Open Access Theses & Dissertations

2019-01-01

Novel Methods For The Inactivation Of Cruzain For The Prevention And Treatment Of Chagas Disease

Veronica Gonzalez
University of Texas at El Paso

Follow this and additional works at: https://digitalcommons.utep.edu/open_etd

Part of the Biochemistry Commons, and the Biophysics Commons

Recommended Citation

NOVEL METHODS FOR THE INACTIVATION OF CRUZAIN FOR THE PREVENTION

AND TREATMENT OF CHAGAS DISEASE

VERONICA GONZALEZ

Doctoral Program in Chemistry

APPROVED:

_____

Mahesh Narayan, Ph.D., Chair

_____

Dino Villagran, Ph.D.

_____

Brenda Torres, Ph.D.

_____

Socorro Arteaga, Ph.D.

_____

Karina Castillo, Ph.D.

_____

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

DEDICATION

This work is dedicated to my mother, Martha P. Camacho for her support through all my

endeavors. Mom, thank you for being an amazing parent, role model and human being. I am

very lucky to have called you my mother. Though you are not physically here to accompany me

in this step of my life, I know that you are by my side.   Love you always.

NOVEL METHODS FOR THE INACTIVATION OF CRUZAIN FOR THE PREVENTION

AND TREATMENT OF CHAGAS DISEASE

by

VERONICA GONZALEZ, M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

THE UNIVERSITY OF TEXAS AT EL PASO

December 2019

AKNOWLEDGEMENT

ABSTRACT

Chagas Disease, also known as American Trypanosomiasis, is endemic to Latin American and is gradually making its way to other countries. The disease is caused by *Trypanosoma cruzi*, a protozoan parasite that causes lesions and inflammation of organs including the heart, esophagus, and colon. Over 30% of people infected will develop heart disease which can be fatal if untreated. While no vaccine is yet available, research is currently being conducted in finding new treatments for the disease. One target is an enzyme that is believed to help the progression of the disease. Cruzain, a cysteine protease expressed as a pro-enzyme, has been a target since it was discovered that inhibition of the enzyme reduces proliferation of the parasite. Current research is targeted on finding small molecule inhibitors that will interact with fully active cruzain. In our study, we describe a bottom-up approach for the arrest of protein folding. The role of the pro-domain in the folding and maturation of the enzyme is studied by comparing oxidative folding of procruzain and cruzain and by identifying structure stabilizing amino acid residues. The result of these studies can provide a novel approach to the development of Anti-Chagistic treatments and may even be useful in combating other protease-dependent diseases.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION TO CHAGAS DISEASE AND CRUZAIN

1.1 Chagas Disease

1.1.1 Epidemiology and Incidence

Chagas Disease is endemic to many Central and South American countries affecting millions of people. The disease is caused by the protozoan flagellate parasite *Trypanosoma cruzi* most commonly found in *Triatoma* insects that cohabitate with the human population in rural, tropical regions[1]. It was first described in a small female child by Carlos Chagas, a Brazilian physician, in 1909 but was not considered endemic until 40 years later.[2] Today, more that 11 million people worldwide are infected with *T. cruzi* and 10-30 percent will develop the disease. Of the Chagistic patients, 50,000-100,000 will die due to heart failure or swelling of gastrointestinal and nervous system organs cause by the parasite.[1] Over 70 million people are susceptible to the disease, causing at least 10,000 deaths per year.[3]

Of great concern is the migration of infected individuals to the United States. These individuals spread the disease by providing blood for transfusions and organs for transplants which were not previously screened for *T. cruzi*. The parasite has also been transferred congenially and accidentally in laboratory studies. The vector has also been found in non-tropical regions through migration. In addition to the Unites States; Canada, Europe and Australia are among other regions where the disease has been detected but to a much lesser extent than in South America.[4] Of the counties outside of the endemic regions, the US is the most widely affected country.[5]

The disease, vector, and etiological agent are making their way to the US by the migration of infected Latin American people to states such as Texas, California, and Florida. Table 1

1

provides an estimate of cases of infected individuals by state in 2012 and cases where infected

blood was detected (AABB cases) between 2013 and 2017.[3] The AABB, previously known as the

American Association of Blood Banks, is an organization that provides the results of blood

screening. The cases of infected blood are believed to come from individuals that contracted the

parasite in their native countries and later migrated to border states of the US. In 2012, California

was believed to have the highest incidence with over 70,000 estimated cases. Texas, follows with

an estimated 37,000 (Table 1).[3]

Table 1: Estimated cases of T. cruzi in 2012 and infected blood in 2007-2013

State	Est. Cases	AABB Cases	State	Est. Cases	AABB Cases
Alabama	1,116	8	Montana	46	1
Alaska	110	——	Nebraska	855	3
Arizona	6,440	28	Nevada	3,712	25
Arkansas	1,161	25	New Hampshire	159	3
California	70,860	707	New Jersey	8,686	32
Colorado	3,219	4	New Mexico	1,752	4
Connecticut	1,924	8	New York	17,403	160
Delaware	339	——	North Carolina	5,408	41
D.C.	745	2	North Dakota	23	1
Florida	18,096	260	Ohio	1,142	9
Georgia	5,681	37	Oklahoma	1,407	17
Hawaii	139	——	Oregon	1,995	13
Idaho	611	——	Pennsylvania	1,804	7
Illinois	9,316	22	Rhode Island	641	1
Indiana	1,705	12	South Carolina	1,486	15
Iowa	716	5	South Dakota	82	——
Kansas	1,273	9	Tennessee	1,900	14
Kentucky	618	9	Texas	36,977	176
Louisiana	1,427	15	Utah	1,767	24
Maine	49	1	Vermont	36	——
Maryland	5,926	29	Virginia	7,346	103
Massachusetts	3,346	9	Washington	3,144	18
Michigan	1,258	7	West Virginia	88	1
Minnesota	1,443	2	Wisconsin	1,239	3
Mississippi	434	11	Wyoming	112	——
Missouri	927	17	TOTAL	238,091	1,908

The AABB data included 6 confirmed infections whose state was unknown and 4 confirmed infections in Puerto Rico, not included in Table 1.

Manne-Goehler, 2016

Though the presence of both the vector and parasite has been declining in Latin America

over the years due to measures taken by the Pan American Health Organization and other entities,

the incidence in the North America has increased.[6] In Texas, before 2005, very few people were diagnosed with Chagas disease but many people, pets, and animals in the wild were found to be infected with *T. cruzi*[4] Today, over 200,000 cases of the disease have been detected across the country.[5,3] Though, the number of documented cases has not increased significantly, this data does not show information provided from undocumented migrants which has recently been estimated at over 11 million people, 300,000 of whom may be infected. As migration continue to rise, so will the incidences of disease however due to screening of donated blood, it is believed that fewer incidences will be found to have been spread domestically.[6]

Currently, ELISA (enzyme-linked immunosorbent assay) based tests are approved by the FDA as screening tools for *T. cruzi*. The ORTHO *T. cruzi* ELISA Test System, manufactured by Ortho-Clinical Diagnostics is the current method of detection of the parasite[7]. Other mechanisms for detection are currently being developed and some exist to dismiss false results given by the current method.[1] Figure 1 illustrates the cases of Chagas disease reported in Texas by 2005 and the distribution of individuals and non-human hosts that have tested positive for *T. cruzi*.
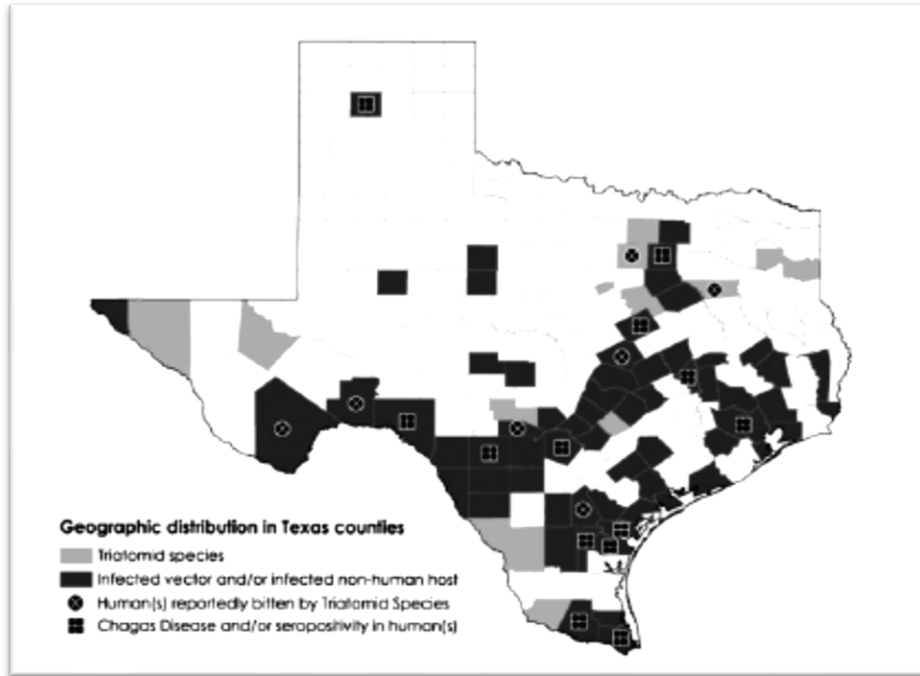
Figure 1: Incidents of *T. cruzi* and/or Chagas Disease in Texas

The infected vector has been found near the southeast and far west regions of Texas. Gray: Triatoma species, Black: Infected vector or non-human source found, Black with diamond: Humans bitten by Triatomine bugs, Black with square: Chagas disease detected.[4]

As shown in Figure 1, cases of Chagas or presence of the Triatomine bugs in Texas is localized to the southern and southeastern regions of the state, however the Triatomine species carrying the parasite have also been documented in West Texas and have additionally been found in New Mexico, Arizona and a high amount of cases have been recorded in New York[3]. With the migration of immigrants to the Unites States continuously increasing, the disease is sure to spread over time if not controlled.

1.1.2 Etiological Agent and Infection

The Triatomine insect is the principal vector of the protozoan parasite. There are over 130 species of Triatomine insects, many of them sylvatic, usually do not infect domestically[8] however

4

in endemic countries, the population co-resides with the insect making them susceptible to exposure. The primary species that carries the parasite, *Triatomine infestans*[9]*,* can be seen in Figure 2. Several Triatoma species have been found in the Unites States. Figure 3 is a photograph of 11 species of Triatoma found in the US that have been proven carriers of the parasite.[8]

Figure 2: Picture of *Triatoma infestans*

Triatoma infestans is the main vector that carries the parasite that causes Chagas disease.[9]
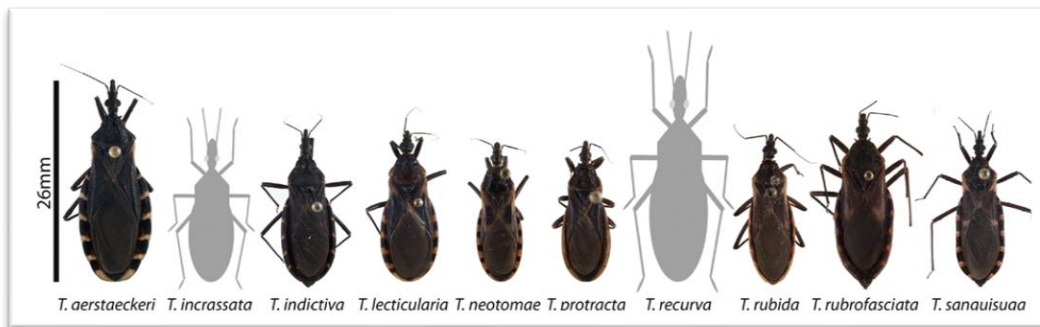
Figure 3: Vectors of the parasite *T. Cruzi* found in the United States.

Most of the vectors found in the US are of the Triatomine species[8] photo by S. Kjos.

The infected blood sucking Triatomine insects commonly known as "kissing bugs" contain the parasite in its mid gut as the epimastigote (insect) form where it can multiply. The parasite

will then travel to the insect's hind gut as a trypomastigote. As the insect feeds on its host it defecates leaving *T. cruzi* on the skin. A person who has been bitten will scratch the affected area, causing lesion in the skin allowing the parasite to enter the blood stream in a trypomastigote form[10]. Figure 4 is a picture of a Giemsa stained trypomastigote form of *T. cruzi,* the etiological agent. The parasite can also enter the body via mucosa membranes of the eyes and other tissues[11]. After the human host is inoculated with *T. cruzi* it will then infect lymphocytes, the white blood cells responsible for responding to the foreign antigens[1].
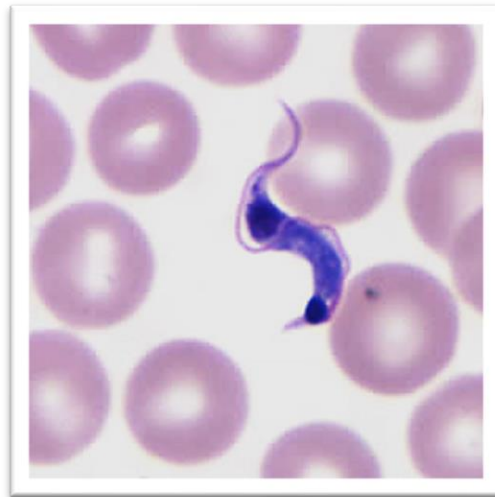
Figure 4: Picture of the *T. cruzi*

Blood smear of *T. cruzi* trypomastigote, a protozoan parasite, is the causative agent of Chagas Disease stained with Giemsa.[12]

Macrophages are the primary immune response to defend against the parasite, therefore along with eosinophils are the most widely infected cells. Within the cells, the trypomastigotes are transformed to the amastigote form, which proliferate by binary fission. The cells then rupture and the amastigotes are released into the blood stream to prepare for the next cycle. Other blood-

sucking insects become infected as they feed on infected hosts. Eventually, the parasite reaches the endothelial cells of the cardiac, nervous, and smooth tissues resulting in cardiomyopathy, megaencephalitis, and megacolon.[1] Figure 5 is a H&E stained slide of cardiac tissues infected with amastigotes of *T. cruzi*.[12] Figure 6 summarizes the life cycle of *T. cruzi* within the parasite and the human host.[8]
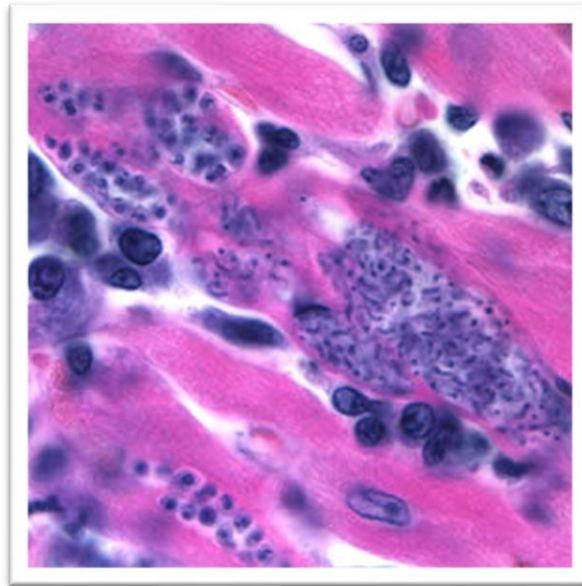
Figure 5: Heart muscle tissues infected with *T. cruzi*

Cardiac muscles tissues were stained with hematoxylin and eosin (H&E). The amastigote form visible in purple is found in tissues where trypomastigote form are found in the blood.[12]
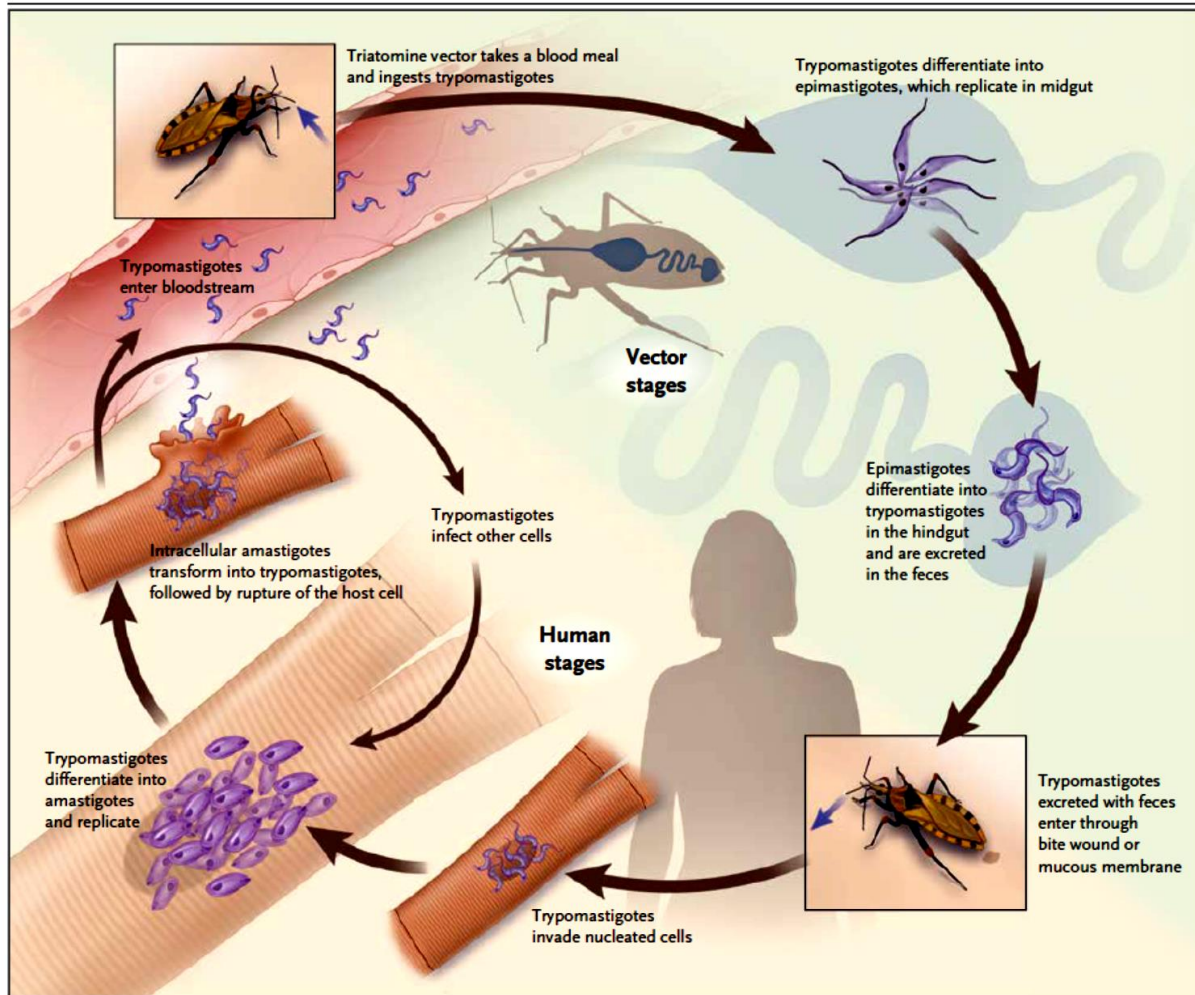
Figure 6: Life cycle of *Trypanosoma cruzi* within the vector and host

The life cycle of the parasite in the Triatoma insect (red arrows) and in the human host (blue arrows) are shown.[8]

The parasite infects and survives within the cells that are responsible for engulfing and destroying antigens and survives even in the presence of oxidative stress. *Trypanosoma cruzi* has evolved to become resistant to reactive oxygen species and mechanisms by which the cells eliminate foreign substances. Studies show that *T. cruzi* contains specialized enzymes that allow the parasite to exist within the harsh environment of macrophages and other white blood cells.[1]

1.1.3   Clinical Manifestations

Chagas Disease is characterized by two phases: An acute and chronic phase. It affects women and men equally from the ages of 20-50 but can also infect children. The disease in the acute phase is usually undetected.[13] It can be characterized by irritation, swelling at the site of transmission, nausea, diarrhea, and fever.[11] In some cases, swelling of the eyes called Romaña's sign (Figure 7) is visible near the site of infection.[14] The Triatoma species of the insect tend to bite around the nose and mouth regions which is why it is commonly referred to as the "kissing bug". In rare cases the acute phase progresses to heart disease, gastrointestinal and nervous system failure. The parasitemia is highest in the acute phase and can be detected in the blood. This stage lasts approximately 6-8 weeks however, patients usually do not experience severe symptoms therefore do not seek medical attention and screening.[11]

Figure 7: Romaña's signs of infection

A swollen eye is one of the symptoms of infection.[13]

An intermediate phase follows which can be asymptomatic. This phase can last years and even decades without ever showing signs of chronic infection. Many people living with Chagas disease spend their entire lifetime in the intermediate phase. Clinical manifestations of the chronic phase include cardiomyopathy, megaencephalitis, and megacolon. Heart disease caused by Chagas is the main cause of death in Latin America. Other symptoms include swelling of heart valves and gastrointestinal tract.[13] Swelling and lesions in the affected organs may be due to autoimmune response or the persistence of parasites in the circulatory system[15] however, since parasite load at this stage is relatively low, it is more difficult to detect Chagas disease at the intermediate or chronic phase.[13]

1.1.4 Current Treatment

The treatment of Chagas disease includes administration of two main drugs: nifurtimox and benznidazole, Figure 8. The therapeutics treatments are most effective during the acute phase of the disease however the disease is not usually detected until the chronic stage is reached, a point at which it is likely organ damage and swelling has already occurred. The effectiveness of the treatment decreases as time passes. These two drugs are not free of side effects therefore development of other modes of treatment is essential[16]. Some of the many side effects include fever, neurological toxicity, gastrointestinal irritation and allergic reactions[17]. The first line of treatment is benznidazole, a nitroimidazole drug, since the side effects are less severe that those from nifurtimox.[18] Other azole-based drugs such as itraconazole have been tested as anti-Chagistics are based on anti-fungal medication but are not as effective as the current treatment.[19,16] The mode of action of these azole anti-parasitic drugs is not fully understood but it is believed that benznidazole is metabolized by the parasite to produce a reduced hydroxylamine compound. The reduced compound breaks apart to produce a cytotoxic and mutagenic substance which may attack

the parasite's DNA. This drug is also believed to prevent expression of or inhibit enzymes that are responsible for parasitic survival during oxidative stress. Nitrofurtimox, a derivative of nitrofurane, is metabolized by parasitic enzymes to produce radicals that also cause oxidative stress and destabilize the parasitic DNA.[19]
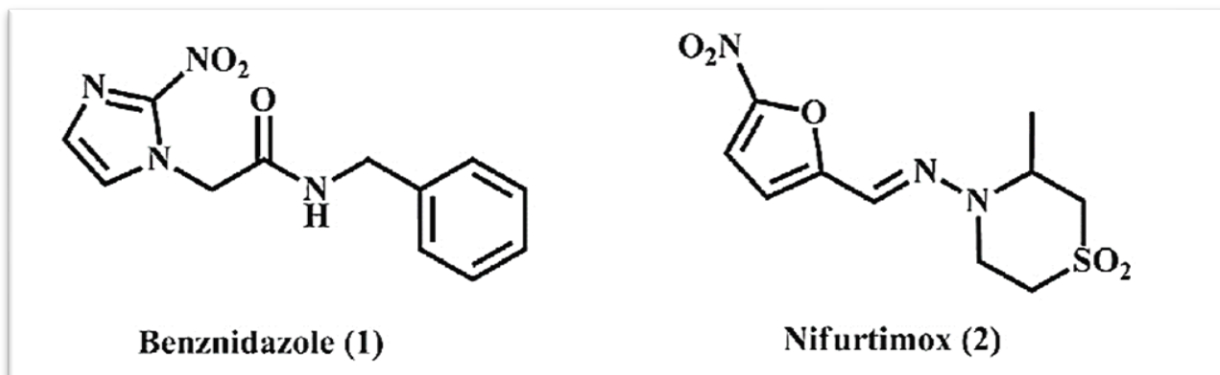
Figure 8: Current drugs used to treat Chagas disease

Benznidazole (1) is the primary drug used for the treatment of Chagas disease and Nifurtimox (2) is an alternate treatment.[19]

Both compounds contain a nitro group at the imidazole ring that is reduced to highly reactive metabolites. Additional side effects may result from metabolism of the drugs are extensive and can be found in the Table below. The side effects are dose dependent and while some are rare, they can be damaging.

