



LUND UNIVERSITY
Faculty of Medicine

LU:research

Institutional Repository of Lund University

This is an author produced version of a paper published in Journal of molecular biology. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the published paper:

Ljunggren, Anna and Redzynia, Izabela and Alvarez-Fernandez, Marcia and Abrahamson, Magnus and Mort, John S and Krupa, Joanne C and Jaskolski, Mariusz and Bujacz, Grzegorz.

"Crystal structure of the parasite protease inhibitor chagasin in complex with a host target cysteine protease."
J Mol Biol, 2007, Vol: 371, Issue: 1, pp. 137-53.

<http://dx.doi.org/10.1016/j.jmb.2007.05.005>

Access to the published version may
require journal subscription.
Published with permission from: Elsevier

Crystal structure of the parasite protease inhibitor chagasin in complex with a host target cysteine protease

Anna Ljunggren¹, Izabela Redzynia², Marcia Alvarez-Fernandez¹, Magnus Abrahamson¹,
John S. Mort³, Joanne C. Krupa³, Mariusz Jaskolski^{4,5}, Grzegorz Bujacz^{2,5}

¹Department of Laboratory Medicine, Division of Clinical Chemistry and Pharmacology, Lund University, Sweden

²Faculty of Biotechnology and Food Sciences, Technical University of Lodz, Lodz, Poland

³Joint Diseases Laboratory, Shriners Hospital for Children, Montreal, Quebec, Canada

⁴Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland

⁵Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

RUNNING TITLE: Chagasin-cathepsin L complex structure

Corresponding author:

Prof. Grzegorz Bujacz

Lodz, Poland

Phone: (48-42)-6313431

Fax: (48-42)-6313402

Email: gdbujacz@p.lodz.pl

FOOTNOTES

This work was supported in part by a grant from the State Committee for Scientific Research (T09A 039 25) and by a subsidy from the Foundation for Polish Science to M.J. The work in Lund was supported by grants from the Crafoord, A. Österlund and Magn. Bergvall Foundations, the Swedish Research Council (project no. 09915) and the Research School in Pharmaceutical Chemistry ("FLÄK"), Lund University. J.C.K. and J.S.M. were supported by the Shriners of North America. We thank Santosh Panjikar from EMBL Hamburg for his help with the data collection and Patrizia Mason for the mutagenesis of the procathepsin L construct.

E-mail address of the corresponding author: gdbujacz@p.lodz.pl

Protein Data Bank accession codes: Atomic coordinates, together with the structure factors, have been deposited in the Protein Data Bank (PDB) under the accession codes 2NNR (chagasin) and 2NQD (chagasin-cathepsin L complex).

The residues of the inhibitor are labeled without a chain designator. The residues of the enzyme are marked "e", except for the first residue, marked "p" (E1p), which belongs to the prosegment sequence.

Abbreviations used: E-64, *L-trans*-epoxysuccinylleucylamido-(4-guanidino)butane; MAD, multiwavelength anomalous diffraction; MS, mass spectrometry; PDB, Protein Data Bank; NHMec, 7-(4-methyl)coumarylamide; r.m.s., root mean square; Z-, carboxybenzoyl-; GST, glutathione-S-transferase

