Scholar@UPRM

Detection and characterization of ######## ######### lethal factor interacting proteins using human stomach T7 Phage Display cDNA libraries

Item Type	Thesis
Authors	Cardona-Correa, Albin A.
Download date	2025-06-19 06:17:54
Link to Item	

Detection and Characterization of *Bacillus anthracis* Lethal Factor Interacting Proteins using Human Stomach T7 Phage Display cDNA Libraries

By

Albin A. Cardona-Correa

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

MASTER OF SCIENCE
IN
BIOLOGY
UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2015

Approved by:

_____   _____
Juan C. Martínez-Cruzado, Ph. D   Date
Member graduate committee

_____   _____
Carlos Rodríguez-Minguela, Ph. D   Date
Member graduate committee

_____   _____
Mildred Zapata-Serrano, Ph. D   Date
Representative, Office of Graduate Studies

_____   _____
Carlos Ríos-Velázquez, Ph. D   Date
President graduate committee

_____   _____
Nannette Diffoot, Ph. D   Date
Chairperson of the Department

ABSTRACT

The 2001 bioterrorism attacks to the USA using envelopes containing *B. anthracis* spores increased the concern of scientists to fully understand the mode of action of the anthrax disease. *B. anthracis* produces an exotoxin composed of three proteins. The first protein is the protective antigen (PA), which interacts with the human cell receptor TEM8 & CMG2 to allow the entrance of the two other subunits, edema factor (EF) and lethal factor (LF), to the cell by endocytosis. The EF interacts with the ATP molecules, causing an excess of cAMP which results in edema. On the other hand, the LF was found to cleavage most mitogen activated protein kinase (MAPK) kinases (MKKs) and then the NOD-like receptor proteins (NLRPs), resulting in cells apoptosis. If additional substrates suffer a LF-mediated cleavage, resulting in cell apoptosis, then we suggest novel LF-interacting partners to promote this biological process in gastrointestinal (GI) anthrax. T7 Phage Display (T7PD) is an i*n vitro* technique of combinatorial chemistry that allows the quick selection, amplification, and identification of potential ligands of specific biomolecules. The technique consists on the expression of cloned DNA sequences as fusion proteins on the T7 phage surface, allowing the study of protein-protein interactions. In this work, the technique was used to identify a profile of proteins that interact with LF, as additional substrates, in order to understand GI anthrax *in vitro*. Four T7PD screenings, consisting of three biopanning rounds each one, were performed using a human stomach cDNA library in order to identify proteins that binds LF. A lethal factor wild-type (LF-WT) with metalloprotease activity and a mutant type (LF-MT) without protease activity were used as target to identify different sites for interaction. The T7PD LF-interacting partners were tittered; eventually, 192 were isolated for extract their DNA for a PCR reaction using specific primers to the cloning site. The DNA fragment sizes varied from 300-900bp. After sequencing 124 clones a total of 33 proteins were identified *in silico*, being mostly

from protein families such as peptidase A1, lipase, and kruppel C2H2-type zinc-finger protein motifs. From these proteins, 18 were selected for a specificity test consisting of allowing the interaction of individual candidates with each type of LF and the blocking agent (5% casein). A total of 10 T7PD candidates were found to bind at least one type of LF tested. The candidate peptides belong to proteins from families such as peptidase A1, cytochrome, and Cation transport ATPase. The PAP46, a candidate from the peptidase A1 family, was found to interact with both types of LF tested with a target minimal concentration for interaction (MCI) of 1µg/mL. These findings could represent the identification of additional substrates for LF, besides the MKKs and NLRPs. In short, these putative novel peptides that bind LF *in vitro* can be used to better understand anthrax disease at the molecular level, through the characterization of these substrates, and to develop new therapeutic agents.

RESUMEN

Los ataques de bioterrorismo en el 2001 a los EE.UU. utilizando sobres con esporas de *B. anthracis* aumentaron la preocupación de los científicos en comprender plenamente el modo de acción de la enfermedad del ántrax. *B. anthracis* produce una exotoxina compuesta de tres proteínas. La primera proteína es el antígeno protector (PA), que interactúa con los receptores de la célula humana TEM8 y CMG2 para permitir la entrada de las otras dos subunidades, factor edémico (EF) y factor letal (LF), a la célula por endocitosis. El EF interactúa con las moléculas de ATP, causando un exceso de cAMP resultando en edema. Por otro lado, el LF fue encontrado de llevar a cabo escisión en la mayoría de la proteínas quinasas activadas por mitógenos (MKKs) y luego en las proteínas receptoras-NOD (NLRPs), resultando en apoptosis de las células. Si sustratos adicionales sufren una escisión mediada por LF, resultando en apoptosis celular, entonces sugerimos nuevos socios interactuantes de LF para promover este proceso biológico en ántrax gastrointestinal (GI). *T7 Phage Display* (T7PD) es una técnica *in vitro,* de la química combinatoria, que permite la rápida selección, amplificación e identificación de ligandos potenciales para una biomolécula específica. La técnica consiste en la expresión de secuencias de ADN clonados como proteínas de fusión en la superficie de los fagos T7, permitiendo el estudio de interacciones proteína-proteína. En este trabajo, la técnica fue utilizada para identificar un perfil de proteínas que interactúan con LF, como sustratos adicionales, con el fin de entender el ántrax GI *in vitro*. Cuatro monitoreos de T7PD, consistiendo cada uno en tres rondas de bioselección, se realizaron utilizando una biblioteca de ADNc de estómago humano con el fin de identificar aquellas proteínas que se unen a LF. Un LF tipo silvestre (LF-WT) con actividad de metaloproteasa y uno tipo mutante (LF-MT) sin actividad de proteasa, fueron utilizados para diferenciar sitios de interacción. Los interactuantes de LF se cuantificaron; eventualmente, 192 fueron aislados para extraer su ADN

para una reacción de PCR utilizando iniciadores específicos para el sitio de clonación. Los fragmentos de ADN variaron de 300-900pb. Después de la secuenciación de 124 clones se identificaron un total de 33 proteínas *in silico*, siendo en su mayoría de las familias de proteínas tales como peptidasa A1, lipasa, y dedos de zinc tipo C2H2 Kruppel. De estas proteínas, 18 fueron seleccionadas para la prueba de especificidad que consistía en permitir la interacción de candidatos individuales con cada tipo de LF y con el agente bloqueador (caseína al 5%). Un total de diez candidatos de T7PD fueron capaces de enlazar al menos un tipo de LF. Estos péptidos pertenecen a proteínas de las familias peptidasa A1, citocromo y ATPasas de transporte de cationes. La PAP46, un candidato de la familia peptidasa A1, fue encontrado como un enlazador de ambos tipos de LF ensayados con una concentración mínima de interacción para el objetivo (LF) de 1µg/mL. Estos hallazgos podrían representar la identificación de nuevos sustratos, además de las MKKs y los NLRPs para LF. En resumen, estos posibles péptidos noveles que enlazan al LF *in vitro* pueden ser utilizados para entender mejor el ántrax a nivel molecular, a través de la caracterización de estos sustratos, además, para desarrollar nuevos agentes terapéuticos.

DEDICATION

I want to dedicate this work to my beloved parents Jannette Correa and Angel Cardona. Thank you for being present in every step of my life. All the effort employed in my career will never compare the sacrifices that you have done for me. Both of you were the reason to continue when adversities came, when the experiments were hard, and at the moments of failure. Counting with your advice, help, and arms, were the scaffold for me to become the person that I am today.

Jannette (mami), thank you for being a mother 24 hours since the day I was born. I will never forget all your love, care, and sacrifices. I remember clearly the day that you said "you are intelligent and you can reach your goals with effort, no matter the stones on the road". At that point, I realized that my dreams can be true with perseverance and passion. I hope to make you feel proud of yourself. Of course, thank you for cooking so well. I love you!

Angel (papi), I feel proud to be your son. Despite you did not have the opportunity of being a college student, the methodologies used for study were the ones taught by you during my childhood. For this reason, the success of my studies are the reflection of all that time employed to make me understand my surroundings on the start. Thank you for make me feel happy by highlighting my virtues when the things were going wrong. That is what a son expect from a father. As a result, I had more courage to face every problem. Thank you for all your support!

I want also to dedicate this work to my "ginger" siblings Alberto and María Cardona. I thank God all the days of my life for having such good persons as family. All the moments lived with you are the marrow of happiness in my life. You are not only family, both are also my friends. Alberto, thank you for protect and guide me to the right ways. Currently, life circumstances made us to be far from me due to work issues. However, the constant communication makes me feel that

you are in every step I take. María, what an exceptional young lady! Because you are the smallest of the family, that fact allowed me to experience mentoring. I admire the adventurous form of seeing life that you have, which brought me a lot of times a laugh despite the adversities. Thank you, for being so positive, funny, and lovely…

God, thanks for watching every step that I took. Bless all our beloved ones and to those that bring happiness to the world.

ACKNOWLEDGEMENTS

I should start to thank God, because he accompanied me through this travel and allowed the generation this thesis work. He gave me the forces and the wisdom to persevere in every aspect. In addition, he gave me a beautiful biological family and the B266 family. Thank you for everything!

In addition, I want to say thank you Dr. Carlos Ríos-Velázquez because when I was without course he was there to guide me as an exceptional mentor and friend. The day that you allowed me to be part of the B266 on august 2012, you gave me one of the best gifts of my life. It was not only an opportunity to grow as scientists, but also to establish a family relationship in our laboratory. Thank you for being there for us all the days of the year and because you put confidence on me to be your graduate student. Through the time I understood that being a mentor is more than knowledge, power or leadership, it is the passion and time employed to guide people. You served for me as the model of Scientist that I want to become. Moreover, I took from you the phrase ''Joven, usted debería…'' because when I feel that something is missing on what I am doing, it reminds me that I should improve that.

To my friend of B266 Frank Ferrer, my first graduate student mentor, I want to thank you for believing in me from the first time I visited the lab. All the hundreds of tubes of MIC that you gave me as work, was the fire proof to demonstrate the love that I have for Sciences. The skills that you taught to me on 2012, today are the fruits of this thesis work. I really appreciate the friendship that we established and I wish you the best as a human being and Scientist.

To my soulmate Jesie Rullán, all the words in this world cannot explain how really I am happy because I met you. Because it was friendship from the first sandwich. Thank you for

listening, helping, and mentoring me during this master degree. I feel admiration to your experience, seriousness and curiosity in every aspect of life. Never change the way you are because that is what distinguishes you from others, being unique and special. My best wishes for you!

To Dr. Martínez Cruzado, Dr. Rodríguez-Minguela and Dr. Acosta-Mercado, thank you for being exceptional professors from the Biology Department of UPRM. The things and skills that I received from all of you I am sure that will be the key for successful in my career. All of you helped me to improve as a best Scientist in different aspects and that is why it was a pleasure to meet you. Thank you for increase my passion for what I am doing.

Also, thanks to the all Biology department staff, especially Mrs. Magaly Zapata, for their support these years. To Dr. Jaime Acosta, Dr. Diffoot, Mrs. Vilmarie Rivera and Mrs. Mary Jiménez, thank you for your availability and time.

Ms. Selimar Ledesma and Ms. Lorein Moya, thank you to both of you for all the support. Selimar you were my research mentor for this project and I expect that you feel proud about the product of your efforts. To my ''research daughter'' Lorein, you do not have idea of how proud I am of you. Thank you for being so comprehensive, intelligent, and responsible.

To all the Microbial Biotechnology and Bioprospecting laboratory (B266) because of their strong support during this time. Especially I want to mention, Wilmer, Ricky, Edgar, Luis, Laura, Moisés, and Robert. All of you were key for me to grow as researcher.

To the RISE2BEST-UPRM program for their economic support with grant NIH-R25GM088023. Thank you especially to the staff Mairim Romero, Rosalie Ramos and Dr. López.

TABLE OF CONTENTS

LIST OF FIGURES

xi

LIST OF TABLES

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

The 2001 attacks through envelopes containing *Bacillus anthracis* spores, in order to cause terror, increased the interest of scientists to fully understand anthrax toxin mode of action. According to the CDC, these acts resulted in the death of five US citizens from 22 reported cases of inhalational anthrax. Therefore, the 2001 acts described are the most recent events of bioterrorism, which consist in the use of pathogenic germs to cause terror. Currently, the three types of anthrax disease known are the inhalational, gastrointestinal (GI), and the cutaneous (Liu et al., 2014). In 1997 the generation of antibiotics against *B. anthracis* became an important step, at state defense level, guaranteeing the military personnel safety, for those at high risk of exposure (Asa et al., 2002). However, depending upon the type of exposure (inhalation of spores and direct bacterial contact, among others) *B. anthracis* infections can require prolonged treatment often for six months with a variety of antibiotics (Jang et al., 2013). Moreover, in most cases, the use of antibiotics is not enough to ensure the safety of individuals. The virulence factors, such as the *B. anthracis* toxin's components and capsule, play an important role in the pathogenicity and proliferation of this kind of microorganisms. As a result, the understanding of these factors has challenged molecular biology by promoting the development of new techniques that provide more specific information about chemical interactions between toxins and ligands.

As mentioned above, *B. anthracis* secretion of an exotoxin enhances its virulence by causing host death in several days if not treated. The anthrax toxin is a tripartite protein system, where the two enzymatic subunits edema factor (EF) and lethal factor (LF), require the protective antigen (PA), to enter to the cell cytoplasm by endocytosis (Abrami et al., 2010). Once inside the cell, the EF which is a Ca^{2+} /calmodulin-dependent adenylate cyclase appears to disrupt immune function, while the LF subunit acts as a zinc dependent metalloproteinase (Leppla 1982; Klimpel

et al., 1994). The LF cleaves the N-terminal of the mitogen activated protein kinase (MAPK) kinases (MKKs/MEK), resulting in the inhibition of the MAPK pathways (Duesbery et al., 1998). The disruption of the MKKs signaling pathways stimulates a cascade of reactions, principally mediated by caspase-1 and activation of interleukins in macrophages, which finishes in cell apoptosis. Levinsohn et al. (2012) demonstrated the direct proteolytic cleavage of LF at the N-terminal of the protein families with NOD-like receptors proteins (NLRPs), resulting in inflammasome response in mouse NLRP1B. The inflammasomes are a multimeric protein complex that form as a response to danger signals within the cytoplasm. As a result, these inflammasomes provide a scaffold for the activation of caspase-1, which is involved in apoptosis (Moayeri et al., 2012). Considering this, if additional substrates of LF, besides the MKKs, were involved in cell apoptosis, then we suggest novel LF-interacting partners to promote this biological process in GI anthrax. This study is focused on GI anthrax because the mechanisms of this type of the disease are not well understood due to low number of clinical cases, resulting in mortality rates of 20-60% of cases (Tonry et al., 2013).

In 1985, a revolutionary technique named Phage Display was introduced by George P. Smith, to facilitate the study of protein-protein or protein-ligand interactions. Phage display is an *in vitro* technique that allows the expression, selection, and amplification of proteins on the surface of viral particles, such as M13 and T7, linking the phenotype to its genotype (Smith 1985; Clackson & Wells 1994; and Kay et al., 1996). The isolation of nucleic acids, like mRNAs, from different human tissues was an important step for display technologies. The conversion of mRNAs to complementary DNA (cDNA), mediated by a reverse transcriptase, resulted in the generation of genomic libraries from different human tissues. The cDNA is the reverse transcriptase product of

mRNA and represents the coding sequence of all transcribed genes at the time of mRNA isolation (Sche et al., 1999).

T7 Phage Display can be used to find treatment for infections, generate vaccines to objectionable microorganisms, develop biosensors, and as a detection method of a desired biomolecule, among others. As a result, phage display can be considered as a powerful technique that allows the quick identification of therapeutic antibodies and peptides, such as those proteins associated with diseases such as cancer (Brissette et al., 2006). In order to identify peptides or ligands that have high affinity to the anthrax LF the T7 Phage Display technique can be used as a tool to allow an understanding of the anthrax disease at molecular level on an *in vitro* manner. By performing T7 Phage Display, it is expected to discover novel peptides from human stomach with specific binding properties to the toxin component under study. The bioselection of these peptides may result in the generation of consensus sequences of amino acids which can be used for the development of peptides that can interfere with the cleavage activity of LF, resulting in higher survival rates.

1.2 LITERATURE REVIEW

1.2.1 *Bacillus anthracis* as a potential pathogenic microorganism and clinical aspects

Bacterial diseases frequently involve the secretion of different types of toxins. In a large extent, interactions between toxins and target human proteins results in the denaturation or cleavage of proteins causing cell damage. Anthrax is considered a zoonotic infection, having an occurrence in human of 20,000-100,000 cases in the first half of the 20th century and later the incidence declined with approximately 2,000 cases annually through the second half of that century. Herbivores are natural hosts of the disease (Kamal et al., 2011). The study of anthrax lethal toxin, produced by *B. anthracis*, has become an important issue since the 2001 attacks to USA citizens, driving research to the development of alternative treatments when antibiotics are no longer an option. In October 2001, the first inhalational anthrax case in the USA since 1976 was identified in a media company worker from Florida; in addition, from October 4 to November 20, 2001, 22 cases of anthrax (11 inhalational, 11 cutaneous) were identified; 5 of the inhalational cases were fatal (Jernigan et al., 2002). These concerns in populations due to the usage of *B. anthracis* as a biological weapon led the biomedical scientists to develop novel therapeutics. As a result, refining pharmacokinetic properties in protein-based therapeutics, as prophylactic countermeasures, became an option (Wu et al., 2011). Despite previous studies in anthrax disease, the complete mechanism that promotes human cells death is not well understood.

In 1850, Pierre Raver and Casimir Joseph Davain, *B. anthracis*, the microorganism causing anthrax was discovered; subsequently, in 1876 Robert Koch first described the complete life cycle of *B. anthracis* (Kamal et al., 2011). Anthrax disease has three common forms: cutaneous, inhalational and gastrointestinal, being the inhalational the rarest. Furthermore, due to spore formation, persistence in the environment, easy dissemination, inhalation route of

transmission, and associated high rates of death makes *B. anthracis* one of the most harmful bioterrorism agents (Belay & Monroe, 2014).

The pulmonary anthrax results when the spores are inhaled and deposited on lung's alveoli. After this, alveolar macrophages phagocytose and transport the spores to the hilar and mediastinal lymph nodes where they germinate and produce the bacterial toxins (Inglesby et al., 2002). The inhalational anthrax can be divided in two stages: the first is characterized by fever, malaise, fatigue, and nonproductive cough lasting for several days; the second phase is accompanied of dyspnea, cyanosis, fever, accelerated pulse rate and obstruction of the trachea followed by death (Brachman, 1980). The dissemination of the bacteria through the body, once they evade the immune system may cause hemorrhagic meningitis complicating the inhalational anthrax. Symptoms associated with anthrax related meningitis are meningismus, delirium, and obtundation (Hicks et al., 2012). The mortality rate from untreated inhalational anthrax cases approaches 100% and the costs associated with a real or perceived *B. anthracis* bioterrorist attack have been estimated at over $26 billion per 100,000 persons exposed (Watson & Keir D, 1994; Dixon et al., 1999; Kaufmann et al., 1997).

The cutaneous anthrax is the most common form of the disease rounding the 95% of cases worldwide (CDC, 2001). This last one can be acquired by direct contact with infected cattle, mostly on farm workers with previously injured and exposed skin. However, no cases of human-human contamination with anthrax have been reported. When anthrax is suspected the appropriate biosafety precautions must be performed; in addition, the diagnosis can be done by culture methods, either from blood in disseminated disease or from the ulcer in the cutaneous anthrax and PCR from the blood of infected area must be done for identification (Murthy et al., 2013). Between years 2009-2010, approximately 55 cases of injection anthrax cases were reported in heroin users

from Scotland, England, and Germany, being 47 from Scotland (Grunow et al., 2013). Only one fatal victim from Bavaria was documented. However, this study only presents 69 published reported cases of anthrax caused by direct inoculation of a single strain of *B. anthracis* to the sub-cutaneous, intramuscular or intravenous tissues. In figure 1 the anthrax distribution around the world as for 2014 is shown. The high number of countries facing the disease represents a concern because these only where the disease was well documented. Well developed countries such as USA were found to be facing the disease.
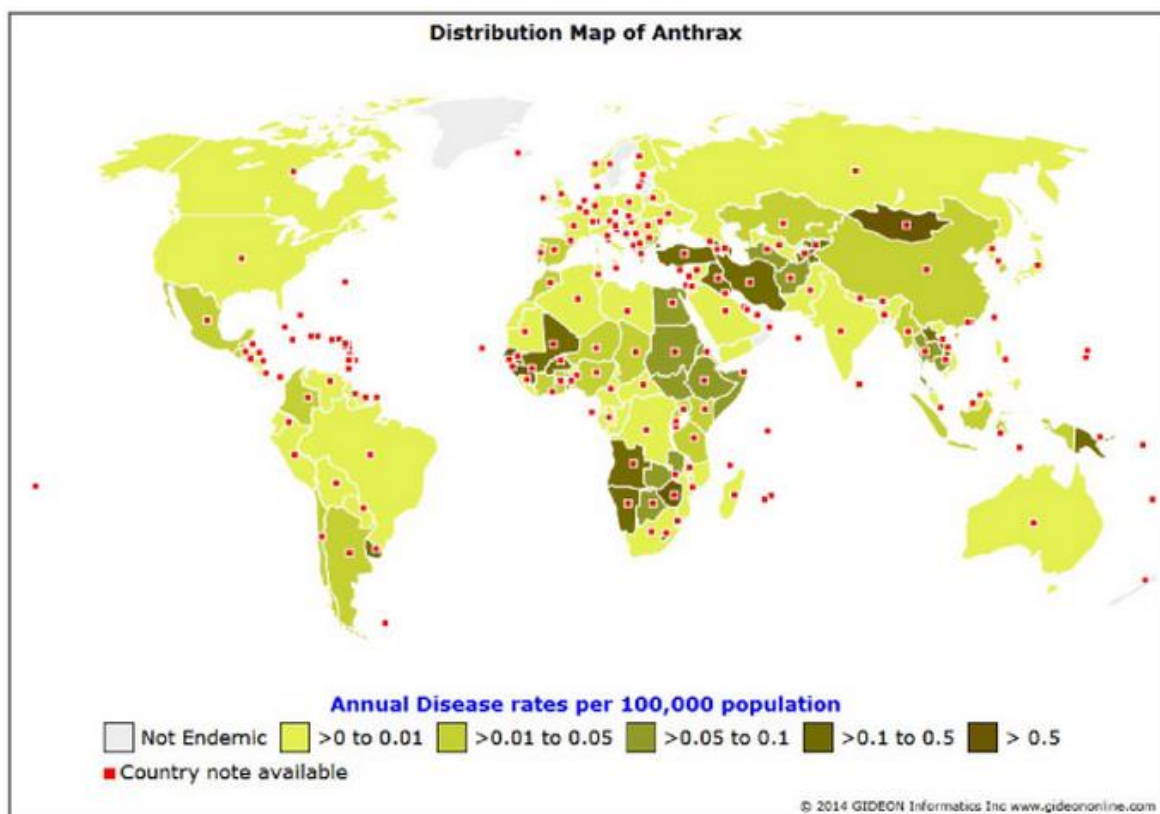
Figure 1. Anthrax global distribution. The disease is endemic on 147 countries. Adapted from: Berger, S. (2014).

Only in Georgia USA, a total of 251 cases of cutaneous anthrax were reported from 2010-2012 in humans working with cattle, raising the interest of studying the spread of the disease through the country (Kracalik et al., 2014). The reported cases for cattle were only 74 and the

human cases were from workers such as the cattle slaughterers, meat processors, and harvesters of crops. For this reason, it is of public concern to study anthrax cases no matter the source where they come, the early detection of the disease is crucial (Griffith et al., 2014).

Statistical data can be interpreted in several ways, in order to help human beings to understand tendencies through determined populations or samples. The US Army Corps of Engineers believes in the usage of the Bayesian statistical approaches to fill those gaps of limited empirical data at the time of determining the safety of contaminated buildings with anthrax. Bayesian methods can support direct statements about: probabilities, e.g., "there is a 95% probability on non-contamination in the room"; probability ranges, e.g., "are 95% confident that the probability of the room being clean is between 2% and 4%"; or probability distributions, e.g., "the probability distribution over the number of spores follows a beta distribution with these parameters" or even "there is a 1% chance that there are ten or more spores remaining" (Linkov et al., 2011). The last type of the disease is the GI anthrax which can be mostly related to a bioterrorist attack because of the mode of occurrence.

1.2.2 Gastrointestinal anthrax

GI anthrax is the rarest type of the three kinds of anthrax disease and in is caused by eating raw food inoculated with *B. anthracis*, which is frequently in dormant stage (endospores). Once the *B. anthracis* spores are ingested, they attach to the epithelial cells of tonsilla and M cells over the payer plaques of the small intestine causing toxin-mediated ulcers. Later the disease can be divided into three clinical phases: Phase I fainting accompanied of fever, phase II abdominal pain with vomiting and the phase III intensified abdominal pain accompanied with bleeding. (Glomski et al., 2007; Akbulut et al., 2012). The symptoms may appear from 2-5 days; the bacteria are

transported from the bowels to mesenteric and finally to the regional lymph nodes, oral and esophageal ulcers can occur; however, the lack of clinical experience on this disease may cause delays in diagnosis (Fowler & Shafazand, 2011).

Reported cases specify that lesions passed down the gastrointestinal tract, in the mid-jejunum, terminal ilium, or cecum, result in a single or several ulcerations and edema (Sirisanthana & Brown, 2002). An additional study demonstrated that the particular proteolytic activity of LF, on the MKKs, is necessary for the anti-proliferative and pro-apoptotic effects of the toxin at the intestinal epithelium, determined through the collection of intestinal samples from mouse infected with PA-LF (Sun et al., 2012). Despite the efforts to identify the LF targets in the human cell, the mechanisms that causes macrophages death and the impact that it performs to the innate immune response are not well understood (Liu et al., 2014). For this reason it is crucial a complete understanding of GI anthrax at molecular level to allow the generation of new therapeutic agents (Lightfoot et al., 2014; Baldari et al., 2006).

1.2.3 *Bacillus anthracis,* virulence factors and toxin interactions.

Bacillus anthracis is an aerobic gram-positive bacterium associated to soils, animals or laboratory culture for bioterrorism. The complete genome of *B. anthracis* H9401, isolated from a GI anthrax patient, was sequenced showing a total of ~5.2Mbp of chromosomal DNA with 5,309 encoding genes, and G+C content of 35%. In addition, two plasmids were sequenced, the pXO1 (~181Kbp, 202 ORFs) which encode the toxin and the pXO2 (~94Kbp, 110 ORFs) encoding the poly-γ-D-glutamic acid capsule (Chun et al., 2012). The toxin produced by *B. anthracis* is a tripartite one, whose proteins are associated to one another in order to cause cell death. These

components are the protective antigen (PA), edema factor (EF), and lethal factor (LF), being this last one as a potential pathogenic metalloproteinase.

These two last proteins penetrate the cell with the help of the protective antigen, a protein which recognizes the cellular receptors and binds to it, allowing entrance of LF and EF to the cytosol by endocytosis (Fig. 2). The entrance of LF and EF to the cell cytosol starts with binding of PA (83 kDa) to the cell anthrax toxin receptor (ATR) known as Tumor Endothelial Marker 8 (TEM8). The association of these two proteins occurs through a specific interaction between the VWA/I domain of the ATR and the receptor-binding domain of PA (Bradley et al., 2001). Moreover, taking in consideration that the VWA/I domain of TEM8 is 60% identical to the VWA/I domain of CMG2, another receptor was confirmed by the generation of retroviral vectors encoding the CMG2 human gene previously amplified by PCR of placental mRNA. The same study confirmed the PA binding to CMG2 by CHOR1.1 cells flow cytometry, anti-PA serum and an allophycocyanin-conjugated 2 antibody (Scobie et al., 2003).