Table 2: Side effects associated with recommended dosage of anti-Chagistics[8]

Benznidazole
Dosage regimen*
    Age <12 yr: 5–7.5 mg/kg per day orally in 2 divided doses for 60 da
    Age ≥12 yr: 5–7 mg/kg per day orally in 2 divided doses for 60 days
Side effects in adults†
    Allergic dermatitis (frequent: 29 to 50%)‡
    Paresthesia (frequent: 0 to 30%)
    Peripheral neuropathy (frequent: 0 to 30%)§
    Anorexia and weight loss (frequent: 5 to 40%)
    Nausea or vomiting (infrequent: 0 to 5%)
    Leukopenia (rare: <1%)§
    Thrombocytopenia (rare: <1%)§
    Early discontinuation because of side effects (frequent: 7 to 20%)
Nifurtimox
Dosage regimen¶
    Age ≤10 yr: 15–20 mg/kg per day orally in 3 or 4 divided doses for 90
    Age 11–16 yr: 12.5–15 mg/kg per day orally in 3 or 4 divided doses for 9(
    Age ≥17 yr: 8–10 mg/kg per day orally in 3 or 4 divided doses for 9(
Side effects in adults†
    Anorexia and weight loss (very frequent: 50 to 75%)
    Nausea (frequent: 15 to 50%)
    Vomiting (frequent: 15 to 26%)
    Abdominal discomfort (frequent: 12 to 40%)
    Headache (frequent: 13 to 70%)
    Dizziness or vertigo (frequent: 12 to 33%)
    Mood changes (frequent: 10 to 49%)
    Insomnia (frequent: 10 to 54%)
    Myalgia (frequent: 13 to 30%)
    Peripheral neuropathy (infrequent: 2 to 5%)§
    Decreased short-term memory (infrequent: 6 to 14%)
    Leukopenia (rare: <1%)§
    Early discontinuation because of side effects (frequent: 6 to 40%)

The most common means of prevention is through the administration of insecticides in endemic regions. This can also be problematic because insecticides can be toxic. Research on vaccines as a mode of prevention and small molecule inhibitors of the major cysteine protease of *T. cruzi*, cruzain, is currently being conducted.[20] Administration of a vaccine may cause an over active immune response, so introduction of the parasite though vaccine can cause the same immune response leading lesions in organs that result similar to those that *T. cruzi* produces as it

proliferates.[1] A major chemotherapeutic target for the treatment of Chagas disease is cruzipain, the major cysteine protease localized in lysosomes of *T. cruzi*. This enzyme was named for its

homology to papain and after Oswald Cruz, mentor of Carlos Chagas. It belongs to the papain superfamily of proteins and is homologous to papain and the cathepsins. The name cruzipain is given to the full enzyme within the parasite and the recombinant form without a C-terminal domain extension is known as cruzain. Cruzipain is present in all stages of the protozoan life cycle[21] which makes it an attractive target. The proteolytic enzyme takes part in the infection process, provides nutrition to the parasite and combats the immune response. Papain-like proteases are known to hydrolyze regions of immunoglobulin G (IgG) antibodies that present antigens to phagocytes for degradation. In epimastigotes, cruzipain is a 60kDa protein composed of two domains, one mostly comprised of alpha helices and another of beta sheets.[22] Current research is being conducted to produce inhibitors of cruzipain.

1.2 Cruzain as a Target

1.2.1 Proteases

Proteases are enzymatic proteins that catalyze the hydrolysis of proteins. Also known as peptidases or proteinases, these enzymes break the amide bond between amino acids in a polypeptide chain. They may break bonds at certain residues producing fragments or may be non-specific and degrade at all peptide bonds to produce individual amino acid residues.[23] Proteases are present in all organisms and contribute to a variety of processes including but not limited to cell death, growth, repair and digestions depending on location and type. Proteases are of the mostly widely studied enzymes and are highly produced using recombinant techniques. It is estimated that proteinases make up about 60% of enzymes expressed organisms.[24] Proteases are normally synthesized as inactive zymogens for regulation and often undergo autoactivation by

removing the prodomain that inhibits catalysis. The prodomains are believed to behave as chaperone-like region of proteases which may allow the protein to achieve the native structure. These regions of proteases involved in disease are currently being studied as a target of inhibition.

Hydrolysis of proteins is important for the degradation of aged cellular material and production on newly synthesized peptides. Peptidases can be classified in several ways depending on the identity of the catalytic moiety, the chemical mechanism of degradation and homology with other proteins. There are two main types of proteinase: endopeptidases and exopeptidases. Endopeptidases catalyze the hydrolysis of a peptide bond within the substrate peptide strand at certain amino acid residues. Exopeptidases cut peptide bonds at the amino terminus (aminopeptidases) or carboxyl terminal (carboxypeptidases) of a target protein[23]. It is estimated that approximately 2% of expressed proteins are proteases.[24]

These enzymes can be isolated from plants, animals, microbes and are often mass produced by recombinant DNA technology. Most of proteinases initially come from animal sources but considering regulations and concerns with animals as a source, many advances in recombinant DNA technology have been improved over time to produce higher yields of proteins. Figure 9 provides a comparison between extraction of proteinases from living organisms versus those produced by biotechnological advances.

14
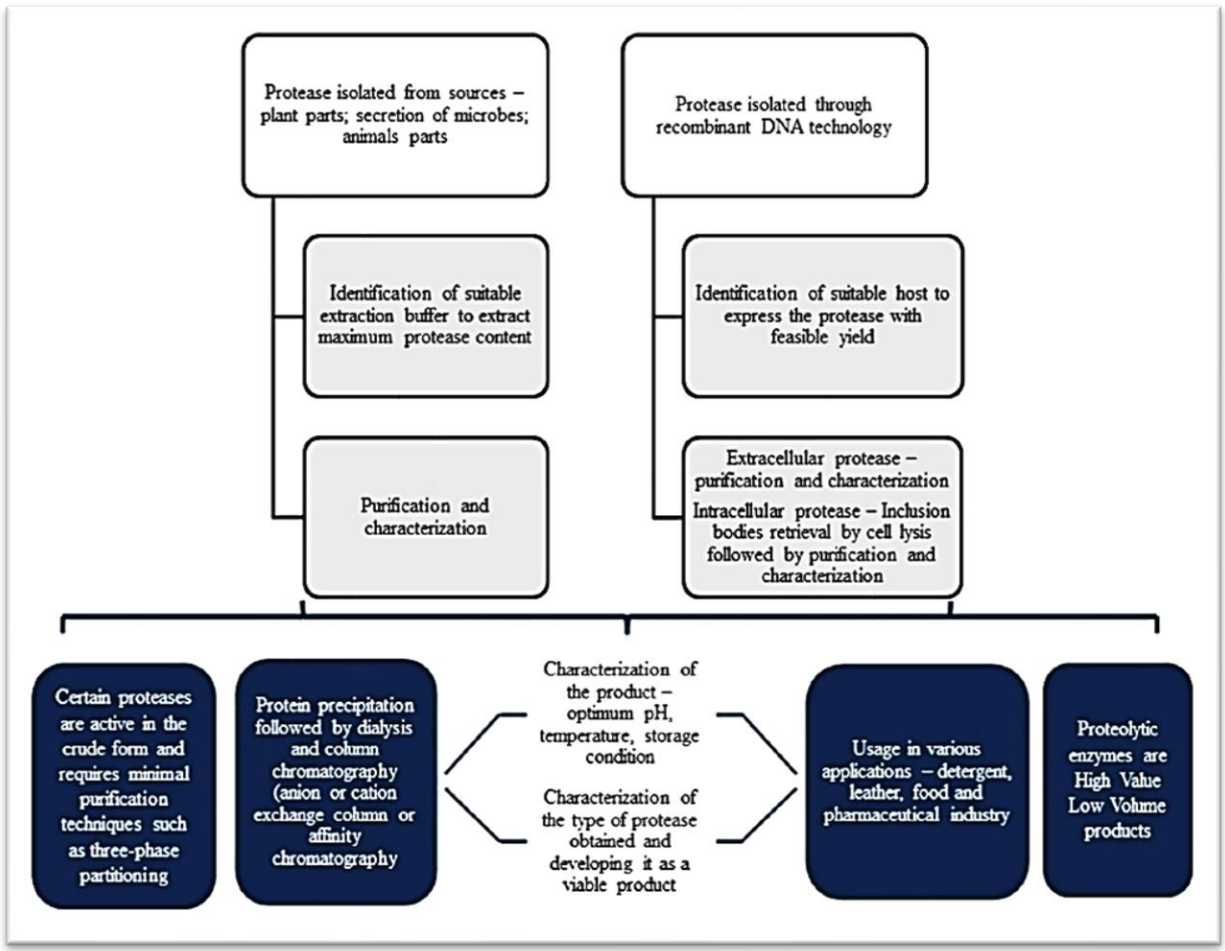
Figure 9: Production of Proteolytic Enzymes

Comparison between production of proteinases from living organisms and through recombinant technology.[24]

Both methods will involve processing of cells containing the enzymes of interest, as well as purification of and characterization of the enzyme. One advantage of extracting proteinases directly from the source is that after separation of the gene products, the exact genetic makeup of proteins can be determined. To express proteins via DNA recombinant technology, the amino acid sequence must be known which should be determined directly from the source.   After determination of the amino acid sequence plasmids can be produced and inserted into vectors that

will produce proteins in mass quantities, often as inclusion bodies. This method is commonly used in industry.[24] Proteases are widely studied however they are costly.

Proteinases can be subclassified according to catalytic amino acids with in the active site. The subclasses of proteases include: threonine, aspartate, glutamic acid, serine proteases, cysteine and metalloproteases. Many of the subclass of proteases have highly conserved catalytic dyads and often triads that carry out the enzyme activity.

Threonine proteases contain a threonine residue in their catalytic domain. The secondary alcohol of the residue behaves as a nucleophile to form an intermediate that is subsequently hydrolyzed. Threonine proteases are a major component of proteasomes, large complexes that degrade aged or damaged proteins. Aspartate proteases contain an aspartic acid residue in the catalytic region. These proteases produce an activated water molecule that can attack the substrate during hydrolysis. Examples of aspartic acid proteinases include pepsin and cathepsin D. They are also involved in digestion and blood pressure control. Glutamic acid proteases contain a glutamic acid residue in the catalytic core.[24] Metalloprotease contain a divalent metal ion[25] such as Zn^{2+}. One example is thermolysin, a metalloproteinase commonly found in *Bacillus* species. Metalloproteinases that are found in the tissues of vertebrates include collagenases, others are involved in would healing and tumor metastasis.

There are 20 families of the serine proteases, all which have a highly conserved catalytic triad containing aspartate, histidine and serine residues. Similar to threonine proteinases, serine proteinases have nucleophilic alcohol in the active site.[26] One example of serine protease is subtilisin, a bacterial serine protease first isolated from *Bacillus subtilis*. The role of the prodomain of these proteases is widely studied. Chymotrypsin and trypsin are additional examples of serine peptidases. Chymotrypsin hydrolyzes at phenylalanine, tyrosine and tryptophan while trypsin

cleaves at the C-terminal of arginine and lysine residues. These enzymes create fragments of protein that can be separated by liquid chromatography and studied by mass spectroscopy. Serine proteases are often found in the animal pancreas, assist in digestion, blood clotting and immunity.[27]

Cysteine proteases, also known as thiol proteinases, contain a cystine residue in the catalytic core. Many contain a highly conserved catalytic triad that includes cysteine, histidine and asparagine. These proteases have been widely studied as a target for a variety of diseases[28]. Some examples include papain, calpain and cathepsins B, H, L. These enzymes are divided into four subtypes: papain-like, glutamic acid-like, trypsin-like and others that do not fit any of those classifications[24]. Papain, isolated from the papaya fruit, was the first cysteine proteinase to be isolated and characterized.[29] Calpain are cysteine proteases that contain a calcium ion in the catalytic domain[24]. Human cathepsins are part of the papain-like family of cysteine proteases and can be found in epithelial cells, thymus and testis. Some cathepsins play a significant role in immunity and are expressed on antigen presenting cell, T-cells and natural killer cells while others are involved in bone resorption.[30] The cathepsins and other cysteine proteases are a target of study for their involvement in cancers, arthritis and immune system related disorders. Parasitic and viral cysteine proteases are among the most widely studied for their role in diseases such as Chagas disease, African sleeping sickness, AIDS, and malaria.[29] Parasitic cysteine proteases are important to the survival and proliferation of the parasite and are present in all stages of the parasitic life cycle which make parasitic cysteine proteinases an important therapeutic target.

1.2.2   Cysteine Protease Structure, Activation, and Proteolytic Mechanism

Cysteine proteases are often found within lysosomes, the degradation compartment of the cell. Lysosomal cysteine proteases are active under acidic conditions and tend to lose their activity at physiological pH therefore it is not common to find an active cysteine protease outside of the

17

lysosome.   In fact, in neutral conditions, these enzymes tend to denature and aggregate. Structurally, they contain a prodomain, catalytic domain, c-terminal extension and signal sequence. Similar to other proteases, cysteine proteases are synthesized a precursor protein. The prodomain of these zymogenes covers the catalytic domain to prevent auto-degradation and can range from 30 to over 200 amino acids in length. The sequences and structures of papain-like cathepsins are widely characterized.   Figure 10 is an illustration of different types of cathepsins in which the prodomains range in size. There are over 10 members of human cathepsins and many of the structures have been characterized by x-ray crystallography. Other cysteine proteases have been studied using techniques applied to isolate and characterize the cathepsins.[30]
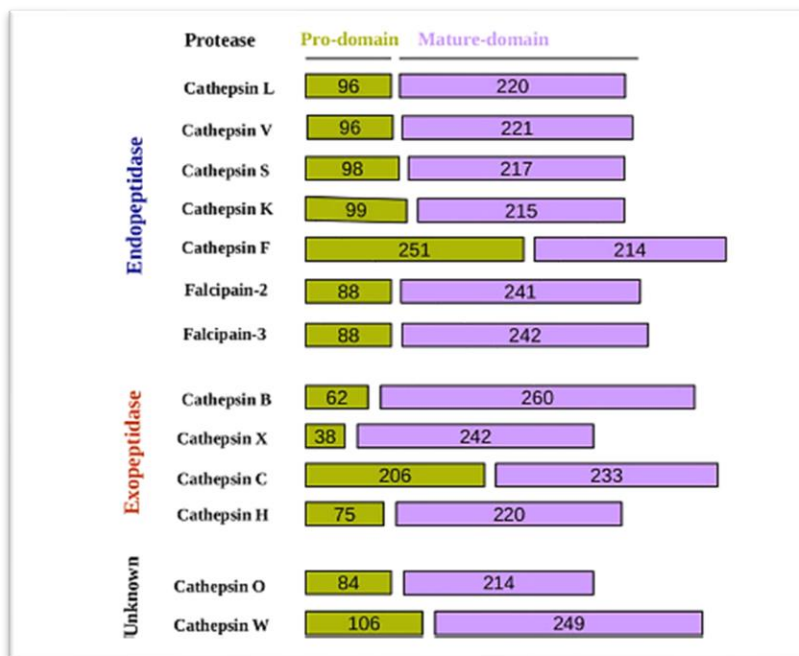
Figure 10: Comparison of the pro- and catalytic domains of different cystine proteases

Green regions denote prodomains while the purple regions designate the catalytic domains. Of the examples listed, most are Cathepsins.[29]

18

Cysteine proteases can mature by both auto-activation and *trans*-activation[31]. The pro-domain of many papain-like proteases such as the cathepsins interacts with the catalytic domain through salt bridges, hydrogen bonding and hydrophobic interactions. Based on previous studies on cathepsin L, it is believed that many papain-like proteases undergo activation in acidic or reducing conditions causing a disruption in the salt bridges formed between charged residues in the prodoamin which in turn causes a disruption of hydrogen bonding and hydrophobic interactions. The interruption of the interactions within the prodomain cause structural changes in the protein which no longer allows the prodomain to block the catalytic core. Exposing the catalytic core may then allow for autoactivation. Once activated, mature enzyme can further activate the zymogens in a trans-activation mechanism thus, activation can be unimolecular and bimolecular.[32]

Once mature, the active site is free to catalyze protein degradation. Cysteine proteases break peptide bonds by nucleophilic attack of the cysteine thiol. The sulfur of cysteine is not very nucleophilic however the histidine in the catalytic triad increases the nucleophilicity of the thiol by the interaction of the imidazole ring with the hydrogen of the cysteine sulfhydryl resulting in a thiolate-imidazolium dyad. The thiolate- imidazolium interaction may be stabilized by the asparagine of the catalytic triad. The thiolate being a stronger nucleophile than the sulfhydryl attacks the carbonyl carbon of the scissile bond within the substrate protein displacing the pi elections of the carbonyl, creating tetrahedral intermediate oxyanion hole. As the double bond of the carbonyl regenerates the nitrogen of the scissile bond leaves. Water attacks the carbonyl, once again displacing the pi elections, producing a tetrahedral oxyanion hole. As the double bond regenerates, the thiolate is cleaved. Glutamate may help generate and stabilize the oxyanion holes.[33] Figure 11 summarizes the catalytic mechanism of cysteine proteases. Cysteine proteases often cleave at phenylalanine and arginine residues.
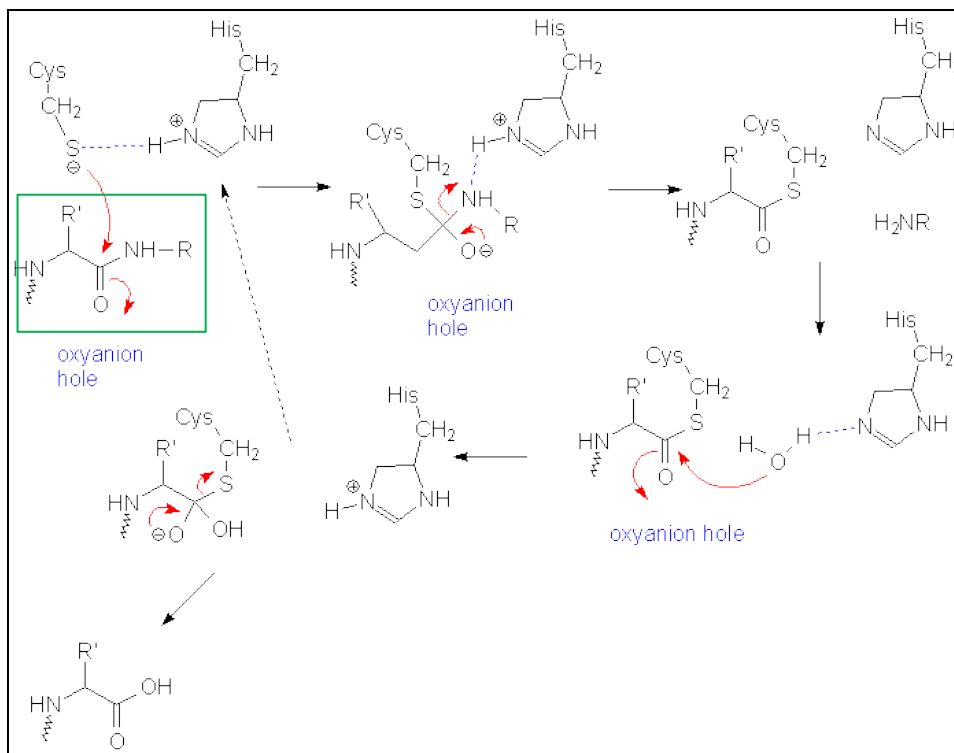
Figure 11: Cysteine protease mechanism.

Red arrows the transfer of electrons from nucleophile to electrophile. The green box highlights the substrate protein.[34]

Parasitic cysteine proteases are currently a target for the prevention and treatment of diseases. Parasites that cause neglected tropical diseases (NTDs) affect over a billion people worldwide. Latin America, Africa and Asia are among the regions most affected[35]. Chagas disease, a NTD caused by *Trypanosoma cruzi*, is the main cause of heart disease in Latin America. Human African trypanosomiasis is an NTD cause by *Trypanosoma brucei* that eventually leads to neurological disorders.[36] Over 10,000 cases of infection are documented each year, leaving over 70 million susceptible to the disease[37]. Cysteine proteases are present in every stage of the parasitic life cycle and are involved in invasion, nutrition, reproduction and progression of disease[33] therefore are considered a valuable target for prevention and treatment of diseases caused by the

parasites. Current drug discovery efforts involve the development of small molecule inhibitors that arrest progression of disease by preventing cysteine protease catalytic activity.[38] Though many small molecules have been developed over the years, none have reached as far as human clinical trials.

1.2.3   Structure and Function of Cruzipain

Cruzipain is a cathepsin-like cysteine protease of the papain family found in *T. cruzi*. Because if its presence in all stages of the parasitic life cycle, it is a major therapeutic target. The DNA sequence was deduced by Eakin et.al. in 1992 and the crystal structure of the mature proteinase was produced shortly after. The endogenous enzyme was expressed within each form of the parasite but the expression of cruzipain, the full prepropeptide, was highest in the amastigote form. A 1.85 kbp DNA sequence from *T. cruzi* was isolated, sequenced and compared to other cysteine proteases for classification. The full sequence of prepropeptide was found to be most homologous to the cysteine proteases found in *T. brucei* and a murine cathepsin-like protease both belonging to the papain family; hence the full protein was named cruzipain. It was found that 6 copies of the gene are located within the parasite's genome, giving an insight to its significance. Figure 12 provides the organization of the vector that was constructed to contain the cruzipain domains, which was expressed in bacteria and purified for analysis of enzymatic function. The major domains include the CheY signal sequence, prodomain, protease core and a c-terminal extension.[39]
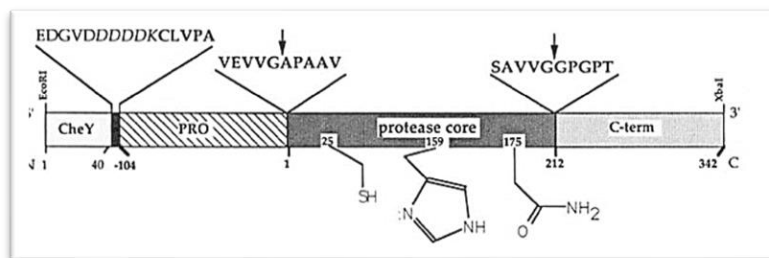
Figure 12: Genetic organization of cruzipain

The regions of the genome that are highlighted include the pro-domain, protease core and the c-terminal extension.[39]

The truncated version of cruzipain, cruzain, is commonly used for anti-Chagistic studies. The proenzyme is expressed as a 215 amino acid residue-long protein lacking the c-terminal extension, which undergoes autocatalytic activation to produce the mature enzyme. Over 25 crystal structures of cruzain inhibited with a small molecule are available in the Protein Data Bank, however no crystal structure of procruzain exists to date.[40] Figure 13 is an x-ray crystal structure of the active enzyme containing a small molecule inhibitor in the catalytic domain. The structure of mature cruzain is composed of two main types of domains, an alpha helical domain and a beta pleated domain. The active site located between the major domains.[41,42] Also highlighted is the catalytic triad, a highly conserved region among cysteine proteases, in which cysteine and histidine are essential to the function. Figure 13 is a structure generated by PyMol of the catalytic triad consisting of Cys25, His159, Asn175, common in papain family members.[42] The enzyme also contains three disulfide bonds that can be of great importance when studying the folding mechanism of the protein as it is being synthesized with in the cells.

Figure 13: X-ray crystal structure of cruzain

Cruzain crystallized with a small molecule inhibitor in the active site.[41]
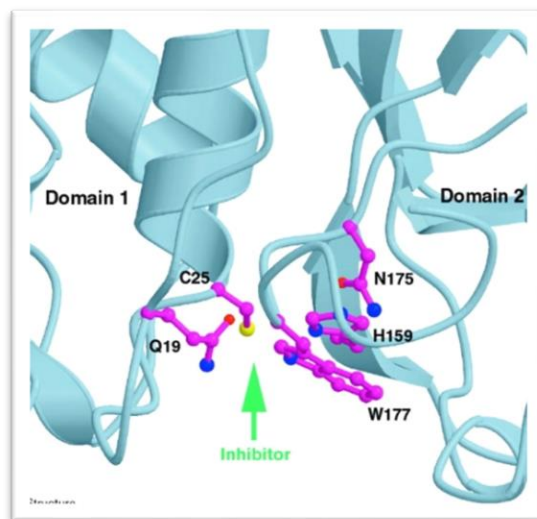
Figure 14: Cruzain catalytic triad

The cruzain catalytic triad consists of Cys25, Hist159, and Asn175 (pink). Glutamin19 and Tryptophan177 are

essential for the efficient protease activity but not part of the catalytic triad. Green arrow indicates active site.[42]

It has been found that cruzain is not only located in lysosomes but also on the surface of the cell and the flagellar region.[44] Cruain is believed to play a significant role in invasion of cells, differentiation of the parasite, and blood break down for nutrition.[45] High titers of IgG were found in the blood of infected individuals however, cruzain uses its proteolytic activity to escape the immune system.[46] Studies have proven that cruzain degrades immunoglobulins leaving free Fab segments that can localize around the parasitic surface, allowing them to escape phagocytes.[47] Figure 15 highlights the hinge region of IgG, the location of degradation of Fab from Fc regions of the antibody. Another mechanism for evasion of the immunes system is cleave of transcription factors by cruzain causing a deactivation of macrophages.
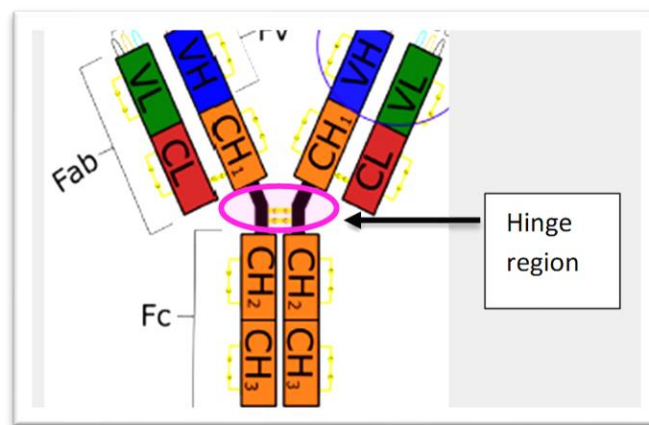
Figure 15: Structure of IgG

Cruzain cleaves at the hinge region of IgG (highlighted in pink), releasing Fab domains (structure of IgG from[45] which then are available to coat the parasite and avert the immune response[47].

