2016-01-01

Identifying Non-Classical Active Sites As A Tool For Enzyme Inhibition

Marisol Serrano
University of Texas at El Paso, mari_colet@hotmail.com

IDENTIFYING NON-CLASSICAL ACTIVE SITES AS A TOOL FOR ENZYME INHIBITION

MARISOL SERRANO

Doctoral Program in Chemistry

APPROVED:

Mahesh Narayan, Ph.D., Chair

Russell Chianelli, Ph.D.

Suman Sirimulla, Ph.D.

Jose Peralta-Videa, Ph.D.

Charles Ambler, Ph.D.

Dean of the Graduate School

DEDICATION

First of all, I would like to thank God for allowing me to finish my doctoral studies, for have given me the strength and wisdom I needed. To my husband Victor and daughter Valentina for allowing me taking their valuable time so I could finish my career. To my family and friends for their support and help throughout this journey.

Philippians 4:13

IDENTIFYING NON-CLASSICAL ACTIVE SITES AS A TOOL FOR ENZYME INHIBITION

by

MARISOL SERRANO, 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 2016

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Mahesh Narayan, for his support, wisdom advice, his patience, and for allowing me to grow in my career while pursuing my doctoral studies. I personally loved his sense of humor which made this journey less stressful. I would like to thank by lab colleagues from past and present: Veronica, Emmanuel, Pal, Parijat, Jose, Louis, Gabriela, Alejandra, Denisse, Maricarmen, Rene, Karina, Marisela, Brian, and Cristina. Thank you all for your help. I would like to extend my gratitude to Veronica, Emmanuel, Pal, and Parijat for being more than a lab mate; for their friendship, helpful suggestions, and scientific discussions. Thank you also to all the graduate students from other labs: Aidee Medina, Alfredo Ornelas, Gustavo Avila, Andrew Pardo, and Adrian Enriquez who were always willing to aid.

I would also like to thank my committee members Dr. Russel Chianelli, Dr. Jose Peralta-Videa, and Dr. Suman Shirimulla for accepting being part of my committee. Thanks to Dr. River, Dr. Bernal, and Dr. Michael for allowing me to use some of their facilities to undergo this research. Special thanks to Dr. Ray and Dr. Fratev for their collaboration in this study and their vital suggestions to enrich this research project.

I would also like to thank the Chemistry Department, for providing financial support to perform my research. Thanks to Lucema and Grace for their service to the chemistry department. Thanks to the BBRC at the Biology Department for the facilities provided and to the staff: Ana Betancourt, Dr. Kyle Johnson and Dr. Armando Varela. Many thanks to El Paso Pain Center for the monetary support.

Lastly, I would like to thank my lovely family and friends for their continuing support and love. I thank my husband Victor for his patience, support, wisdom advices, and endless love. Thanks to my mother, mother and law, sister, and sister in law for taking care of my daughter. This journey would have been very hard without your help. Infinite thanks to all my family and close friends.

ABSTRACT

Chagas disease, caused by the parasite Trypanosoma cruzi, is an endemic life-threatening disease that affects mainly the heart. It remains the leading cause of heart failure in Latin American countries. Since current treatments against this parasite are highly toxic and somewhat ineffective, novel and more efficacious types of interventions are desired. Cruzain, identified as the major cathepsin for *T. cruzi*, plays a major role in the parasite's life cycle; making this enzyme very attractive for potential trypanocidal drugs discovery. The recombinant cruzain is synthesized as a zymogenic pro-protein (PCZN) which possesses a pro domain and a catalytic domain. In this study we worked with the zymogen (inactive) form of the protease and, by utilizing bioinformatics tools, we identified undiscovered structure-impacting sites within the pro domain of pro-cruzain. PCZN DNA constructs containing point site mutations were successfully expressed and purified. Real-time catalytic activity assays and secondary structure analysis of mutants of PCZN revealed a structural and functional impact compared to the wild- type (WT-PCZN). These findings could be used to identify promising targets to aid in the development of future anti-chagastics agents as well as to support the development of a host of small-molecule compounds that could interfere in the maturation of the catalytically active state of the enzyme by disrupting critical aminoacid-aminoacid interactions within the zymogen. The outcome will possibly open the field of inhibitor design by advancing a novel, bottom-up approach, rather than the classical top-down approach that targets only the active-site of the enzyme.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

CHAGAS DISEASE AND IMPORTANCE OF PRO-CRUZAIN

1.1 Background

1.1.1 What is Chagas disease?

Chagas disease was named after the Brazilian physician Carlos Justiniano Ribeiro Chagas (1879 - 1934), who first described it in 1909 [1]. Also known as "The American Trypanosomiasis", it is caused by the flagellated Protozoan *Trypanosoma cruzi* (**Error! Reference source not found.**). The infection occurs when the parasite is transmitted by the feces of a triatomine insect vector (or "kissing" bug) which defecates after sucking blood at night [2] (**Error! Reference source not found.**). Common triatomine vector species for trypanosomiasis belong to the genera *Triatoma*, *Rhodnius*, and *Panstrongylus.* The insect resides in crevices in the walls and roofs of poorly constructed houses, usually in rural and periurban areas throughout Latin America. The parasite is transmitted when a person inadvertently enables the parasite-contaminated feces to make contact with any break in the skin (including the bite), the eyes or mouth. Other modes of transmission include transfusion of infected blood, oral transmission through contaminated food, vertical transmission, organ transplantation and even in laboratory accidents [3].

Figure 1. *Trypanosoma cruzi* (Gruby, 1843).

Figure 2. Sylvatic form of Triatoma infestans, the main vector of Chagas disease in the Southern Cone countries [4]

Infection by *T. cruzi* is characterized by an accumulation of host tissue damage over several years as the parasites sustain their life cycle by infecting host cells, multiplying intracellularly, and rupturing the cells to reinfect new cells. The complex disease has both acute and chronic phases [5]. The acute phase, which may pass almost unnoticed, starts with a local inflammatory lesion appearing in the site where the metacyclic trypomastigotes enter and undergo their first rounds of multiplication. After dissemination,

symptoms of cardiac insufficiency develop in a small number of patients, occasionally it can lead to fatal meningoencephalitis or acute myocarditis (mostly in very young children). Most deaths from acute Chagas disease are due to heart failure because the heart is the organ most commonly involved, and sudden death due to cardiac dysrhythmias may occur. Also, an indeterminate asymptomatic phase may be present, which can last for about 10 - 20 years or even for the whole lifetime of the infected person. In perhaps 20 - 30 % of the cases, a chronic phase may occur, consisting of heart problems eventually leading to sudden death, and/or enlarging of hollow viscera, like the esophagus and the colon (megaesophagus and megacolon), which can also be fatal [2,5–7]. In addition, individuals with Chagas heart disease often develop mural thrombi which embolize and cause cerebrovascular accidents [8, 9].

The use of synthetic irreversible cysteine protease (CP) inhibitors in animals experimentally infected with *T. cruzi* has significantly reduced parasitemia and mortality, validating these enzymes as promising targets for the development of new drugs[10]. Several studies have supported the fact that the main target of the development of an inhibitor is the well-characterized major CPs of *T. cruzi*, cruzipain (or cruzain) [11–13].

In many pathogenic parasites, the activity of papain like CPs seems to be crucial for growth, development and tissue ⁄ host cell penetration and several targets within *T. cruzi* have been exploited by the current arsenal of antichagasics which work by inhibiting different targets within the parasite [11, 14]. However, currently available trypanocidals present side effects or lack the desired potency. Several have become obsolete via the parasite defense system (through drug resistance and evasion).

For example, nitrofurans and nitroimidazoles are unsatisfactory due limited efficacy and are linked to mammalian host toxicity [15, 16]. Nitroheterocyclics like Benznidazole and nifurtimox are only available medications for acute-phase Chagas' disease, but are highly toxic and have poor efficacy in long-lasting chronic infections [17–19]. Antifungal sterol biosynthesis inhibitors are not powerful enough to

induce parasitological cures of human or experimental infections [14, 15, 20]. Previous efforts have identified vinyl sulfones, sulfonates, and sulfonamides as high-affinity cruzain inhibitors [21, 22]; one of these vinyl sulfones, K11777, is currently undergoing Investigational New Drug enabling studies [23, 24]. α-ketoamide-, α-ketoacid, α -ketoester-, aldehyde-, and ketone-based inhibitors have also been described [25, 26]. No extensive studies of the long-term sequellae of these therapeutics have been conducted in humans, but several reports of neuropathy and tumorigenic or carcinogenic effects have been described [17, 18].

While these successes are encouraging, many potential drugs, including those that enter clinical trials, ultimately fail to gain approval [27], and those that are approved are subject to growing parasitic resistance. Furthermore, studies have demonstrated that cysteine proteinase inhibitors have trypanocidal activity with negligible mammalian toxicity [28]. Efforts to develop a vaccine against *T. cruzi* have also failed thus far, likely because the disease pathology has an autoimmune component [17]. Consequently, a diverse set of inhibitory scaffolds that can be optimized into distinct therapeutic candidates is urgently needed.

1.1.2 Epidemiology

Chagas disease is prevalent in most of Latin America, where it affects an estimated number of 16-18 million people. Although Chagas disease continues to persist, the estimated number of infected people has fallen from approximately 20 million in 1981 to around 10 million in 2009. The risk of transmission has been reduced by introducing vector-control measures and safer blood transfusions in Latin America. Population mobility has carried the disease to regions where the disease was previously unknown. In past decades, it has been increasingly detected in other non-endemic countries in the Region of the Americas (Canada and the United States of America), the Western Pacific Region (mainly Australia and Japan) and the European Region (mainly in Belgium, France, Italy, Spain, Switzerland and the United Kingdom, but

also in Austria, Croatia, Denmark, Germany, Luxembourg, the Netherlands, Norway, Portugal, Romania and Sweden) [3]. It is also endemic in Texas with an autochthonous canine cycle, abundant vectors (Triatoma species) in many counties, and established domestic and peridomestic cycles which make competent reservoirs available throughout the state. Yet, Chagas disease is not reportable in Texas, blood donor screening is not mandatory, and the serological profiles of human and canine populations remain unknown [29]. Although, reports estimate between 300,000 – 1 million Chagas-carrying men, women and children currently residing in the Unites States of America making it "The new HIV/AIDS of the Americas" [20].

1.1.3 Cruzain or cruzipain

The name cruzipain was proposed to denote that it constitutes the major cysteine protease activity in *T. cruzi* [30], and that it belongs to the C1 papain-like enzyme family [31]. The name cruzain was suggested afterwards by Eakin *et al* denoting the recombinant active truncated form of the enzyme expressed in *Escherichia coli* [32]. The protease was initially purified by Rangel *et al.* (1981) and was immunologically labeled on the three developmental stages of *T. cruzi*, epimastigotes (insect forms), trypomastigotes (bloodstream forms), and amastigotes (intracellular forms) [33, 34]. Therefore cruzain is differentially expressed in all stages of the parasite's life cycle, and its activity being at least 10-fold higher in epimastigotes, as compared with the other parasite forms [35–37]. During the life cycle of *T. cruzi*, cruzain, can be found in the lysosomes, reservorosomes, and the cell membrane [11, 38].

The three-disulfide-bond containing enzyme, as a major lysosomal proteinase, plays an important role in the parasite's nutrition. Cruzain's participation in blood breakdown makes it critical for parasites survival, replication and proliferation [39]. Additionally, there is indirect evidence for a possible participation of the enzyme in penetration of the trypomastigote into the host cell in the defense of the

parasite against the immune system of the mammal [40]. Recent studies have shown that cruzain is able to activate the cell death program, by direct activation of caspase zymogens 3 and 7 [41].

Cruzipain is expressed as pre-pro-cruzain, which is an inactive precursor that possesses a signal peptide, a pro domain, a mature domain (that retains catalytic activity), and a C-terminal extension (CTE). Cruzipain refers to the full-length native enzyme while cruzain refers to the mature recombinant form (without pre-pro and CTE) [32]. In this work, for simplicity, we will call the recombinant pro and catalytic (mature) domain complex as pro-cruzain (PCZN).

The unusual C-terminal extension (CTE) of cruzipain has been found only in cysteine proteases from related parasites and certain plants [43]. The function of the CTE is currently unknown, but it apparently does not play a role in catalysis, enzyme inhibition, protein folding, or trafficking to the lysosome [42, 44, 45]; however this domain accounts for the high antigenicity of intact glycoprotein [46]. On the other hand, the pro domain of cysteine proteases has two well defined functions: 1) maintaining the enzyme in an inactive form (zymogen) until it reaches an appropriate site of protease function by binding the enzyme's active site, and 2) functioning as a structural template to ensure proper folding during translation [32, 47]. The pro domain can also act as a reversible inhibitor and a stabilizer of the mature protease [48] and has also been implicated in protease precursor trafficking in the endosomal-lysosomal system [49, 50].

Cruzain, like all cysteine proteases, contains a catalytic triad composed of three residues: Cys 25, His 159, and Asn 175[11, 38]. The activation is achieved by proteolytic excision of the prodomain, an event believed to occur in vivo by a multistep process that may involve multiple endosomal / lysosomal peptidases or even proteases present in the extracellular environment [51].

It has been known for decades that molecular chaperones are involved in protein folding. Although, some proteins have evolved to contain a specific sequence as an intramolecular chaperone

(IMC) or pro-peptide, which is essential for protein folding but not required for protein function, as it is removed after the protein is folded by autoprocessing or by an exogenous protease. To date, a large number of pro-peptides from various proteins have been identified to function as an intramolecular chaperone to assist the folding of the respective functional domains [52, 53]. An increasing amount of evidence has revealed that pro-peptides play an important role in protein folding both *in vivo* and *in vitro*. Therefore, mutational studies within these pro-peptides are necessary to explore its importance in more detail.

1.2 Specific Aims

As mentioned in the introduction of this chapter, cruzain is considered the main target for the development of anti-chagastics drugs. However, the mature enzyme and the pro domain have been well investigated as separate peptides while the full length pro-cruzain (zymogenic) has been understudied. The long term aim of this project is to generate recombinant pro-cruzain using a modified protein expression procedure.

Specific aim one: To create plasmids encoding for wild type and mutants of pro-cruzain. Traditional cloning and sub cloning techniques were utilized.

Specific aim two: To generate recombinant proteins by customized expression and purification techniques using transformed *Escherichia coli* cells.

1.3 Materials and Methods

1.3.1 Molecular Cloning of PCZN WT and Mutants

The recombinant PCZN-WT construct was produced based on the PCZN gene sequence obtained from the *T. cruzi.,* Tulahein 2 strain (Genebank accession M84342)[32]. The coding sequence of 1011bp of PCZN-WT was first cloned into pQE30Xa (Qiagen) for expression and to pUC57 for cloning by GeneScript Piscataway, NJ. USA. However, a signal peptide (underlined sequence in Figure 3) present in the PCZN-WT

cDNA insert had to be excluded in order to produce soluble protein. DNA sequence encoding for pro and catalytic domains was copied by PCR using the following primers: Forward- StuI: 5'- GTGT AGG CCT TGT CTG GTC CCG GCT -3' and reverse-HindIII: 5'- GTGT AAG CTT TTA ACC CAC AAC TGC ACT AGA GGC -3' (Eurofins MWG Operon). PCR components were: 75µL of DNase free water, 2uL of forward and reverse primers (10uM), 20µL of Taq 5x master mix (New England Biolabs), and 1 µL of DNA (0.01ng/µL final concentration). The PCR conditions were as follows: one cycle of denaturation for 5 min at 95˚C, followed by 40 cycles of 1 min of denaturation at 95˚C, 1 min of annealing at 60˚C, and 1 min of extension at 72˚C; then 10 min of long extension at 72˚C, to finish with hold at 4˚C. The PCR product was observed by agarose gel electrophoresis. The gene fragment was isolated from the agarose gel and purified with QIAquick Gel Extraction Kit (Qiagen).

ATGTCTGGCTGGGCGCGTGCGCTGTTGCTCGCGGCCGTCCTGGTCGTCATGGCG**TGCCTCGTCCCCGCGGCGACGGC**
GAGCCTGCATGCGGAGGAGACGCTGACGTCGCAATTCGCAGAATTCAAGCAGAAGCATGGCAGGGTGTACGAGA
GCGCCGCGGAGGAGGCGTTCCGCCTGAGCGTGTTCAGGGAGAACCTGTTTCTTGCGAGGCTGCACGCCGCGGCAA
ACCCACACGCGACCTTCGGCGTCACGCCCTTCTCGGACCTCACGCGCGAGGAATTCCGGTCCCGCTACCACAACGGC
GCGGCGCACTTTGCGGCGGCGCAGGAGCGCGCGAGAGTGCCGGTGAAGGTGGAGGTAGTTGGC*GCGCCCGCGGC*
AGTGGATTGGCGTGCGAGAGGCGCCGTGACAGCCGTCAAGGACCAGGGCCAATGCGGTTCGTGCTGGGCCTTCTCC
GCCATTGGCAACGTTGAGTGCCAGTGGTTTCTTGCCGGCCACCCGCTGACGAACCTGTCGGAGCAGATGCTCGTGTCG
TGCGACAAAACGGACTCTGGCTGCAGTGGTGGCCTGATGAACAACGCCTTTGAGTGGATTGTGCAGGAGAATAACG
GCGCCGTGTACACGGAGGACAGCTACCCTTATGCGTCGGGCGAGGGGATATCGCCGCCGTGCACGACGTCAGGCCAC
ACGGTGGGTGCCACGATTACCGGTCACGTTGAATTACCTCAGGACGAGGCCCAAATAGCCGCATGGCTTGCAGTCAA
TGGCCCGGTTGCCGTTGCCGTCGACGCCAGCAGCTGGATGACCTACACGGGCGGCGTTATGACGAGCTGCGTCTCCG
AGCAGCTGGATCACGGCGTTCTTCTCGTCGGCTACAATGACAGCGCCGCAGTGCCGTACTGGATCATCAAGAACTCGT
GGACCACGCAGTGGGGCGAGGAAGGCTACATCCGCATTGCAAAGGGCTCGAACCAGTGCCTTGTCAAGGAGGAGGC
GAGCTCCGCGGTGGTCGGT

Figure 3. DNA sequence of PCZN obtained from the *T. cruzi.,* Tulahein 2 strain [32]**.** Underlined: signal peptide, bold: pro domain, and cursive: catalytic domain.