When entering the cell, the edema factor remains bound to lethal factor. The edema toxin is a calmodulin-dependent adenylate cyclase that interacts with ATP molecules, leading to conformational changes of them to cAMP, resulting in a substantial increase of cAMP levels in the cell. As a result, the EF disrupt the normal intracellular levels causing edema (Leppla, 1982, 1984). For this reason, people presenting anthrax disease may have visible swelling areas. A study suggest a principle role for EF in the disruption of the integrity of the blood-brain barrier endothelium prior to the eventual loss of cell monolayer integrity (Ebrahimi et al., 2011). In addition, decreased mean arterial pressure, central venous pressure, and systemic vascular resistance index, combined with increased heart rate symptoms were detected (Sweeney et al., 2010).

TRENDS in Microbiology
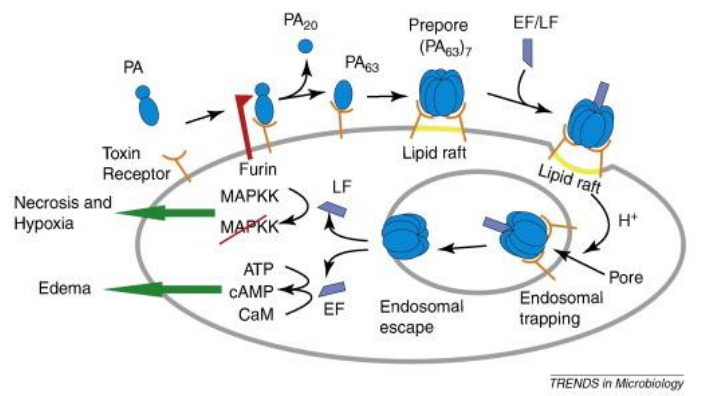
Figure 2. Current model of the cellular effects of anthrax toxins. The protective antigen (PA) interacts with cellular receptors TEM8/CMG2 allowing the entrance of the lethal factor (LF) and edema factor (EF) by endosome formation. Once those toxic proteins enter the cell the (EF) hydrolyses ATP causing an excess of cAMP seen as edema; moreover, the (LF) acts as a metalloprotease, by denaturing cell signaling proteins from the MEK family causing cell apoptosis. Image Adapted from Bouzianas (2009).

1.2.4 The structure and function of anthrax lethal factor

The LF is a zinc-dependent metalloprotease that cleaves most of the MKKs; resulting in a cell apoptotic pathway that leads to tissue hypoxia and cell lysis, usually within 1−2 hours
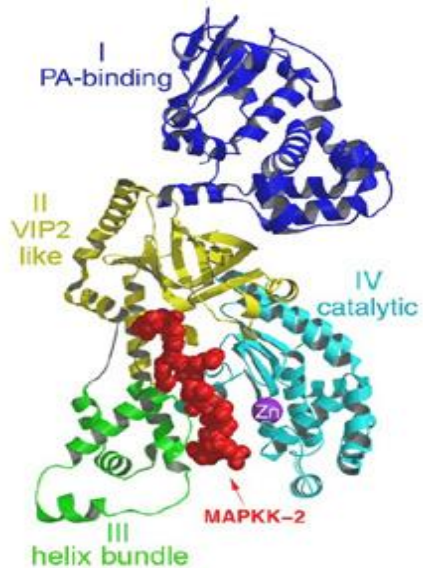
Figure 3. Crystal structure of LF. LF comprises four domains: domain I bind to the protective antigen (PA); domains II, III and IV binds to the N-terminal tail of MKK-2. The fourth domain is Zn-dependent catalytic center (Pannifer et al., 2001).

(Bouzianas, 2009). According to Pannifer et al. (2001) LF consists of four domains (Fig. 3): domain I binds the membrane-translocating component, the PA; domains II (involved in substrate recognition), III and IV together create a long deep groove that holds the 16-residue N-terminal tail of MKK-2 before cleavage. The same scientific group described domain III as a duplication of a structural element of domain II; on the other hand, domain IV was described as distantly related to the zinc metalloprotease family, and contains the catalytic center. A research, involving LF_N-PA spin-spin interactions, revealed that the Phe427 residue

located at the PA prepore, in the lumen and near the base of the structure, converge upon formation of the pore resulting in a particular structure. In this structure, considered the Φ-clamp, LF residues, 2, 5 and 10 cleave to the PA (Jennings et al., 2011). Once the LF is inside the macrophage cell, protein-protein interactions are the major explanation for the subsequent events.

The anthrax LF promotes macrophage apoptosis both by disrupting MKKs dependent pathways and by activating a proteasome and inflammasome dependent death pathway, involving NALP1b cleavage. In addition, the LF and EF inhibit expression of TNF-a and modulate expression of cytokines (IL-10, IL-12) implicated in helper T cell differentiation in dendritic cells stimulated with toll-like receptor (TLR) agonists (Fig. 4) (Tournier et al., 2005). The Interleukin 1 (IL-1) family of ligands and receptors is mostly associated with hematopoiesis, immune response and inflammation; moreover, the cytosolic fragment of each IL-1 receptor family member contains the Toll-IL-1-receptor domain also existing in each TLR, the receptors that react to microbial products and viruses (Dinarello, 2011). Cell pyroptosis, which results in cellular lysis and release of the cytosolic contents to the extracellular space, is performed in macrophages because of the ability of pathogens to survive and replicate in these; furthermore, this cellular process is driven by Caspase-1 during infectious and inflammatory reactions by activating the IL-1β and IL-18 secretion (Miao et al., 2011). Moreover, additional studies suggest that LF cleaves the N-terminus of the NOD-like receptor proteins (NLRPs) to induce the inflammasome response, resulting in novel LF substrates (Levinsohn et al., 2012).
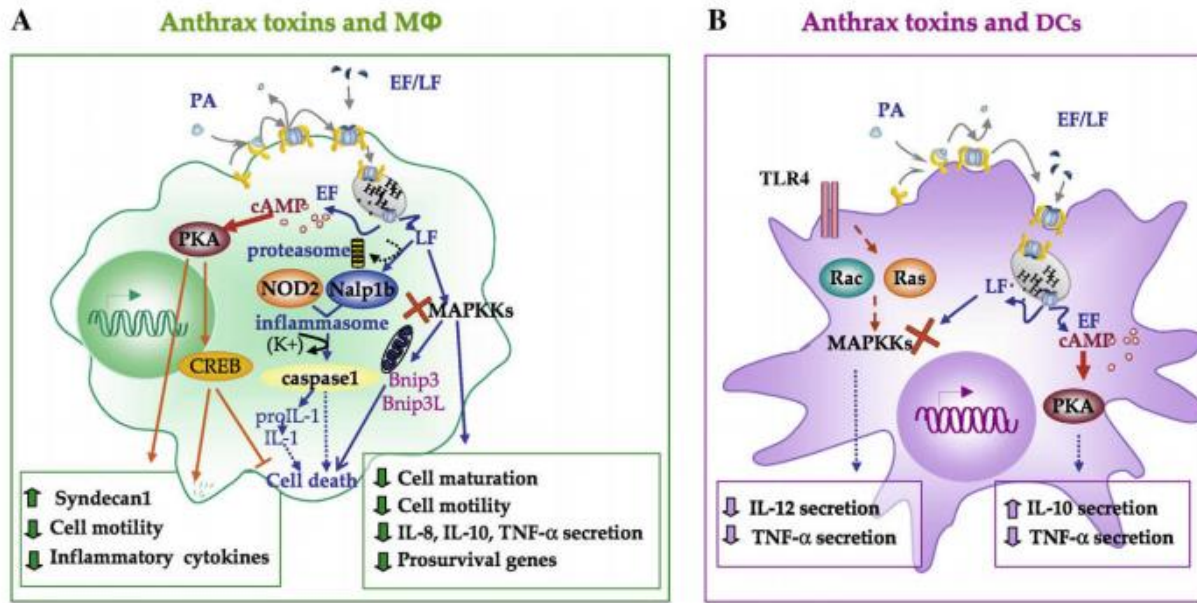
Figure 4. Effects of the anthrax toxins on macrophages and dendritic cells. (A) LT inhibits macrophage activation, expression of pro-inflammatory cytokines and cell motility by disrupting the pathways regulated by MKKs. ET suppresses expression of pro-inflammatory cytokines and cell chemotaxis while antagonizing apoptosis through pathways involving PKA dependent activation of CREB and inhibition of MAPKs. (B) Anthrax toxins inhibit expression of TNF-α and modulate expression of cytokines (IL-10, IL-12). Adapted from Tournier et al. (2009).

1.2.5 Anthrax (LF) interactions with mitogen activated protein kinase kinases (MKKs).

The mitogen-activated protein kinase (MAPK) kinases (MKKs) were known as the unique proteins that suffer direct cleavage by the LF, resulting in cell death and its effect is seen as necrotized tissue. The MKKs are characteristically dual specificity kinases that catalyze the phosphorylation of MAPKs at the tyrosine and threonine residues; in addition, by 1999 twelve member proteins of the MAPK family had been identified in mammalian cells and grouped into 5 subfamilies, on the basis of sequence homology and function (Garrington & Johnson, 1999). The MKKs are involved in several important cellular pathways such as cell proliferation, embryogenesis (Pearson et al., 2001), and angiogenesis (Yang et al., 2000).

The disruption of these proteins promotes macrophage apoptosis due to the inhibition of p38 MAPK pathways, suggesting the need of p38 to allow the transcription of genes involved in

apoptosis inhibition (Park et al., 2002). The docking sites by LF for MKK1 and MKK2 were described by Duesbery et al. (1998), located at the amino terminus for both which contain similar sequences, two proline residues and three positively charged amino acids. The same group suggested the mentioned features of these signaling proteins as necessary for the endopeptidase activity of LF on the mentioned MKKs. Further studies established that the cleavage site for LF on the MKKs requires the consensus sequence of amino acids (++++XφX↓φ) being the + for basic residues, φ for hydrophobic ones, X for any residue and the cleavage site is shown by an arrow (fig. 5) (Vitale et al., 2000). Taking advantage of this knowledge additional studies were directed in order to detect interactions between LF and other MKKs.

The anthrax LF cleaves MEK1, MEK2, MKK3 and MKK6 once, while MKK4 and MKK7 two times. Furthermore, all of the anthrax LF cleavage sites are within the N-terminal region of the MKKs, outside the protein kinase domains (Bardwell et al., 2004).

Nevertheless, studies were guided to identify the lethal factor cleavage in distal residues from the N-termini and resulted in the requirement of a region that confined either in whole or in part within residues 292–318 at the

Figure 5. Docking sites in MKKs by (LF) toxin from *Bacillus anthracis*. (A) D-site peptides previously proposed, residues comprising the basic submotif (+++) are shown in bold and blue; residues comprising the hydrophobic-X-hydrophobic submotif (φXφ) are shown in bold and red. (B) Control peptides. Adapted from Bardwell et al. (2009).

carboxyl-terminal of MKK-1; in addition, the LF cleavage sites on MKKs 1–4, 6, and 7 expose homology features. In all cases, the cleavage site is preceded by a series of basic residues and followed by an aliphatic residue (Chopra et al., 2003). These events promote an impairment on the phosphorylation of MAPK p38 and ERK1/2 on dendritic cells avoiding an effective immune response to the infection through the cleavage performed by LF (Agrawal et al., 2003). As mentioned above this cytotoxic and apoptotic process is proteasome-dependent; however, cell lysis was not exclusively related to the presence of caspases which brought the idea of additional substrates for LF (Tang et al., 1999). For this reason, further studies were driven in order to detect novel anthrax LF-binding proteins to counteract the disease.

1.2.6 NOD-like receptor proteins as additional identified substrates of LF protease.

The NOD like receptors contain nucleotide oligomerization domains (NODs) and a leucine-rich repeat (LRR); on the other hand, NOD1 protein was associated to apoptosis events through the activation of Caspase-9 and Nuclear Factor-κB (NF-κB.), in order to maintain homeostasis. Further studies demonstrated that Nod1 acts as a possible intracellular receptor of bacterial Lipopolysaccharides (LPS) as an innate immune response to pathogens in epithelial cells (Inohara et al., 1999, 2001). Later Nod1 and Nod2 were confirmed as cytosolic receptors for pathogen molecules, promoting the activation of MKKs/NF-κB and inflammasome complexes which result in cell death driven by caspase-1 (Franchi et al., 2008). Caspase-1 plays a role in the activation of multiple pro-inflammatory cytokines such as the Interferon-γ-inducing factor (IGIF) once a pathogen LPS is recognized (Ghayur et al., 1997).

Boyden & Dietrich (2006) described the *Nalp1b* gene as the mediator of macrophage's susceptibility to the anthrax lethal factor; however, at that point it was unclear if the product of

this gene contains a cleavage site for LF. Both Nalps and NODs, are subfamilies of the NBS-LRR family which differ in their N-terminal domains, being a PYD for NALP and a CARD for NODs (Chamaillard et al., 2003). As a result, later, it was discovered by Levinsohn et al (2012) that a direct proteolytic cleavage by LF on NLRPs is necessary to induce the inflammasome complex that results in apoptosis. These findings identified additional substrates that interact with LF. For this reason, more studies are needed to understand the complete cellular pathway that induces the cell death and the complete profile of proteins that experiences a direct proteolytic cleavage by LF in order to develop therapeutic agents. To reach that goal combinatorial chemistry tools can be used as a method. These consist of high-throughput techniques that allow the quick synthesis of a large extent of biomolecules in order to study aspects such as protein-peptide interactions to develop new drugs (W Gruber et al., 2010; Macarron et al., 2011).

1.2.7 Combinatorial chemistry as a strategy to study interactions.

The combinatorial chemistry is an application that allows the synthesis of large amounts of biological molecules such as peptides and oligonucleotides in order to allow the quick screening of diverse biological activities (Jacobs et al., 1994). As a result, the development of therapeutics started to be performed by applying more cost-effective techniques (Houghten et al., 2000). This approach is also used by pharmaceuticals to design new drugs with less time and money using such amount of compounds as a whole for screenings; however, it has the limitation of the generation of unexpected products (Pandeya & Thakkar, 2005).

The two hybrid system, considered a screening platform of several compounds, identifies proteins capable of interacting with a known protein resulting in the availability of the cloned genes for these interacting peptides. In essence, the system utilizes two plasmids with gene fusions

that are cotransformed into a yeast strain having inducible reporter genes, such as *lac*Z, that are transcribed when interactions between the proteins of interest occur (Chien et al., 1991). This powerful technique allows scientists to understand protein-protein and also protein-DNA interactions increasing their understanding in molecular genetics of several species (Wu & Lieber, 1996). Another study was guided using yeast-based system in order to understand the clinical importance of integral membrane proteins. As a result, the identification and characterization of the interactions of these proteins will provide information to recognize their function and find out novel targets that can be used as therapeutics (Snider et al., 2010). For this reason, the two-hybrid system became an important tool in drug discovery; in addition, other emergent techniques such as protein-chip and display systems were crucial for this step (Burbaum & Tobal, 2002).

A display system can consist on the expression of fusion proteins on a microorganism surface for screening (Jahns & Rehm., 2012). These systems are based on peptide libraries that are derived from the cloning and expression of random sequences of oligonucleotides, providing a great amount of ligands (Mattheakis et al., 1994). The first fusion proteins, displayed on viral entities, consisted on foreign DNA fragments cloned on the filamentous phage gene III (Smith, 1985). For this reason, phage display is considered as the physical linkage between the phenotypes of peptides to its corresponding genotype and the DNA encoding the displayed fusion protein is packaged inside the phage particles, allowing the generation of phage display libraries up to 10^{10} members (Paschke, 2006). As a result, a great amount of molecules can be used to study protein-protein interactions between the generated libraries and a desired target being this a fast process of selection, amplification, and identification.

1.2.8 T7 vs M13 Phage Display: as an *in vitro* technique for the study of protein-protein or protein-ligand interactions.

      Phage-display is a high-throughput technique for screening molecules with affinity to a specific target and the process consist in washing and elution steps, for advanced enrichment and amplification of the phage population carrying molecules with affinity (Castel et al., 2011). Considering molecular genetics, for eukaryotes, genes are divided in different regions like the 5'/3'UT, introns, and exons, among others. However, introns are removed from the RNA to give a mature messenger RNA that consists only of series of exons. This process, RNA splicing, involves the precise deletion of introns from the primary transcript and joining the ends of the RNA on either side of each exon to form a covalently intact molecule (Krebs, Goldstein, and Kilpatrick 2010). Moreover, proteomics takes advantage of the mRNAs, from different human tissues, by transforming them in cDNA through reverse transcriptase and this represented the generation of genomic libraries.

      Genomics is based on the study and manipulation of the genomes; in addition, Proteomics is considered as the study of the function of all expressed proteins (Tyers & Mann, 2003). The process of translation, conducted by mRNA, ribosomes, and tRNA, finishes the proteins production by joining together the chains of amino acids in peptide bonds between the C- terminal of one amino acid and the N-terminal of other. The approaches of functional proteomics have been associated to the explanation of large protein–protein interaction networks centered on: protein-protein interaction maps, an accessible database, and simplified functional assays that confirm new proteins identified (Colland et al., 2004). The chemical properties of proteins allows them to interact freely with each other; however, sometimes these interactions can cause the inactivation or denaturation of an important protein of human beings.

A phage display library is composed of a heterogeneous mixture of phage clones, each carrying different foreign DNA inserts displayed as fusion peptides on its surface, allowing the study of protein-protein interactions *in vitro* (Smith & Petrenko, 1997). For instance, previous studies used M13PD to identify peptides that bind the anthrax toxin *in vitro*, in order to generate polyvalent inhibitors as therapeutics (Mourez et al., 2001; Basha et al., 2006). M13 is a filamentous bacteriophage 1µm in length (<10 nm in diameter), specific to *Escherichia coli*, with single stranded DNA and five coat proteins. Moreover, polypeptides fused to M13 coat proteins will be displayed on the surface of the particle as a hybrid phage and this is included as an additional component of a phage coat that contains wild-type copies of all five coat proteins (Sidhu, 2001). The phage M13 is the most commonly used system of presentation of the desired molecules through coupling to the minor coat protein pIII (consisting of 3 domains). However, this type of expression of cDNA libraries is challenging due to the presence of regular translational stop codons in the 3'-region of reverse transcribed mRNAs, and the inability of *E. coli* host system to perform post-translational modifications (Georgieva & Konthur, 2011). In addition, this phage requires that all components of the particle be exported through the host inner membrane before the assembly; as a result, only proteins that perform this pathway can be displayed on the M13 phage capsid, being necessary not only a signal peptide, but also a specific length (Danner & Belasco, 2001). In contrast, the T7 phage performs a lytic cycle that results in the release of phage particles, expressing diverse proteins, from the host (Fig. 6).
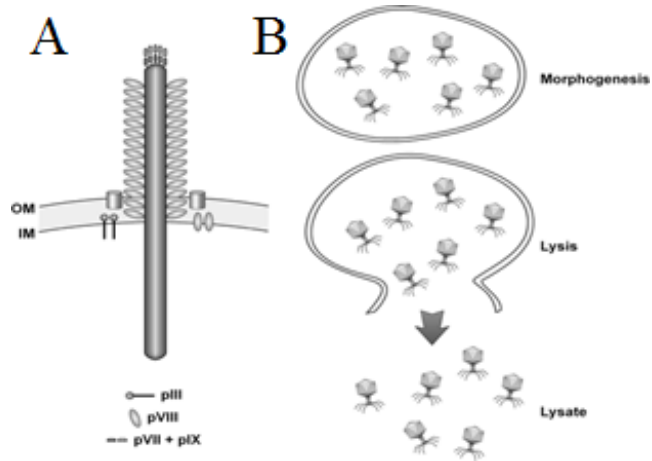
Figure 6. Differences between the M13 filamentous phage and T7 lytic phage morphogenesis. (a) Filamentous phage assembly takes place at the *E. coli* inner-membrane (IM) and is secreted through the outer-membrane (OM) outside the cell, safeguarding host viability. (b) Lytic phage assembles within the *E. coli* cytoplasm, and mature phage particles are released by host cell lysis. Adapted from Krumpe & Mori (2006).

The procedure of M13 Phage Display consists in several rounds of the biopanning (Fig. 7), enrichment phages having affinity to the target is then measured by phage titering and enzyme-linked immunosorbent assay (ELISA); finally, the peptide sequence displayed on the phage can be analyzed by performing DNA sequencing (Fukunaga & Taki, 2012).
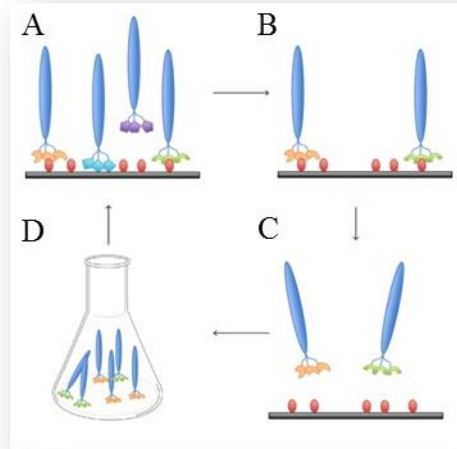
Figure 7. M13 Phage Display biopanning step. (A) Phage particles expressing the cloned proteins are exposed to an immobilized target. (B) Phages having non-interacting peptides are washed. (C) Phages with high affinity proteins are eluted. (D) The phages having peptides with high affinity to the target are amplified by infecting *an Escherichia coli.* Adapted from Fukunaga & Taki (2012).

T7 Phage display is an *in vitro* technique that helps in the identification of peptides or proteins that have specific binding properties to a target; briefly, a peptide or protein is displayed on the capsid of a phage as a fusion to a protein normally found on the phage capsid (Rosenberg et al., 1996). As indicated by the manufacturer (Novagen) target proteins are expressed as fusions to the carboxy-terminus of the 10B capsid protein (near amino acid 348) and are displayed on the virion surface where they are accessible for study interactions with other proteins or ligands. To promote stability of phage lysates the RNase I gene from the host strain (BLT5615) was disrupted; furthermore, the synthesis of the phage capsid 10A protein encoded by pAR5615 must be induced by adding isopropyl β-D-thiogalactoside (1mM) (Danner & Belasco, 2001). The *E. coli* used [Rosetta 5615 Novagen (R5615)] is a less inhibitory host than, where the 10A gene expression of the phage T7 is controlled by a *lac*UV5 promoter present on the host plasmid. Because of the difference in codon usage, Rosetta 5615 was designed to allow the successful expression of phage T7.

As cited on Burgos (2010) the T7 capsid head has an icosahedral shape and the cDNA libraries from human tissues are composed of 415 copies of capsid protein 10; this protein exists in two forms: the shorter 10A and the longer 10B, the latter resulting from a translational frameshift. In contrast with M13, the T7 phage particles are released from host cells by lysis so membrane transport cannot affect phage particles assembly; moreover, displayed peptides have less interference in phage replication, which also decreases the possibility for the existence of radically faster disseminating phage clones (Vodnik et al., 2011). In addition, the T7 virus (~40Kbp) is not difficult to cultivate and replicates more quickly than either bacteriophage λ or filamentous phage. Its plaques form within 3 hours at 37 °C and cultures lyse 1–2 hours after

infection; in addition, the capsid shell of the phage is composed exclusively of the capsid/peptide fusion protein (Rosenberg et al., 1996).

The procedure for biopanning consist of several stringent washes to discard those phages particles with no interaction to the target, followed by an elution step where those phages carrying peptides with high affinity to the target are collected (Fig. 8). Once the phages that had affinity to the target are eluted, a process of bioamplification is performed by infecting *Escherichia coli* R5615 (Novagen) yielding a vast number of phage particles expressing candidate peptides. The process of biopanning and bioamplification can be repeated several times in order to promote or increase specificity. Despite the versatility of T7PD, no studies were found to be focused on the study interactions between LF and proteins from human stomach. Considering this, T7 Phage Display technique can be performed in order to identify anthrax LF-binding proteins using human stomach libraries.
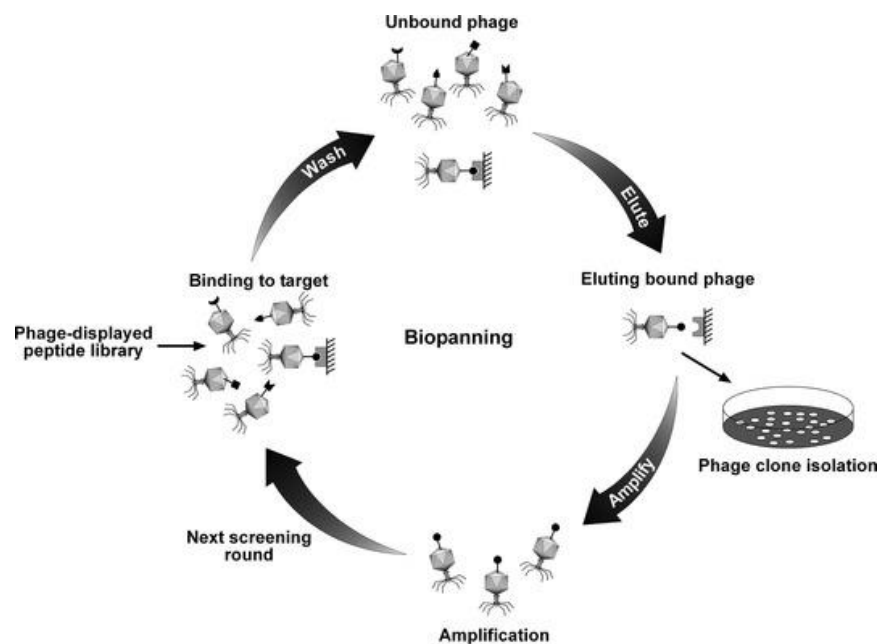
Figure 8. Biopannings rounds for Phage Display. Lethal factor was adsorbed at the bottom of an ELISA plate and then exposed to cDNA libraries. Non-interacting phages were washed and those having specific binding properties with the LF were eluted to infect *Escherichia coli* 5615. Adapted from Krumpe and Mori (2006).

CHAPTER 2

ISOLATION OF LETHAL FACTOR INTERACTING PEPTIDES USING HUMAN STOMACH T7 PHAGE DISPLAY cDNA LIBRARIES

2.1 INTRODUCTION

The study of protein-protein interactions can be done by performing several methods previously employed such as protein-chip and yeast two hybrid, among others. However, understanding the complete set of proteins that serve as target to toxins components may represent a substantial challenge due to the existence of a large number of human proteins. For this reason, it is necessary to use a technique that provides a large number of proteins to generate a selective screening of proteins with desirable properties. In Phage Display millions of biomolecules are expressed as peptides on the phage T7 capsid surface, sometimes M13 and lambda, making possible the analysis of protein-protein interactions between one target and a larger number of putative ligands. High throughput techniques such as T7PD allows the quick selection and identification of protein ligands by performing several rounds of biopannings using the lytic phage T7 (Rosenberg et al., 1996).

In order to find a more comprehensive profile of proteins involved in GI anthrax T7PD cDNA libraries from human stomach were used. The high number of different cloned DNA fragments generated allows a selection not exclusive to proteins but also interacting-domains (Sche et al., 1999). Despite the existence of M13PD, the need of the interacting phages of being exported through the host membrane may result on extracellular proteins' bias. Taking advantage of the lytic properties of the phage T7 the population of interacting-clones can be selected from the whole library. For this reason, the T7-select system provided by Novagen serves as platform to study protein-protein interactions in order to find new anthrax LF ligands from the human stomach. Because of the recent studies, focused on proposing a whole proteome analysis of cutaneous anthrax using antibodies (Popova et al., 2014), here another proteomic tool such as T7PD was performed in order to generate a profile of proteins that interacts with LF *in vitro*.