The C-terminal extension is not believed to be involved the protein's catalytic activity therefore "cruzain" is suitable for studies. In our investigation, procruzain is expressed as a 35kDa inactive zymogen that must be activated to the 26kDa mature enzyme.[48,49] Currently, small

molecule inhibitors of this protease are being synthesized and analyzed for treatment as a point of inhibition of the enzyme however, problems exist with irreversibility, bioavailability and tolerance of the novel chemotherapeutics. Aside from these challenges, cruzain is homologous to human proteins like Cathepsin L which can create side effects. Recent studies on numersou inhibitors claim that the specificity is high enough to not interfere with human gene products and do not create side effects, but further studies must be conducted to determine long term effects.[50]

1.2.4 Implications in Treatment and Prevention of Chagas Disease

Due of the lack of efficiency of current treatment of Chagas disease, cruzain is a highly studied target for the development of novel therapeutics. Benznidazol and nifurtimox are only effective in treatment during the early onset of the disease (the acute stage) however, most cases go undiagnosed until the chronic stage is reached and the illness become more debilitating. Nifurtimox (a nitrofuran) and benznidazole (a nitroimidazole) and were discover in 1965 and 1971 respectively and later approved for the treatment of fungal infections. Other antifungals were tested against *T. cruzi* but were proven ineffective. Recently, fexinidazole, a nitroimidazole similar in structure to benznidazole has been tested for the treatment of Chagas disease in both the acute and chronic phases. It was shown to cure over 90 percent of acute and chronically affected mice however the sample in the study was small (n=6). Surprisingly, fexinidazole (Figure 16) was more effective at treating infected mice in the chronic stage than the acute stage. This may be because the parasite is restricted to fewer locations within the body and not widely systemic like in the acute phase.[51] In 2014, fexinidazole was under Phase II of clinical trials. In Bolivia, the drug was administered to 47 patients but the study was interrupted due to safety issues and drug intolerability. In 2018, the drug re-entered Phase II in Spain and with significant efficacy will continue to 2019 and conclude in 2020.[52]
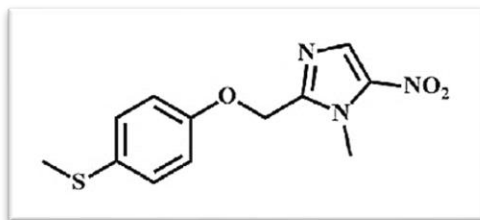
Figure 16: Structure of fexinidazole

The structure shows that this molecule is a nitroimidazole derivative. It is currently being investigated as treatment for Chagas disease and other parasitic infections.[52]

As previously mentioned, current research is being conducted on inhibiting the active site of cruzain with small molecules. High-throughput screening methods have identified several classes of inhibitors, many of which have been synthesized and tested against cruzain. Upon finding effective inhibitors, pharmokinetics are constantly being improved. While effective in interacting with the active site of cruzain, issues with the bioavailability and half-life of the molecules still remain.[38] Fluoromethyl ketones were among the first inhibitors synthesized to block activity of cruzain.[53] Vinyl sulfonic molecules were also developed against cruzain, the most effective being K777 (Figure 17). This vinyl sulfone derivative irreversibly inhibits cruzain in the nanomolar range arresting proliferation of the parasite however the drug is intolerable and the study ended in the preclinical trials.[45] Tetrafluorophenoxy methyl ketones have also shown efficacy in inhibition at low concentration (IC_{50} = 3nM) and peptidyl nitroalkanes inhibit cruzain (IC_{50} = 0.44nM). An example of a Tetrafluorophenoxy methyl ketones is seen in Figure 18 and peptidyl nitroalkanes in Figure 20.
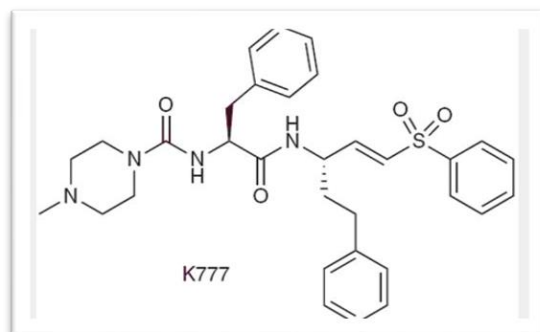
Figure 17: Structure of K777

A vinyl sulfone derivative studied for the inhibition of cruzain.[45]
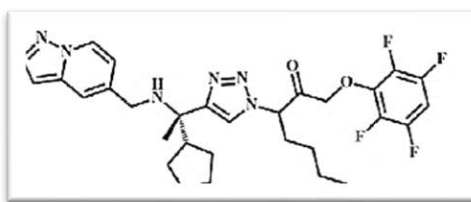
Figure 18: Example of a tetrafluorophenoxymethyl ketone
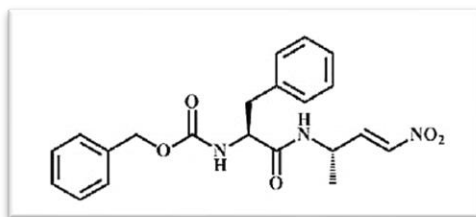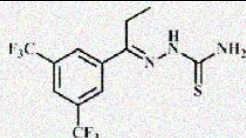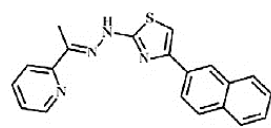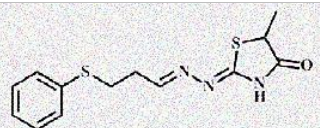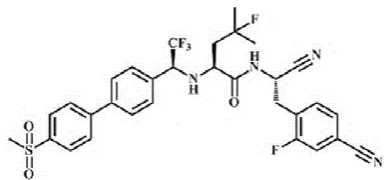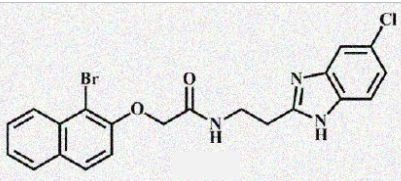
Figure 19: Example of a peptidyl nitroalkene

The following table is a list of additional inhibitors that have been studied for their activity against mature cruzain but have not yet enter clinical trials.

Table 3: Additional cruzain inhibitors previously and currently studied

Inhibitor type	Structure
Thiosemicarbazones	
Thiazoles	
Thiazolidines	
Nitriles	
Benzimidazole derivatives	

Structure found in Ferriera[38]

Our aim is to avoid small molecule interaction with host proteins by preventing cruzain from reaching its native state. We call this a bottom up approach because the protein will be arrested early in its expression, minimizing the proliferation of the parasite and the damage to the host. Folding rate comparison between cruzain and procruzain and identification of structure stabilizing residues can lead us to produce small molecules that may prevent three-dimensional

folding of procruzain. These finding will provide a different angle and novel approach to creating treatment for Chagas disease.

1.3 Protein Synthesis

Proteins are biological macromolecules composed of amino acids and are responsible for carrying out many of the cell's functions within an organism. Proteins are the gene products encoded by DNA. Each protein has a specific function that is determined by its primary amino acid sequence. Some biological processes proteins take part in include catalysis of thermodynamically unfavorable biochemical reactions, signaling and trafficking of substrates, defense against pathogens and folding of other proteins. Protein synthesis involves the transcription of a single stranded DNA template into messenger RNA followed by the translation of mRNA to a primary amino acid sequence. Translation is the last step of gene expression but can be considered the first step in the formation of a fully functional protein. Synthesis of peptide chains determined by mRNA occurs within free ribosome or those located outside of the endoplasmic reticulum (ER). [54]

There are four levels of organization of a protein that will determine its structure and function. The primary structure of a protein is the linear sequence of amino acids[54]. The first full length amino acid sequence of a protein was determined by Frederick Sanger in 1953. Sanger discovered the primary structure of insulin, but more importantly revealed that each protein consists of a specific amino acid sequence that determines the function. The secondary structure of a protein is the regular arrangements of amino acids within localized regions of the polypeptide chain. Alpha helices and beta sheets are formed and stabilized by hydrogen bonds and hydrophobic interactions. The tertiary structure of a protein is achieved by the interactions between the side chains of amino acids that lay in different regions of one single peptide chain.

Hydrophobic amino acids will aggregate to the interior of the protein while hydrophilic amino acids will be exposed to the aqueous environment surrounding the protein. Also contributing to tertiary structure is the formation of disulfide bonds between thiol-containing amino acid residues. Finally, the fourth level of protein organization is the quaternary structure in which two or more peptide chains come together as subunits to form one functional protein[54]. Not all proteins reach a quaternary structure. Figure 20 summarizes the different steps in protein synthesis and the formation of a functional protein.
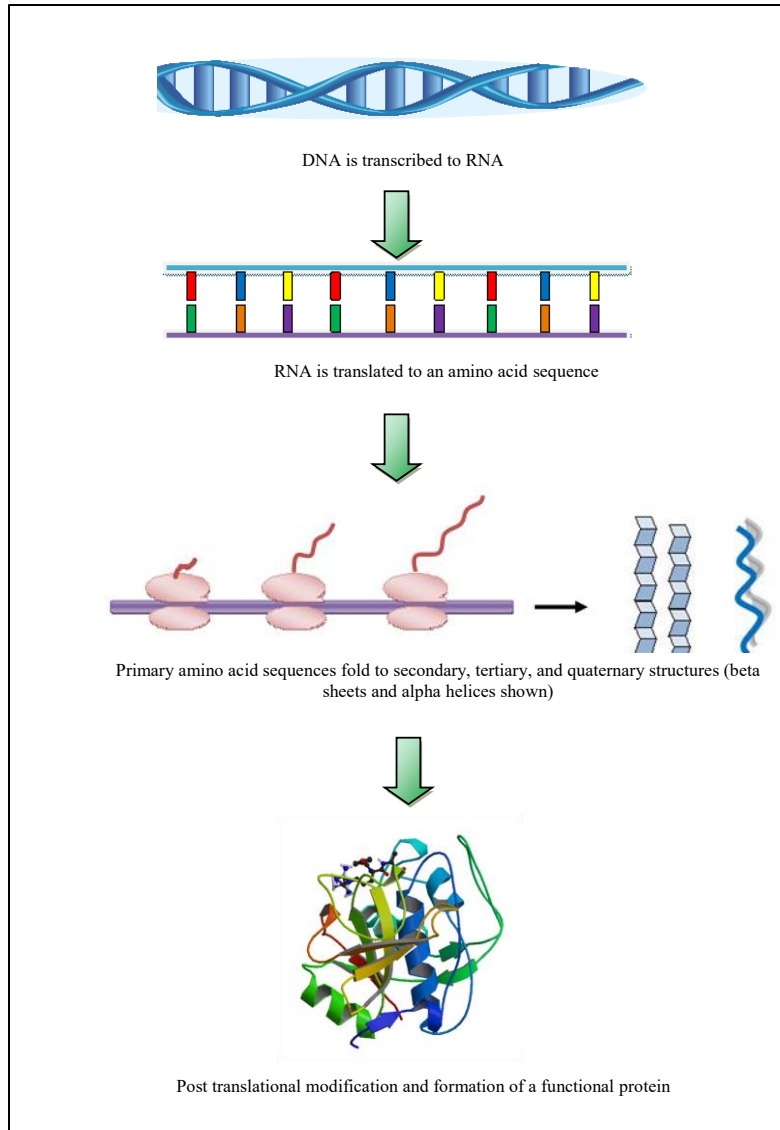
Figure 20: Summary of protein synthesis.

Transcription, translation, and three-dimensional organization of a protein

Nobel Prize recipient Christian B. Anfinsen, known for the "Anfinsen Dogma" also known as the "thermodynamic theory", demonstrated that denatured proteins can spontaneously fold back to their native three-dimensional conformation and regain activity. This indicates that the native structure of the protein is determined by the information given in the primary amino acid

sequence[56]. Under physiological conditions, the primary amino acid sequence will give rise to the most thermodynamically stable conformation.[57] Catalysts and chaperones facilitate the protein in achieving a native structure, but it is the protein's amino acid sequence that determines conformation[56].

During and after synthesis, proteins undergo modification before becoming fully functional. Secreted, plasma membrane, and lysosomal proteins contain a signal sequence which allow them to translocate into the ER for post-translational modifications such as disulfide bond formation and to their final destination. Disulfide bond formation occurs when sulfur from a cysteine is oxidized and forms a covalent bond with the sulfur from another thiol-containing moiety. The ER is an oxidizing environment relative to the cytosol allowing disulfide bond formation to occur.[57] Aside from disulfide bonding, other functions of the ER include further protein folding, multi-subunit formation, initiation of glycosylation, and addition of glycolipid anchors. ER proteins such as protein disulfide isomerase (PDI) will catalyze the assembly and formation of secreted proteins by stabilizing unfolded proteins and isomerizing the correct disulfide bridge formations.[58] When the native three-dimensional structure is achieved, the protein is ready for transport to its intended location. Secretory proteins are transported from the ER to the Golgi and finally secreted out of the cell.[59] Proteins that are to reside in the ER are retained or transported back to the ER via the KDEL signal sequence.[60] Enzymatic activity and protein function in general, is dependent upon pH. Cytosolic proteins will function at physiological pH, those in the rough ER are in a slightly basic environment and proteins found in degradation organelles like the lysosomes will function under acidic conditions. Figure 21 illustrates the organization of the cell and pathway secreted proteins undergo.
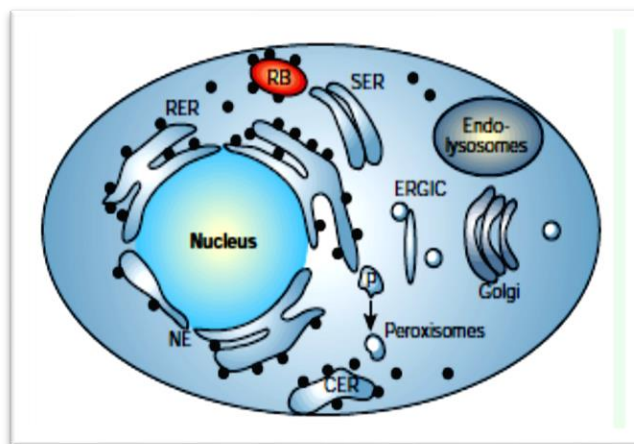
Figure 21: Organization of the cell

Secretory proteins are synthesized on ribosomes, translocated into the ER, sent to the Golgi and secreted[59].

Correct disulfide bond formation is significant to the structure and function of a protein. Without proper disulfide bond formation, the peptide is degraded or can form aggregates that will disrupt the overall function of the cell. When unfolded proteins cannot quickly find their conformation, hydrophobic regions of individual peptides cause insoluble aggregates to form. Not only can aggregation lead to a loss of protein function but the accumulation can lead to cell death. Endoplasmic reticulum associated degradation (ERAD), a process where misfolded proteins are labeled and sent to proteasomes for degradation, helps alleviate protein aggregation.   If accumulation of misfolded proteins occurs and the cell machinery cannot handle the load, the cell will experience stress and may become apoptotic.[61]

Cruzain is considered the main target for the development of anti-Chagastic drugs. The mature enzyme and the prodomain have been well investigated as separate peptides while the full length procruzain has been understudied. The purpose of this project is to determine the involvement and significance of the pro-domain in protein folding of cruzian to impede folding and arrest the progression of Chagas Disease.

33

CHAPTER 2

REFINEMENT OF EXPRESSION, PURIFICATION AND ACTIVATION

OF PROCRUZAIN

2.1   Specific Aim

Specific aim one: To refine the expression, purification and activation method of cuzain.

2.2   Background

Previously, a low yield of mature cruzain was obtained with protocols provided in the literature therefore our goal was to refine our method to produce a higher yield of active cruzain for kinetic studies. Production of procruzain and activation methods were adapted from Eakin et.al.[21]. Refinement was made in the solubilization of the protein, purification methods and to the activating conditions. The enzyme, when activated, is extremely catalytic and is subject to autodegradation, therefore storing conditions were also refined to prevent degradation or inactivation before experiments were conducted. The expression, purification and activation to obtain our protein of interest is described here.

2.3   Materials and methods: Expression, purification and activation

2.3.1   Transformation of *E. coli* DH5 α for Subcloning Efficiency

The plasmid (pQE-30) containing the gene that codes for procruzain with a polyhistidine tag at the N-terminus kindly provided by Dr. Ana Paula Lima from Federal University of Rio de Janeiro, Brazil; Biophysics Institute Carlos Chagas Filho was inserted into competent *E. coli* DH5α for subcloning efficiency and DNA purification. Figure 22 shows the pQE-30 plasmid map highlighting the location of the insert containing a PT5 promotor region, lac operon, multiple

cloning sites, histidine tag (6x), and restriction sites[62]. The ampicillin resistant gene (amp^r) and origin of replication are also shown. The amp^r gene codes for beta-lactamase, an enzyme that hydrolyzes the beta-lactam ring of ampicillin inactivating the antibiotic. The lac-operon regulates gene expressionn[62].

To produce a DNA stock solution for multiple expression trials, 2 uL of the plasmid with the procruzain gene inserted at the MCS provided by A.P. Lima was added to 50 uL of DH5α cells (Invitrogen) and incubated on ice for 30 minutes to allow for mixing. The mixture was heated at 42°C for 50 seconds and placed back on ice for a minimum of two minutes for a heat shock treatment. The transformed cells were added to 200 uL of SOC media, a variant of super optimal broth (SOC) media and incubated under slight agitation for 2 hours at 37°C to allow for recovery. 100 uL of the incubated cells were spread on a lysogeny broth (LB) nutrient agar plate containing 100 uM ampicillin (amp) and 50uM kanamycin (kan) final concentration. The plate was incubated for 14 hours at 37°C then sealed and stored at 4°C to prevent overgrowth and contamination.

The plasmid was extracted and purified from the transformed DH5α cells using the QIAprep Spin Miniprep kit (QIAGEN)[64]. To begin plasmid purification, 5 mL of SOC media containing 100uM amp and 50 uM kan was inoculated with one colony from the previously transformed DH5α cells and was incubated under agitation at 37°C for 14 hours. The cells were harvested by centrifugation at 3500 rpm for 7 minutes with an Eppendorf 5804R Benchtop Refrigerated Centrifuge. The supernatant was discarded and the QIAGEN protocol for plasmid purification was followed. The harvested pellet was solubilized with a buffer and lysed and the cellular debris was removed by spinning the cells at 13,000 rpm for 7 minutes with an Eppendorf 5415R Refrigerated Benchtop Centrifuge. The supernatant was applied to the QIAGEN spin column and eluted with 50 uL of 0.2 um filtered/UV treated Millipore water. The protocol

produced between 10 and 20 ug of DNA.   The purified DNA concentration was quantified with a

NanoDrop instrument and was either directly stored at -20°C or lyophilized until needed for

transformation of bacterial cells and protein expression.
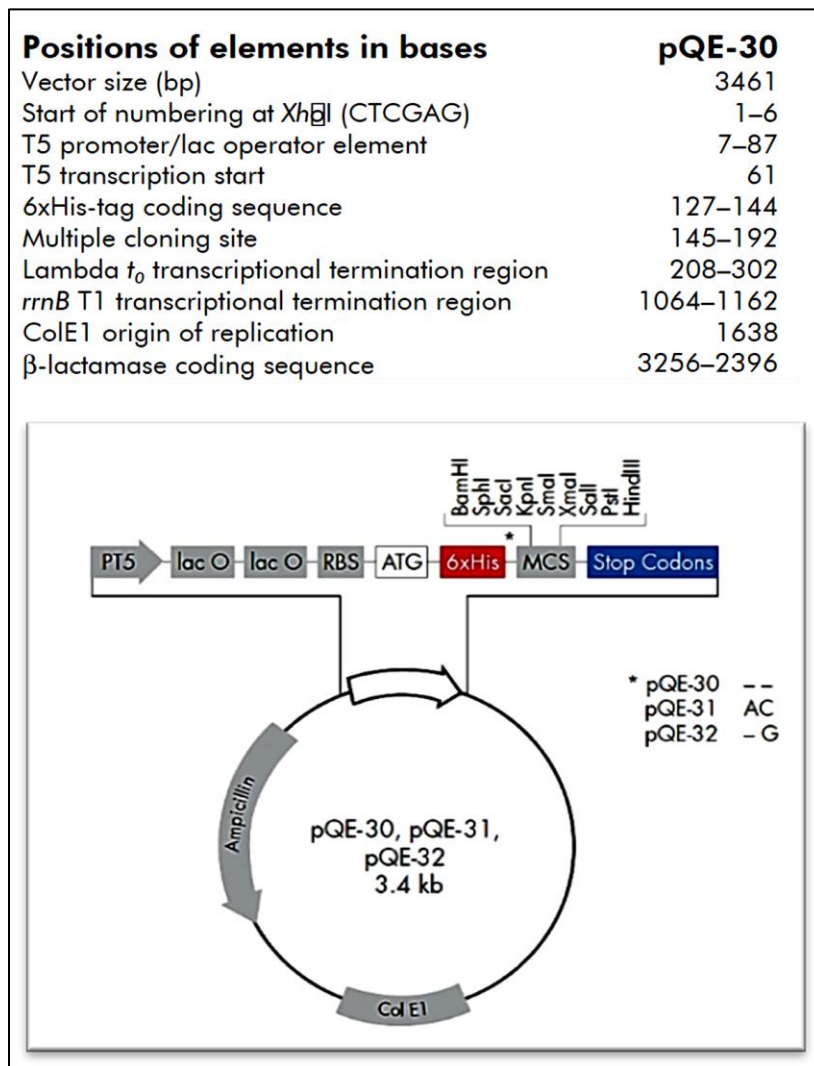
Figure 22: pQE-30 Plasmid Map

Plasmid, pQE-30 (QIAGEN), contained the gene that codes for pczp at the multiple cloning site (MCS)

Chemically competent cells were prepared by growing a 5mL overnight culture of *E. coli*

M15 and BL21 expression cells and DH5α subcloning efficiency cells (Invitrogen). The culture

was diluted the following day in 100 mL LB after growing overnight for no more than 16 hours at 37°C under slight agitation. The inoculated 100 mL culture was incubated at 37°C while shaking until the optical density (OD) reached 0.40. The absorbance was measured at 600 nm with a Thermo GENESYS™ 10UV UV-Vis spectrophotometer. After reaching the optimal OD the cells were harvested at 4°C by centrifugation at 3,700 rpm for 7 minutes. The pellet was resuspended with 10 mL of 0.10 M calcium chloride/1.0 M pipes buffer at pH 7.4 and incubated on ice for 30 minutes. The cells were harvested and the supernatant was discarded. This step was repeated for two more cycles and finally the cells were resuspended in 2.0 mL of $CaCl_2$/pipes buffer, incubated overnight at 4°C, then stored as 50 uL aliquots 1 mL sterile microcentrifuge tubes. Two 50 uL aliquots of cells were immediately used for transformation and the rest were stored at -80°C for up to 6 months. This protocol was used to prepare chemically competent DH5α, BL21 and M15 strains of *E. coli* cells.

2.3.2 Transformation of *E. coli* BL21 and M15 Cells for Expression of Procruzain

E. coli BL21 is commonly used for overexpression of proteins and M15 contain the genetic conditions that allow for expression of proteins found on pQE30 plasmids therefore these strains were each transformed to compare expression levels. 4uL of plasmid obtained from the miniprep extraction was added to 100uL of BL21 or M15 chemically competent cells prepared as in section 2.31. of this manuscript and incubated on ice for 30 minutes to allow for mixing. Heat shock as conducted by heating the mixture at 42°C for 1 minute then placed on ice for a minimum of two minutes. The transformed cells were added to 200 uL of SOC media, a variant of super optimal broth (SOB) media, and incubated under slight agitation for 2 hours at 37°C to allow for recovery. 100 uL of the incubated cells were spread on a lysogeny broth (LB) nutrient agar plate containing

100 uM ampicillin (amp) and 50 uM kanamycin (kan) final concentration. The plate was incubated for 14 hours at 37°C then sealed and stored at 4°C to prevent overgrowth and contamination.

2.3.3   Expression of Procruzian

A small-scale culture was grown by inoculating 100 mL of LB broth containing 100 uM amp and 50 uM kan with a small colony of transformed BL21 or M15 cells. The culture was grown overnight at 37°C while shaking. The small-scale culture was diluted 10 times into 1 L of LB broth containing the same final concentration of antibiotics. The large-scale culture was grown at 37°C and carefully monitored by checking the optical density at 600 nm periodically. Once the optical density of the culture reached 0.5 (approximately 2 hours after dilution), 250 uL of bacteria was removed from the culture and stored as a glycerol stock to reduce the need for additional transformations. A glycerol stock of transformed cells, either BL21 or M15, was prepared by adding 250 uL of the bacterial culture to 250 uL of a 60% glycerol stock for long-term storage at -80°C in a 1.8 mL capacity cryogenic tube (Nunc® cryotube, Sigma-Aldrich). The glycerol stock was prepared by adding 125 uL of 100% heat sterilized glycerol to 250 uL of 0.2 um filtered water. The glycerol stocks were stored at -80°C for future expressions. Glycerol stock preparation ensured the viability of transformed cells during long-term storage while reducing the need for transformations.

After the 1 L culture reached the optimal OD and the aliquot for the glycerol stock was removed and stored, expression was induced by adding 1M isopropyl-B-D-1-thiogalactopyranoside (IPTG) (1 mM final concentration) and was incubated for 4 hours at 37°C. The cells were harvested by centrifugation on a Sorvall Model RC-5C Refrigerated Superspeed Centrifuge and a GSA rotor at 5000 rpm for 15 minutes and the supernatant removed and discarded. At this point the cells were either lysed or stored at -80°C for a maximum of 6 months.

