisms through PHI networks with high-throughput experimental data. This kind of thorough analysis of bacterial infection mechanism was missing due to the insufficient data. In the framework of this thesis, bacterial infection mechanisms were investigated and the properties of the targeted human proteins and targeting pathogen proteins were analyzed to enlighten infection mechanisms. Nonetheless, a lot more studies should be done in order to understand the pathogenesis of infections completely. The following analysis can be performed for the future work:

- The types of bacteria should be compared as intracellular versus extracellular to see whether their targeting behaviours differ or not.

- To enlighten the infection strategies deeply, graph theoretical properties (degree, betweenness centrality, closeness centrality, etc.) of the bacteria-targeted human proteins within the human intraspecies protein interaction network should be compared to the non-targeted ones.

- In addition to investigation of targeted human proteins, pathogen proteins, which have crucial roles in infection (e.g. highly connected pathogen proteins in pathogen-human protein networks or pathogen proteins interacting with a human hub protein), can be investigated within the intraspecies protein interaction network of bacterial pathogens.

- Motif analysis of the PHI networks should be performed to provide insights on the interaction patterns between proteins of bacteria and human.

APPENDIX A: GO MAPS

A.1. GO Enrichment Map of Human Proteins Targeted by All Bacteria

GO enrichment map of human proteins targeted by all bacteria obtained by using WebGestalt software (Zhang *et al*., 2005) can be found in the attached CD.

A.2. GO Enrichment Map of Human Proteins Targeted by Both Gram Negative and Gram Positive Bacteria

GO enrichment map of human proteins targeted by both gram negative and gram positive bacteria obtained by using WebGestalt software (Zhang *et al*., 2005) can be found in the attached CD.

A.3. GO Enrichment Map of Human Proteins Targeted by Only Gram Negative Bacteria

GO enrichment map of human proteins targeted by only gram negative bacteria obtained by using WebGestalt software (Zhang *et al*., 2005) can be found in the attached CD.

A.4. GO Enrichment Map of Human Proteins Targeted by Only Gram Positive Bacteria

GO enrichment map of human proteins targeted by only gram positive bacteria obtained by using WebGestalt software (Zhang *et al*., 2005) can be found in the attached CD.

A.5. GO Enrichment Map of Three-Gram Negative Bacteria Set

GO enrichment map of human proteins targeted by three gram negative bacteria obtained by using WebGestalt software (Zhang *et al*., 2005) can be found in the attached CD.

A.6. GO Enrichment Map of Two-Gram Positive Bacteria Set

GO enrichment map of human proteins targeted by two gram positive bacteria obtained by using WebGestalt software (Zhang *et al*., 2005) can be found in the attached CD.

REFERENCES

Perrin, A. J., X. Jiang, C.L. Birmingham, N.S. So, J.H. Brumell, 2004, "Recognition of Bacteria in the Cytosol of Mammalian Cells by the Ubiquitin System", *Current Biology*, Vol. 14, pp. 806–811.

Abbas, A. K., A. H. Lichtman, and S. Pilla, 2010, *Cellular and Molecular Immunology*, Saunders/Elsevier, Philadelphia.

Abbas, A. K., 2011, *Basic Immunology: Functions and Disorders of the Immune System*, Saunders, Philadelphia, Pa.

Acuner Özbabacan, S. E., S. Durmuş Tekir, K. Ülgen, 2011, "Crucial Components in Toll-like Receptor Signalling Network as Key Regulators for Immunological Disorders", *International Journal of Physical Sciences*, Vol. 6, pp. 2016-2026.

Alberts, B., A. Johnson, J. Lewis, 2002, "Molecular Biology of the Cell", 4th edition, *Garland Science*, New York.

Alonso, A., N. Bottini, S. Bruckner, S. Rahmouni, S. Williams, S. P. Schoenberger, T. Mustelin, 2004, "Lck Dephosphorylation at Tyr-394 and Inhibition of T Cell Antigen Receptor Signaling by *Yersinia* Phosphatase YopH", *The Journal of Biological Chemistry*,Vol. 279, pp. 4922-4928.

Arifuzzaman, M., M. Maeda, A. Itoh, K. Nishikata, C. Takita, R. Saito, T. Ara, K. Nakahigashi, H. C. Huang, A. Hirai, K. Tsuzuki, S. Nakamura, M. Altaf-Ul-Amin, T. Oshima, T. Baba, N. Yamamoto, T. Kawamura, T. Ioka-Nakamichi, M. Kitagawa, M. Tomita, S. Kanaya, C. Wada, H. Mori, 2006, "Large Scale Identification of Protein-Protein Interaction of *Escherichia coli* K-12", *Genome Research*, Vol. 16, pp. 686-691.

