Production of Crystallizable Cruzain, the Major Cysteine Protease from Trypanosoma cruzi*

(Received for publication, October 8, 1992)

Ann E. Eakin‡§, Mary E. McGrath¶, James H. McKerrow§, Robert J. Fletterick||, and Charles S. Craik‡

From the Departments of ‡Pharmaceutical Chemistry, ¶Pathology, and ||Biochemistry, University of California, San Francisco, California 94143-0446

The major cysteine protease of Trypanosoma cruzi, cruzain, has been previously expressed in Escherichia coli as a fusion polypeptide. The proteolytic processing events required to obtain active, mature cruzain from the recombinant expression system have been characterized using mutational analysis of the cloned gene. An inactive variant of cruzain (cruzain-C25A) revealed that the proteolytic cleavage of the COOH-terminal domain from the recombinant cruzain is independent of cruzain activity. This cleavage event, presumably performed by another protease, was reduced, although not completely eliminated, in a variant in which the cleavage recognition site was altered (cruzain-E2219P). To obtain a homogeneous COOH terminus of the recombinant enzyme, a truncated form of cruzain (cruzain-ΔC) was engineered by insertion of a stop codon in the gene at a site corresponding to auto-proteolysis observed with the native enzyme, purified from epimastigotes. Diffraction quality crystals of the recombinant cruzain (cruzain) and the truncated variant (cruzain-ΔC) have been produced and characterized. Cruzain and cruzain-ΔC were cocrystallized with the peptide fluoromethyl ketone (FMK) inhibitors, Z-Phe-Arg-FMK and Z-Phe-Ala-FMK, respectively, (where Z is benzoxycarbonyl). The crystals are monoclinic, space group P2₁, with a = 45.5 Å, b = 61.6 Å, c = 45.7 Å, and β = 116.1°. One cruzain molecule is present in the asymmetric unit. The crystallographic data reveal that the high resolution structure determination is feasible. This system will facilitate the three-dimensional structure determinations and biochemical analyses of cruzain and cruzain variants.

A potential strategy to develop new chemotherapy for the parasitic diseases that are major health problems in underdeveloped parts of the world is inhibition of crucial parasite enzymes. Parasite proteases are attractive targets, because they are involved in parasite survival, replication, and the production of disease (McKerrow, 1989). A protozoan parasite, Trypanosoma cruzi, is the etiologic agent of Chagas' disease, which affects more than 24 million people in Latin America. The gene for the major cysteine protease from T. cruzi, known as cruzain, has been cloned and sequenced and found to be a member of the papain family of sulfhydryl proteases (Eakin et al., 1992). Several studies indicate that this cysteine protease may be pivotal for the parasite's development and survival within the host. The addition of cysteine protease inhibitors to a cell culture model of the parasite life cycle blocks the development of T. cruzi and disrupts the replicative cycle (Bonaldo et al., 1991; Meirelles et al., 1992; Harth et al., in press).

An accurate three-dimensional structure of the target protease, cruzain, would greatly facilitate design efforts for specific drugs. Toward this end, a bacterial expression system was designed, which produces reagent levels of the recombinant enzyme. Cruzain is initially expressed in this system as an inactive, insoluble fusion protein that is easily isolated from the bacterial lysate and successfully refolded and processed to yield active cruzain (Eakin et al., 1992). We report herein the characterization of the processing events required to produce active enzyme from the recombinant expression system as well as the crystallization and preliminary X-ray diffraction data of a variant enzyme inhibited with a covalently bound peptide fluoromethyl ketone.

RESULTS AND DISCUSSION

Analysis of the Processing Events Required to Yield Active Cruzain—Cruzain is initially expressed in bacteria as an insoluble fusion polypeptide with the first 40 amino acids of the Escherichia coli CheY protein. This insoluble fusion protein can be solubilized in urea, refolded, and partially purified as previously described (Eakin et al., 1992). Active cruzain can then be recovered from the fusion by incubation in pH 5.5 buffer with dithiothreitol for several hours at 37 °C. This incubation period permits the proteolytic processing events that remove the CheY fusion, the prodomain, and the COOH-terminal domain of the protease. The gene sequence for cruzain indicates a COOH-terminal domain of 130 amino acids, which is not common to other members of the papain family of cysteine proteases. A similar domain was noted in the sequence of a cysteine protease from the related parasite Trypanosoma brucei (Pamer et al., 1990), as well as sequences from two plants, tomato and rice (Schaffer and Fischer, 1988; Watanabe et al., 1991).

