Displacement of the Occluding Loop by the Parasite Protein, Chagasin, Results in Efficient Inhibition of Human Cathepsin B*

Received for publication, March 14, 2008, and in revised form, May 8, 2008 Published, JBC Papers in Press, May 30, 2008, DOI 10.1074/jbc.M802064200

Izabela Redzynia1, Anna Ljunggren1, Magnus Abrahamsson1, John S. Mort2, Joanne C. Krupa2, Mariusz Jaskolski3, and Grzegorz Bujacz2

From the 1Faculty of Biotechnology and Food Sciences, Technical University of Lodz, 90–924 Lodz, Poland, 2Department of Laboratory Medicine, Division of Clinical Chemistry and Pharmacology, Lund University, SE-221 85 Lund, Sweden, 3Joint Diseases Laboratory, Shriners Hospital for Children, 1529 Cedar Avenue, Montreal, Quebec H3G1A6, Canada, **Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61–704 Poznan, Poland

Cathepsin B is a papain-like cysteine protease showing both endo- and exopeptidase activity, the latter due to a unique occluding loop that restricts access to the active site cleft. To clarify the mode by which natural protein inhibitors manage to overcome this obstacle, we have analyzed the structure and function of cathepsin B in complexes with the Trypanosoma cruzi inhibitor, chagasin. Kinetic analysis revealed that substitution of His-110e, which anchors the loop in occluding position, results in 3-fold increased chagasin affinity (Ki for H110A cathepsin B, 0.35 nM) due to an improved association rate (kassN, 5 × 108 M⁻¹ s⁻¹). The structure of chagasin in complex with cathepsin B was solved in two crystal forms (1.8 and 2.67 Å resolution), demonstrating that the occluding loop is displaced to allow chagasin binding with its three loops, L4, L2, and L6, spanning the entire active site cleft. The occluding loop is differently displaced in the two structures, indicating a large range of movement and adoption of conformations forced by the inhibitor. The area of contact is slightly larger than in chagasin complexes with the endopeptidase, cathepsin L. However, residues important for high affinity to both enzymes are mainly found in the outer loops L4 and L6 of chagasin. The chagasin–cathepsin B complex provides a structural framework for modeling and design of inhibitors for cruzipain, the parasite cysteine protease and a virulence factor in Chagas disease.

Cysteine proteases that belong to the papain family (MEROPS family C1) (1) are essential for the normal body functions and are also implicated in many disease processes, making them candidate therapeutic targets (1–3). In addition to their action as endopeptidases, several family C1 enzymes also act as exopeptidases. This specialization is achieved in several ways by incorporation of additional structural elements that restrict access to specific regions of their active site cleft. For instance, aminopeptidases retain fragments of the propeptide that block part of the unprimed side of the active site (S4, S3, S2, and S1), whereas access to the primed sites (S1’, S2’, and S3’) in carboxypeptidases is controlled by an extra sequence that is inserted into the main chain of the mature enzyme and loops back into the active site cleft. In the case of the aminopeptidase cathepsin H, an eight-residue portion of the propeptide, the mini-chain, is held in the unprimed side of the active site (4). In cathepsin C, a large segment of the proregion, termed the exclusion domain, remains associated with the mature peptidase and blocks the active site cleft beyond the S2 site, rendering cathepsin C an aminopeptidase (5). In both enzymes, the additional elements provide a strategically positioned negative charge to accept the substrate N terminus. In cathepsins X and B, an occluding loop is present that blocks the primed side of the active site, restricting access to one (cathepsin X) (6) or two (cathepsin B) (7) residues.

In cathepsin B, a histidine residue (His-111e)4 provides a positive charge to accept the C-terminal carboxylate of a substrate. At low pH, the occluding loop is held in place by a salt bridge between His-110e in the middle of the loop and Asp-22e in the main body of the protease (8). The uniqueness of this arrangement has been exploited in the design of specific cathepsin B inhibitors, such as CA-074 (9), which are based on the epoxy-succinyl reactive center first recognized in the Aspergillus japonicus product, E-64 (10).

The occluding loop is not, however, fixed in one conformation. As the pH is raised, His-110e becomes deprotonated, and the loop is free to move, allowing cathepsin B to act as endopeptidase (11). The ability of the occluding loop to adopt multiple structures of in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

1 To whom correspondence may be addressed: Clinical Chemistry, University Hospital, SE-221 85 Lund, Sweden. Tel.: 46-46-173445; E-mail: Magnus.Abrahamsson@med.lu.se.
2 Supported by the Shriners of North America.
3 To whom correspondence may be addressed: Institute of Biochemistry, Stefaniowskiego 4/10, 90-924 Lodz, Poland. Tel.: 48-42-6313431; E-mail: gdbujacz@p.lodz.pl.
4 The residues of the inhibitor are labeled without a chain designator. The residues of the enzyme are marked e, except for the residues belonging to the prosegment sequence, which are marked p.
conformations is apparent in the structure of the proenzyme. Here, the proregion passes through the active site cleft, its peptide chain adopting the opposite direction relative to that required for substrate cleavage (12). The occluding loop is deflected by the propeptide, as a part of it occupies the catalytic crevice together with the occluding loop.