Ashburner, M., C. Ball, J. Blake, D. Botstein, 2000, "Gene Ontology: Tool for the Unification of Biology", *Nature*, Vol. 25, pp. 25-29.

Atlas, R. M., 1995, *Microorganisms in Our World*, Mosby-Year Book, St. Louis, MO.

Barnett, J. A., 2003, "Beginnings of Microbiology and Biochemistry: The Contribution of Yeast Research", *Microbiology*, Vol. 149, pp. 557–567.

Beaglehole, R., 2004, "The World Health Report 2004 - Changing History", *World Health Organization*, pp. 120-124.

Beek, D. V. D., 2012, "Progress and Challenges in Bacterial Meningitis", *The Lancet*, Vol. 380, pp. 1623-1624.

Beutler, B., 2004, "Innate Immunity: An Overview", *Molecular Immunology*, Vol. 40, pp. 845–859.

Borriello, S. P., 1983, "Relative Carriage Rates of Nuclear Dehydrogenating *Clostridia* in Two Populations of Different Colorectal Cancer Risk", *Journal of Clinical Pathology*, Vol. 36, pp. 93-95.

Brook, I., 2007, "Anaerobic Infections Diagnosis and Management", *Informa Healthcare*, Inc. New York.

Buchsbaum, R.J., 2007, "Rho Activation at a Glance", *Journal of Cell Science*, Vol. 120, pp. 1149-1152.

Buck, J. D., 1982, "Nonstaining (KOH) Method for Determination of Gram Reactions of Marine Bacteria", *Applied And Environmental Microbiology*, pp. 992-993.

Buonaguro, L., E. Wang, M. L. Tornesello, F. M. Buonaguro, Marincola F. M., 2011, "Systems Biology Applied to Vaccine and Immunotherapy Development", *BMC Systems Biology*, Vol. 5, p. 146.

Butland, G., Peregrin-Alvarez J. M., Li J., Yang W., 2005, "Interaction Network Containing Conserved and Essential Protein Complexes in *Escherichia coli*", *Nature*, Vol. 433, pp. 531–537.

Calderwood, M. A., K. Venkatesan, L. Xing, M. R. Chase, A. Vazquez, A. M. Holthaus, A. E. Ewence, N. Li, T. Hirozane-Kishikawa, D. E. Hill, M. Vidal, E. Kieff, E. Johannsen, 2007, "Epstein-Barr Virus and Virus Human Protein Interaction Maps", *Proceedings of the National Academy of Sciences*, Vol. 104, pp. 7606-7761.

Casadevall, A., L. Pirofski, 2000, "Host-Pathogen Interactions: Basic Concepts of Microbial Commensalism, Colonization, Infection, and Disease", *Infection and Immunity*, Vol. 68 pp. 6511–6518.

Cavaillon, J. M., 2011, "The Historical Milestones in the Understanding of Leukocyte Biology Initiated by Elie Metchnikoff", *Journal of Leukocyte Biology*, Vol. 90, pp. 413-424.

Cho, H. S., D. J. Leahy, 2002, "Structure of the Extracellular Region of Her3 Reveals an Interdomain Tether", *Science*, Vol. 297, pp. 1330–1333.

Choi, Y., 2005, "Role of Traf6 in the Immune System", *Mechanisms of Lymphocyte Activation and Immune Regulation X, Advances in Experimental Medicine and Biology*, Vol. 560, pp. 77-82.

Ciechanover, A., O. Amir, L. S. Alan., 2000, "Ubiquitin-mediated Proteolysis: Biological Regulation via Destruction", *BioEssays*, Vol. 8, pp. 442-451.

Conti S., S. S. F. Santos, C. Y. Koga, and A. O. C. Jorge, 2009, "*Enterobacteriaceae* and *Pseudomonadaceae* On The Dorsum Of The Human Tongue", *Journal Appied Oral Science*, Vol. 17, No. 5, pp. 375–380.

Croft D., G. O'Kelly, G. Wu, R. Haw, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kalatskaya I, Mahajan S, May B, Ndegwa N, Schmidt

E, Shamovsky V, Yung C, Birney E, Hermjakob H, D'Eustachio P, Stein L. "Reactome: A Database of Reactions, Pathways and Biological Processes", *Nucleic Acids Research,* pp. 691-697.

Das, U., 2000, "Critical Advances in Septicemia and Septic Shock", *Critical Care,* Vol. 4, pp. 290-296.

