A CYSTEINE PROTEASE IS A TARGET FOR THE ENZYME STRUCTURE-BASED DESIGN OF ANTI-PARASITIC DRUGS

ANN E. EAKIN, PhD
Departments of Pharmaceutical Chemistry and Pathology

JAMES H. MCKERROW, MD, PhD
Department of Pathology

CHARLES S. CRAIK, PhD
Department of Pharmaceutical Chemistry
University of California, San Francisco, California

Proteases have been shown to be factors in the pathogenicity of many parasitic diseases, either by inducing tissue damage and facilitating invasion or by enabling the parasites to salvage metabolites from host proteins. To study genes encoding cysteine proteases of parasites, a general method for cloning fragments of thiol protease genes was developed using the polymerase chain reaction (PCR) with degenerate oligonucleotide primers. Subsequently, a PCR-amplified gene fragment of the protozoan parasite, Trypanosoma cruzi, was used to isolate a full-length gene encoding a cysteine protease. At least six copies of the gene are organized in the genome as a tandem array. The high degree of sequence identity with the papain family of enzymes suggested the name "cruzain." A copy of the gene was expressed in bacteria as an inactive, insoluble fusion polypeptide. Subsequently, the fusion protein was solubilized in urea and refolded to produce a polypeptide which processed autocatalytically to yield active, recombinant enzyme. This expression method generated recombinant protease of sufficient quality and quantity for crystallization. Diffraction quality crystals of recombinant cruzain, inactivated with a peptide inhibitor shown to block growth of the parasites in infected human cells, have been produced and characterized. The studies presented herein will provide insight into the mechanism of action and structure of cruzain and may enable the development of specific inhibitors for antiparasitic chemotherapy in the treatment of Chagas' disease.

Key Words: Cysteine protease; Parasite; Bacterial expression; Drug design; Crystallization

The work was supported by National Institute of Health Training Grant GM07175 and the University of California, San Francisco Biotechnology Program in Protein Crystallization (to A.E.E.), by World Health Organization TDR on Chagas' disease and funds from the American Heart Association (to J.H.M.), and by National Science Foundation Grants DMB8904956 and EET8807179 (to C.S.C.).

Reprint address: Ann E. Eakin, PhD, Department of Biochemistry, University of Puerto Rico School of Medicine, P.O. Box 365067, San Juan, PR 00936-5067.
THE TRADITIONAL METHOD of discovering drugs for the treatment of parasitic diseases has been limited to the random screening of natural or synthetic compounds. Therefore, the discovery of a useful drug has been dependent upon the more or less fortuitous identification of a successful reagent from among a vast assortment of available compounds. American trypanosomiasis, or Chagas’ disease, is a serious health problem in South and Central America, affecting more than 24 million individuals. Two drugs that are in use today for the treatment of this disease, Nifurtimox and Benznidazole, were discovered by traditional screening methods. These compounds kill the trypomastigote forms of the parasite in the bloodstream and, therefore, are usually effective in increasing a patient’s chance to survive the acute phase of the infection. These drugs do not eliminate all of the parasites from the blood, however, and have no effect on the intracellular parasites, which represent the replicative form in the human host. Also, these compounds have toxic side effects, making them unsatisfactory for the long-term treatment of Chagas’ disease and thus, do not prevent the subsequent development of a life-long disease.

An alternative approach to discovering drugs for the treatment of diseases caused by parasites involves the study of a potential target molecule and the subsequent design of compounds that will specifically interfere with the activity of this target. For a new drug to successfully kill a parasite and cure the disease, the target molecule must be pivotal for survival of the parasite. Also, to reduce the potentially toxic side effects of a drug, the compound must be designed to selectively inhibit the target molecule without affecting the enzymes of the host. Therefore, knowledge of the molecular structure and biochemical properties of the target enzyme should facilitate the design of new drugs for the treatment of the disease.

Proteases have been shown to be key factors in the pathogenicity of many parasitic diseases, either by inducing tissue damage and facilitating invasion or by empowering parasites to salvage metabolites from proteins of the host. Therefore, parasite proteases may be potential targets for the design of new drugs. Elucidating the biological roles of these proteases and attaining a thorough understanding of their biochemistry and structure could lead to the design of antiparasitic drugs which specifically block proteolytic action and kill the parasites.