Naturally occurring inhibitors of cysteine proteases, such as the cystatins, bind to most of the papain family members with extremely high affinity (13). Much lower affinity is seen for cathepsin B due to obstruction of the binding site by the occluding loop (14). However, inhibition of cathepsin B by natural protein inhibitors does occur; hence, the structure of such complexes can provide interesting information on the mechanism of occlusion.

In this work we have determined two crystal structures of a complex between human cathepsin B and chagasin, a cysteine protease inhibitor from the protozoan pathogen Trypanosoma cruzi. The properties of chagasin are similar to those of cystatins, but its structure is very different (15–17). Chagasin belongs to a new inhibitor family (MEROPS family I42) (1) and its mode of binding to the cysteine endopeptidases, cathepsin L and falcipain, has been recently elucidated (16, 18, 19). The present structures of the cathepsin B inhibitor complex provide a detailed atomic view of how the parasite protein inhibits a cysteine exopeptidase that may be in the first line of the host defense. The high level of structural and functional similarity of cathepsins B and L with cruzipain also offers clues on how the cysteine protease activity of the parasite enzyme could be regulated. This may be of biomedical relevance since cruzipain is a potent virulence factor in T. cruzi infections and, hence, a possible therapeutic target in patients with Chagas disease.

**EXPERIMENTAL PROCEDURES**

**Proteins**—The recombinant chagasin construct included an extra N-terminal GPLGS sequence as a cloning artifact, with the remaining sequence corresponding to the native protein (GenBank™/EMBL accession number AJ299433) (20). The protein was expressed in a glutathione S-transferase fusion system (Amersham Biosciences) and purified by affinity and ion exchange chromatography as described earlier (16). The chagasin-containing fractions eluted in the final purification step were pooled, dialyzed against 100 mM Hepes buffer, pH 7.0, and concentrated to ~5 mg/ml using a Vivaspin column with a cut-off limit of 5 kDa (Vivasience, Lincoln, UK).

Human cathepsin B variants were produced by Pichia pastoris expression (Invitrogen). All recombinant variants were derived from a full-length pro-cathepsin B cDNA insert in which the codon for Ser-115e in an Asn-Xaa-Ser glycosylation consensus sequence had been changed by site-directed mutagenesis to an Ala-115e codon (21). By further mutagenesis, the variant (H110A,S115A)-cathepsin B was produced, with removal of the His-110e side chain that is believed to be critical for the conformation of the occluding loop (22). The mutant protein was originally designed to facilitate purification of the enzyme to homogeneity to avoid different glycoforms of the recombinant protein when produced in yeast cells. The S115A substitution has no effect on the enzymatic properties of cathepsin B; recombinant S115A-mutated cathepsin B has highly similar activity as the wild type protein purified from liver (23). The S115A- and (H110A,S115A)-cathepsin B variants were expressed as inactive pro-proteins. Autocatalytic activation was accomplished by incubation of the purified proenzyme in 20 mM sodium acetate, pH 5.0, 1 mM EDTA, for 5 days after which the active enzyme was purified by CM-Sepharose FastFlow column chromatography (22). For crystallization, inactive protein with the catalytic Cys-29e residue mutated to Ala was generated. Maturation of the resulting (C29A,H110A,S115A)-cathepsin B was achieved as described previously (14). Highly purified (C29A,H110A,S115A)-cathepsin B was dialyzed against 50 mM sodium acetate, pH 5.0, and concentrated to ~3 mg/ml for crystallization trials.

Active papain was prepared from the commercial papaya latex enzyme (Sigma-Aldrich) through affinity purification on Sepharose 4B-coupled Gly-Gly-Tyr-Arg column, as detailed previously (24, 25). Purified this way, the enzyme could be activated to at least 65% even after storage at −80 °C for up to 3 months.

**Enzyme Inhibition Assays**—Active-site titration of papain for the determination of the molar enzyme inhibitory concentrations of chagasin preparations was accomplished as described (26), demonstrating that the recombinant chagasin used was properly folded and essentially fully active as a protease inhibitor (16). The equilibrium dissociation constants (K_i) of chagasin complexes were determined for non-glycosylated S115A- and (H110A,S115A)-cathepsin B. The continuous-rate method used previously for studies of cystatin interactions with target proteases was followed (26). The fluorogenic substrate used was 10 μM Z-Phe-Arg-NHMec (Bachem Feinchemikalien, Bubendorf, Switzerland), and the assay buffer contained 133 mM sodium phosphate, pH 6.5, 1.3 mM EDTA, and 2.8 mM dithiothreitol. Steady-state velocities were measured, and apparent K_i values (K_i (app)) were calculated according to Henderson (27) using [I] values from active-site titrations. Corrections for substrate competition were made using a K_m value of 55 μM for cathepsin B hydrolysis of Z-Phe-Arg-NHMec (28). To calculate the association rate constants for chagasin interaction with S115A- and (H110A,S115A)-cathepsin B, the pseudo-first-order rate constants (k_0) in continuous-rate assays with different concentrations of chagasin were determined from the pre-steady-state phase of the velocity curve by non-linear regression using FLUSYS (obtained from Dr. Neil Rawlings, the Babraham Institute). The association rate constant (k_0) was then calculated from the slope of the k_0 values versus inhibitor concentration (Fig. 1).