SUMMARY

Chagasin is a protein produced by *Trypanosoma cruzi*, the parasite causing Chagas disease. This small protein belongs to a recently defined family of cysteine protease inhibitors. Although resembling well-known inhibitors like the cystatins in size (110 amino acid residues) and function (they all inhibit papain-like (C1 family) proteases), it has a unique amino acid sequence and structure. We have crystallized and solved the structure of chagasin in complex with the host cysteine protease, cathepsin L, at 1.75 Å resolution. An inhibitory wedge composed of three loops (L2, L4 and L6) forms a number of contacts responsible for high affinity binding (K_i , 39 pM) to the enzyme. All three loops interact with the catalytic groove, with the central loop L2 inserted directly into the catalytic center. Loops L4 and L6 embrace the enzyme molecule from both sides and exhibit distinctly different patterns of protein-protein recognition. Comparison with a 1.7 Å structure of uncomplexed chagasin, also determined in this study, demonstrates that a conformational change of the first binding loop (L4) allows extended binding to the non-primed substrate pockets of the enzyme active site cleft, thereby providing a substantial part of the inhibitory surface. The mode of chagasin binding is generally similar, albeit distinctly different in detail, when compared to those displayed by cystatins and the cysteine protease inhibitory p41 fragment of the invariant chain. The chagasin-cathepsin L complex structure provides details of how the parasite protein inhibits a host enzyme of possible importance in host defense. The high structural and functional similarity between cathepsin L and the *T. cruzi* enzyme, cruzipain, gives clues to how the cysteine protease activity of the parasite can be targeted. This information will aid in the development of synthetic inhibitors for use as potential drugs for the treatment of Chagas disease.

Keywords: *Trypanosoma cruzi*; Chagas disease; cysteine peptidases; cysteine proteinases; lysosomal enzymes; protein inhibitors

INTRODUCTION

An estimated 18 million people in Central and South America are infected with the protozoan pathogen *Trypanosoma cruzi* and about 120 million are at risk. Up to 30% of those infected with the parasite will develop the irreversible medical condition known as Chagas disease, with fatal prognosis. The protozoan causing Chagas disease is related to those causing African sleeping sickness and leishmaniasis, but the insect vector, *Triatoma infestans*, is different.

T. cruzi expresses a potent papain-like (C1 family) cysteine protease, called cruzipain or cruzain, which is the main antigen recognized by the immune system at infection and is essential both for the life-cycle of the parasite and for its interaction with host cells. Because of these properties, the enzyme is a promising target for therapy and it has indeed been demonstrated that synthetic cysteine protease inhibitors strongly limit the parasitic replication both in cell culture and in mice.¹ The natural regulation of cruzipain in *T. cruzi* is largely unknown. It can be assumed that, as in other uni- and multicellular organisms, a balance exists between cysteine protease(s) and specific inhibitor(s). In a variety of species, from mammals, fish, plants, and insects to simpler eukaryotic organisms such as the filarial parasites *Onchocerca volvulus* and *Acanthocheilonema viteae*, cysteine proteases are in equilibrium with inhibitors of the cystatin family I25.^{2,3,4,5} Most cystatins are single-domain proteins composed of 100-120 residues with a characteristic fold presenting the N-terminal segment and two β -hairpin loops for complementation of the active site cleft of the target enzyme.⁶ They show high affinity for their target enzymes, with equilibrium constants for dissociation of the formed complexes (K_i) typically in the range of 10^{-9} - 10^{-11} M, but these can even be as low as 10^{-14} M, as in the case of the human cystatin C – papain complex.⁷ A completely different structure resulting in a similar high-affinity inhibition of certain papain-like cysteine proteases is represented by the p41 fragment of the MHC class II-associated invariant chain, originally isolated from human kidney, as a complex with cathepsin L. The p41 fragment is smaller than cystatins (64 amino acid residues), but exhibits a similar mode of enzyme inhibition that utilizes three loop fragments forming a wedge to cover the active site cleft; it binds with high affinity (K_i 10^{-9} - 10^{-12} M for cathepsin L, cruzipain, and papain).^{8,9}

Evidence has been presented that trypanosomatids, such as various *Trypanosoma* and *Leishmania* species, contain cystatin-like inhibitors of their own family C1 proteases.¹⁰ Searching for such an inhibitor in *T. cruzi*, we identified and cloned chagasin, a potent reversible inhibitor of cruzipain, with an appropriate size (110 amino acid residues), but no obvious sequence similarity to cystatins (GenBank/EMBL Acc. No. **AJ299433**).¹¹ Subsequent identification and partial characterization of chagasin in *T. cruzi* extracts showed that it indeed functions very much like a cystatin, showing high-affinity and tight-binding inhibition of papain and cruzipain.¹² Early molecular modeling studies predicted an immunoglobulin-like fold for chagasin,¹³ which was essentially confirmed by a recent NMR study¹⁴ and an X-ray analysis of chagasin crystallized in the presence of polyethylene glycol (PEG).¹⁵ Despite this progress in the elucidation of the chagasin structure, the key issue of the mechanism of inhibition has been left to conjecture, based only on modeling^{15,16} and cursory NMR data¹⁴.