To study the biochemical properties of a target protease and to determine its molecular structure, an abundant source of the active enzyme is required. The purification of a protease from parasites, however, can be very time consuming, difficult and dangerous, particularly in cases where no effective treatment for an accidental infection is available (e.g., Chagas’ disease). Furthermore, some parasites cannot be obtained in adequate quantities to permit the efficient purification of the protease. An alternative method for procuring large quantities of reagent-quality protease is to employ recombinant DNA techniques to clone a gene from the parasite and express the recombinant enzyme in laboratory organisms, such as Saccharomyces cerevisiae or Escherichia coli. In addition, the recombinant approach enables the future manipulation of the gene encoding the enzyme to test the effects of modifications on the structure or function of the enzyme. Also, this ability allows alteration of the protease to facilitate purification and crystallization or to enable the exploration of the role of amino acids that may interact with ligands or inhibitors.

For doctoral research one of the authors has used recombinant DNA techniques to study the major cysteine protease from the human parasite, Trypanosoma cruzi, with the goal of discovering or designing inhibitors that selectively target this enzyme and thus, might be used as drugs to treat Chagas’ disease. The cloning and recombinant expression of the gene encoding the cysteine protease of T. cruzi has provided en-
zyme of sufficient quality and quantity for biochemical analysis, as well as for the determination of the three-dimensional structure of the parasite enzyme. Already, one promising inhibitor has been identified and this work should provide the foundation for the discovery of additional lead compounds and for the design of drugs which may be useful for the treatment of Chagas’ disease.

MATERIALS AND METHODS

Gene Cloning

The polymerase chain reaction methods were as described (1). Southern blot, construction, and screening of a bacteriophage library of genomic DNA; and assembly and sequencing of a copy of the gene encoding the cysteine protease from *T. cruzi* ("cruzain") from a λ phage clone was as described (2). Sequence analysis and amino acid sequence alignments were performed using programs available in the Biocomputational Laboratory at the University of California, San Francisco.

Bacterial Expression of Recombinant Cruza

The development of a bacterial expression system was as described (2). The method of solubilization and refolding of the fusion protein was a modified procedure of that originally described by Marston et al. (3). These modifications and the scheme for purification of active, recombinant cruzain were as described (2). The partially purified fusion protein activated autocatalytically during incubation at 37°C. Details concerning the steps and proteolytic cleavages required for the recovery of active enzyme were examined by time course incubation experiments and aminoterminal sequencing of the resultant peptides (2). Kinetics and inhibition studies were performed using chromogenic substrates as described (2).

Mutagenesis and Mass Spectrometry

Site-directed mutagenesis experiments were performed using the PCR with an oligonucleotide primer carrying the desired base changes. Secondary primers were homologous to sequences located near convenient restriction sites within the original DNA sequence (4). Mass spectrometry was performed by Zhonghua Yu using electrospray and tandem MS equipment in the University of California, San Francisco Mass Spectrometry facility.

Crystallization and X-ray Diffraction Analysis

The crystallization of recombinant cruzain and preliminary X-ray diffraction analyses were as described previously (4). For gel analysis, single crystals were isolated from the hanging drops and dissolved in loading buffer. The refined structure of cruzain is currently being completed by Dr. Mary McGrath at the University of California using molecular replacement methods where the search model is derived from papain coordinates.

Biotin-labeled Inhibitor Studies

Biotin-labeled Phe-Ala fluoromethyl ketone inhibitor was purchased from Enzyme Systems Products. *T. cruzi* epimastigotes were lysed by repeated freeze/thaw steps in lysis buffer (10mM Tris, pH 8, 1mM EDTA, 0.1% SDS). The biotin-labeled inhibitor was added to 10 μM to both the cell lysates and to separate samples of active, recombinant cruzain before incubation for 30 minutes at 25°C to allow covalent inactivation of the enzyme by the inhibitor. The proteins were concentrated by TCA precipitation and resolved by SDS/polyacrylamide gel electrophoresis before transfer to nitrocellulose. Duplicate gels were run so that one blot could be developed by standard Western methods (5) using rabbit antiserum raised against the recombinant cruzain and the other blot
could be developed using an avidin/ biotin-horse radish peroxidase detection kit as per the manufacturer's instructions. The avidin/ biotin detection system is specific for the biotin label on the inhibitor which was covalently bound to the protease in the cell lysate or to the purified recombinant enzyme.