1.3.1.1 Ligation of pGEM-T and pQE30-Xa vectors with PCZN-WOSP.

To prepare the pGEM-T cloning vector and pQE30-Xa expression vector and the insert PCZN-WOSP for ligation they were digested at HindIII and StuI sites in separate reactions. The digestion reaction mix was composed of 16.5µL of steril miliQ water, 5µL of 10X multicore buffer, 0.5µL of acetylated beta serum albumin (BSA) (10µg/µL), DNA sample (1µg) 4µL HindIII and StuI (10 units/µL) from Promega, incubation time of 1.5 hrs at 37°C. Then the digested products were gel purified as well. Ligation reaction of vector and insert was performed as follows: 5µL of 2X rapid ligation buffer, 2 µL of PGEM-T vector or pQE30Xa vector (10ng), 2µL of the insert (16 ng), 1 µL of T4 DNA ligase (3 units/µL) was mixed and incubated for 5 min at room temperature. A control insert DNA (4ng/µL) was used as a positive control; water was used as negative control. *E. coli* DH5alpha cells were then transformed with the ligated product and many replicates of the plasmids were purified (Qiagen miniprep kit), and subsequently a double digestion HindIII and StuI was performed. The digested products were then visualized in a 1% agarose gel.

In order to select transformed colonies containing the pQE30-Xa-PCZN-WOSP construct, a PCR colony selection was performed. Three freshly transformed colonies were selected and minipreped, then the cDNA was subjected to PCR using forward Type III/IV: 5'-CGGATAACAATTTCACACAG-3' and reverse: 5’-GTTCTGAGGTCATTACTGG-3’ primers. PCR products were then visualized in a 1% agarose gel. Colonies showing the PCZN-WOSP insert were preserved with 20% glycerol at -80C. The DNA sequence was confirmed using forward Type III/IV primer: 5'-CGGATAACAATTTCACACAG-3' and reverse primer: 5’-GTTCTGAGGTCATTACTGG-3’ on a 3130xl Genetic Analyzer (GeneAmp PCR System 9700, Life Technologies). DNA concentration was measured using NanoDrop instrument (Thermo Scientific Nano Drop 1000, software ND 100 v3.6.0).

Since PCZN-WT-pQE30-Xa expression system was unable to produce recombinant protein. We decided to utilize a plasmid donated by Dr. Ana Paula Lima from Federal University of Rio de Janeiro, Brazil; Biophysics Institute Carlos Chagas Filho. Plasmid pQ30-PCZN was sequenced with forward Type III/IV primer: 5'-CGGATAACAATTTCACACAG-3' and reverse primer: 5'-GTTCTGAGGTCATTACTGG-3'. PCZN-WT DNA sequence was inserted into pQE30 vector system (Qiagen, pQE trisystem vector) at the HindIII and BamHI restriction sites (Promega) by Genescript USA. Colonies containing the sub-cloned products PCZN-pQE30 were also selected by PCR colony selection as previously described. The DNA sequence was confirmed using pQE vector primers: forward Type III/IV primer 5'-CGGATAACAATTTCACACAG-3' and reverse primer 5'-GTTCTGAGGTCATTACTGG-3'. PCZN-WOSP will be named PCZN-WT (wild type) for the rest of the thesis. PCZN-WT cDNA was subject to site-directed mutagenesis to generate mutants: E25V, F32A, F35A, E49V, F56A, D82V, E86V, E87V, and C141A by GeneScript. DNA sequence of each construct was confirmed using primers: forward 5' CGGATAACAATTTCACACAG 3', and reverse 5' GTTCTGAGGTCATTACTGG 3' (Eurofins MWG Operon, AL, USA).

1.3.2 Protein expression of PCZN-WT and mutants (D82V, E87V, E25V, C141A).

Plasmid pQE-30 containing 6xhistidine-tagged PCZN (WT and mutants) at the N-terminus was introduced into competent *E. coli* DH5α for cloning and into *E. coli* M15 for expression (Invitrogen). A modified version of a method previously described[54] was used to identify and obtain high-expressing colonies for clones containing the 6xHis-tagged protein. A 5mL LB culture with antibiotics (amp 100µM/kan 50µM) 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/km plate and grown O.N. at 37˚C. From this plate, 3 colonies were selected and cultured each one in a 5mL LB/amp/kan culture and grown O.N. at 37˚C. A small scale expression, described below, was performed for each 5mL culture. Before induction with

Isopropyl-B-D-1-thiogalactopyranoside (IPTG), 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.

1.3.2.1 Small scale expression

Luria Broth (LB) media (5mL) 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 min at 1500rpm. 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 (1mM) and incubated at 37˚C for 4 hrs. Expression levels were checked by spinning down 250uL of cells suspension at 4400 rpm for 5mim. The pellet was re-suspended in 100µL lysis buffer, then 20µL of lysate was mixed with 20µL of sodium dodecyl sulfate (SDS) loading buffer and heated at 95˚C for 5 min. The sample was centrifuged and supernatant was loaded into SDS-PAGE to check the expression level.

Minimal medium was composed of M9 salts (64g Na_2HPO_4, 15g KH_2PO_4, 2.5g NaCl, and 5g NH_4Cl into 1000mL of MiliQ water), 1M $MgSO_4$, 20% glucose, 1M $CaCl_2$. The $MgSO_4$, glucose, and $CaCl_2$ were sterilized with a 0.2µm Milipore Stericup vacuum filtration system. To prepare the minimal media, 200ml of M9 salts were added to 700ml of sterile MilliQ, then 2ml of 1M $MgSO_4$, 20ml of 20% glucose, 100µl of 1M $CaCl_2$ were added last then the volume was adjusted to 1000ml with sterile miliQ water.

1.3.2.2 Large scale expression

One colony selected from the plate with the highest expression levels was inoculated into 100mL of LB media with amp(100µM)/kan(50µM) antibiotics and grown O. N. at 37˚C, then the cells were gently spun down at 4400 rpm for 15 min at room temperature and the pellet was re-suspended in 500mL 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 hrs (1mM IPTG).

1.3.3 Cell Lysis and Protein purification

After induction the cells were collected by centrifuging at 5000rpm for 15min at 4˚C using a Sorvall Centrifuge RC28S (Wilmington, DE, USA) with GSA rotor, and lysed in 100mM NaH_2PO_4, 10mM Tris-HCl, 10mM imidazole, and 8M urea (pH 8), under agitation, for 30 min at 4°C. The cell lysate formed was further disrupted by sonication at 60 W for 3 x 30s pulses (with 1min delay between pulses) on ice bath, with a Branson Sonifer 450. The soluble material was recovered by centrifugation at 11000rpm for 20 min 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 washing buffer (100mM NaH_2PO_4, 10mM Tris-HCl, and 8M urea (pH 6.3). The recombinant proteins were then eluted with elution buffer using gradient concentrations of imidazole (10, 30, 50, and 200mM). Then the purity of recombinant proteins were 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 min, 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 µm 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 63,473$cm^{-1}M^{-1}$and molecular weight of 35.403kDa.

1.3.3 Protein Refolding

Aliquots with pure PCZN protein were combined and a final concentration of 30uM 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 (250mM Arginine, 100mM Tris HCl, 1mM EDTA and 20% Glycerol, pH 8.0) and incubated under slight agitation at 4˚C for 24 hrs overnight as previously described [11]. The sample was dialyzed using a 3kDa MWCO Ultracel membrane filter (Millipore) prior to being concentrated to a final volume of 20mL. The refolded protein was split into two: one batch was used in the activation studies and the other for the CD studies (Chapter 3).

1.4 Results and Discussion

1.4.1 Molecular Cloning of WT and Mutants of PCZN

PCZN-WT DNA was inserted into pQE30-Xa initially by GenScript which contains a Xa recognition site used for cleavage of the His-tag (Figure 9A). This plasmid was intentionally used to remove the 6xHis-tag after recombinant protein was expressed. Unfortunately, the cDNA sequence recognized to encode for a signal peptide caused the protein to be expressed as inclusion bodies (Figure 11). Hence we choose to keep the DNA fragments encoding for pro and catalytic domains only and sub-clone them into pQE30-Xa.

1.4.1.1 Removal of the signal peptide from PCZN-pQE30-Xa plasmid

First, we had to remove the signal peptide from the PCZN-pQE30-Xa construct. This was done amplifying only the DNA fragments encoding for pro and catalytic domains from the full-length PCZN that contained the signal peptide. The PCZN gene without signal peptide portion (PCZN-WOSP) was successfully amplified via PCR, as shown in Figure 4, where the 957bp corresponding to the PCZN gene

was visualized in a 1% agarose gel. The amplified PCR product was then cut by StuI and HindIII for future ligation with pGEM-T.

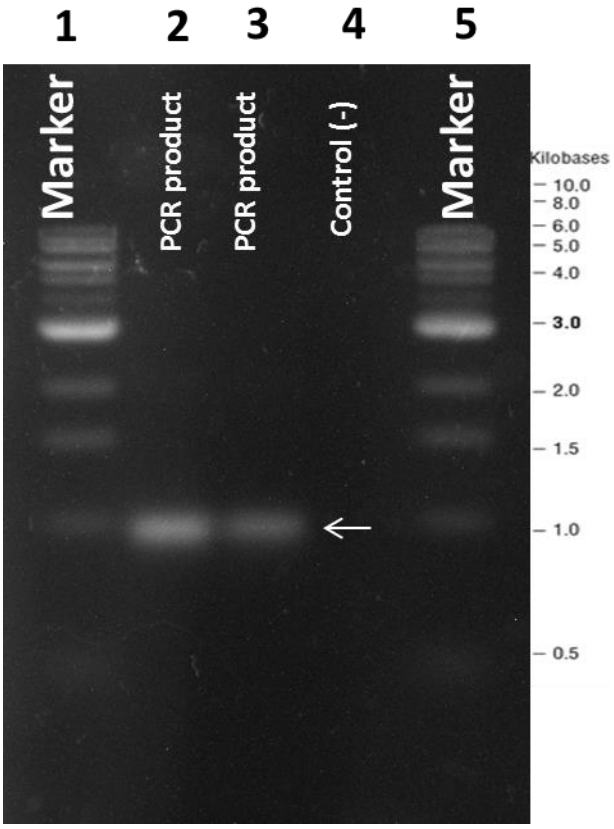

Figure 4. **PCR product of PCZN-WOSP.** 1% agarose gel showing the cDNA encoding for pro-cruzain without the signal peptide fragment. Line 1 and 5: DNA marker (1Kb DNA ladder, New England Biolabs), line 2-3: PCR product, line 4: DNase free water used as negative control.

1.4.1.2 Ligation of cloning vector pGEM-T easy with insert PCZN-WT-WOSP:

The pGEM-T cloning vector was digested at HindII and StuI sites prior to ligation. Then the ligation of pGEM-T and PCZN-WOSP was performed using T4 Ligase as mentioned above. PGEM-T was successfully

cloned and the construct was confirmed by digestion with the above-mentioned sites and visualized in a 1% agarose gel as shown in Figure 5. Bands belonging to pGEM-T vector and PCZN_C25A_WOSP gene are observed at around 3015bp and 957bp respectively (lines 3-5). Line 2 pGEM-T corresponds to the undigested pGEM vector, showing a pattern characteristic to the different forms that plasmids make in agarose gels when they are undigested.

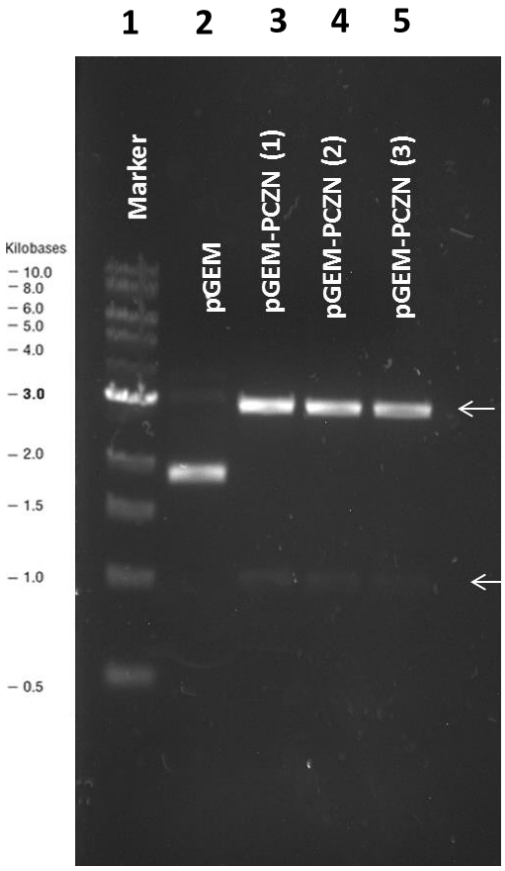

Figure 5. Enzyme digestion of pGEM-T-PCZN_WOSP ligation reaction. Line 1: DNA marker, line 2: undigested pGEM-T vector, lines 3-5: pGEM-T-PCZN_WOSP plasmid digested at HindII and StuI sites.

1.4.1.3 Ligation of pQE30-Xa with insert PCZN-WT-WOSP

Before ligation of pQE30-Xa with insert PCZN-WT-WOSP was performed, the two components had to be digested. Figure 6 shows the successfully digested products at HindIII and StuI sites. Band at around 3Kb belongs to the digested vector while the bands at around 1Kb correspond to the undigested and digested PCZN-WT-WOSP gene, lines 6 and 7 respectively. Digested components were then ligated using T4 Ligase as well. Ligation of plasmid was verified by PCR colony selection from the transformed cells with ligated product. Successfully amplified products were then subject to digestion (HindIII/StuI). The PCR colony selection using vector primers enabled us to verify the PCZN-WT-WOSP was surely inserted into the pQE30-Xa vector. Line 1 of Figure 7 shows the band corresponding to the DNA that encodes for PCZN-WT-WOSP. On the other hand, digestion of the ligated product also confirmed the presence of insert and vector as shown in line 2, bands at 1Kb and 3Kb respectively. Therefore, a suitable construct was produced.

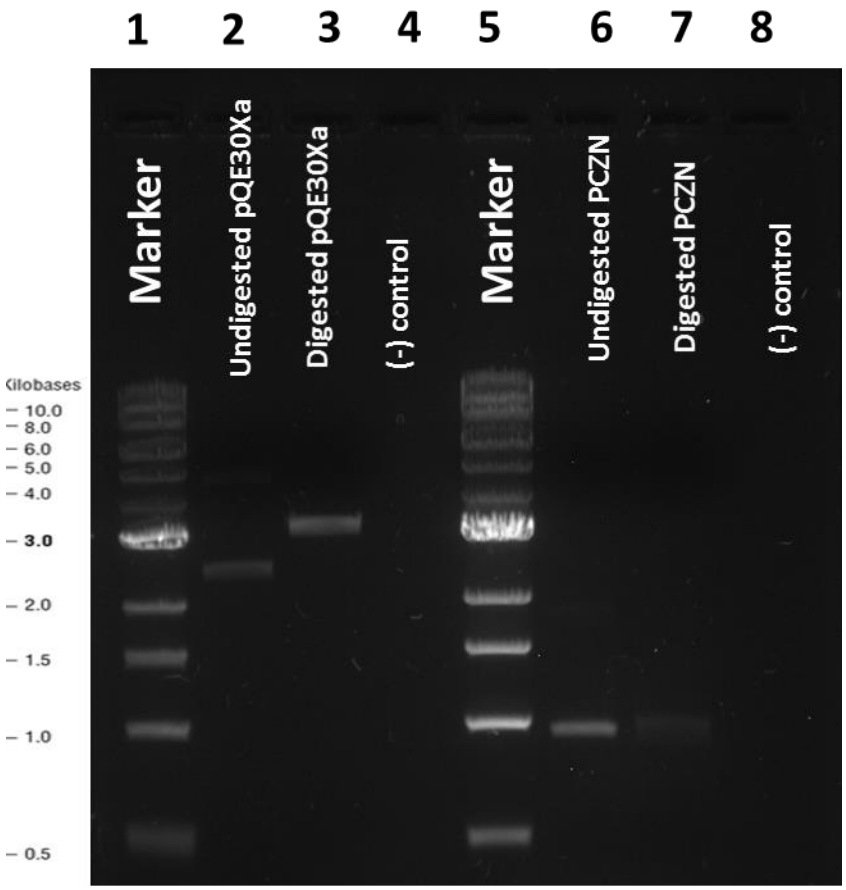

Figure 6. Digestion of PCZN encoding for DNA and pQE30-Xa vector. Line 1 and 5: DNA markers, line 2: undigested vector, line 3: digested vector, line 4 and 8: negative controls (no DNA template), line 6: undigested insert, line: 7 digested insert.

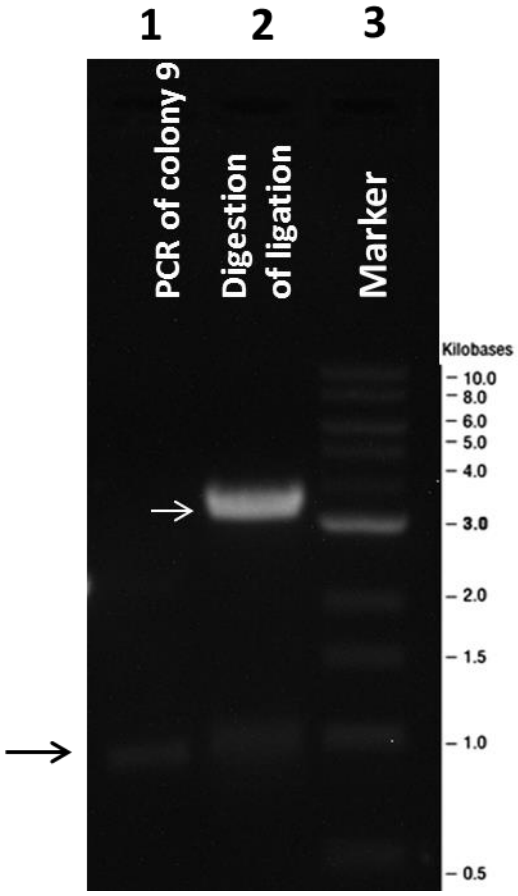

Figure 7. Digestion reaction of PCZN-WT-pQE30Xa. 1% agarose gel showing the HindIII and StuI digested products of ligated PCN-WT-pQE30Xa. Line 1: band showing the band of insert PCZN-WT obtained from PCR of colony 9, line 2: product of digestion reaction with HindIII and StuI, bands corresponding to the vector pQE30Xa =3.5Kb and PCZN-WT= 0.957Kb (pointing arrow). Line 3: DNA marker.