In the present study, we present a high-quality crystallographic structure of chagasin in complex with human cathepsin L, which is a potential target enzyme during *T. cruzi* infection. The results confirm mapping of the enzyme-binding epitope to three loops and illustrate the degree of structural adjustments as well as precise atomic contacts formed during enzyme binding. These aspects of chagasin inhibition, in particular the conformational adjustment of one of the inhibitory loops, were impossible to accurately predict in the earlier modeling studies. Chagasin inhibits the enzyme by inserting one loop into the active site and by using two lateral loops for docking on target surfaces, with a different mode of interaction exploited by each loop. Detailed knowledge of the structure and inhibition mode of chagasin will guide the development of drugs for possible prevention and treatment of Chagas disease.

RESULTS

Properties of chagasin

A cDNA fragment encoding the 110 residues of native chagasin was amplified from the full-length *T. cruzi* clone Tc18¹¹ and subcloned in the expression vector pGEX-6P to produce a GST-chagasin fusion protein in *E. coli*. Complete DNA sequencing of the insert in the expression vector verified that the recombinant protein was encoded as intended, i.e. with 110 residues of native chagasin (Fig. 1a) sequence plus five extra residues at the N-terminus (GPLGS) derived from the expression vector construct. After glutathione affinity chromatography and proteolytic cleavage of the fusion protein, chagasin could be purified to homogeneity by ion exchange chromatography. Physicochemical characterization of the isolated recombinant chagasin revealed the correct size and charge according to electrophoretic analyses. The results of N-terminal protein sequencing revealed a homogenous sequence GPLGSMH- and MS analyses produced a single peak corresponding to to expected 12,440 Da.

The function of recombinant chagasin was analyzed by titrating dilutions of the isolated protein into a papain solution of known active enzyme concentration, with results supporting that the preparation contained properly folded and fully active protein. The concentration of active inhibitor established this way allowed for the determination of kinetic parameters for the interaction of chagasin with cathepsin L using a continuous fluorimetric assay with the substrate Z-Phe-Arg-NHMec. By varying the assay concentration of chagasin, under dilute conditions so that the equilibrium between free and inhibitor-bound enzyme could be monitored, we could estimate both the equilibrium constant (K_i) and rate constants for the interaction. The apparent K_i value was 0.16 nM. Compensating for effects of the competing substrate, the true K_i value was calculated as 39 pM, demonstrating a very tight interaction between chagasin and cathepsin L, comparable to those of the best cathepsin L inhibitors of the cystatin family (Table 1). The apparent rate constant for association (k_{on}) of chagasin with cathepsin L was determined under pseudo-first order conditions to be $2.5 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$. Calculated from the determined apparent K_i and k_{on} values, the rate constant for dissociation (k_{off}) was $4.0 \times 10^{-4} \text{ s}^{-1}$, agreeing with that estimated from the progress curves ($7.3 \times 10^{-4} \text{ s}^{-1}$).¹⁷ Analysis of the chagasin interactions with

other cysteine proteases shows that the inhibition of cathepsin L and the cathepsin L-like *T. cruzi* enzyme, cruzipain, is in the range of 0.0039 to 0.018 nM (Table 1). Although with lower affinity, chagasin can also inhibit human cathepsin B, which several cystatins fail to do (Table 1).