Expression of a mutant of cruzain was done in a similar manner to the expression of procruain. The gene that codes for the mutant lacking the cysteine of the catalytic triad was inserted into pQE-30 with a histidine tag at the N-terminal domain. E. coli (DH5α, M15 and BL21) cells were transformed to produce the mutant. Purification, refolding and storage was done using the protocol below. The C25A mutant was expressed to determine the role of the pro-domain in activation and folding.

2.3.4 Cell Lysis and Purification

The chaotropic agent urea at high concentration (8M) in a lysis buffer (10 mM Tris, 100 mM NaH_2PO_4 pH 8.0) was used to lyse the cells and disrupt the expressed procruzain inclusion bodies. For a 1L culture, 100 mL of lysis buffer (1/10 total volume of inoculated media) was applied to the pellet. The cells were lysed at 4°C under slight agitation for 1 hour. To increase solubilization of the proenzyme and maximize lysis, the cell lysate was sonicated with 40 W output Branson Ultrasonics Sonifier 450 for 3 x 30 second pulses while on ice. The lysate was allowed to recover for 1 minute in between 30 second pulses. A final concentration of 1 uM phenylmethylsulfonyl fluoride (PMSF) was added to prevent degradation of the protein by serine proteases. Active cruzain is a cysteine protease therefore protein cocktail inhibitors (PIC) and cysteine protease inhibitors such as trans-epoxysuccinyl-L-leucylamido-(4-guanidino) butane (E-64) and leupeptin were avoided during lysis when activation was desired otherwise were added to a 1 uM final concentration. The cellular debris was removed by centrifugation at 11,500 rmp for 20 minutes using a Sorvall Model RC-5C Refrigerated Superspeed Centrifuge with a SS-34 Superspeed centrifuge rotor. The supernatant containing procruzain was applied to glass Econo-Column® Chromatography low pressure/gravity flow column containing 3mL divalent nickel

coordinated nitrilotriacetic acid agarose resin (Ni-NTA His•Bind®, EMB Biosciences) pre-equilibrated with lysis buffer for purification.

The Ni-NTA metal chelating resin stored in 20% ethanol was rinsed with water and equilibrated with lysis buffer containing 10mM imidazole. The mixture was incubated with the supernatant at room temperature under slow agitation for 20 minutes to allow for binding then added to the gravity column. The flow through was released, collected and stored at 4°C. The beads were washed with 5 mL of wash buffer (10mM Tris, 100 mM NaH_2PO_4, and 8 M urea pH 6.3) for a total of 5 washes. The protein was eluted in 2 mL aliquots with elution buffer (10 mM Tris, 100 mM $NaHPO_4$, 8.0 M urea, 200 mM Imidazole at pH 8.0) for a total of 15 fractions. The nickel resin was be cleaned by incubating the beads with 5 mL of 0.1M acetic acid-acetate buffer at pH 4.0 for 5 minutes then re-equilibrated with excess lysis buffer. After 5 uses, the resin was regenerated as directed by the supplier. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) described below was used to examine purification. Pure fractions were collected and stored at -20°C until protein folding studies were conducted.

Aliquots (25 uL) of each stage of the purification process were analyzed by SDS-PAGE. The samples were mixed with and equal volume of 2x sample buffer under reducing and denaturing conditions (4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris HCl, pH 6.8) and incubated at 90°C for 2 minutes. The samples were allowed to cool and 15 ul of each were loaded on a hand casted 12% polyacrylamide gel (37.5:1) along with 5 uL of Precision Plus protein standard (BioRad) then applied to a BioRad Mini-Protean II vertical electrophoresis system. Migration of protein samples was achieved by applying 100 V of current until the dye front of the sample buffer reached 1cm from the bottom of the gel. The current was provided by a BioRad PowerPac Basic Power Supply. The gels were stained with

Coomassie dye (Sigma-Aldrich) containing 1 g of Coomassie Brilliant Blue R250, 100 mL of glacial acetic acid, 400 mL methanol and 400 mL of Millipore treated water and placed on a platform rocker for 20 minutes. The gels were rinsed with doubly distilled (ddH$_2$O), resolved with destaining solution (staining solution lacking Coomassie powder) and viewed on a UVP LMW20 UV/White Light Transilluminator. Images were taken with a UVP BIODOC-IT UV Transilluminator Imaging System.

2.3.5   Refolding of Procruzain

The pure aliquots shown by SDS-PAGE were pooled together and the concentration was determined with the BioRad DC Protein Assay Kit. This assay for protein concentration determination was used in place of Bradford assay or Lowry method because it is compatible with a high concentration of urea (4 M urea and above). 10-15 mM dithiothreitol (DTT^{red}) per milligram of purified protein was added to the pooled fractions and incubated at 37°C under slight agitation for 1 hour to fully reduce disulfide bonds.

Procruzain was refolded by diluting the reduced and denatured protein 100 times in refolding buffer (250 mM arginine, 100 mM Tris HCl, 1 mM ethylenediaminetetraacetic acid and 20% Glycerol, pH 8.0) and incubated under slight agitation at 4°C overnight. The solutes were partially removed by displacing the refolding buffer with 1L of phosphate buffer using an Amicon Dialfiltration unit. The sample was concentrated to 20 mL, the concentration determined with the Bradford assay and prepared for activation. Dialysis was avoided due to precipitation of the enzyme.

2.3.6 Activation and Purification of Mature Enzyme

Precipitates that may have formed within the purified product were removed with a 13 mm, polyvinylidene fluoride, 0.22 um pore size syringe filter (Whatman). Initial activity of the of the purified protein was analyzed by fluorescence spectroscopy on DM45 Spectrofluorimeter (On-line Instrument Systems, Inc.) using Spectra Works software in a quartz cuvette (Fisher Scientific) with a 1cm path length containing four transparent walls. Fluorescence was recorded as a function of time at 380nm excitation and 440 nm emission wavelengths set for 0.100 integration time with 600 photomultiplier tube (PMT HV) for 300 seconds at 37°C. To detect initial activity of the protein, 50 uL of refolded procruzain was added to 2 mL of activity buffer containing 30 uM of the fluorogenic substrate N-carbobenzyloxy-phenylalanine-arginine-7-amino-4-methyl coumarin (Z-F-R AMC, BACHEM), 10 mM DTT^{red} and 100 mM sodium acetate (Sigma) at pH 5.5. It was expected that refolded procruzain would exhibit little to no activity prior to activation as indicated by a lack of fluorescence emission.

Previously described activation methods[21] were ineffective in producing workable quantities of active enzyme therefore refinement of the protocol was necessary. The refolded protein was mixed with 1:1 2X activation buffer (200mM NaCH$_3$COO, 1.8M NaCl, 10mM DTT^{red} and 10mM EDTA, pH 5.2) and incubated at 37°C for 2 to 4 hours. 50 uL aliquots of activation mixture was added to 2 mL of activity buffer and applied to the fluorimeter every 30 minutes to check activation progress. The test samples were lyophilized and analyzed on a 15% SDS-PAGE and compared to a Fisher Bio-Reagents EZ Run protein ladder to visualize the amount of procruzain remaining in the sample over the collected timepoints. As maximum activity was observed, the mature enzyme was inhibited with a final concentration of 10 uM leupeptin or 1 uM

E-64 (Sigma), a reversible cysteine protease inhibitor, to prevent self-digestion of active cruzain. Previous work shows the active protein has a molecular weight of 26.5 kDa[21,42].

After activation, additional impurities were removed by reverse phase HPLC a Discovery® BIO Wide Pore C5 Column (Sigma) column eluted with a propanol gradient (0-60%) or ion exchange chromatography (Mono Q column, GE Healthcare) eluted with a NaCl gradient pH 6.5 (0-1M). High salt concentration can affect protein affinity to the column therefore the sample was desalted using gel filtration on a G-25 column (GE Healthcare), filtered with a sterile 0.22 um syringe filter and immediately applied to HPCL or IEX columns. Ultra-pure cruzain was obtained by applying to a gravity column containing 1 gram of thiopropyl sepharose resin (GE Healthcare) equilibrated with activation buffer lacking DTT[red], eluted with 25 mM DTT[red] at pH 5.6.

Purification of cruzain by revers phase HPLC was done by applying 1.5 mL (35 uM) of the filtered protein to a C5 (Sigma) column with a 0 to 60% n-propanol gradient in 0.1% formic acid[48] at a flow rate of 0.3 mL/min detected at 280 nm. Purification using ion exchange chromatography was done by applying the desalted sample onto a Mono Q anion exchange column (GE). The protein was eluted with a 0 to 1 M NaCl gradient at pH 5.6 and a flow rate of 1.0mL/min detected at 280nm. The sample was collected from C5 column, concentrated with a Millipore-Sigma™ Amicon™ Ultra-0.5 centrifugal filter unit, lyholized and run on 15% SDS-PAGE to determine column retention times of cruzain and procruzain. Samples collected from the anion exchanger, Mono-Q, were desalted on a G-25 column at 1.5 mL/min flow rate with 1% formic acid before lyophilization.

2.4   Results and discussion

2.4.1   Transformation of *E. coli* subcloning and expression strains

The *E. coli* strains used for expression and subcloning produced a high level of transformation efficiency. To prevent overgrowth of colonies, transformed cells spread on LB agar plates were restricted to an incubation period 14 hours. Overgrown colonies were not further utilized since they are not properly isolated and often grow into other colonies which can lead to inactivation of the antibiotic resistance genes. Production of small satellite colonies stemming from overgrown colonies can evolve to survive in the presence of antibiotics even if they lack the plasmid that codes for resistance. These satellite colonies will not express the protein of interest. To prevent this from occurring, small colonies were incubated for a maximum of 14 hours. No green fluorescent protein (GFP) or other florescent tags were co-expressed with the procruzin to maintain the integrity of the protein structure. Though very useful in identifying successfully transformed cells, GFP tags were omitted because mature cruzain may degrade the tag and create unwanted waste increasing the number of purification steps while reducing yield.

Chemically competent cells were used for both types of transformation. It is believed that the calcium in the buffer prepares the cell for uptake of the plasmid by making the bacterial cell membrane more permeable. Heat shock also increases permeability of the cell however heating at a temperature higher than 42°C or for a long period of time may kill the cells giving less expression product. To reduce the need for frequent transformations, cells were stored in a glycerol stock. Glycerol protects the cells that are being stored at -20 and -80°C. Aliquots of purified DNA and transformed cells were stored in individual sterilized cryogenic tubes to prevent freeze-thaw damage that large stock solutions may undergo.

2.4.2 Expression and Purification of Procruzain

The inactive protein was expressed as a proprotein containing the pro- and catalytic domains. The genetic organization of the non-recombinant protein is shown in Figure 23. The c-terminal extension is not required for catalytic function or folding and the signal peptide is not compatible with prokaryotic expression systems therefore the two domains were omitted from the DNA construct. The prodomain was degraded by autoactivation under the appropriate conditions (see activation). Table 4 provides the expected molecular weights of each domain provided by a molecular weight calculator based on the amino acid sequence of the protein (http://www.bioinformatics.org/sms/) discovered by Eakin et.al.[21]
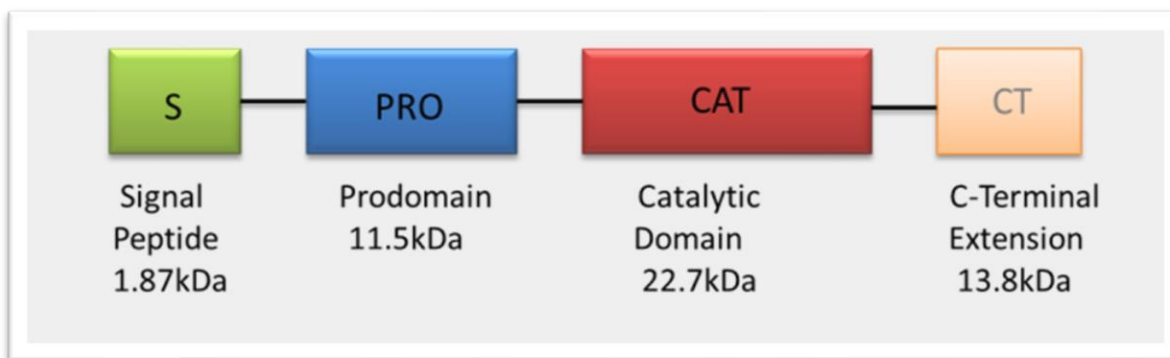
Figure 23: Sequence organization of recombinant procruzain

The protein used in our studies lack the c-terminal extension

Table 4: Estimated molecular weights of cruzain domains.

Region	Molecular Weight (kDa)
S	1.87
S + PRO	13.4
S + PRO + CAT	36.1
S + PRO + CAT + CT	49.8
CAT	22.7

Procruzain was expressed as an inclusion body and was solubilized sonification in addition to lysis with 8M urea. The addition of sonification to our method based on published protocols helped improve our yield. A significant amount of proenzyme with high purity was obtained (Figure 24) after lysis and purification. The fractions were pooled, divided and stored for future studies. Figure 25 is a picture of the SDS-PAGE results of a second expression and purification off a nickel chelation chromatography eluted with 2mL aliquots of elution buffer. A dominant band believed to be procruzain can be seen in both figures. This band is estimated to have a molecular weight of approximately 36 kDa which correlates well with the estimated molecular weight of procruzain given by the molecular weight obtained from the literature[42] (Figure 23, Table 4). In Figure 25 a faint band above the 25 kDa marker may represent a very small quantity of active enzyme. In fact, some gel results showed a faint band near 10kDa possibly representing autocatalytic degradation product, the prodomain. For use in subsequent experiments, no additional purification was necessary. The pH of the samples was maintained at pH 8.0 and immediately stored at 20°C to prevent full activation. Because of the possibility that cruzain is co-eluted from the column, folding studies were conducted soon after to prevent degradation of the proenzyme.
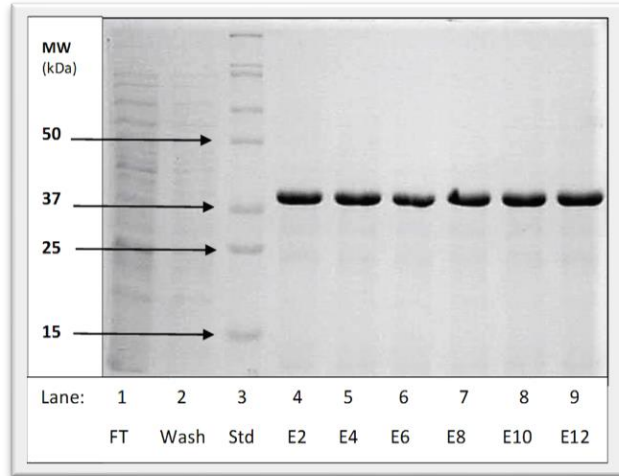
Figure 24: Procruzain purification results.

12% SDS-PAGE stained with Coomassie blue results of elutions off the NiNTA colunm, Elution 1 (E1) to Elution 12 (E12) collected from the nickel column. A BioRad molecular weight maker was used to estimate the molecular weight of the expressed protein. Molecular weights are labeled with arrows.
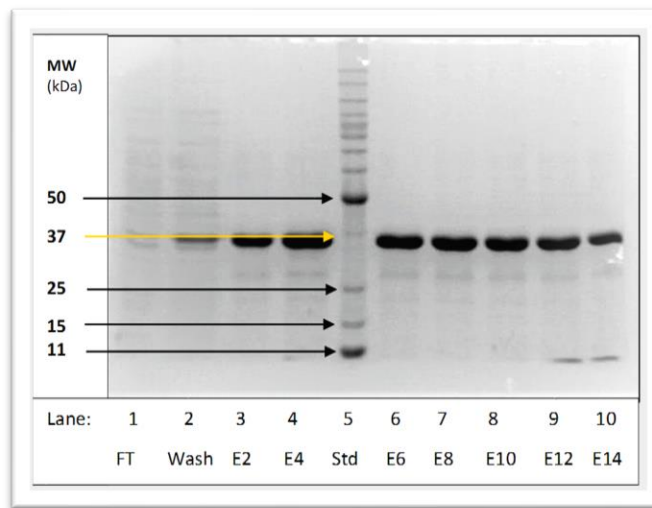
Figure 25: Additional procruzain purification results.

12% SDS-PAGE stained with Coomassie blue results of elutions off a NiNTA colunm, Elution 1 (E1) to Elution 14 (E14) collected from the nickel column. This picture shows a slight impurity above 25 kDa Molecular weights are labeled with arrows.

47

To increase purity, 10mM imidazole was added to the buffer used to lyse the cells, solubilize the protein, and equilibrate the column. Using too much resin can increase non-specific binding therefore excess protein was applied to the column. A high concentration of urea (8 M) was added to all the purification buffers (lysis/equilibration, wash and elution) to prevent precipitation of the protein. Dialyzing the sample for removal of solutes resulted in precipitation therefore the concentration of the solutes was reduced by dialfiltration.

An activity check was conducted on refolded protein to check for initial activity. The fluorogenic substrate, N-carbobenzyloxy-phenylalanine-arginine-7-amino-4-methyl coumarin (Z-Phe-Arg-AMC) is cleaved at the amine group of the arginine residue by the cysteine protease, which releases the AMC group. Fluorescence is caused by the conjugated pi system of electrons, resulting delocalization allowing electrons to absorb energy and reach an excited state where fluorescence is emitted. Within in the substrate, the fluorogenic AMC group is quenched buy the arginine residue in the dipeptide substrate, however once cleaved by the protease, fluorescence can be observed. Figure 26 is the structure of z-Phe-Arg AMC with the fluorogenic region highlighted. Fluorescence released by the cleavage of the peptide at the AMC region was detected with a fluorescent spectrophotometer at an absorbance of 380nm and emission at 440 nm.
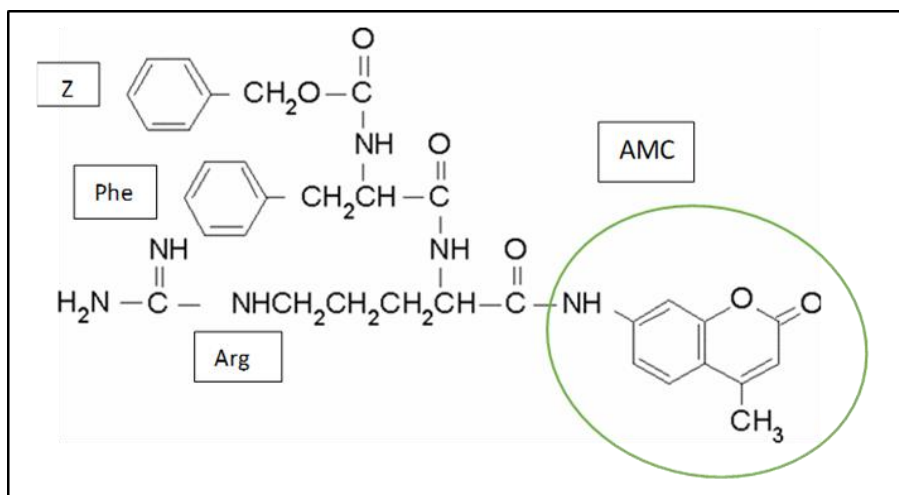
Figure 26: Structure of Z-Phe-Arg-AMC

All regions of the substrate are labeled, the fluorogenic region, AMC, is highlighted in green

The results of the fluorescence assay seen in Figure 27. The slight fluorescence may be due to the presence of small amount of mature enzyme. This information correlates with the gel picture seen in Figure 26. A light band with approximately 27 kDa in size was visualized on the gel which could be the producing the slight fluorescence. The enzyme showed a very slight and not full activity because the majority of the enzyme remains in the inactive zymogenic form until the optimal conditions for activation are optimized, however after refolding a small amount has reached maturation. The sample was compared to a negative control containing activation buffer containing the substrate lacking the protein. The control shows no increase in fluorescence hence no activity. When compared to active protein (Figure 28), the amount of fluorescence produced by premature activation (after Ni NTA purification and refolding) is minimal however the presence of active enzyme, even in a very small amount, can trans-catalyze the maturation of additional protein over time if left uninhibited or if not stored in conditions that prevent activation.

49
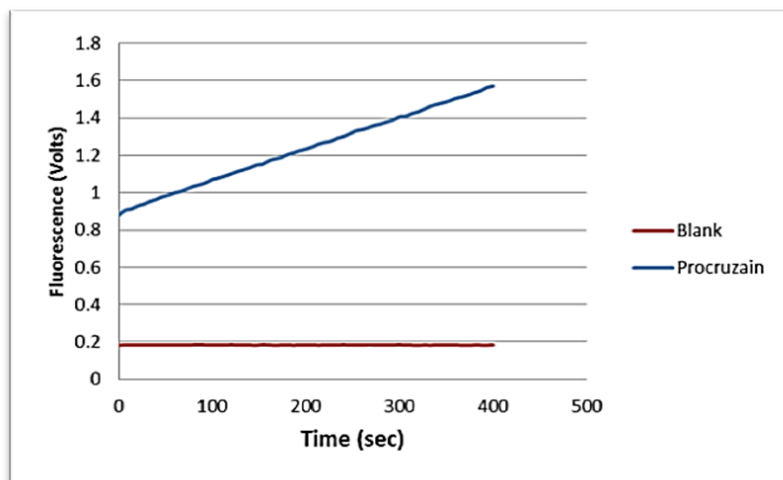
Figure 27: Initial activity check of purified procruzain.

The activity of the purified protein was analyzed through fluorescence spectroscopy. Fluorescence was analyzed at an excitation wavelength of 380 nm and emission wavelength of 440 nm.

Though fluorescence is still detected after refolding, it is very minimal in comparison to the fluorescence produced by the fully matured enzyme. The graph (Figure 27) shows an increase in fluorescence over the time assayed to be less than difference of 1 fluorescence volt and often produces even less of a difference. In comparison, active cruzain gives a difference of over 5 fluorescence volts. The next section will describe the comparison of the activity in procruain and cruzain.

2.4.3   Activation of Cruzain

Activation was detected by fluorescence detected by the hydrolysis of the fluorogenic substrate, z-Phe-Arg AMC. Once the maximum amount of fluorescence was observed, an inhibitor was added to the sample to prevent self-degradation. The fluorescence released from the hydrolysis of the substrate was significantly larger compare to the activity check done in freshly purified protein (after refolding). Figure 28 is the chromatogram of the maximum activation

50

achieved as well as the inactivation by leupeptin. The maximum activation was achieved after 4 hours of incubation at 37°C which remained constant for up to 6 hours. After 6 hours, the protein showed a decline in activity due to autodegradation. At 4 hours the protein was highly active (Figure 28) and incubation with the inhibitor at 37°C rendered the enzyme inactive. A 500uL aliquot of the protein sample was removed at different time points and applied to electrophoresis to determine if the active protein possessed the expected molecular weight of active cruzain and to also visualize the activation progress over time (Figure 29). By the last time point, the protein was fully activated and the prodomain (approximately 11 kDa) that is believed to cover the active site of the proenzyme was cleaved (Figure 30). The size difference between the mature and proprotein along with the results of fluorescence assay are proof that the protein has matured.
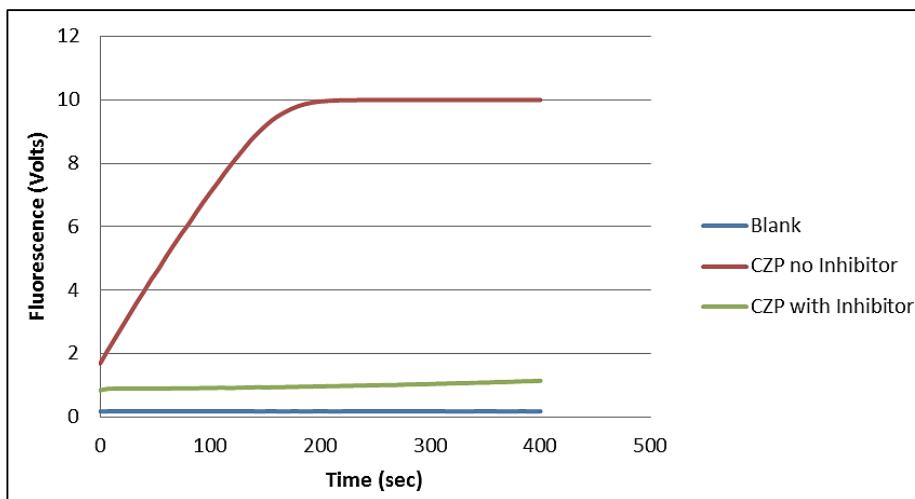
Figure 28: Fluorescence analysis of active and inhibited cruzain

The activity check of mature enzyme (red) is compared to a blank (blue) and enzyme with an inhibitor (green)
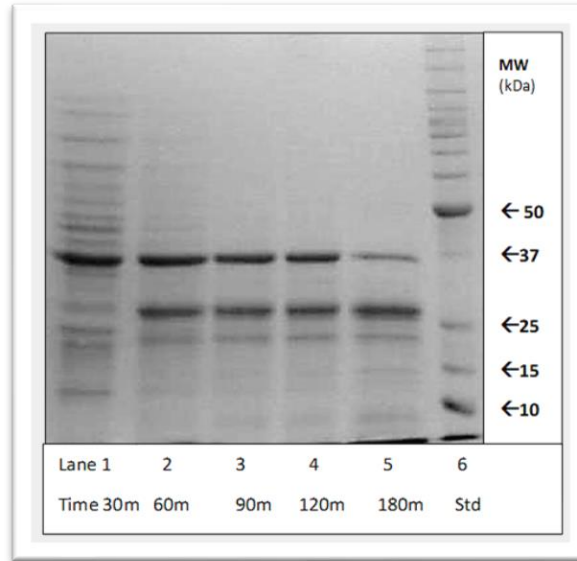
Figure 29: Activation progress

Samples collected at different time points were analyzed on 15% SDS-PAGE. Different time points were collected to track activation progress. Lane 1: Time = 30 min after initial activation, Lane 2: Time = 60 min, Lane 3: Time = 120 min, Lane 5: Time = 180 min and Lane 6: Molecular weight marker.
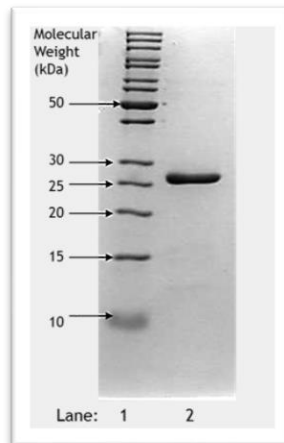
Figure 30: Full activation of cruzain

Activation after 4 hours. Lane 1: MW marker (Fisher EZ-Run ladder). Lane 2: active cruzain

After activation, most of the procruzain matured to the active enzyme. The gel seen in Figure 29 shows a decrease in the 36 kDa proenzyme and an increase in cruzain at approximately 27 kDa. Along with the fluorescence assay, the result show that the prodomain has been removed from the proenzyme. Since the prodomain is small (11 kDa) and subject to additional degradation by the mature enzyme, it is difficult to visualize it on SDS-PAGE however very faint bands can be seen above the 11kDa protein marker in the ladder and a band is also visible near 15 kDa. These may be degradation products that remain after the autocatalytic activity of the enzyme is complete.