GGATCCTGCCTGGTGCCGGCGGCGACCGCGAGCCTGCATGCGGAGGAAACCCTGGCGAGCCAATTTGC
GGAGTTCAAACAGAAACACGGTCGTGTGTACGAGAGCGCGGCGGAGGAAGCGTTCCGTCTGAGCGTTT
TTCGTGAAAACCTGTTCCTGGCGCGTCTGCATGCGGCGGCGAACCCGCACGCGACCTTTGGCGTGACCCC
GTTCAGCGACCTGACCCGTGAGGAATTTCGTAGCCGTTATCACAACGGTGCGGCGCACTTTGCGGCGGC
GCAGGAGCGTGCGCGTGTGCCGGTTAACGTGGAAGTGGTTGGT*GCGCCGGCGGCGGTTGATTGGCGTG*
CGCGTGGTGCGGTGACGCGGTTAAGGATCAGGGCCAATGCGGTAGCTGCTGGGCGTTTAGCGCGATCGG
CAACGTGGAGTGCCAGTGGTTCCTGGCGGGTCACCCGCTGACCAACCTGAGCGAACAAATGCTGGTTAGC
TGCGACAAAACCGATAGCGGTTGCGGTGGCGGTCTGATGAACAACGCGTTTGAGTGGATCGTGCAAGAA
AACAACGGCGCGGTTTACACCGAGGGTAGCTACCCGTATGCGAGCGGCGAAGGTATTAGCCCGCCGTGCA
CCACCAGCGGTCACACCGTGGGTGCGACCATCACCGGCCACGTTGAGCTGCCGCAGGACGAAGCGCAAAT
TGCGGCGTGGCTGGCGGTGAACGGTCCGGTTGCGGTGGCGGTTGATGCGAGCAGCTGGATGACCTACAC
CGGCGGTGTGATGACCAGCTGCGTTAGCGAGCAGCTGGACCATGGTGTGCTGCTGGTTGGTTACAACGAT
AGCGCGGCGGTGCCGTATTGGGTTATCAAGAACAGCTGGACCACCCAATGGGGCGAAGATGGTTATATTC
GTATTGCGAAGGGCAGCAATCAACTGGTGAAAGAAGAGGCGAGCAGCGCGGTGGTGGTTAAGCTT

Figure 8. PCZN-WT sequence from plasmid donated by Dr Lima. Bold nucleotides correspond to the Pro domain and cursive nucleotides correspond to Catalytic domain.

Various attempts to express protein from the newly construct were all failures (data not shown). Due to this issue, we decided to continue working with a plasmid that was kindly donated by Dr. Lima (PCZN-WT-pQE30). PCZN-WT-pQE30 plasmid was successfully sequenced using primers for pQE vectors previously described in the methodology section. The resulted PCZN-WT sequence is shown in Figure 8. PCZN-WT-pQE30 construct (Figure 9B) successfully expressed a soluble recombinant protein (Figure 12, line 1) in contrary to the PCZN-WT- pQE30-Xa construct containing the signal peptide that produce insoluble protein. Therefore, we decided to clone all mutants of PCZN (E25V, F32A, F35A, E49V, F56A, D82V, E86V, E87V, and C141A) into pQE30 as well (Figure 9C). DNA sequence of all pQE30 constructs were confirmed using pQE primers mentioned above and the resulted protein sequence of PCZN-WT is shown in Figure 10**.** Additionally, we want to mention that the mutated residues mentioned above (except for C141A), were selected based on data obtained from molecular dynamics studies of homology modeled PCZN-WT protein. These data is found in chapter 2.

.

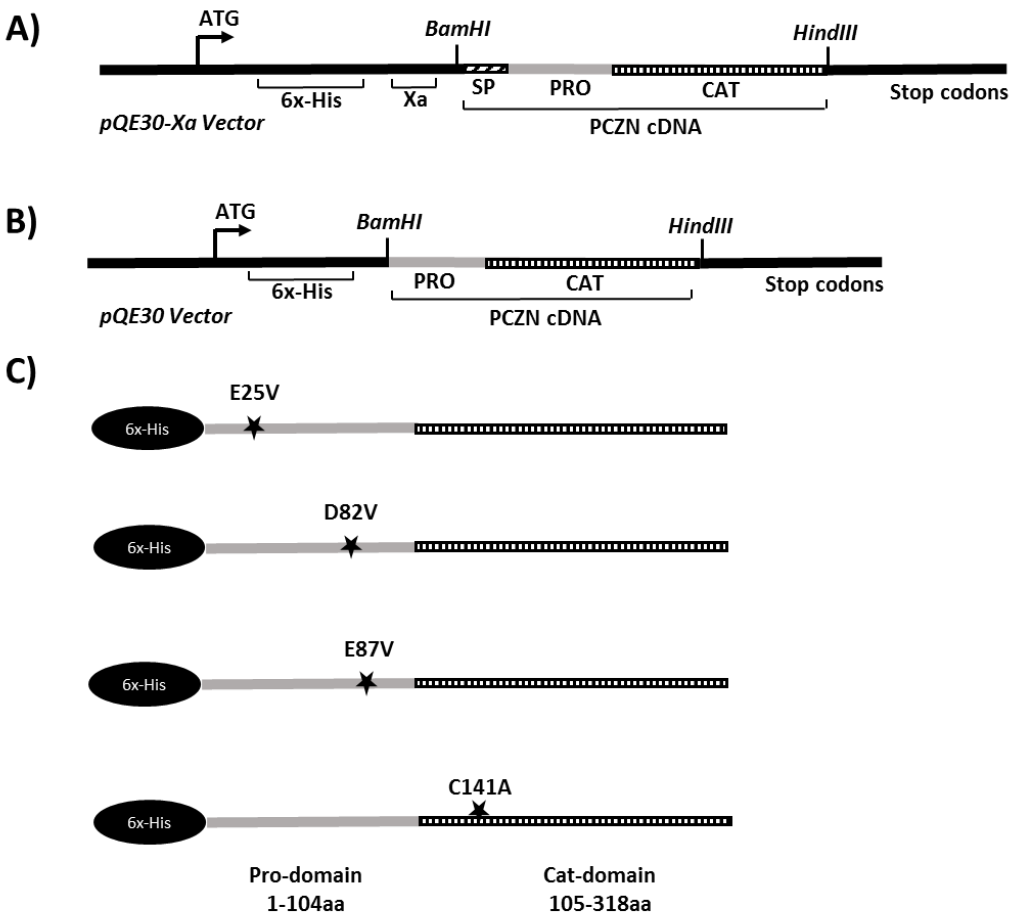

Figure 9. Plasmid organization. (A) pQE30-Xa-PCZN vector map, (B) pQE30-PCZN vector map (C) Mutant constructs. Star shows a single residue mutation.

MRGS*HHHHHH*GS**CLVPAATASLHAEETLASQFAEFKQKHGRVYESAAEEAFRLSVFRENLFLARLHAAANPHATFG VTPFSDLTREEFRSRYHNGAAHFAAAQERARVPVNVEVVG***APAAVDWRARGAVTAVKDQGQCGS**C***WAFSAIGNV ECQWFLAGHPLTNLSEQMLVSCDKTDSGCGGGLMNNAFEWIVQENNGAVYTEGSYPYASGEGISPPCTTSGHTVGA TITGHVELPQDEAQIAAWLAVNGPVAVAVDASSWMTYTGGVMTSCVSEQLD**H**GVLLVGYNDSAAVPYWVIK**N**SWT TQWGEDGYIRIAKGSNQLVKEEASSAVVV*

Figure 10. Aminoacid sequence of PCZN-WT including 6-histidine tag at the N-terminus. Pro-segment composed of 104 residues (underlined); catalytic domain composed of 214 residues; the catalytic triad of this enzyme is composed of cysteine 25, Histidine 169, and Asparagine 182 (bold and cursive C, H, and N respectively; numbered starting from first residue of catalytic segment, "papain numbering"). Recombinant protein is composed of 330 residues.

1.4.2 Protein expression of PCZN-WT and mutants (E25V, D82V, E87V, C141A)

As mentioned earlier, previous attempts at generating soluble recombinant protein containing signal peptide was not successful. Figure 11 shows the result of protein expression for PCZN-WT-pQE30Xa construct containing the signal peptide. Supernatant of cell lysate (line 2) illustrates the lack of recombinant protein, whereas the lysed *E.coli* cells contains the band at around 36kDa corresponding to the recombinant PCZN protein. We believe that the insolubility of this protein is the result of the signal peptide (MSGWARALLLAAVLVVMA) present in PCZN that produced inclusions bodies causing the protein to remain insoluble therefore it was removed. This may be due to the many hydrophobic residues characteristic of all signal peptides which is the portion responsible for protein cell trafficking[55]. On the other hand, the PCZN-WT-pQE30 construct (Figure 9B) successfully expressed a soluble recombinant protein (Figure 12, line 1), and because we used the pQE30 expression vector instead of the pQE30-Xa the His-tag was kept in all recombinant proteins. Additionally, protein expression using the pQE30 vector expression system was successfully accomplished due to a combination of a previously described protocol[54] with our protocol (un-published data). Moreover, the use of plasmid pREP4 present in M15

cells was key to prevent leakage of expression. Figure 12 shows the bands at around 35 to 38kDa corresponding to the his-tagged recombinant PCZN-WT and its mutants.

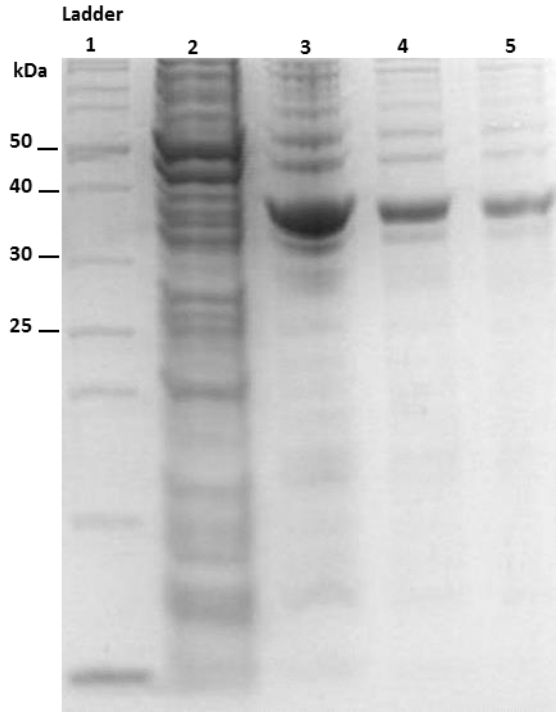

Figure 11. Recombinant PCZN with signal peptide. 12 % SDS PAGE gel showing the cell lysates of expressed PCZN. Line 1: MW marker, line 2: supernatant of cell lysate, and lines 3-5: cell pellet re-suspended in lysis buffer. Lines 3-5 indicate bands of approximately 37kDa corresponding to the 6xHis-PCZN with signal peptide.

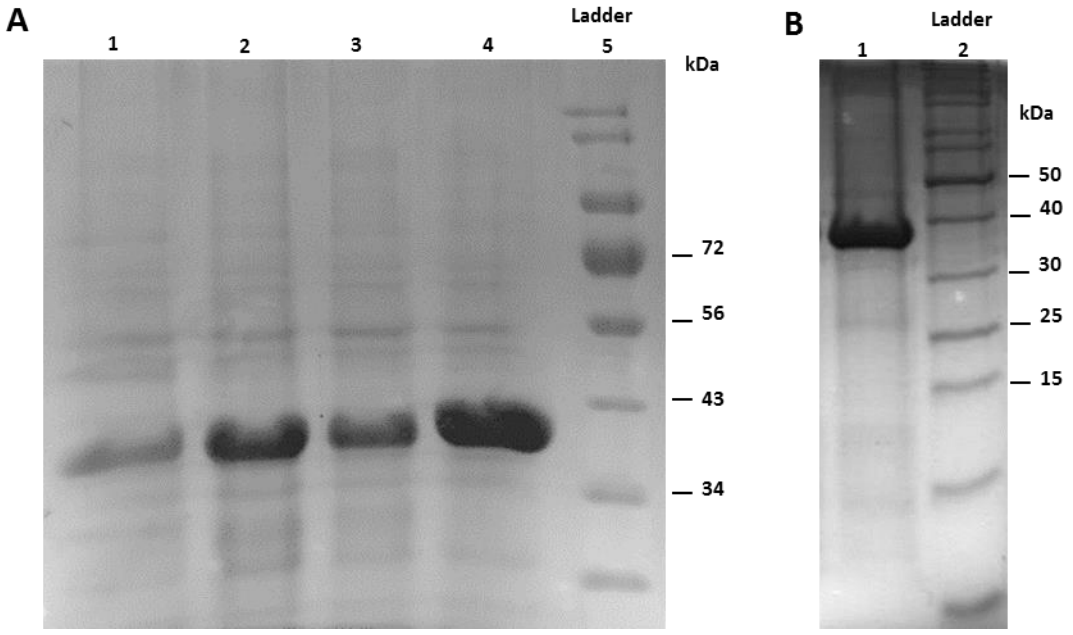

Figure 12. Recombinant PCZN WT and Mutants. 12 % SDS PAGE gel showing refolded and concentrated WT and mutants of PCZN. Figure (A) line 1: PCZN-E25V, line 2: PCZN-E87V, line 3: PCZN-D82V, line 4: PCZN-WT, and line 5: MW marker. Figure (B) line 1: PCZN-C141A, and line 2: MW marker.

1.4.3 Protein purification of PCZN and mutants

Based on conditions described earlier, we generated 1L of IPTG induced culture bacteria. The *E. coli* M15 cells were harvested, and bacterial pellet was re-suspended in lysis buffer. The soluble protein was collected by sonication followed by centrifugation. Purification of recombinant proteins was carried out using Ni-NTA affinity chromatography using a gradient concentration of imidazole as eluting agent. Adjustments on imidazole concentration had to be done to recover a higher amount of pure protein. In Figure 13A, lines 2-4, we can observe a thick band at around 40kDa corresponding to PCZN-C141A protein along with undesired protein. These fractions were eluted with 50mM imidazole elution buffer followed by elution with 100mM imidazole elution buffer shown in lines 5-7. Then purer fractions were observed as we increased the imidazole concertation to 200mM. In order to avoid early elution of his-tagged

protein we utilized elution buffers with lower imidazole concentrations. After 2 cycles of washes, 3 fractions with 10mM were collected, as well as with 30mM (Figure 13E) to finally elute with 200mM (5 fractions of 1mL and 2 fractions of 5mL). The rest of the recombinant mutant proteins and WT were purified following this elution process as shown in Figure 13. The final yield of protein for 1L of cell culture was around 1.6mg/mL. Unfortunately more than half of the protein precipitated out during the dialysis processes resulting in around 0.7mg/mL of refolded protein.

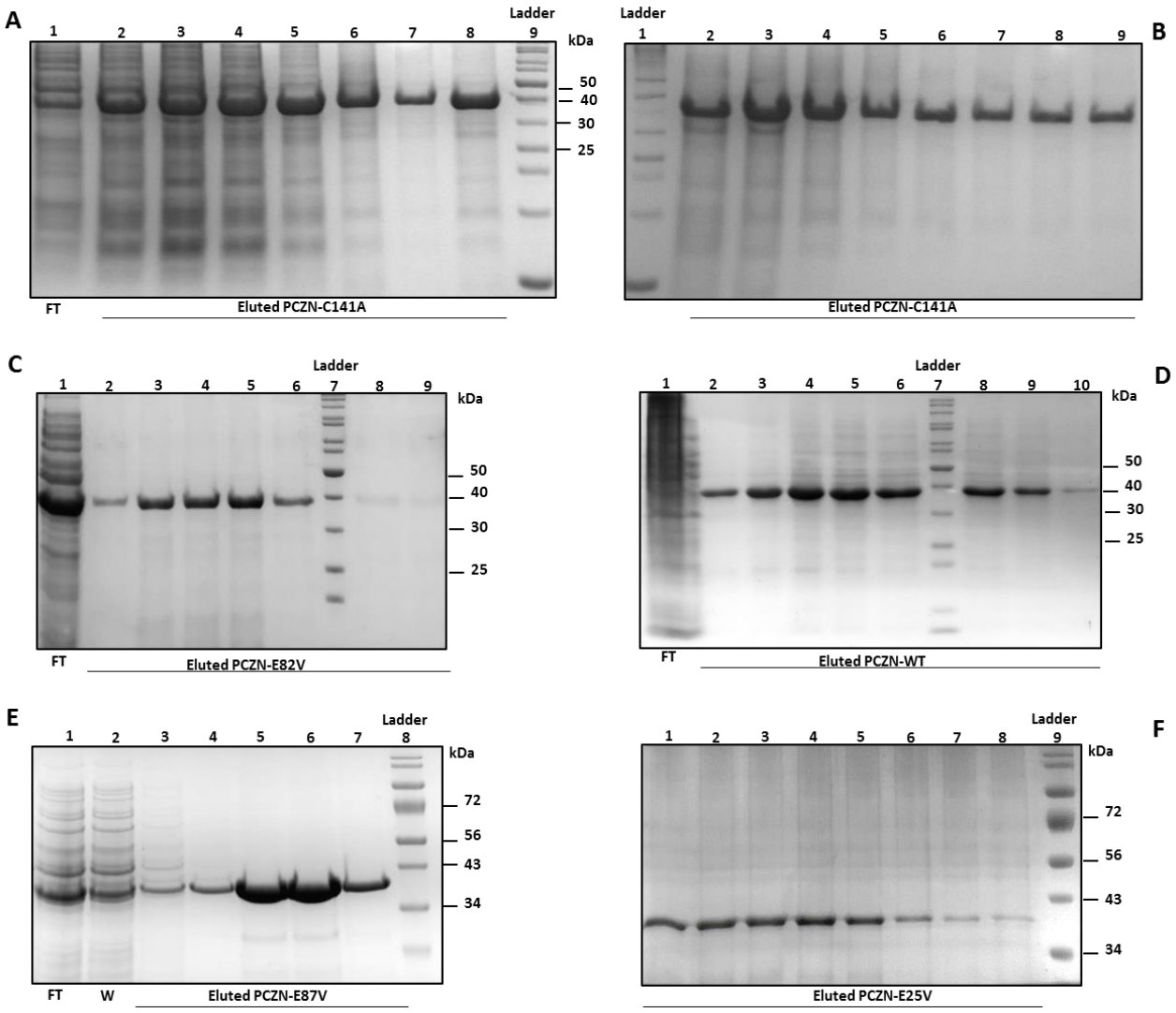

Figure 13. Purification of recombinant PCZN WT and Mutants. Figure (A) PCZN-C141A protein purification. Line 1: FT (cell lysate proteins from flow through fraction after His-tag recombinant protein has been bounded to resin), lines 2-4: mix of PCZN-C141A protein with undesired protein (eluted with 50mM imidazole elution buffer), lines 5-7: mix of PCZN-C141A protein with undesired protein (eluted with 100mM imidazole elution buffer), line 8: shows band at approximately 36kDa corresponding to the 6xHis-PCZN-C141A (fraction eluted with 200mM imidazole), and line 9: MW marker (EZ-Run Rec Protein Ladder, Fisher BioReagents). Figure (B) line 1: MW marker, lines 2-3: mix of PCZN-C141A protein with undesired protein (eluted with 200mM imidazole elution buffer), lines 4-9: purified PCZN-C141A protein fractions

(eluted with 200mM imidazole elution buffer). Figure (C) PCZN-E82V purification. Line 1: FT (showing unbound PCZN-E82V protein), lines 2-6 and 8-9 purified PCZN-E82V protein fractions (eluted with 200mM imidazole elution buffer), line 7 MW marker. Figure (D) PCZN-WT purification. Line 1: FT, lines 2-6 and 8-9 purified PCZN-WT protein fractions (eluted with 200mM imidazole elution buffer), line 7 MW marker. Figure (E): PCZN-E87V purification. Line 1: FT, line 2: washed fraction, line 3: fraction eluted with 10mM imidazole buffer, line 4: fraction eluted with 30mM imidazole buffer, lines 5-7: fractions eluted with 200mM imidazole buffer, line 8: MW marker. Figure (F) PCZN-E25V purification. Line 1-8: Purified fractions eluted with 200mM imidazole buffer, line 9: MW marker (EZ-Run Pre-stained Rec Protein Ladder, Fisher BioReagents).