Crystallization and structure determination

Recombinant chagasin, Se-Met-substituted chagasin, and a complex of native recombinant chagasin with the human cysteine protease cathepsin L were all subjected to crystallization trials. Cathepsin L used in these studies was produced by *Pichia pastoris* expression and purified to homogeneity. The expression vector insert was mutated to replace the catalytically active C25 residue with an alanine (C25A), but otherwise encoded native procathepsin L. The procathepsin L was converted to the mature form by addition of catalytic amounts of active recombinant wild-type cathepsin L, after which the mature cathepsin L was isolated by ion exchange chromatography. Complex formation between chagasin and cathepsin L was carried out directly in the crystallization drops, as detailed in the Materials and Methods section.

Crystals of native chagasin were obtained using either lithium sulfate or ammonium sulfate as precipitants. Although the crystals from ammonium sulfate were larger and better formed, they cracked on flash-freezing. Consequently, X-ray diffraction data were collected for the crystals obtained with lithium sulfate, using 10% glycerol as cryoprotectant. The crystals of Se-Met chagasin were grown from protein solutions of lower concentration and with a higher precipitant concentration; they did not require any additional cryoprotection.

The crystal structure of chagasin solved from Se-MAD data revealed two molecules (A and B) in the asymmetric unit. The refinement was carried out using the native data set extending to 1.7 Å resolution. The main chain of both chagasin molecules corresponding to the native amino acid sequence could be traced end-to-end without breaks in the $2F_o - F_c$ electron density map at the 1.2σ level and all the side chains are visible at the 0.8σ level.

Single crystals of the chagasin-cathepsin L complex were obtained using the initial screening (see Materials and Methods). The 1.75 Å resolution structure of the complex was solved by molecular

replacement with two separate starting structures. The coordinates of the cathepsin L molecule with the truncated proenzyme (PDB code 1CS8)¹⁸ were used as the first model. The second run of molecular replacement was performed with fixed cathepsin L molecule and molecule A of chagasin as the search model. In the chagasin-cathepsin L complex, the main chain is visible in the 2F_o-F_c electron density map without breaks at the 1.3σ level, except for the loop S174e-D178e and E1p of the cathepsin L molecule. All side chains and the above-mentioned segments of the main chain are visible at the 0.7σ contour level.

Description of the chagasin structure

Although the structures of *T. cruzi* chagasin and of its *Leishmania mexicana* homologue have been reported recently,^{14,15,16} a brief overview of our high-resolution crystal structure of free chagasin is useful. The chagasin molecule is composed almost exclusively of β-strands (Fig. 1a) and can be considered as a split β-barrel formed by two β-sheets, whose edges do not meet. Four β-strands, β1, β3, β6 and β5, on one side of the barrel, create a sheet with a predominately antiparallel architecture, in which only strand β1 is parallel to its neighbor (Fig. 1b). The second β-sheet, created by strands β2, β8, β7 and β4, is also predominantly antiparallel, with only β2 parallel to its neighbor. The space between the sheets is filled with a large number of hydrophobic side chains. The only exception is E71, forming a contact with the main-chain N atom of W36. It has been noted before that this topology resembles a distorted immunoglobulin fold¹³. The distortion, however, is quite extensive and can be described as an insertion of the β1-loop-β2 motif, which violates the parallel character of both constituent β-sheets. A prominent feature of this folding scheme is the clustering of loops at each end of the barrel. Loops L1, L3 and L5 on one side of the molecule are compact and do not stick out from the molecule. Loops L1 and L5 are short, composed of four and three residues, respectively. The L3 loop, consisting of residues E39-M50 is much longer; it traces a semi-helical turn in its N-terminal part, but despite this it does not protrude from the molecular outline. Loop L3 assumes two clearly distinguishable conformations in molecule A, but both of them are different from the single conformation observed in molecule B.