**Protein Analyses**—Protein concentration in preparations used for crystallization trials was estimated by A_280 measurement using theoretical extinction coefficients calculated in PROTPARAM (28). N-terminal sequencing was carried out after electrophoresis in SDS-PAGE gels, blotting to a polyvinylidene difluoride membrane (Millipore, Bedford, MA), and staining with 0.05% Coomasie Brilliant Blue. Edman degradation was carried out by an Applied Biosystems Procise system (Sheldon Biotechnology Centre, McGill University). Matrix-as-

---

*The abbreviation used is: Z-Phe-Arg-NHMec, carbobenzoxy-phenylalanyl-arginine-7-amido-4-methylcoumarin.*
TABLE 1

Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>MAX-la, I911-5</th>
<th>MAX-la, I911-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation source</td>
<td>0.907</td>
<td>1.041</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Temperature of measurements (K)</td>
<td>32.0</td>
<td>32.0</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Cell parameters (Å,°)</td>
<td>a = 153.3, b = 30.8, c = 79.7, β = 92.7</td>
<td>a = b = 85.1, c = 115.7</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.0-1.80 (1.86-1.80)</td>
<td>20.0-2.67 (2.77-2.67)</td>
</tr>
<tr>
<td>Reflections collected</td>
<td>300,829</td>
<td>473,202</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>30,592 (2,955)</td>
<td>11,903 (1,175)</td>
</tr>
<tr>
<td>Completenez (%)</td>
<td>99.9 (97.2)</td>
<td>94.4 (96.5)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2.6 (2.6)</td>
<td>5.7 (5.0)</td>
</tr>
<tr>
<td>(h/ct)</td>
<td>14.5 (2.3)</td>
<td>15.0 (2.1)</td>
</tr>
<tr>
<td>Rint</td>
<td>0.076 (0.487)</td>
<td>0.109 (0.576)</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.038 (0.346)</td>
<td>0.042 (0.230)</td>
</tr>
</tbody>
</table>

Refinement

| No. of reflections in working/test set | 29,055/1537 | 11,311/573 |
| R(int)/Rfree (%)                     | 16.3/21.0 | 19.3/24.2 |
| Number of atoms (protein/solvent/PO4⁻) | 2847/243/20 | 2827/63/20 |
| Root mean square deviations from ideal |           |           |
| Bond lengths (Å)                     | 0.017 | 0.024 |
| Bond angles (°)                      | 1.53 | 2.36 |
| (B) (Å²)                             | 26.1 | 52.5 |
| Residues in Ramachandran plot (%)   | 85.7/97.3³ | 82.3/93.7³ |
| Most favored regions                 | 14.0 | 17.3 |
| Allowed regions                      | 0.3 | 0.3 |
| Generously allowed                   | 3CBJ | 3CBK |
| PDB code                             |            |            |

² Values in parentheses correspond to the last resolution shell.
³ Ramachandran favored from MolProbity (43).

sisted laser desorption/ionization time-of-flight mass spectrometry (Refract III, Bruker) analysis was used to verify the correct mass of recombinant chagasin and cathepsin B variants. Complex formation after mixing of purified chagasin and cathepsin B was studied using agarose gel electrophoresis (29). The fate of chagasin in such complexes after prolonged incubation was studied by SDS-PAGE using 4–12% gradient gels (Novex; Invitrogen) and the buffer system recommended by the gel supplier.

Crystalization—All crystals used in this study were grown by the hanging drop vapor diffusion method at 18 °C. Initial screening for crystallization conditions was done using Peg/Ion Screen and Crystal Screen (Hampton Research, Aliso Viejo, CA). Crystals of the chagasin–cathepsin B complex could be grown under three different conditions, and those used for diffraction experiments were obtained in the following crystallization trials. Crystal form I was obtained when the chagasin–cathepsin B complex was prepared by direct mixing in the drops of the following solutions: 1 μl of chagasin (5 mg/ml), 2 μl of cathepsin B (3 mg/ml), 2 μl of precipitant (0.2 M NH₄H₂PO₄, 20% polyethylene glycol 3350, pH 4.6). After 2 weeks clusters of rod-shaped crystals appeared that were tightly associated and very hard to separate. To prevent crystal cracking and degradation due to drop evaporation, the separation was performed under oil. A single crystal with the dimensions of 0.15 × 0.05 × 0.02 mm was used for data collection. Mineral oil was also used for cryoprotection. Crystal form II was obtained when the chagasin–cathepsin B complex was prepared by mixing chagasin at 5 mg/ml concentration with cathepsin B solution at 3 mg/ml concentration in a 1:3 volume ratio. Drops were mixed from 1.5 μl of protein solution and 1 μl of precipitant solution containing 0.14 M NH₄F and 14% polyethylene glycol 3350, pH 6.2. The crystals, in the form of thin square plates, reached maximum dimensions of 0.2 × 0.2 × 0.01 mm within 1 week. A mixture of the well solution with 50% polyethylene glycol 400 (v/v) in a 1:1 ratio was used as cryoprotectant.