RESULTS

Gene Cloning

Homologous proteins, which have evolved from a common ancestral gene, often share conserved motifs in their amino acid sequences. Therefore, isolation of genes encoding homologous proteins can be accomplished if functionally and/or structurally conserved regions of the proteins can be identified. Based upon conserved amino acid sequences in the regions flanking the active site cysteine-25 and asparagine-175 of aligned cysteine proteases, two sets of degenerate oligonucleotide primers were designed for the amplification of cysteine protease gene fragments (Figure 1). These primers were used in PCR to amplify fragments of cysteine protease genes from the genomic DNA of Trypanosoma cruzi, T. brucei, and Entamoeba histolytica (1). In addition to the partial sequence data obtained, these fragments provided homologous probes with which the genes encoding these cysteine proteases could be isolated from bacteriophage libraries (2,6,7).

The PCR fragment from T. cruzi was used to probe Southern blots of partially digested, immobilized genomic DNA from T. cruzi (2). The blot (Figure 2) reveals a ladder of evenly spaced (~2 kb) bands in the lanes with partially digested DNA, indicating that the protease is encoded by a tandem repeat of at least six copies of the gene. The completely digested sample (lane 1) shows that the size of a single copy of the gene is approximately 1.85 kb. This type of organization of tandemly repeated genes is common to many trypanosomal genes and the entire tandem array is likely to be transcribed as a single polycistronic message that is subsequently cleaved into single-copy mRNAs by trans-splicing (8).

Northern blot analysis, where total RNA from each of the three developmental stages of T. cruzi was immobilized on nitrocellulose and hybridized with the radiolabeled PCR-amplified gene fragment, indicated that the gene is constitutively expressed in all developmental stages (data not shown; 2). This lack of developmental regulation suggests that the protease may be required throughout the life cycle of the parasite and is consistent with both immunologic (9,10) and activity studies (11) of the endogenous protease from the parasites.

A library of T. cruzi genomic DNA constructed in λ EMBL phage was screened with the PCR-amplified cysteine protease gene fragment. One phage clone was isolated which contained the entire tandem array of the six cysteine protease genes. One copy of the gene was subcloned into a plasmid vector and sequenced (2). This sequence is shown in Figure 3 along with the deduced amino acid sequence of the proposed coding region. The entire coding region predicts a preproform of the protein with a mass of 49,800 Da. Both the DNA and deduced amino acid sequences contain regions which are identical to PCR-amplified gene fragments reported previously (1,12) and to the several reported peptide sequences of the endogenous protease that was purified from T. cruzi parasites (9,13,14).

Sequence Comparisons

Amino acid alignments revealed sequence identities of 59.3%, 42.2%, and 32% for regions of overlap between the deduced amino acid sequence of the T. cruzi protease gene versus the T. brucei, mouse cathepsin L and papain sequences, respectively (2). Because of the high degree of sequence similarity with members of the papain family of cysteine proteases, the
Cysteine Protease as a Drug Target

CYS PRIMER

5' ACA GAA TTC CAR GGI CAR TGY GGI TCI/AGY TGY TGG 3'
ECORI GLN GLY GLN CYS GLY SER CYS TRP

Nterm-------------------C-------------------H-------------------N----------------Cterm
25                   159                   175

PAPAIN  QGCGSGCW........................................VDHAVAAGV....YILKNSW
CHICK CATH L QGCGSGCW........................................LDHGYLNVG....YWIVKNPS
MOUSE CYS  QGCGSGCW........................................LDHGYLNVG....YWIVKNSW
RAT CATH H  QGCGSGCW........................................VDHAVAAGV....YWIVKNSW
RAT CATH B  QGCGSGCW........................................GDHAILLG.....YWIVKNSW
HUMAN CATH B  QGCGSGCW........................................GDHAILLG.....YWIVKNSW
SLIME MOLD  QGCGSGCW........................................LDHGYLNVG....YWIVKNSW

ASN PRIMER

3' ATR ACC TAR CAY TTY TTR AGI/TCR ACC TTC GAA ATT 5'
TYR TRP ILE VAL LYS ASN SER TRP HINdIII

R = PURINE
Y = PYRIMIDINE
I = INOSINE

FIGURE 1. Design of consensus oligonucleotide primers used to amplify cysteine protease gene fragments with the polymerase chain reaction.