Loops L2, L4 and L6 are located at the opposite end of the molecule. The connection between

strands $\beta 1$ and $\beta 2$ on the same side of the β -barrel forms a tight turn with 3_{10} geometry (Fig. 1c). The three broad loops L4, L2 and L6 are located one after another in an approximately coplanar alignment resembling a fan. Loop L4 is created by residues P59-G68 and loop L2 by residues N29-F34, with a single 3_{10} -helical turn at its apex. Loop L6, made up of residues R91-S100, is large and takes the form of two turns.

Overall, the shape and the functional elements of the chagasin molecule are reminiscent of a “three-prong plug”, with the prongs created by the loops L4, L2 and L6 as the enzyme-binding region (Fig. 1).

Comparison of the two crystallographic copies of the free chagasin molecule

The molecules A and B of free chagasin are very similar (Fig. 1b), superimposable with an r.m.s. deviation between their $C\alpha$ atoms of 0.35 Å (Table 2). The largest differences are seen at the N-terminus, where the tips of the chains diverge by 5 Å. Starting from residue H3, the superposition is almost complete, except for loop L3, where deviations up to 2.4 Å are observed. The variable conformation is attributed to increased flexibility of this loop. In molecule A, the E39-T42 segment of the main chain assumes two conformations. The majority of the side chains of the two chagasin molecules form identical rotamers. Different conformations of the side chains are observed only in the regions where there are measurable discrepancies between the main chains, *i.e.* at the N terminus and in loop L3. In addition, a few surface side chains, such as E44, E48, N47, E54, K56, K63, H98, D99, R102, and N110, have slightly different conformations.

Comparison of free chagasin and chagasin in complex with cathepsin L

Comparison of the free chagasin structure with the chagasin molecule in complex with cathepsin L shows that the molecule does not change its general fold upon binding to the enzyme (Fig. 1b). The overall r.m.s. deviation between the corresponding $C\alpha$ atoms is 0.5 Å (Table 2). An important exception is loop L4, which undergoes a large conformational change corresponding to 4.5 Å at the $C\alpha$ atom of K63. An even larger movement is observed for the side chain of this lysine residue, where the terminal nitrogen atom moves by 13.1 Å. In the altered conformation, the L4 loop can form a number of contacts with the enzyme, as described below. In particular, the

amino group of K63 forms, in one of two alternate conformations, a hydrogen bond with the OH group of Y72e. The other elements of the enzyme-binding epitope (loops L2 and L6) do not change conformation upon docking to cathepsin L. The movement of other side chains is not large and is limited to some solvent-exposed residues.

The chagasin-cathepsin L interface

The polypeptide chain of cathepsin L in the chagasin complex starts with residue E1p, which is the last residue of the propeptide sequence¹⁹ (UniProtKB/Swiss-Prot P07711).²⁰ This residue is highly mobile and has been modeled in two conformations. The next residue, A1e, which is the first residue of the native cathepsin L sequence, is well defined in the electron density. The last residue, V220e, is stabilized by a hydrogen bond with the $-\text{NH}_3^+$ group of K41e. The primary structure of the enzyme corresponds to the native sequence, except for the catalytic C25e residue, which has been mutated to alanine. The residue N108e, which is part of a glycosylation consensus sequence²¹, is partially glycosylated and two NAG sugar units with fractional occupancy have been modeled.

The cathepsin L structure in the present chagasin complex is very similar to the high-resolution structure of procathepsin L (PDB code 1CS8), the r.m.s. deviation between the $\text{C}\alpha$ atoms being 0.35 Å. The only significant discrepancy is observed in the loop formed by residues E173e-N180e. This loop is very flexible as attested by the high atomic B-factors in this region. It has a blunt tip and forms two turns, E173e-E176e and S177e-N180e. This conformation is influenced by a crystal contact of the side chain amide group of N179e with the carboxyl group of E192e from a symmetry-related cathepsin L molecule. Additionally, the disordered side chain of E176e forms close contacts with the main-chain carbonyl of residue N179e and with the side chain carboxyl of E173e in its two orientations. The side-chain hydroxyl group of S177e forms a hydrogen bond with the main-chain carbonyl of T175e, serving as a "clip" in the middle of the flexible loop and creating the two turns mentioned above.