Data Collection and Processing—X-ray diffraction data were collected for crystal forms I and II. The measurements were performed using synchrotron radiation generated at the MAX-lab (Lund, Sweden) beamlines I-911-2 and I-911-5. Crystals of form I are monoclinic, space group C2, and diffract X-rays to 1.80 Å resolution. Crystals of form II are tetragonal, space group P4₂₁₂, and diffract X-rays to 2.67 Å resolution. In both cases the diffraction data were indexed, integrated, and scaled with Denzo and Scalepack from the HKL program package (30).

Table 1 shows the data collection and processing statistics.

Structure Determination and Refinement—A partial solution of the structure of the chagasin–cathepsin B complex in crystal form I was obtained by molecular replacement using MolRep (31). The structure of human liver cathepsin B (PDB code 1HUC) (7) provided the initial search model. The complete model of the complex was obtained by superposition of the chagasin–cathepsin L model (PDB code 2NQD) (16) on the oriented cathepsin B portion as the target. The chagasin molecule derived from the cathepsin L complex fit the electron density map very well. The resulting model of the chagasin–cathepsin B complex was refined using a rigid body procedure, with the enzyme and the inhibitor chains treated as separate rigid blocks. The structure of the tetragonal crystal (form II) was solved by molecular replacement using MolRep and the form I model of the chagasin–cathepsin B complex. In both cases, structure refinement was carried out in Refmac5 (32) from the CCP4 package (33) using the maximum-likelihood targets. TLS parameters (34), defined separately for each polypeptide chain,
**Cathepsin B-Chagasin Complex**

were also optimized during the refinement. Model rebuilding was carried out in Coot (35) with water molecules introduced manually. Progress of the refinement was monitored, and the models were validated using R<sub>free</sub> testing (36). The quality of the final structures was assessed with Procheck (37). The final refinement statistics are shown in Table 1. The refined atomic coordinates and structure factors for the monoclinic and tetragonal structures have been deposited in the PDB under accession codes 3CBJ and 3CBK, respectively.

**RESULTS AND DISCUSSION**

**Chagasin Is an Efficient Inhibitor of Cathepsin B**—Chagasin is a small protein of 110 amino acid residues (M<sub>r</sub> 12,440 for the recombinant protein used in the present study), showing tight, reversible binding and inhibition of human cathepsin L in 1:1 molar ratio in competition with substrates but without being cleaved by the protease (16). It also inhibits papain-like cysteine proteases, such as cruzipain from the T. cruzi parasite itself or cathepsin B from the human host (15, 16). The function of chagasin as a cathepsin B inhibitor was analyzed in detail using two recombinant forms of the human enzyme. One was mature, non-glycosylated cathepsin B (S115A-cathepsin B), and the other contained an additional H110Ae substitution in the occluding loop (H110A,S115A-cathepsin B). The occluding loop is a unique structural feature of cathepsin B that partly covers the active site cleft and is responsible for the carboxy- dypeptidase activity of the enzyme under certain pH conditions (7). Logically, the occluding loop could also affect binding of a protein inhibitor to the active site cleft of the enzyme. Dilute cathepsin B assays with the substrate Z-Phe-Arg-NHMec at pH 6.5 allowed the determination of the equilibrium between free and chagasin-bound enzyme at steady state when varying concentrations of chagasin were added. Compensated for substrate competition, the calculated K<sub>i</sub> values for chagasin inhibition of wild type and H110Ae cathepsin B were 0.93 and 0.35 nM, respectively. Thus, removal of the His-110e side chain from the occluding loop results in noticeable improvement of inhibitor binding. However, the affinity for the mutated cathepsin B is still lower than that measured for chagasin binding to cysteine endopeptidases, papain, cruzipain, and cathepsin L (15, 16). The rate constants for association (k<sub>on</sub>) of chagasin with the two cathepsin B forms were determined under pseudo-first-order conditions from the rates of inactivation of the enzyme (Fig. 1). The calculated k<sub>on</sub> value for wild type human cathepsin B was 8 × 10<sup>4</sup> M<sup>−1</sup>s<sup>−1</sup>. The association rate for H110Ae cathepsin B was clearly faster with a k<sub>on</sub> value 5 × 10<sup>5</sup> M<sup>−1</sup>s<sup>−1</sup>, estimated from a linear relation between the rate and inhibitor concentration in the assay (Fig. 1), approaching the fast association rate observed for cathepsin L (2.5 × 10<sup>6</sup> M<sup>−1</sup>s<sup>−1</sup>) (16).