The sequence of the processing events that occur during purification and activation of the fusion protein is dia-
The removal of the COOH-terminal domain was assumed to be an autoproteolytic event, as are the other observed processing events, because a similar cleavage process was observed for the native protease purified from T. cruzi epimastigotes (Hellman et al., 1991). The native protease was shown to partially degrade over time by specifically releasing a protein fragment with an NH₂-terminal sequence that marked the cleavage site at Gly²¹². This position is the beginning of the COOH-terminal domain as predicted by alignment with papain. The accurate molecular mass of the fully processed recombinant cruzain was determined by electrospray mass spectrometry to be 23,534 ± 6.6 daltons. By relating this mass to the sequence of the gene, the recombinant cleavage domain was cruzain-C₂⁵₄ in which Glu²¹⁹ is replaced with a proline (cruzain-E₂₁⁹P) in an attempt to prevent the proteolytic removal of the COOH-terminal domain. Cruzain-E₂₁⁹P is inhibited with Z-Phe-Arg-FMK. The lanes marked "std" are molecular mass standards whose sizes are indicated to the left and right of the gel.

The removal of the COOH-terminal domain was assumed to be an autoproteolytic event, as are the other observed processing events, because a similar cleavage process was observed for the native protease purified from T. cruzi epimastigotes (Hellman et al., 1991). The native protease was shown to partially degrade over time by specifically releasing a protein fragment with an NH₂-terminal sequence that marked the cleavage site at Gly²¹². This position is the beginning of the COOH-terminal domain as predicted by alignment with papain. The accurate molecular mass of the fully processed recombinant cruzain was determined by electrospray mass spectrometry to be 23,534 ± 6.6 daltons. By relating this mass to the sequence of the gene, the recombinant cleavage domain was cruzain-C₂⁵₄ in which Glu²¹⁹ is replaced with a proline (cruzain-E₂₁⁹P). The resultant protein expressed and refolded similarly to the wild type protein. However, during the activation process, the protease accumulated at an apparent molecular weight that indicated the COOH-terminal domain remained intact. This form of the protein was stable for at least 10 h at 45 °C. After prolonged incubation (>16 h), partial processing resulted in a shift of the 36-kDa band on SDS-PAGE to one which comi-

The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; FMK, fluoromethyl ketone; Z, benzoyloxy carbonyl.

1 The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; FMK, fluoromethyl ketone; Z, benzoyloxy carbonyl.

The removal of the COOH-terminal domain was assumed to be an autoproteolytic event, as are the other observed processing events, because a similar cleavage process was observed for the native protease purified from T. cruzi epimastigotes (Hellman et al., 1991). The native protease was shown to partially degrade over time by specifically releasing a protein fragment with an NH₂-terminal sequence that marked the cleavage site at Gly²¹². This position is the beginning of the COOH-terminal domain as predicted by alignment with papain. The accurate molecular mass of the fully processed recombinant cruzain was determined by electrospray mass spectrometry to be 23,534 ± 6.6 daltons. By relating this mass to the sequence of the gene, the recombinant cleavage domain was cruzain-C₂⁵₄ in which Glu²¹⁹ is replaced with a proline (cruzain-E₂₁⁹P). The resultant protein expressed and refolded similarly to the wild type protein. However, during the activation process, the protease accumulated at an apparent molecular weight that indicated the COOH-terminal domain remained intact. This form of the protein was stable for at least 10 h at 45 °C. After prolonged incubation (>16 h), partial processing resulted in a shift of the 36-kDa band on SDS-PAGE to one which comi-

1 The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; FMK, fluoromethyl ketone; Z, benzoyloxy carbonyl.

The removal of the COOH-terminal domain was assumed to be an autoproteolytic event, as are the other observed processing events, because a similar cleavage process was observed for the native protease purified from T. cruzi epimastigotes (Hellman et al., 1991). The native protease was shown to partially degrade over time by specifically releasing a protein fragment with an NH₂-terminal sequence that marked the cleavage site at Gly²¹². This position is the beginning of the COOH-terminal domain as predicted by alignment with papain. The accurate molecular mass of the fully processed recombinant cruzain was determined by electrospray mass spectrometry to be 23,534 ± 6.6 daltons. By relating this mass to the sequence of the gene, the recombinant cleavage domain was cruzain-C₂⁵₄ in which Glu²¹⁹ is replaced with a proline (cruzain-E₂₁⁹P). The resultant protein expressed and refolded similarly to the wild type protein. However, during the activation process, the protease accumulated at an apparent molecular weight that indicated the COOH-terminal domain remained intact. This form of the protein was stable for at least 10 h at 45 °C. After prolonged incubation (>16 h), partial processing resulted in a shift of the 36-kDa band on SDS-PAGE to one which comi-