DeChassey, B., V. Navratil, L. Tafforeau, M. S. Hiet, A. Aublin-Gex, S. Agaugue, G. Meiffren, F. Pradezynski, B. F. Faria, T. Chantier, M. Le Breton, J. Pellet, N. Davoust, P. E. Mangeot, A. Chaboud, F. Penin, Y. Jacob, P. O. Vidalain, M. Vidal, P. Andre, C. Rabourdin-Combe, and V. Lotteau, 2008, "Hepatitis C Virus Infection Protein Network", *Molecular Systems Biology*, Vol. 4, p. 230.

Donald, A., T. V. Henderson, T. O'Toole, and P. Philip, 2003, "Can Postexposure Vaccination against Smallpox Succeed?", *Clinical Infectious Diseases,* Vol. 36, pp. 622-629.

Dong, C., R.J. Davis, and R.A. Flavell, 2002, "Map Kinases in the Immune Response*", Annual Review of Immunology*, Vol. 20, pp. 55-72.

Durmuş Tekir, S., 2013, *Investigation of Infection Mechanisms Through Pathogen-Human Protein-Protein Interactions by Bioinformatics Approaches*, Ph.D. Thesis, Boğaziçi University.

Durmuş Tekir, S., T. Cakir, and K. Ulgen, 2012, "Infection Strategies of Bacterial and Viral Pathogens through Pathogen-Human Protein-Protein Interactions.," *Frontiers in Microbiology*, Vol. 3, pp. 46.

Dyer, M. D., C. Neff, M. Dufford, C. G. Rivera, D. Shattuck, J. Bassaganya-Riera, T. M. Murali, and B. W. Sobral, 2010, "The Human-Bacterial Pathogen Protein Interaction Networks of *Bacillus Anthracis*, *Francisella Tularensis*, and *Yersinia Pestis*", *PLoS One*, Vol. 5, pp. 120-131.

Dyer, M. D., T. M. Murali, and B. W. Sobral, 2008, "The Landscape of Human Proteins Interacting with Viruses and Other Pathogens", *PLoS Pathogens*, Vol. 4, p. 32.

Edwin, S. V. A., J. C. Theo, B. Van, and K. Johan, 2003, "Receptors, Mediators, and Mechanisms Involved in Bacterial Sepsis and Septic Shock", *Clinical Microbiology Reviews*, Vol. 16, pp. 379–414.

Egi, Y., M. Ito, S. Tanaka, S. Imagawa, S. Takata, M. Yoshihara, K. Haruma, and K. Chayama, 2007, "Role of *Helicobacter Pylori* Infection and Chronic Inflammation in Gastric Cancer in the Cardia", *Japanese Journal of Clinical Oncology*, Vol. 37, pp. 365–369.

Ehrlich, P., 2013, "Biographical", *Nobel Media A. B.*, www.nobelprize.org, accessed at June 2013.

Elek, S.D., 1959, "*Staphylococcus pyogenes* and Its Relation to Disease", *Edinburgh: E. & S. Livingstone*, p. 3.

Elliott, E. J., 2007, "Acute Gastroenteritis in Children", *British Medical Journal*, Vol. 334, pp. 35–40.

Etienne-Manneville, S., and A. Hall, 2002, "Rho GTPases in Cell Biology", *Nature*, Vol. 420, pp. 629-635.

Fauci, A. S., 2001, "Infectious Diseases: For Considerations the 21st Century", *Clinical Infectious Diseases*, Vol. *32*, pp. 675-685.

Ferlay, J., I. Soerjomataram, and M. Ervik, 2012, "Cancer Incidence and Mortality Worldwide: IARC", *Globocan*, Vol. 1, No. 11, pp. 1133-1145.

Finegold, S. M., 1977, "Anaerobic Bacteria in Human Disease", *Academic Press*, Vol. 23, pp. S97-S101.

Forst, C. V, 2006, "Host-Pathogen Systems Biology," *Drug Discovery Today*, Vol. 11, No. 5–6, pp. 220–227.

Fritz G., I. Just, and B. Kaina, 1999, "Rho GTPases are Over-expressed in Human Tumors", *International Journal of Cancer*, Vol. 81, No. 5, pp. 682–687.

Gallagher R.M., C. W. Gross, and C. D. Phillips, 1998, "Suppurative Intracranial Complications of Sinusitis", *Laryngoscope*, Vol. 108, pp. 1635-1642.

Galli, T., A. Zahraoui, V.V. Vaidyanathan, G. Raposo, J. M. Tian, M. Karin, H. Niemann, and D. M. Louvard, 1998, "A Novel Tetanus Neurotoxin-insensitive Vesicle-associated Membrane Protein in SNARE Complexes of the Apical Plasma Membrane of Epithelial Cells", *Biology of the Cell*, Vol. 9. pp. 1437-1448.