**Structural Determinants of the Chagasin-Cathepsin B Interaction**—The crystal structure of the chagasin complex with cathepsin B was determined for an inactive recombinant enzyme variant. Its primary structure corresponds to the native human cathepsin B sequence with the occluding loop substitutions H110Ae and S115Ae and with an additional C29Ae mutation of the catalytic cysteine residue introduced to prevent potential autocatalytic degradation under the extended incubation time required for crystallization.

![Inhibition of cathepsin B by chagasin](Image)

The chagasin-cathepsin B complex was crystallized in two forms, monoclinic (I) and tetragonal (II). Form I crystals were obtained from complex made by placing drops of the two proteins on the coverslip and allowing them to mix, whereas form II crystals were obtained from a complex prepared by preincubating both proteins before crystallization trials. Because the crystal structure of form I (Fig. 2) was determined to a significantly higher resolution, it will be used in the discussion unless stated otherwise.

The polypeptide chain of the cathepsin B model starts with residue Asp-60p, which is the third amino acid of an N-terminal extension derived from the propeptide attached to the mature sequence. This residue is not visible in the tetragonal model of the chagasin-cathepsin B complex (form II), where the polypeptide chain of cathepsin B starts with residue Leu-61p. The conformation of the N and C termini of cathepsin B is different in both structures. The residues from the N-terminal extension are well defined in the electron density in contrast to the C terminus, where the electron density map is fragmented and the two terminal residues, Asp-254e and Gln-255e are exposed to solvent and disordered.

The model of the chagasin molecule starts with either Met-1 (form II) or Ser-2 (form I) of the authentic sequence. The extension residues at the N terminus introduced as a cloning artifact are completely disordered.

The overall shape of the chagasin-cathepsin B complex resembles a rider on a horse, where the “rider” is formed by the cylindrical chagasin molecule and the “horse” by the globular cathepsin B (Fig. 2). The active site of the enzyme is located in a deep, V-shaped cleft that runs across the whole molecule and divides it into two distinct domains, termed the L and R lobe (7). The inhibitory epitope of chagasin is composed of three loops,
L4 (residues Pro-59—Gly-68), L2 (Asn-29—Phe-34), and L6 (Arg-91—Ser-100), protruding at one end of the molecule (Fig. 3a). The loops are nested deep in the active site groove of cathepsin B, with the central loop L2 anchored directly above the catalytic center and loops L4 and L6 embracing the enzyme laterally. Each of the loops has a different docking mode on the target enzyme surface (Fig. 3, b–d). The main hydrogen bonds between chagasin and cathepsin B in both complex structures are summarized in Table 2.

The L6 loop of chagasin interacts with the enzyme near the site used for docking the C-terminal portion of the substrate, at the outlet of the catalytic cleft. It forms three different types of interactions with the enzyme, hydrogen bonds (Arg-91), hydrophobic contacts (Pro-92), and π interactions (Trp-93) (Fig. 3b). To understand the intricate nature of these interactions, it is important to first realize that the classic interpretation of the enzymatic apparatus of a cysteine protease as a catalytic triad (Cys−29e—His−199e—Asn−219e in cathepsin B numbering) should be extended to include a fourth component, a cluster of aromatic residues (16), two of which (Trp-221e and Trp-225e) should be extended to include a fourth component, a cluster of aromatic residues, one (Trp-221e) of which is located near the catalytic center and forms hydrogen bonds with the carbonyl group of Thr-31. Another segment of the guanidinium group of Arg-91 forms a salt bridge with the side-chain carboxylate group of Asp-22e from cathepsin B. The presence of an acidic partner provided for this interaction by the enzyme reflects the situation in cruzipain, in contrast to the situation found in cathepsin L, where this position is held by asparagine. Finally, the guanidinium group of Arg-91 is also hydrogen-bonded to the carbonyl group of Gly-24e. The third element of L6, Pro-92, which serves to shape the loop for optimal interactions with the enzyme, is located in a hydrophobic cavity formed by the side chains of two leucine residues in the enzyme sequence, Leu-181e and Leu-182e. A structurally important Tyr-89 residue (19) bridges strands β3–β7 and is responsible for the stabilization and shape of loop L6 by forming an OH-N hydrogen bond with the main-chain N atom of Asn-29. In addition to the direct interactions of loop L6 described above, there are also contacts mediated by water molecules.

The interactions of loop L4 with the enzyme are based on the formation of an intermolecular antiparallel β-sheet. Three residues from chagasin, Leu-65, Gly-66, and Ala-67, interact with cathepsin B residues Gly-74e, Gly-73e, and Asn-72e. This β-sheet formation is supported by additional polar contacts involving the side chains of Asn-72e (hydrogen-bonded to Gly-68) and Glu-245e (interacted with Lys-63) (Fig. 3c).