This chapter describes the procedures performed to detect protein-protein interactions, based mostly on Novagen protocols with few modifications of Burgos, 2010. The procedures for phage titering and the amplification of the cloned cDNA fragments by PCR are also described.

2.2 METHODOLOGY

The research was performed at the Dr. Carlos Ríos-Velázquez's Microbial Biotechnology and Bioprospecting laboratory located at the Biology Department of the University of Puerto Rico, Mayaguez Campus. The utilized system was based on Novagen T7Select^{TM} libraries procedures. The pre-made cDNA libraries (by Novagen) were prepared by using mature mRNA that encodes human proteins and through reverse transcriptase, the messenger was converted into cDNA. This cDNA was inserted into the cloning sites of the gene 10. The T7 capsid gene (gene 10) is located at about position 60 in the T7 genome, within the region of genes coding for proteins involved in the structure and assembly of T7 (Novagen, 1996). For this reason, the cloned cDNA can be expressed as a fusion protein on the phage T7 surface, combined with the natural capsid protein. Pre-made human stomach T7 Phage Display cDNA libraries, acquired at 8.9×10^7 clones $(1.4 \times 10^{10}$pfu/mL), were used to perform this research. As described by the company, the libraries were generated from two persons: a 62 years old male, and a 70 years old woman. Two types lethal factor were used as ''baits'', one is the LF wild type (LF-WT) having metalloprotease activity due to a zinc dependent catalytic center in the fourth domain. The other LF is a site directed mutant in the active site (LF-MT) E687A, lacking cleavage activity (Klimpel et al., 1994). **The toxin-component was purified and provided by Dr. Stephen Leppla's Laboratory of Parasitic Diseases, National Institutes of Health/NIAID Bethesda, MD.** The complete procedure was performed four times for each type of LF at four separated events of biopanning cycles, in order to identify repeatedly binding clones. A biopanning cycle consist of three consecutive rounds of biopanning, which was performed in four independent experiments. **It is important to mention that during this research it was not necessary to cultivate or manipulate *B. anthracis* at UPRM laboratories.**

2.2.1 Adsorption and coating of anthrax Lethal Factor protein

Sterile 96-well microplates (Greiner Bio One, USA) were used for the coating step. First, the wells were washed four times using 400μL of Tris buffered saline (TBS) (Tris-HCl, pH 8 150mM NaCl). Aliquots of the target protein, each lethal factor, were prepared with TBS to 4.5μg/mL and were applied to each well. The two types of anthrax LF were coated and the microplate was covered with parafilm to avoid evaporation and was incubated overnight at 4ºC. After the incubation period, the microplates plates were washed 3 times with 300μL of TBS [1X] to remove unbounded protein. A solution of 5% of casein (skimmed powdered milk) in water was prepared as a blocking agent, and 200μL were transferred into each well followed by overnight incubation at 4º C. Finally, the microplate wells were washed five times with TBS, and 200μL of TBS were poured into each well. The washed plates were covered with parafilm and stored at 4ºC until first round of biopanning.

2.2.2 Biopanning and bioamplification of peptides having affinity to LF

Aliquots of approximately 1.4×10^7 clones (1.4×10^5 pfu/mL) from the stomach T7 cDNA library were prepared on 100μL of Tris buffered saline tween 20 (10 mM Tris-HCL, pH 8.0, 150 mM NaCl, 0.2% TWEEN 20) (TBST) and pipetted in to the microplate wells to start protein-protein interactions. After one hour, the wells were washed five times with 200μL of (TBST) and 200μL of sodium dodecyl sulfate (SDS 1%) were added as an elution solution for 30 minutes. From the eluted phage collected from the wells of each type of LF, wild or mutant type, 250μL were discharged on a particular culture (fig. 9).
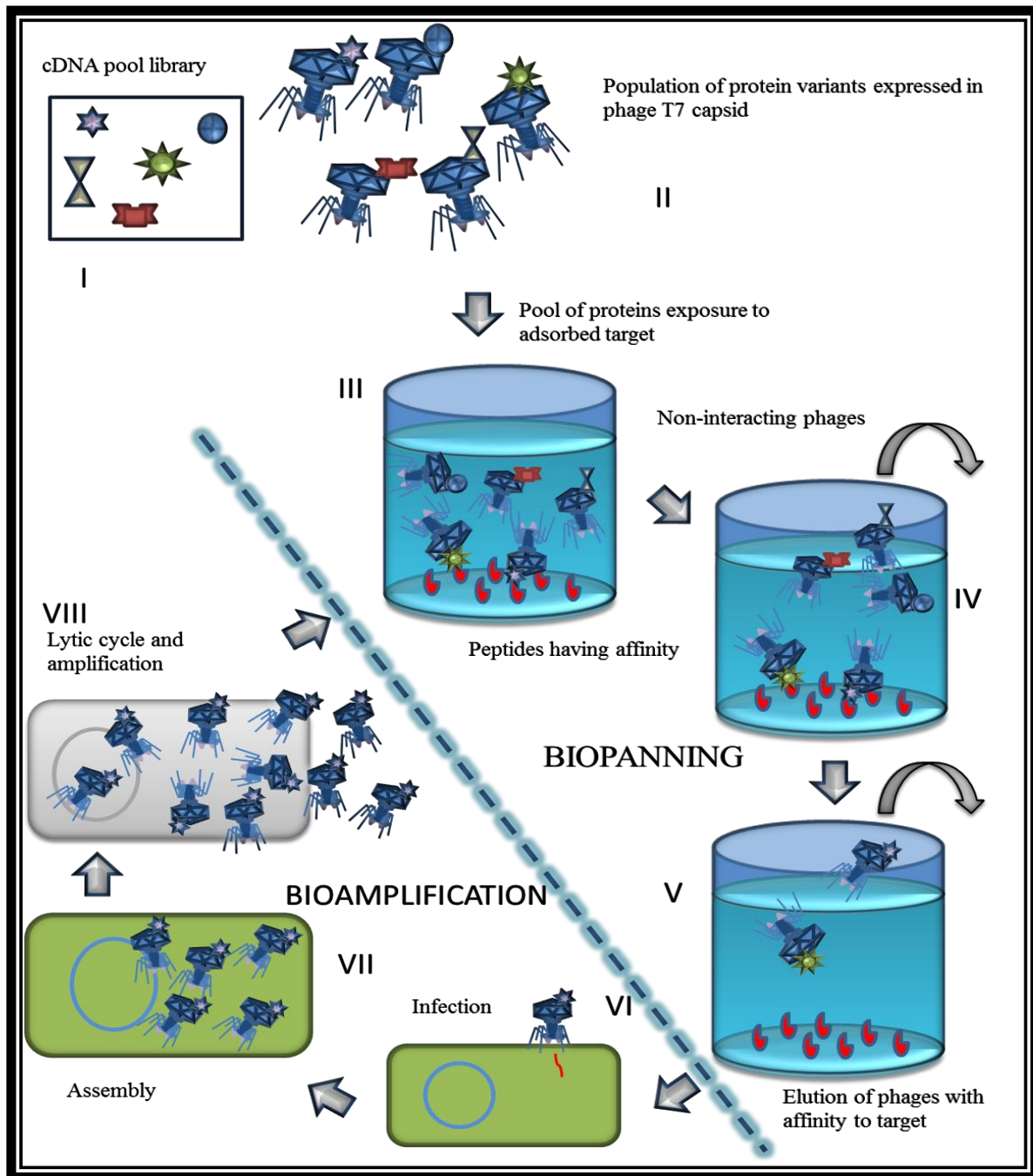
Figure 9. T7 Phage Display biopanning step. I. Pre-made human cDNA libraries, cloned at the T7 phage capsid gene. II. The pool of proteins, expressed at the T7 phage capsid, are exposed to the immobilized target to start protein-protein interactions. III & IV. Several stringent washes are performed to discard those phages having no affinity to the target, only those high affinity peptides will persist through washes. V. By using detergents protein-protein interactions are disrupted. VI. Infection of the host cell R5615. VII. Assembly of new phages takes place on host cell cytosol. VIII. The host cell lysis promoted by the phage T7 amplification. Several rounds of biopanning are performed to enrich for ligands with high affinity to the target.

An *Escherichia coli* Rosetta 5615 (R5615) was cultivated overnight, in 50mL of Luria Bertani broth with 50µg/mL ampicillin (Sigma, USA) at 37ºC shaking in an orbital shaker at 150rpm, before starting biopannings. The cultures were grown in Erlenmeyers at 37ºC shaking at 150rpm until they reached an optical density of 0.5. Prior infection, a host cell induction to allow expression of the phage capsid protein, controlled by the *lac*UV5 promoter at the pAR5615 plasmid, was performed. The induction was achieved by adding isopropyl-beta-D-thiogalactopyranoside (IPTG) 1mM to the incubating cultures for a period of 30 minutes. The infected R5615 was incubated in an orbital shaker at 150rpm for three hours at 37°C. After this, 1,500µL of each lysed culture was transferred to a clean tube. To clarify the lysates 200µL of NaCL 0.5M were added. The mixture was centrifuged at 8,000xg for 10 minutes, and then transferred 1mL from the supernatant to microtubes of 2mL.

The isolated phages were stored at 4ºC until next round of biopanning. Aliquots of ~1×10^7 clones from the first round of biopanning were prepared in 100µL TBST [1X] and were used to perform a second round of biopanning. Finally, a third round of biopanning was performed by using ~1×10^9 clones from the second bioamplification in order to promote specificity of the interacting partners.

2.2.3 Phage titering

To determine the number of T7PD interacting phages a plaque assay (overlay technique) was performed. The elution of every first biopanning (in the four cycles performed) were quantified in order to determine the number of interacting phages prior the bioamplification step. Serial dilutions, $(10^{-2}-10^{-10})$ on Luria Bertani Broth (Difco, USA), of the amplified interacting phages of each biopanning was done to determine the number of PFU/mL obtained. The cultivation of the R5615 was performed by following the biopanning steps described before. After 30 minutes of the host cell induction with IPTG 1mM, selected dilutions $(10^{6}-10^{7})$ were mixed with a fresh culture of R5615 $(OD_{600} 0.5)$ on top agar (Luria Bertani broth 0.7% agar). The mix was poured into a Petri dish containing bottom agar (LB agar with 50mg/mL ampicillin). The plates were incubated at 37°C for 4-6 hours or until plaques were observed.

2.2.4 Isolation of phages expressing peptides with putative affinity to the anthrax LF.

The R5615 was grown following the protocols described in the biopanning step (Sec. 2.2.2). After 30 minutes of the induction with IPTG 1mM addition, 600uL of the cultures were transferred to clean microtubes. A sterile loop or pipette tip was used for take a sample of the plaque at the top agar and the transferred cultures were infected using individual plaques to each microtube. A non-infected cell culture was used as control, to allow comparisons between lysates. The microtubes containing the infected R5615 and control were incubated at 37°C in an orbital shaker at 150rpm for 3-6 hours. After this period of time lysis must be observed on those tubes containing an infected R5615. To clarify the lysates 200µL of NaCL 0.5M were added. The tubes were centrifuged at 8,000xg for 10 minutes and the supernatants were transferred to clean microtubes. Phage titering to the isolated T7PD candidates was performed following sec. 2.2.3.

2.2.5 DNA extraction and amplification (PCR)

Using a sterile tip or wood stick, a sample of the desired tube containing lysates were transferred to a clean tube containing 100µL of 10mM ethylenediamine tetra acetic acid (EDTA), pH 8.0. The microtubes were mixed briefly using a Vortex and then exposed to heat at 65°C during 15 minutes. The samples were left to reach room temperature (~10 minutes) and subsequently clarified by centrifugation at ~8,000xg for 3 minutes.

All the PCR reactions were performed at 25µL. The amplification of the cloned cDNA was performed by polymerase chain reaction (PCR) using GoTaq Green Master Mix (Promega WI, USA), following manufacturer recommendations. Primers specific to the cloning site at (10pmol) were added to the mix: T7 Select up 5'-GGAGCTGTCGTATTCCAGTC-3' (Novagen) and T7 Select down 5'-GCTGATACCACCCTTCAAG-3' (Novagen). 1-2µL of phage DNA were added as template.

The thermal cycler (T100 Bio-Rad) conditions consisted of a first denaturation step at 94°C for 3 minutes, followed by 30 cycles of 94°C for 30 sec, 51°C annealing for 30 sec, 72°C of extension for 30 sec, and a final extension at 72°C for 6 min. The stop step consisted on 4°C incubation. The PCR control reactions consisted of a known and already amplified PD clone DNA sample (positive control) and a reaction tube without DNA sample (negative control). Amplicons were observed performing an electrophoresis in a 1% agarose gel (Sigma, USA) using 100bp ladder 0.1µg/uL (ThermoScientifc). The amplicons were purified and sent for sequencing to Macrogen USA (http://orders.macrogenusa.com/) and MCLAB USA (https://www.mclab.com/home.php). .

2.3 RESULTS

2.3.1. Biopannings phage titering of LF binding partners.

After four different events of biopanning cycles, each one consisting of 3 biopanning rounds, the population of LF binding partners was quantified. The first phage titering performed was for the eluted viral particles interacting with each type of LF in biopanning 1 (BPI) prior and after (BPIA) the amplification step for each event of biopanning cycle (see table 1). The amplification of the eluted phages in biopanning 2 was quantified too for every performed cycle (BPII). In addition, after each third round of biopanning, at four different cycles, the last biopanning's amplification products were quantified (BPIII). The values shown in table 1 are means of the assayed quadruplicates for each type of LF. The numbers of interacting phages were close for both types of LF tested at BPI elution, showing a decrease from the initial number discarded on each well 10^5 to 10^2pfu/mL.

Table 1. Number of phages obtained through three rounds of biopannings in four separate events by using the human stomach cDNA library.

Events LF-WT	BPI	BPIA	BPII	BPIII
1	$1.6x10^2$	$9x10^8$	$1.4x10^9$	$9x10^9$
2	$1.3x10^2$	TNTC	$2.1x10^9$	$5.3x10^9$
3	$3.6x10^2$	$4.5x10^8$	$9x10^9$	$9.9x10^8$
4	$4.6x10^2$	$2.1 x10^9$	$2x10^9$	$9.3x10^9$
Events LF-MT	**BPI**	**BPIA**	**BPII**	**BPIII**
1	$2.4x10^2$	$6 x10^8$	$1.4x10^9$	$9.5x10^9$
2	$3.4x10^2$	TNTC	$2.1x10^9$	$6.7x10^9$
3	$6x10^2$	$1.2 x10^9$	$2.1x10^9$	$8.9x10^8$
4	$4x10^2$	$5.5 x10^9$	$1.4x10^9$	$2x10^9$

*Number of phages are in term of (PFU/mL)
*Each event consist of three rounds of biopanning

The mean for the phage titering assays are shown in figure 10. A constant increase on LF-WT interacting partners was seen; nevertheless, the LF-MT candidates decreased in number after the second round of biopanning.

	BPI	BPIA	BPII	BPIII
LF-WT	2.90E+02	1.20E+09	3.60E+09	6.10E+09
LF-MT	4.00E+02	2.40E+09	1.75E+09	4.70E+09
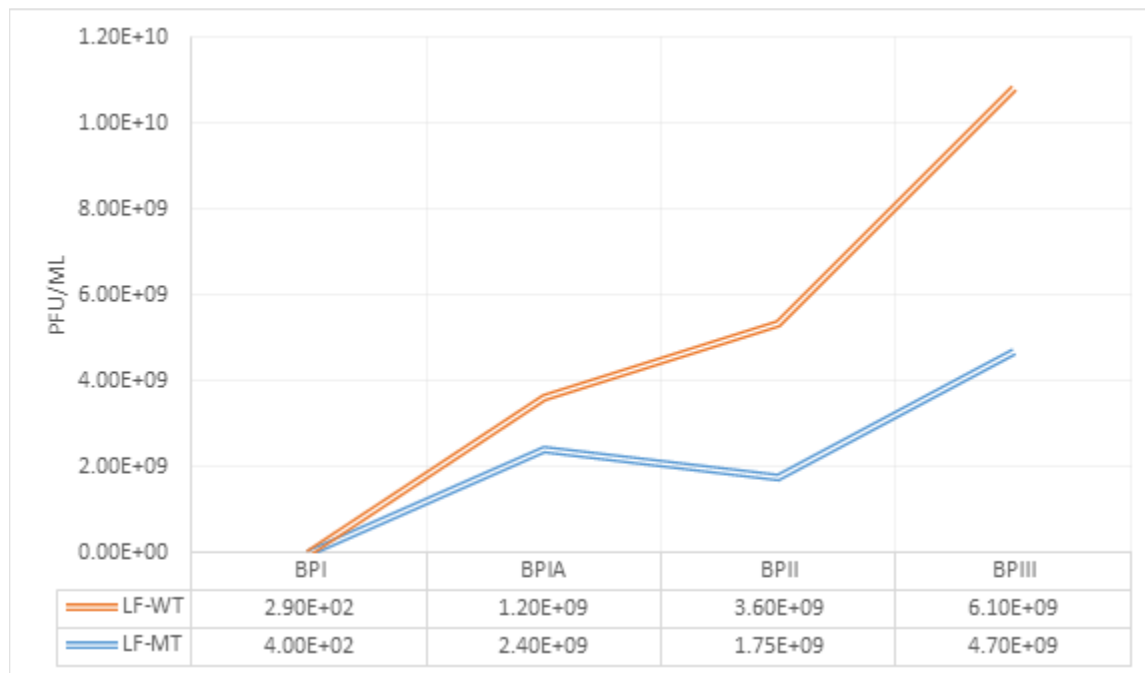
Figure 10. Phage population vs. biopanning round. Mean of phage number for each biopanning cycle that binds to LF-MT (blue) and LF-WT (orange). The BPI represents the eluted phages having interaction with each type of LF tested. The BPIA is the amplification of the BPI. The mean for tittered phages for biopanning 2 & 3 amplification products are shown as BPII and BPIII respectively.

The phage titering, for all the T7PD selected proteins that binds LF, was performed by an overlay essay as the shown in figure 11. Plaques were observed after 3 hours of infection and their sizes varied from 1-4mm. Once the amount of interacting partners were determined for both "baits" the next step was to determine the size of the cDNA fragments within the viral particles.
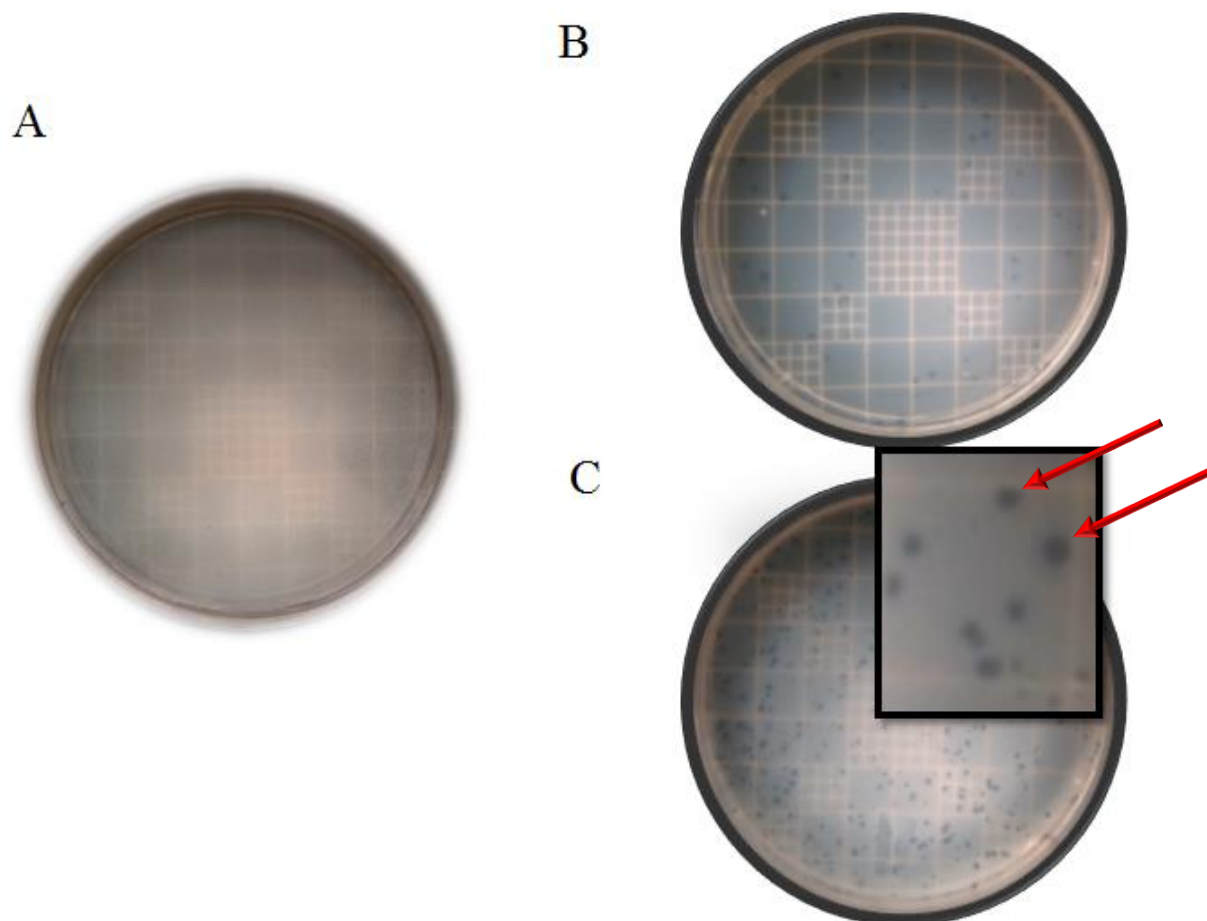
Figure 11. Titering of phages with specific binding properties to the LF through overlay assay. After the three rounds of biopannings those amplified phages having affinity to both types of LF were quantified. (A) A control overlay consisting on the elution of phages on wells without LF was performed to detect plastic interacting proteins. (B, C) In order to determine the pfu/mL of phage particles from the human stomach library having interaction with LF-WT & LF-MT, another overlays assays were performed.

2.3.2. Screening of the cloned cDNA in candidate phages by performing PCR

After four different events of selection, through the three rounds of biopannings a total of 192 interacting phages were randomly isolated and amplified, from the overlay assays. From these isolated candidates, 124 putative LF-interacting clones were successfully amplified by PCR and verified in a 1.8% agarose gel (Sigma, USA). The number of identified clones for LF-WT were 54. Moreover, a total of 70 clones were identified as LF-MT interacting partners (See table 2). Amplicons size varied from 300-1000bp in most cases for both types of LF (figure 12).

Table 2. Number of clones identified per LF type at different events of biopanning cycles.

Events	LF-WT	LF-MT
1	18	22
2	NA	13
3	14	13
4	22	22
Total	**54**	**70**
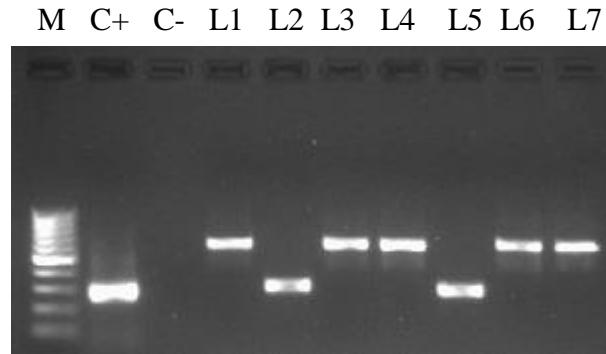
Figure 12. Amplicons verified by an electrophoresis gel (1.8% agarose) showed that fragments of cDNA from 300-1000bp (L1-L7). The DNA was measured in base pair by using 100bp marker (M) (ThermoScientific, USA). A positive control (C+) consisting of previously amplified cDNA exposed to a second round of PCR was loaded. As negative control (C-) a lane was loaded with a sample exposed to PCR consisting only on master mix without DNA template.

2.4 DISCUSSION OF RESULTS

2.4.1 Isolated T7PD LF-binding partners quantification and PCR

Currently, the known anthrax LF targets that undergoes a direct cleavage belongs to the NLRPs and from the MKKs family (Levinsohn et al., 2012; Hellmich et al., 2012). However, the mechanisms involved in the GI anthrax have not been well elucidated (Bishop et al., 2010) leaving the question if all the toxin components targets have been identified. In this study we present several proteins as putative lethal factor's targets obtained through the T7 Phage Display Technique. The data suggest the detection of a more diverse group of proteins or isolated phages with affinity to the LF-MT than LF-WT. This could be due to the fact that LF-MT has the catalytic center mutated. A LF with no metalloproteinase activity allows the selection and identification of those interacting proteins with affinity to the mutated catalytic domain, avoiding the cleavage of candidates. An interaction with the active form of this domain could result in loss of valuable data after the washing steps. The decrease of phages number in biopanning 1 elution's (table 1), from 10^5 to 10^2pfu/mL, implies the loss of proteins from the whole population that does not bind LF *in vitro*. Figure 10 shows how the mean number of phage particles increases after amplifications steps; however, for biopanning 2 the mean number for phages (LF-MT) showed a slight decrease which can be due to the loss of non-specific binding partners. Considering the size of the amplicons of most of the interacting partners (300-900bp) verified by electrophoresis gel (fig. 12), the length of the peptides was expected to be no more than 300 residues.

CHAPTER 3

IN SILICO IDENTIFICATION OF THE PUTATIVE

T7PD ISOLATED INTERACTING PARTNERS

3.1 INTRODUCTION

An *in silico* biological assay refers to the use of computers to carry out studies in biological sciences (Palsson, 2000). The *in silico* analysis represented an important step in molecular biology by facilitating the study of biomolecules such as proteins and nucleic acids, among others. In addition, this type of study has provided the tools to address the development of new drugs to counteract diseases such as anthrax (Rahman et al., 2014). Globalization of bioinformatics allowed the generation of web pages such as the National Center for Biotechnology Information (NCBI). This web page is a place where most of biological and biomedical information of several species is accessible to scientists worldwide. Through NCBI, it is possible to compare the obtained data with the available resources of the database; in addition, it is possible to deposit DNA and protein sequences.

In this chapter the DNA sequences from the isolated T7PD candidates were analyzed using the available nucleotide and protein databases on NCBI. Sequences were identified at the DNA and protein level. The idea was to generate a putative profile of proteins that interacts with LF *in vitro*. Indeed, it is available *in silico* tools, such as *Vacceed*, capable of generating vaccines against eukaryotic pathogens (Goodswen et al., 2014).   For this reason, through these assays, an idea of the molecular pathogenesis of GI anthrax will be closer, if the identity of the putative LF substrates is determined. Two profiles were generated in order to identify the peptides that binds each type of LF that were used as target. Because of the lack of activity of the peptidase domain of LF E687A, it was used to determine different putative epitopes for each candidate. On the other hand, protein families were provided with the idea of giving an overview of the data.

3.2 METHODOLOGY

3.2.1 Identification of the T7PD selected interacting partners

Amplicons sequenced by Macrogen USA and MCLAB USA companies were refined by using the computer program BioEdit Sequence Alignment Editor, available on (http://www.mbio.ncsu.edu/bioedit/bioedit.html), according to the chromatograms. The nucleotide sequence were identified at nucleotide level by performing a BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) using the available NCBI nucleotide databases. Once the DNA sequences were refined, the joining or cloning sites were determined according to the T7 Select System Manual provided by Novagen. These consist on a 5'AATT and 3' AGCTT ''sticky ends''. The mentioned sticky ends were previously generated using *Eco*RI/*Hind*III restriction enzymes (Novagen). The open reading frame were detected starting from the 5'AATT for each sequenced clone.