Goel, A., 2014, "Microbes In Agriculture And Medicine", *Bio Evolution*, Vol. 1, pp. 5-8.

Greaves, J. E., and O. E. Greaves., 1946, *Elementary Bacteriology*, W.B. Saunders, Philadelphia, London.

Guha, M., and N. Mackman, 2001, "LPS Induction of Gene Expression in Human Monocytes", *Cell Signal*, Vol. 13, pp. 85-94.

Hanahan D., and R. A. Weinberg, 2011, "Hallmarks of Cancer: The Next Generation", *Cell*, Vol. 144, pp. 646–674.

Hawker, L. E., 1960, *An Introduction to the Biology of Micro-organism*, London.

Hiss, P. H., 1928, *A Textbook of Bacteriology*, New York, London.

Horseman, M. A., R. Bray, F. B. Lujan, and E. Matthew, 2013, "Infections Caused by *Vibrionaceae:* A Review of the Epidemiology, Pathogenesis, Clinical Presentations, and Treatment", *Infectious Diseases in Clinical Practice*, Vol. 21, pp. 222-232.

Huang, H.R., Z. J. Chen, S. Kunes, G.-D. Chang, and T. Maniatis, 2010, "Endocytic Pathway Is Required for Drosophila Toll Innate Immune Signaling.," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 18, pp. 8322-8327.

Humphrey, T., S. O'Brien, and M. Mads, 2007, "Campylobacters as Zoonotic Pathogens: A Food Production Perspective", *International Journal of Food Microbiology*, Vol. 117, pp. 237-257.

Hyun, Y., K. P. Amit, P. Chhinder, D. J. Hackam, and S. H. Abigail, 2012, "Novel Role for the Innate Immune Receptor Toll-Like Receptor 4 (TLR4) in the Regulation of the Wnt Signaling Pathway and Photoreceptor Apoptosis", *PLoS ONE*, Vol. 7, pp. 365-370.

Irgens, L., 2002, "The Discovery of the *Leprosy Bacillus*", *Tidsskr nor Laegeforen,* Vol. 122, p. 7.

Iwasaki, A., and R. Medzhitov, 2004, "Toll-like Receptor Control of the Adaptive Immune Responses", *Nature Immunology*, Vol. 5, pp. 987-995.

Jager, S., P. Cimermancic, N. Gülbahçe, J. R. Johnson, K. E. McGovern, S. C. Clarke, M. Shales, G. Mercenne, L. Pache, K. Li, H. Hernandez, G. M. Jang, S. L. Roth, E. Akiva, J. Marlett, M. Stephens, I. D'Orso, J. Fernandes, M. Fahey, C. Mahon, A. J. O'Donoghue, A. Todorovic, J. H. Morris, D. A. Maltby, T. Alber, G. Cagney, F. D. Bushman, J. A. Young, S. K. Chanda, W. I. Sundquist, T. Kortemme, R. D. Hernandez, C. S. Craik, A. Burlingame, A. Sali, A. D. Frankel, and N. J. Krogan, 2012, "Global Landscape of HIV-Human Protein Complexes", *Nature*, Vol. 481, pp. 365-370.

Janda J. M., and L. S. Abbott, 2010, "The Genus Aeromonas: Taxonomy, Pathogenicity, and Infection", *Clinical Microbiology Reviews*, Vol. 23, No. 1, pp. 35–73.

Jassim S. S., A. Malik., and A. Aldridge, 2011, "Small Bowel Perforation: An Unusual Cause", *Grand Rounds*, Vol. 11, pp. 17–19.

Jones, J., 2000, "What Is the Role of Bacteria In Cancer Carcinogenesis?", *Journal of National Cancer Institute*, Vol. 92, p. 1713.

Kenneth, R., C. G. Ray, N. Ahmad, W. L. Drew, and J. Plorde, 2010, "Sherris Medical Microbiology", 5th edition, *McGraw Hill Medical*, New York.

Kaufmann, S. H., 1993, "Immunity to intracellular bacteria", *Annual Reviews of Immunology*, Vol. 11, pp. 129-163.

Kocazeybek, B., 2003, "Chronic Chlamydophila Pneumoniae Infection in Lung Cancer, a Risk Factor: A Case-control Study", *Journal of Medical Microbiology*, Vol. 52, pp. 721–726.

Kumar, V., A. K. Abbas, N. Fausto, and R. N. Mitchell, 2007, *Robbins Basic Pathology*, Saunders Elsevier, 8th edition, pp. 102-103.