1.5 Conclusions

Cloning and expression of mutants of recombinant wild type of PCZN was successfully performed using existing methodologies in combination with our protein expression and purification process modifications. It is important to mention that although plasmid constructs of PCZN-mutants E25V, F32A, F35A, E49V, F56A, D82E, E86V, and E87V were generated, only PCZN-E25V, PCZN-D82, PCZN-E87V were expressed. The combination of pQE30 expression vector and *E. coli* M15 competent cells aided to effectively express PCZN-WT and the aforementioned mutant proteins. Additionally, the modification on the protein purification process helped in the recovery of high amounts of pure protein. Methods described herein will be of benefit for future studies on this protein.

CHAPTER 2

DISCOVERING CRITICAL AMINO ACID-AMINO ACID INTERACTIONS WITHIN THE PRO DOMAIN OF THE ZYMOGEN FORM OF CRUZAIN

2.1 Background

Recent drugs design that exploit the binding site of cruzain and the known mechanism of cysteine proteases has allowed for the development of one small molecule inhibitor that has made it to clinical trials [56]. Advances in a class of vinyl-sulfone inhibitors are encouraging; however, as most potential therapeutics fail in clinical trials and both disease progression and resistance call for combination therapy with several drugs, the identification of additional classes of inhibitory molecules is essential. Recent studies have identified small-molecule cruzain inhibitors. Further optimization of these chemical scaffolds could lead to the development of novel drugs useful in the treatment of Chagas' disease. Although progress has been made in the treatment of the disease, especially given the recent success of K11777, multiple cruzain inhibitors are needed given the difficulty of obtaining FDA approval and ever-progressing drug resistance [57]. Despite that several crystal structures of complexes of mature cruzain with inhibitors have been determined and are available [23, 58], none of these inhibitors have passed the clinical trials. Therefore, studies of the interface between catalytic and pro domains, and within domains of pro-cruzain (PCZN) could lead to smaller and yet unexploited molecule inhibitors in the future. As suggested by Shinde *et al.,* 1997 [59] it is possible that point mutations within the pro-domain may result in altered folding, and subsequent changes in substrate specificity of the enzyme may lead to malfunction of the protein [59]. Other studies have also demonstrated the ability of PCZN's pro-domain to inhibit other cysteine and serine proteases [60].

2.1.1 Homology modeling

Homology modeling is a newly discovered computational method to determine the 3D structure of proteins. This technique utilizes available high-resolution protein structures to produce a 3D model of a protein of related, but unknown, structure. The process of homology modeling a protein of unknown 3D structure is divided into four steps: template identification, alignment, model building and refinement, and validation [61, 62]. Here we present a three-dimensional homology-modeled structure of the precursor form of cruzain, determined by homology modeling structure prediction.

2.2 Specific Aims

Previous studies have focused in the developments of cruzain function, specificity, and structure, specifically aimed for the inhibition of active cruzain. Herein, we propose to discover weakening points within the zymogen form of cruzain (pro-cruzain) to act via prophylaxis by inhibiting the maturation of the enzyme. Therefore, we hypothesize that critical residues within this weakening sites will play an important role in the structure stability of pro-cruzain.

Specific Aim One: To generate a homology-modeled structure of the zymogenic cruzain by the utilization of bioinformatics tools.

Specific Aim Two: To discover crucial residue-residue interactions at the not yet exploited pro-catalytic interface of zymogen form of cruzain and to create in-silico mutagenesis and assess the impact of those mutations by molecular dynamic analysis.

2.3 Materials and Methods

2.3.1 Protein sequence identity and alignment

BLAST (Basic Local Alignment Search Tool) [63] analysis of the zymogenic cruzain amino acid sequence was performed. The resulted sequences were selected for further alignment using the following criteria: total score >350, query coverage >90%, and maximum identity >50%. Then, proteases with the highest scores were aligned with cruzain for comparison using Clustalw Omega [64].

2.3.2 Homology modeling:

To get insight into the significance of a single amino acid substitution on protein function, a good 3D structure of the protein is very important. A homology-model of PCZN was developed using the I-Tasser [65] server. One of the templates used by I-TASSER was another cysteine protease proenzyme (procathepsin K, PDB code 7PCK_A) which has homology identity of 37% with PCZN and query coverage of 97%. The predicted structure of PCZN was further superposed with another well-known pro-cathepsin L (PDB ID:1CJL) and the crystal structure of mature cruzain (PDB ID:1ME4) for comparisons.

2.3.3 Molecular Dynamics simulations

In order to obtain high quality structures of wild type (WT) cruzain and selected mutants, a high-throughput MD simulations were executed on the resulted model from I-Tasser (model1). The system was setup by Amber Tools 15 suite [66] and was performed using conventional molecular dynamics (cMD) and accelerated molecular dynamics (aMD). The cMD and aMD were carried out using the Amber 14 program and the Amber14SB force field[66]. Initially, the systems were energy-minimized in two steps. First, only the water molecules and ions were minimized in 4000 steps while keeping the protein structure restricted by weak harmonic constrains of 2 kcal mol^{-1} $Å^{-2}$. Second, a 4000 steps minimization with the conjugate gradient method on the whole system was performed. Furthermore, the simulated systems were

gradually heated from 0 to 310 K for 50 ps (canonical ensemble, NVT) and equilibrated for 3 ns (isobaric-isothermal ensemble, NPT). The production runs were performed at 310 K in a NPT ensemble. Temperature regulation was done using a by Langevin thermostat with a collision frequency of 2 ps^{-1}. The time step of the simulations was 2 fs with a nonbonded cutoff of 8Å using the SHAKE algorithm [67] and the particle-mesh Ewald method [68].The production runs were performed at 310 K in a NPT ensemble. The output from the performed and above described cMD simulations were used as an input for executed aMD simulations. Temperature regulation was done using a Langevin thermostat with a collision frequency of 2 ps^{-1}. The time step of the simulations was 2 fs with a nonbonded cutoff of 8 Å using the SHAKE algorithm and the particle-mesh Ewald method. Two 250ns long aMD simulations on Model 1 was performed. The aMD simulations provided the possibility to sample the conformational space much better and to detect the local energy minima that remain hidden in the cMD calculations [69]. A total of 17 simulations with total length of 4.3μs were executed. Five simulations of the PCZN structure (model1) were executed, of which 4 lasted 250ns using (cMD) and one lasted 250ns using accelerated MD (aMD) giving a total length of 1.3 μs. The last 12 simulations were done on the output model from first set of simulations using cMD and lasting 250 ns each for a total length of 3.0 μs. MD simulations were done in collaboration with Dr. Fratev from the Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences.

2.3.4 Protein Residue-Residue Interaction analysis

Ionic and hydrophobic interactions were calculated by Protein Interaction Calculator (PIC) [70]; 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. Homology-modeled structure, inter-protein interactions, and residue substitutions were visualized using PyMol software [71].

2.3.5 Stability studies of PCZN mutants

The stability of mutants of PCZN was accessed using servers FoldX3, FoldX4, ENCoM, and Com with the help of Dr. Fratev.

2.4 Results and Discussion.

2.4.1 Protein sequence identity and multiple sequence alignments

PCZN domains possess significant similarities in the sequence homology in comparison to several other Papain-like cysteine proteases when compared separately decipher by running a BLAST analysis (data not shown). The pro domain is somewhat variable in homogeneity (85-50%) to several other Papain-like cysteine proteases, such as cysteine peptidases from *T. rangeli* and *T. brucei*, whereas the catalytic domain shares lower sequence identity (72-65%). However, although some pro parts may differ in length and sequence identity their overall fold appears to be similar, as it has been shown by the resolved X-ray structures [72]. Nonetheless, when PCZN sequence is compared as a pro-protein complex (pro-cat) it shows lower sequence identity as shown in Table 1.

Table 1. Results of BLAST analysis of pro-cruzain amino acid sequence (pro and catalytic domains).

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

Then, the pro domains of all the cysteine proteases that resulted to share a high sequence identity to PCZN were aligned as shown in Figure 14. Residues E25, F32, F35, E49, F56, D82, E86, and E87 are highly conserved within the pro domain sequence of cysteine proteases from homologous parasites; allowing us to point out several residues as possible candidates for single amino acid substitution. Interestingly, the importance of these residues were confirmed by the molecular dynamics studies that will be shown later in this chapter.

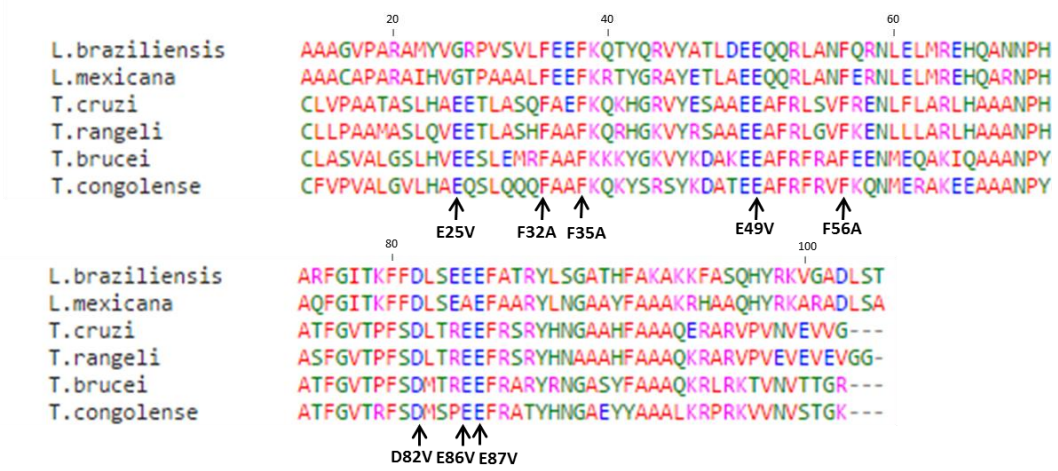

Figure 14. Pro-region sequence alignment of PCZN-WT and cysteine proteases. Sequence alignment pro-domains of *T. cruzi*, *T. rangeli, T. brucei, T. congolense, L. braziliensis,* and *L. mexicana.* Arrows indicate the residues selected for substitution: E25, F32, F35, E49, F56, D82, E86, and E87.

A predicted secondary structure of PCZN-WT was performed by inputting the amino acid sequence into phyre2 server (Figure 15). The secondary structure of PCZN was predicted to have more percentage of alpha-helical content (30%), followed by 18% of beta strands and 17% disordered, while the rest remained as coil conformation. From this predicted secondary structure we can determine the localization of the selected residues within the different motifs in the PCZN structure. Residues F32, and F35; E49, and F56; E86, and E87 belong to a stable helix spanning E26 to H39 (α1p); A46 to A67 (α2p); R85 to Y92 (α3p) respectively. Nevertheless, residues E25, E82 were not part of a motif in the secondary structure. The structural impact of these residues will be demonstrated later in this chapter. The predicted secondary structure of the pro domain region is in concordance with the secondary structure represented in a sequence alignment of Cathepsins pro-peptides, including cruzain and bruzain [60], except for the first alpha helix shown in our prediction. This disagreement may be due to the 6x-his tag at the N-terminus kept in our sequence.

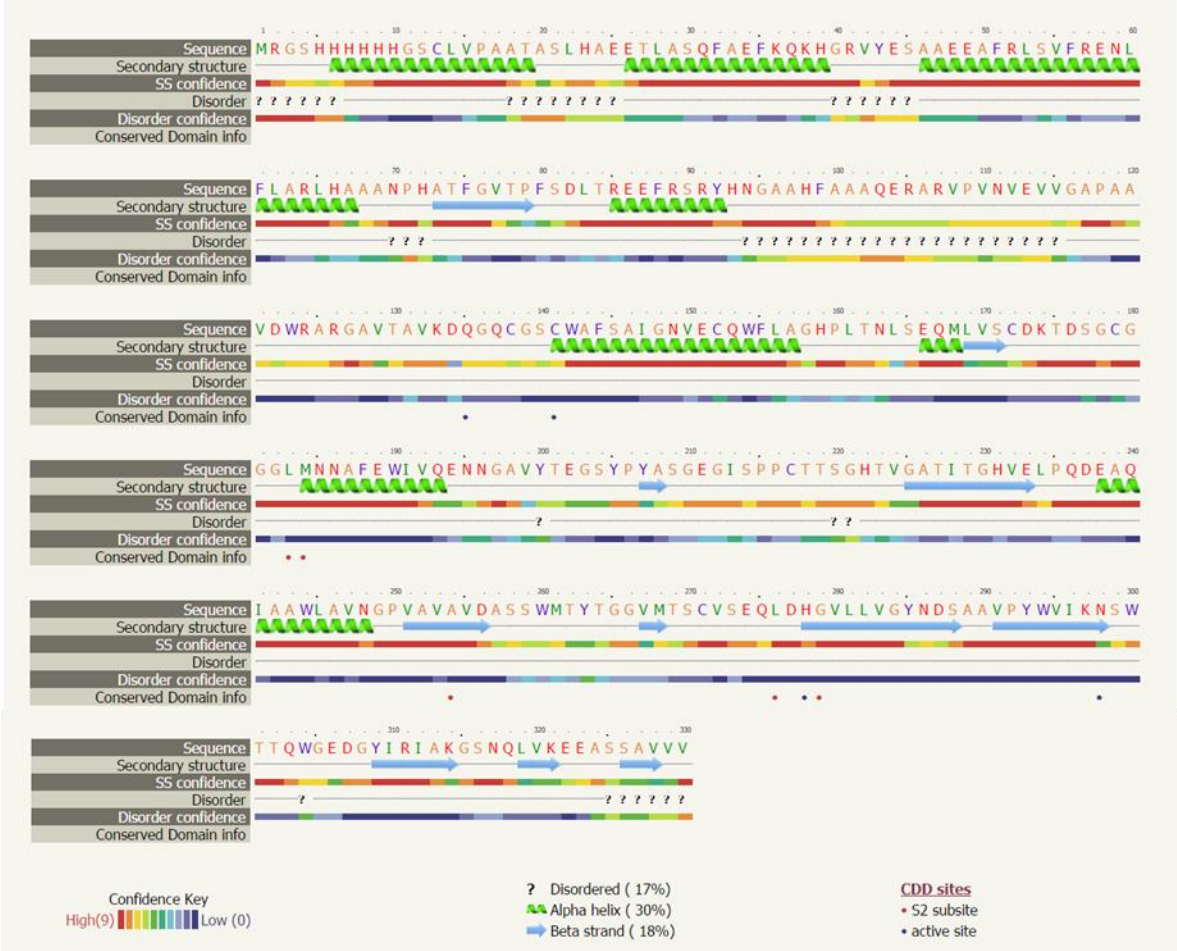

Figure 15. Predicted secondary structure of PCZN-WT. Secondary structure of PCZN-WT reproduced based on amino acid sequence. Regions not covered are shown along with the predicted secondary structure. Overall, α-helical content is the predominant secondary structure for pro-cruzain.

2.4.2 Homology modeling

A 3D homology-modeled structure of the zymogen form of cruzain was successfully created using I-Tasser software as mentioned above. The I-Tasser model (model 1) was based only on the available X-ray data from the pro-catalytic complexes (zymogen protein). For instance, the best scored model was constructed using the as a template which was another cysteine protease proenzyme (Procathepsin L,

PDB ID entry 1CS8) with homology identity of 37% with PCZN and query coverage of 90% (Figure 16). The confidence score (C-score) was 0.75, which estimates quality of predicted models by I-TASSER and it is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5,2], where a C-score of higher value signifies a model with a high confidence and vice-versa. Our C-score of 0.75 resulted the highest value compared with the other predicted models. On the other hand, estimated TM-score and RMSD (root-mean-square deviation of atomic positions) are standards to measure structural similarity between, in this case, the native structure based on the C-score and the predicted model [73, 74]. For instance, values of TM >0.5 and a small RMSD in a predicted model signifies a high quality homology-modeled structure. Our values for TM and RMSD values were 0.81±0.09 and 4.8 ±3.1Å, respectively, for which we can state that our predicted model is of high quality.

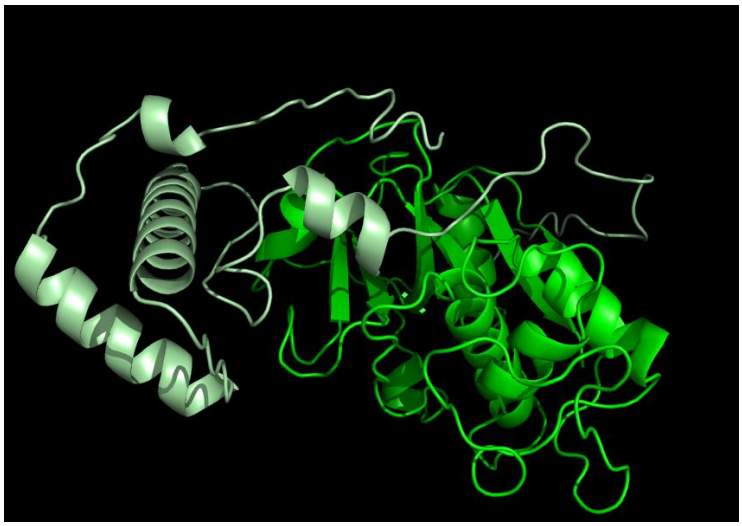

Figure 16. PCZN structure homology-modeled by I-Tasser (Model1). Light green represents the pro domain and bright green represents the catalytic domain.