A sequence alignment of cysteine proteases in the regions of the active site amino acids is shown in the center of the figure. Above this alignment is a schematic representation of the linear positioning of the active site amino acids (in bold) in the protein sequences. Based upon conserved amino acid sequences in the regions flanking the active site cysteine-25 and asparagine-175 residues of the aligned cysteine proteases, two sets of degenerate oligonucleotides were designed for the amplification of any cysteine protease gene using the polymerase chain reaction. The sequences of the Cys and Asn primers are shown above and below the alignment, respectively. The combination of bases synthesized in degenerate positions are symbolized as R for purines, Y for pyrimidines, and I for inosine, which will form hydrogen bonds with any of the four bases. At the 5' end of each oligo, recognition sites for the restriction endonucleases EcoRI and HindIII were added to allow directional subcloning of the amplified fragments.

trivial name, "cruzain," was chosen for the protease from T. cruzi (2). Interestingly, the trypanosomal sequences extend 114 amino acids (T. brucei) and 130 amino acids (T. cruzi) beyond the carboxy-terminus of related cysteine proteases. Similar domains have been described in sequences encoding cysteine proteases of two plants (15,16). The biological function of these unusual domains is unknown.

Expression of Recombinant Cruzain

A diagram of the plasmid vector that was used to express the T. cruzi cysteine protease in E. coli is shown in Figure 4. The proform of cruzain (beginning with amino acid Cys-104) was expressed as a fusion with 40 amino acids of the bacterial Che Y protein (18). An enteropeptidase site was included at the junction of this fusion to enable its removal from the amino terminus of the T. cruzi protease.

The Che Y-cruzain polypeptide was expressed in bacteria in an insoluble form (Figure 4). This insoluble fusion protein was solubilized in urea, refolded, and partially purified (2). Incubation of the purified fusion protein at 37°C in 0.1M sodium acetate, pH 5.5, 5mM dithiothreitol (DTT) resulted in an increase in activity that could be monitored by the release of the fluorogenic leaving group, amino methyl coumarin (AMC) from the peptide
substrate Z-Phe-Arg-AMC (Z = benzyl-oxycarbonyl). Maximal activity was attained in four hours and remained stable for up to a total of seven hours incubation (Figure 5, graph). The incubation permitted autoactivation and a concomitant reduction in the size of the fusion protein (Figure 5, gel). A duplicate sample, in which an irreversible inhibitor of cysteine proteases (Z-Phe-Arg-fluoromethyl ketone) was added to the fusion protein, showed no increase in activity during the incubation period (Figure 5, graph, dark circle). In the inhibited sample, no proteolytic processing of the fusion protein was observed (Figure 5 gel, right lane).

The proteolytic events required for the processing and activation of the fusion protein were examined in time-course experiments (2). The first cleavage event to occur was the removal of the C-terminal domain (Figure 5). This cleavage event occurred early in processing, apparently while the polyprotein was bound to DEAE-Sepharose and before incubation at 37°C. Subsequent processing events occurred during incubation at 37°C. These include cleavages at two sites within the pro domain of the protease to release the
FIGURE 3. DNA sequence of one copy of the cysteine protease gene repeat showing the deduced amino acid sequence for the encoded enzyme.