In previous activation attempts, very little mature enzyme was produced with protocols found in the literature[28]. The amount of mature protein produced was less than the nanogram range and only resolved with silver staining after running a 12% gel. To improve our yield, we optimized the conditions of expression and activation. The protein was expressed as inclusion bodies and a high concentration of urea was added to harvested cells to help dissolve the inclusion bodies and lyse the cells. To release a higher concentration of protein from the cells, sonic waves were applied to the cells containing lysis buffer. In addition, 10 mM imidazole was added to the lysis buffer to prevent non-specific binding on the nickel column. Activation methods using different conditions were conducted to determine the optimal conditions for activation. Concentrations of DTT^{red}, urea and sodium chloride were manipulated. The optimal pH was found to be between 5.2 and 5.5 and activation was initiated at a pH of 5.3. Though the role of the buffer is to maintain pH, over time, the pH of the protein drifted due to the maturation of the acidic proteinase, therefore pH was checked every hour to ensure that the pH was maintained at the optimal range. Procruzain is moderately hydrophobic therefore a high concentration of sodium chloride was added to the activation buffer to prevent aggregation of the protein. They high hydrophobic index of proruzian is also the reason dialysis was avoided. Approximately 10 mg of procruzain was produced from

1L of cell culture and 100 ug of mature enzyme was produced after activation. Both procruzain and cruzain were stored at -20°C and used for kinetic studies describes in the next chapters.

High concentrations of sodium chloride and urea were used in buffers at all stages of expression and purification to prevent insoluble aggregates from forming during purification. The auto-catalytic activation of cruzain is not well understood however the process is highly pH dependent. Many cysteine proteases are often localized in lysosomes where the pH is often below physiological conditions. Recent studies conducted on the mechanism of activation of the cathepsins suggest the activation is both a unimolecular and a bimolecular process meaning the initial activation is done by the zymogen alone which in turn, activate other inactive enzymes in a chain reaction. A low pH disrupts salt bridges within the pro-domain of the protein by protonating carboxylate groups which can cause change in structure that prevents hydrophobic interactions. The changes in tertiary structure likely removes the pro-domain from blocking the active site, allowing the active site to cleave the prodomain. Figure 31 illustrated the significance of pH on the tertiary structure and enzyme activity of falcapain-2, a papain family cysteine protease found in *Plasmodium falciparum* parasites.[29]
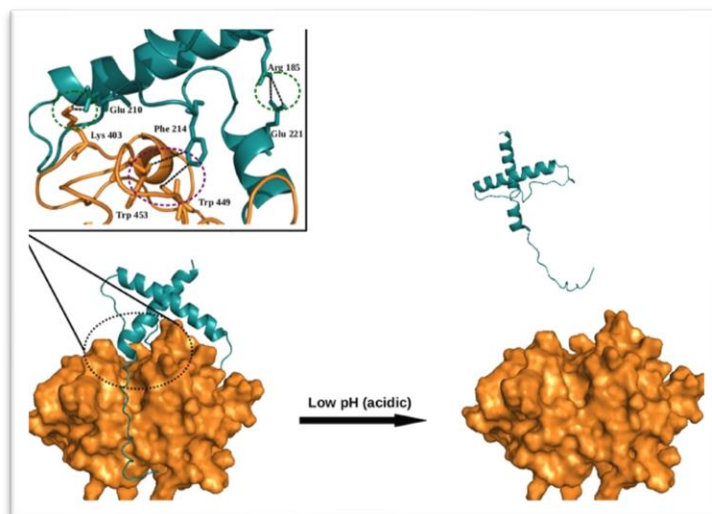
Figure 31: Interactions between pro-domain and catalytic domain of falcapain-2

The green dotted circle highlights the salt bridges allowing the pro-domain to black the active site. Reducing the pH disrupts these salt bridges and as a result the pro-domain is cleaved.[29]

Reduced and native protein were each applied a reverse phase Discovery BIO Wide Pore C5 column (5um pore size, Sigma) and an ion exchange column to determine retention times for future studies. Figure 32 is the chromatogram of the purification using the HPLC C5 column. Initially, a C18 column was used to attempt separation of the peaks, however it was difficult to elute the protein off the column due to strong interactions of the hydrophobic residues exposed on the surface of the enzyme with the column. Procruzain and mature cruzain were then applied to the C5 (less hydrophobic) column which was eluted with a gradient of n-propanol. The peaks were collected and run on 15% SDS-PAGE, Figure 33. It was difficult to separate reduced from native enzyme however we were able to identify the peaks on the chromatogram. With further manipulation of the method, we plan to visualize all the intermediate species during oxidative folding and applying aliquots over of the 24 hours to HPLC to calculate folding rates constants.
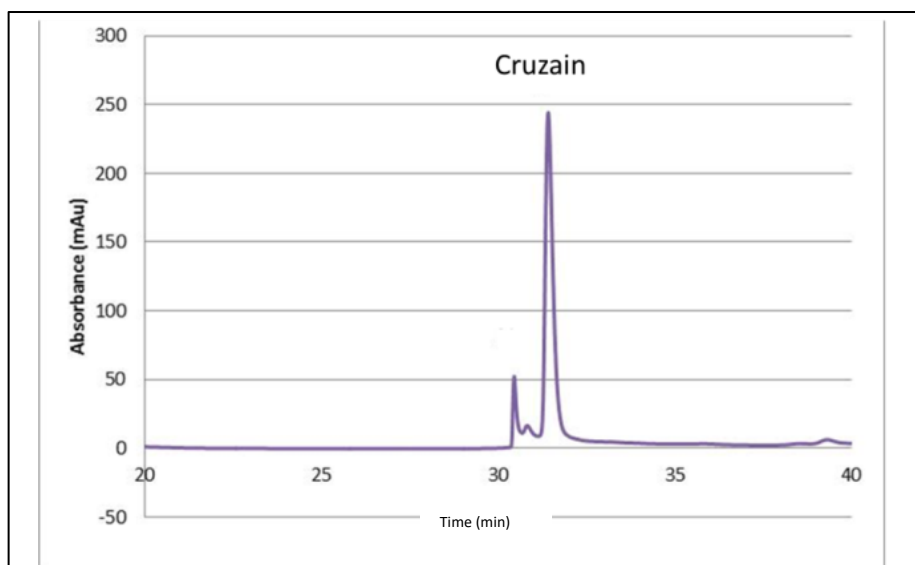
Figure 32: HPLC Chromatogram of active cruzain.

The small peak is the native protein that has a slightly shorter retention time than the reduced form of cruzain.
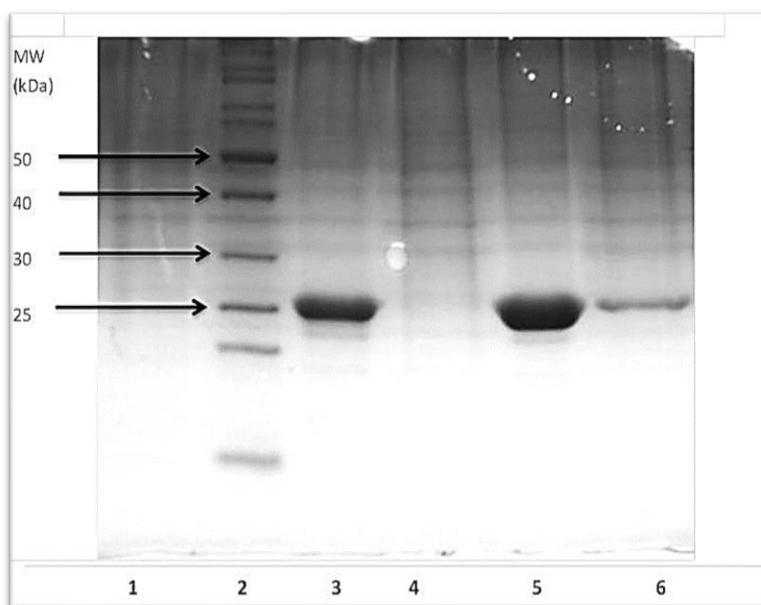
Figure 33: Purified Cruzain on HPLC C-5 Column

Lane 1: Empty well, Lane 2: Molecular weight marker, Lane 3: HPLC purified native cruzain, Lane 4:

Empty well, Lane 5: Concentrated HPLC purified native cruzain, Lane 6: HPLC purified reduced cruzain.

Figure 34 represents the superimposed chromatogram of the purification of native and reduced cruzain by ion exchange chromatography, Mono Q anion exchange column. Native and reduced cruzain elute at similar times and are difficult to separate. Reduced and native procruzain had similar retention times. The gel picture, Figure 35, is the large peak collected from the HPLC and run on 15% SDS-PAGE proving the large peak obtained from the column is mature cruzain. The reduced peak was also run and visualized on a gel.
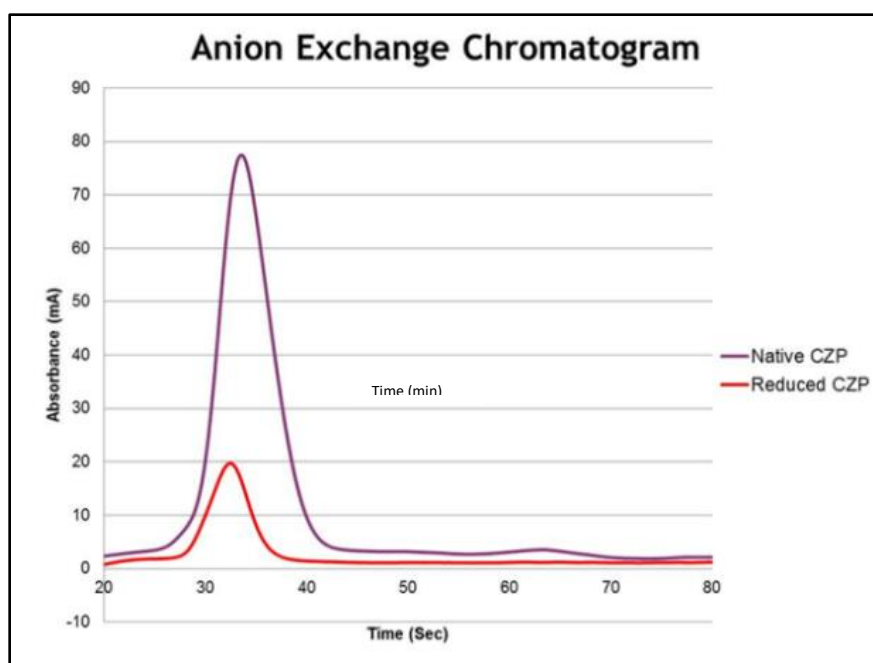
Figure 34: Purification by anion exchange chromatography

The red peak is reduced cruzian and the purple peak is native cruzain.
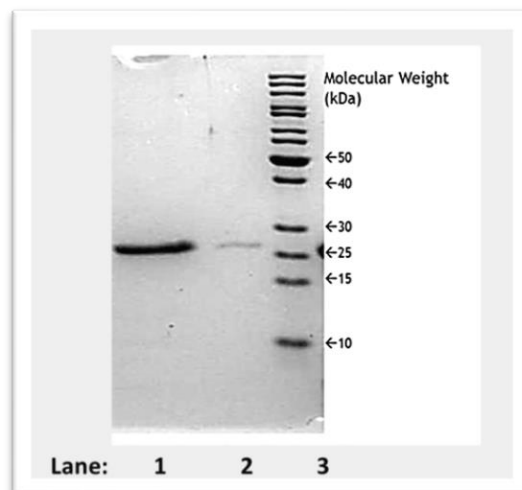
Figure 35: Purification of CZP by anion exchange chromatography

After separation on IEX, the samples were collected and applied to a 15% SDS-PAGE. Lane 1 cruzain before purification. Lane 2 cruzain after purification on a Mono Q column.

Purified protein was then collected and used for studies that are discussed in the next chapters. Several challenges expressing, purifying and activating the enzyme were experienced however, the method of purification and activation were refined to obtain a higher yield of protein. The high degree of hydrophobicity of the protein contributed to the formation of insoluble aggregates thus a loss of protein. These aggregates were prevented by keeping the sample in an 8.0 M urea buffer at pH 6.5 until the was to be used. While further purification of mature enzyme was not necessary, the samples were applied to the HPLC to determine retention rates for oxidative protein folding studies. The goal was to compare folding rate constants of procruzain and cruzain and determine the role of the prodomian in protein folding however this method was not effective therefore a novel method was developed to overcome these challenges. Several expressions were done successfully and the protein produces was used to refine the methods and subsequent expression were used for protein folding studies.

CHAPTER 3

DEVELOPMENT OF A NOVEL METHOD TO DETERMINE FOLDING

RATES

3.1   Specific Aim

Specific aim two: To develop a novel technique to compare the folding rate of procruzain and

cruzain.

3.2   Background

Traditionally, folding rates are calculated by UV spectroscopy. For thiol containing

proteins, the rate limiting step during oxidative folding is the formation of disulfide bonds.

Proteins that do not contain disulfide bonds often fold in milliseconds.[55] In comparison, those that

do contain disulfide bonds will fold slowly. Within the cell, oxidizing agents and chaperones assist

with folding to prevent aggregation and malfunction. The more disulfide bonds a protein contains,

the longer it will take to obtain its native structure due to constant isomerization of the disulfides

until the native structure is achieved. We can utilize this rate limiting step to help us detect folding

progress over time, *in vitro*.

A denatured protein lacking disulfide will achieve its correct 3-dimensional structure as

soon as denaturants are removed (often by dilution). The amount light absorbed by the protein

will change, often decreasing as hydrophobic residues with aromatic rings evade an aqueous

environment. The folding rates can be calculated by analyzing the absorbance of a protein at

280nm assuming the protein contains tryptophan residues (and phenylalanine and tyrosine to a

lesser extent). Intrinsic tryptophan fluorescence can also be used to detect protein folding at 280

nm excitation and 350 nm emission. Exposed tryptophan quenches fluorescence therefore, florescence should increase as the protein folds depending on the identity of the protein.[65]

A method described by Narayan et. al[66-68] utilizes disulfide bond formation as a rate limiting step to calculate folding rates. The method collects aliquots at different time points during oxidative folding until the folding is complete. The aliquots collected are then applied to an HPLC and elution off a reverse phase column (often C-18) is detected at 280 nm. Reduced and native proteins often have different retention times. The reduced protein is retained on the column longer due to exposure of hydrophobic resides, while native protein will have less hydrophobic resides available to interact with the column. Over time, the peak representing the quantity of reduced protein will decrease and the native peak will increase. *In vitro* conditions are made to mimic the cell however the cell is much more efficient at oxidative folding provided individuals are healthy. The data obtained from kinetic studies can provide valuable information about the protein such as structure, enzyme activity and malfunction.

Though the method previously describe is exceptionally useful in observing oxidative folding, highly hydrophobic proteins will not benefit from this type of study. Protein aggregation can be one problem while strong attraction to the column can is another. Finally, sensitivity may also pose a problem due to low concentration of proteins. Even though the method for activation of cruzain was refined (Chapter 2), precipitation of the both procruzain and cruzain occurred, therefore a method with high sensitivity is was developed. Equally important, procruzain and cruzain do not elute off a of reverse phase chromatography or ionic exchange columns easily therefore a new method was developed

3.3　Materials and method: Determination of folding rates

3.3.1　Oxidative Folding of Ribonuclease A

A model protein with a fully known folding patter was used to develop a novel method for the determination of folding rates. Bovine Pancreatic Ribonuclease A (Sigma, molecular biology grade, >70 Kunitz) was reduced and denatured with a reducing buffer (8 M guanidine HCl, 30 mM DTT^{red} and 100 mM Tris buffer pH 8.0) by incubating 2mg of ribonuclease A in 5 mL of reducing buffer and incubated for 45 minutes at room temperature. Refolding was prevented by decreasing the pH to 4.0 with an acid quench of glacial acetic acid (20 uL). The reduced protein was dialyzed in a 3.5 kDa cellulose dialysis tube (Spectra Por) against 0.010 M acetic acid buffer under slight agitation at 4°C for 3 buffer changes. After dialysis was complete, the sample was filtered with a 0.22 um syringe filter (Whatman) and the concentration was calculated using Beer's law by analyzing at an absorbance of 280 nm.

Reduced RNase A (30uM final concentration) was added to a refolding buffer containing 0.50 M urea, 5% glycerol and 125 mM arginine at pH 8.0 for a total volume of 10.0 mL. 2.0 mL of the reaction mixture was separated for HPLC studies conducted in parallel to the novel method. 250 uL aliquots were collected every hour for 20 hours. These samples were reacted with Ellman's reagent as described below to determine the extent of folding for each time point.

Oxidative folding was carried out in parallel on the HPLC using an Agilent HPLC and autosampler in parallel to the method described above to verify the results of the novel method. Reduced and native RNase A were run on a reverse phase HPLC C-18 (Sigma Discovery) column to determine retention times and extent of folding over time. 100 uL samples were run on the HPLC every hour for 16 hours. The time points were eluted with a 0-100% acetonitrile gradient in trifluoroacetic acid at a flow rate of 0.40mL/min. The area under the peaks were calculated and

plotted versus time and the equation of the line provided the rate constant. The rate constant obtained by the novel method was compared to the HPLC method to determine validity.

3.3.2 Procruzain and cruzain denaturation and reduction

To reduce the protein reducing buffer (8 M urea, 30mM DTT^{red} and 100mM Tris buffer pH 8.0) was added to the protein and incubated at 37°C for 2 hours. After the 2-hour incubation period, the pH was decreased to 4.0 to prevent refolding. The reducing agents and salts were removed with a G-25 size exclusion column by eluting with 20 mM formic acid at pH 4.0. The protein concentration was determined with the Bradford assay (BioRad) and bovine serum albumin as the standard. The absorbance was recorded at 595 nm and the concentration of the protein was calculated with the Beer-Lambert law then stored at -20˚C. The increasing concentrations of BSA were used for the standard curve were: 0.125 mg/mL, 0.500 mg/mL, 0.750 mg/mL, 1.00 mg/mL, 1.50 mg/mL and 2.00 mg/mL. The removal of the solutes from the protein lead to aggregates therefore the processing of the protein was be done carefully to prevent loss of yield. The protocol above is the most effective in removing the denaturing and reducing agents while preventing precipitation of the protein. Dialysis was avoided when possible, however when it was the only means for removal of solutes, dilute sample was placed in 3-5 kDa a cellulose membrane dialysis tube (Spectra Por) against an acetic acid/acetate buffer at pH 4.0 (3.0 L) with three changes of buffer once every 8 hours at 4°C under slight agitation. The figure below is a diagram of the denaturing/reduction process.
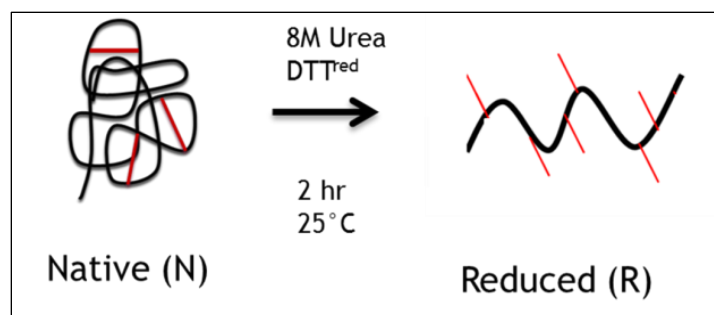
Figure 36: The reduction of native protein

Reducing buffer was added to the protein to denature and reduce procruzain and cruzian.

3.3.3   Ellman's test for determination of folding rates of procruzain and cruzain

Both procruzain and cruzain were reduced and refolded to determine the rate of folding and to identify any stable intermediates during regeneration. Regeneration devoid of oxidizing and reducing agents was essential since DTNB reacts equally with both. In place of oxidizing agents like glutathione or DTT^{ox}, refolding was done by adding a low concentration of chaotropic agents and salts. The proteins were then allowed to fold by air oxidation. Regeneration was done by adding 1:1 refolding buffer containing 250 mM L-arginine (Sigma), 100 mM Tris HCl (Sigma), 1mM EDTA (Fischer Scientific) 1 M urea (Sigma) and 20% Glycerol (Fischer Scientific) to the desalted, reduced and denatured protein for a final concentration of 30 uM and the pH was adjusted to 8.0. The protein was exposed to air and refolded at room temperature for 15 hours under slight agitation. 500 uL aliquots were collected and stored at -20°C. Oxidative folding was analyzed with fluorescence spectroscopy and with liquid chromatography.

500 uL time points were reacted with excess Ellman's reagent, 0.4 mg/mL final concentration, in a reaction buffer (100 mM Tris HCl 1 mM EDTA pH 8.0) and incubated at

25°C for 15 minutes. After the incubation period, the absorbance was taken at 412 nm (molar extinction coefficient 14,500 M^{-1}· cm^{-1}) on a Genesys 10 Visible Spectrophotometer. The folding rate was calculated by finding the slope of the line of best fit when plotting the absorbance versus concentration. Oxidative folding lacking oxidizing agents was conducted for the mature and the proenzyme. The average of the folding constants was calculated and compared to determine the role of the pro-domain in oxidative folding. Figure 37 is a schematic of the reaction and the protocol developed to determine folding rates constants for cruzain and procruzain.
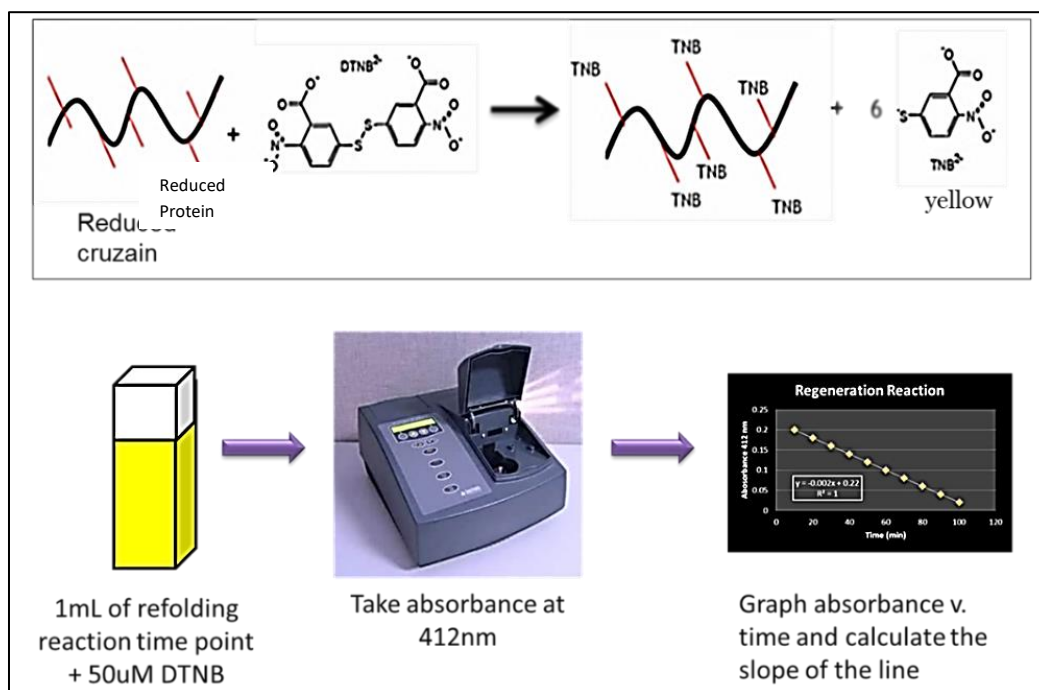
Figure 37: Schematic of Ellman's test for thiol concentration determination

Top: DTNB mechanism. Bottom: Reaction porotocol. Bottom schemic of folding analysis protocol.