2.4.3 Molecular dynamics simulations

The final structure of the Pro-cruzain WT, model2, (Figure 17) and mutants were described by a total of 17 simulations with total length of 4.3μs. We contrasted our results to both the pro-cathepsin L (PDB ID: 1CJL) and the catalytic part of Cruzain (pdb id: 1ME4) X-ray structures (Figure 18). Such a comparison was useful due to the fact that Pro segment of cathepsin L, protease that belong to the same papain family, is one of best described to the moment whereas structural and dynamical information about cruzain is available only for the catalytic part.

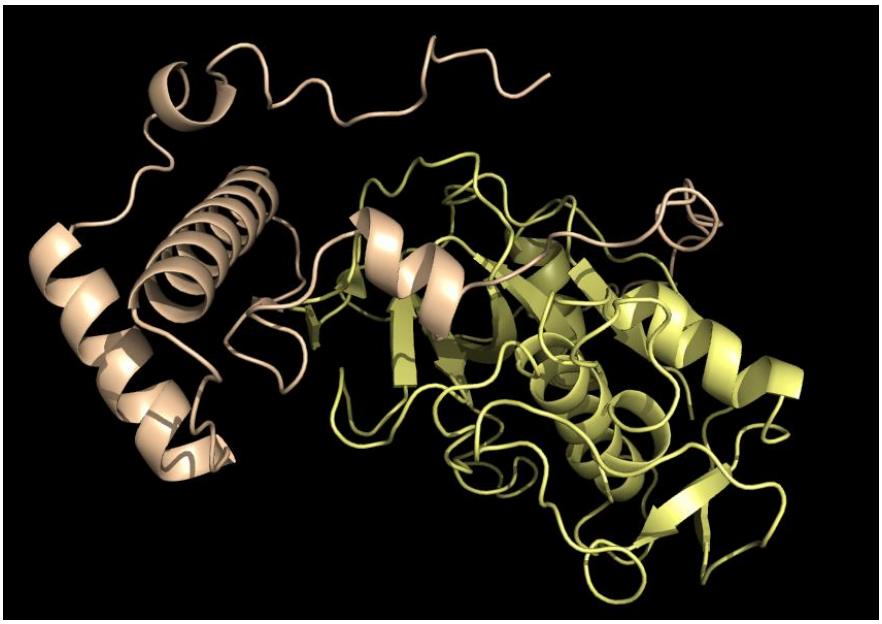

Figure 17. High resolution homology-modeled PCZN (Model2) showing the two domains. Catalytic domain colored yellow and pro domain colored light brown.

Figure 18. High resolution structure of PCZN (Model2). Pro-cruzain (light brown colored) juxtaposed to the structures of Cruzain (PDB ID: **1ME4**; green colored) and pro-cathepsin L (PDB ID:1CJL; blue colored).

The MD data revealed RMSD of 1.1Å observed between the catalytic parts of PCZN and pro-cathepsin L whereas the RMSD of only 0.7Å was registered between the MD obtained catalytic cruzain average structure (model2) and those resolved by the X-ray technique. Considering that the RMSD between the X-ray structures of pro-cathepsin L and cruzain is 1.1Å one can conclude that the MD results precisely reproduced the difference between catalytic parts of these papaine proteases. In addition, the catalytic triad constituted by Cys 25, His 169, and Asn 182 was properly aligned by strong hydrogen bond network and remained very stable (>95% live time) during all performed MD simulations, including those of the selected mutants (Figure 19). This is opposite to the unstable H-bond observed by MD simulations of Procathepsins L and O [75]. Considering that the same His and Asn residues had an identical conformation in the 1CJL X-ray structure it seems that the last results are presumably due to the homology

modeling and inadequate sampling (less than 10ns) rather than some unusual properties of the Procathepsins L and O.

Aligning the imidazole and thiol groups for hydrogen bonding assist catalysis making the enthalpy of activation less negative thus the imidazole acts as a general base [76]. The role of Asn182 is to orient the His169 side chain in the optimum positions for various steps of the catalytic mechanism. In the resting state of the enzyme, the His side chain would be coplanar to the Cys25 residue while during acylation, the protonated imidazole ring would rotate to act as a proton donor to the nitrogen atom of the leaving group of the substrate. A comparison of pro domains of mutants with WT PCZN revealed a slight difference among the pro-segment binding loop (PBL) regions. In Figure 20 we can observe the loop Tyr92 to Glu103 of each mutant are positioned differently in comparison to the WT where the mutants are closer to the catalytic triad cleft. This might suggest that mutants could possible block the catalytic cleft stronger than in the WT enzyme.

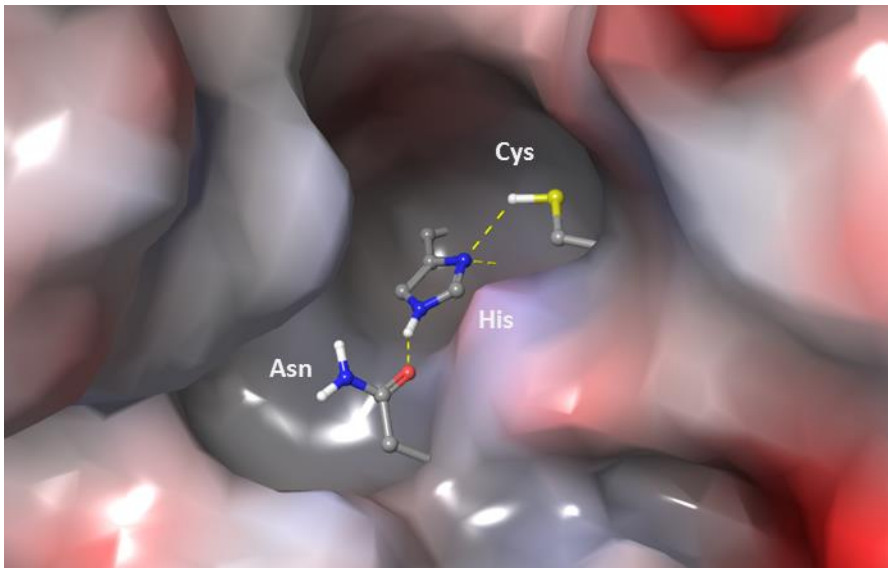

Figure 19. Catalytic triad of homology modeled PCZN. Triad is consisted of a cysteine at position 25, a histidine 169, and an Asparagine 182 (count starting from the first residue in the catalytic domain).

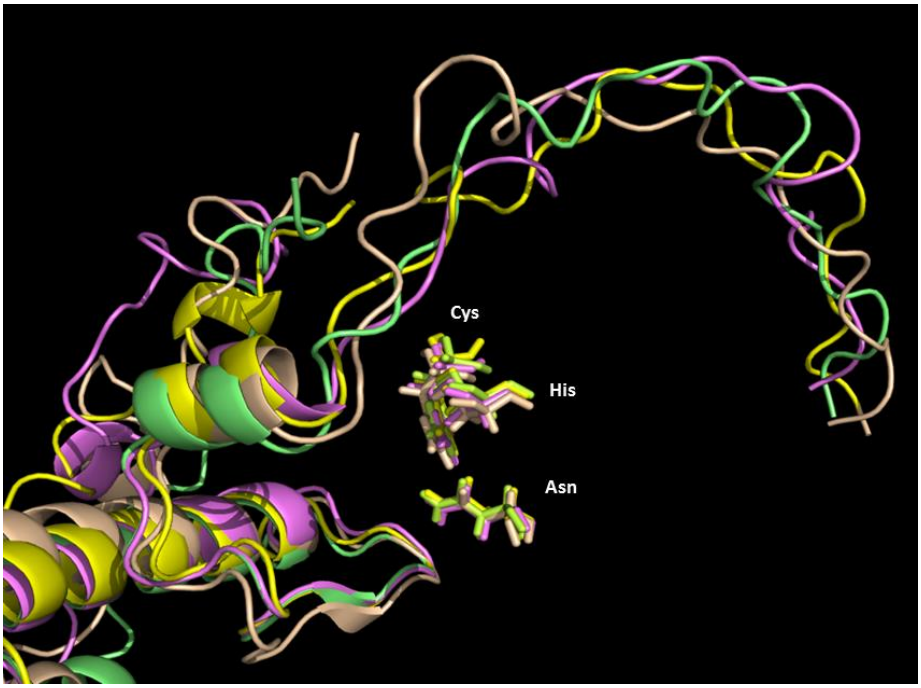

Figure 20. A visual comparison of pro domains with the catalytic triad of mutants and WT. Wild type colored in beige; E25V colored in green; D82V colored in yellow; E87V colored in pink.

Although the Pro segments of procathepsin L and Procruzain exhibited larger difference characterized by the RMSD value of 2.0Å, they also shared significant structural similarity. The difference of the Pro fragments between these papains is mainly due to the insertion of 4 additional residues in PCZN sequence (Figure 21). In addition, the central two helices of the Pro part, α1p and α2p, rotates by about 15 degree compared to pro-cathepsin L (Figure 21). This effect was reproduced in all performed simulations in our study. The Procruzain, similarly to procathepsin L, contains two main type core interactions which are believed to stabilize the Pro domain. The hydrophobic contacts of the three phenylalanines (Phe32, Phe35 and Phe55) stabilize the α1p and α2p helices (Figure 22)**.** Another important key of the stabilization of the pro domain was found in the strong hydrogen-bond network composed by hydrogen bond interactions among Arg52, Asp82, Tyr43 and Arg41 (Figure 23). They are

incorporated into a common H-bond network that presumably keeps the integrity of the Pro part and in particular interactions between helices α2p and helices α3p. In fact, this network is also significant for the α3p conformation and it positioning into the binding cleft. A significant reorganization of the catalytic site cleft occurs after the pro part binding, which provides the structural basis of the strong inhibition observed. The Trp177 rotates by about 70 degree in an open position, which allows the α3p helix binding to the catalytic grove **(**Figure 21**)**. In such a way the Tyr92 of the Pro segment buries deep into the cleft, makes a strong H-bond with the Asn175, π-π stacking with the catalytic His159 and also establishes several hydrophobic contacts within the site. In turn the Asn175 forms a hydrogen bond with His159 too. Despite above mentioned α3p helix stabilization factors the Phe80 and Phe88 were also incorporated into these pro-catalytic hydrophobic interactions providing both much stronger inhibitions by α3p segment and full obstruction of the catalytic site.

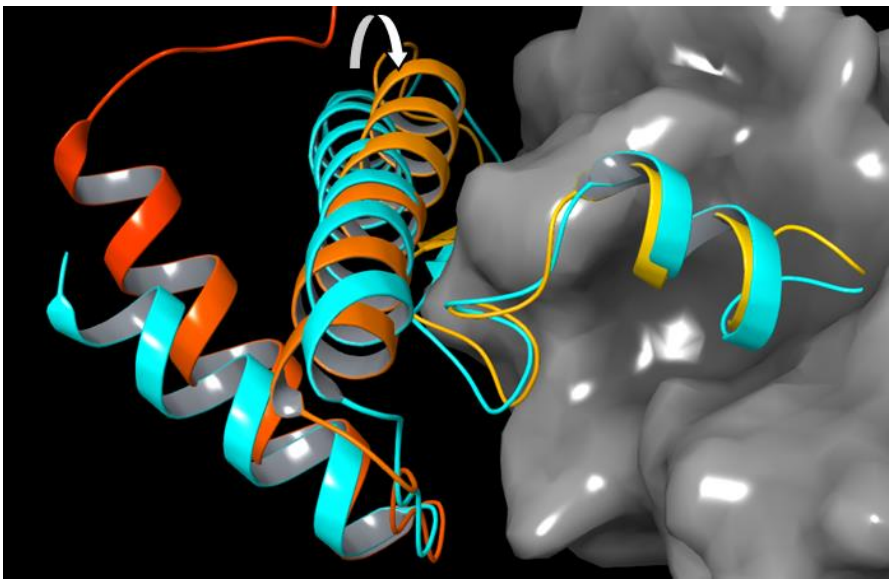

Figure 21. **The pro domain of PCZN structure (multicolor) compared to Pro-Cathesin L (pdb id:1cjl; blue colored).**

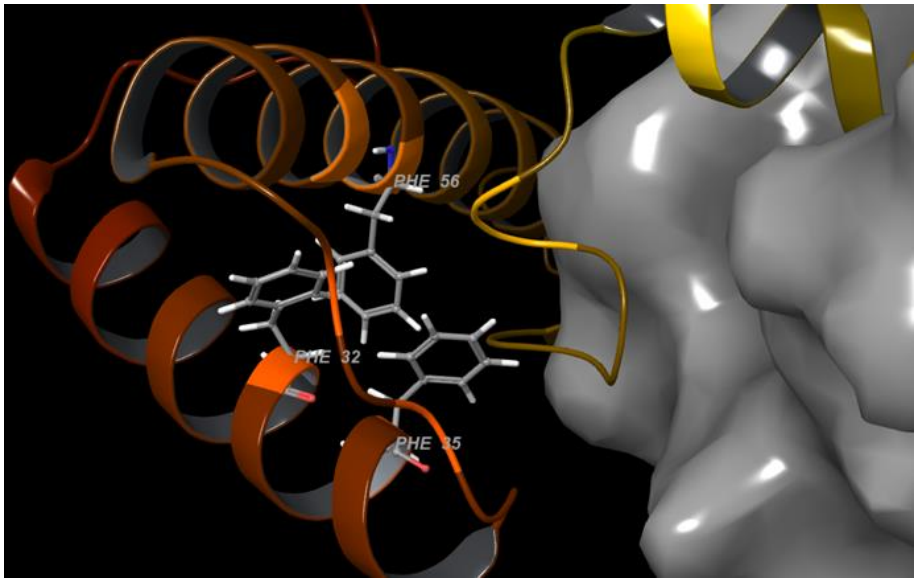

Figure 22. Hydrophobic core within the alpha helices of pro domain of PCZN.

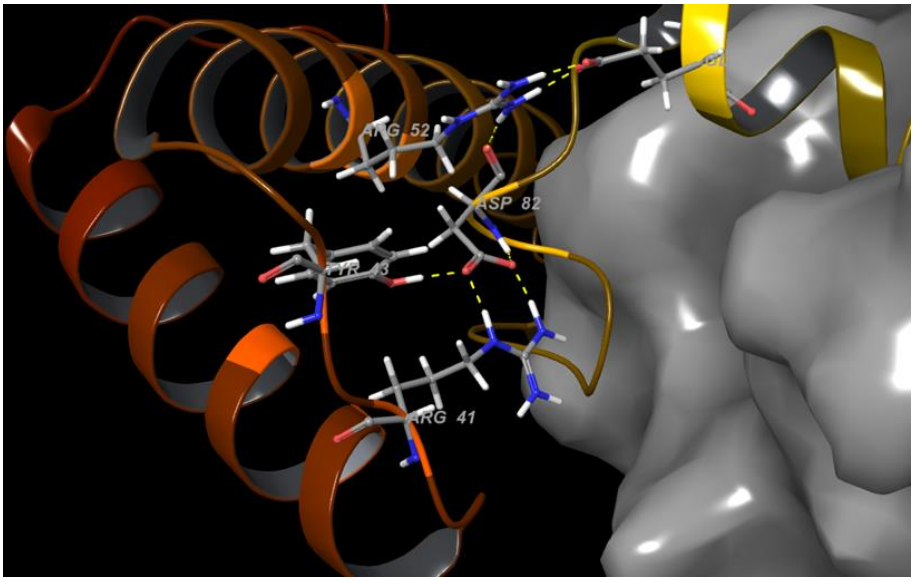

Figure 23. Hydrogen-bond network shown at the pro domain of PCZN.

2.4.4 Protein Residue-Residue Interaction analysis

Interactions within a protein are essential for stability and function of the protein. There are several weak as well as strong interacting forces that typically govern interactions mediated by a protein. In our aim to enumerate critical residues-residue non-covalent-interactions that can occur within the pro domain, primarily ionic and hydrophobic interactions, we utilized PIC server. Enlisted in Table 2 are the intra-protein residue interactions in pro domain of PCZN structure predicted by PIC server. Results revealed several ionic interactions hydrophobic interactions within the pro domain. Intra-protein interactions between the cat and pro domains as well as within the catalytic domain were also observed, but are not included in this study since our main focus is the impact of the pro domain. The underlined residues correspond to some of the residues found in the amino acid sequence alignment in Figure 14.

Table 2. Inter-protein ionic and hydrophobic interactions found within pro-segment of PCZN.

Interactions done by PIC server and found within 6Å.

Ionic	Hydrophobic
Glu87-Arg91	Phe32-Phe35
Glu25-Arg57	Phe32-Phe56
Arg41-Asp82	Phe35-Phe56
Lys36-Glu49	
Glu86-Arg89	
Arg52-Glu87	

The inter-protein interactions of the pro domain of PCZN obtained from the PIC server were visualized using PyMol software. Distances between residues were also measured with this server. Nevertheless, only the imagining of interactions Glu25-Arg57, Asp82-Arg41, and Glu87-Arg91 were shown here (Figure 24Figure 25,Figure 26) due to their implications in protein stability according to data which will be shown in chapter 3. Each residue-residue interaction showed a distance of 1.6, 1.5, and 1.5Å for

Glu25-Arg57, Asp82-Arg41, and Glu87-Arg91, respectively. Surprisingly, while visualizing each of the residue-residue interactions and residue substitutions within the pro domain, we discovered a new ionic interaction. Arg41 that was ionically interacting with Asp82 in the wild type structure shifted its position when the amino acid Asp82 was substituted to Val, causing Arg41 get attracted to the closest negatively charged side chain of Glu44 (Figure 27).

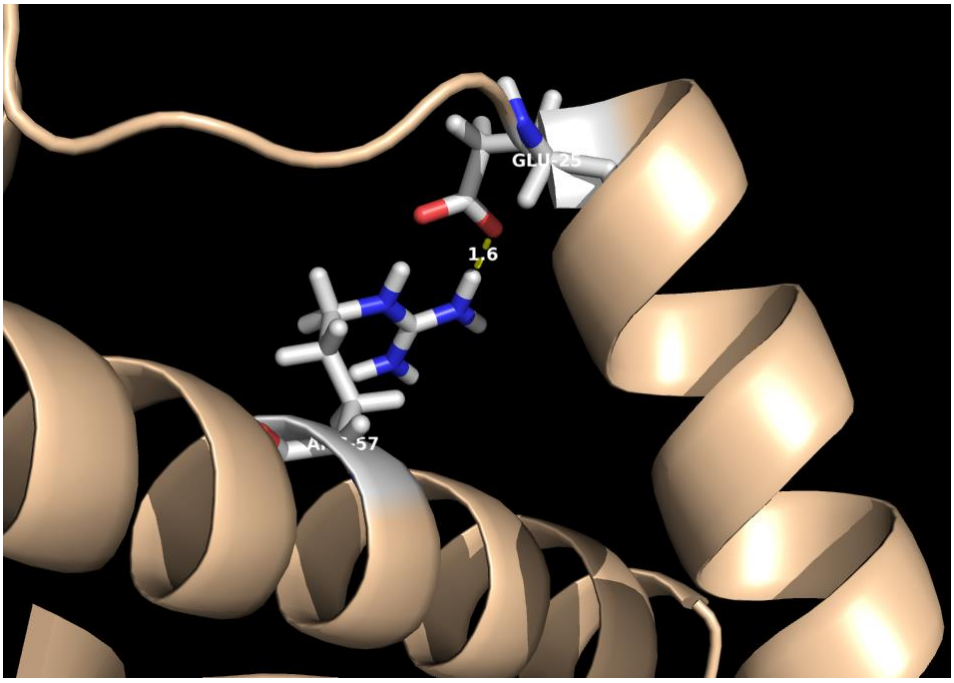

Figure 24. Ionic interaction between Glu25-Arg57 of pro domain of PCZN (light brown).