Then, the DNA sequences were translated in all 6 ORFs to identify the encoded proteins. The translations were done using the web page (http://web.expasy.org/tools/translate/). The protein sequences were identified by using protein BLAST (BLASTp), from the third residue to the end of the peptide. BLAST compares the selected sequences with those available on the databases. The BLASTp parameters were adjusted to reference protein databases, human as organism, an expected threshold of 0.001, and the matrix used was BLOSUM45. From the result of BLASTp, it was also obtained putative conserved domains, if detected. Because the libraries were generated by cloning random cDNA sequences to the T7 vector, frameshifts can occur; for this reason, sequences that were a product from a frameshift in ORF were also analyzed. As a result, peptides fragments can fluctuate in length. Peptides without similarities and higher E-values were also analyzed because of their putative potential as inhibitors.

3.3 RESULTS

3.3.1 Profile of T7 candidate proteins identified through *in silico* analysis

After sequencing 124 clones that showed putative affinity to LF *in vitro*, tables 3 & 4 represent the isolated candidates for LF-MT and LF-WT. The DNA and protein sequences were submitted to *in silico* analysis using tools such as the available databases on NCBI BLAST. A total of 53 different peptides were detected as LF-interacting partners *in vitro* through T7PD technique. These peptides are part or fragments of 33 proteins from a human source. From these, 11 proteins are exclusive to LF-WT and 19 are to LF-MT. Three additional proteins showed putative binding properties for both types of LF tested. The three identified proteins that showed possible interaction with LF *in vitro* are Pepsin A, Gastric triacylglycerol lipase (GL), and Calcium/integrin-binding family member 2 (CIB2). Despite the great number of clones of CIB2, it was only detected on one biopanning cycle. Two forms of 60S ribosomal proteins L (RPL), such as 31 (RPL31) and 35a (RPL35A), were detected as possible binders of LF.

Protein families more frequently detected *in silico* were: peptidase A1, lipase, and kruppel C2H2-type zinc-finger protein (ZNF) (tables 3 & 4, pp. 42, 43, 45 & 47).These consisted of 16%, 8%, and 6%, respectively, from the 53 identified peptides. The rest of the families are <4% of the peptides amount. Putative conserved domains such as xylanase inhibitor terminal, aspartyl protease, Abhydrolase 1, BRICHOS, and subunit 1 of cytochrome *c* oxidase were found. The peptide for LF-MT with the largest length was of 170 residues, corresponding to Gastricsin a member of the peptidase A1 family; on the other hand, the smallest peptide was of 3 residues without protein similarity. The peptide with the highest length, showing interaction with LF-WT, was of 94 residues of Pepsin A5 which is member of the peptidase A1 family too. The smallest LF-WT interacting peptide consisted of 4 residues corresponding to CIB2, a member of the

calcium and integrin binding family. Moreover, a frameshift of the *H. sapiens* pepsinogen 4 from

the LF-MT isolated clones resulted on a single Ig IL-1-related receptor isoform X1 (SIGIRR). A

few candidates showed E-values higher than 1 at protein level and those that also were less than 4

residues are not shown on the data.

Table 3. T7PD human stomach proteins identified as LF-MT interacting partners.

Peptide-MT	Clones #	Number of amino acids	Blast identity Protein level	Blast identity DNA level	Query coverage %	Identity %	E-value	Protein Family	Protein location and accession number
NSANVTAMNVPIAVWNGGKDPLAD PQDVGLLLPKLPNLIYHKEIPFYNHL DFIWAMDAPQEVYNDIVSMISEDKK	1	73	Gastric triacylglycerol lipase isoform 4 precursor	*H. sapiens* lipase, gastric (LIPF), transcript variant 4, mRNA	98	99	3e-47	Lipase family	Secreted NP_001185759.1
NSAKDLLADPQDVGLLLPKLPNLIY HKEIPFYNHLDFIWAMDAPQEVYND IVSMISEDKK	1	58	Gastric triacylglycerol lipase isoform 4 precursor	*H. sapiens* lipase, gastric (LIPF), transcript variant 4	100	98	2e-36	Lipase family	Secreted NP_001185759.1
NSAGGKDLLADPQDVGLLLPKLPNL IYHKEIPFYNHLDFIWAMDAPQEVY NDIVSMISEDKK	2	60	Gastric triacylglycerol lipase isoform 4 precursor	*H. sapiens* lipase, gastric (LIPF), transcript variant 4	98	100	3e-38	Lipase family.	Secreted NP_001185759.1
NSASSYYTGSLNWVPVTVEGHWQIT VDSITMNGEAIACAEGCQAIVDTGTS LLTGPTSPIANIQSDIGASENSDGDM VVSCSAISSLPDIVFTINGVQYPVPPS AYILQSEGSCISGFQGMNLPTESGEL WILGDVFIRQYFTVFDRANNQVGLA PVA	1	156	Pepsin A preprotein	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	99	99	5e-113	Peptidase A1	Secreted NP_001073275.1
NSATSRATSGASENSDGDMAVSCSAI SSLPDIVFTINGVQYPVPPSAYILQSE GSCISGFHGMKL	1	65	Pepsin A3 preprotein	*H. sapiens* pepsinogen 5, group I	100	86	9e-33	Peptidase A1	Secreted NP_001073275.1
NSACISGFQGMNLPTESGELWILGD VFIRQYFTVFDRANNQVGLAPVA	1	46	pepsin A preproprotein	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	100	98	9e-28	Peptidase A1	Secreted NP_001073275.1
NSADLPTESGELWILGDVFIRQYFTV FDRANNQVGLAPVA	1	38	pepsin A 3 preproprotein	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	97	97	2e-20	Peptidase A1	Secreted NP_001073275.1
NSALWILGDVFIRQYFTVFDRANNQ VGLAPVA	2	30	Pepsin A-3 preproprotein	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	96	100	4e-24	Peptidase A1	Secreted XP_005276461.1

<u>NS</u>= Not considered for BLAST

Table 3. Continuation. T7PD human stomach proteins identified as LF-MT interacting partners.

Peptide-MT	Clones #	Number of amino acids	Blast identity Protein level	Blast identity DNA level	Query coverage %	Identity %	E-value	Protein Family	Protein location and accession number
NSAMHVPLCLASFCIFSRDWVSPHWPG WSRTPDFR	3	33	High mobility group protein HMGI-C	*H. sapiens* coiled-coil domain containing 90B	94	74	6e-10	Anion exchanger	Membrane NP_001287848.1
NSAQHDRHPRIRHCPGWRHL	1	18	11-cis retinol dehydrogenase precursor	*H. sapiens* pepsinogen 5, group I (pepsinogen A)	61	73	0.2	Short-chain dehydrogenases/ reductases	endoplasmic reticulum lumen NP_002896.2
NSDSGADTRNEKPVSMVPSRFSEAWRK	1	25	ras-associated and pleckstrin homology domains-containing protein 1 isoform 3	Human DNA sequence from clone RP4-718D20 on chromosome 20	92	47	1.3	MRL	Cell membrane, Cytoplasm, Cytoskeleton NP_976241.1
NSADPAWKYRFGDLSVTYEPMAYMDAA YFGEISIGTPPQNFLVLFDTGSSNLWVPS VYCQSQACTSHSRFNPSESSTYSTNGQTF SLQYGSGSLTGFFGYDTLTVQSIQVPNQE FGLSENEPGTNFVYAQFDGIMGLAYPAL SVDEATTAMQGMVQEGALTSPVFSCVC LISS	1	170	Gastricsin isoform 2 preproprotein	*H. sapiens* progastricsin (pepsinogen C) (PGC)	97	99	1e-124	Peptidase A1	Secreted NP_001159896.1

<u>NS</u>= Not considered for BLAST

Table 3. Continuation. T7PD human stomach proteins identified as LF-MT interacting partners.

Peptide-MT	Clones #	Number of amino acids	Blast identity Protein level	Blast identity DNA level	Query coverage %	Identity %	E-value	Protein Family	Protein location and accession number
NSARAAALGFAIFL	2	12	Protocadherin Fat 2 precursor	*H. sapiens* DENN/MADD domain containing 1B	100	100	1.6	Cadherin	Cell membrane; NP_001438.1
NSAEKRNEDEDSPNKLYTLVTYVPVTTF KNLQTVNVDEN	1	37	60S ribosomal protein L31 isoform 1	*H. sapiens* ribosomal protein L31 (RPL31)	100	95	8e-20	Ribosomal protein L31e	Cytosol NP_000984.1
NSAGTDRK	1	6	Homeobox protein Hox-B7	associated lung adenocarcinoma transcript 1	83	100	120	Antp homeobox	Nucleus NP_004493.3
NSASDNDEGTEGEATEGLEGTEAVEKGS RVDGEDEEGKEDGREEGKEDPEKGALT TQDENGQTKRKLQAGKKSQDKQKKRQ RDRMVERAPALALASSASDSCSCWTFLS LAGARPLGSPDKA	1	120	A kinase (PRKA) anchor protein 8-like, isoform CRA_c	*H. sapiens* A kinase (PRKA) anchor protein 8-like	71	99	2e-51	AKAP95	Nuclear matrix NP_001278407.1
NSARGEGAVSAEQIQKIIETLTQQLQAK GKELNEFREKHNIRLMGEDEKPAAKENS EGAGAKASSAGVLVS	1	69	Prefoldin subunit 2	*H. sapiens* prefoldin subunit 2	86	100	4e-38	Prefoldin subunit beta	Nucleus. Cytoplasm. Mitochondrion NP_036526.2

NS= Not considered for BLAST

Table 3. Continuation. T7PD human stomach proteins identified as LF-MT interacting partners. These peptides did not showed any similarity through BLAST or showed E-values higher than 0.001.

Peptide-MT	Clones #	Number of amino acids	Blast identity Protein level	Blast identity DNA level	Query coverage %	Identity %	E-value	Protein Family	Protein location and accession number
NSAECWRECEWVCAGGHGGAVCKIG VANHRTRAWSGYPPPTQRGRASPHTLT AEFALGRVKKA	1	61	No similarity found	*H. sapiens* chromosome 11 open reading frame 96	-	-	-	-	-
NSAGKTWPVGTVSACREESEDGSLSLG WSLLPSPVGLGAVLA	1	40	No similarity found	*H. sapiens* ORM1-like 3	-	-	-	-	-
NSAVKRRAISRSNNKKVIATKKNFMEK GTRAGDIGSKPHS	1	40	No similarity found	*H. sapiens* isolate 4811480 mitochondrion, complete genome	-	-	-	-	-
NSALNIFCSHRLPRTSRHYWLNFPHYL LHPPTNISLYIQTSLWLRSRRLILAFCR CGLTISVCLHLLMRVL	1	71	No similarity found	*H. sapiens* isolate LHON patient Le1407 mitochondrion	-	-	-	-	-
NSAVMGRLRAPLHCPSPSSLRH	4	20	Zinc finger protein 142	*H. sapiens* zinc finger protein-like 1 (ZFPL1)	80	56	0.18	krueppel C2H2-type zinc-finger protein	Nucleus NP_001099007.1
NSLQGYPGSSSTSLRPHSSN	1	18	zinc finger protein Rlf	*H. sapiens* glycine receptor, alpha 4 (GLRA4)	94	59	2.1	Krueppel C2H2-type zinc-finger protein	Nucleus NP_036553.2
NSLNLVIVQVSSGPGVSSAPA	1	19	GDNF-inducible zinc finger protein 1	*H. sapiens* lipase, gastric (LIPF)	73	71	0.61	krueppel C2H2-type zinc-finger protein	Nucleus XP_005260860.1
NSALEFKPSTASSTGCRSGQW	2	19	Death ligand signal enhancer precursor	*H. sapiens* solute carrier organic anion transporter family, member 2A1	52	60	13	TPR repeats	Mitochondrion XP_006714875.1

NS= Not considered for BLAST

*Table 3. Continuation. T7PD human stomach proteins identified as LF-MT interacting partners.

Peptide-MT	Clones #	Number of amino acids	Blast identity Protein level	Blast identity DNA level	Query coverage %	Identity %	E-value	Protein Family	Protein location and accession number
NSARIQSSPDSVGRFMPWKPLMQLPSLC RM	4	28	M-phase inducer phosphatase 2 isoform 8	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	60	55	0.46	MPI phosphatase family	Cytosol, nucleoplasm NP_001274451.1
NSALLTYWENSLC	1	11	probable phospholipid-transporting ATPase IK isoform 2	*H. sapiens* isolate H3-11 mitochondrion	54	100	1.5	Cation transport ATPase (P-type) (TC 3.A.3)	Membrane; Multi-pass membrane protein. XP_006722719.1
NSAGGLGPVPCPTNNVE	1	15	PRAME family member 10-like	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	46	100	1.1	PRAME	Nucleus, cell membrane NP_001278310.1
NSSRQLLMT	1	7	Unconventional myosin-Ic isoform c	*H. sapiens* staufen double-stranded RNA binding protein 1	100	86	12	TRAFAC class myosin-kinesin ATPase superfamily. Myosin family.	Cytoplasm, nucleus NP_203693.3
NSAHQWLPGHEPPHRIWRALDPG	4	21	PREDICTED: single Ig IL-1-related receptor isoform X1 *H. sapiens*	*H. sapiens* pepsinogen 4, group I	85	64	1.5	Interleukin-1 receptor	Plasma membrane, extracelular, XP_005253101.1
NSAFLLLPGSPEAGKVRL	1	16	Cholinesterase precursor	*H. sapiens* chromosome 17, clone RP11-794C22	87	80	1.6	Type-B carboxylesterase /lipase	Secreted NP_000046.1
NSFVDM	22	4	Calcium and integrin-binding family member 2 isoform 3	*H. sapiens* phosphodiesterase 10A (PDE10A)	100	100	126	Calcium and integrin binding	Photoreceptor inner segment By similarity. Cell membrane › sarcolemma NP_001258818.1
Total	70	22							

NS= Not considered for BLAST., *= These peptides did not show any similarity through BLAST or showed E-values higher than 0.001.

Table 4. Putative LF-WT interacting proteins from human stomach identified through T7PD.

Peptide-WT	Clones #	Number of amino acids	Blast identity Protein level	Blast identity DNA level	Query coverage %	Identity %	E-value	Protein Family	Protein location and accession number
NSGPPYYNVTAMNVPIAVWNGGKDLLAD PQDVGLLLPKLPNLIYHKEIPFYNHLDFIW AMDAPQEVYNDIVSMISEDKK	1	77	Gastric triacylglycerol lipase isoform 4 precursor	*H. sapiens* lipase, gastric (LIPF)	98	100	5e-52	Lipase family	Secreted NP_001185759.1
NSARIGASENSDGDMVVSCSAISSLPDIVFT INGVQYPVPPSAYILQSEGSCISGFQGMNV PTESGELWILGDVFIRQYLTVFDRANNQV GLAPVA	1	94	pepsin A-5 preproprotein	*H. sapiens* pepsinogen 5, group I (pepsinogen A) (PGA5)	100	97	2e-62	Peptidase A1	Secreted NP_055039.1
NSADGDMVVSCSAISSLPDIVFTINGVQYP VPPSAYILQSEGSCISGFQGMNLPTESGEL WILGDVFIRQYFTVFDRANNQVGLAPVA	1	86	Pepsin A3 preprotein	*H. sapiens* pepsinogen 4, group I (pepsinogen A) (PGA4)	100	99	3e-58	Peptidase A1	Secreted XP_005276461.1
NSACISGFQGMNLPTESGELPLQGDVFIR QYFTVFHRENNQVGLAPVA	1	46	Pepsin A3 preprotein	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	100	87	6e-18	Peptidase A1	Secreted XP_005276461.1
NSANINVNDDNNNAGSGQQSVSVNNEHNV ANVDNNNGWDSWNSIWDYGNGFAATRLF QKKTCIVHKMNKES	1	69	Gastrokine-1	*H. sapiens* gastrokine 1 (GKN1),	97	100	6e-45	Gastrokine	Secreted NP_062563.3

<u>NS</u>= Not considered for BLAST

Table 4. Continuation. Putative LF-WT interacting proteins from human stomach identified through T7PD.

Peptide-WT	Clones #	Number of amino acids	Blast identity Protein level	Blast identity DNA level	Query coverage %	Identity %	E-value	Protein Family	Protein location and accession number
NSFGQRCAYVYKAKNNTVTPGGKPNKTRV IWGKVTRAHGNSGMVRAKFRSNLPAKAIG HRIRVMLYPSRI	1	68	60S ribosomal protein L35a	*H. sapiens* ribosomal protein L35a, mRNA	100	97	2e-45	Ribosomal protein L35Ae	Cytosol > ribosomal NP_000987.2
NSACNIIPLQR	1	11	26S proteasome non-ATPase regulatory subunit 9 isoform 2	*H. sapiens* proteasome (prosome, macropain) 26S subunit, non-ATPase, 9	88	100	0.018	Proteasome subunit p27	Nucleus, cytoplasm, cytosol NP_001248329.1
NSAELGQPGNLLGNDHIYNVIVTAHAFVIIF FIVIPIIIGGFGN	1	44	cytochrome oxidase subunit 1, partial (mitochondrion)	*H. sapiens* isolate H3-11 mitochondrion,	100	90	5e-21	Cytochrome *c* oxidase subunit 5B	Mitochondrion inner membrane YP_002124304.1
NSAVTPSEKAYEKPPEKKEGEEEEENTEEP PQGEEEESMETQE	1	43	interleukin enhancer-binding factor 2 isoform 2	Interleukin enhancer binding factor 2	97	100	5e-19	DNF domain	Nucleus, cytoplasm NP_001254738.1
NSVEPSHKSTQRPPPPQGRQRWGGSLGSHN RVCVTHNH	1	36	NEDD4-binding protein 2-like 2 isoform 2	*H. sapiens* NEDD4 binding protein 2-like 2 (N4BP2L2)	100	97	2e-19	N4BP2	Nucleus › nucleolus By similarity. Nucleus › PML body NP_055702.1

NS= Not considered for BLAST

*Table 4. Continuation. Putative LF-WT interacting proteins from human stomach identified through T7PD.

Peptide-WT	Clones #	Number of amino acids	Blast identity Protein level	Blast identity DNA level	Query coverage %	Identity %	E-value	Protein Family	Protein location and accession number
NSADTWAPLFLLTLHPPRALYLSFHSGFSLL FLDSGLSLIINSSSL	1	44	No similarity found	*H. sapiens* progastricsin (pepsinogen C)	-	-	-	-	-
NSARHRSQRELRWRHGGQLLSHQQPARHR LHHQWSPVPRATQCLHPAERGELHQWLP GHEPPHRIWRALDPG	16	70	No similarity found	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	-	-	-	-	-
NSALDCFGSKRPTSWGSANRSLPPFHTAIGT FMAVTL	1	35	No similarity found	*H. sapiens* lipase, gastric (LIPF), transcript variant 4	-	-	-	-	-
NSAWSPVPRATQCLHPAERGELHQWLPGH EPPHRIWRALDPG	1	40	No similarity found	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	-	-	-	-	-
NSAGGWGLLLG	2	9	Telomerase reverse transcriptase isoform 2	*H. sapiens* proline-rich nuclear receptor coactivator 2 (PNRC2)	77	86	15	Reverse transcriptase	nucleolus› telomere. PML body NP_001180305.1
NSAREQGQDLHTPLGRTVRRIRYI	2	22	Copper-transporting ATPase 2 isoform b	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	31	100	2.5	Cation transport ATPase (P-type) (TC 3.A.3).	Golgi apparatus XP_006719902.1
NSAGTGPRPPHSPGKNSQEN	1	18	NAC-alpha domain-containing protein 1	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	55	69	2.9	NAC-alpha	Cytoplasm, nucleus NP_001139806.1
NSSKKKKT	1	8	transcription initiation factor IIE subunit beta	*PREDICTED: Homo sapiens strawberry notch homolog 1 (Drosophila) (SBNO1)*	100	100	35	TFIIE beta subunit family	Nucleus XP_011542812.1
NSFVDM	18	4	Calcium and integrin-binding family member 2 isoform 3	*H. sapiens* phosphodiesterase 10A (PDE10A)	100	100	126	Calcium and integrin binding	Cell membrane › sarcolemma NP_001258818.1
NSAMPWKPLMQLPSLCRM	1	16	toll-like receptor 8 precursor	*H. sapiens* pepsinogen 4, group I (pepsinogen A)	50	88	0.63	Toll-like receptor	Membrane; NP_619542.1
Total	**54**		**14**						

\- = Databases did not provide this information., <u>NS</u>= Not considered for BLAST.,   *= These peptides did not show any similarity through BLAST or showed E-values higher than 0.001.

3.4 DISCUSSION OF RESULTS

2.4.2 Identified proteins that binds LF *in vitro* by performing T7PD technique.

The largest peptide in our findings is part of gastricsin (table 3, p. 43), a protein that belongs to the peptidase A1 family. This peptide contained domains that are part of the aspartyl protease and a xylanase inhibitor and did not have conserved regions with pepsin A. Because pepsin A contained an aspartyl protease domain this suggests the idea that the protease domain could be recognizing LF or vice versa. The *in silico* analysis identified 4 shared putative interacting partners for LF-MT and LF-WT; for instance, a gastric lipase (GL), pepsin A pre-protein (PAP), calcium binding and integrin (CIB2), and 60S ribosomal protein (60SRP) L31/L35a. Considering this, these proteins could be interacting with domains II and III of LF because of their presence in both types of LF tested. As mentioned by Pannifer et al. (2001) these domains are involved in the recognition of substrate by the metalloproteinase. For this reason, any interactions between the substrate recognition domains and human stomach proteins should be detected by PD. However, isolation of putative catalytic region interacting peptides can only be obtained or isolated by PD if multiple interactions with a non-catalytic region of LF are present (eg, domains II and III). The pepsin A and GL (tables 3 & 4, pp. 42 & 47) showed similar amino acid sequences having possible interaction with LF in this study. For this reason, the relevance of these proteins will be discussed in chapter 5 with their corresponding figures. As a result, peptides from these protein families can be tested in order to identify a possible potential as LF inhibitors.

It was suggested in previous studies a relationship between the mitogen activated protein kinase kinases (MKKs) stress signaling pathways and time of their expression, as an explanation of the alteration of these proteins on *in vivo* models, but no *in vitro* (Kandadi et al., 2010). The same study suggested a difference in the regulation of these proteins during stress conditions. For

this reason, time of interactions may have played an important role identifying interactions, explaining the absence of these proteins in our data. Other explanation can be due to the low amount of mRNA coding for this protein in the stomach, compared to other tissues, based on the human protein Atlas (Web tool). However, protein levels are more related to other tissues, which can be due to protein stability. Considering that the T7PD cDNA libraries were done using mRNAs from human stomach, this could be an explanation to the non-detection and identification of the MKKs as partners. In addition, if the N-terminal of the MKKs were not part of the fusion proteins produced, the interaction with LF could be interrupted. In this study, other proteins from different families were identified as putative LF-binding partners, but not as high in frequency as the previously mentioned ones. From all the detected proteins by T7PD only a few of them will be discussed in terms of their relevance at biological processes, especially if a role in apoptosis or host immunity are possible.

The 60S RPL35A (table 4, p. 48) has been identified as an initiation and elongation tRNAs-binding protein at the mRNA translation step (Ulbrich et al., 1980). Further studies suggest the activation of this ribosomal protein due to cytotoxic damage stimuli underlining its role in inhibiting cell death (Lopez et al., 2002).

Currently, potential protein toxins such as ricin are strongly associated to the inhibition of 60S ribosomal activity which results in the detention of protein synthesis followed by apoptosis (Antoine et al., 2012). The peptide for RPL35A did not share similar residues with the previously proposed cleavage site for LF; nevertheless, this peptide has a positive charge that could have interacted with the negative groove of LF (data not shown). Considering this we can suggest RPL35A as a candidate for being a LF's target in human cells. In addition, another 60S ribosomal

protein was detected, RPL31 (table 3, p.44). These two peptides did not share conserved residues, this could be due to possible multiple interaction sites for LF in the 60S subunit.

The peptide, part of an A kinase anchor protein (AKAP) (table 3, p. 44), was ranked third in size in our findings, showing a basic region with residues such as lysine and arginine which were possibly recognized as the cleavage amino acid sequences. The *in silico* analysis of this peptide revealed that it contains a substantial region of helix structure which contains mostly hydrophilic residues, based on PRALINE predictions (data not shown). Proteins from the AKAP's families, including the AKAP95, are considered scaffolds for antigen presentation and activation of T cells in mammalian dendritic cells (Schillace et al., 2009). In addition, this family is characterized for their activity in chromatin condensation when cells are in mitosis (Collas et al., 1999). Furthermore, the inactivation of these AKAP95 may result in low concentrations of nuclear Ribosomal S6 kinase 1 (RSK1) which are necessary for the MAPK/ERK signaling pathways (Gao et al., 2012). This factor added to the well-studied proteolytic effects of LF on the MAPPKs could serve as a platform to elucidate how the cell signaling pathways are disrupted, resulting in apoptosis. NLRPs and Toll like receptors (TLRs) are known by the role in protection of pathogenic infections (Abreu et al., 2005). Cellular location of TLRs varies depending of the type; for instance, TLR8 (table 4, p. 49) was found to be part of the endosome promoting cellular signaling through MYD88 protein which is involved in the expression of pro-inflammatory cytokines (Brown et al., 2011). Nevertheless, studies *in vivo* by injection of the lethal toxin, which is composed of LF and PA, demonstrated that TLRs are not necessary for the inflammasome formation (Moayeri et al., 2012). For this reason, being this an *in vitro* study, consisting of a high number of expressed peptides, the relationship between the TLR8 and anthrax LF may be determined by the conformation of the peptide, its electrical charge or chemical groups.

Another interacting partner was the 26S proteasome regulatory subunit (table 4, p. 48), which degrades >80% of cell proteins, being essential for different metabolic pathways, including cell cycles; for this reason, the inhibition of this enzyme results in cell death (Adams, 2002). Currently, this 26S proteasome inhibitor is used as a therapeutic for targeting cancer cells to induce the disruption the proteasome activity. Considering this, the *in vivo* inhibition of this proteasome inhibitor by the anthrax LF may represent an alternate pathway in order to achieve immune cell apoptosis. The amino acid sequence for this protein was basic with mostly hydrophobic residues. The other candidate, cytochrome *c* oxidase (C*c*O), is a metalloprotein that maintains the proton gradient and synthesis of ATP; in addition, activated oxygen as the terminal acceptor of the transport electron chain (Brunori et al., 1987). Previous studies postulated that the s-nitrosylation process, through the *B. anthracis*-derived nitric oxide (NO), promotes the inactivation of C*c*O and other proteins involved in energy pathways. (Chung et al., 2013). The identified protein has a neutral charge leaving interrogations on this kind of interactions with LF (data not shown), but it contained a substantial region of hydrophobic uncharged amino acids which may play a role in binding. These results suggest a possible inhibition of C*c*O, followed by its release, through the proteolytic activity of LF, in GI anthrax, causing a decrease in efficiency of cellular respiration. Death ligand enhancers (DELE) (table 3, p. 45) promote apoptosis by activating TNF-α and TRAIL; however, this process is due to the binding of death associated protein (DAP3) to DELE (Harada et al., 2010). The same study suggested that the disruption of DELE decreases the apoptosis rate in cells; for this reason, it is unclear the *in vitro* interaction between DELE and LF. One reason can be that LF changes DELE's conformation, increasing its affinity to DAP3, followed by the rest of apoptosis signaling pathways.