The DNA sequence shown begins and ends arbitrarily at the PvuII sites which occur in a noncoding region of the repeat. Within the noncoding sequence preceding the gene, the putative spliced leader recognition signal, including a polypyridine stretch followed by an AG splice acceptor sequence, is indicated in bold italics. The putative start methionine and amino acid sequence are shown above the DNA sequence. The numbering of the amino acids is based on the papain numbering system. The active site amino acids are identified with boxes. Three possible N-linked glycosylation sites are marked with a star above the sequence. Arrow "a" denotes the predicted site of signal peptidase cleavage. This results in a proform of the protease beginning at Cys-104. Arrow "b" denotes the location at which the proregion of the protease is cleaved to yield the mature form of the enzyme as identified by the published N-terminal sequence (14, 9). Arrow "c" indicates the site of proteolysis where the C-terminal extension is cleaved (13, 20). The arrows marked as "x" indicate the sites of autocatalysis that remove the prodomain in the recombinant fusion protein activation process (refer to Figure 5). The stop codon terminates the protease at Leu 342. A putative poly-A addition signal is in bold and underlined in the noncoding region following the gene. Reprinted from reference 2.
FIGURE 4. The expression of cruzain in bacteria.
The bacterial expression vector is shown in the diagram above the gel. The protein domains are labeled within the diagram, and the corresponding amino acid numbers of the Che Y protein (in plain type) and cruzain (in bold) are indicated below the diagram. The amino acid sequence at the junction between the Che Y protein (in plain type), the unique enteropeptidase cleavage recognition site (in italics), and the beginning of cruzain (in bold) are indicated above the diagram at the Che Y/protease junction. The amino acid sequences at the predicted cleavage sites which remove the proragon and C-terminal extension domain of the protease are also shown above the diagram (in bold). The arrows above these sequences indicate the predicted positions for autoproteolytic cleavage. The side chains of the amino acids of the catalytic triad are displayed below the diagram at their approximate positions within the protease core. The remainder of the expression plasmid (not shown) contains the lac promoter and the β-lactamase gene, which confers resistance to ampicillin. Proteins from two strains of E. coli (dg06 and DH5α) hosting the expression vector (pCHeYTC) or a nonexpression plasmid that only provides resistance to ampicillin (pBS) were divided into soluble ("S") and insoluble ("I") fractions and run on a 12% polyacrylamide gel to monitor the expression of the fusion protein. Sizes of the molecular weight standards are indicated to the left of the gel, and a bold arrow on the right indicates the position of the fusion protein (60,000 Da). Reprinted from reference 2.
FIGURE 5. Autoproteolysis and activation of the purified fusion protein.
The fusion protein was incubated for several hours at 37°C in 0.1 M sodium acetate, pH 5.5 and 5 mM DTT. Samples were removed hourly, the activity was tested, and the proteins were acid precipitated. The activity, as monitored by a 60-second fluorometric assay with the peptide substrate Z-Phe-Arg-AMC, was plotted versus the corresponding incubation time, 0-7 hours (graph). The point marked with a solid circle is the activity of a sample which was incubated seven hours in the presence of the irreversible fluoromethyl ketone inhibitor (Z-Phe-Arg-FMK) at a concentration of 10 μM. Below the activity graph is a coomassie-stained polyacrylamide gel of the acid precipitated samples loaded in order of their incubation times. The right-most lane (7 hr + 1) is the sample which was incubated seven hours in the presence of the inhibitor. The sizes of the molecular weight standards are indicated to the left of the gel. To the right are the amino-terminal sequences of the corresponding proteins on the gel. The 38 kDa protein band provided two sequences—a major sequence (comprising 90% of the signal) and a minor sequence ((10% of the signal). The 28 kDa protein yielded a single sequence. Below each N-terminal sequence is amino acid numbering in the protein sequence (refer to Figure 3). Reprinted from reference 2.

Che Y fusion and the subsequent cleavage of the remaining pro domain to yield a fully active protease with an amino terminal sequence identical to that of the endogenous protease purified from the parasites (2,9,14). This activation by autocatalysis obviated the need for the enteropeptidase cleavage site to remove the Che Y fusion.