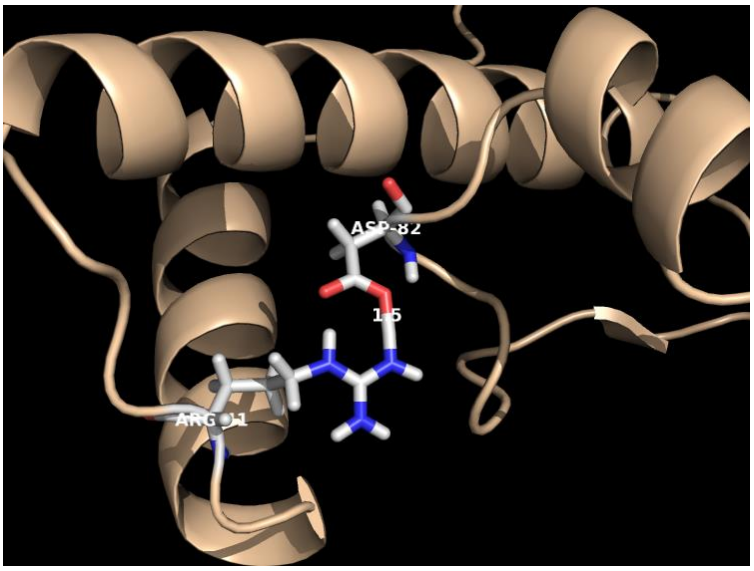

Figure 25. Ionic interaction between Asp82-Arg41 of pro domain of PCZN. Light brown backbone represents the pro-domain.

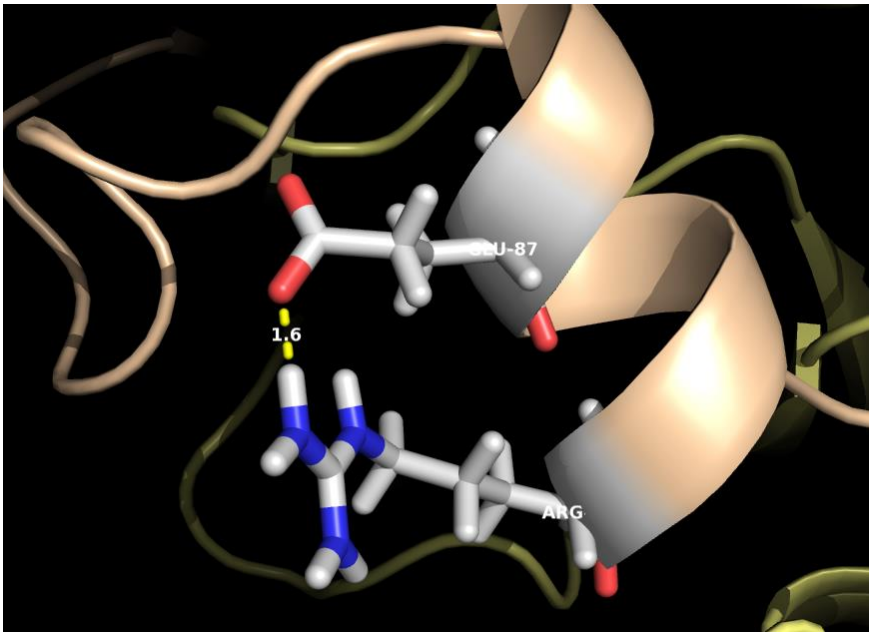

Figure 26. Ionic interaction between Glu87-Arg91 of pro domain of PCZN. Light brown and yellow backbones represent the catalytic-domain and the pro-domain respectively.

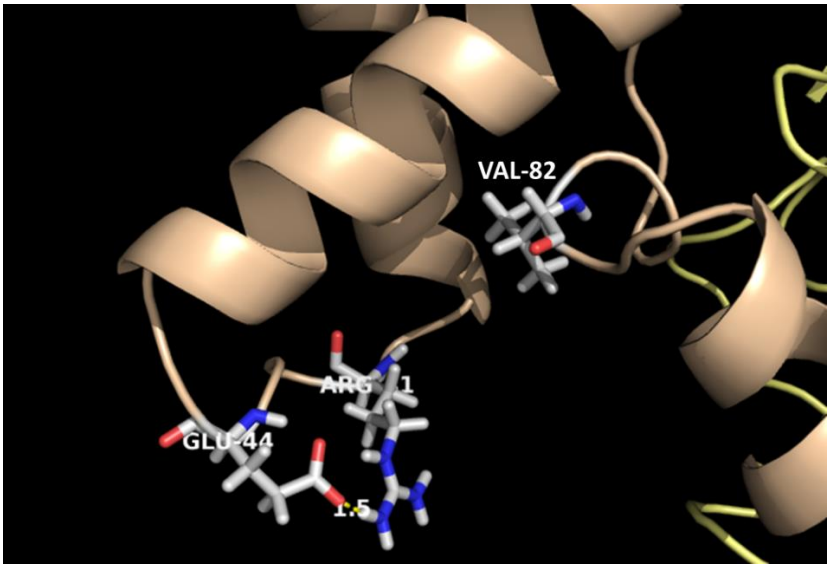

Figure 27. New interaction (Glu44-Arg41) found in PCZN-D82V. Light brown and yellow backbones represent the catalytic-domain and the pro-domain respectively.

2.4.5 Structural impact of mutants of homology-modeled PCZN

The impact of already selected mutations were calculated by some analytical tools. Table 3 provides with data obtained for each mutant denoting the overall protein structure stability. The higher the free energy values the less stable the structure was. Mutants F35A and F56A, showed the highest instability of all, followed by D82V, E87V, F32A, and D176V with a considerable degree of instability as well. Results from this data analysis confirmed the important of the ionic residue selected using PIC in the structure stability of PCZN.

Table 3. Stability of mutants of PCZN calculated by some analytical tools. *Weighted combination between ENCoM and FoldX 3.0. All values are in kcal/mol

Mutant	FoldX3	FoldX4	ENCoM	Com*	Unstable
F35A	1.47	3.97	1.68	3.16	+++
F56A	1.59	4.19	1.54	3.14	+++
D82V	1.95	5.10	0.27	2.22	++
E87V	1.69	4.41	0.15	1.85	+
F32A	1.42	3.67	1.43	2.85	+

The thermodynamic parameters estimated from the above bioinformatics tools are very helpful in predicting protein stability. Entropy in particular can be considered the most useful parameter in understanding the stability of proteins because when a protein molecule is deactivated the randomness of the system increases which is a direct measure of entropy [77]. Previous studies have found that the free energy of inactivation of mutants was found to be less (more entropy energy as consequence) than wild type indicating that mutants are less stable than the wild-type enzyme [78].

Structural alignment of wild type and mutants revealed a slight difference among the structures as observed in Figure 28. Nonetheless, a more in-depth structural analysis of mutants generated in silico are further necessary in order to validate and complement these results. Bioinformatics tools offers scientist the opportunity to characterize structures without the need to spend years in the lab bench trying to crystalize a protein. It is important to recall that not all mutations generate dramatic changes in structure or large effects on protein function, but we could see that the mutations observed in this study did have some effect on protein maturation rate and secondary structure.

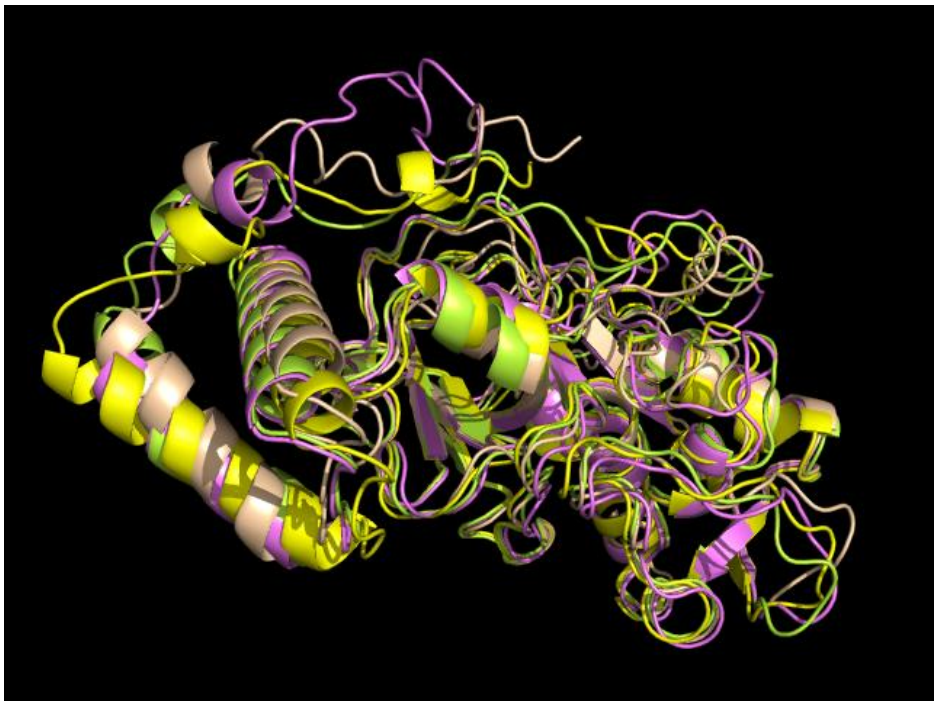

Figure 28. Structural alignment of wild type and mutants of PCZN. Wild type colored in light brown; E25V colored in green; D82V colored in yellow; E87V colored in pink.

2.5 Conclusions

A homology predicted model of PCZN (model1) was developed using I-Tasser by the utilization of X-Ray structures of related protein sequence. Model1 was subsequently subjected to MD simulation studies to generate a high-resolution 3D structure of PCZN WT (model2) and its mutants: PCZN-E25V, PCZN-D82, PCZN-E87V. A well-reproduced PCZN structure was confirmed by the high similarity (small RMSD values) with the already solved Pro-cathepsin (PDB ID:1CJL). Crucial residue-residue interactions predicted by PIC, were then evaluated by using servers FoldX3, FoldX4, ENCoM, and Com. These findings led us to access undiscovered structural-weakening sites within the pro and catalytic domains of procruzain. The sites identified in this study could furthermore help in the development of small-molecule compounds that could bind these spots to create an allosteric effect and/or disrupting critical aminoacid-

aminoacid interactions within the proenzyme. Further experimental data is necessary to corroborate these findings. These computational data will help to reduce the vast number of possible sites that could be considered as candidates to be tested experimentally in order to find critical interactions within the full PCZN structure.

CHAPTER 3

EFFECT OF POINT MUTATIONS AND IMPORTANCE OF PRO DOMAIN IN ACTIVITY AND STRUCTURE OF PRO-CRUZAIN

3.1 Background

As mentioned in chapter 1, cysteine proteases are not synthesized as active enzymes to avoid uncontrolled self-digestion before they reach their destination. Instead, they are synthesized as latent pro-proteins that differ from their mature counterparts by an additional N-terminal pro domain fragment, which runs through the active-site cleft along the direction opposite to that of normal substrates. Although many studies have been performed in the mature (catalytic) domain of the main cysteine protease of T. cruzi [24, 79–82], very little information has been found on the impact of the pro domain in the maturation of this cysteine protease. In this chapter we will focus in one of the pro domain functions, which is the assistance in folding of the pro-protein. Our main interest is to decipher the importance of pro domain of PCZN via the analysis of mutants of PCZN.

3.1.1 Pro domain

An increasing number of proteases have been found to be synthesized as pro-enzymes, which are known as zymogens. Essentially all known extracellular bacterial proteases are made as pre-pro-proteins (the pre-region being the signal sequence), and a growing number of intracellular and extracellular eukaryotic proteases have also been shown to be synthesized as pro-enzymes where the amino-terminal pro-peptides are the most common. Pro-enzymes are activated by intramolecular, autoproteolytic cleavage of the pro domain from its precursor in order to create the mature form of the enzyme[32, 83, 84].

3.1.2 Role of pro-domain in protein folding rate

In several well characterized cases, it has been conclusively demonstrated that the pro-regions (pro domains) are, indeed, required for the folding of their associated protease domains. Enzymes like aspartic proteinases [85], cysteine proteinases such as cathepsin D and papain [86], serine proteinases such as subtilisin and alpha-lytic [83, 87], and pro-hormone convertases [88] have been found to rely on their pro-segment to achieve their native folds. Studies on subtilisin demonstrated that isolated pro-peptide greatly accelerates the *in vitro* folding of this enzyme. The 77 amino acid pro-peptide of subtilisin has been shown to promote subtilisin folding, and mutations have been identified which abolish its function *in vivo* [89]. In addition to this, it has been reported that the isolated pro-peptide of subtilisin does not fold autonomously [90]. Although, when the isolated and unstructured pro-peptide is added to the mature subtilisin, folding of both of them occurs spontaneously [91]. Similar studies on alpha-lytic protease showed that pro-peptide binds to, and stabilizes a native-like transition state in the folding reaction of this enzyme [87].

There must be several other proteases which have pro-domains that may play important and specific roles in protein folding. In fact, expression experiments with the cysteine proteases papain, cruzain, and cathepsin L also imply a requirement for pro-regions. Although, while the serine proteases alpha-lytic and subtilisin, are the two most extensively studied cases of pro-region assisted folding, studies pertaining this topic in cruzain are still without deep investigation.

While recent research convincingly demonstrates that external introduction of the folded 10 kDa pro-domain effectively inhibits mature cruzain function cementing the pro-petides' role as a potential antitrypanocidal [60], logistical and practical barriers limit its administration as a potential therapeutic. Here we are focused on therapeutic and prophylactic compounds that are aimed to the interface between pro-domain and cruzain for the prevention of PCZN maturation.

3.2 Specific Aims

As stated before, pro domain plays a very important role in the folding and maturation of pro enzymes, making the pro domain of cruzain a key target for a more in-depth exploration. Normally, when there is a mutation in one or more residues of a protein, it causes the folding to change. This will then cause the protein to not being able to function properly. Therefore, the possibility to identify critical residues within the pro domain of cruzain that may alter the enzyme's function is our main goal. Herein, we hypothesize that residue substitutions within the pro-domain of cruzain will destabilize the structure and ultimately its function.

Specific Aim One: to investigate the implication of cruzain's pro domain in the maturation of the enzyme by testing the autoactivation rate of PCZN mutants.

Specific Aim Two: to analyze the effect of residue substitutions at the pro domain on the secondary structure of PCZN by circular dichroism studies.

3.3 Materials and Methods

3.3.1 Activation assays of recombinant mutants and wild type PCZN.

All recombinant proteins were purified, refolded and dialized as mentioned in chapter 1. After protein was refolded and dialyzed with cold water, the protein concentration was adjusted to 0.3mg/mL. Protease activity of recombinant mutants and WT PCZN was measured using Z-Phe-Arg-7-amido-4-methyl-coumarin-HCl (AMC) as substrate (where Z is benzyloxycarbonyl), and performed as previously described [32] 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 440nm emission wavelength and 380nm excitation wavelength and set for 0.100 integration time and 600 PMT HV. The reaction cuvette, with a total volume of 2050uL, was filled with activation buffer (5mM EDTA, 100mM

NaH_2PO_4, 200mM NaCl, 2mM DTTred, pH 4.5) and substrate (40uM AMC), 5mM EDTA, 50mM NaH_2PO_4, 100mM NaCl, 5mM DTTred) and adjusted to pH 5.3. Reaction was then initiated by addition of 50uL of refolded protein at a concentration of 0.3mg/mL. The fluorescence was recorded promptly after protein addition and incubated at 37˚C for a period of 5000sec. When maximum activity was observed, the mature enzyme was inhibited with 10uM 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 Coommassie brilliant blue R-250 (BioRad).

3.3.2 Circular dichroism studies of folded recombinant mutants and wild type PCZN.

Refolded protein was dialyzed against Phosphate-buffered saline buffer (PBS) pH 7.4 using 3kDa MWCO tubes (GE Healthcare) and further concentrated using centrifugal filter tubes (Vivaspin2, 3kDa MWCO). Resultant dialyzed protein was passed through PVDF 0.22µm 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). Jasco 1500 spectropolarimeter was used for CD measurements that has Peltier temperature controller. Data was documented at 4°C and 37°C and samples were incubated for 10 minutes prior recording of the data. CD studies were performed in collaboration with Dr. Supriyo Ray from the University of Texas at El Paso.

3.4 Results and Discussion

3.4.1 Activation assays of recombinant mutants and wild type PCZN.

Although pro domains of several cysteine proteases have shown to be potent inhibitors of other papain-like enzymes [60], 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 8). Slope of the curve of the autoactivation of PCZN and its mutants resulted as follows: WT= $8.43x10^{-4}$, E25V= $2.56x10^{-4}$, E87V= $3.73x10^{-4}$, D82V= $1.31x10^{-5}$. 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[90], we considered mutant C25A (C141A) as negative control for the studies for autoactivation of mutants. Therefore, C141A mutant showed not 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 than the rest of the mutants (Figure 27).

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 [92]; or in the case of cathepsin L that resulted in improper folding after deletions in its pro domain [93]. Nevertheless, not all mutations produced a negative effect on the activity of proteases as demonstrated in another study done by Ruan *et al.,* 1999 [94] 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.

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.