3.3.4 Expression of the protein prodomain

The prodomain of procruzain was expressed, purified and refolded as described in 2.3.3-2.3.4 of this manuscript. The protocol used to express, purify and refold procruzain and the mutant C25A were followed without any modifications. Removal of solutes was done with dialysis against phosphate buffered saline (PBS) pH 8.0 with a 3.5 kDa Fisherbrand regenerated cellulose dialysis tubing. After three dialysis changes, purity was visualized by 18% SDS-PAGE against a Fisher BioReagents EZ-Run Rec protein ladder and stained with Coomassie brilliant blue R250 as described in section 2.3.4. The dialyzed product was concentrated with a GE Healthcare Vivaspin 20 Sample Concentrator centrifugal unit at 4000 rpm and 4°C. Figure 38 is a schematic of the protocol used to prepare the proteins. The Bradford assay was used to find the concentration of the prodomain.
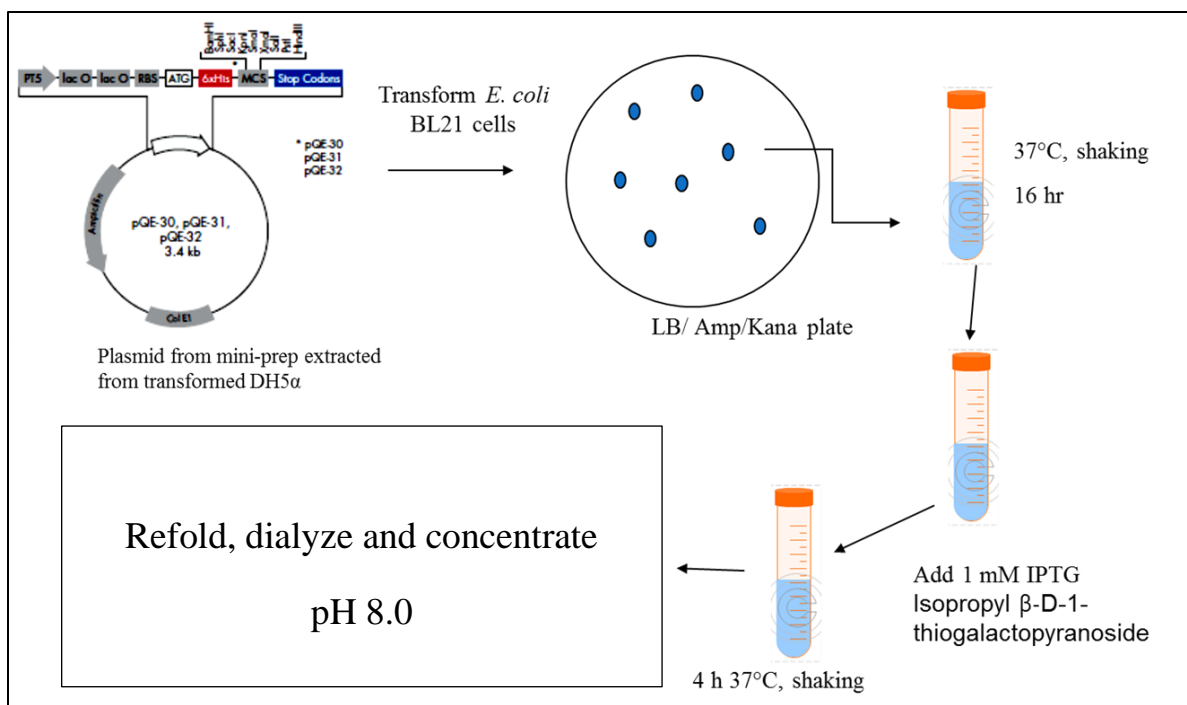
Figure 38: Expression protocol of the proteins used in this study

BL21, M15 or DH5α cells were transformed, plated and grown at 37°C with the appropriate antibiotics. Purification was achieved with a Ni-NTA chelating column. The pure protein was dialyzed and concentrated.

3.3.5   Role of Prodomain in Oxidative Folding

To determine if the pro-region is necessary for the mature protein to regain its native conformation and activity, regeneration and activity was simultaneously detected by fluorescence spectroscopy. The reduced and denatured enzyme (section 3.3.3 for denaturation and reduction method) was desalted on a G-25 (GE Healthcare) column eluted with 1% formic acid at pH 4.0 and was quickly added to a refolding buffer containing the substrate to a 30uM final concentration. The refolding buffer containing 500 mM L-arginine (Sigma), 200 mM Tris HCl (Sigma), 2 mM EDTA (Fischer Scientific) and 40% Glycerol (Fischer Scientific) was diluted 1:3 with ddH$_2$O. The substrate Z-Phe-Arg AMC and reduced cruzain was added to the buffer at final concentration of 50 uM and 30 uM respectively (2 mL total reaction volume). The

sample was placed in a quartz cuvete (Fisher scientific) and inserted in the fluorimeter with an excitation set at 380nm and emission at 440 nm. Fluorescence was detected at 37°C over a period of 16 hours. A negative control lacking mature enzyme was also run to determine if the reagents used in the reaction were fluorogenic.

The reaction was repeated with the addition of expressed and purified prodomain to the reaction mixture in equal concentrations (30 uM) to reduced mature cruzain to see if the folding rate changed.

3.4 Results and discussion

3.4.1 Ellman's test for determination of folding rates

The data obtained from traditional oxidative folding methods was compared to the data obtained from the novel methods to test the validity. Because the folding pattern of RNase A is well known and the folding rates have been determined,[66-68] it was used a model protein. As opposed to the method previously used, oxidative and reducing reagent were completely omitted due to reactivity with DTNB. The oxidative folding results from the HPLC were as expected and the folding rate extrapolated from graphical data. The chromatogram in Figure 39 show the HPLC results of the regeneration of RNase A using an absorbance of 280nm. Several time points were selected and the graphs superimposed to visualize the folding progress. In general, native protein is eluted first from the column and the reduced protein follows second. The separation of native and reduced protein was achieved by choosing a C-18 column and refining the method. The native enzyme binds with a reduced affinity because less hydrophobic amino acids are exposed to interact with the hydrophobic column. On the other hand, the hydrophobic residues are exposed when the protein is reduced and denatured making them available for interactions with the column. The amount of reduced RNase A (second peak) decreased over

time producing more protein in its native state (first peak). The area under the curve was calculated and converted to concentrations with Beer's law.
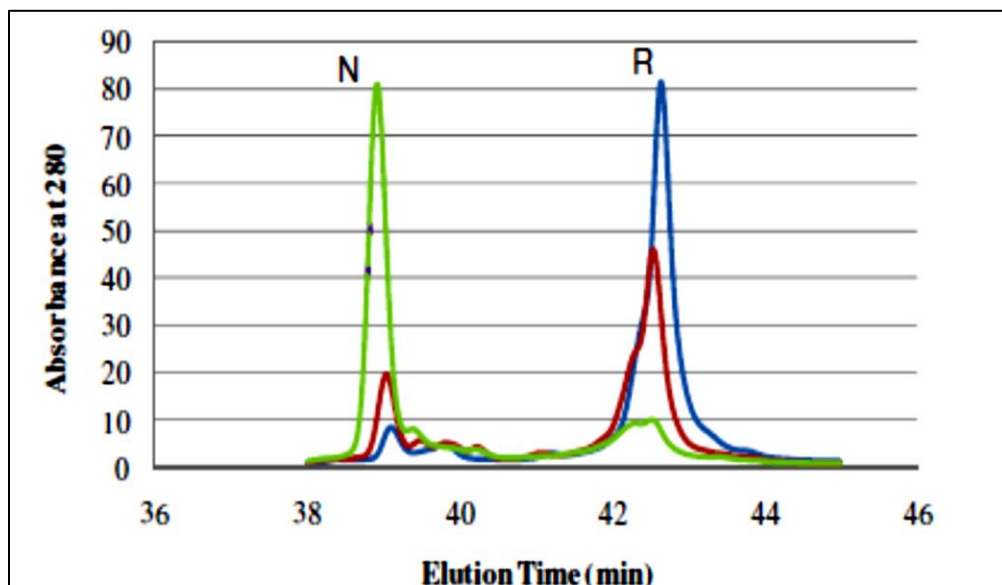
Figure 39:   Regeneration of RNase A

Chromatograms of select time points super imposed to show regeneration over 20 hours. Blue: Time 0, Red: Tine 6 hours, Green: Time 16 hours after initiation of reaction. N=native, R=reduced.

To test the validity of our method with Ellman's reagent, we compared the results obtained from traditional oxidative folding experiments to the results obtained with our novel method. To determine if the protein was folding over time, we reacted an aliquot of each time point collected during oxidative folding with Ellman's reagent and recorded the absorbance at 412 nm.   The absorbance was plotted versus the time to demine a folding rate. As protein folding occurs, less thiols are available to react with DTNB, providing a decrease in intensity of the yellow color (due to free NTB^{2+}). Though spectroscopy provides us with absorbance, the decrease in color intensity can was observed. Figure 40 is a photograph of the results showing a decrease in yellow color over time.

Figure 40: Results of Regeneration Detected with Ellman's Reagent

The intensity of the yellow color at T1 (1 hour after initiating folding) is deeper that T10 (at 20 hours) when regeneration was complete.

This data was compared to the results obtained from the HPLC to determine the validity of the novel method. The folding rates were calculated by taking the area under the curve of each peak from individual time points provided by the HPLC chromatograms. Protein folding follows first order kinetics therefore the data was plotted as the natural log of the fraction of native protein formed over time. The calculations used to find the points plotted can be found in Table 5.

Table 5: Calculating the fraction of native RNase A formed

Time Point	Time (min)	Native (N) (area of first peak on chromatogram)	Reduced (R) (area of second peak on chromatogram)	N+R (total protein)	N/N+R =n (fraction of native)	((1-n) (fraction of reduced)	ln(1-n) (1st order kinetics)
T0	M0	N1	R1	X1	Y1	Z1	F1
T1	M1	N2	R2	X2	Y2	Z2	F2
T2	M2	N3	R3	X3	Y3	Z3	F3
T3	M3	N4	R4	X4	Y4	Z4	F4
T4	M4	N5	R5	X5	Y5	Z5	F5
T5	M5	N6	R6	X6	Y6	Z6	F6
T6	M6	N7	R7	X7	Y7	Z7	F7
T7	M7	N8	R8	X8	Y8	Z8	F8
T8	M8	N9	R9	X9	Y9	Z9	F9
T9	M9	N10	R10	X10	Y10	Z10	F10

Since protein folding is believed to occur under first order kinetics the natural log of 1-n was graphed versus the time in seconds. Figure 41 is an example of a graph of the plotted points lnR (reduced fraction) versus time in minutes. The experiment was replicated at least 6 times and an average of the rate constants was obtained. Other proteins such as lysozyme were used to

test the validity of the method as well (data not shown). The equation of the line provided the

folding rate constant which is equivalent to the slope and the R^2 value shows the correlation

value of the line. This data was compered to data obtained from the method developed with

Ellman's reagent to test the validity of the novel method (Figure 42) which was run in parallel
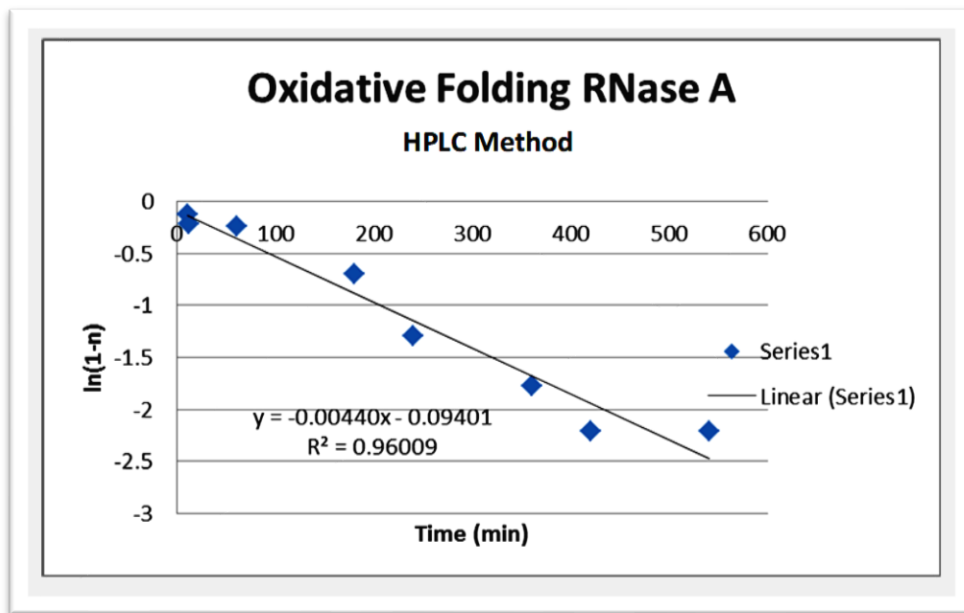
Figure 41: Traditional Regeneration Method

The folding rate constant was obtained from studied done directly on the HPLC
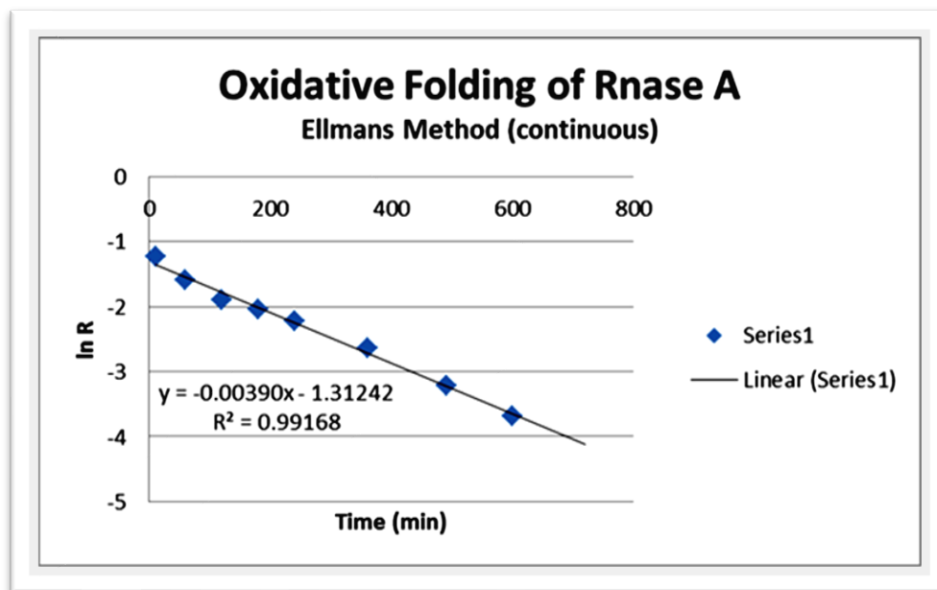
Figure 42: Novel Method for regeneration

Ellman's reagent was used to determine the fraction of folded protein at different time points

Upon conducting several trials of each method that were run in parallel to each other using the same source of reaction mixture, it was determined that novel method produced similar results to the method that employs liquid chromatography and UV absorbance. The Table 6 provides an average of the folding rate constants using the HPLC method in comparison to the novel method including standard deviations for each.

Table 6: Comparison of folding rate constants obtained from the two methods

Method	Avg Folding Constant	Deviation
RNase A HPLC	0.00520	8.983×10^{-4}
RNase A Ellman's	0.00480	6.379×10^{-4}

A slight difference in folding may be due to the storage and preparation of the samples. Aliquots that were collect at different times were quickly frozen (flash freeze with liquid nitrogen) and stored at -80°C until all time points were collected. The samples were allowed to thaw for 15 minutes. DTNB was added and incubated at room temperature for an additional 15 minutes to allow the reaction to occur. A yellow color was observed and the intensity (absorbance) of the yellow color decreased over time due to less thiolates available to react with DTNB as the protein folds.

Because it is difficult track regeneration of procruzain and cruzian by HPLC and UV spectroscopy, we detected the folding progress with Ellman's reagent to determine the concentration of thiolates remaining at different times during regeneration. Ellman's reagent or DTNB (Sigma) is a compound that contains a disulfide bond and creates a mixed disulfide with free thiolates at pH 8.0. In the reaction, DTNB (5,5'-dithio-bis-[2-nitrobenzoic acid]) becomes reduced and releases TNB^{2-} which can be detected by spectroscopy in the visible range. With high concentration of thiolates for example at time zero when a protein is fully reduced, an intense yellow color is observed and the free due to a high concentration of TNB^{2-} (absorbs at 412 nm) released from DTNB. The amount of TNB^{2-} released is proportional to the number of thiolates present according to Beer's law. The absorbance due to the presence of free TNB^{2-} ($\varepsilon=14{,}150$ $M^{-1}cm^{-1}$) decreased over time as the protein folds.[69]

There are several benefits to using this method to determine folding rates. First, highly hydrophobic proteins tend to precipitated when desalted and cannot be solubilized for use in experiments giving a low yield of available protein. Another benefit is that the method has a high sensitivity. The reaction was conducted with protein at 10-30 uM. Finally, the method is cost effective. Reagents are economical and a visible spectrophotometer can be found in most

laboratories, whereas the UV spectrophotometer is much more costly. One flaw of the method is that it is more labor intensive than the traditional method. Most HPLC are equipped with an auto sampler and the reactions can be set up to take time points automatically, every hour for a long period of time. Our method requires and individual to physically remove aliquots at different time points store them and conduct Ellman's test as soon as folding is complete. Time management is essential since proteins lacking oxidative reagents can possibly take days to fold. This method would not be effective for calculating folding rates constants of proteins containing more than 5 disulfides as it may take long to fold the protein without oxidizing agents.

It was previously believed that proenzyme do not achieve a native conformation once the pro-region is cleaved however the results (Figure 41) indicate that both procruzain and mature cruzain will fold over time. Studies on cathepsin and subtilisin have concluded that the prodomain assists in the formation of the native structure *in vivo*, however it has not yet been determined if the pro-region is required.[70, 71] In this work we show that the prodomain is not required for regeneration and that procruzain and cruzain have slightly different rates of regeneration probably due to an additional cysteine found in the prodomain. Figure 43 provides the graphical results for one of the trials of the regeneration of both the mature and the proenzyme. Mature cruzain had a faster rate of regeneration as shown by the slope of the line.
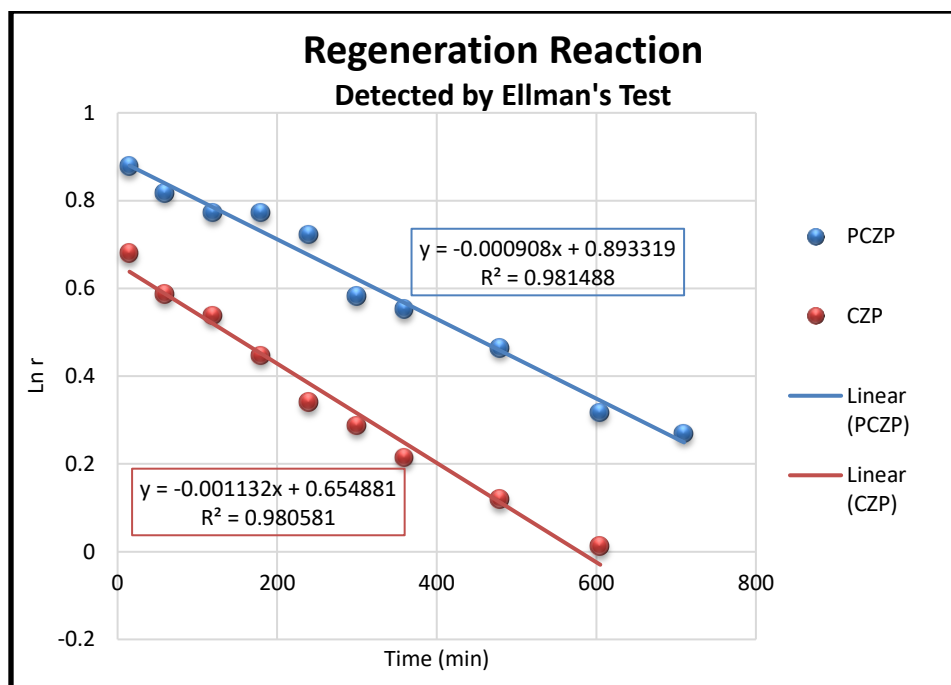
Figure 43: Regeneration results with Ellman's reagent

Comparison of the folding rate of procruzain was compared and cruzain

The average of the trials along with standard deviations are provided in Table 7. Cruzain is slightly faster at reaching the native conformation in comparison to procruzian. Procruzain contains an additional cysteine in the pro-region that can interfere with disulfide bond shuffling, taking longer to reach the native state.

Table 7: Comparison of folding rates of procruzain and cruzain

Protein	Avg Folding Constant	Deviation
Procruzain	0.00095	3.923×10^{-5}
Cruzain	0.0011	1.342×10^{-4}

The folding rate constant at near standard conditions, was similar for each trial and the results of all trials showed that cruzain folded slightly faster than procruzain.

3.4.2 Production of the Prodomain

The prodomain was expressed with the exact same protocol that produced procruzain. The prodomain was refolded and dialyzed with less precipitation when compared to dialysis of procruzain therefore this was the preferred method for the removal of salts. The results were visualized on an 18% gel. The increased concentration of acrylamide was used to reduce run-off of the protein by allowing the small protein to migrate slowly through the gel. Figure 44 is the results of the expression and purification given by SDS-PAGE.
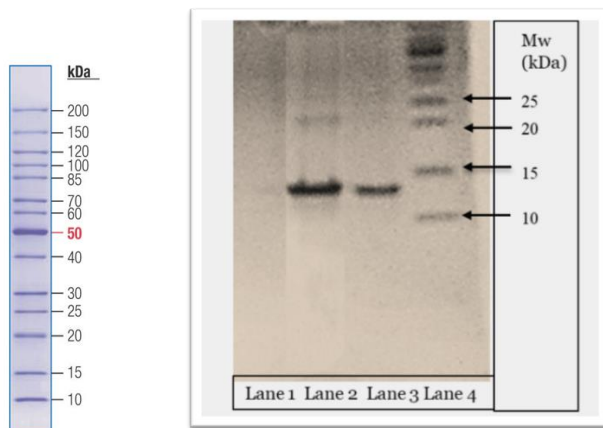
Figure 44: Results of the Prodomain purification

Lane 1: Empty well. Lane 2: Elution 2. Lane 3: Elution 4 off the Ni-NTA column. Lane 4: Protein ladder.

The dominant band that can be seen on the gel has a molecular weight between 10 and 15 kDa. There is a faint band that can be seen above 20 kDa which may represent a dimer between

the prodomains though the amount was negligible. The band was no longer visible in subsequent elutions off the column. The purified prodomian was used to determine the impact on catalytic activity on protein folding of the mature enzyme.

3.4.3 Dependence of the prodomain on protein folding

In addition to regeneration with Ellman's test, we detected the folding progress with fluorescence spectroscopy over several hours to see if the protein would refold and regain its activity. It is believed that proteases expressed as zymogens will not gain their full activity or properly fold since the prodomain exhibits endogenous chaperone-like activity. The results below show an increase in florescence of the reaction mixture over time as compared to the negative control (blank). The blank does not exhibit fluorescence so we can assume that all fluorescence provided by the reaction mixture is due the activity of cruzain. It can be concluded that the protein can regain the native conformation even in the absence of the prodomain, and more importantly, regains the protease catalytic activity.
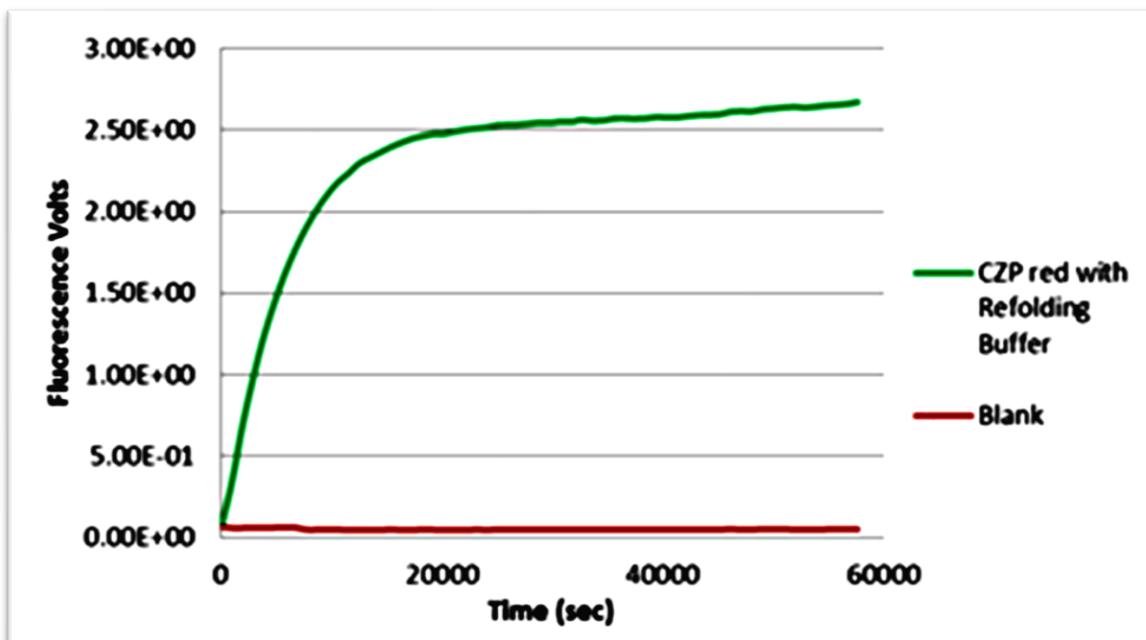
Figure 45: Regeneration of reduced and denatured cruzain

Regeneration of previously activated reduced and denatured enzyme was detected by fluorescence spectroscopy over 16 hours.