Kühner, S., V. van Noort, M. J. Betts, A. Leo-Macias, C. Batisse, M. Rode, T. Yamada, T. Maier, S. Bader, P. Beltran-Alvarez, D. Castano-Diez, W. -H. Chen, D. Devos, M. Güell, T. Norambuena, I. Racke, V. Rybin, A. Schmidt, E. Yus, R. Aebersold, R. Herrmann, B. Böttcher, A. S. Frangakis, R. B. Russell, L. Serrano, P. Bork, and A. C. Gavin, 2009, "Proteome Organization in a Genome-Reduced Bacterium", *Science*, Vol. 326, pp. 1235-1240.

LaCount, D. J., M. Vignali, R. Chettier, A. Phansalkar, R. Bell, J. R. Hesselberth, L. W. Schoenfeld, I. Ota, S. Sahasrabudhe, C. Kurschner, S. Fields, and R. E. Hughes, 2005, "A Protein Interaction Network of the Malaria Parasite *Plasmodium falciparum*", *Nature*, Vol. 438, pp. 103-107.

Lai, X. and I. Golovliov, 2001, "*Francisella tularensis* Induces Cytopathogenicity and Apoptosis in Murine Macrophages via a Mechanism that Requires Intracellular Bacterial Multiplication", *Infection and Immunity*, Vol. 69, pp. 4691-4694.

Li, S., A. Strelow, E. J. Fontana, and H. Wesche, 2002, "IRAK-4: A Novel Member of the IRAK Family with the Properties of an IRAK-Kinase", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 99, No. 8, pp. 5567–5572.

Li, S., 2004, "A Map of the Interactome Network of the Metazoan *C. Elegans*", *Science*, Vol. 303, pp. 540–543.

Liberzon, A., A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, and J. P. Mesirov, 2011, "Molecular Signatures Database (MSigDB) 3.0.," *Bioinformatics*, Vol. 27, No. 12, pp. 1739–1740.

Lombard, M., P. P. Pastoret, and A. M. Moulin, 2007, "A brief history of vaccines and vaccination", *Revue Scientifique et Technique*, Vol. 26, pp. 29-48.

Lozano, R., M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, J. Abraham, T. Adair, R. Aggarwal, and S. Y. Ahn, 2012, "Global and Regional Mortality from 235 Causes of Death for 20 Age Groups in 1990 and 2010: A Systematic Analysis for the Global Burden of Disease Study 2010", *Lancet*, Vol. 380, pp. 2095–2128.

Lund, O., M. S. Nielsen, and C. Lundegaard, 2005, *Immunological Bioinformatics*, MIT Press.

Luo, M. and J. Guan, 2010, "Focal Adhesion Kinase: A Prominent Determinant in Breast Cancer Initiation, Progression and Metastasis", *Cancer Letters*, Vol. 289, No. 2, pp. 127–139.

Madigan, M. T., and J. M. Martinko, 2006, *Brock Biology of Microorganisms*, Pearson Prentice Hall Imprint, Upper Saddle River, NJ.

Maere, S., K. Heymans, and M. Kuiper, 2005, "BiNGO: A Cytoscape Plugin to Assess over Representation of Gene Ontology Categories in Biological Networks", *Bioinformatics*, Vol. 21, pp. 3448-3449.

Mahon, C. R., D. C. Lehman, G. Manuselis, 2007, *Textbook of Diagnostic Microbiology*, Saunders-Elsevier.

Manjunath, M., 2007, "Fatal Septicaemia due to *Chromobacterium Violaceum*", *West Indian Medical Journal*, Vol. 56, pp. 380-381.

Mann, J., 1996, *Bacteria and Anti-bacterical Agents*, University Science Books Imprint, Oxford.

Marchadier, E., R. Carballido-Lopez, S. Brinster, and C. Fabret, 2011, "An Expanded Protein-protein Interaction Network in *Bacillus subtilis* Reveals a Group of Hubs: Exploration by an Integrative Approach", *Proteomics*, Vol. 11, pp. 2981–2991.

Mikael, R., E. Sofia, C. Mark, B. Sven, and J. N. Staffan, 2003, "The Basis of Persistent Bacterial Infections", *Trends in Microbiology*, Vol. 11, No. 2, pp. 80-86.

Murray, W. R., A. Blackwood, J. M. Trotter, K. C. Calman, and C. Mackay, 1980, "Faecal Bile Acids And Clostridia In The Aetiology", *British Journal of Cancer*, Vol. 41, pp. 923–928

Namiki, K., S. Goodison, and S. Porvasnik, 2009, "Persistent Exposure to *Mycoplasma* Induces Malignant Transformation of Human Prostate Cells", *PLoS ONE*, Vol. 4, pp. 68-72.