The overall shape of the complex resembles an inverted mushroom, with the stalk formed by the cylindrical chagasin molecule and the cap by the globular cathepsin L (Fig. 2a). The active site of

cathepsin L is located in a long cleft running across the whole molecule, with the sides constituting the L- and R-domains of a typical papain-like cysteine protease.²² One side of this cleft is formed by residues H140e-Y146e, in helical conformation, and the loop S157e-D162e of the L-domain. The other side of the cleft is created by loops N18e-A25e and N62e-G68e of the R-domain. The bottom of the cleft is made up of loops E159e-H163e and N187e-E191e. The catalytic triad C25e-H163e-N187e is located in the middle and at the bottom of the cleft. In the construct used in the present study, the nucleophilic C25e residue had been mutated to an alanine.

The binding region of chagasin, composed of loops L2, L4, and L6, binds very tightly to the cathepsin L molecule. The main hydrogen bonds between chagasin and cathepsin L in the complex are listed in Table 3. All three loops are located in the catalytic groove, with the central loop L2 inserted directly into the catalytic center. Loops L4 and L6 embrace the enzyme molecule from both sides (Fig. 2b).

The interactions of each loop have distinctly different character. Compared to the very strong and extended interactions of loops L4 and L6, the interactions of loop L2 are very limited. Its 3_{10} -helical turn is inserted directly into the active site but makes only two contacts with the enzyme (Fig. 2c). An effect of a repulsive contact is seen between the carbonyl O atom of T31 and the N δ 1 atom of the imidazole ring of the catalytic H163e residue. A much longer, attractive contact exists between the same T31 carbonyl and the N ϵ 1 atom of W189e (Table 3). The repulsive interaction between T31 and the imidazole ring of H163e blocks any access to the active site of the enzyme.

The tight interaction of loop L4 with the enzyme is based on formation of an intermolecular antiparallel β -sheet. Three residues from chagasin, L65, G66 and A67, interact with cathepsin L residues from G68e through N66e (Fig. 2d). This β -sheet formation is supported by two polar contacts, namely the side chain carbonyl O δ 1 atom of N66e forms a hydrogen bond (mediated by a chloride anion) with the main-chain N atom of G68, and the K63 amino group of chagasin interacts with the hydroxyl group of Y72e. Additionally, the main-chain G66 of chagasin forms a hydrogen bond with the main-chain carbonyl group of D162e from cathepsin L.

Loop L6 forms three different types of interactions with the enzyme: hydrogen bonds (R91), hydrophobic contacts (P92), and π interactions (W93) (Fig. 2e). For the understanding of the intricate nature of these interactions it is important to first realize that the classical interpretation of the enzymatic apparatus of a cysteine protease as a catalytic triad (C25-H163-N187 in cathepsin L numbering) should be extended to include a fourth component of this system, namely a cluster of aromatic residues which serves to anchor the asparagine residue (N187e). This aromatic cluster includes a number of residues, two of which (W189e and W193e) provide N-H... π interactions for the anchoring of the N187e side-chain amide group (Fig. 2e). The essence of the interactions of the chagasin L6 loop with the enzyme is in the recognition of the crucial residues in the catalytic cleft. Residue W93 becomes part of the enzyme's aromatic cluster through a C-H... π interaction with W193e. Another element of loop L6, R91, assumes a fully extended conformation reaching to the catalytic site and to loop L2, which is also present there. The guanidinium group of R91 forms two H bonds with the carbonyl group of T32 in loop L2, which is located next to the active-site-blocking T31 (*vide supra*). The other segment of the guanidinium group of R91 forms a pair of H bonds with the side-chain C=O group of the enzyme's N18e residue. It is interesting to note that the equivalent position in cruzipain is occupied by an aspartate, making the interaction with R91 even better. Finally, the guanidinium group of R91 is also hydrogen bonded to the carbonyl group of G20e. The third element of L6, P92, serves to shape the loop for optimal interactions with the enzyme. It is located in a hydrophobic cavity formed by the side chains of F145e and L144e. In addition to the interactions of loop L6 described above, other contacts are mediated by water molecules. Wat37 forms three hydrogen bonds, with the side-chain nitrogen atom of W93, with the carbonyl oxygen atom of W189e, and with the carboxylate group of E192e. Wat133 bridges the edges of the loop by interacting with the hydroxyl group of T94 and with the main-chain nitrogen atom of R91.