Mutagenesis

Site-specific mutagenesis was used to examine the proteolytic processing events during activation of the fusion protein (4). Expression of an inactive mutant of cruzain, in which the active site cysteine was replaced with an alanine (cruzan-
C25A], revealed that although the initial processing event (removal of the C-terminus) was not dependent on the activity of cruzain, subsequent cleavage events did require autocatalysis. Cleavage of the C-terminal domain could thus be due to a bacterial protease or some other, as yet unidentified, hydrolytic event (19). Initially, removal of the C-terminal domain was assumed to be an autoproteolytic event since the protease purified from T. cruzi epimastigotes has been shown by others to autodegrade, releasing a protein fragment with an N-terminal sequence that marked the cleavage site at Gly212 (13). This position is at the beginning of the C-terminal domain as predicted by alignment with papain. However, the molecular mass of the fully processed recombinant cruzain, as determined by electrospray mass spectrometry, was 23,534 ± 6.6 daltons. By relating this mass to the sequence of the gene, the cleavage site of the recombinant protein was assigned to Thr221, nine amino acids further into the C-terminal domain than that observed for the native enzyme. To mimic the native protease, a truncated form of the gene was created by insertion of a stop codon after Gly212 (4). This truncated variant (cruzain-Δc) expressed, refolded, and activated comparably to the recombinant protein from the wild type gene (cruzain). Mass spectrometry of this variant indicated a molecular mass of 22,700 ± 5.69 daltons. The difference between the relative molecular masses of cruzain and cruzain-Δc is 834 daltons and is consistent with the predicted nine amino acid difference at the C-terminus.

Activity and Inhibition Analyses

The kinetic parameters for the recombinant protease were comparable to the published kinetic data for the native protease purified from the parasites (2). The endogenous enzyme has been shown to be highly glycosylated (20,21); therefore, the difference in structure of the glycosylated native protein and the recombinant protease may contribute to the observed differences in kinetic parameters. The kinetic parameters determined for a Z-Phe-Arg-AMC substrate indicate that the activity of the truncated mutant (cruzain-Δc) is within two-fold that of cruzain, despite the nine amino acid difference at the C-terminus [cruzain: $K_m = 0.96 \mu M, k_{cat} = 269.5 \text{min}^{-1}, k_{cat}/K_m = 279.7 \mu M^{-1} \text{min}^{-1}$; cruzain-Δc: $K_m = 0.80 \mu M, k_{cat} = 129.7 \text{min}^{-1}, k_{cat}/K_m = 163.3 \mu M^{-1} \text{min}^{-1}$ (4)].

Inhibitor studies support the classification of the recombinant protease as belonging to the papain family of cysteine proteases (2). The recombinant protease demonstrated sensitivity to leupeptin and E64, and the activity persisted in the presence of PMSF, pepstatin, or 1,10-phenanthroline. Also, the increased activity in the presence of DTT supports the classification as a thiol-dependent protease. The fluoromethyl ketone substrate analogues, Z-Phe-Ala and Z-Phe-Arg-FMK, were the most potent inhibitors tested. Tandem mass spectrometry demonstrated that these substrate analogues inactivate the recombinant protease by covalently alkylating the active site thiol (Cys25) by a sulfide-methylene bond with concomitant release of the fluoride group of the inhibitor (Yu, Eakin, Jones, Craik, Burlingame, unpublished data, 1993). Similar results have been observed in crystal structures of papain with peptide chloromethyl ketone inhibitors (22).

Crystallography and X-ray Diffraction Analyses

Protein of sufficient quality for crystallization was obtained by anion exchange chromatography of the refolded, activated recombinant enzyme (4). The protease was inactivated by the addition of an irreversible inhibitor (either Z-Phe-Arg or Z-Phe-Ala fluoromethyl ketone). The purified, inactivated recombinant enzyme was concentrated to 10–20 mg/ml in water by diafiltration and stored at 4 ºC.
Cysteine protease as a drug target

Crystallization trials, using hanging drop vapor diffusion with various precipitants, salts, and buffer pH values, were carried out for both cruzain and cruzain-Δc. A highly birefringent and twinned crystal formed after six months in a sodium citrate drop with cruzain. Tests showed that the crystal was composed of cruzain and diffracted to at least 2.8Å resolution (4). Since the initial crystal nucleation required several months, a protocol (23), using the sodium citrate-derived crystal to streak-seed equilibrated hanging drops of both cruzain and cruzain-Δc, was employed to refine optimal crystallization conditions for the formation of larger crystals. Three-dimensional, highly birefringent and twinned crystals were obtained in many variations of the original crystallization conditions with each protein, but the crystals of cruzain-Δc were consistently more three-dimensional and less twinned than those of cruzain. Therefore, the truncated mutant which had been inactivated with the fluoromethyl ketone inhibitor (Z-Phe-Ala-FMK) was used for subsequent crystallization studies. Large, uniform crystals were obtained at 4°C using a dilutional microseeding protocol with the best crystals being obtained at a 10³ dilution of the microseeds. Figure 6 shows a coomassie-stained SDS/polyacrylamide gel loaded with crystals of cruzain (lane 2) and cruzain-Δc (lane 3) grown by these methods.