Ning, J., and C. Shou, 2004, "*Mycoplasma* Infection and Cancer", *Chinese Journal of Cancer*, Vol 23, pp. 602–604.

O'Riordan, M., and D.A. Portnoy, 2002, "The Host Cytosol: Front-line or Home Front?", *Trends in Microbiology*, Vol. 10, pp. 361–364.

Oda, K. and H. Kitano, 2006, "A Comprehensive Map of the Toll-Like Receptor Signaling Network", *Molecular Systems Biology*, Vol. 2, pp. 1-20.

Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, 1999, "KEGG: Kyoto Encyclopedia of Genes and Genomes", *Nucleic Acids Research*, Vol. 27, pp. 29-34.

Ohneck, E.A., Y.M. Zalucki, and P. J. Johnson, 2011, "A Novel Mechanism of High-level, Broad-spectrum Antibiotic Resistance Caused by a Single Base Pair Change in *Neisseria Gonorrhoeae*", *mBio Journal*, Vol. 2, pp. 187-197.

Palusińska-Szysz, M. C., 2009, "Pathogenicity of the Family *Legionellaceae*", *Archivum Immunologiae et Therapiae Experimentalis*, Vol. 57, pp. 279-290.

Pappenheimer, A. M., 1984, "The *Diphtheria Bacillus* and Its Toxin: A Model System", *Journal of Hygiene*, Vol. 93, pp. 397-404.

Park, J. M., F. R. Greten, Z. -W Li, and M. Karin, 2002, "Macrophage Apoptosis by Anthrax Lethal Factor through P38 MAP Kinase Inhibition", *Science*, Vol. 297, pp. 2048-2051.

Parri, M., and Chiarugi P., 2010, "Rac and Rho GTPases in Cancer Cell Motility Control", *Cell Communication and Signaling*, Vol. 8, No. 23, pp. 1-14.

Parrish, J. R., J. Yu, G. Liu, and J. A. Hines, 2007, "A Proteome-wide Protein Interaction Map for *Campylobacter jejuni*", *Genome Biology*, Vol. 8, p. R130.

Penn, R. L., G. L. Mandell, J. E. Bennett, and R. Dolin, 2005, *Francisella tularensis* (tularemia), Douglas and Bennett's Principles and Practice of Infectious Diseases, Elsevier Churchill Livingston, pp. 2674-2685.

Peter, S., and C. Beglinger, 2007, "*Helicobacter Pylori* and Gastric Cancer: The Causal Relationship", *Digestion*, Vol. 75, pp. 25–35.

Phillips, T., and L. Hoopes, 2008, "Transcription Factors and Transcriptional Control in Eukaryotic Cells", *Nature Education*, Vol. 1, p. 119.

Picard, C., A. Puel, M. Bonnet, C. L. Ku, J. Bustamante, K. Yang, C. Soudais, S. Dupuis, J. Feinberg, C. Fieschi, C. Elbim, R. Hitchcock, D. Lammas, G. Davies, A. Al-Ghonaium, H. Al-Rayes, S. Al-Jumaah, S. Al-Hajjar, I. Z. Al-Mohsen, H. H. Frayha, R. Rucker, T. R. Hawn, A. Aderem, H. Tufenkeji, S. Haraguchi, N. K. Day, R. a Good, M.-A. Gougerot-Pocidalo, A. Ozinsky, and J.L. Casanova, 2003, "Pyogenic Bacterial Infections in Humans with IRAK-4 Deficiency," *Science*, Vol. 299, No. 5615, pp. 2076–2079.

Pickering L.K., C. J. Baker, D. W. Kimberlin, and S. S. Lang, 2009, *Group A Streptococcal Infections*, Red book: Report of the Committee on Infectious, American Academy of Pediatrics, 28th edition.

Porter, J. R., 1976, "Antony van Leeuwenhoekl: Tercentenary of His Discovery of Bacteria", *American Society for Microbiology*, Vol. 40, pp. 260-269.

Pruitt, K. D., T. Tatusova, and D. R. Maglott, 2005, "NCBI Reference Sequence (RefSeq): A Curated Non-Redundant Sequence Database of Genomes, Transcripts and Proteins," *Nucleic Acids Research*, Vol. 33, pp. D501-D504.

Rain, J. C., L. Selig,, H. DeReus, and V. Battaglia, 2001, "The Protein-Protein Interaction Map of *Helicobacter Pylori*", *Nature*, Vol. 409, pp. 211–215.

Ridley, J., 2013, "RhoA, RhoB and RhoC Have Different Roles in Cancer Cell Migration," *Journal of Microscopy*, Vol. 251, No. 3, pp. 242–249.