The E141e residue, located at the edge of the catalytic groove of cathepsin L, forms two additional hydrogen bonds with two chagasin residues that do not belong to the L4-L2-L6 binding epitope. The first contact is formed with the hydroxyl group of Y37 from strand β 4, and the second contact is a salt bridge formed with the fully extended side chain of K43 from loop L3 (Table 3). These two strong hydrogen bonds may be responsible for the skewed approach of the

inhibitor into the catalytic cleft, visible for example in Fig. 2a.

Inhibition mode presented by chagasin in comparison with cystatins

The most studied group of cysteine protease inhibitors are the cystatins, which are small proteins with a size of 10-12 kDa.²³ The mode of inhibition of cystatins can be described on the basis of the crystal structure of cystatin B (also called stefin B) in complex with papain (PDB code 1STF)²⁴ or cystatin A (also called stefin A) in complex with cathepsin H (PDB code 1NB5).²⁵ For comparison with the chagasin cathepsin L complex, we decided to use the cystatin B-papain complex²⁴ rather than the cystatin A-cathepsin H complex,²⁵ because of the higher structural and enzymatic similarities of papain to cathepsin L and cruzipain.

Superposition of the chagasin-cathepsin L complex with that of cystatin B-papain complex²⁴ shows a completely different fold for the two inhibitors, namely the all- β structure of chagasin and a five-stranded antiparallel β -sheet wrapped around a long α -helix of the cystatin. Although both inhibitors use a three-element epitope for enzyme binding, the angle of approach for the bulk of the inhibitors is different (Fig. 3a).

Despite of the lack of overall similarity of the inhibitors' sequences and structures, one notes a striking analogy of the inhibition mode and interactions of the three binding elements (Fig. 3b). First of all, however, one should note that the three-element epitopes of these two classes of enzymes have different topology. While the epitope of cystatins is sequential, *i.e.* consists of the N-terminus, loop L1 and loop L2 in this order, these elements are mapped, respectively, onto loops L4, L2, and L6 of chagasin.

The three crucial residues R91, P92, and W93 of loop L6 of chagasin correspond, respectively, to L102, P103, and H104 in the cystatin B molecule. It is noteworthy that the pattern Pro-aromatic is conserved in cystatins and in chagasin-like inhibitors, despite the lack of overall sequence similarity. The role of the proline residue seems to be to maintain the specific shape of the loop. The aromatic residue, on the other hand, interacts with the aromatic cluster of the enzyme. The residue preceding the Pro-aromatic motif, which is invariably an arginine in chagasin-type inhibitors of protozoan origin, is replaced by an aliphatic residue in cystatins. This difference

may be an important element regulating the enzyme specificity of these two groups in inhibitors. The R91 residue of chagasin provides direct communication between loops L6 and L2, and also interacts with the crucial N18e of the cathepsin L molecule. An additional interaction between loops L2 and L6 of chagasin is provided by the carbonyl group of the main chain of M90 and the nitrogen atom of A35. A similar stabilizing contact between cystatin B loops L2 and L1 is formed by the main-chain carbonyl of Q101 and the peptide nitrogen of T58.