Proteins form the P-type ATPases family (tables 3 & 4, pp. 46 & 49) are considered as anion, cation and lipid pumps located in cell membranes; these enzymes generate electrochemical gradients using metabolic energy and also mediate cell signaling pathways (Palmgren et al., 2011). Due to their ability to maintain homeostasis through the transport of different ions such as Ca^{2+}, the lack of activity from these enzymes can result in neurological disorders (van Veen et al., 2014). Accordingly, the disruption of these ion pumps may cause an electrochemical imbalance within the cell, probably, promoting the circulatory shock stage in GI anthrax patients. In fact, a study of cardiomyocytes infected with the anthrax lethal toxin suggested the disruption of their function and of the intracellular calcium transport; however, the mechanism that promotes that event was not clear (Kandadi et al., 2010). Proteins from the MRL family, such as Ras-associated and pleckstrin (RAPH1) (table 3, p. 43), are strongly related to several cellular processes such as cell signaling, adhesion and regulation of Ena/VASP protein family (Legg et al., 2004). In addition, RAPH1 and Ras-related protein (Rab-34) are known to be proteins involved in signal transduction; for this reason, their disruption may serve as platform to destabilize the cell. Despite being the smallest identified peptide, our data suggest Rab34 as a putative target of LF on early endosomes, resulting in the interference of its maturation, allowing the release of the bacteria to the cytosol. Our data not only presents proteins, but also motifs that have specific properties such as DNA binding.

The zinc finger proteins (ZNFs) (table 3, p. 45) are considered as residues of cysteine and histidine, whose center binds zinc and provides a structural stability to the rest of the protein in order to allow interactions with DNA (Brown et al., 1985). These protein motifs are known for their involvement in transcription processes. The amino acid sequences for these proteins varied in length from 20-22 amino acids and were identified as the mutant-interacting peptides. Despite

of finding two types of ZNFs, further studies are needed in order to understand the origin of these motifs. Nevertheless, a study suggests the need of ZNFs, such as the 64, for the activation of pro-inflammatory response in macrophages through TLRs (Wang et al., 2013). Based on this, the found ZNFs can be a representation of transcription factors that are activated during immune responses. Another particular candidate is gastrokine-1 (GKN1), which is a secreted protein from the human stomach, shown in table 4, p. 47. GKN1 was first found in gastric antrum serving as a mitogen at the surface epithelial cells and it was suggested to function as a gastric growth factor (Martin et al., 2003). Another study postulated the decrease of this protein after infection of gastric cells with *Helicobacter pylori* (Nardone et al., 2007). Moreover, this protein is in charge of maintaining the homeostasis of gastric epithelium, regulates tissue repair by the inhibition of cancer cells proliferation through apoptosis, and maintains mucosal integrity (Rippa et al., 2011; Mao et al., 2012). According to this, during an anthrax infection the disruption of GKN1 may result on the inability of cells from epithelial tissue to regenerate, probably, allowing the release of the bacteria out of the GI system.

Considering the properties of the identified GKN1, which contained a BRICHOS domain, these domains are known to interact with trefoils peptides that are regulated by pro-inflammatory cytokines, in order to protect the mucosa barrier and promote mucosa healing (Tebbutt et al., 2002). For this reason, a disruption of this portion of the protein by the proteolytic activity of LF could significate an additional framework to identify the escape of *B. anthracis* from the GI system. Mechanisms of GI anthrax are poorly understood due to the deficit of *in vivo* models of the disease (Xie et al., 2013). *B. anthracis* response in murine macrophages revealed changes in the regulation of several genes, including those related to cell adhesion such as protocadherin 10, in order to transport macrophages to lymph nodes (Bergman et al., 2005). Considering the

disruption of protocadherins in gastric cells may represent an instability in the epithelial cells migration and adhesion. Previous studies using T7PD to determine LF-binding proteins from different human tissues revealed the presence of proteins such as cytochrome *c* oxidase, ribosomal proteins, kinase anchor protein, cadherins, and ATPase (Burgos, 2010). This supports the provided data and brings an idea of the identity of the profile of proteins involved in GI anthrax, especially to understand how this pathogenic bacteria evade immune system, breaks down epithelial barriers, and its dissemination out from the GI system. Finally, in peptides such as AKAP, TLR8, 26S proteasome inhibitor, and protocadherin Fat 2, putative conserved domains were not detected on the *in silico* analysis. Considering this the interactions between LF and these peptides could be due to the chemical properties of these.

Candidates that showed higher E-values were analyzed too because of their possible potential as inhibitors. By using high throughput techniques it is possible to obtain random products; however, these could be selected because of their properties of interest. For this reason, these candidates were exposed on tables 3 and 4, since they showed to be LF-interacting biomolecules. Considering that most of these proteins were only identified at one biopanning cycle the nature of the interactions are not clear. For this reason, the next chapter described if these interactions occur using single candidates to bind LF (WT, MT or both) or to the blocking agent (casein 5%). The idea was to determine if the interactions detected in this chapter occurred randomly during the biopanning step (sec. 2.2.2).

CHAPTER 4

T7PD INDIVIDUAL INTERACTION SPECIFICITY

TEST AND MINIMUM CONCENTRATION FOR

INTERACTION

'4.1 INTRODUCTION

Fully understanding protein-protein interactions could represent a challenge when a high-throughput technique such as T7PD is performed. Previous studies, using the yeast two-hybrid systems, allowed to detect specificity in protein-protein interactions; for instance, for the E2 protein present on pestiviruses, which is involved in virulence (Gladue et al., 2014). In fact, it is important to determine the specificity of proteins to identify their functional activity (Schreiber & Keating, 2011), especially considering that the goal in this work is to better understand GI anthrax molecular pathogenesis. For this reason, it is important to define whether the isolated T7PD candidates are specific to the anthrax LF or if the interactions are due to non-specific properties of the selected biomolecules. In this chapter, 18 T7PD candidates were tested individually to determine if their interactions to LF are specific to each type of ''bait'' used or to the blocking agent (casein 5%). Since the tested candidates could express affinity to LF and casein, varying the percent of Tween 20 on the washing solution Tris buffered saline tween 20 (TBST) was the scaffold to determine whether the interactions to LF were strongest. Tween 20 serves as a surfactant agent that avoids non-specific interactions and increasing it results on stringent washes.

Other scientists performing T7PD used 3% bovine serum albumin (BSA), as a blocking agent (Gnanasekar et al., 2009), for this reason, candidates having a relative equal affinity to LF and casein could be tested with BSA to differentiate interactions. Because recent studies demonstrated that a determined concentration of LF ($1\mu g/mL$) was enough to allow the detection of the NLRPs cleavage (Levinsohn et al., 2012), it was necessary to establish the minimum concentration for interaction (MCI). For this reason, in this chapter a T7PD candidate that binds both types of LF was tested to determine MCI by varying the amount of target. This could provide a better idea at the time of designing therapeutics against LF.

4.2 METHODOLOGY

4.2.1 Specificity test

In order to determine if the interacting peptides are specific to the lethal factor and to determine the minimal concentration of the target protein to allow interactions, a specificity test was done by following the procedure performed by Burgos-Muñiz in 2010. Eighteen putative T7 phage candidates having affinity were used for this test. Each candidate was exposed to interact in wells of 96-well microplates (in duplicates) previously coated with (LF-WT) & (LF-MT) at 11.25µg/mL (sec. 4.2.2) & 22.5µg/mL (sec. 4.2.3) per well and also non-coated but blocked was the negative control. In addition, a well that was blocked but not exposed to the candidates was used as a control to detect crossed contamination. The interaction in each well with each candidate was performed following the adsorption, coating and biopanning protocol described previously (Sec. 2.2.1-2.2.2); however, a modified version of the biopanning step was done.

4.2.2 Detection of individual T7PD candidates that remains on the wells by an *in situ* infection

The first modification consisted in eliminating the elution step and performing an *in situ* infection by inoculating 270µL of a previously induced Rosetta 5615 (OD$_{600}$ 0.5-0.8) on the microplate's wells, after washing steps (fig. 13). The inoculated microplates were incubated at 37°C for six hours. The OD$_{600}$ was measured on time of infection 0hrs, 3hrs and 6hrs. The idea was to determine if a population of phages remains on each well or only at the experimental ones.
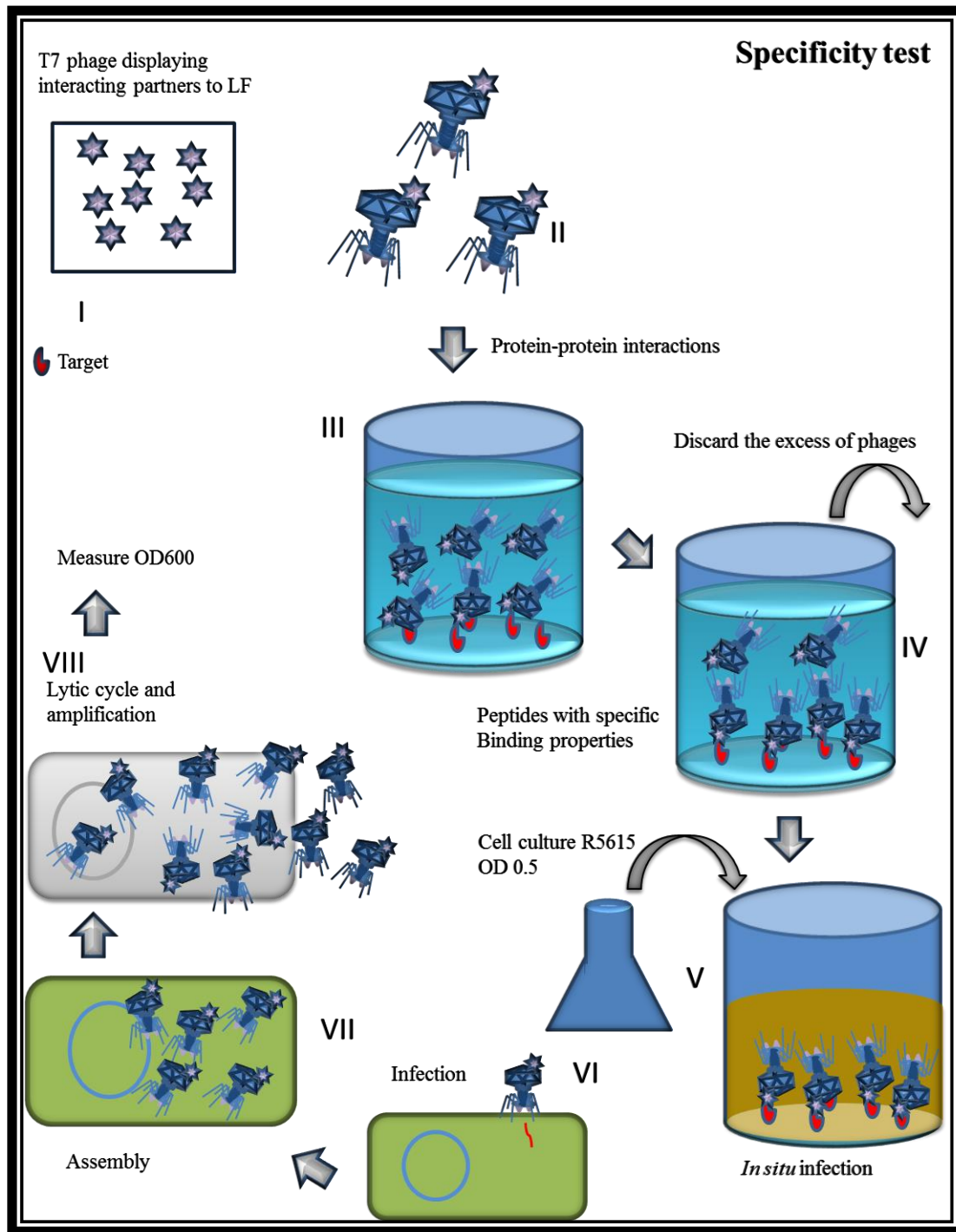
Figure 13. Specificity test. I. Isolated clones expressing peptides with high affinity to LF. II & III. Putative interacting partners to LF are exposed to the previously adsorbed target at different concentrations; in addition, other proteins are going to be used to determine if the candidates are specific ligands of LF. IV. Several stringent washes are performed to discard the excess of T7 phages candidates. V. *In situ* infection performed by adding R5615 to the microplate wells. VI. Transduction. VII. Assembly of phage particles inside the host cells. VIII. Release of mature phage T7 from the R5615 through lysis, measured by OD_{600}.

4.2.3 Elution of the specificity test product to infect a major volume of *Escherichia coli*

The second modification, an alternate protocol, consisted in changing the amount of bacterial cultures to be infected to 5mL of Luria Bertani containing Rosetta 5615 (OD$_{600}$ 0.5-0.8, induced with IPTG). After the specificity test was performed, a total of 25µL from the wells' elution were discarded to the 5mL cultures. The tubes were incubated at 37ºC in an orbital shaker at 150rpm per six hours or until lysis was observed. The OD$_{600}$ of the cultures was measured, after overnight incubation, using an Eppendorf UV biophotometer (Eppendorf, USA). Non-specific binders that showed affinity to casein (control) were subjected then to the same test but varying the percent of Tween 20 from 0.1-0.15%. Once it was determined the percent of Tween 20 needed to disrupt the interactions to the blocking agent, the candidates were then exposed to both types of LF, in order to identify if the interactions to the target protein were stronger. For those candidates that bind equally LF and casein, the same protocol was performed using 3% Bovine Serum Albumin (BSA) as a blocking agent.

4.2.4 Determination of minimal concentration of target protein (LF) needed to detect interaction after the elution of specificity test product

The minimal concentration of LF needed to allow interaction with the candidates was determined by immobilizing different amounts of target protein from 1-10µg/mL (see sec. 2.2.1). For this test, it was chosen the candidate that binds the two types of LF tested (pepsin A3 pre-protein 46aa) (table 3 & 4, pp. 42 & 47). A blocked well without LF was used as control. The candidate was subjected to these wells, following the protocols described on section 2.2.2. The elution from these wells were collected to the corresponding microtubes and then discarded on a previously induced R5615 (5mL, OD$_{600}$ 0.5-0.8). The R5615 cultures were incubated on assay tubes using an orbital shaker (150rpm) at 37°C for 4-6 hours or until lysis was detected.

4.3 RESULTS

4.3.1 Determination of T7PD candidates' presence on microplate's wells after washing steps by an *in situ* infection

After testing all the candidates, in triplicates, with this methodology it was obtained that all the microplate's wells containing R5615 experienced lysis, including controls. When increasing the percent of surfactant agent Tween 20 from 0.1-1% no lysis was detected on the controls; however, the wells containing LF did not show phages presence either. The OD_{600} measured on 0, 3 and 6 hours was analyzed using bar graphs. These graphs (figs. 14 & 15) demonstrated a substantial decrease of the OD_{600} at 6 hours. The shown candidates pepsin A pre-protein 156aa (PAP156) (table 3, p. 42) length and gastric lipase 77aa (GL77) (table 4, p. 47) were taken as an example of the collected data. The average OD_{600} of R5615 exposed to pepsin A phage at 3 hours on experimental wells was of 0.77, its control had a 0.81 and the well not exposed to phages 1.54. At 6 hours, the average optic density for pepsin on experimental wells decreased, from the OD_{600} at 3 hours of infection, a 57% for LF-MT and 70% on LF-WT. The R5615 control experienced an OD_{600} decrease of 56% and the non-T7 well a 13%, from the 3 hours measurements (fig. 14).

On the other hand, R5616 average OD_{600}, when exposed to gastric lipase T7 phage, at 3 hours on experimental wells was of 0.63, the control 1.58 and the non-infected well 1.54. After six hours of infection the average OD_{600} measured for *E. coli* on LF wells decreased a 46% for MT and 56% for WT. The R5615 control optical density decreased a 37%, while the well not exposed to the candidate phage a 13% (fig. 15). These decreasing percent values are taken considering the change in OD_{600} between 3 and 6 hours of infection. The error bars for pepsin A varied from 0.08-0.12 and for gastric lipase 0.14-0.43 at 3 hours of infection. At 6 hours of infection the error bars for pepsin A were from 0.02-0.04 and gastric lipase 0.02-0.44.
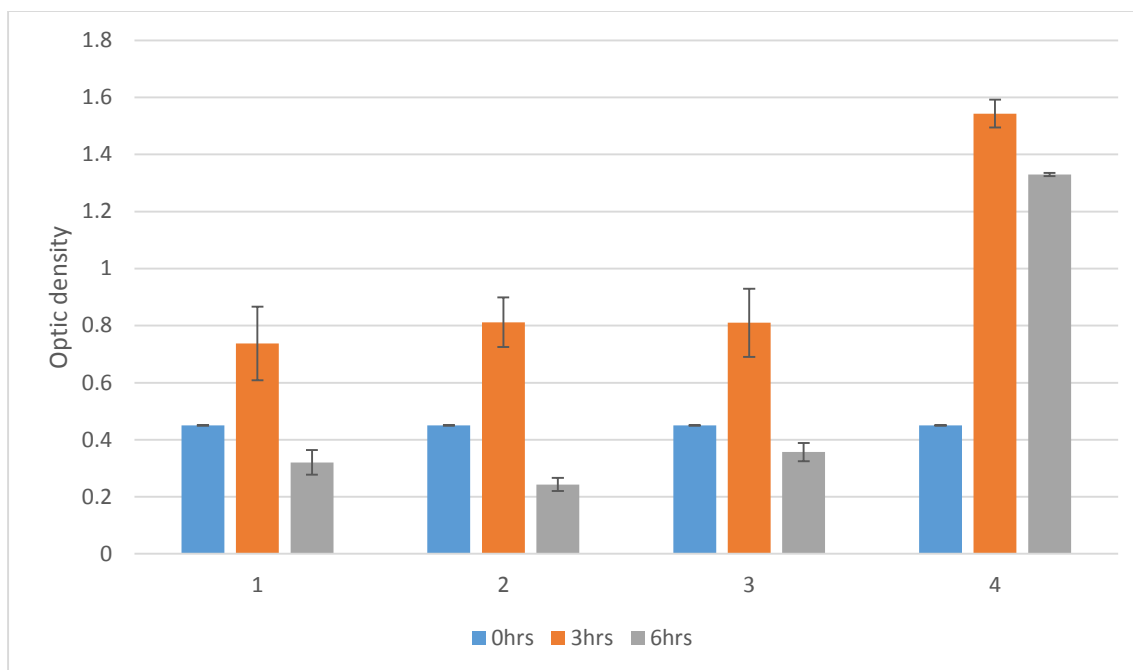
Figure 14. Optic density vs. time (Pepsin A). Pepsin A specificity test bar graphs for R5616 optic density vs. time (hours). Error bars represent standard error for the triplicates used for each graphed average value. (1 & 2) Represent the values for a R5615 inoculated on a well containing LF-MT and LF-WT (11.25µg/mL), respectively, that were exposed to PAP156 phage. (3) Average values for a non-LF well (casein 5%) exposed to the candidate. (4) Graphed average values for a R5615 on microplate's wells not exposed to the candidate phage.
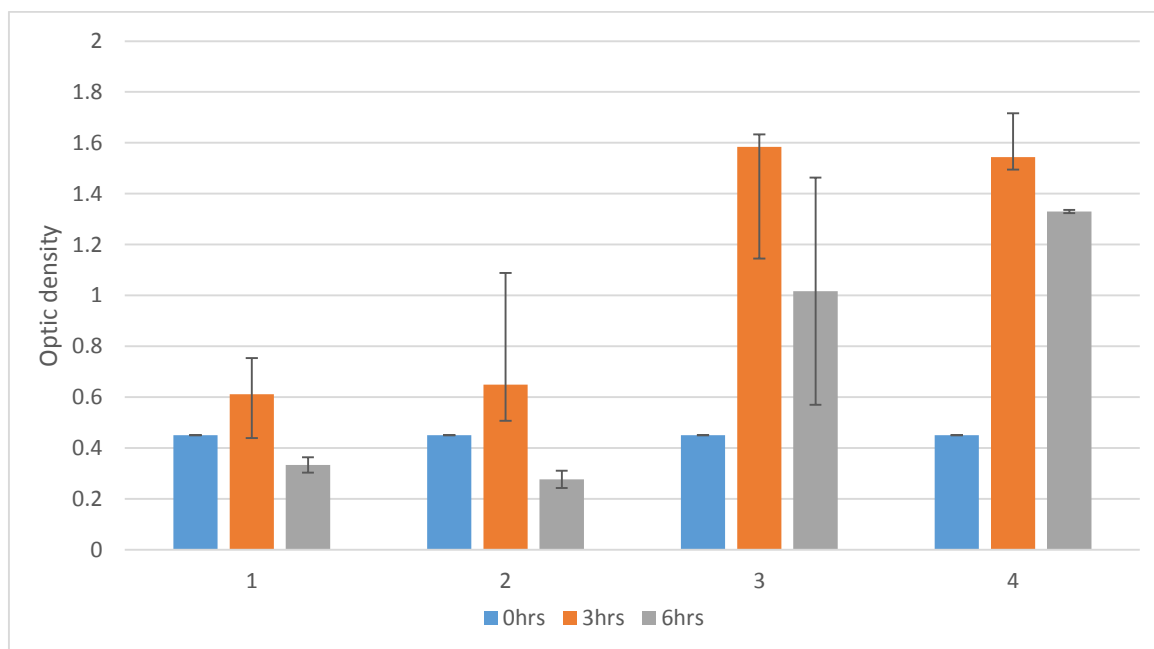
Figure 15. Optic density vs. time (Gastric lipase).Gastric lipase specificity test bar graphs for R5616 optic density vs. time (hours). Error bars represent standard error for the triplicates used for each graphed average value. (1 & 2) Represent the values for a R5615 inoculated on a well containing LF-MT and LF-WT (11.25µg/mL), respectively, that were exposed to GL77 phage. (3) Average values for a non-LF well (casein 5%) exposed to the candidate. (4) Graphed average values for a R5615 on microplate's wells not exposed to the candidate phage.

4.3.2 Elution and infection of the specificity test product to a major volume of R5615

For this test, 18 candidates were selected by their conservation in open reading frame (ORF) at sequencing, which consisted in obtaining the correct amino acid sequences for each isolated clone. A total of 10 T7PD candidates showed affinity to at least one type of LF tested (table 5). From these candidates, a total of 4 proteins were exclusive to LF-WT (fig. 16) and 2 to LF-MT (fig. 17). Three of these peptides showed affinity to both types of LF tested and two showed affinity to the blocking agent and LF (fig. 18). These proteins were GDNF inducible zinc finger protein (GZNFP), open reading frame 96 (ORF96), and pepsin A3 pre-protein (PAP46).

The first seven proteins shown on table 5 were the results of using casein 5% as a blocking agent and a 0.1% of Tween 20 on the TBS, for the washing steps. For these candidates, lysis was evident after 4 hours of infection and the OD_{600} was measured after overnight incubation at 37°C (table 6). For PAP156, at 0.1% of Tween 20, lysis occurred on all the assay tubes containing the test product of wells with LF and without it. Increasing the Tween 20 percent to 0.15 disrupted the interaction between PAP156 and the blocking agent allowing the detection of interaction with LF-MT. On the other hand, PAP46 (table 3 & 4, pp. 42 & 47) was only tested with 3% BSA as blocking agent and it was found that this protein interacted with both types of LF tested. For gastric lipase (GL) it was unclear the nature of interactions because of its interactions with both types of blocking agents used.

Table 5. Specificity test performed to the isolated T7PD candidates. Lysis detection after 4-6hrs of incubation.

Protein	Original interacting ''bait''	LF-MT	LF-WT	Type of blocking agent
Interleukin enhancer-binding factor 2 (IEBF2)	LF-WT	NL	+	NL C
Copper transporting ATPase (67)	LF-WT	NL	+	NL C
Death ligand signal enhancer (56)	LF-MT	+	NL	NL C
Probable phospholipid ATPase (29)	LF-MT	NL	+	NL C
Cytochrome *c* oxidase (18WT)	LF-WT	NL	+	NL C
GDNF-inducible zinc finger protein (15)	LF-MT	+	+	NL C
Homo sapiens chromosome 11 open reading frame 96 (16)	LF-MT	+	+	NL C
Pepsin A preprotein (156aa)	LF-MT	+ +	+ NL	+C NL C/T
Pepsin A3 preprotein (46aa)	LF-WT	+	+	*-B*
Gastric triacylglycerol lipase	LF-WT	NL +	+ +	+C/T +B
Total of interacting proteins	-	6	8	2

+= Lysed cultures., NL= No lysis detected., C= 5% casein., B= 3% BSA., C/T= Increase in Tween 20 using casein

Table 6. Optical density (OD600) of specificity test candidates after overnight incubation.

Protein	LF-MT OD$_{600}$	LF-WT OD$_{600}$	Blocking agent
Interleukin enhancer-binding factor 2	2.490	0.551	2.656 C
Copper transporting ATPase (67)	2.503	0.676	2.624 C
Death ligand signal enhancer (56)	0.704	2.527	2.622 C
Probable phospholipid ATPase (29)	2.412	0.529	2.320 C
Cytochrome *c* oxidase (18WT)	2.401	0.401	2.249 C
GDNF-inducible zinc finger protein (15)	0.903	0.610	2.167 C
Homo sapiens chromosome 11 open reading frame 96 (16)	0.190	0.209	1.982 C
Pepsin A3	0.345	0.401	2.383 B
Pepsin A preproteins (156aa)	0.411 0.355	0.456 2.121	0.429 C 2.233 C/T
Gastric triacylglycerol lipase	0.45 0.390	0.387 0.5	0.366 C 0.445 B

C= 5% casein., B= 3% BSA., C/T= 5% casein and increase in Tween 20
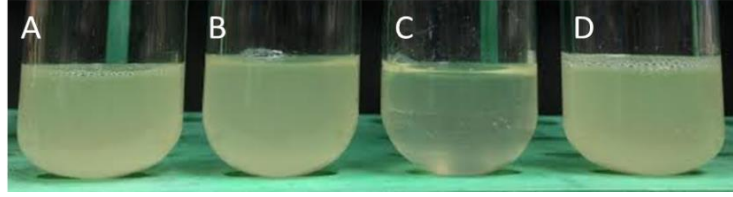
Figure 16. Interaction confirmation by performing a specificity test to candidate Probable phospholipid ATPase. (A) Control R5615 which consisted on the addition of elution solution from a microplate well no exposed to candidate. (B) *E. coli* with the eluted product of candidates exposed to LF-MT. (C) Lysis detected with an *E. coli* exposed to the eluted product of the T7PD candidates exposed to LF-WT. (D) *E. coli* exposed to the elution of phages from the control wells, consisting of 5% casein as blocking agent.
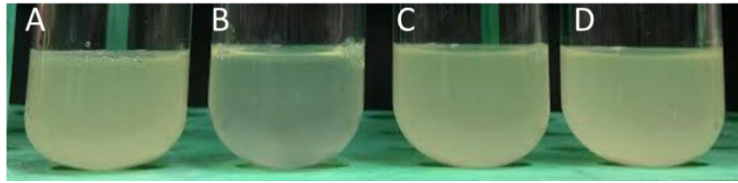
Figure 17. Interaction confirmation by performing a specificity test to candidate death ligand signal enhancer. (A) Control R5615 which consisted on the addition of elution solution from a microplate well no exposed to candidate. (B) Lysis detected with an *E. coli* exposed to the eluted product of candidates exposed to LF-MT. (C) *E. coli* with the eluted product of the T7PD candidate exposed to LF-WT. (D) *E. coli* exposed to the elution of phages from the control wells, consisting of 5% casein as blocking agent.
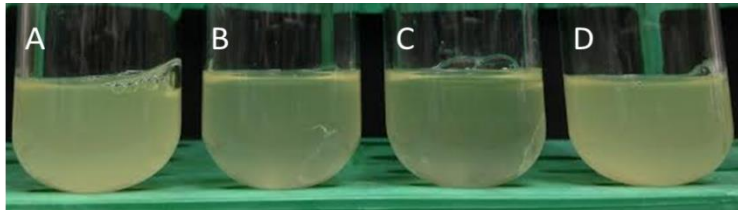
Figure 18. Interaction confirmation by performing a specificity test to candidate GZNFP. (A) Control R5615 which consisted on the addition of elution solution from a microplate well not exposed to candidate. (B) Lysis detected with an *E. coli* with the eluted product of candidates exposed to LF-MT. (C) Lysis detected on *E. coli* exposed to the eluted product of the T7PD candidate exposed to LF-WT. (D) *E. coli* exposed to the elution of phages from the control wells, consisting of 5% casein as blocking agent.
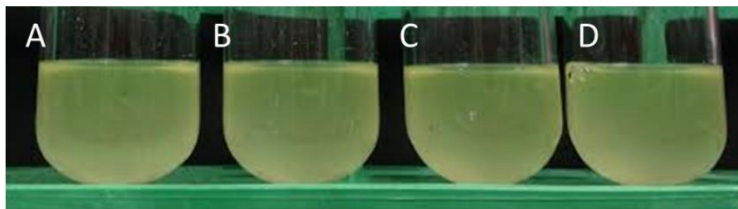
Figure 19. A T7PD candidate that does not bind any of the LF tested. (A) Control R5615 which consisted on the addition of elution solution from a microplate well not exposed to candidate. (B & C) No lysis detection by exposing *E. coli* to the elution product of LF-MT and LF-WT, respectively. (D) *E. coli* exposed to the elution of phages from the control wells, consisting of 5% casein as blocking agent.