Roper, W. L., 2008, "Diphtheria, Tetanus, and Pertussis: Recommendations for Vaccine Use and Other Preventive Measures", *Immunization Practices*, Vol. 28, pp. 1-28.

Rual, J. F., K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxern, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Liamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, and M. Vidal, 2005, "Towards a Proteome-Scale Map of the Human Protein-Protein Interaction Network", *Nature*, Vol. 437, pp. 1173-1178.

Ruuskanen, O., E. Lahti, L.C. Jennings, and D.R. Murdoch, 2011, "Viral Pneumonia", *Lancet*, Vol. 377, No. 9773, pp. 1264-1275.

Ryan K.J., and C. G. Ray, 2004, *Sherris Medical Microbiology* , 4th edition, McGraw Hill, pp. 463–470.

Sakula, A., 1983, "Robert Koch: Centenary of the Discovery of the *Tubercle Bacillus*, 1882", *The Canadian Veterinary Journal*, Vol. 24, No. 4, pp. 127–131.

Salle, A. J., 1996, *Fundamental Principles of Bacteriology*, Tata McGraw - Hill Publishing Company, 7th edition.

Santos S. S.F.D., C. K. Koga-Ito, A. O. C. Jorge, and B. H. Iglewski, 1996, *Pseudomonas*, Medical Microbiology, 4th edition.

Schaechter, M., J. L. Ingraham, and F. C. Neidhardt, 2006, *Microbe*, American Society for Microbiology, Washington, D.C.

Schnelzer, A., D., U. Prechtel, K. Knaus, G. M. Dehne, H. Graeff, N. Harbeck, M. Schmitt, and E. Lengyel, 2000, "Rac1 in Human Breast Cancer: Overexpression, Mutation Analysis, and Characterization of a New Isoform, Rac1b", *Oncogene*, Vol. 19, pp. 3013–3020.

Schroder, K., P. J. Hertzog, T. Ravasi, and D. A. Hume, 2004, "Interferon: An Overview of Signals, Mechanisms and Functions", *Journal of Leukocyte Biology*, Vol. 75, No. 2, pp. 163-189.

Schwartz, M., 2004, "Rho Signalling at a Glance", *Journal of Cell Science*, Vol. 117, pp. 5457-5458.

Shapira, S., I. Gat-Viks, B. Shum, and A. Dricot, 2009, "A Physical and Regulatory Map of Host-Influenza Interactions Reveals Pathways in H1N1 Infection", *Cell*, Vol. 139, pp. 1255-1267.

Shiotani A., and D. Y. Graham, 2002, "Pathogenesis and Therapy of Gastric and Duodenal Ulcer Disease", *Medical Clinics of North America*, Vol. 86, pp. 1447–1466.

Sosa-Pineda, B., K. Chowdhury, M. Torres, G. Oliver, and P. Gruss, 1997, "The Pax4 Gene is Essential for Differentiation of Insulin-producing β Cells in the Mammalian Pancreas", *Nature*, Vol. 386 pp. 399–402.

Stelzl, U., U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksöz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, and E. E. Wanker, 2005, "A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome", *Cell*, Vol. 122, pp. 957-968.

Stuart B. S., C. E. Hooi, R. C. Jennifer, and S. G. Michael, 1999, "Paired-Homeodomain Transcription Factor PAX4 Acts as a Transcriptional Repressor in Early Pancreatic Development", *Molecular and Cellular Biology*, Vol. 19, No. 12, pp. 8272–8280.

Susan L. F., M. Arnett, and C. P. Sinave, 2014, "*Enterobacter* Infections", *Medscape*, Vol. 25, p.15.

Titz, B., S. V. Rajagopala, J. Goll, and R. Hauser, 2008, "The Binary Protein Interactome of *Treponema pallidum* - the Syphilis Spirochete", *PLoS One*, Vol. 3, p. 229.

Valcheva-Komitska, I., L. Dimitrova, Y. Borisova, R. Nenova, I. N. Ivanov, and P. Kircheva, 2007, "A Case of Septic Anthrax - Diagnostic Problem and Perspectives in the Contemporary Microbiological Diagnosis", *Medical Review*, Vol. 43, No. 2, pp. 94-98.

Van Eijk, J. P., A. C., Giessen., S. J., Toxopeus, 1976, "Perspectives of Breeding for Resistance to 'yellow disease', *Xanthomonas hyacinthi* (Wakker) Dowson, in hyacinths (*Hyacinthus L.*)", *Euphytica*, Vol. 25, p. 131.

Verhoef J., and E. Mattsson, 1995, "The Role of Cytokines in Gram-positive Bacterial Shock", *Trends in Microbiology*, Vol. 3, pp. 136-140.