Compared with the very strong and extended interactions of loops L4 and L6, the interactions of loop L2 are rather limited. The 310-helical turn at its apex is inserted directly into the active site but makes only two contacts with the enzyme. A repulsive contact is seen between the carbonyl O atom of Thr-31 and the Nε1 atom of the imidazole ring of the catalytic His-199e residue (Fig. 3d). An attractive contact exists between the same Thr-31 carbonyl and the Ne1 atom of Trp-221e. The presence of loop L2 directly above the catalytic center physically blocks any access to the active site of the enzyme.

The shape of the loop His-190e—Gly-198e is supported by interactions with the chagasin molecule. There is a strong hydrogen bond (2.89 Å) between the carbonyl O atom of Tyr-57 of chagasin and Oε1 of Glu-194e of cathepsin B (form I, Table 2). The residues of this cathepsin B loop also interact with chagasin with the mediation of water molecules. Oε2 of Glu-194e forms water-related hydrogen bonds with Oε1 of Glu-71 and Nδ2 of Asn-55. The carbonyl groups of two consecutive methionine residues, Met-195e and Met-196e, interact via water molecules with the C=O and OH groups of Tyr-57, respectively. Only one very weak hydrogen bond interaction (between Glu-194e and Lys-56 of chagasin) is seen in this region in the crystal form II.
comparison of the two chagasin-cathepsin b models — although loops l4, l2, and l6, defining the enzyme-binding epitope have an essentially identical conformation in the two complexes, the angle of approach of the inhibitor relative to the catalytic cleft is visibly different (fig. 4a). the interactions of loop l6 in both structures are based on aromatic-stacking and hydrophobic or polar contacts. the aromatic cluster in cathepsin l is created by trp-189e, trp-193e, and phe-143e, whereas in cathepsin b the analogous group of aromatic residues (trp-221e, trp-225e, and phe-180e) is extended by phe-196e.

comparison of chagasin complexes with cathepsin b and cathepsin l — a comparison of the chagasin-cathepsin b structure with the cathepsin l complex of the same inhibitor (16) shows that the overall architecture and mode of inhibitor–enzyme interactions are similar (fig. 4b). the interactions of loop l4 in both structures are based on antiparallel β-sheet formation. the enzymes have identical sequences in this region, including two consecutive glycine residues. a small conformational change of loop l4 is visible in a 1.9 Å shift of the co atom of lys-63, where the mode of interaction is not conserved.

the interactions of loop l6 are based on aromatic stacking and hydrophobic or polar contacts. the aromatic cluster in cathepsin l is created by trp-189e, trp-193e, and phe-143e, whereas in cathepsin b the analogous group of aromatic residues (trp-221e, trp-225e, and phe-180e) is extended by phe-
174e and Phe-231e, suggesting a stronger interaction for this hydrophobic patch. The hydrophobic contacts in cathepsin L are based on Pro-92 interactions with Leu-144e and Phe-145e. These interactions are mimicked by interactions with Leu-181e and Leu-182e in cathepsin B. The polar interactions of Arg-91 are stronger in the cathepsin B complex, where Asp-22e is found in place of Asn-18e.

The interactions between loop L2 and the catalytic center are identical in all complexes. Even the positions of three water molecules in the catalytic center are the same.

There are some additional interactions of the inhibitor with enzyme residues outside of the catalytic cleft. The Tyr-37—Glu-141e interaction seen in the chagasin-cathepsin L complex is not present in the chagasin-cathepsin B complex, but there is another interaction of this type, between Tyr-57 in strand β5 of chagasin and Glu-194e of cathepsin B (form I). This interaction is possible because of the movement of the loop His-190e—Gly-198e in the chagasin-cathepsin B complex. The central loop of the inhibitor molecule has only one repulsive contact, with the catalytic His residue functioning to physico-chemical block the catalytic center.

The catalytic pockets of cathepsin B and cruzipain are superposed almost perfectly (Fig. 4c), suggesting that the interactions of loop L2 of chagasin in this area will be essentially identical. The central loop of the inhibitor molecule has only one repulsive contact, with the catalytic His residue functioning to physically block the catalytic center.

The formation of an intermolecular β-sheet between loop L4 of chagasin and cathepsin B is facilitated by the presence of a flexible glycine segment (Gly-73e—Gly-74e) in cathepsin B, which corresponds to Gly-65e—Gly-66e in cruzipain. The polar contact of Lys-63e—Glu240e is mimicked in the cruzipain complex by interaction of Lys-63 with Glu205e, and additionally the same amino group of Lys-63 may interact with the side chain of Gln-156e. This interaction may be supported by the hydrogen bond between Asp-61 and the side chain of Gln-56e—Cys-98e. The other four are characteristic for cathepsin B and cross-link polypeptide fragments of either slightly (Cys-14e—Cys-43e, Cys-63e—Cys-67e, Cys-100e—Cys-132e) or completely different architecture (Cys-108e—Cys-119e). The cysteine residues Cys-29e and Cys-240e in cathepsin B are unpaired. Cys-29e, corresponding to Cys-25e in cathepsin L, is the active-site nucleophile. Although Cys-240e is equivalent to Cys-209e in cathepsin L, it is unable to pair in a cathepsin L-like fashion (Cys-209e—Cys-156e) because in the position of Cys-156e, Gly-193e is present. The missing disulfide bond in cathepsin B may be responsible for an entirely different course of the loop His-190e—Gly-198e.