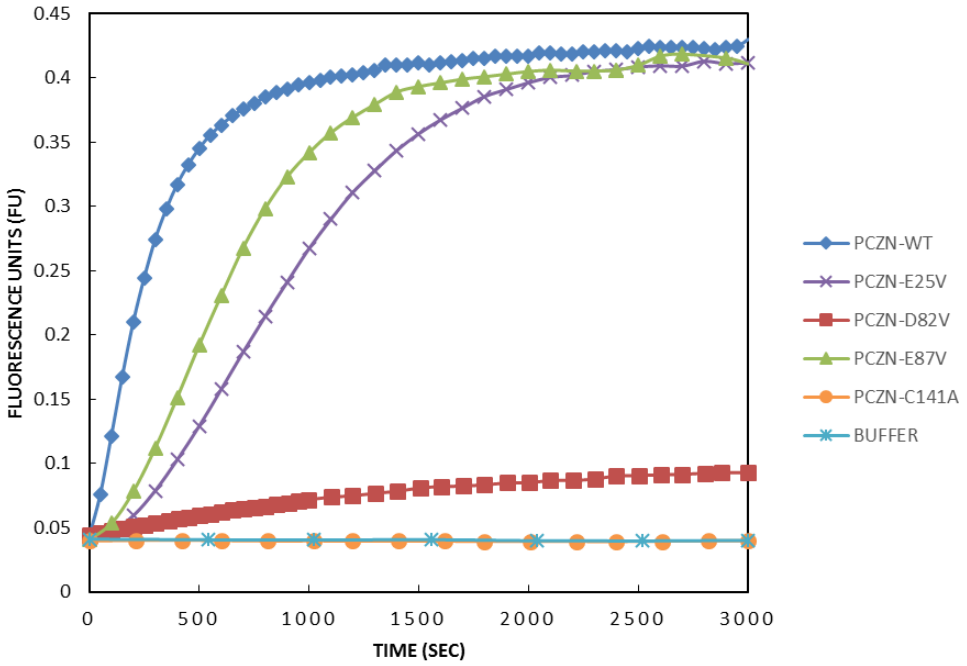

Figure 29. Activation of recombinant PCZN-WT and mutants. The kinetics of autoactivation of recombinant PCZN was monitored from cleavage of AMC from Z-Phe-Arg-AMC substrate. Diamonds: PCZN-WT, cross marks: E25V, squares: PCZN-D82V, triangles: PCZN-D87V, circles: PCZ-C141A, stars: activation buffer alone.

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 [92]; or in the case of cathepsin L that resulted in improper folding after deletions in its pro domain [93]. Nevertheless, no all mutations produced a negative effect on the activity of proteases where in another study done by Ruan *et al.,* 1999, 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.

3.4.2 Characterization of PCZN and mutants

To determine the effect of mutation on the overall structure of PCZN, we did far-UV CD of the PCZN- mutants and compared it with the structure of PCZN-WT. Putative 3D homology modeling predicted predominant helical structure (30%) for the PCZN-WT (Figure 15). WT far-UV CD data confirmed our prediction that the PCZN WT structure is mainly α-helical with strong minima at 222 nm and 208 nm. At 4˚C we looked at the native confirmation of the WT and all the mutants (Figure 30). PCZN-WT has a stronger minima at 208 nm than at 222 nm signifying rigidity of the helix. Overall, all mutations were shown to decrease the total helicity of the protein. D82V, E25V and E87V mutations increased the hydrophobic content of the protein as they replaced negatively charged amino acid residues involved in hydrogen bonding. This may partially explain decrease in helical content compared to the PCZN-WT.

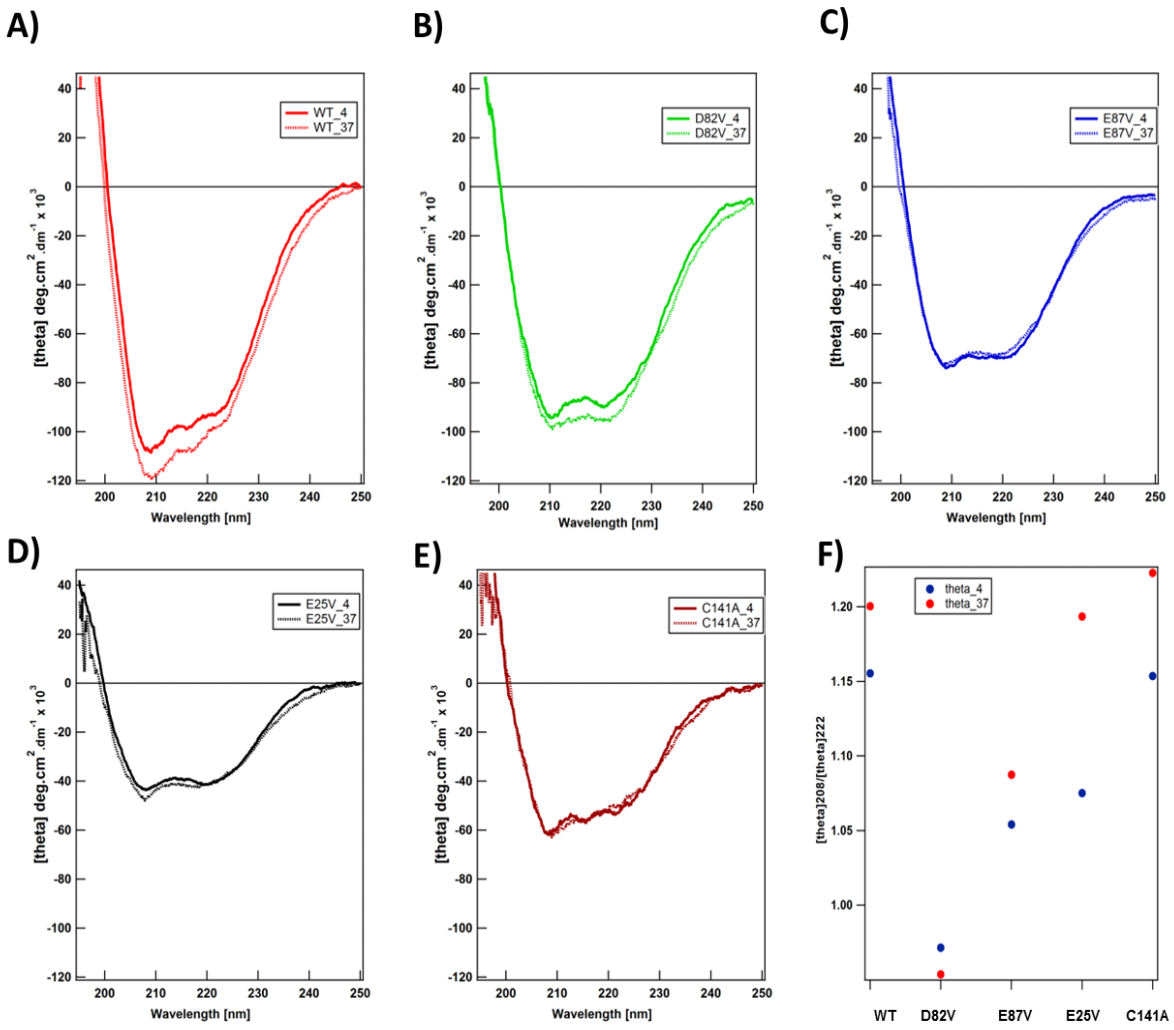

Figure 30. Far-UV circular dichroism of PCZN-WT and compared with that of the mutants at 37˚C. A-E: secondary structure determination at 37˚C of WT in comparison with the PCZN mutants; F: ratio of $[\theta]_{208}/[\theta]_{222}$ (minimas for α-helical conformation) at 4˚C (native conformation) in comparison with that at 37˚C (conformation at physiological temperatures).

Figure 31 also shows the helical wheel comparison between WT and the mutants. In all helices analyzed, except for E25V, no other helices have a long stretch of hydrophobic residues (5 and above). The distribution of hydrophobic residues and hydrophilic mostly alternate each rather than forming a

patch. E25V mutation shows a hydrophobic stretch facing outside. The mutation increases the hydrophobicity of the peptide stretch that was already rich in hydrophobic residues.

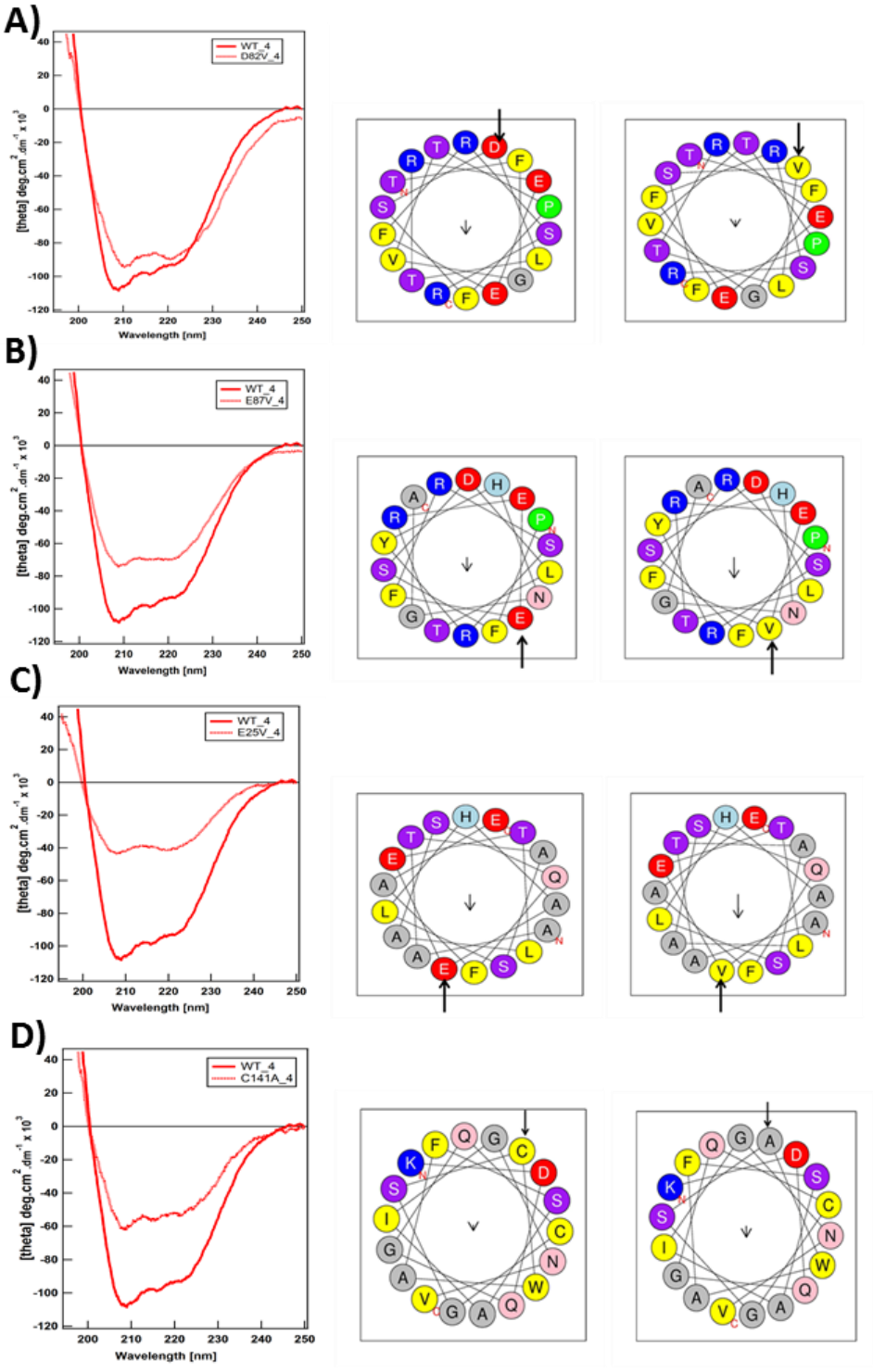

Figure 31. Circular dichroism spectra and helical wheels for PCZN-WT and mutants. CD data demonstrates the native structure of PCZN-WT at 4˚C and that is compared with the mutants at same

temperature. The helical wheel diagram shows the change in the helical structure post mutation. Arrows indicate the residue mutated. "N" indicates N-terminal and "C" C-terminal positions.

In Figure 30 (A-E), we looked at the effect of physiological temperature on the overall structure of PCZN-WT and their mutants. In Figure 30 (F) we looked at the ratio of $[\theta]_{208}$ and $[\theta]_{222}$ since they are the minimas for α-helical signature conformation. As PCZN-WT had a stronger minima at 4˚C (the native conformation), we compared it with that at 37˚C (conformation at physiological temperatures). From the activity data, PCZN-WT had the highest initial activity as well as highest V_{max} compared with the other mutants at 37˚C. So, we compared the native conformational change in the mutants when the temperature was increased from 4˚C to 37˚C and compared it with the WT. The change in conformation at 37˚C denotes the effect of temperature on the structure of the enzyme. We compared the change in the ratio of $[\theta]_{208}$ and $[\theta]_{222}$ at 4˚C to that of 37˚C. The WT showed increase in the ratio because there was an overall increase in the helical content at the physiological temperature. This was also seen in E87V and E25V mutants and they do show high enzymatic activity. Nonetheless, the initial activity (first five minutes) was less in both the mutants and had an extended lag phase. This could be attributed to the reduced helical content in their native conformation at 4˚C when compared to the WT. At 37˚C there is an increase in helical content in both the mutants and E25V shows highest increase in the ratio. D82V shows a reverse trend where the ratio decreases when switched to 37˚C. This was because there was an overall increase in $[\theta]_{222}$ signifying relaxation of the helix. It can be speculated that the rigidity of the helix could be important for higher enzymatic activity. Interestingly, C141A mutant in spite of increase in ratio of $[\theta]_{208}$ and $[\theta]_{222}$ like the WT and enzymatically active mutants, it showed little or no activity. This is due to the fact that the mutation directly affects the active site unlike other mutations that affect other parts of the protein that are not involved directly in substrate processing. The other mutations may affect substrate acquisition partially explaining the extended lag phase. Thus, conformational adjustment at physiological temperatures by PCZN plays a critical role in enhanced enzymatic activity.

3.5 Conclusions

Changes in secondary structure of each of the mutants was observed during the CD studies. This structural changes are attributed to the breakage in the ionic interaction between the residues observed within the pro domain of cruzain. These results were in concordance with the rate reduction of the auto-activation of each mutant with respect to the wild type. These finding are of great significance and provide evidence that pro domain plays an important role in the maturation of cruzain and could possibly open the field for novel inhibitor design.

FUTURE PLANS

As future plans we propose to perform a multi-set of mutagenesis compromised of a combination of all residues mutated in this study in one plasmid as an additional evaluation of these sidechains in the structural impact of PCZN structure. In addition, mutagenesis of observed hydrophobic cluster within the pro domain, compromised of phenilalanines 32, 35 and 56, may be valuable to determine its implication in protein destabilization. Finally, an in-depth study within the crucial residue-residue interactions described in this study via bioinformatics tools could be very beneficial, since the contribution of the surrounding residues may be important for the characterization of pockets that can aid in the development of anti-chagastics drugs.

REFERENCES

1. Lewinsohn, R. (1981). Carlos Chagas and the discovery of Chagas' disease (American trypanosomiasis). *Journal of the Royal Society of Medicine*, *74*(6), 451–5. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1438770&tool=pmcentrez&rendertype=abstract

2. Rassi, A., & Marin-Neto, J. A. (2010). Chagas disease. *Lancet*, *375*(9723), 1388–402. doi:10.1016/S0140-6736(10)60061-X

3. Who, F. (2010). Working to overcome the global impact of neglected tropical diseases First WHO report on neglected tropical diseases. *World Health*, *86*(13), 1–184. Retrieved from http://whqlibdoc.who.int/publications/2010/9789241564090_eng.pdf

4. Noireau, F., Diosque, P., & Jansen, A. M. (2009). Trypanosoma cruzi: adaptation to its vectors and its hosts. *Veterinary research*, *40*(2), 26. doi:10.1051/vetres/2009009

5. Kirchhoff, L. V. (1955). American trypanosomiasis (Chagas disease). *Meditsinskaia Parazitologiia I Parazitarnye Bolezni*, *7*(3), 58–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22632639

6. Tanowitz, H. B., Kirchhoff, L. V, Simon, D., Morris, S. A., Weiss, L. M., & Wittner, M. (1968). Chagas' disease. *Proceedings of the Royal Society of Medicine*, *5*(5), 444–445. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=358257&tool=pmcentrez&rendertype=abstract

7. Kirchhoff, L. V, & Pearson, R. D. (2007). The emergence of chagas disease in the United States and Canada. *Current Infectious Disease Reports*, *9*(5), 347–350. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17880844

8. Engman, D. M., & Leon, J. S. (2002). Pathogenesis of Chagas heart disease: role of autoimmunity. *Acta Tropica*, *81*(2), 123–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11801219

9. Leon, J. S., Wang, K., & Engman, D. M. (2003). Captopril ameliorates myocarditis in acute experimental Chagas disease. *Circulation*, *107*(17), 2264–9. doi:10.1161/01.CIR.0000062690.79456.D0

10. 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. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2213346&tool=pmcentrez&rendertype=abstract

11. Cazzulo, J. J., Stoka, V., & Turk, V. (2001). The major cysteine proteinase of Trypanosoma cruzi: a valid target for chemotherapy of Chagas' disease, Curr. *Pharm Des*, *7*, 1143e1156.

12. McKerrow, J. H., Engel, J. C., & Caffrey, C. R. (1999). Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorganic & Medicinal Chemistry*, *7*(4), 639–644. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10353643

13. McKerrow, J. H., McGrath, M. E., & Engel, J. C. (1995). The cysteine protease of Trypanosoma cruzi as a model for antiparasite drug design. *Parasitology today Personal ed*, *11*(8), 279–282. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15275323

14. Avila, A. R., Araujo, P. R., Bern, C., Blangero, J., Buckner, F. S., Carvalho, A. C. C. De, … Weiss, L. M. (2011). *Chagas Disease, Part A*. (L. M. Weiss, H. B. Tanowitz, & L. V Kirchhoff, Eds.)*Advances in Parasitology* (Vol. 62). Academic Press. doi:10.1016/S0065-308X(05)62017-9

15. Urbina, J. A. (2002). Chemotherapy of Chagas disease. *Current Pharmaceutical Design*, *8*(4), 287–295. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11860367

16. Urbina, J. A., & Docampo, R. (2003). Specific chemotherapy of Chagas disease: controversies and advances. *Trends in Parasitology*, *19*(11), 495–501. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S1471492203002411

17. Rodriques Coura, J., & De Castro, S. L. (2002). A critical review on Chagas disease chemotherapy. *Memórias do Instituto Oswaldo Cruz*, *97*(1), 3–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11992141

18. Castro, J. A., De Mecca, M. M., & Bartel, L. C. (2006). Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). *Human experimental toxicology*, *25*(8), 471–479. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16937919

19. Jackson, Y., Alirol, E., Getaz, L., Wolff, H., Combescure, C., & Chappuis, F. (2010). Tolerance and safety of nifurtimox in patients with chronic chagas disease. *Clinical Infectious Diseases*, *51*(10), e69–e75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20932171

20. Hotez, P. J., Dumonteil, E., Woc-Colburn, L., Serpa, J. a, Bezek, S., Edwards, M. S., Bottazzi, M. E. (2012). Chagas Disease: "The New HIV/AIDS of the Americas". *PLoS neglected tropical diseases*, *6*(5), e1498. doi:10.1371/journal.pntd.0001498

21. Brinen, L. S., Hansell, E., Cheng, J., Roush, W. R., Mckerrow, J. H., & Fletterick, R. J. (n.d.). Site to define a target within the target : probing cruzain ' s P1 ' structural determinants for the Chagas ' disease protease, 831–840.