The mature enzyme was allowed to fold in the presence of the prodomain to verify the prodomain was not needed for regeneration and function. Equimolar of purified prodomain was added to reduced cruzian and allowed to fold over the course of 16 hours. Regeneration and activity were detected by the release of fluorescence of the fluorogenic substrate z-Phe-Arg AMC. The figure below shows that activity decreases when the purified prodomain is added to the regeneration mixture. It seems that once the prodomain is added to cruzain, it actually impedes folding and inhibits the enzymatic activity. Though the peptide bond between the mature enzyme and the prodomain does not exist, some of the interactions made may block folding and catalytic activity. When observing the mixture after the reaction was complete, a slight precipitation was visible. This could mean that the prodomain prevented the protein from

78

folding and caused the protein to aggregate, greatly reducing the amount of refolded active enzyme.

Another reason a decline in fluorescence could have occurred is that the enzyme may hydrolyze the prodomain in addition or in place of the fluorogenic substrate as it enters the reaction mixture. The prodomain contains several arginine residues at which the enzyme can catalyze degradation. This activity would not be visible on the spectrofluorometer due to the lack of fluorogenic groups on the prodomain. The enzyme may have a higher affinity for the prodoamin than z-Phe-Arg AMC. One way to verify if the prodomain attaches to the mature enzyme through non-covalent interactions would be to run the samples on a native gel under non-denaturing conditions however proving this by electrophoresis proved to be difficult.
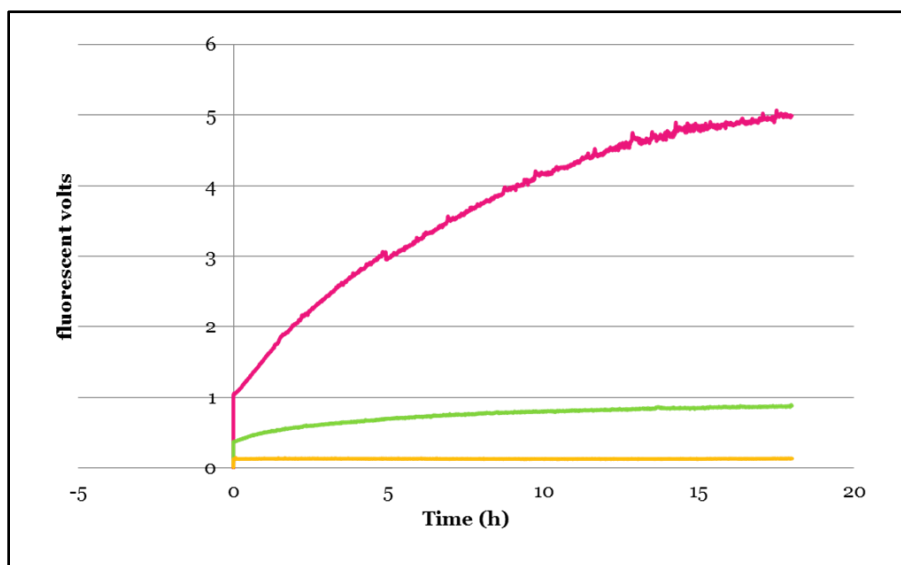
Figure 46: Catalytic Activity of Cruzain after the Addition of the Prodomain

The yellow line is the blank with substrate and prodomain but no substrate. The green line is the enzyme added to the substrate with the prodomain. The pink line is the enzyme with substrate lacking the prodomain

Work conducted on proenzymes such as the procathepsins and prosubtilisin suggest that the prodomain acts as an intramolecular chaperone and is essential in protein folding. Because cruzain regains activity when regenerated after being reduced and denatured, we can conclude that the correct conformation can be achieved without the prodomain. It can also be concluded that free prodomain may actually inhibit the mature enzyme. The role of the prodomain in the folding of procruzian is studied in the upcoming chapters.

CHAPTER 4

EXPRESSION OF PROCRUZAIN WITH STURCTURE STABILIZING MUTATIONS

4.1 Specific Aim

To determine the role of the prodomain in conformational folding of procruzain by identifying structural stabilizing residues with in the prodomain.

4.2 Background

As previously mentioned, the prodomain of zymogenic enzymes is believed to play a significant role in protein folding and is essential to regulation of enzymatic function[65]. Enzymatic activity is gained by the autocatalytic cleavage of the prodomain. Once cleaved, the prodomain may undergo additional degradation to prevent inactivation of the mature enzyme[29]. Though not well understood, it has been proposed that cysteine proteases contain a highly conserved catalytic triad, in addition, residues near that catalytic site may also interact with the proregion to achieve the catalytic function. These interactions are essential to regulation. Interestingly, the prodomains different classes of proteases differ from family to family however differences may even exist within a family. The pro-regions span from 30 amino acid residues long to over 200 residues long. Some have several alpha helical regions while others very few. Procathepsin-B contains an occluding loop that may help regulate enzymatic activity, while procathepsin-L lacks this loop. Procathepsin-B contains two short alpha helix domains while the alpha-helix found within procathepsin-L are significantly larger. Figure 45 compares the prodomain of procathepsin-B and procathepsin-L highlighting the occluding loop in procathepsin-B. The occluding loop of procathepsin-B may enhance inhibition of the zymogen

but is also believed to take part in the exogenous protease activity. This activity is less prominent in procathepsin-L, which lacks the occluding loop.
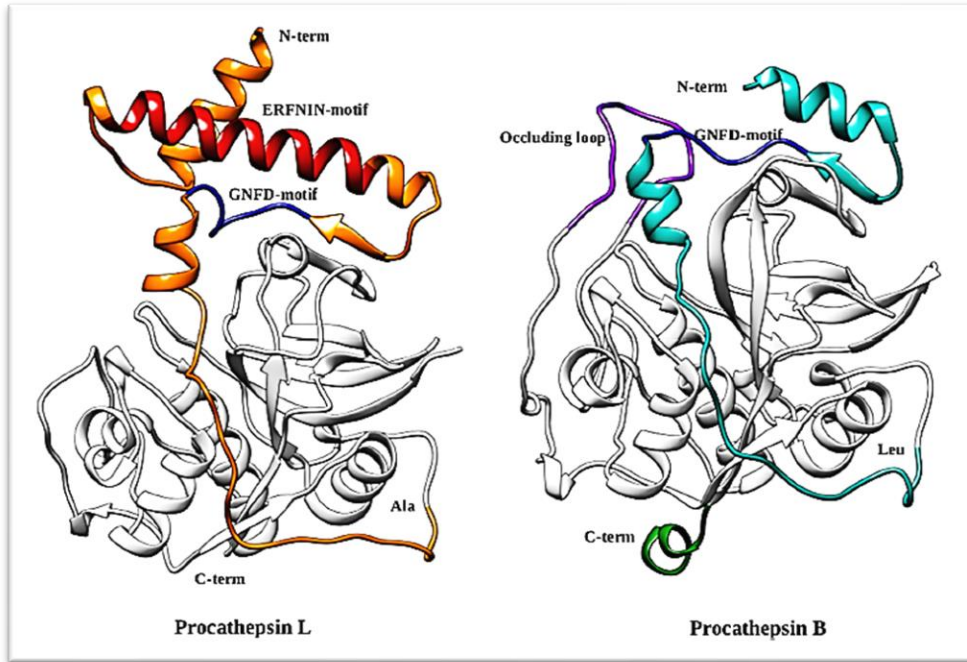
Figure 45: Comparsion of the procathpsin-L and procathpsin-b prodomains
Procathepsin-L domain is larger by approximately 33 amino acids and procathepsin-B contains an occluding loop (purple) that L lacks.[30]

Most cysteine proteinases reside in the lysosomes of the cell which is an acidic organelle, therefore the optimal pH for activation is essential for maturation and function. In *in vitro* studies, the pH must be maintained between a pH of 4.5 and 5.5 for autoactivation and full catalytic activity. At a low pH, the salt bridge interactions found between the prodomin and the active site are disrupted by the protonation of carboxylate containing residues such as aspartic acid and glutamic acid. Disrupting these interactions can lead the disruption of other prodomain-active site interactions. It has been observed that an interaction of a highly conserved glycine

residue pushes the c-terminal towards the active site, occludes the catalytic regions and obstructs activity. A decrease in pH and disruption of salt bridges can cause conformational changes that allow the c-terminal domain to delocalize from the catalytic site. The catalytic site, being free from obstruction cleaves the prodomain.[72] A crystal structure for procruzain is not available however modeling studies and point mutations may give us an insight to the interaction of the prodomain with catalytic domain. If these residues are essential for protein maturation and function, then they could also be significant to achieving a native conformation as the protein is being synthesized. Structure stabilizing residues can be targeted to prevent folding before the protein matures in an effort to combat Chagas disease, an immensely different approach to inhibiting cruzain's active site directly with small molecules.

4.3 Methods: Identification of target residues, expression and purification

4.3.1 Molecular modeling and alignment studies

Molecular dynamics on procuzain was conducted to identify structure stabilizing residues that reside within the prodoamin. Charged amino acid residues were replaced with neutral, nonpolar residues. Folding and activity of the mutants was compared to the wilt type proenzyme to determine the role of the mutated residues on protein folding.

Homology studies to identify conserved residues was conducted with a BLAST (Basic Local Alignment Search Tool)[73] analysis using the known amino acid sequence of procruzain. There are human cysteine proteases such as cathepsins that have significant roles in biochemical processes and are also implicated in many diseases such as cancer. Homology studies can help us determine which amino acids are conserved and the effect of mutations on essential human cysteine proteinases such as the cathepsins. Another reason for sequencing homology studies is

to determine which amino acids are conserved across different species of parasites to target other disease-causing parasites containing cysteine proteinases essential to parasitic survival.

Since no crystal structure for PCZN has been deciphered yet, a homology modeled of PCZN was constructed using I-Tasser.[74] Inter-protein ionic and hydrophobic interactions were further calculated by Protein Interaction Calculator (PIC), where salt bridges were assigned when distance between the two atoms of opposite charge was less than 6Å and side chain-side chain hydrophobic interactions were also calculated within 6Å distance. The homology-modeled structure, inter-protein interactions, and residue substitutions were visualized using PyMol MAESTRO, and CHIMERA software.

4.3.2   Expression and purification

4.3.2.1 Double selection expression

Expression of the mutants each containing a single mutation required a double selection process. A plasmid pQE-30 containing 6xhistidine-tagged procruzain (WT and mutants) at the N-terminus was introduced into competent *E. coli* DH5α for subcloning and into *E. coli* M15 for expression (Invitrogen). A modified version of a method previously described in this manuscript was used produce a high yield of protein. A 5 mL LB culture with antibiotics (amp 100 uM/kan 50 uM) was inoculated with a freshly transformed colony and grown to an O.D.$_{600}$ of 0.7-0.9. Grown cells were then inoculated onto a LB/amp/kan plate and grown overnight at 37˚C. From the transformed plate, 3 colonies were selected inoculate 3x 5 mL LB/amp/kan culture and grown overnight. at 37˚C. A small-scale expression, described below, was performed for each 5 mL culture. Before induction with Isopropyl-B-D-1-thiogalactopyranoside, IPTG (Fisher Scientific), cells from each expression were inoculated into a LB/amp/kan plate. The plates were identified as "colony A", "colony B" and "colony C". Expression levels were checked and 3

colonies from the plate with the highest expression level were selected. The second colony selection repeated the aforementioned procedure followed by glycerol preservation.

4.3.2.2 Small scale expression

Luria Broth (LB) media (5 mL) was inoculated with one colony from overnight grown plate and incubated at 37˚C. When the O.D.$_{600}$ was between 2 and 3, the culture was gently spun down for 5 minutes at 1500 rpm. The resulted cell pellet was re-suspended into minimal medium. When the O.D.$_{600}$ reached around 1.0 cells were induced by addition of IPTG (1 mM) and incubated at 37˚C for 4 hrs. Expression levels were checked by spinning down 250 uL of cells suspension at 4400 rpm for 5mim. The pellet was re-suspended in 100 uL lysis buffer, then 20 uL of lysate was mixed with 20 uL of sodium dodecyl sulfate (SDS) loading buffer and heated at 95˚C for 5 minutes. The sample was centrifuged and supernatant was loaded into SDS-PAGE to check the expression level.

Minimal medium was composed of M9 salts (64 g Na$_2$HPO$_4$, 15 g KH$_2$PO$_4$, 2.5 g NaCl, and 5 g NH$_4$Cl into 1000 mL of MiliQ water), 1M MgSO$_4$, 20% glucose, 1M CaCl$_2$. The MgSO$_4$, glucose, and CaCl$_2$ were sterilized with a 0.2 um Milipore Stericup vacuum filtration system. To prepare the minimal media, 200 ml of M9 salts were added to 700 ml of sterile MilliQ, then 2 ml of 1 M MgSO$_4$, 20 ml of 20% glucose, 100 ul of 1 M CaCl$_2$ were added last then the volume was adjusted to 1000 ml with sterile miliQ water. Minimal medial was only used if expression levels appeared to be low, otherwise cultures were grown in LB.

4.3.2.3 Large scale expression

One colony selected from the plate with the highest expression levels was inoculated into 100mL of LB media with amp(100 uM)/kan(50 uM) antibiotics and grown overnight at 37˚C,

then the cells were gently spun down at 4400 rpm for 15 minutes at room temperature and the

pellet was re-suspended in 500 mL of minimal medium (2:10 ratio of bacteria culture to flask

volume) and incubated at 37˚C. Once O.D.$_{600}$ reached 0.6-0.7 the culture was induced for 4 hours

(1 mM IPTG).

4.3.2.4 Cell Lysis and Protein purification

After induction the cells were collected by centrifuging at 5000 rpm for 15 minutes at

4˚C using a Sorvall Centrifuge RC28S (Wilmington, DE, USA) with GSA rotor, and lysed in

100 mM NaH$_2$PO$_4$, 10 mM Tris-HCl, 10 mM imidazole, and 8 M urea (pH 8), under agitation,

for 30 minutes at 4°C. The cell lysate formed was further disrupted by sonication at 60 W for 3 x

30 s pulses (with 1min delay between pulses) on ice bath, with a Branson Sonifer 450. The

soluble material was recovered by centrifugation at 11000 rpm for 20 minutes using a Sorvall

Centrifuge RC28S with F28∕36 rotor.

Recombinant mutants and wild type PCZN proteins were purified using affinity

chromatography. The solubilized His-tagged recombinant proteins were mixed with pre-

equilibrated Ni-nitrilotriacetic acid (NTA)-agarose resin (loaded with Ni^{2+}) (ThermoFisher) for

15 min. Unbound protein was allowed to pass through the column and the column was washed

with wash buffer (100 mM NaH$_2$PO$_4$, 10 mM Tris-HCl, and 8 M urea pH 6.3). The recombinant

proteins were then eluted with elution buffer using gradient concentrations of imidazole (10, 30,

50, and 200 mM). Then the purity of recombinant proteins was checked by 12% SDS-PAGE.

The eluted samples were pooled and submitted to in vitro refolding by incubation with 5 mM

dithiothreitol at 37 °C for 45 minutes, followed by 20-fold dilution in ice-cold 100 mM Tris-HCl

(pH 8), 1 mM EDTA, 1 M KCl, and 20% glycerol, and further incubation at 4 °C for 24 h. The

solution was concentrated in a mini-reservoir (RC800 Millipore)) under nitrogen pressure to a final volume of 30 mL using 0.2 um cellulose acetate filters (Millipore, Bedford, MA, USA) at 4°C. Protein concentration was measured with Nano Drop 1000 instrument (Thermo Scientific, software ND 100 v3.6.0) by using extinction coefficient of 63,500 $cm^{-1}M^{-1}$ and molecular weight of 35.4 kDa.

4.3.2.5 Protein Refolding

Aliquots with pure PCZN protein were combined and a final concentration of 30 uM dithiothreitol (DTTred) was added to the pooled fractions and incubated at 37˚C under slight agitation for 1 hour. The resulted denatured and reduced protein was diluted 20 times in refolding buffer (250 mM Arginine, 100 mM Tris HCl, 1 mM EDTA and 20% Glycerol, pH 8.0) and incubated under slight agitation at 4˚C overnight as previously described[21]. The sample was dialyzed using a 3 kDa MWCO Ultracel membrane filter (Millipore) prior to being concentrated to a final volume of 20 mL. Refolded protein was used for kinetic studies obtained by fluorescence spectroscopy.

4.3.3   Kinetic Study of Mutants in comparison with wild type procruzain

All recombinant proteins were purified, refolded and dialyzed as mentioned in chapter 2. After protein was refolded and dialyzed with cold water, the protein concentration was adjusted to 0.3 mg/mL. Protease activity of recombinant mutants and WT pczp was measured using Z-Phe-Arg-7-amido-4-methyl-coumarin-HCl (AMC) as substrate (where Z is benzyloxycarbonyl), and performed as previously described[75] with modifications. Activation assays were performed in a DM45 Spectrofluorimeter (On-line Instrument Systems, Inc.) using Spectra Works software and were carry out in a quartz cuvette 1cm path length (Fisher Scientific). Fluorescence was

recorded as a function of time at 440 nm emission wavelength and 380 nm excitation wavelength and set for 0.100 integration time and 600 PMT HV. The reaction cuvette, with a total volume of 2050 uL, was filled with activation buffer (5 mM EDTA, 100 mM NaH_2PO_4, 200 mM NaCl, 2 mM DTTred, pH 4.5) and substrate (40 uM AMC), 5 mM EDTA, 50 mM NaH_2PO_4, 100 mM NaCl, 5 mM DTTred) and adjusted to pH 5.3. Reaction was then initiated by addition of 50 uL of refolded protein at a concentration of 0.3 mg/mL. The fluorescence was recorded promptly after protein addition and incubated at 37˚C for a period of 5000 seconds. When maximum activity was observed, the mature enzyme was inhibited with 10 uM Leupeptin (Sigma) for 15 min at 37 ˚C. An aliquot of the active protein was lyophilized and run on 12% SDS-PAGE and gels were stained in Coomassie brilliant blue R-250 (BioRad).

Refolded protein was dialyzed against Phosphate-buffered saline buffer (PBS) pH 7.4 using 3 kDa MWCO tubes (GE Healthcare) and further concentrated using centrifugal filter tubes (Vivaspin2, 3 kDa MWCO). Resultant dialyzed protein was passed through PVDF 0.22 um filters (Fisherbrand) to remove any debris of precipitated protein. Protein concentration was measured using NanoDrop instrument (Thermo Scientific Nano Drop 1000, software ND 100 v3.6.0). Assays were performed in 0.1 cm path length quartz cuvette (Fisher Scientific). After protein folding and determination of concentration the extent of protein folding, and activity were analyzed with fluorescence spectroscopy.

Protease activity of recombinant mutants and PCZN-WT was measured using Z-Phe-Arg-7-amido-4-methyl-coumarin-HCl (AMC) as substrate, and performed as previously described[16] with modifications. Activation assays were performed in a DM45 Spectrofluorometer (On-line Instrument Systems, Inc.) using Spectra Works software and were carried out in a quartz cuvette 1cm path length (Fisher Scientific). Fluorescence was recorded as a function of time at 440 nm

emission wavelength and 380 nm excitation wavelength and set for 0.100 integration time and 600 PMT HV. Recombinant protein was added to a 1:1 activation buffer (200 mM NaCH$_3$COO, 1.8 M NaCl, 10 mM DTT^{red} and 10 mM EDTA, pH 5.2) to substrate buffer (40 uM Z-F-R-AMC, 5 mM EDTA, 50 mM NaH$_2$PO$_4$, 100 mM NaCl, 5 mM DTT^{red}) for a final protein concentration of 7.32 ug/mL and pH of 5.3. The fluorescence was recorded promptly after protein addition and incubated at 37°C for a period of 50,000 seconds. When maximum activity was observed, the mature enzyme was inhibited with 10 uM Leupeptin (Sigma) for 15 minutes at 37 °C to prevent auto-degradation of the protein. An aliquot of the active protein was lyophilized and run on 12% SDS-PAGE and gels were stained in Coomassie brilliant blue R-250 (BioRad).

4.4 Results and discussion

4.4.1 Molecular dynamics results

Amino acid residues that are highly conserved among parasitic cysteine proteases but not human are targeted in our studies. The following results were provided by sequencing alignment.

Protein	Accession #	Total score	Query coverage (%)	Maximum identity (%)
Trypanosoma brucei (brucipain) Cathepsin B-like protein	XP_845225.1	470	98	66
Trypanosoma rangeli Cathepsin L-like protein	AFA34859.1	463	100	75
Trypanosoma congolense (congopain). Cathepsin L-like protein	AAA18215.1	436	96	64
Leishmania braziliensis Cathepsin L-like protease	XP_001562140.1	371	93	57
Leishmania mexicana Cathepsin L-like protease	CBZ24131.1	362	91	55

Figure 46: Result of sequence alignment studies

The molecular modeling studies identified the following mutations as potential structure stabilizing targets.

90

Table 8: List of Targets and their Mutations

Target Number	Mutation (including his-tag)
1	D70V (D82V)
2	E75V (E87V)
3	F20A (F32A)
4	F23A (F35A)
5	F44A (F56A)
6	E13V (E25V)
7	E37V (E49V)
8	E74V (E86V)
Negative Control	C129A (C141A)

4.4.2   Expression results

All eight targets were expressed and purified. They were run on a12% SDS-PAGE determine expression level and purity. The diagrams below are photos of the gels after the purification of each target.   The purest fractions were pooled, the concentration of each protein was determined and prepared for fluorescence analysis of protein folding and activity.
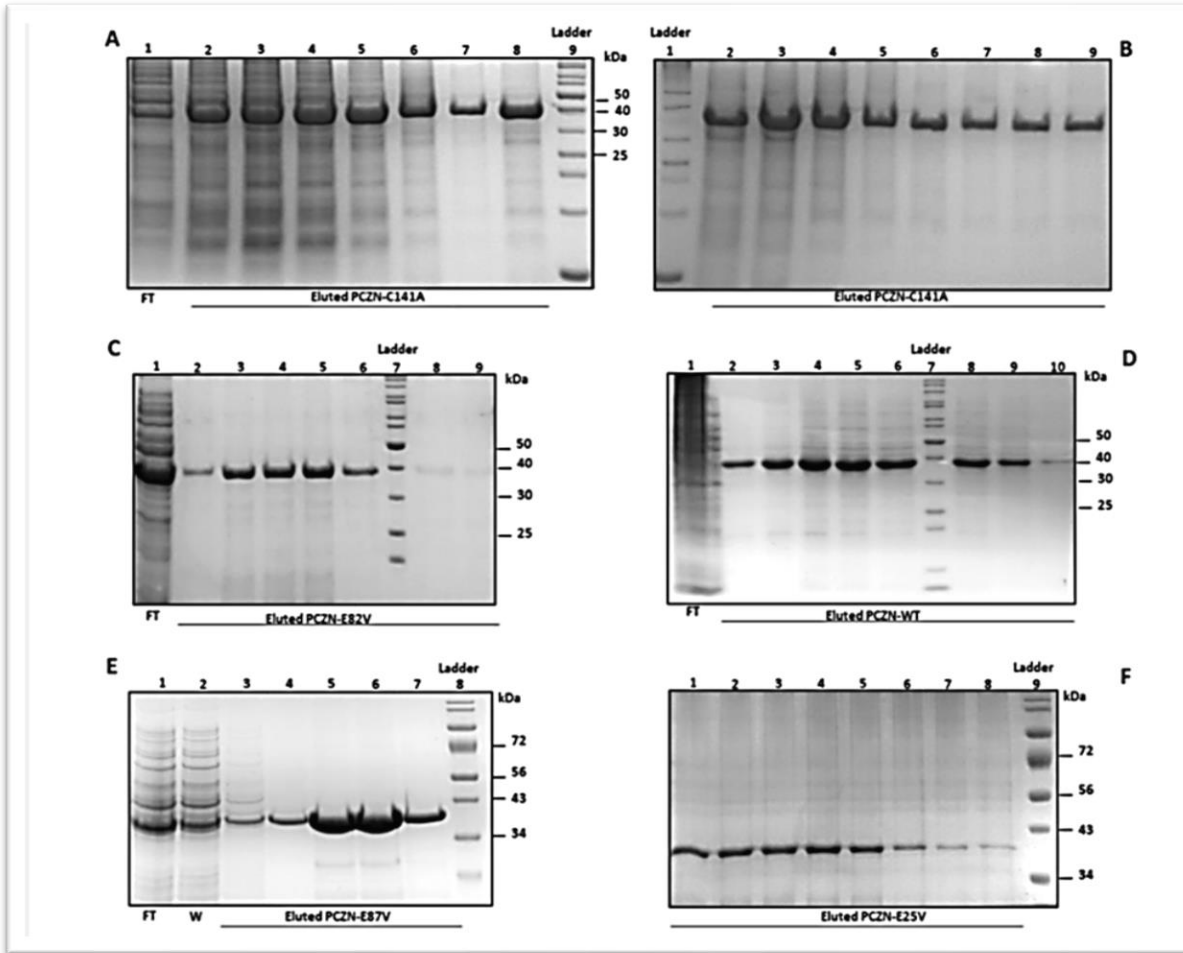
Figure 47: Expression Results of Selected Targets

Figure 47 A and B =C129A, C=Target 1 (D70V), D=Wild-type, E=Target 2 (E75V), and F=Target 6 (E13V) were expressed for protein folding studies. The remaining targets 4 (F23A), 5 (F44A), 7 (E37V), and 8 (E74V) were expressed at a later time.
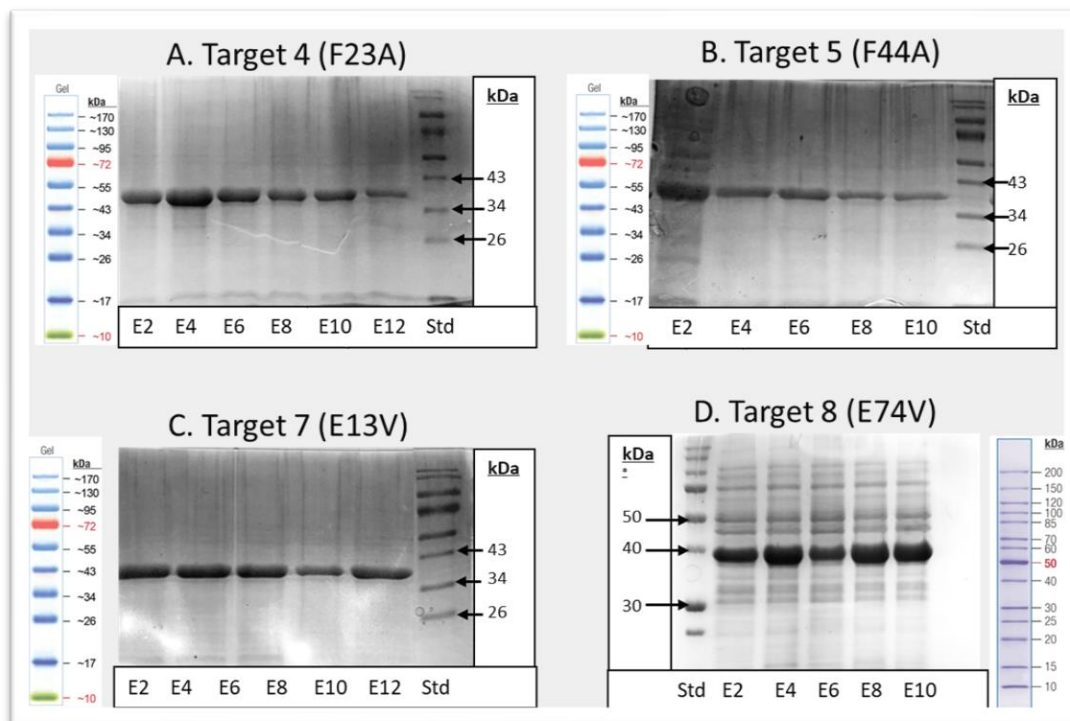
Figure 48: Expression results of Targets 4, 5, 7 and 8

The remaining targets were expressed and also analyzed with fluorescence spectroscopy. A = Target 4, B =Target 5, C= Target 7, D=Target 8. A low yield of Target 3 was obtained (not shown).