**Predicted Chagasin-Cruzipain Interactions**—Although the complexes of chagasin with cathepsins B (this work) and L (16) are important for *T. cruzi* pathology, the chagasin-cruzipain complex is obviously essential for the physiology of the pathogen. Because no structural data are so far available, one may use the existing cathepsin complexes as templates for modeling chagasin-cruzipain interactions. Superposition of the catalytic part of cruzipain (PDB code 1ME3) on the cathepsin B molecule from form I of the chagasin-cathepsin B complex shows a possible mode of interaction of the parasite proteins.

The catalytic pockets of cathepsin B and cruzipain are superposed almost perfectly (Fig. 4c), suggesting that the interactions of loop L2 of chagasin in this area will be essentially identical. The central loop of the inhibitor molecule has only one repulsive contact, with the catalytic His residue functioning to physically block the catalytic center.

The formation of an intermolecular β-sheet between loop L4 of chagasin and cathepsin B is facilitated by the presence of a flexible glycine segment (Gly-73e—Gly-74e) in cathepsin B, which corresponds to Gly-65e—Gly-66e in cruzipain. The polar contact of Lys-63e—Glu240e is mimicked in the cruzipain complex by interaction of Lys-63 with Glu205e, and additionally the same amino group of Lys-63 may interact with the side chain of Gln-156e. This interaction may be supported by the hydrogen bond between Asp-61 and the side chain of Gln-
156e. An additional hydrophobic contact of Leu-64 of chagasin with Tyr-75e of cathepsin B is preserved because cruzipain has Leu-67e in this position.

The interactions of loop L6 of chagasin with cruzipain should be the same as those with cathepsin B. In particular, the interaction of Trp-93 with the aromatic cluster of the enzyme is conserved. The residues forming the aromatic cluster of cruzipain are Trp-181e, Trp-177e, and Trp-141e. The crucial salt bridge between Arg-91 of chagasin and Asp-22e of cathepsin B is preserved because cruzipain has Asp-18e in this position.

Movement of the Occluding Loop of Cathepsin B on Enzyme Inhibition—The occluding loop is a specific feature of cathepsin B. The 22 residues between Ile-105e and Pro-126e, creating this loop (Fig. 5, a and b), have no equivalent in cathepsin L, where there is only a short connection formed by residues Ala-93e and Thr-94e. The occluding loop is cross-linked at Cys-108e—Cys-119e (Fig. 5c), and this disulfide bond creates a covalently closed circular structure that starts and ends with the same Pro-Pro-Cys sequence. These two tripeptides may help stabilize the ends of the occluding loop in the L domain of the enzyme. In the native form of cathepsin B, this loop blocks access to the active site cleft (7). In the inhibitory conformation, the tip of the occluding loop is created by two histidine residues, His-110e and His-111e. Pushing the occluding loop out of the active-site cleft is connected with a reshaping of its tip. In procathepsin B as well as in both chagasin complexes, the tip of the loop is formed by Asn-113e and Gly-114e (Fig. 5c).

The occluding loop has a different conformation in the two chagasin complex structures (Fig. 5b). To measure the relative movement of the occluding loop, we arbitrarily chose the Ca position of Asn-113e. Relative to native cathepsin B (PDB code 1HUC) (7), the movement of the Asn-113e Ca marker in the tetragonal structure is 22.5 Å. The dis-
two loops run in the same direction, and the shape of both loops is also similar. The parallel orientation starts at Thr-94 of chagasin and Ala-110e of cathepsin B and ends at the pair Glu-101/Thr-120e. In the tetragonal structure, the interaction between loop L6 and the occluding loop is slightly different. The occluding loop is pushed out of the catalytic groove, and only two short fragments are parallel to loop L6 of chagasin. The first one is formed by Ser-97—His-98 from chagasin and Glu-109e—Ala-110e from cathepsin B. The second one is created between residues Asp-99—Glu-101 and Pro-118e—Thr-120e. There are hydrogen bond interactions between the side chain of Ser-97 and the O atom of Cys-108e and between the side chain of Asp-99 and the main-chain O atom of His-111e. At the C-terminus of the interacting loops, Oe1 of Glu-101 interacts with the N atom of Gly-121e. In the monoclinic structure a water molecule bridges the carbonyl groups of Asp-99 and Cys-119e. In the tetragonal structure there are probably some interactions mediated by water molecules, but the poor quality of the electron density maps in this flexible region precludes unambiguous modeling.