Vos, P., Garrity, G., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., Schleifer, K.H., and Whitman, W., 2009, *Bergey's Manual of Systematic Bacteriology*, Williams & Wilkins, 2nd edition.

Walker T., 1998, *Microbiology*, WB Saunders Company, Philadelphia.

Walker, L., H. Levine, and M. Jucker, 2006, "Koch's Postulates and Infectious Proteins", *Acta Neuropathologica*, Vol. 112, pp. 1-4.

Wang J., Rao Q., Wang M., Wei H., Xing H., Liu H., Wang Y., Tang K., Peng L., Tian Z., 2009, "Overexpression of Rac1 in Leukemia Patients and Its role in Leukemia Cell Migration and Growth", *Biochemical and Biophysical Research Communications*, Vol. 386, pp. 769–774.

Wang, Y., Cui, T., Zhang, C., Yang, M., 2010, "A Global Protein-Protein Interaction Network in the Human Pathogen *Mycobacterium tuberculosis* H37Rv", *Proteome Res*earch, Vol. 9, pp. 6665–6677.

Wendt, M. K. and W. P. Schiemann, 2009, “Therapeutic Targeting of the Focal Adhesion Complex Prevents Oncogenic TGF-Beta Signaling and Metastasis.,” *Breast Cancer Research*, Vol. 11, No. 5, p. R68.

William J., T. Berger., and D. Elston, 2005, *Andrews' Diseases of the Skin: Clinical Dermatology*, Saunders, 10th edition.

Wilson J. W., M. J. Schurr, C. L. LeBlanc, R. Ramamurthy, and K. L. Buchanan, 2002, “Mechanisms of Bacterial Pathogenicity”, *Postgrad Medical Journal*, Vol. 78, pp. 216–224.

Wingender E., 2008, “The Transfac Project as an Example of Framework Technology that Supports the Analysis of Genomic Regulation”, *Brief Bioinformatics*, Vol. 9, pp. 326–332.

Woese, C.R., O., Kandler, and M. L., Wheelis, 1990, “Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya (Euryarchaeota/Crenarchaeota/kingdom/evolution*)”, Proceedings of the National Academy of Sciences of the United States of America*, Vol. 87, pp. 4576-4579.

World Health Organization, 2009, "World Health Statistics", *WHO Press*, Geneva.

World Health Organization, 2012, "World Health Statistics-2012", *WHO Press*, Geneva.

Wyllie D, D. Crook, and T. Peto, 2006, “Mortality after *Staphylococcus aureus* bacteraemia in Two Hospitals in Oxfordshire, 1997–2003: Cohort Study”, *British Medical Journal*, Vol. 333, No. 7562, p. 281.

Yang, H., Y. Ke, J. Wang, Y. Tan, S. K. Myeni, D. Li, Q. Shi, Y. Yan, H. Chen, Z. Guo, Y. Yuan, X. Yang, R. Yang, and Z. Du, 2011, "Insight into Bacterial Virulence Mechanisms against Host Immune Response via the *Yersinia Pestis*-Human Protein-Protein Interaction Network", *Infection and Immunity*, Vol. 79, pp. 4413-4424.

Yan-lei W., A. B. Malik, Y. Sun, S. Hu, A. B. Reynolds, R. D. Minshall, and H. Guochang, 2001, "Innate Immune Function of the Adherens Junction Protein p120-Catenin in Endothelial Response to Endotoxin", *Journal of Immunology*, Vol. 186, pp. 180-187.

Yarden, Y., 2001, "The EGFR Family and Its Ligands in Human Cancer: Signalling Mechanisms and Therapeutic Opportunities", *European Journal of Cancer*, Vol. 37, pp. 3-8.

Yu-Chao W., C. Lin, M. Chuang, W. Hsieh, C. Lan, Y. Chuang, and B. Chen, 2013, "Interspecies Protein-Protein Interaction Network Construction for Characterization of Host-Pathogen Interactions: a Candida Albicans-Zebrafish Interaction Study", *BMC Systems Biology*, Vol. 7, p. 79.

Zhang, G. and S. Ghosh, 2000, "Molecular Mechanisms of NF-κB Activation Induced by Bacterial Lipopolysaccharide through Toll-Like Receptors", *Journal of Endotoxin Research*, Vol. 6, pp. 453-457.

Zhang, Y., A. T. Ting, K. B. Marcu, B. James, and J. B. Bliska, 2005, "Inhibition of MAPK and NF-KB Pathways is Necessary for Rapid Apoptosis in Macrophages Infected with Yersinia", *Journal of Immunology*, Vol. 174, pp. 7939-7949.