It is now fully known why expression levels were different for each target. In fact, with in the same target, the quantities of purified product differed between expression. One issue could have occurred with the plasmid construction. The purest fractions of the expression were collected, refolded and used for fluorescence spectroscopy studies.

4.4.3   Activation assays of recombinant mutants and wild type procruzain

Although pro domains of several cysteine proteases have shown to be potent inhibitors of other papain-like enzymes[69], no studies have been done on identification of critical residues of these pro domains with the final objective of protein early inactivation. In our study, enzymatic

activity of recombinant cysteine proteases was accessed by the liberation of the fluorescent leaving group 7-amino-4-methyl-coumarin (AMC) from the peptide substrate Z-Phe-Arg-AMC and the fluorogenic molecule was measured (Figure X). Slope of the curve of the autoactivation of PCZN and its mutants resulted as follows: WT= 8.43×10^{-4}, E25V= 2.56×10^{-4}, E87V= 3.73×10^{-4}, D82V= 1.31×10^{-5}. From the observed results, D82V mutation was found to be the most deleterious compared to WT, followed by E25V and lastly E87V. As it has been stablished that mutation of active site of subtilisin (S22C) does not affect the folding rate[70], we considered mutant C25A (C141A) as negative control for the studies for autoactivation of mutants. Therefore, C141A mutant showed no activity as expected. The low rate of autoactivation of mutant D82V may be as a consequence of a change in the backbone conformation of Glu44 and Val82 that destabilized the protein even more so than the rest of the mutants.

Although mutant E25V was proposed as important residue by sequence alignment studies and by folding characterization, it showed a decrement in autoactivation rate during the above-mentioned experimental analysis. Although neither of the mutants are not completely inactive, residue mutations have reduced the proteinase activity by delaying the reach of the substrate. Previous studies have shown the effect of mutated pro domains of proteases. That is the case of subtilisin E, where mutations generated within the pro-peptide were incapable of producing an active subtilisin,[75] or in the case of cathepsin L that resulted in improper folding after deletions in its pro domain.[70] Nevertheless, not all mutations produced a negative effect on the activity of proteases as demonstrated in another study done by Ruan *et al.,* 1999[76] where mutated pro-peptide increases subtilisin's folding rate when compared to WT pro-peptide. Indeed, our mutants caused a deleterious effect on the rate of PCZN's autoactivation proving important information for the identification for structural destabilization sites within the pro domain.

94

Pro-domain mutations can indeed have critical effect on enzyme activity. The almost complete lack of activity in D82V mutant signifies that it could act as a potential therapeutic target. Other mutations, had increased autoactivation time, but, once they got activated, they reached similar V_{max} compared to the PCZN-WT. D82V activation mimicked the C141A active site mutant which had hardly any activity. What effect does D82V mutation has on the enzyme leading to complete lack of activity needs further in-depth study.
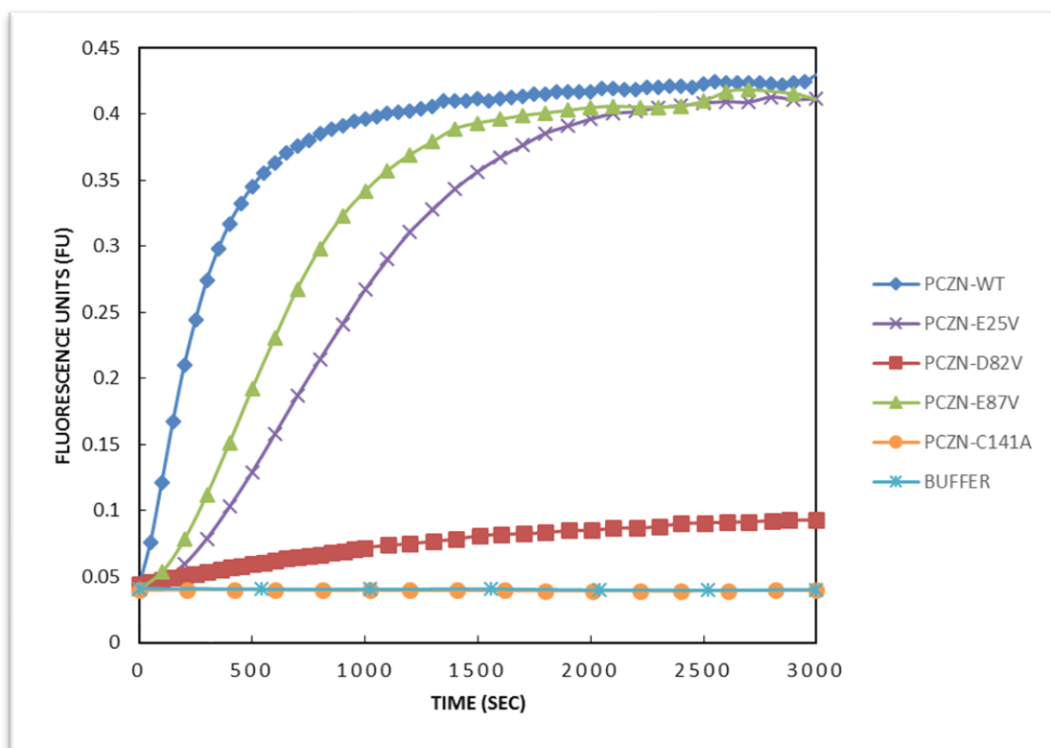
Figure 49: Activation of recombinant PCZN-WT and mutants.

The kinetics of autoactivation of recombinant PCZN was monitored from cleavage of AMC from Z-Phe-Arg-AMC substrate. Diamonds: PCZN-WT, cross marks: E25V, squares: PCZN-D82V, triangles: PCZN-D87V, circles: PCZ-C141A, stars: activation buffer alone.

95

Previous studies have shown the effect of mutated pro domains of proteases. That is the case of subtilisin E, where mutations G2R, M20T, K35E, and V67A generated within the pro-peptide were incapable of producing an active subtilisin; or in the case of cathepsin L that resulted in improper folding after deletions in its pro domain.[74,75] Nevertheless, no all mutations produced a negative effect on the activity of proteases where in another study done by Ruan *et al.,* 1999[76], mutated pro-peptide increases subtilisin's folding rate when compared to WT pro-peptide**.** Indeed, our mutants caused a negative effect on the rate of PCZN's autoactivation proving important information for the identification for structural destabilization sites within the pro domain.

CHAPTER 5

CONCLUSION

To date we have been able to refine the methods of expression, purification and activation of cruzain. It is believed that the prodomain can help the folding of the protein *in vivo* however no data has been given on the effect of the folding rate. Our initial folding data suggests that the prodomain is not necessary for folding and may the slow regeneration process however several amino acid residues within the prodomain are imperative to the structure and essential the function of the cruzain. Mutations of structure stabilizing amino acid residues provided in this work resulted in a decrease in enzymatic activity for all the targets but one. Target 5 (F44A) showed an increase in activity possibly due to the increased interaction of arginine or phenylamine residues with the active site, promoting autodegradation. Future studies can be done on combining mutations to further impede protein folding and activity. Thereafter, small molecules can be synthesized to target those structure stabilizing reactions to prevent protein folding, enzymatic activity and the progression of the Chagas disease.

FUTURE PLANS

As future plans, additional mutations, possibly not limited to the prodomain, can be carried out to determine the role of the reside in protein folding. In the future, small molecules cans be developed to target these resides to prevent folding and maturation of procruzain to cruzain. Point mutations that showed to slow protein folding in our studied can be combined to analyze the additive effect of both mutations. In addition, mutagenesis of observed hydrophobic cluster within the pro domain, compromised of phenylalanine 32, 35 and 56, may be valuable to determine its implication in protein destabilization. Insight to amino acid residue interactions that destabilization of these interactions can give us great insight in producing anti-chagastics drugs.

REFERENCES

1. Tanowitz H.B., Machado F.S., Jelicks L.A., Shirani J., Campos de Carvalho A.C., Spray D.C., Factor S.M., Kirchohoff L.V., Weiss L.M. (2009) *Progress in Cardiovascular Disease* 51:524-539.

2. Lewinsohn, Rachel. (1981) Carlos Chagas and the discovery of Chagas' disease (American trypanosomiasis). Journal of the Royal Society of Medicine, 74(6), 451–5.

3. Manne-Goehler, J., Umeh, C. A., Montgomery, S. P., and Wirtz, V. J. (2016) Estimating the burden of Chagas disease in the United States. PLoS Neglected Tropical Diseases 10:e0005033.

4. Hanford E.J., Zhan B.F., Lu, Y., and Giordano J. (2007) Chagas disease in Texas: recognizing the significance and implications of evidence in the literature. *Social Science and Medicine* 65:60-79.

5. Bern, C., Kjos, S., Yabsley, M.J., and Montgomery, S.B. (2011) *Trypanosoma cruzi* and Chagas' Disease in the United States. *Clinical Microbiology Reviews*. 24(4):655–681

6. Dodd R.Y., and Leiby D.A. (2004) *Ann. Rev. Med* 55:191-207.

7. Food and Drug Administration. Updated Feb 2018. Accessed Jun 2018. http://www.fda.gov/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/LicensedProductsBLAs/BloodDonorScreening/InfectiousDisease/ucm085846.htm

8. Bern, Caryn. (2015). Chagas' disease. *The New England Journal of Medicine.* 373:456–466.

9. Ceballos, L.A., Piccinali, R.V., Berkunsky, I., Kitron, U., and Gurtler, R.E. (2009) First finding of melanic sylvatic *Triatoma infestans*. *Journal of Medicinal Entomology*. 46(5): 1195-1202.

10. Rassi, A., and Marin-Neto, J. A. (2010). Chagas disease. *Lancet*. 375(9723):1388–402

11. Center for Disease Control website. Parasites - American Trypanosomiasis (also known as Chagas Disease). Updated May 2016. Accessed Aug 2016. www.cdc.gov/parasites/chagas/

12. Center for Disease Control website. Laboratory Identification of Parasites of Public Health Concern: America Trypanosomiasis. Updated Dec. 2017. Accessed Nov. 2018. https://www.cdc.gov/dpdx/trypanosomiasisamerican/index.html

13. Teixeira A.R.L., Nascimento R.J., Sturm N. (2006) Evolution and pathology in Chagas disease-a review. *Mem Inst Oswaldo Cruz* 101:463-491.

14. Bastien Joseph William. (1998) *The Kiss of Death-Chagas Disease in the Americas*. Salt Lake City, University of Utah Press.

15. Tarleton, Rick. (2001) Parasite persistence in the etiology of Chagas disease. *International Journal for Parasitology* 31:550-554.

16. Guedes P.M.M., Fietto J.L.R., Lana M., Bahia M.T. (2006) Advances in Chagas disease chemotherapy. *Anti-Infective Agents in Medicinal Chemistry* 5:175-186.

17. Muñoz, M. J., Murcia, L., & Segovia, M. (2011). The urgent need to develop new drugs and tools for the treatment of Chagas disease. *Expert Review of Anti-Infective Therapy*. 9:5–7.

18. Castro, J. A., deMecca,M. M., & Bartel, L. C. (2006). Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). *Human & Experimental Toxicology* 25: 471–479.

19. Urbina J.A., Docampo R. (2003) Specific chemotherapy of Chagas disease: controversies and advances. *TRENDS in Parasitology* 19:495-501.

20. Camargo E.P. (2009) Perspectives of vaccination in Chagas disease revisited. *Mem Inst Oswaldo Cruz.* 104 (Suppl. I): 275-280.

21. Eakin A.E., McGrath M.E., McKerrow J.H., Fletterick R.J., Craik C.S. (1993) Production of crystallizable cruzain, the major cysteine protease from *Trypanosoma cruzi*. *Journal of Biological Chemistry.* 268:6115-6118.

22. Cazzulo B.M.F., Martinez J, North M.J., Coombs, Cazzulo J (1994) Effects of proteinase inhibitors on the growth and differentiation of Trypanosoma cruzi. *FEMS Microbiology Letters* 124:81-86.

23. Clark, D. P. and Pazdernik, N.J. Biotechnology 2nd Ed. **2016.**

24. Gurumallesh, P., Alagu, K., Ramarkrishnan, B., and Muthusamy, S. (2019) A systematic reconsideration on proteases. *International Journal of Biological Molecules*.**,** *128,* 254–267.

25. Baird, T.T., Wright, W.D. and Craik, C.S. (2006) Conversion of trypsin to a functional threonine protease. *Protein Sci.* 15:1229–1238

26. Di Cera, E. (2009) Serine proteases. *IUBMB Life* 61:510–515

27. Graycar, T.P., Bott, R.R., Power, D., and Estell, A. (2013) Handbook of Proteolytic Enzymes Chapter 693-Subtilisns. 3:3148-3155.

28. Siklos, M., BenAissa, M., and Thatcher, G.R.J. (2015) Cysteine proteases as therapeutic targets: does selectivity matter? A systematic review of calpain and cathepsin inhibitors. *Acta Pharm. Sin. B* 5:506–519,

29. Verma, S. Dixit, R. and Pandey, K.C. (2016) Cysteine Proteases: Modes of activation and future prospects as pharmacological targets. *Front. Pharmacol*. 7:1-12.

30. Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., and Turk, D. (2012) Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochimica et Biophysica Acta*. 1824:68–88.

31. Schroder, B.A., Wrocklage, C., Hasilik, A., and Saftig, P. (2010) The proteome of lysosomes. *Proteomics*. 10:4053–4076.

32. Rozman, J., Stojan, J., Kuhelj, R., Turk, V., and Turk, B. (1999). Autocatalytic processing of recombinant human procathepsin B is a bimolecular process. *FEBS Lett*. 459:358–362.

33. Sajid, M. and McKerrow, J. H. (2002) Cysteine proteases of parasitic organisms. *Molecular and Biochemical Parasitology.* 120:1-21.

34. Department of Chemistry, University of Maine website. The Cysteine Proteases. Updated March 2014. Accessed June 2016. http://chemistry.umeche.maine.edu/CHY431/Peptidase10.html

35. Molyneux, D. H., Savioli, L., & Engels, D. (2017). Neglected tropical diseases: Progress towards addressing the chronic pandemic. *Lancet*. 389:312–325.

36. MacLean, L., Reiber, H., Kennedy, P. G., and Sternberg, J. M. (2012). Stage progression and neurological symptoms in Trypanosoma brucei rhodesiense sleeping sickness: Role of the CNS inflammatory response. *PLoS Neglected Tropical Diseases*. 6:1857.

37. Franco, J. R., Simarro, P. P., Diarra, A., and Jannin, J. G. (2014). Epidemiology of human African trypanosomiasis. *Journal of Clinical Epidemiology.* 6:257–275.

38. Ferreira, L. G. and Andricopulo, A. D. (2017) Targeting cysteine proteases in trypanosomatid disease drug discovery. *Pharmacology and Therapeutics*. 180:49-61

39. Eakin A.E., Mills, A.A., McGrath M.E., McKerrow J.H., Fletterick R.J., Craik C.S. (1992) The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. *Journal of Biological Chemistry*. 267(11): 7411-7420.

40. Berman, H. M. et al. (2000). The Protein Data Bank. Nucleic Acids Research. 28:235–242.

41. Gillmor S.A., Craik C.S., Fletterick R.J. (1997) Structural determinants of specificity in the cysteine protease cruzain. *Protein Sci*. 6:1603-1611.

42. McGrath M.E., Eakin A.E., Engel, J.C., McKerrow, J.H., Craik, C.S. and Fletterick, R.J. (**1995**) The crystal structure of cruzain: a therapeutic target for Chagas' disease. *Journal of Molecular Biology* 247:251-259.

43. Brinen, L.S., Hansell, E., Cheng, J., Roush, W.R., McKerrow, J.H. and Fletterick, R.J. (2000) A target within the target: probing cruzain's P1′ site to define structural determinants for the Chagas' disease protease. *Structure*. 8:831–840.

44. da Silva, E. B., do Nascimento Pereira, G. A., & Ferreira, R. S. (2016). Trypanosomal cysteine peptidases: Target validation and drug design strategies. In S. Müller, R. Cerdan, & O. Radulescu (Eds.), *Comprehensive analysis of parasite biology: From metabolism to drug discovery: From metabolism to drug discovery* (pp. 121–145). Weinheim: Wiley-VCH Verlag GmbH & Co.

45. Doyle, P. S., Zhou, Y. M., Engel, J. C., and McKerrow, J. H. (2007). A cysteine protease inhibitor cures Chagas' disease in an immunodeficient-mouse model of infection. *Antimicrobial Agents and Chemotherapy*, 51(11), 3932–3939.

46. Berasain, P., Carmona, C., Frangione, B., Cazzulo, J. J., and Goni, F. (2003). Specific cleavage sites on human IgG subclasses by cruzipain, the major cysteine proteinase from Trypanosoma cruzi. *Molecular and Biochemical Parasitology* 130:23–29.

47. Kovaltsuk A., Krawczyk K., Galson J.D., Kelly D.F., Deane C.M. and Trück J (2017) How B-cell receptor repertoire sequencing can be enriched with structural antibody data. *Front. Immunol*. 8:1753.

48. Helman U., Wernstedt C., and Cazzullo J.J. (1991) *Molecular and Biochemical Parasitology* 44:15-22.

49. Martinez J. and Cazzulo J.J. (1992) Anomalous electrophoretic behaviour of the major cysteine proteinase (cruzipain) from *Trypanosoma cruzi* in relation to its apparent molecular mass. *FEMS Microbiology Letters* 95: 225-230.

50. McKerrow J.H., Engel J.C., Carrfey, C.R. (1999) Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorganic and Medicinal Chemistry* 7:639-644

51. Francisco, A. F. et al. (2016) Nitroheterocyclic drugs cure experimental *Trypanosoma cruzi* infections more effectively in the chronic stage than in the acute stage. *Sci. Rep*. 6:35351;

52. DNDi. Drugs for neglected diseases initiative. Fexinidazole (Chagas). Updated Feb. 2019. Accessed Feb. 2019. http://www.dndi.org/diseases-projects/portfolio/fexinidazole-chagas/

53. Engel, J. C., Doyle, P. S., Hsieh, I., & McKerrow, J. H. (1998). Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection. *The Journal of Experimental Medicine*. 188(4):725–34.

54. Russell, P. J*., iGenetics.* Pearson Education, 2002, Chapters 6 and 7.

55. Cooper, G.M. *The Cell: A Molecular Approach* 2nd ed.; Geoffrey M. Cooper, 2000, Ch 2.

56. Anfinsen, C.B.; Haber, E., Sela, M. and White, Jr. F.H. (1961) The kinetics of the formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A*. 47:1309-1314.

57. Buck, T.M., Wright, C.M. and Brodsky J.L. (2007) The activities and function of

molecular chaperones in the endoplasmic reticulum. *Seminars in Cell and Developmental Biology* 18:751-761.

58. Appenzeller-Herzog, C.B. and Ellgard, L. (2008) The human PDI family: Versatility packed into a single fold. *Biochim. Biophys. Acta* 1783:535-548

59. Sitia, R. and Braakman, I. (2003) Quality control in the endoplasmic reticulum protein factory. *Nature.* 426:891-894.

60. Ellgard, L. and Ruddock, L.W. (2005) The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep.* 6:25-32

61. Yoshida, Hiderou (2007) ER stress and diseases. *FEBS J*. 274:630-658

62. QIAGEN® pQE-30 Plasmid Map. Accessed Dec 2015. Copyright by QIAGEN 2015-2019. https://www.qiagen.com/us/resources/resourcedetail?id=f057bb24-5b3e-441c-aca8-d34b0b039418&lang=en

63. Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*. 3(3):318-356

64. QIAGEN® Plasmid Purification Handbook. April 2012. Copyright by QIAGEN 2005-2012.

65. Ghisaidoobe, A.B.T. and Chung, S.D. (2014) Intrinsic tryptophan fluorescence in the detection and analysis of proteins: A focus on Förster resonance energy transfer techniques. *Int. J. Mol. Sci.* 15:22518-22538

66. Narayan M, Welker E, Scheraga HA (2001) Development of a novel method to study the rate-determining step during protein regeneration: application to the oxidative folding of RNase A at low temperature reveals BPTI-like kinetic traps. *J Am Chem Soc* 123:2909–2910

67. Narayan, M., Welker, E., Wedermeyer, W.J., and Scheraga, H. (2000) Oxidative Folding of Proteins. *Accounts of Chemical Research* 33:805-812

68. Rothwarf, D.M. and Scheraga, H.A. (1993) Regeneration of bovine pancreatic ribonuclease A steady-state distribution. *Biochemistry* 32:2671-2679

69. Ellman GL (1959) *Arch. Biochem. Biophys*. 82 (1):70–77

70. Reif, M. M., Mach, L., & Oostenbrink, C. (2012). Molecular insight into propeptide-protein interactions in cathepsins L and O. Biochemistry. 51(43):8636–53.

71. Eder, J., Rheinnecker, M. and Fersht, A.R. (1993) Folding of Subtilisin BPN': Characterization of a Folding Intermediate. *Biochemistry*. 32:18-26.

72. Groves,M.R., Coulombe,R., Jenkins,J., and Cygler,M.(1998).Structural basis for specificity of papain-like cysteine protease proregions toward their cognate enzymes. *Proteins*. 32: 504–514.

73. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*. *215*(3):403–10.

74. Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*. *5*(4):725–38.

75. Lerner, C. G., Kobayashi, T., & Inouye, M. (1990). Isolation of subtilisin pro-sequence mutations that affect formation of active protease by localized random polymerase chain reaction mutagenesis. *The Journal of Biological Chemistry*. 265(33):0085–20086.

76. Ruan, B., Hoskins, J., and Bryan, P. N. (1999) *Biochemistry* 38(26):8562-8571

VITA

Veronica Gonzalez was born only daughter of M.P. Camacho and granddaughter to Mr. and Mrs. E. Camacho in El Paso, TX. She graduated from Stephen F. Austin high school in May of 1999 and entered college at the University of Texas at El Paso in the fall of that year as a chemistry major. In 2003 she participated in the NASA Faculty Fellowship summer program as a student research participant along with Dr. Elizabeth Gardner at White Sands Test Facility in Las Cruces, NM. In the fall of 2003, Veronica entered Dr. Boguslaw Stec's protein crystallography laboratory as an undergraduate research assistant to study the role of inositol monophosphatese (IMPase), an enzyme believed to play a role in bipolar disorder. She also participated in AmeriCorps national volunteer program, completing over 1500 community service hours. She graduated from UTEP in the spring of 2005 with bachelors of science degrees in both Chemistry and Biology. In the spring of 2005, Veronica entered graduate school at UTEP to study chemistry and conducted research in Dr. Mahesh Narayan's biochemistry laboratory. Her thesis titled "The Role of Protein Disulfide Isomerase (PDI) in Oxidative Folding" was a study conducted to better understand PDI's contribution to neurodegenerative disorders. She entered the doctoral program soon after to study the oxidative folding of cruzain, the cysteine protease involved in the progression of Chagas disease. She became a faculty member of El Paso Community College as an adjunct professor in 2009, was hired as tenure-track faculty member in 2014 and earned tenure in 2019. In 2016 she was awarded the NIH Faculty Summer Sabbatical Fellowship and conducted research at the University of Arizona Cancer Center under the mentorship of Dr. Marty Pagel. She has mentored LSAMP and RISE scholars from both EPCC and UTEP, some of which are currently graduate students. She plans to continue to conducting research and mentoring students.