Crystals of cruzain-Δc were characterized by precession photography analysis and the results were corroborated by data collection on a Siemens area detector. Figure 7 shows a 110° screened precession photograph demonstrating the repeating absences of reflections with indices 0k0, where k = 2n + 1. The periodicity of these absences indicates that the monoclinic crystal has the symmetry of space group P2₁. Cell constants were found to be a = 45.4Å, b = 51.0Å, c = 45.7Å, and β = 116.1°. The cell volume and molecular mass of cruzain-Δc are consistent with there being one molecule in the asymmetric unit. This assumes an average crystal density and results in a V_m of 2.1 and an estimated 41% solvent content. A 2.4Å X-ray data set has been collected for cruzain-Δc. Because of the high degree of sequence similarity to papain, the structure of cruzain is being solved by molecular replacement where the search model is derived from papain coordinates (17). The small, low symmetry unit cell and the lack of noncrystallographic symmetry has greatly facilitated the structure determination using molecular replacement.

Labeled Inhibitor Studies

Using a biotin-labeled fluoromethyl ketone inhibitor (Biotin-Phe-Ala-FMK; Figure 8), the cysteine protease was specifically labeled and identified in lysates of epimastigotes (Figure 9). This inhibitor also labeled the active, recombinant cysteine protease (Figure 9). The same protein bands were identified on a separate blot using antibodies raised in rabbits against the recombinant protease (Figure 9).

DISCUSSION

The discovery of new drugs for the treatment of Chagas' disease would be very desirable. There is currently no satisfactory treatment for infection by T. cruzi, making Chagas' disease a serious health problem that is the leading cause of premature death due to heart failure in many countries in South America. The work presented here describes the development of a system for drug discovery that employs recombinant DNA manipulations, homologous expression in bacteria, and crystallography of an enzyme that may be a potential target for antiparasitic drugs.

In this work, a cysteine protease was selected as a potential target for the development of new drugs. The cloning of genes encoding cysteine proteases was expedited by the design and use of degenerate oligonucleotide primers (Figure 1) in the polymerase chain reaction to amplify frag-
FIGURE 6. SDS/PAGE analysis of crystals of cruzain and cruzain-Δc.
Crystals of cruzain and cruzain-Δc grown from dilutional seeding using the sodium citrate-
derived seed stock were analyzed on a coomassie-stained, 12% polyacrylamide SDS gel. The sizes of the molecular weight standards (std) are indicated to the left of the gel.

ments of genomic DNA. These gene fragments were subsequently used as homologous probes to isolate the full-length genes from bacteriophage libraries. The gene encoding the cysteine protease of the protozoan parasite, T. cruzi, was cloned using this method and was shown to be organized as a tandem repeat of six copies within the parasite genome (Figure 2). This type of genomic organization is common to trypanosomes and may provide a mechanism for the amplification of gene expression by the parasite in response to harsh environmental changes encountered in a complex life cycle.

Amino acid sequence alignments indicate that the protease from T. cruzi belongs to the papain family of cysteine proteases, justifying the name “cruzain” for the enzyme. Heterologous expression of the gene proved to be very difficult, but a bacterial system was developed that produced the recombinant protease as an insoluble, inactive fusion protein (Figure 4). Solubilization, in vitro folding, and activation conditions were established that
FIGURE 7. Precession photograph of a cruzain-Δc crystal.
A 100° screened, zero-layer precession photograph of a cruzain-Δc crystal is shown. The size of two dimensions of the unit cell can be observed in the spacing of the zero-layer reflections (those spots nearest to the shadow of the beam stop). The systematic absences (every other spot) observed in these spots describe the symmetry within the unit cell of the crystal. The occurrence of focused spots at the limits of the X-ray film (outer edges of ring) indicates a resolution of X-ray diffraction to at least 2.5Å.