22. Kerr, I. D., Lee, J. H., Farady, C. J., Marion, R., Rickert, M., Sajid, M., Brinen, L. S. (2009). Vinyl sulfones as antiparasitic agents and a structural basis for drug design. *The Journal of Biological Chemistry*, *284*(38), 25697–25703. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2757971&tool=pmcentrez&rendertype=abstract

23. Bryant, C., Kerr, I. D., Debnath, M., Ang, K. K. H., Ratnam, J., Ferreira, R. S., … Renslo, A. R. (2009). Novel non-peptidic vinylsulfones targeting the S2 and S3 subsites of parasite cysteine proteases. *Bioorganic & Medicinal Chemistry Letters*, *19*(21), 6218–6221. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19773167

24. Sajid, M., Robertson, S. A., Brinen, L. S., & McKerrow, J. H. (2011). Cruzain : the path from target validation to the clinic. *Advances in experimental medicine and biology*, *712*, 100–115.

25. Huang, L., Brinen, L. S., & Ellman, J. A. (2003). Crystal structures of reversible ketone-Based inhibitors of the cysteine protease cruzain. *Bioorganic & Medicinal Chemistry*, *11*(1), 21–29. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12467703

26. Choe, Y., Brinen, L. S., Price, M. S., Engel, J. C., Lange, M., Grisostomi, C., … Craik, C. S. (2005). Development of alpha-keto-based inhibitors of cruzain, a cysteine protease implicated in Chagas disease. *Bioorganic & Medicinal Chemistry*, *13*(6), 2141–2156. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15727867

27. DiMasi, J. A., Feldman, L., Seckler, A., & Wilson, A. (2010). Trends in risks associated with new drug development: success rates for investigational drugs. *Clinical Pharmacology & Therapeutics*, *87*(3), 272–277. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20130567

28. McKerrow, J. H., Rosenthal, P. J., Swenerton, R., & Doyle, P. (2008). Development of protease inhibitors for protozoan infections. *Current Opinion in Infectious Diseases*, *21*(6), 668–672. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2732359&tool=pmcentrez&rendertype=abstract

29. Sarkar, S., Strutz, S. E., Frank, D. M., Rivaldi, C.-L., Sissel, B., & Sánchez-Cordero, V. (2010). Chagas disease risk in Texas. *PLoS neglected tropical diseases*, *4*(10), 14. doi:10.1371/journal.pntd.0000836

30. Cazzulo, J. J., Cazzulo Franke, M. C., Martínez, J., & Franke de Cazzulo, B. M. (1990). Some kinetic properties of a cysteine proteinase (cruzipain) from Trypanosoma cruzi. *Biochimica et biophysica acta*, *1037*(2), 186–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2407295

31. Rawlings, N. D., & Barrett, A. J. (1994). Families of cysteine peptidases. *Methods in enzymology*, *244*, 461–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7845226

32. Eakin, A. E., Mills, A., Harth, G., Mckerrowo, J. H., & Craiks, C. S. (1992). The Sequence , Organization , and Expression of the Major Cysteine Protease (Cruzain) from Trypanosoma cruzi. *The Journal of biological chemistry*, *267*(1990), 7411–7420.

33. Rangel, H. A., Araújo, P. M., Repka, D., & Costa, M. G. (1981). Trypanosoma cruzi: isolation and characterization of a proteinase. *Experimental Parasitology*, *52*(2), 199–209.

34. Tamashiro, W. M., Repka, D., Sakurada, J. K., Camargo, I. J., Araújo, P. M., Atta, A. M., & Rangel, H. A. (1983). Trypanosoma cruzi: surface antigenic determinants. *Zeitschrift fur Parasitenkunde Berlin Germany*, *69*(4), 425–434. Retrieved from http://dx.doi.org/10.1007/BF00927698

35. Franke de Cazzulo, B. M., MartÃ-nez, J., North, M. J., Coombs, G. H., & Cazzulo, J.-J. (1994). Effects of proteinase inhibitors on the growth and differentiation of Trypanosoma cruzi. *FEMS Microbiology Letters*, *124*(1), 81–86. doi:10.1111/j.1574-6968.1994.tb07265.x

36. Campetella, O., Martínez, J., & Cazzulo, J. J. (1990). A major cysteine proteinase is developmentally regulated in Trypanosoma cruzi. *FEMS Microbiology Letters*, *67*(1–2), 145–150. doi:10.1111/j.1574-6968.1990.tb13852.x

37. Tomás, a M., & Kelly, J. M. (1996). Stage-regulated expression of cruzipain, the major cysteine protease of Trypanosoma cruzi is independent of the level of RNA1. *Molecular and biochemical parasitology*, *76*(1–2), 91–103. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8919998

38. Gillmor, S. A., Craik, C. S., & Fletterick, R. J. (1997). Structural determinants of specificity in the cysteine protease cruzain. *Protein Science*, *6*(8), 1603–1611. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2143760&tool=pmcentrez&rendertype=abstract

39. Sajid, M., & McKerrow, J. H. (2002). Cysteine proteases of parasitic organisms. *Molecular and Biochemical Parasitology*, *120*(1), 1–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11849701

40. Souto-Padrón, T., Campetella, O. E., Cazzulo, J. J., & De Souza, W. (1990). Cysteine proteinase in

Trypanosoma cruzi: immunocytochemical localization and involvement in parasite-host cell interaction. *Journal of Cell Science*, *96 (Pt 3)*(2), 485–490. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2229199

41. Stoka, V., Turk, B., Schendel, S. L., Kim, T. H., Cirman, T., Snipas, S. J., … Salvesen, G. S. (2001). Lysosomal protease pathways to apoptosis. Cleavage of bid, not pro-caspases, is the most likely route. *The Journal of Biological Chemistry*, *276*(5), 3149–3157. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11073962

42. Protease, M. C., Eakin, A. E., Mcgrath, M. E., Mckerrow, J. H., Fletteric, R. J., & Craiks, C. S. (1993). Production of Crystallizable Cruzain, the Major Cysteine Protease from, *268*(9), 6115–6118.

43. Aslund, L., Henriksson, J., Campetella, O., Frasch, a C., Pettersson, U., & Cazzulo, J. J. (1991). The C-terminal extension of the major cysteine proteinase (cruzipain) from Trypanosoma cruzi. *Molecular and biochemical parasitology*, *45*(2), 345–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8711424

44. Stoka, V., McKerrow, J. H., Cazzulo, J. J., & Turk, V. (1998). Substrate inhibition of cruzipain is not affected by the C-terminal domain. *FEBS Letters*, *429*(2), 129–133. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9650575

45. Stoka, V., Turk, B., McKerrow, J. H., Björk, I., Cazzulo, J. J., & Turk, V. (2000). The high stability of cruzipain against pH-induced inactivation is not dependent on its C-terminal domain. *FEBS Letters*, *429*(2), 129–133.

46. Martínez, J., Campetella, O., Frasch, A. C., & Cazzulo, J. J. (1993). The reactivity of sera from chagasic patients against different fragments of cruzipain, the major cysteine proteinase from Trypanosoma cruzi, suggests the presence of defined antigenic and catalytic domains. *Immunology Letters*, *35*(2), 191–196.

47. Tao, K., Stearns, N. a, Dong, J., Wu, Q. L., & Sahagian, G. G. (1994, May 15). The proregion of cathepsin L is required for proper folding, stability, and ER exit. *Archives of biochemistry and biophysics*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8185316

48. Mach, L., Mort, J. S., & Glössl, J. (1994). Noncovalent complexes between the lysosomal proteinase cathepsin B and its propeptide account for stable, extracellular, high molecular mass forms of the enzyme. *The Journal of Biological Chemistry*, *269*(17), 13036–13040.

49. Conner, G. E. (1992). The role of the cathepsin D propeptide in sorting to the lysosome. *The Journal of Biological Chemistry*, *267*(30), 21738–21745.

50. McIntyre, G. F., Godbold, G. D., & Erickson, A. H. (1994). The pH-dependent membrane association of procathepsin L is mediated by a 9-residue sequence within the propeptide. *The Journal of Biological Chemistry*, *269*(1), 567–572.

51. Ishidoh, K., & Kominami, E. (2002). Processing and activation of lysosomal proteinases. *Biological Chemistry*, *383*(12), 1827–1831. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12553719

52. Eder, J., & Fersht, A. R. (1995). Pro-sequence-assisted protein folding. *Molecular Microbiology*, *16*(4), 609–614. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7476156

53. Shinde, U., & Inouye, M. (2000). Intramolecular chaperones: polypeptide extensions that modulate protein folding. *Seminars in cell developmental biology*, *11*(1), 35–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10736262

54. Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., & Li, Q. (2009). Practical protocols for production of very high yields of recombinant proteins using Escherichia coli, *18*(1), 936–948. doi:10.1002/pro.102

55. Loh, Y. P. (1993). *Mechanisms of intracellular trafficking and processing of proproteins*. CRC Press.

56. Doyle, P. S., Zhou, Y. M., Engel, J. C., & 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. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2151429&tool=pmcentrez&rendertype=abstract

57. Rogers, K., Keränen, H., Durrant, J. D., Ratnam, J., Doak, A., Arkin, M. R., & McCammon, J. A. (2012). Novel cruzain inhibitors for the treatment of Chagas' disease. *Chemical Biology Drug Design*, no-no. doi:10.1111/j.1747-0285.2012.01416.x

58. Ferreira, R. S., Dessoy, M. A., Pauli, I., Souza, M. L., Krogh, R., Sales, A. I. L., … Andricopulo, A. D. (2014). Synthesis, Biological Evaluation, and Structure–Activity Relationships of Potent Noncovalent and Nonpeptidic Cruzain Inhibitors as Anti- *Trypanosoma cruzi* Agents. *Journal of Medicinal Chemistry*, *57*(6), 2380–2392. doi:10.1021/jm401709b

59. Shinde, U. P., Liu, J. J., & Inouye, M. (1997). Protein memory through altered folding mediated by intramolecular chaperones. *Nature*, *389*(6650), 520–2. doi:10.1038/39097

60. Reis, F. C. G., Costa, T. F. R., Sulea, T., Mezzetti, A., Scharfstein, J., Brömme, D., … Lima, A. P. C. a. (2007). The propeptide of cruzipain--a potent selective inhibitor of the trypanosomal enzymes cruzipain and brucipain, and of the human enzyme cathepsin F. *The FEBS journal*, *274*(5), 1224–34. doi:10.1111/j.1742-4658.2007.05666.x

61. Vyas, V. K., Ukawala, R. D., Ghate, M., & Chintha, C. (2012). Homology modeling a fast tool for drug discovery: current perspectives. *Indian journal of pharmaceutical sciences*, *74*(1), 1–17. doi:10.4103/0250-474X.102537

62. Özlem Tastan Bishop, A., de Beer, T. A. P., & Joubert, F. (2008). Protein homology modelling and its use in South Africa. *South African Journal of Science*.

63. 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. doi:10.1016/S0022-2836(05)80360-2

64. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Notredame, C. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, *7*(1), 539. doi:10.1038/msb.2011.75

65. 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. doi:10.1038/nprot.2010.5

66. D.A. Case, J.T. Berryman, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T. J. G., H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. L., P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I., Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. S., & R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D. M. Y. and P. A. K. (2015). AMBER 2015.

67. Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. . (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of*

Computational Physics, *23*(3), 327–341. doi:10.1016/0021-9991(77)90098-5

68. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, *103*(19), 8577. doi:10.1063/1.470117

69. Miao, Y., Feixas, F., Eun, C., & McCammon, J. A. (2015). Accelerated molecular dynamics simulations of protein folding. *Journal of Computational Chemistry*, *36*(20), 1536–1549. doi:10.1002/jcc.23964

70. K. G. Tina, R. B. and N. S. (2007). Protein Interactions Calculator. *Nucleic Acids Research*, *35*, W473–W476.

71. PyMOL Molecular Graphics System. Schrödinger, LLC. (n.d.).

72. Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S., & Cygler, M. (1996). Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *The EMBO journal*, *15*(20), 5492–503. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=452294&tool=pmcentrez&rendertype=abstract

73. Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*. doi:10.1038/nprot.2010.5

74. Jianyi Yang, Renxiang Yan, Ambrish Roy, Dong Xu, Jonathan Poisson, Y. Z. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, *12*, 7–8.

75. Reif, M. M., Mach, L., & Oostenbrink, C. (2012). Molecular insight into propeptide-protein interactions in cathepsins L and O. *Biochemistry*, *51*(43), 8636–53. doi:10.1021/bi300802a

76. Vernet, T., Tessier, D. C., Chatellier, J., Plouffe, C., Lee, T. S., Thomas, D. Y., … Ménard, R. (1995). Structural and functional roles of asparagine 175 in the cysteine protease papain. *The Journal of Biological Chemistry*, *270*(28), 16645–16652. Retrieved from http://www.jbc.org/cgi/doi/10.1074/jbc.270.28.16645

77. Gummadi, S. N. (2003). What Is the Role of Thermodynamics on Protein Stability? *Biotechnology and Bioprocess Engineering*, *8*, 9–18.

78. Flory, N., Gorman, M., Coutinho, P. M., Ford, C., & Reilly, P. J. (1994). Thermosensitive mutants of Aspergillus awamori glucoamylase by random mutagenesis: inactivation kinetics and structural interpretation. *Protein engineering*, *7*(8), 1005–12. doi:10.1093/PROTEIN/7.8.1005

79. Lima, A. P. C. A., dos Reis, F. C. G., Serveau, C., Lalmanach, G., Juliano, L., Ménard, R., … Scharfstein, J. (2001). Cysteine protease isoforms from Trypanosoma cruzi, cruzipain 2 and cruzain, present different substrate preference and susceptibility to inhibitors. *Molecular and Biochemical Parasitology*, *114*(1), 41–52. doi:10.1016/S0166-6851(01)00236-5

80. McGrath, M. E., Eakin, a E., Engel, J. C., McKerrow, J. H., Craik, C. S., & Fletterick, R. J. (1995). The crystal structure of cruzain: a therapeutic target for Chagas' disease. *Journal of molecular biology*, *247*(2), 251–9. doi:10.1006/jmbi.1994.0137

81. Durrant, J. D., Keränen, H., Wilson, B. a, & McCammon, J. A. (2010). Computational identification of uncharacterized cruzain binding sites. *PLoS neglected tropical diseases*, *4*(5), e676. doi:10.1371/journal.pntd.0000676

82. Jose Cazzulo, J., Stoka, V., & Turk, V. (2001). The major cysteine proteinase of Trypanosoma cruzi: a valid target for chemotherapy of Chagas disease. *Current pharmaceutical design*, *7*(12), 1143–56. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11472258

83. Ikemura, H., & Inouye, M. (1988). In vitro processing of pro-subtilisin produced in Escherichia coli. *The Journal of Biological Chemistry*, *263*(26), 12959–12963. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11413931

84. Baker, D., Shiau, a K., & Agard, D. a. (1993). The role of pro regions in protein folding. *Current opinion in cell biology*, *5*(6), 966–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15700296

85. Winther, J. R., & Sørensen, P. (1991). Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of the enzymatic activity. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(20), 9330–9334. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=52708&tool=pmcentrez&rendertype=abstract

86. Vernet, T., Berti, P. J., De Montigny, C., Musil, R., Tessier, D. C., Ménard, R., … Thomas, D. Y. (1995). Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. *The Journal of Biological Chemistry*.Retrievedfrom http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1900296&tool=pmcentrez&rendertype=abstract

87. Baker, D., Silen, J. L., & Agard, D. A. (1992). Protease pro region required for folding is a potent inhibitor of the mature enzyme. *Proteins*, *12*(4), 339–344. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/prot.340120406/abstract

88. Steiner, D. F. (2011). Proprotein Convertases. *Current Opinion in Chemical Biology*, *768*(1), 31–9. doi:10.1007/978-1-61779-204-5

89. Kobayashi, T., & Inouye, M. (1992). Functional analysis of the intramolecular chaperone. Mutational hot spots in the subtilisin pro-peptide and a second-site suppressor mutation within the subtilisin molecule. *Journal of molecular biology*, *226*(4), 931–3. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1355566

90. Eder, J., Rheinnecker, M., & Fersht, a R. (1993, September 20). Folding of subtilisin BPN': role of the pro-sequence. *Journal of molecular biology*. doi:10.1006/jmbi.1993.1507

91. Strausberg, S., Alexander, P., Wang, L., Schwarz, F., & Bryan, P. (1993). Catalysis of a protein folding reaction: thermodynamic and kinetic analysis of subtilisin BPN' interactions with its propeptide fragment. *Biochemistry*, *32*(32), 8112–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8347611

92. 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), 20085–20086.

93. Smith, S. M., & Gottesman, M. M. (1989). Activity and deletion analysis of recombinant human cathepsin L expressed in Escherichia coli. *The Journal of biological chemistry*, *264*(34), 20487–95. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2684978

94. Ruan, B., Hoskins, J., & Bryan, P. N. (1999). Rapid folding of calcium-free subtilisin by a stabilized

pro-domain mutant. *Biochemistry*, *38*(26), 8562–71. doi:10.1021/bi990362n

VITA

Marisol Serrano, daughter of Margarita Marquez and Victor Romero, was born and raised in Delicias, Chihuahua, Mexico. She left her natal city to pursuit a higher education at The Autonomous University of Chihuahua, where she received her Bachelors of Science, with specialty in chemistry and microbiology. She then continued her graduate studies at the University of Texas at El Paso, where she received her Master of Science under the mentorship of Dr. Russell Chianelli with a specialty in microalgae as a source for biofuels. Right after, she started her doctoral studies at the University of Texas at El Paso under the supervision of Dr. Mahesh Narayan. During her doctoral studies she served as a teacher assistant for biochemistry laboratory and as a research associate in the Chemistry Department. She also grew professionally by working as a chemistry instructor and tutor at El Paso Community College, and as a Spanish instructor in medical terminology at Paul L. Foster School of Medicine-Texas Tech University. Currently, she is working as a laboratory manager at the El Paso Pain Center, where she started up a pharmacogenomics and a clinical analysis laboratory. In addition, Marisol have two first author publications in progress. She is very passionate about researching in chemistry and biology and its applications in health sciences industry. Marisol currently lives in El Paso, where she will continue working at El Paso Pain Center.