The minimum concentration for interaction (MCI) revealed that using 1µg/mL of each type of LF was enough to allow interaction with PAP46 (figs. 20 & 21). The blocking agent used for this test was 3% BSA and the candidate did not show affinity to it. The OD_{600} values obtained after overnight incubation for this test are shown on table 7. These measurements indicated lysis on wells treated with LF.
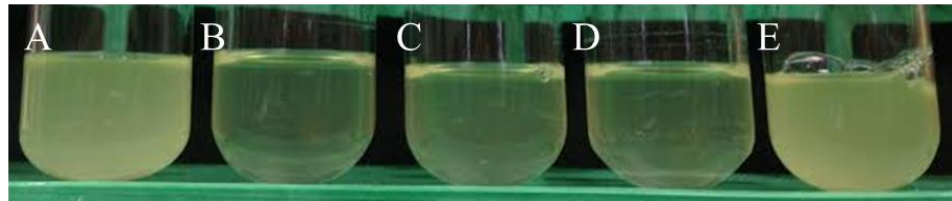
Figure 20. Minimum concentration for interaction (MCI) for PAP46 with LF-MT. (A) An assay tube with R5615 that was subjected to the product of a non-phage treated well. (B, C & D) R5615 with the product of wells subjected to PAP46 against LF-MT at 1, 2 and 3µg/mL, respectively. (E) R5615 with the elution of a control well (3% BSA).
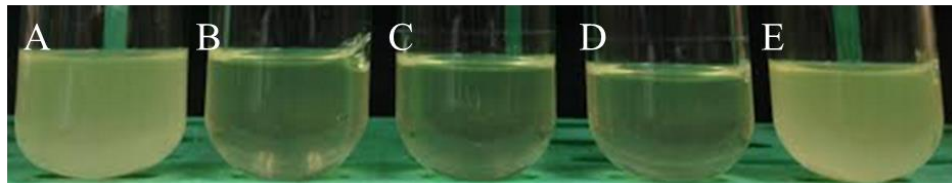
Figure 21. Minimum concentration for interaction (MCI) for PAP46 with LF-WT. (A) An assay tube with R5615 that was subjected to the product of a non-phage treated well. (B, C & D) R5615 with the product of wells subjected to PAP46 against LF-WT at 1, 2 and 3µg/mL, respectively. (E) R5615 with the elution of a control well (3% BSA).

Table 7. Optic density (OD_{600}) value for the minimum concentration of LF for interaction with PAP46.

Values measured after overnight incubation.

Bait	1μg/mL	2μg/mL	3μg/mL	Controls
LF-MT	0.278	0.343	0.278	-
LF-WT	0.310	0.282	0.266	-
3% BSA	-	-	-	2.113
None	-	-	-	2.183

- = An OD_{600} measurement does not apply.

4.4 DISCUSSION OF RESULTS

After testing the specificity test by adding the R5615 to the microplate's wells, subjected to aliquots at 1.4×10^5 pfu/mL of individual T7PD candidates, it was found that a small portion of phages remains on each well. In figures 14 & 15 it is evident how the OD_{600} of R5615 decreases after 6 hours of inoculation on all the samples; however, the *E. coli* inoculated on a non-treated with phage wells maintained an higher optic density. For pepsin A pre-protein 156aa (PAP156), the standard error values were small indicating a constancy on the triplicates obtained for measure the average OD_{600} of each bar. On the other hand, standard error values were higher for gastric lipase (GL77) and its control consisting of 5% casein was close to the R5615 on non-T7 treated wells. Considering that all the candidates produced lysis after six hours of inoculation of R5615 on the wells (data not shown), this suggests that a small portion of phages remains on the wells. Probably these phages could be weakly interacting with the microplate material or with a component present on the blocking agent. Increasing the Tween 20 percent up to 0.2% to solve the problem resulted in no lysis detection, neither on experimental wells. Decreasing the number of phages produced the same results. If this tests were taken as real data, all the tested candidates can be considered as casein or plastic binding partners. For this reason, it was considered that the volume of R5615 (270μL) in each well was a small amount that can be lysed by a random present phage on the sample.

To avoid changing the percent of surfactant agents or the number of phages, the same protocol was performed using LF at 22.5μg/mL to allow detection of interaction and the product was transferred to 5mL R5615 cultures. Considering that lysis was detected on heterogeneous mixtures of T7 candidates using 50mL cultures (sec. 2.3.1), at that point if these phages interacted with different sites of LF allowed to detect lysis. On the specificity test, single candidates were tested

and were thought to interact with a specific site of each type of LF tested, decreasing on the washing steps the number of phages necessary to allow lysis.

After using this methodology, it was found that at least 10 candidates showed putative affinity to LF *in vitro*. On table 5 it is shown whether this T7PD candidates were specific to LF-MT, LF-WT or to the blocking agent consisting of 5% casein. Because of the lack in protease activity in LF-MT it was expected that binders to LF-WT should be present on mutant type, through the recognition of domains II and III. As mentioned before, the LF contains 4 domains, being the I necessary to interact with PA, the II & III for substrate recognition and the IV is the catalytic center (Pannifer et al., 2001). Considering this, binders to the IV domain in LF-MT could be detected and LF-WT partners not, unless it binds far from the active site sequence (HEXXH) proposed to be essential for metalloproteases activity (Jongeneel et al., 1989; Klimpel et al., 1994). This, could be an explanation to proteins such as IEBF2, copper transporting ATPase, Probable phospholipid ATPase, and cytochrome *c* oxidase, which were exclusive to LF-WT. Candidates, such as Death ligand signal enhancer and PAP156, could be interacting with the active site of LF and its cleavage promotes their detection using LF-MT as bait. However, using PAP46, the short version in length of PAP156, allowed interaction with both types of LF tested.

On table 5, it was shown that PAP156 showed affinity to the blocking agent (5% casein) when using Tween 20 at 0.1%. Nevertheless, an increase of this reagent to 0.15% allowed to demonstrate that this protein has more affinity to LF-MT than the blocking agent. It is important to mention that this protein was originally found as a partner of the mutant type LF. For this reason, when performing T7PD technique it must be considered the presence of this particular group of candidates on the controls and also on experimental wells, at the biopanning steps. However, in

this case it was demonstrated that the population of phages recovered from experimental well was higher.

Determining the minimum concentration of LF needed to allow interaction with at least one of the T7PD candidates was an important step to characterize the binding properties. On figures 20 & 21, it was shown that lysis occurred on R5615 that was subjected to the elution product of wells containing from 1-3µg/mL of each type of LF tested. This data suggest that a small amount of the target protein is recognized by PAP46 (table 3 & 4, pp. 42 & 47), being this a potential candidate to bind LF at least *in vitro.*

Once the methodology described on sec. 4.2.2 was repeated several times, obtaining the same results, it was determined that a small fraction of phages remains on each well after biopannings. These random entities could result in posterior problems when a specificity tests are performed. In this case, this could explain why after alternating to the protocol described in sec. 4.2.3, from 18 proteins only 10 were found as LF-interacting partners. The optic density values shown on table 6 and figures 16-19 were the necessary evidence to determine the candidates that best binds to LF by performing this alternate protocol of T7PD. In this chapter, not only it was found LF interacting partners, it was also demonstrated that the technique can be also used for single candidates. In addition, the presence of non-specific binders to LF and casein, allowed the development of a protocol to determine which of these interactions are stronger. Nevertheless, the nature of these interactions remains unclear. For this reason, in the next chapter an additional *in silico* analysis to the T7PD candidates was performed in order to better understand the interactions.

CHAPTER 5

IN SILICO ANALYSIS OF THE T7PD PEPTIDES

THAT SHOWED AFFINITY TO LF: ALIGNMENTS,

PREDICTED PHYSICAL AND CHEMICAL

PROPERTIES

5.1 INTRODUCTION

Determining peptides, from the whole population of libraries, that binds a specific target could represent a challenge if conserved amino acid sequences are not identified *in silico*, at separated events. This chapter is based on the candidates' sequence comparisons and their predicted chemical & structural properties. This analysis of the putative interactions could elucidate the mode of action of the GI anthrax disease *in silico*. For this reason, it is important the use of these tools to address the detection of conserved amino acid residues that interact *in vitro* with the anthrax LF.

Protein alignments allow comparisons between species, drug targets, and also in the understanding of the current public health problem of antibiotic resistant bacteria. Moreover, proteomics' techniques such as Phage Display take advantages of *in silico* tools in order to develop therapeutics, biosensors, and to determine protein ligands. ClustalX is a multiple alignment program for medium sized data in a short period of time and easy (Larkin et al., 2007). For this reason, all the selected T7PD candidates can be analyzed first using this bioinformatics program to identify conserved peptides having putative interactions with LF. On the other hand, once the common amino acid regions are detected on the whole population of possible LF binding peptides, a subsequent analysis of these peptides is necessary to understand the nature of the interactions. PRALINE is a multiple sequence alignment application that allows us to compare sequences conservation, hydrophobicity, secondary structure predictions and the types of residues present on the data (Simossis et al., 2005). Using ScanProsite, common motifs on the candidates helped to identify similarities. Considering this, the chemical properties of the candidates that show more possible affinity to LF *in vitro* can be determined.

5.2 METHODOLOGY

5.2.1 Determination of conserved amino acid sequences of the T7PD isolated candidates and predicted chemical properties

Once all the T7PD selected candidates were identified then were subjected to ClustalX 2.1 alignment in order to identify conserved amino acid sequences. The program was obtained from (http://www.clustal.org/clustal2/). The ID given to the peptides were based on their accession number obtained using BLAST, combined with the original target of isolation. The multiple alignment of the sequences was done using BLOSUM series as the protein weigh matrix with a gap opening and extension cost of 3 and 1.8, respectively. In addition, a multiple alignment was done to all the peptides that belong to the proteins that showed affinity *in vitro* to the anthrax LF on chapter 4.

From the above *in silico* assays, peptides with conserved amino acid sequences were aligned separatedly using ClustalX 2.1. Then web pages such as PRALINE, from the Center for Integrative Bioinformatics VU (IBIVU) (http://www.ibi.vu.nl/programs/pralinewww/), allowed us to establish a prediction of the chemical and structural properties of the T7PD candidates. Conserved regions were annotated as the most interacting amino acid residues having affinity to the (LF-WT) or (LF-MT). Full length reference's sequences for peptides with highly conserved amino acids were used for comparisons and these were obtained from Uniprot databases. These peptides were pepsin A (EC:3.4.23.1) and gastric lipase (EC:3.1.1.3). PRALINE was also used to determine the chemical properties of those T7PD candidates that showed affinity to LF in chapter 4.

To determine cleavage sites for available proteinases PeptideCutter (http://web.expasy.org/peptide_cutter/) from Expasy was used for four of the candidates. Considering that LF shares the motif sequence HEXXH on the active site with thermolysin, according to Klimpel et al., 1994, this peptidase was used as a model for cleavage detection. The MAPKK1 (EC:2.7.12.2) full length sequence, obtained from Uniprot databases, was used as a control for this *in silico* assay. Through this assay a prediction of the putative cleavage sites performed by LF to the candidates was determined. For PAP46 and cytochrome *c* oxidase from tables 3 & 4, pp. 42, 47 & 48, a proposed 3D structure model was designed by using SWISS-MODEL (http://swissmodel.expasy.org/). The motifs present on the isolated T7PD candidates were identified both by excluding motifs with a high probability of occurrence and without excluding, through Prosite (http://prosite.expasy.org/).

5.3 RESULTS

5.3.1 ClustalX and PRALINE alignments and hydrophobicity prediction for the T7PD LF-interacting partners

After comparing all the candidates through alignments using ClustalX 2.1, three regions of conserved amino acid sequences were detected (fig. 22). These regions coincided with proteins such as pepsin, GL, and SIGIRR (tables 3 & 4, pp. 42, 46 & 47) combined with unknown proteins. Peptides having possible affinity with LF-MT and LF-WT at the specificity test were then aligned, resulting in two conserved amino acids regions for each type of LF tested (fig. 23). However, no similarities in sequences were detected for the other candidates. The ClustalX alignments for the conserved regions demonstrated that the amino acid sequences were highly conserved for the region that belongs to candidate pepsin (fig. 24). In comparison with the reference sequence obtained from Uniprot (EC:3.4.23.1) which consists of 388 residues, isolated T7PD peptides matches from residue 233 to 388. All the peptides that were identified as pepsin coincided on the C-terminus of the reference protein. On the other hand, the further conserved amino acids sequences belongs to the gastric lipase. These were compared with a full length reference protein from Uniprot (EC:3.1.1.3) by performing ClustalX alignments, resulting in highly conserved residues (fig. 25). The reference protein consists of 365 amino acids, being the residues from 289 to 365 present on the isolated LF partners. As the pepsin, these peptides coincided with the C-terminus of the reference protein. The hydrophobicity scale for pepsin and GL (figs. 26 & 27) revealed that the peptides consist mostly of hydrophobic residues according to the Eisenberg scale. Secondary structure predictions for pepsin revealed that the peptides consist mostly of strand structures from 21 to ≥66 amino acids, depending upon the peptide (fig. 28). However, the gastric lipase peptides showed substantial regions of helix structure, varying from 32-35aa (fig. 29).
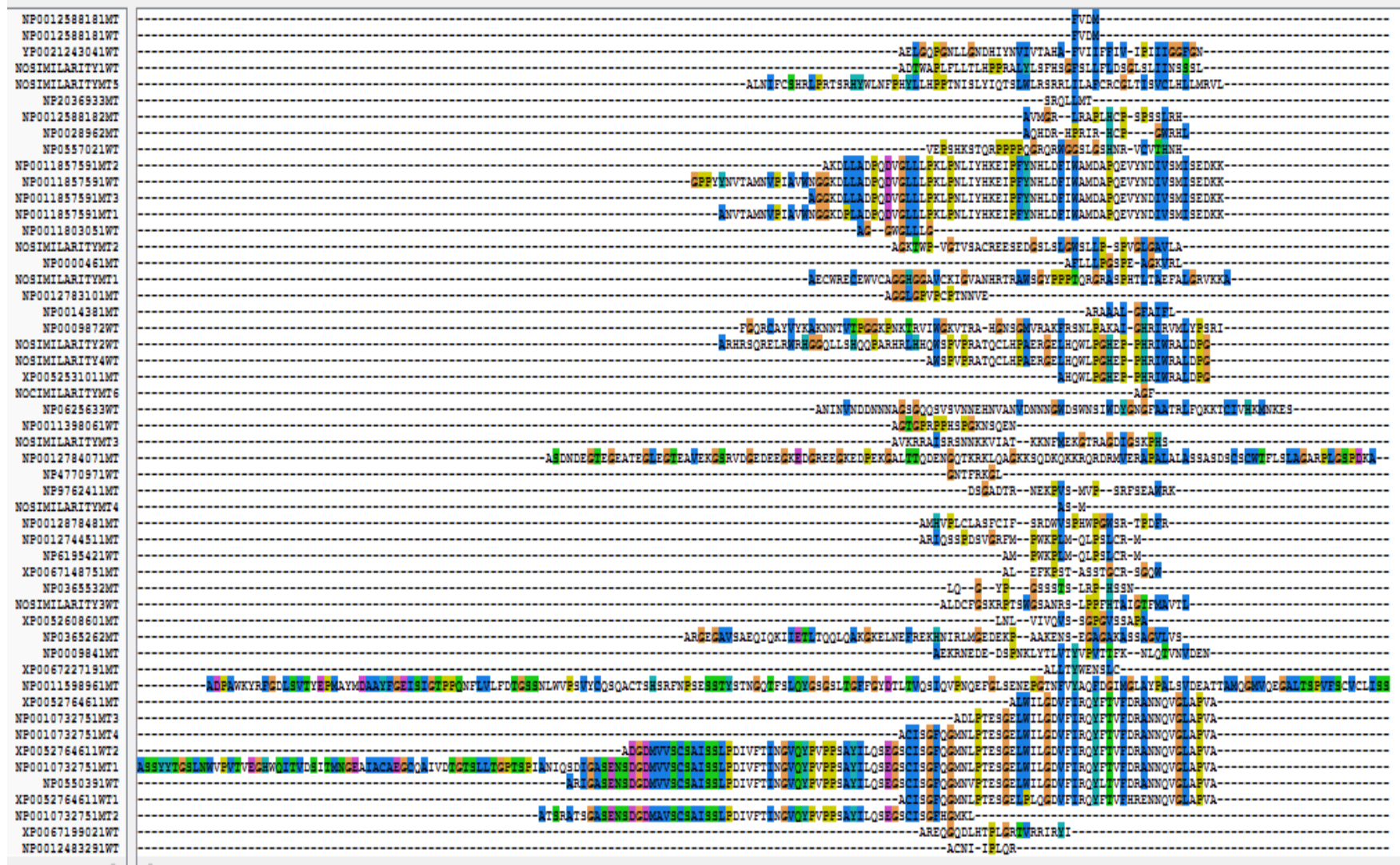
Figure 22. ClustalX alignments for both types of LF-interacting partners used as ''baits''. A total of 52 peptides' sequences from human stomach were subjected to ClustalX alignments to determine consensus sequences between proteins that binds LF-MT and those partners for LF-WT. Common regions of residues between the candidates are shown by red rectangles. The candidates are named by their accession code on NCBI BLAST with the respective type of LF of isolation.

78
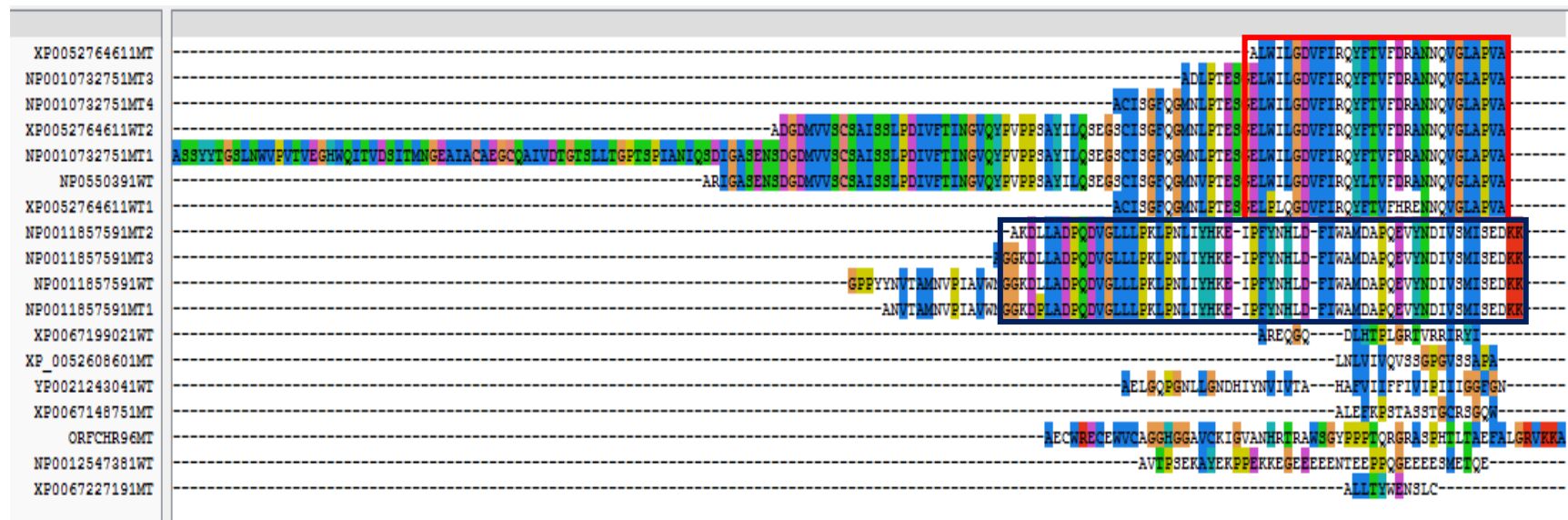
Figure 23. ClustalX 2.1 multiple alignment for T7PD candidates that showed putative affinity to LF on chapter 4. Two conserved regions were detected, being these peptides part of two different proteins. No similarity was found for the rest of the candidates. The candidates are named by their accession code on NCBI BLAST with the respective type of LF of isolation.
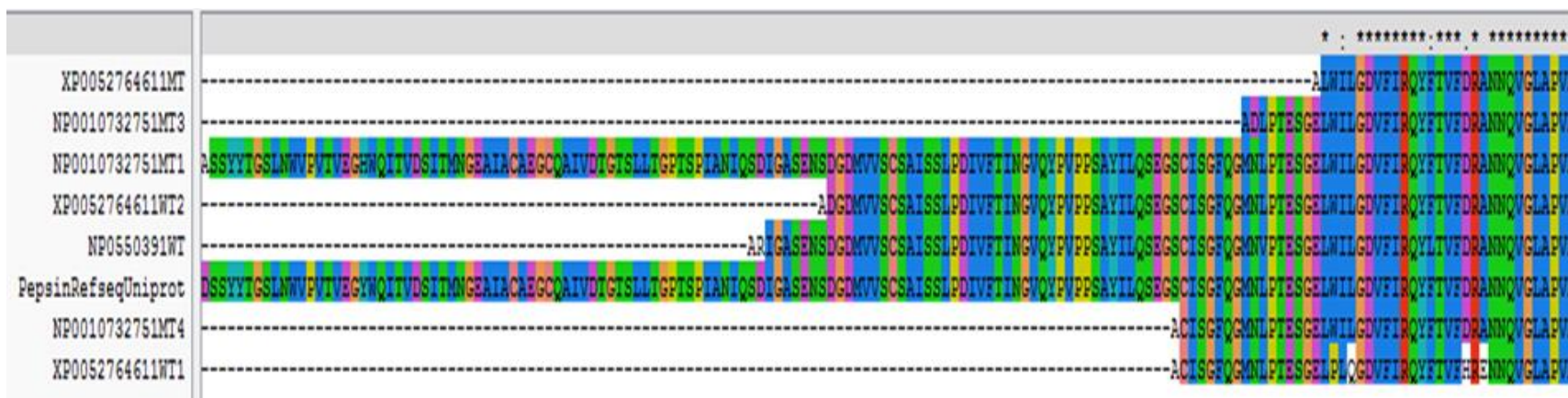
Figure 24, Amino acids' ClustalX alignment for Pepsin, a putative binding peptide for both LFs tested. The region marked with (*) on the top represents the amino acids that were highly conserved or present between all the peptides tested. A consensus sequence is evident on the shown data. The pepsin clones also showed conserved amino acids with the C-terminal of the full length reference protein (EC:3.4.23.1). The candidates are named by their accession code on NCBI BLAST with the respective type of LF of isolation.
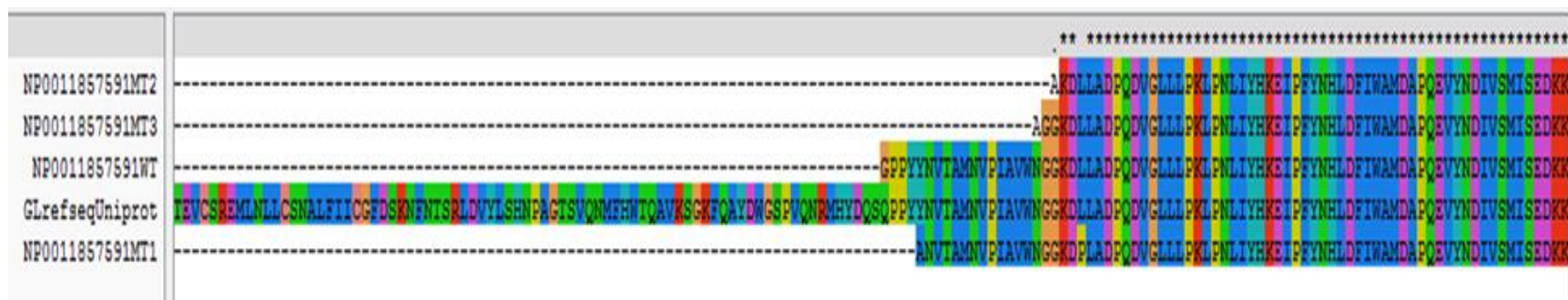
Figure 25. Amino acids' ClustalX alignment for Gastric lipase (GL), a putative binding peptide for both LFs tested. The region marked with (*) on the top represents the amino acids that were highly conserved or present between all the peptides tested. A consensus sequence is evident on the shown data. The alignments for GLs are shown, compared with a reference sequence obtained from Uniprot databases (EC:3.1.1.3). The candidates are named by their accession code on NCBI BLAST with the respective type of LF of isolation.

Figure 26. Hydrophobicity predictions for the candidate Pepsin different length peptides. A color scheme is shown as a prediction of the hydrophobic or hydrophilic state, from left to right, of the residues. The predictions provided by PRALINE are based on an Eisenberg scale showed on the top (IBIVU 2015).