FIGURE 8. Biotin-labeled fluoromethyl ketone inhibitor.
The structure of the biotin-labeled fluoromethyl ketone inhibitor (biotin-Phe-Ala-FMK) used to detect the cysteine protease in parasite lysates is depicted. The biotin at the amino-terminus of the irreversible, substrate-analogue inhibitor does not interfere with the binding nor the inactivation of the protease. This inhibitor can be used to label the protease within cell lysates and subsequently allow the visualization of the labeled proteins on Western-type blots using biotin/avidin detection reagents (see Figure 9).
yield active cruzain at sufficient levels to permit extensive biochemical analyses and crystallization of the recombinant enzyme. The development of this system provided the necessary reagents with which enzyme structure-based drug design can be conducted.

A major criteria for the success of enzyme structure-based drug design is the identification of a target that is pivotal for survival of the pathogen. The precise biological function of the cysteine protease from T. cruzi, and its involvement in the parasitic infection, remains obscure. Sev-
eral experiments, however, suggest that the activity of a cysteine protease may be required for the maintenance of the parasite's life cycle. The addition of cysteine protease inhibitors (Z-Phe-Ala-fluoromethyl ketone or Z-Phe-Arg-fluoromethyl ketone) to media containing mammalian cells and infective trypomastigotes results in a dramatically reduced number of intracellular amastigotes as compared to control cells without inhibitor (24). These inhibitors do not block cell invasion by the parasites but seem to disrupt the transformation process by which the parasites differentiate into the intracellular, replicative form. Also, others have shown that related cysteine protease inhibitors can block differentiation in the parasite life cycle (25,26).

Experiments with a biotin-labeled fluoromethyl ketone (Figure 8) demonstrate that the cysteine protease of T. cruzi is specifically labeled in cell lysates by this inhibitor with no observable cross-reactivity to other parasite proteins (Figure 9). Antibodies specific to recombinant cruzain recognize the identical proteins in a duplicate blot (Figure 9). Unfortunately, the colorimetric biotin detection system is not sensitive enough to visualize the label when intact, infected mammalian cells are exposed to the inhibitor. Future studies are in progress using $^{125}$I-labeled avidin to increase the sensitivity of the inhibitor detection methods. An advantage of using a biotin-labeled inhibitor is that future techniques, including fluorescent and electron microscopy, might be used to localize the protease within the parasites using available avidin/biotin detection systems. These types of experiments could provide firm evidence to support the results of earlier studies indicating that cruzain is a pivotal metabolic enzyme that is specifically targeted by these inhibitors, resulting in the interruption of the parasite life cycle.

To design inhibitors with greater selectivity of the protease from the parasite, a thorough understanding of the biochemistry and the three-dimensional structure of the enzyme is essential. Herein, a bacterial expression and purification system that yields high levels of recombinant forms of cruzain has been described. This methodology provided the reagents that set the stage for several avenues of future studies. For example, the system was and can continue to be used to generate and characterize variants of the enzyme to test the primary determinants influencing the structure and/or function of this biomedically important member of the thiol protease family. Also, the cellular location of the protease within the parasite may be examined using the specific antibodies raised against the recombinant enzyme. By combining this information with that obtained using the biotin-labeled inhibitor, the protease can be identified as cruzain both immunologically and by activity. Also, the purified recombinant protease might be useful as an antigen in the development of a badly needed kit for the diagnosis of Chagas' disease.

The forthcoming three-dimensional structure of cruzain has a peptide inhibitor covalently bound to the active site of the enzyme. Therefore, the structure will provide important details regarding the amino acids in the binding pocket of the protease that are in contact with the inhibitor. This information will be useful for designing new peptide inhibitors with enhanced selectivity for the parasite protease versus related mammalian thiol proteases. Also, the structural information can be analyzed using programs, such as “DOCK” (27,28), for the identification of nonpeptide, small-molecule inhibitors that could subsequently be tested as lead compounds that inhibit the cysteine protease and, therefore, might be developed into drugs. These lead compounds could be further refined using information provided by the structural analysis of recombinant cruzain co-crystallized with the leads. Such reiterative enzyme structure-based drug design experiments are now possible using the recombinant expression system described in this dissertation summary.Hopefully,
these studies will eventually lead to the discovery of new drugs for the treatment of Chagas' disease.

REFERENCES


26. Meirrelles MNL, Juliano L, Carmona E, Silva
Cysteine Protease as a Drug Target