1 The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; FMK, fluoromethyl ketone; Z, benzoyloxy carbonyl.
grated with the original activated cruzain (Fig. 2, lanes D and E), suggesting that the cleavage site had been sufficiently altered to delay the proteolysis but not prevent it. Interestingly, the gel pattern suggests that the protease did not cleave at Gly192 and where the native enzyme was seen to be digested. Additionally, fully activated cruzain will not cleave itself at Gly192 even after 20 h of incubation at 50 °C. Glycosylation of the native enzyme may alter the structure or accessibility of the protein and may account for processing differences. Alternatively, a protease in the native enzyme preparation may be responsible for the observed cleavage pattern.

To minimize the COOH-terminal heterogeneity, a truncated form of the gene was created by insertion of a stop codon after Gly219. This truncated variant (cruzain-Δc) expressed, refolded, and activated comparably to the 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 weights of cruzain and cruzain-Δc is 834, reflecting the nine-amino acid difference at the COOH terminus.

Kinetic parameters of the recombinant enzymes on a Z-Phe-Arg-aminomethyl coumarin substrate indicate that the activity of cruzain-Δc is within 2-fold of that of cruzain, despite the nine-amino acid difference at the COOH terminus (cruzain: $K_m = 0.96 \mu M$, $k_{cat} = 269.5 \text{ min}^{-1}$; cruzain-Δc: $K_m = 279.7 \mu M^{-1}$, $k_{cat} = 163.3 \text{ min}^{-1}$). In contrast to the cruzain-Δc enzyme, a variant in which the prodomain of cruzain was deleted (cruzain-Δp) was inactive, and the fusion protein was unstable after refolding and precipitated out of solution during cold storage. This observation is similar to the recombinant cathepsin L expression system where deletion of the prodomain also resulted in inactive enzyme (Smith and Gottesman, 1989). Therefore, the prodomains of these cysteine proteases seem to be involved in the proper folding or stabilization of the enzymes as demonstrated for the bacterial proteases subtilisin (Zhu et al., 1989) and α-lytic protease (Silen et al., 1989).

All mutageneses described in this report were performed using the polymerase chain reaction with a mutagenic oligonucleotide carrying the desired genetic variation and an opposing oligonucleotide located near convenient restriction sites within the original sequence (Vallette et al., 1989). The plasmid encoding the cloned, wild type cruzain gene was linearized using a restriction endonuclease that cuts outside the coding sequence and was used as the template DNA in the polymerase chain reactions. The mutagenic oligonucleotides were designed to include both the mutation and a nearby restriction site already present in the cloned gene. The opposing oligonucleotides were chosen based upon the proximity to a restriction endonuclease recognition site and the distance from the mutagenic oligonucleotide. The polymerase chain reaction amplifications resulted in the production of fragments of the gene, containing the desired mutations that were between 300 and 600 base pairs in length with flanking restriction sites. These fragments were easily subcloned into the expression vector to replace the wild type sequence with the altered sequence. Each mutation was confirmed by DNA sequencing analysis of both strands and checked for inadvertent mutations.

To obtain protein of sufficient purity for crystallization, the degradation peptides resulting from activation needed to be removed. The mature protease was first inactivated by the addition of an irreversible inhibitor (either Z-Phe-Arg or Z-Phe-Ala fluoromethyl ketone), which covalently modifies cysteine proteases at their active sites (Rasnick, 1985). Because the PI of activated cruzain is 3.5 (as determined by isoelectric focusing gel electrophoresis, data not shown), inactivated cruzain and cruzain-Δc were purified to homogeneity using an anion exchange column (Pharmacia LKB Biotechnology Inc. MonoQ 10/10), at pH 5.8, with a sodium chloride elution gradient. This purified material was concentrated to 10–20 mg/ml in water by diafiltration and stored at 4 °C to be used for crystallization procedures.