Figure 27. Hydrophobicity predictions for the candidate GL different length peptides. A color scheme is shown in figure 26 as a prediction of the hydrophobic or hydrophilic state, from left to right, of the residues. The predictions provided by PRALINE are based on an Eisenberg scale showed on the top (IBIVU 2015).

```
                  .........210.........220.........230.........240.........250
(PRED) PepsinRefseqUni NIWNQGLVSQ  DLFSVYLSAD  DQSGSVVIFG  GIDSSYYTGS  LNWVPVTVEG
(PRED) NP0010732751MT1 ----------  ----------  ----------  --ASSYYTGS  LNWVPVTVEG
(PRED) NP0550391WT     ----------  ----------  ----------  ----------  ----------
(PRED) XP0052764611WT2 ----------  ----------  ----------  ----------  ----------
(PRED) NP0010732751MT4 ----------  ----------  ----------  ----------  ----------
(PRED) XP0052764611WT1 ----------  ----------  ----------  ----------  ----------
(PRED) NP0010732751MT3 ----------  ----------  ----------  ----------  ----------
(PRED) XP0052764611MT  ----------  ----------  ----------  ----------  ----------


                  .........260.........270.........280.........290.........300
(PRED) PepsinRefseqUni YWQITVDSIT  MNGEAIACAE  GCQAIVDTGT  SLLTGPTSPI  ANIQSDIGAS
(PRED) NP0010732751MT1 HWQITVDSIT  MNGEAIACAE  GCQAIVDTGT  SLLTGPTSPI  ANIQSDIGAS
(PRED) NP0550391WT     ----------  ----------  ----------  ----------  ----ARIGAS
(PRED) XP0052764611WT2 ----------  ----------  ----------  ----------  ----------
(PRED) NP0010732751MT4 ----------  ----------  ----------  ----------  ----------
(PRED) XP0052764611WT1 ----------  ----------  ----------  ----------  ----------
(PRED) NP0010732751MT3 ----------  ----------  ----------  ----------  ----------
(PRED) XP0052764611MT  ----------  ----------  ----------  ----------  ----------


                  .........310.........320.........330.........340.........350
(PRED) PepsinRefseqUni ENSDGDMVVS  CSAISSLPDI  VFTINGVQYP  VPPSAYILQS  EGSCISGFQG
(PRED) NP0010732751MT1 ENSDGDMVVS  CSAISSLPDI  VFTINGVQYP  VPPSAYILQS  EGSCISGFQG
(PRED) NP0550391WT      ENSDGDMVVS CSAISSLPDI  VFTINGVQYP  VPPSAYILQS  EGSCISGFQG
(PRED) XP0052764611WT2 --ADGDMVVS  CSAISSLPDI  VFTINGVQYP  VPPSAYILQS  EGSCISGFQG
(PRED) NP0010732751MT4 ----------  ----------  ----------  ----------  --ACISGFQG
(PRED) XP0052764611WT1 ----------  ----------  ----------  ----------  --ACISGFQG
(PRED) NP0010732751MT3 ----------  ----------  ----------  ----------  ----------
(PRED) XP0052764611MT  ----------  ----------  ----------  ----------  ----------


                  .........360.........370.........380........
(PRED) PepsinRefseqUni MNLPTESGEL  WILGDVFIRQ  YFTVFDRANN  QVGLAPVA
(PRED) NP0010732751MT1 MNLPTESGEL  WILGDVFIRQ  YFTVFDRANN  QVGLAPVA
(PRED) NP0550391WT     MNVPTESGEL  WILGDVFIRQ  YLTVFDRANN  QVGLAPVA
(PRED) XP0052764611WT2 MNLPTESGEL  WILGDVFIRQ  YFTVFDRANN  QVGLAPVA
(PRED) NP0010732751MT4 MNLPTESGEL  WILGDVFIRQ  YFTVFDRANN  QVGLAPVA
(PRED) XP0052764611WT1 MNLPTESGEL  PLQGDVFIRQ  YFTVFHRENN  QVGLAPVA
(PRED) NP0010732751MT3 ADLPTESGEL  WILGDVFIRQ  YFTVFDRANN  QVGLAPVA
(PRED) XP0052764611MT  --------AL  WILGDVFIRQ  YFTVFDRANN  QVGLAPVA
```

HELIX (H) STRAND (E)

Figure 28. Secondary structure prediction for the isolated pepsin T7PD candidate showing affinity to LF. The secondary structure is shown by colors, where the red represents helix structures and blue is for strand structures. Areas without the mentioned colors represents those without a prediction available on the server. The predictions were done using the database of secondary structure assignments (DSSP) and the Protein structure prediction server (PSIPRED).
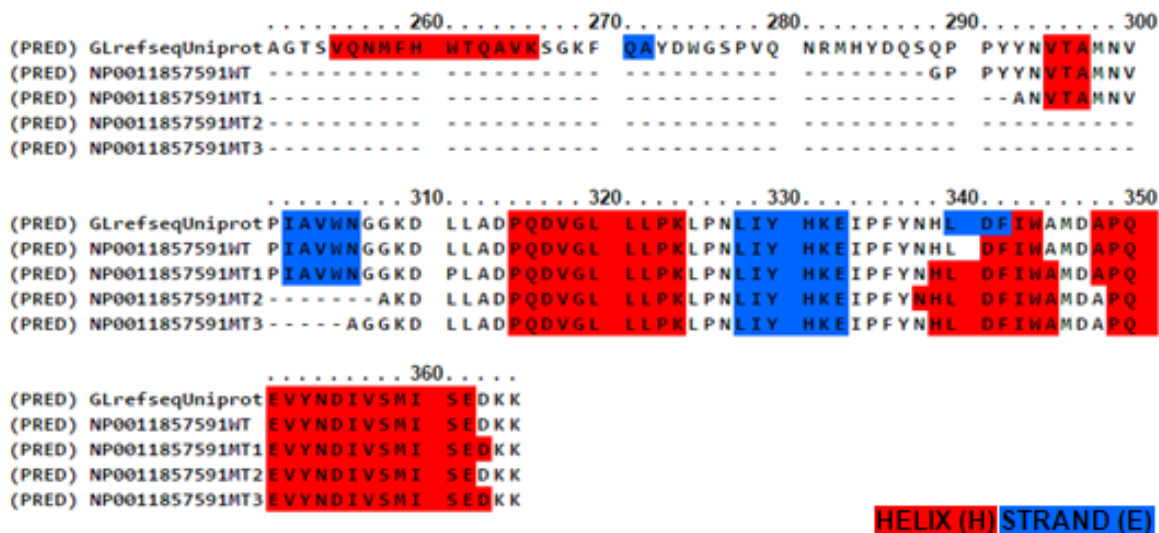
Figure 29. Secondary structure prediction for the isolated GL T7PD candidate showing affinity to LF. The secondary structure is shown by colors, where the red represents helix structures and blue is for strand structures. Areas without the mentioned colors represents those without a prediction available on the server. The predictions were done using the database of secondary structure assignments (DSSP) and the Protein structure prediction server (PSIPRED) (IBIVU 2015).

The Prosite scan revealed that the candidates that showed affinity to LF in chapter 4 contained

mostly N-myristoylation and Protein kinase C phosphorylation sites. After performing an *in silico*

cleavage of the control protein MAPKK1, obtained from Uniprot, several cuts were obtained with

the available proteinases (fig. 30). In addition, it was found a cleavage, done by thermolysin, on

the same site expected by LF which was proposed by Duesbery et al. (1998); Vitale et al. (2000);

Bardwell et al. (2009) (see. fig 5). A total of 108 cleavage sites, on the full length protein (not

shown), for thermolysin on MAPKK1were predicted by PeptideCutter.

```
                                               Therm
                                         Therm   |
             Therm      Therm     Therm    |     |        Therm
         Therm  |      Therm| Therm   |    |     |        Therm|
                                         |  |    |     Therm  || Therm

            | |           ||    |    |  |    |        |  ||      |
         MPKKKPTPIQLNPAPDGSAVNGTSSAETNLEALQKKLEELELDEQQRKRLEAFLTQKQKV
      1  ---------+---------+---------+---------+---------+---------+   60
```

Figure 30. A prediction for cleavage sites on MAPKK1 N-terminal using PeptideCutter from Expasy. The Sequence was obtained from Uniprot databases (EC:2.7.12.2).

85

In addition, 13, 7, 20 and 15 cleavage sites, for thermolysin, were detected on peptides PAP46, GZNFP, cytochrome *c* oxidase, and ORF96 peptides, respectively (fig. 31).

A

B

```
                                          Therm
                                         Therm |
                                         Therm| |
                                         Therm|| |
                                        Therm  |||  |
                                        Therm|   |||  |
                                       Therm||   |||  |
                                       Therm|||   |||  |
                                       Therm|||   |||  |
                                      Therm|||||   |||  |
                                      Therm|||||   |||  |
                                     Therm||||||   |||  |
                                     Therm|||||||   |||  |
                Therm                Therm |||||||   |||  |
               Therm|                Therm |  ||||||||  |||  |
              Therm ||       Therm   Therm| |  |||||||||| |||  |
      Therm   Therm  | ||   Therm|   Therm|| |  |||||||||| |||  |
     Therm |  Therm|   |||   Therm|   Therm  |||  ||||||||| |||  |
    Therm || Therm   ||  | || Therm | ||  Therm    |  |||  ||||||||| ||| |
   Therm ||  Therm    ||  | || Therm | ||  Therm|    |  ||| |||||||||  ||| |
```
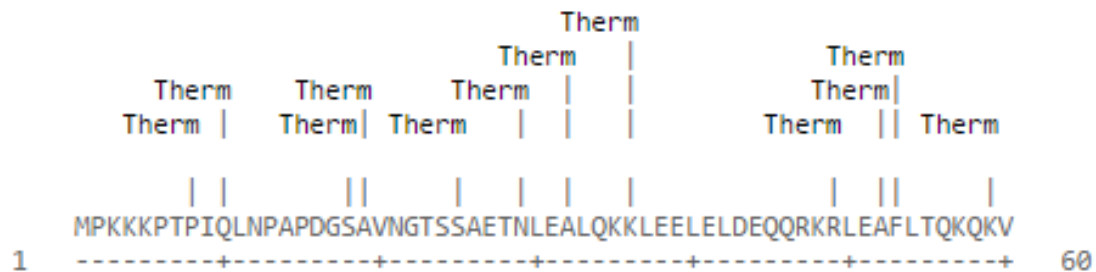```
    | | | | || | || | | || || | | ||| | ||||||||| ||| |
   ACISGFQGMNLPTESGELPLQGDVFIRQYFTVFHRENNQVGLAPVA   AELGQPGNLLGNDHIYNVIVTAHAFVIIFFIVIPIIIGGFGN
   1 --------+---------+---------+---------+------ 46   1 --------+---------+---------+---------+-- 42
```

C

D