Comparison of Chagasin and Propeptide Binding in the Catalytic Cleft of Cathepsin B—The aromatic interactions described above are an important part of the active site architecture of papain-like cysteine proteases. At one end of the catalytic cleft, the aromatic cluster is responsible for binding the propeptidic part or the occluding loop in the native structure or of an inhibitory element such as the L6 loop of chagasin. The cluster of three cathepsin L-like residues (Trp-221e, Trp-225e, Phe-180e) is surrounded by Tyr-188e and His-199e and additionally by Phe-174e and Phe-231e. The interaction between Trp-93 of chagasin and two tryptophan residues (Trp-221e, Trp-225e) in the cathepsin B complex is preserved also in mature cathepsin B and in procathepsin B, where the interacting residues are His-111e (in the occluding loop) and Phe-30p.

FIGURE 4. Stereoviews of superposed enzymes complexes of chagasin. a, comparison of chagasin-cathepsin B models in two different crystal forms. The view is perpendicular to the standard orientation, with the L-domain behind the inhibitor molecule. The figure was prepared by superposition of the Ca atoms of the enzyme molecules. Color code: chagasin in cathepsin B complex, orange; chagasin in cathepsin L complex, gold; cathepsin B, lime green; cathepsin L, blue; chagasin in cathepsin L complex, gold; cathepsin B, lime green; chagasin in cathepsin L complex, orange. b, comparison of chagasin complexes with cathepsin B (form I) and cathepsin L (PDB code 2NQD) (16) in standard orientation. The figure was prepared by superposition of the Ca atoms of the enzyme molecules. Color code: chagasin in cathepsin B complex, orange. c, modeling of the chagasin-cruzain interaction based on the superposition of cruzain (burgundy) on the cathepsin B molecule (lime green) based on form I of the chagasin-cathepsin B complex shown with the L-domain on the front. The side chains of residues crucial for enzyme interactions are shown in ball-and-stick representation.

FIGURE 5. Movement of the occluding loop of cathepsin B in different states of the enzyme. a, cathepsin B from the monoclinic complex is shown in ribbon diagram (lime green) with semitransparent surface. b, zoom-in of the occluding loop region in different states. The occluding loop from the native enzyme (PDB code 1HUC) (7) is colored violet, from the proenzyme (PDB code 3PBH) (39) is raspberry, and from the tetragonal form of the chagasin-cathepsin B complex is forest green. c, a view illustrating the interactions between loop L6 of chagasin (orange) and the occluding loop of cathepsin B (lime green). The side chains of crucial residues are shown in ball-and-stick representation.
(in the propeptide) (Fig. 6, a and b). There is another interaction that is similar in native cathepsin B and in the chagasin complex where Arg-91 forms a strong salt bridge with the carboxylic group of Asp-22e. In native cathepsin B, an analogous interaction with Asp-22e is created by the Ne2 atom of the imidazole ring of His-110e in the occluding loop (Fig. 6a). In procathepsin B, the interaction of Asp-22e with the hydroxyl group of Tyr-37p from the propeptide is mediated by a water molecule. The repulsive interaction in the active site between the catalytic His-110e in the occluding loop (Fig. 6a) due to improvement of the “on” rate ($k_{on} = 5 \times 10^5 \text{m}^{-1}\text{s}^{-1}$). The structures of chagasin in complex with (C29A,H110A,S115A)-cathepsin B solved in two crystal forms reveal that the enzyme binding epitope defined by the loops L4, L2, and L6 has essentially fixed conformation but that the angle of approach of the inhibitor relative to the catalytic cleft of the enzyme is slightly different, resulting in a 5-Å tilt of the back side of the chagasin molecule. The occluding loop is differently displaced in the two structures by up to 14.5 or even 22.5 Å relative to its position in the wild type enzyme, indicating a significant degree of adaptation of the loop conformation under the influence of the inhibitor docked in the catalytic cleft. In both crystals forms there are interactions between chagasin and the occluding loop in the displaced position, contributing to overall affinity and indicating that the parasite inhibitor may have evolved to utilize the loop structure to achieve increased binding selectivity to its host target protease. Comparison with the recently solved structure of chagasin with the host endopeptidase, cathepsin L, shows that the area of contact is larger in the cathepsin B complex. Contact residues of importance for inhibitor–enzyme affinity are mainly found in the loops L4 and L6, although additional specificity-conferring residues could be identified in other parts of the chagasin molecule. The chagasin–cathepsin B complex described in this paper provides structural framework for designing inhibitors of cruzipain, the parasite cysteine protease and virulence factor at $T. cruzi$ infections, and may, therefore, aid the development of therapeutic agents against Chagas disease.

Acknowledgment—We thank Robert Kolodziejczyk (A. Mickiewicz University) for the idea of separating the crystals under oil and for skillfully performing this task.

REFERENCES
Displacement of the Occluding Loop by the Parasite Protein, Chagasin, Results in Efficient Inhibition of Human Cathepsin B
Izabela Redzynia, Anna Ljunggren, Magnus Abrahamson, John S. Mort, Joanne C. Krupa, Mariusz Jaskolski and Grzegorz Bujacz

doi: 10.1074/jbc.M802064200 originally published online May 30, 2008

Access the most updated version of this article at doi: 10.1074/jbc.M802064200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 41 references, 9 of which can be accessed free at http://www.jbc.org/content/283/33/22815.full.html#ref-list-1