**Crystallization and Preliminary X-ray Diffraction Data**—
Crystallization trials using hanging drop vapor diffusion and various precipitants, salts, and pH values were carried out for both cruzain and cruzain-Δc. A crystal formed after 6 months as a highly birefringent and twinned crystalline solid mass in a sodium citrate drop with cruzain. Also after this extended period of time, thin triangular crystalline plates were observed from ammonium phosphate with cruzain. It has not been possible to crystallize cruzain-Δc using these conditions.

The composition (salt or protein) of the sodium citrate-derived crystal was determined by testing the crystal hardness, examining the x-ray diffraction pattern generated by a fragment of the crystal, and SDS-PAGE and Western blot analyses of the diffracted crystal. The tests showed that the crystal was comprised of cruzain, was highly twinned, and diffracted to at least a 2.8-Å resolution. Since crystal nucleation appeared to be problematic, a streak-seeding protocol (using a cat whisker) (Stura and Wilson, 1990) was employed to find optimal crystallization conditions. Three-dimensional, highly birefringent and twinned crystals of apparently identical morphology were obtained in almost all variations of the original crystallization conditions when either the sodium citrate-derived crystal or the ammonium phosphate-derived crystal plates were used to seed equilibrated hanging drops of both cruzain and cruzain-Δc. The crystals of cruzain-Δc were consistently more three-dimensional and less twinned than those of cruzain; therefore, further crystallization studies have focused on the truncated mutant, which is inactivated with the fluoromethyl ketone inhibitor, Z-Phe-Ala-FMK.

Larger crystals were obtained by switching to a dilutional microseeding protocol with the best crystals obtained at a 10∂ dilution of microseeds. Fig. 2 shows a Coomassie-stained SDS-polycrylamide gel with crystals of cruzain (lane D) and cruzain-Δc (lane C) that were grown using this method. Precession photography analysis suggested that most crystals were highly twinned. The majority of the crystals presented a banded pattern of birefringence, while the rare untwinned crystal had uniform birefringence. Attempts were made to increase the percentage of untwinned crystals by inclusion of compounds such as dioxane and β-octyl glucoside in the crystallization solution and by crystallization at lowered temperatures. While the first approach was not helpful, it was possible to obtain untwinned crystals at 4 °C. With the exception of the initial crystallization event, which required several months, it has not been possible to grow crystals in the absence of seeds. Crystals grown from cruzain or cruzain-Δc, and from sodium citrate or ammonium phosphate, appear to be isomorphous. The crystals were characterized by precession

---

2 The following mutagenic oligos were used in the polymerase chain reaction to create the variant cruzains described herein: C25A, 5’ GCCAATGCGATCCGGCTGCGCTTCTCC3’; Δc, 5’ TCATCT- AGATCAACGGACACCAGGAG 3’ (used as antisense oligo in polymerase chain reaction); and E219P, 5’ TAGAACCTCCAGC- CTTGGCTCGTGGTTCCCGCCGCACCTCGCCCGCAACACC- AGC 3’. The codons in boldface encode the mutations.
photography analysis, and the results were corroborated by data collection on a Siemens area detector. A screened precession photograph revealed the periodic absence of reflections with indices \(0k0\), where \(k = 2n + 1\), indicating the symmetry of space group \(P2_1\). Cell constants were found to be \(a = 45.4\ \text{Å}\), \(b = 51.0\ \text{Å}\), \(c = 45.7\ \text{Å}\), and \(\beta = 116.1^\circ\). The cell volume and molecular mass of cruzain-\(\Delta\)c indicated one molecule in the asymmetric unit. This assumes an average crystal density and results in a \(V_n\) of 2.1 and an estimated 41% solvent content. The small, low symmetry unit cell and the lack of noncrystallographic symmetry will greatly facilitate the structure determination.

The amino acid sequence of cruzain-\(\Delta\)c is 35% identical to papain and includes several separate insertions or deletions. A 2.4-Å X-ray data set has been collected on cruzain-\(\Delta\)c and will be used to solve the structure by molecular replacement where the search model is derived from papain coordinates.

CONCLUSION

The methodology herein described readily produces reagent levels of recombinant forms of the major cysteine protease of \(T.\ cruzi\), cruzain. Analysis of the processing events suggested the production of variants of the mature enzyme that were crystallized. This system can now be used to create and characterize variants of the enzyme that explore the primary determinants influencing the structure and/or function of this biomedically important member of the thiol protease family. These studies should provide insight into the mechanism of action of cruzain and aid in developing specific inhibitors for antiparasitic chemotherapy in the treatment of Chagas' disease.

Acknowledgment—We acknowledge John Perona for advice on crystallization.

REFERENCES