```
            Therm   Therm
           Therm | Therm |
           Therm| | Therm   Therm | |
          Therm|| | Therm| Therm| | |
         Therm||| | Therm   Therm || Therm | | |
                              Therm   Therm | || Therm | | |
                        Therm | Therm| | || Therm   Therm | | || | |
```
```
           |||| | | | | | || | || | | | || | |
          LNLVIVQVSSGPGVSSAPA   AECWRECEWVCAGGHGGAVCKIGVANHRTRAWSGYPPPTQRGRASPHTLTAEFALGRVKK
         1 --------+--------- 19   1 --------+---------+---------+---------+---------+---------+ 60
```
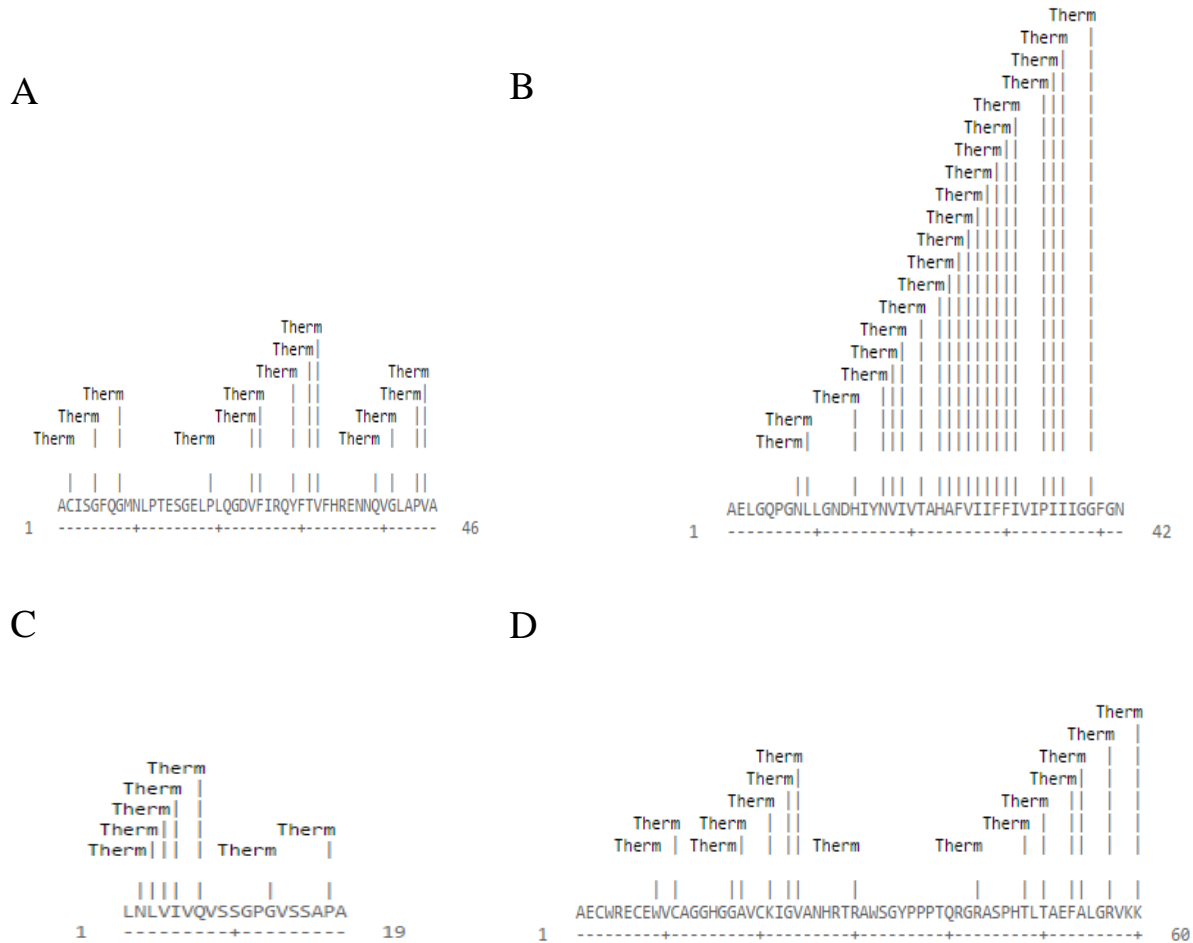
**Figure 31.** Predicted thermolysin cleavage sites for T7PD candidates using PeptideCutter from Expasy. From A-D candidates PAP46, cytochrome *c* oxidase, GZNFP, and ORF96. The thermolysin cleavage sites are marked with lines. The image represents the putative cleavage sites for the anthrax LF on the isolated T7PD candidates.

The predicted 3D structures for PAP46 and cytochrome *c* oxidase, chosen by their affinity to LF and presence of cleavage sites for thermolysin, demonstrated two different protein arrangements as binders to LF (figs. 32 & 33). The PAP46 consisted mostly of a strand structure while the cytochrome was an almost a complete α-helix peptide.
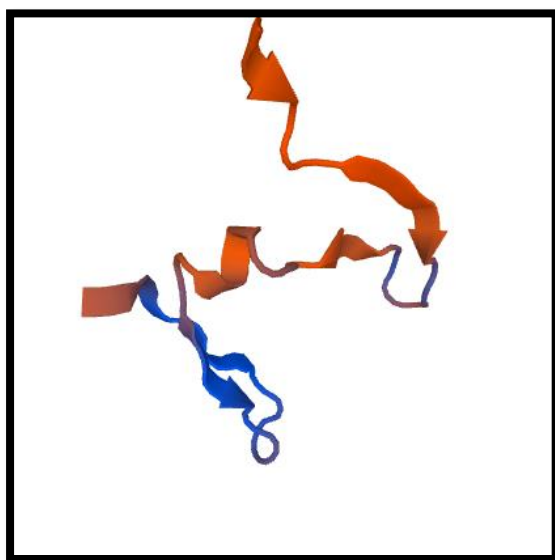


**Figure 32.** Predicted 3D structure for PAP46, obtained using SWISS-MODEL. The peptide consist mostly of strand structures or single chain with a small helicoidally region.
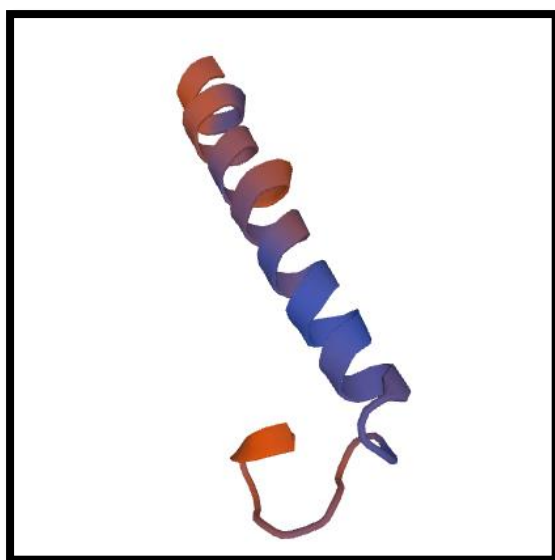


**Figure 33.** Predicted 3D structure for cytochrome *c* oxidase subunit I, obtained using SWISS-MODEL. The peptide consist mostly of an alpha-helix structure.

# 5.4 DISCUSSION OF RESULTS

**5.4.1 Multiple alignment performed to the identified T7PD candidates and biochemical properties for interactions**

A multiple alignment for all the LF-interacting partners isolated through T7PD gave a baseline to identify which peptides had more affinity to LF *in vitro* (fig. 22). The putative conserved domains obtained could represent a step to begin understanding GI anthrax. The detection of common motifs in proteins having affinity to LF-WT and LF-MT, such as the gastric triacylglycerol lipase and pepsin A, at independent events (figs. 23, 24 & 25), represent the most valuable candidates. This data suggest that both sequences could have interactions with the LF at least *in vitro*. For the pepsin A (PAP) (a member of the peptidase A1 protein family) the minimum conserved LF-interacting sequence was of 30 residues that matches with the aspartyl protease domain (fig. 24). Several members from different peptidase families have been proposed as anthrax LF inhibitors, resulting in the survival of macrophage cells due to an interference of LF protease activity (Menard et al., 1996). The pepsins are well known to have a preference for hydrophobic ligands and some of these may serve as inhibitors, such as the pepstatin (Kuzmič et al., 1991). Considering this, the interactions between the hydrophobic residues shown in figure 26 and the hydrophobic amino acids from domain III proposed by Pannifer et al. (2001) could explain and support our interaction data. The secondary structure predictions suggest that pepsin peptides (fig. 28 & 32) are composed of strand structures and this could help to interact with the groove present on the fourth domain of LF.

On the other hand, the largest peptide for gastric lipase (GL) was 77 residues (table 4, p. 47); nevertheless, the putative minimum interaction sequence with LF was of 58 amino acids that did not contain the alpha/beta hydrolase fold (fig 25). Likewise, GLs has been proposed to be

associated to *Helicobacter pylori* secreted proteins, resulting as an alternative for treatment of gastrointestinal diseases (Park et al., 2013). Some lipases are known to be secreted when infectious agents are present and this could explain the presence of the GL in our findings. The lipases contain two lysine residues on the C-terminal (figs, 25, 27 & 29), which are positively charged, suggesting a possible recognition and interaction site by the negatively charged residues of the LF domain III proposed by Dalkas et al., 2009. Despite the constant detection of this interacting partner during the biopanning cycles, the interaction to the LF tested is still currently in doubt because of the lack of specificity of this candidate.

A study was undertaken to determine the needed amino acids for being a LF substrate and then using databases resulted to have similarity with NADH dehydrogenase subunit 1; however, posterior hydrolysis of this protein was not detected (Li et al., 2011). For this reason, it was important to use other properties to understand the interactions, such as the hydrophobicity scale of the candidates (data not shown), which resulted in mostly negatively charged residues. Based on Dalkas et al. (2009) the stabilization of the LF-substrate (MKKs) interaction was determined depending on the residues, the protease active site is acidic; for this reason, interactions with basic residues such as the N-terminal of MKKs is possible. Considering that PAP46 (tables 3 & 4, pp. 42 & 47) consisted mostly of hydrophobic residues, this peptide could be possibly interacting with the aspartic acid clusters present in the active site groove described before. This interaction is mediated by a water molecule. Indeed, it is also difficult to attribute functional activity to the detected domains, present on candidates, because a few cases were only fragments of this essential part of proteins. On the other hand, chromosome 11 open reading frame 96 (ORF96) (table 3, p. 45) contained 12 basic residues that could be interacting directly with the negatively charged active site of LF.

As cited before, metalloproteases such as LF and thermolysin share a motif (HEXXH) necessary to bind zinc on their active site to allow the peptidase activity. For this reason, an *in silico* cleavage was performed to the candidates that were found to interact LF on section 4.3.2. Cytochrome *c* oxidase (table 4, p. 48), which was a LF-WT interacting partner, was found as the candidate we more cleavage sites for thermolysin peptidase (fig. 31). Since the previously mentioned data was predicted *in silico*, a possible cleavage of this protein in nature could be a possibility during GI anthrax disease. However, because this is only a fraction of the whole protein it must be considered that the rest of the protein chemical properties could protect this region from LF. On the other hand, PAP46 showed 13 sites for cleavage, but this result is still unclear until further studies are undertaken, at least *in vitro*. Establishing a prediction of the 3D structure of the candidates was necessary to know their physical properties (figs. 32 & 33). Moreover, the versatility of these peptides could explain why in tables 3 & 4 (sec. 3.3.1) the MKKs were not found as LF binding partners. Possibly, a competitive inhibition between the isolated T7PD partners and the MKKs was another reason for not detecting these proteins.

Nevertheless, taking advantage of the binding properties of the peptide PAP46 by performing T7PD, it could serve as a potential inhibitor to the anthrax LF. In short, further studies *in vitro* and especially *in vivo* are necessary to fully understand the role of the profile of proteins exposed in this research. These studies could help to elucidate better the GI anthrax molecular pathogenesis.

# CHAPTER 6

## CONCLUSIONS, RECOMMENDATIONS, AND

## CITED LITERATURE

# 6.1 SUMMARY AND CONCLUSIONS

✓ T7PD demonstrated to be a powerful tool for study protein-protein interactions, in this case peptides from human stomach and the anthrax toxin component, LF.

✓ A more diverse group of peptides interacted with LF-MT than LF-WT, at sequencing, which can be due to interaction with the active site in its nature state. However, on posterior tests, most of the confirmed binders were specific to LF-WT.

✓ After the *in situ* specificity test, worked by inoculation of R5615 on the microplate's well, it was determined that a small fraction of phages remains on each well, no matter if experimental or controls.

✓ The second specificity test resulted in 10 peptides that interact with at least one type of LF tested. From these, PAP46, GZNFP, and ORF96 interacted with both types of LF. In addition, in this test it was found non-specific binders to LF such as PAP156 and gastric lipase.

✓ For PAP46 minimum concentration of both types of LF necessary for interaction resulted in 1µg/mL.

✓ The *in silico* tools suggested conserved amino acid sequences for pepsin A and gastric lipase. Also, it was found that proteins such as PAP46, cytochrome *c* oxidase, and GZNFP, contain cleavage sites for thermolysin, a metalloprotease that shares a motif sequence for the active site with LF. The provided 3D structures could serve to better understanding the interactions.

✓ A total of 10 LF-interacting partners were confirmed by performing T7PD.

## 6.2 RECOMMENDATIONS

✓ Further screenings are necessary of the isolated candidates identified in this work to determine for specificity properties of the peptides.

✓ Several screenings consisting of different biopanning cycles are needed to determine which peptide binds strongly to the target protein in study.

✓ It is strongly recommended the alternation of blocking agents through the rounds of biopannings at an individual cycle. This will avoid the enrichment of peptides that binds casein or BSA.

✓ Changes in stringency could help to differentiate between weak and strong interactions.

✓ Once the candidates are sequenced or before, a specificity test of these should be performed in order to determine whether the peptides bind the blocking agent or the target protein.

✓ At this test, if lysis is detected on all the treatments, a change in the blocking agent will help to understand if the isolated binders are non-specific to the target.

✓ Increasing the Tween 20 percent helps to differentiate between these non-specific binders. This allows us to identify to which clone expresses more affinity to the target.

✓ *In vivo* models, using lab rats, could help to determine if these peptides contain inhibition properties against the anthrax LF or if novel interactions were discovered in this work.

# 6.3 CITED LITERATURE

Abrami, L., Bischofberger, M., Kunz, B., Groux, R., & van der Goot, F. G. (2010). Endocytosis of the anthrax toxin is mediated by clathrin, actin and unconventional adaptors. PLoS pathogens, 6(3), e1000792.

Abreu, M. T., Fukata, M., & Arditi, M. (2005). TLR signaling in the gut in health and disease. *The Journal of Immunology, 174*(8), 4453-4460.

Adams, J. (2002). Development of the proteasome inhibitor PS-341. *The oncologist, 7*(1), 9-16.

Agrawal, A., Lingappa, J., Leppla, S. H., Agrawal, S., Jabbar, A., Quinn, C., & Pulendran, B. (2003). Impairment of dendritic cells and adaptive immunity by anthrax lethal toxin. *Nature*, *424*(6946), 329-334.

Akbulut, A., Akbulut, H., Özgüler, M., İnci, N., & Yalçın, Ş. (2012). Gastrointestinal Anthrax: A Case and Review of Literature. *Advances in Infectious Diseases*, *2*, 67.

Antoine, M. D., Hagan, N. A., Lin, J. S., Feldman, A. B., & Demirev, P. A. (2012). Rapid detection of ribosome inactivating protein toxins by mass-spectrometry-based functional assays. *International Journal of Mass Spectrometry, 312*, 41-44.

Asa, P. B., Wilson, R. B., & Garry, R. F. (2002). Antibodies to squalene in recipients of anthrax vaccine. *Experimental and molecular pathology*, *73*(1), 19-27.

Baldari CT, Tonello F, Paccani SR, Montecucco C (2006) Anthrax toxins: A paradigm of bacterial immune suppression. *Trends in immunology 27*: 434–440. doi: 10.1016/j.it.2006.07.002.

Bardwell, A., Abdollahi, M., & Bardwell, L. (2004). Anthrax lethal factor-cleavage products of MAPK (mitogen-activated protein kinase) kinases exhibit reduced binding to their cognate MAPKs. *Biochem. J*, *378*, 569-577.

Bardwell, A. J., Frankson, E., & Bardwell, L. (2009). Selectivity of docking sites in MAPK kinases. *Journal of biological chemistry*, *284*(19), 13165-13173.

Basha, S., Rai, P., Poon, V., Saraph, A., Gujraty, K., Go, M. Y., Sadacharan, S., Frost, M., Mogridge, J., & Kane, R. S. (2006). Polyvalent inhibitors of anthrax toxin that target host receptors.*Proceedings of the National Academy of Sciences*, *103*(36), 13509-13513.

Belay, E. D., & Monroe, S. S. (2014). Low-incidence, high-consequence pathogens. *Emerging Infectious Diseases*, *20*(2), 319.

Berger, S. (2014). *Anthrax: Global Status*. GIDEON Informatics Inc.

Bergman, N. H., Passalacqua, K. D., Gaspard, R., Shetron-Rama, L. M., Quackenbush, J., & Hanna, P. C. (2005). Murine macrophage transcriptional responses to *Bacillus anthracis* infection and intoxication. *Infection and immunity, 73*(2), 1069-1080.

Bishop, B. L., Lodolce, J. P., Kolodziej, L. E., Boone, D. L., & Tang, W. J. (2010). The role of anthrolysin O in gut epithelial barrier disruption during *Bacillus anthracis* infection. Biochemical and biophysical research communications, 394(2), 254-259.

Bouzianas, D. G. (2009). Medical countermeasures to protect humans from anthrax bioterrorism. *Trends in microbiology*, *17*(11), 522-528.

Boyden, E. D., & Dietrich, W. F. (2006). Nalp1b controls mouse macrophage susceptibility to anthrax lethal toxin. *Nature genetics*, *38*(2), 240-244.

Brachman, P. S. (1980). Inhalation anthrax. *Annals of the New York Academy of Sciences*, *353*(1), 83-93.

Bradley, K. A., J. Mogridge, M. Mourez, R. J. Collier, and J. A. Young. (2001). Identification of the cellular receptor for anthrax toxin. *Nature 414***:**225-229.

Brissette, R., Prendergast, J. K., & Goldstein, N. I. (2006). Identification of cancer targets and therapeutics using phage display. *Current Opinion in Drug Discovery and Development*, *9*(3), 363.

Brown, J., Wang, H., Hajishengallis, G. N., & Martin, M. (2011). TLR-signaling Networks An Integration of Adaptor Molecules, Kinases, and Cross-talk. *Journal of dental research, 90*(4), 417-427.

Brown, R. S., Sander, C., & Argos, P. (1985). The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS letters, 186*(2), 271-274.

Brunori, M., Antonini, G., Malatesta, F., Sarti, P., & Wilson, M. T. (1987). Cytochrome-*c* oxidase. *European Journal of Biochemistry, 169*(1), 1-8.

Burbaum, J., & Tobal, G. M. (2002). Proteomics in drug discovery. *Current opinion in chemical biology*, *6*(4), 427-433.

Burgos, R. (2010). *Isolation of Interacting Peptides to Bacillus anthracis lethal toxin (LF) by T7 Phage Display*. (Master's thesis). Retrieved from UPRM Dissertations and Theses.

Castel, G., Chtéoui, M., Heyd, B., & Tordo, N. (2011). Phage display of combinatorial peptide libraries: application to antiviral research. *Molecules*, *16*(5), 3499-3518.

Centers for Disease Control and Prevention. (2001). Human anthrax associated with an epizootic among livestock—North Dakota, 2000. MMWR Morb. Mortal. Wkly. Rep. 50:677–680.

Chamaillard, M., Girardin, S. E., Viala, J., & Philpott, D. J. (2003). Nods, Nalps and Naip: intracellular regulators of bacterial-induced inflammation. *Cellular microbiology*, *5*(9), 581-592.

Chien, C. T., Bartel, P. L., Sternglanz, R., & Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proceedings of the National Academy of Sciences*, *88*(21), 9578-9582.

Chopra, A. P., Boone, S. A., Liang, X., & Duesbery, N. S. (2003). Anthrax lethal factor proteolysis and inactivation of MAPK kinase. *Journal of Biological Chemistry*, *278*(11), 9402-9406.

Chun, J. H., Hong, K. J., Cha, S. H., Cho, M. H., Lee, K. J., Jeong, D. H., Yoo, C. K., & Rhie, G. E. (2012). Complete Genome Sequence of Bacillus anthracis H9401, an Isolate from a Korean Patient with Anthrax. *Journal of bacteriology*, *194*(15), 4116-4117.

Chung, M. C., Narayanan, A., Popova, T. G., Kashanchi, F., Bailey, C. L., & Popov, S. G. (2013). *Bacillus anthracis* derived nitric oxide induces protein S-nitrosylation contributing to macrophage death. *Biochemical and biophysical research communications, 430(1),* 125-130.

Clackson, T. & Wells, J.A. (1994). *In vitro* selection from protein and peptide libraries. *Trends Biotechnol.* 12, 173-184.

Colland, F., Jacq, X., Trouplin, V., Mougin, C., Groizeleau, C., Hamburger, A., Meil, A., Wojcik, J., Legrain, P., & Gauthier, J. M. (2004). Functional proteomics mapping of a human signaling pathway. *Genome Research*, *14*(7), 1324-1332.

Collas, P., Le Guellec, K., & Taskén, K. (1999). The A-kinase–anchoring protein AKAP95 is a multivalent protein with a key role in chromatin condensation at mitosis. *The Journal of cell biology, 147(6),* 1167-1180.

Dalkas, G. A., Papakyriakou, A., Vlamis-Gardikas, A., & Spyroulias, G. A. (2009). Insights into the anthrax lethal factor–substrate interaction and selectivity using docking and molecular dynamics simulations. *Protein Science*, *18*(8), 1774-1785.

Danner, S., & Belasco, J. G. (2001). T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries.*Proceedings of the National Academy of Sciences*, *98*(23), 12954-12959.

Dixon Tc, Meselson M, Guillemin J, Hanna Pc (1999) Anthrax. N Engl J Med 341: 815-826.

Dinarello, C. A. (2011). Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood*, *117*(14), 3720-3732.

Duesbery, N. S., Webb, C. P., Leppla, S. H., Gordon, V. M., Klimpel, K. R., Copeland, T. D., Ahn, N. G., Oskarsson, M. K., Fukasawa, K., Paull, K. D., & Woude, G. F. V. (1998). Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science, 280*(5364), 734-737.

Ebrahimi, C. M., Sheen, T. R., Renken, C. W., Gottlieb, R. A., & Doran, K. S. (2011). Contribution of lethal toxin and edema toxin to the pathogenesis of anthrax meningitis. *Infection and immunity*, *79*(7), 2510-2518.

Fowler, R. A., & Shafazand, S. (2011). Anthrax Bioterrorism: Prevention, Diagnosis and Management Strategies. J Bioterr Biodef 2:107. doi:10.4172/2157- 2526.1000107.

Franchi, L., Park, J. H., Shaw, M. H., Marina-Garcia, N., Chen, G., Kim, Y. G., & Núñez, G. (2008). Intracellular NOD-like receptors in innate immunity, infection and disease. *Cellular microbiology*, *10*(1), 1-8.

Fukunaga, K., & Taki, M. (2012). Practical tips for construction of custom peptide libraries and affinity selection by using commercially available phage display cloning systems. *Journal of nucleic acids*, *2012*.

Gao, X., Chaturvedi, D., & Patel, T. B. (2012). Localization and retention of p90 ribosomal S6 kinase 1 in the nucleus: implications for its function. *Molecular biology of the cell, 23(3),* 503-515.

Garrington, T. P., & Johnson, G. L. (1999). Organization and regulation of mitogen-activated protein kinase signaling pathways. *Current opinion in cell biology*, *11*(2), 211-218.

Georgieva, Y., & Konthur, Z. (2011). Design and screening of M13 phage display cDNA libraries. *Molecules*, *16*(2), 1667-1681.

Ghayur, T., Banerjee, S., Hugunin, M., Butler, D., Herzog, L., Carter, A., Quintal, L., Sekut, L., Talanian, R., Paskind, M., Wong, W., Kamen, R., Tracey, D., & Alien, H. (1997). Caspase-1 processes IFN-γ-inducing factor and regulates LPS-induced IFN-γ production. *Nature 386*, 619 - 623

Gladue, D. P., Baker-Bransetter, R., Holinka, L. G., Fernandez-Sainz, I. J., O'Donnell, V., Fletcher, P., Lu, Z., & Borca, M. V. (2014). Interaction of CSFV E2 protein with swine host factors as detected by yeast two-hybrid system. *PloS one*, *9*(1), e85324.

Glomski, I. J., Piris-Gimenez, A., Huerre, M., Mock, M., & Goossens, P. L. (2007). Primary involvement of pharynx and Peyer's patch in inhalational and intestinal anthrax. *PLoS pathogens, 3*(6), e76

Gnanasekar, M., Suleman, F. G., Ramaswamy, K., & Caldwell, J. D. (2009). Identification of sex hormone binding globulin-interacting proteins in the brain using phage display screening. *International journal of molecular medicine*,*24*(4), 421-426.

Goodswen, S. J., Kennedy, P. J., & Ellis, J. T. (2014). Vacceed: a high-throughput in silico vaccine candidate discovery pipeline for eukaryotic pathogens based on reverse vaccinology. *Bioinformatics*, *30*(16), 2381-2383.

Griffith, J., Blaney, D., Shadomy, S., Lehman, M., Pesik, N., Tostenson, S., Delaney, L., Tiller, R., DeVries, A., Gomez, T., Sullivan, M., Blackmore, C., Stanek, D., & Lynfield, R. (2014). Investigation of Inhalation Anthrax Case, United States. *Emerging infectious diseases*, *20*(2), 280.

Grunow, R., Klee, S., Beyer, W., George, M., Grunow, D., Barduhn, A., Klar, S., Jacob, D., Elschner, M., Sandven, P., Kjerulf, A., Jensen, J. S., Cai, W., Zimmermann., R., & Schaade, L. (2013). Anthrax among heroin users in Europe possibly caused by same Bacillus anthracis strain since 2000.

Harada, T., Iwai, A., & Miyazaki, T. (2010). Identification of DELE, a novel DAP3-binding protein which is crucial for death receptor-mediated apoptosis induction. *Apoptosis, 15*(10), 1247-1255.

Hellmich, K. A., Levinsohn, J. L., Fattah, R., Newman, Z. L., Maier, N., Sastalla, I., Liu, S., Leppla, S. H., & Moayeri, M. (2012). Anthrax lethal factor cleaves mouse nlrp1b in both toxin-sensitive and toxin-resistant macrophages. *PLoS One, 7*(11), e49741.

Hicks, C. W., Sweeney, D. A., Cui, X., Li, Y., & Eichacker, P. Q. (2012). An overview of anthrax infection including the recently identified form of disease in injection drug users. *Intensive care medicine*, *38*(7), 1092-1104.

Houghten, R. A. (2000). Parallel array and mixture-based synthetic combinatorial chemistry: tools for the next millennium. *Annual review of pharmacology and toxicology*, *40*(1), 273-282.

"Human Microbiome Project." *Human Microbiome Project*. Web. 12 Feb. (2013). <http://commonfund.nih.gov/hmp/>.

Inglesby, T. V., O'Toole T., Henderson, D. A., Bartlett, J. G., Ascher, M. S., Eitzen, E., Friedlander, A. M., Gerberding, J., Hauer, J., Hughes, J., McDade, J., Osterholm, M. T., Parker, G., Perl, T. M., Russell, P. K., & Tonat, K. (2002). Working Group on Civilian *Biodefense JAMA. 1; 287*(17):2236-52.

Inohara, N., Ogura, Y., Chen, F. F., Muto, A., & Nuñez, G. (2001). Human Nod1 confers responsiveness to bacterial lipopolysaccharides. *Journal of Biological Chemistry*, *276*(4), 2551-2554.

Inohara, N., Koseki, T., del Peso, L., Hu, Y., Yee, C., Chen, S., Carrio, R., Merino, J., Liu, D., Ni, J., & Núñez, G. (1999). Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-κB. *Journal of Biological Chemistry*, *274*(21), 14560-14567.

Jacobs, J. W., & Fodor, S. (1994). Combinatorial chemistry—applications of light-directed chemical synthesis. *Trends in biotechnology*, *12*(1), 19-26.

Jahns, A. C., & Rehm, B. H. (2012). Relevant uses of surface proteins–display on self-organized biological structures. *Microbial biotechnology*, *5*(2), 188-202.

Jang, K. H., Nam, S. J., Locke, J. B., Kauffman, C. A., Beatty, D. S., Paul, L. A., & Fenical, W. (2013). Anthracimycin, a Potent Anthrax Antibiotic from a Marine-Derived Actinomycete. *Angewandte Chemie International Edition*,*52*(30), 7822-7824.
.
Jernigan, D. B., Raghunathan, P. L., Bell, B. P., Brechner, R., Bresnitz, E. A., Butler, J. C., ... & Gerberding, J. L. (2002). Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. *Emerging infectious diseases*, *8*(10), 1019.

Jennings-Antipov, L. D., Song, L., & Collier, R. J. (2011). Interactions of anthrax lethal factor with protective antigen defined by site-directed spin labeling.*Proceedings of the National Academy of Sciences*, *108*(5), 1868-1873.

Jongeneel, C. V., Bouvier, J., & Bairoch, A. (1989). A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS letters*, *242*(2), 211-214.

Kamal, S. M., Rashid, A. K. M., Bakar, M. A., & Ahad, M. A. (2011). Anthrax: an update. *Asian Pacific journal of tropical biomedicine*, *1*(6), 496-501.

Kandadi, M. R., Hua, Y., Ma, H., Li, Q., Kuo, S. R., Frankel, A. E., & Ren, J. (2010). Anthrax lethal toxin suppresses murine cardiomyocyte contractile function and intracellular Ca2+ handling via a NADPH oxidase-dependent mechanism. *PloS one, 5*(10), e13335.

Kaufmann, A. F., Meltzer, M. I., & Schmid, G. P. (1997). The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable?. *Emerging infectious diseases*, *3*(2), 83.

Kay, B.K., Winter, J., & McCafferty, J. (1996). Phage display of peptides and proteins. In *A Laboratory Manual*, Academic Press, San Diego.

Klimpel, K. R., Arora, N., & Leppla, S. H. (1994). Anthrax toxin lethal factor contains a zinc metalloprotease consensus sequence which is required for lethal toxin activity. *Molecular microbiology*, *13*(6), 1093-1100.

Kracalik, I., Malania, L., Tsertsvadze, N., Manvelyan, J., Bakanidze, L., Imnadze, P., Tsanava, S., & Blackburn, J. K. (2014). Human Cutaneous Anthrax, Georgia 2010–2012. *Emerging infectious diseases*, *20*(2), 261.

Krebs, J., Goldstein, E., & Kilpatrick, S. (2010). Lewin's Genes X. print. Jones and Bartlett Publishers. Sudbury, Massachusetts (Boston).

Krumpe, L. R. H. (2006). *Diversity comparison of T7 and M13 phage-displayed peptide libraries.* (Order No. 1439197, Hood College). *ProQuest Dissertations and Theses, ,* 82-82 p. Retrieved from http://search.proquest.com/docview/304918491?accountid=28498. (304918491).

Kuzmič, P., Sun, C. Q., Zhao, Z. C., & Rich, D. H. (1991). Long range electrostatic effects in pepsin catalysis. *Tetrahedron*, *47*(14), 2519-2534.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947-2948.

Legg, J. A., & Machesky, L. M. (2004). MRL proteins: leading Ena/VASP to Ras GTPases. *Nature cell biology, 6*(11), 1015-1017.

Leppla, S. H. (1982). Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells. *Proceedings of the National Academy of Sciences*, *79*(10), 3162-3166.

Leppla, S. H. (1984). *Bacillus anthracis* calmodulin-dependent adenylate cyclase: Chemical and enzymatic properties and interactions with eucaryotic cells. Adv Cyclic Nucleotide Protein Phosphorylation Res 17:189–198.

Levinsohn, J. L., Newman, Z. L., Hellmich, K. A., Fattah, R., Getz, M. A., Liu, S., Sastalla, I., Leppla, S. H., & Moayeri, M. (2012). Anthrax lethal factor cleavage of Nlrp1 is required for activation of the inflammasome. *PLoS pathogens*, *8*(3), e1002638.

Li, F., Terzyan, S., & Tang, J. (2011). Subsite specificity of anthrax lethal factor and its implications for inhibitor development. *Biochemical and biophysical research communications*, *407*(2), 400-405.

Lightfoot, Y. L., Yang, T., Sahay, B., Zadeh, M., Cheng, S. X., Wang, G. P., Owen, J. L., & Mohamadzadeh, M. (2014). Colonic Immune Suppression, Barrier Dysfunction, and Dysbiosis by Gastrointestinal *Bacillus anthracis* Infection. *PloS one, 9*(6), e100532.

Linkov, I., Coles, J. B., Welle, P., Bates, M., & Keisler, J. (2011). Anthrax cleanup decisions: Statistical confidence or confident response. *Environmental science & technology*, *45*(22), 9471-9472.

Liu, S., Moayeri, M., & Leppla, S. H. (2014). Anthrax lethal and edema toxins in anthrax pathogenesis. Trends in microbiology.

Lopez, C. D., Martinovsky, G., & Naumovski, L. (2002). Inhibition of cell death by ribosomal protein L35a. *Cancer letters, 180*(2), 195-202.

Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., Green, D. V. S., Hertzberg, R. P., Janzen, W. P., Paslay, J. W., Schopfer, U., & Sittampalam, G. S. (2011). Impact of high-throughput screening in biomedical research. *Nature reviews Drug discovery*, *10*(3), 188-195.

Mao, W., Chen, J., Peng, T. L., Yin, X. F., Chen, L. Z., & Chen, M. H. (2012). Downregulation of gastrokine-1 in gastric cancer tissues and restoration of its expression induced gastric cancer cells to apoptosis. *J Exp Clin Cancer Res, 31*, 49.

Martin, T. E., Powell, C. T., Wang, Z., Bhattacharyya, S., Walsh-Reitz, M. M., Agarwal, K., & Toback, F. G. (2003). A novel mitogenic protein that is highly expressed in cells of the gastric antrum mucosa. *American Journal of Physiology-Gastrointestinal and Liver Physiology, 285*(2), G332-G343.

Mattheakis, L. C., Bhatt, R. R., & Dower, W. J. (1994). An *in vitro* polysome display system for identifying ligands from very large peptide libraries.*Proceedings of the National Academy of Sciences*, *91*(19), 9022-9026.

Menard, A., Papini, E., Mock, M., & Montecucco, C. (1996). The cytotoxic activity of *Bacillus anthracis* lethal factor is inhibited by leukotriene A4 hydrolase and metallopeptidase inhibitors. *Biochem. J, 320*, 687-691.

Miao, E. A., Rajan, J. V., & Aderem, A. (2011). Caspase-1-induced pyroptotic cell death. *Immunological reviews*, *243*(1), 206-214.

Moayeri, M., Sastalla, I., & Leppla, S. H. (2012). Anthrax and the inflammasome. *Microbes and Infection, 14*(5), 392-400.

Mourez, M., Kane, R. S., Mogridge, J., Metallo, S., Deschatelets, P., Sellman, B. R., Whitesides, G. M., & Collier, R. J. (2001). Designing a polyvalent inhibitor of anthrax toxin. *nature biotechnology*, *19*(10), 958-961.

Murthy, S., Keystone, J., & Kissoon, N. (2013). Infections of the Developing World. *Critical care clinics*, *29*(3), 485-507.

Nardone, G., Rippa, E., Martin, G., Rocco, A., Siciliano, R. A., Fiengo, A., Cacace, G., Malorni, A., Budillon, G., & Arcari, P. (2007). Gastrokine 1 expression in patients with and without *Helicobacter pylori* infection. *Digestive and liver disease, 39*(2), 122-129.

Palmgren, M. G., & Nissen, P. (2011). P-type ATPases. Annual review of biophysics, 40, 243-266.

Palsson, B. (2000). The challenges of in silico biology. *Nature biotechnology*,*18*(11), 1147-1150.

Pandeya, S. N., & Thakkar, D. (2005). Combinatorial chemistry: A novel method in drug discovery and its application. *Indian Journal of Chemistry*, *44*, 335-348.

Pannifer, A. D., Wong, T. Y., Schwarzenbacher, R., Renatus, M., Petosa, C., Bienkowska, J., Lacy, B., Collier, R. J., Park, S., Leppla, S. H., Hanna, P., & Liddington, R. C. (2001). Crystal structure of the anthrax lethal factor. *Nature*, *414*(6860), 229-233.

Park, J. M., Greten, F. R., Li, Z. W., & Karin, M. (2002). Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science, 297*(5589), 2048-2051.

Park, J. S., Lee, S. J., Kim, T. H., Yeom, J., Park, E. S., Seo, J. H., Jun, J. S., Lim, J. Y., Park, C. H., Woo, H. O., Youn, H. S., Ko, G. H., Kang, H. L., Baik, S. H., Lee, W. K., Cho, M. J., & Rhee, K. H. (2013). Gastric Autoantigenic Proteins in *Helicobacter Pylori* Infection. *Yonsei medical journal, 54*(6), 1342-1352.

Paschke, M. (2006). Phage display systems and their applications. *Applied microbiology and biotechnology*, *70*(1), 2-11.

Pearson, G., Robinson, F., Beers Gibson, T., Xu, B. E., Karandikar, M., Berman, K., & Cobb, M. H. (2001). Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions 1. *Endocrine reviews*, *22*(2), 153-183.

Popova, T. G., Espina, V., Zhou, W., Mueller, C., Liotta, L., & Popov, S. G. (2014). Whole Proteome Analysis of Mouse Lymph Nodes in Cutaneous Anthrax. *PloS one*, *9*(10), e110873.

Rahman, M. A., Noore, M. S., Hasan, M. A., Ullah, M. R., Rahman, M. H., Hossain, M. A., Ali, Y., & Islam, M. S. (2014). Identification of potential drug targets by subtractive genome analysis of *Bacillus anthracis* A0248: An *in silico* approach. *Computational biology and chemistry*, *52*, 66-72.

Rippa, E., La Monica, G., Allocca, R., Romano, M. F., De Palma, M., & Arcari, P. (2011). Overexpression of gastrokine 1 in gastric cancer cells induces Fas-mediated apoptosis. *Journal of cellular physiology, 226*(10), 2571-2578.

Rosenberg, A., Griffin, K., Studier, F. W., McCormick, M., Berg, J., Novy, R., & Mierendorf, R. (1996). T7Select® Phage Display System: A powerful new protein display system based on bacteriophage T7. *NEWSLETTER*.

Sche, P. P., Mckenzie, K. M., White, J. D., & Austin, D. J. (1999). Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chemistry & biology*, *6*(10), 707-716.

Schillace, R. V., Miller, C. L., Pisenti, N., Grotzke, J. E., Swarbrick, G. M., Lewinsohn, D. M., & Carr, D. W. (2009). A-kinase anchoring in dendritic cells is required for antigen presentation. *PloS one, 4*(3), e4807.

Schreiber, G., & Keating, A. E. (2011). Protein binding specificity versus promiscuity. *Current opinion in structural biology*, *21*(1), 50-61.

Scobie, H. M., G. J. Rainey, K. A. Bradley, and J. A. Young. (2003). Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor. Proc. Natl. Acad. Sci. USA 100**:**5170-5174.

Sidhu, S. S. (2001). Engineering M13 for phage display. *Biomolecular engineering*, *18*(2), 57-63.

Simossis, V. A., & Heringa, J. (2005). PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Research*, *33*(suppl 2), W289-W294.

Sirisanthana, T., & Brown, A. E. (2002). Anthrax of the Gastrointestinal Tract. Emerg Infect Dis. Available from http://wwwnc.cdc.gov/eid/article/8/7/02-0062.htm

Smith, G.P. (1985). Fimentous fusion phage: novel expression vectors that display cloned antigens on the viron surface. *Science* 228, 1315-1317.

Smith, G. P., & Petrenko, V. A. (1997). Phage display. *Chemical reviews*,*97*(2), 391-410.

Snider, J., Kittanakom, S., Curak, J., & Stagljar, I. (2010). Split-ubiquitin based membrane yeast two-hybrid (MYTH) system: a powerful tool for identifying protein-protein interactions. *Journal of visualized experiments: JoVE*, (36).

Sun, C., Fang, H., Xie, T., Auth, R. D., Patel, N., Murray, P. R., Snoy, P. J., & Frucht, D. M. (2012). Anthrax lethal toxin disrupts intestinal barrier function and causes systemic infections with enteric bacteria. *PloS one*, *7*(3), e33583.

Sweeney, D. A., Cui, X., Solomon, S. B., Vitberg, D. A., Migone, T. S., Scher, D., Danner, R. L., Natanson, C., Mani-Subramanian, G., & Eichacker, P. Q. (2010). Anthrax lethal and edema toxins produce different patterns of cardiovascular and renal dysfunction and synergistically decrease survival in canines. *Journal of Infectious Diseases*, *202*(12), 1885-1896.

Tang, G., & Leppla, S. H. (1999). Proteasome activity is required for anthrax lethal toxin to kill macrophages. *Infection and immunity*, *67*(6), 3055-3060.

Tebbutt, N. C., Giraud, A. S., Inglese, M., Jenkins, B., Waring, P., Clay, F. J., Malki, S., Alderman, B. M., Grail, D., Hollande, F., Heath, J. K., & Ernst, M. (2002). Reciprocal regulation of gastrointestinal homeostasis by SHP2 and STAT-mediated trefoil gene activation in gp130 mutant mice. *Nature medicine, 8*(10), 1089-1097.

Tonry, J. H., Popov, S. G., Narayanan, A., Kashanchi, F., Hakami, R. M., Carpenter, C., Bailey, C., & Chung, M. C. (2013). *In vivo* murine and *in vitro* M-like cell models of gastrointestinal anthrax. *Microbes and Infection, 15*(1), 37-44.

Tournier, J. N., Rossi Paccani, S., Quesnel-Hellmann, A., & Baldari, C. T. (2009). Anthrax toxins: a weapon to systematically dismantle the host immune defenses. *Molecular aspects of medicine*, *30*(6), 456-466.

Tournier, J. N., Quesnel-Hellmann, A., Mathieu, J., Montecucco, C., Tang, W. J., Mock, M., Vidal, D. R., & Goossens, P. L. (2005). Anthrax edema toxin cooperates with lethal toxin to impair cytokine secretion during infection of dendritic cells. *The Journal of Immunology*, *174*(8), 4934-4941.

Tyers, M., & Mann, M. (2003). From genomics to proteomics. *Nature*,*422*(6928), 193-197.

Ulbrich, N., Wool, I. G., Ackerman, E., & Sigler, P. B. (1980). The identification by affinity chromatography of the rat liver ribosomal proteins that bind to elongator and initiator transfer ribonucleic acids. *Journal of Biological Chemistry, 255*(14), 7010-7019.

van Veen, S., Sørensen, D. M., Holemans, T., Holen, H. W., Palmgren, M. G., & Vangheluwe, P. (2014). Cellular function and pathological role of ATP13A2 and related P-type transport ATPases in Parkinson's disease and other neurological disorders. *Frontiers in Molecular Neuroscience*, 7.

Vitale, G., Bernardi, L., Napolitani, G., Mock, M., & Montecucco, C. (2000). Susceptibility of mitogen-activated protein kinase kinase family members to proteolysis by anthrax lethal factor. *Biochem. J*, *352*, 739-745.

Vodnik, M., Zager, U., Strukelj, B., & Lunder, M. (2011). Phage display: selecting straws instead of a needle from a haystack. *Molecules*, *16*(1), 790-817.

W Gruber, C., Muttenthaler, M., & Freissmuth, M. (2010). Ligand-based peptide design and combinatorial peptide libraries to target G protein-coupled receptors.*Current pharmaceutical design*, *16*(28), 3071-3088.

Wang, C., Liu, X., Liu, Y., Zhang, Q., Yao, Z., Huang, B., Zhang, P., Li, N., & Cao, X. (2013). Zinc finger protein 64 promotes Toll-like receptor-triggered proinflammatory and type I

interferon production in macrophages by enhancing p65 subunit activation. *Journal of Biological Chemistry, 288*(34), 24600-24608.

Watson, A., & Keir, D. (1994) Information on which to base assessments of risk from environments contaminated with anthrax spores. Epidemiol Infect 113: 479- 490.

Wu, S. J., Eiben, C. B., Carra, J. H., Huang, I., Zong, D., Liu, P., Wu, C. t., Nivala, J., Dunbar, J., Huber, T., Senft, J., Schokman, R., Smith, M. D., Mills, J. H., Friedlander, A. M., Baker, D., & Siegel, J. B. (2011). Improvement of a potential anthrax therapeutic by computational protein design. *Journal of Biological Chemistry*, *286*(37), 32586-32592.

Wu, X., & Lieber, M. R. (1996). Protein-protein and protein-DNA interaction regions within the DNA end-binding protein Ku70-Ku86. *Molecular and cellular biology*, *16*(9), 5186-5193.

Xie, T., Sun, C., Uslu, K., Auth, R. D., Fang, H., Ouyang, W., & Frucht, D. M. (2013). A new murine model for gastrointestinal anthrax infection. *PloS one, 8*(6), e66943.

Yang, J., Boerm, M., McCarty, M., Bucana, C., Fidler, I. J., Yuan, Z. J., & Su, B. (2000). Mekk3 is essential for early embryonic cardiovascular development.*Nature genetics*, *24*(3), 309- 313.