The interaction of loop L4 of chagasin with cathepsin L is based on an intermolecular β -sheet formation. There is an analogous interaction between the N terminus of cystatin B and papain (Fig. 3b). G9 from the N terminus of cystatin B and G66 from loop L4 of chagasin provide a degree of flexibility, thus allowing for optimal interactions between the two main chains. The same role is played by G65e and G67e in papain and cathepsin L, respectively.

When structurally aligned in their respective enzyme complex structures, loops L2 of chagasin and L1 of cystatin B reveal a trace of sequence similarity. S28 from chagasin corresponds to S52 in cystatin B and G33 from chagasin maps onto G57 in cystatin B. The conformation of these two loops differs but, in both cases, there is a surprisingly small number of specific interactions with the catalytic residues of the enzyme. The repulsive interaction between chagasin and cathepsin L (*vide supra*) is reproduced in the cystatin complex as well.

Comparison of cathepsin L complexes of chagasin and p41

The complex of the p41 fragment of the invariant chain with cathepsin L (PDB code 1ICF) shows a mode of inhibition of papain-like cysteine proteases that is similar in principle but completely different in detail when compared to that displayed by cystatins.⁹ The small p41 fragment consists of two subdomains, where the first one is composed of an α -helix and a β -strand and the second one of a three-stranded antiparallel β -sheet. Similarly to cystatins and chagasin, the molecule interacts with the enzyme through a wedge composed of three hairpin loops (Fig. 3a). When superimposing our chagasin-cathepsin L structure with that of the p41 fragment-cathepsin L complex, distinct differences in the binding mode of the inhibitors to the enzyme are seen (Fig. 3c). Although the overall affinity for cathepsin L is very high and almost identical for chagasin

and the p41 fragment, with K_i of 3.9 (this work) and 1.7 pM,⁸ respectively, the mode of interaction of the particular loops is different.

The largest difference is observed for the loop R248-H252 in the p41 fragment, which corresponds to loop L6 in chagasin. This loop is much shorter and interacts with the enzyme differently. Although the Pro-aromatic motif is missing in the p41 fragment, the guanidinium group of its R250 residue mimics the indole ring of W93 of chagasin. Analogously to the situation observed in the chagasin-cathepsin L complex, there is also an interaction between the inhibitory loops of the p41 fragment, namely, the loop formed by residues T247-N254 interacts with the middle loop via hydrogen bonds between the side chain atoms of R250 and the main chain carbonyls of I231 and G232, respectively.

The interaction motif of the loop in the p41 fragment that corresponds to loop L4 in chagasin is one residue shorter. In addition to the shortened loop, the rigidity of P209 residue (which maps onto L65 of chagasin), makes the intermolecular β -sheet less extensive.

In comparison to the middle L2 loop of chagasin, the equivalent loop Y228-F233 of the p41 fragment interacts more tightly. The tight turn of that loop is supported by a disulfide bridge between C227 and C234. The first hydrogen bond occurs between the hydroxyl group of S230 and a sulfur atom of the catalytic C25e. The additional H-bonds are formed between the carbonyl of S230 with the side chain amide group of Q19e of cathepsin L and with the indole nitrogen of W189e. The carbonyl of G229 can interact with the imidazole ring of H163e. This observation suggests that the less extensive interactions of the loops L4 and L6 are compensated by the tighter interaction of the active center.

DISCUSSION

Comparison of different chagasin models

The present structures of free chagasin and chagasin in complex with cathepsin L, together with the recently published NMR and crystallographic models of free chagasin,^{15,16} allow a critical evaluation of the existing speculations concerning the chagasin structure and mechanism of

inhibition.

Comparison with the NMR model

Although the general fold of the chagasin molecule determined in this study is the same as that derived from solution NMR measurements (PDB code 2FO8),¹⁴ the r.m.s. deviation in C α superposition with the first chain of the NMR ensemble is very high, approaching 3 Å (Table 3). Part of this high discrepancy can be related to the increased dynamic properties of the enzyme-binding loops reported for the NMR model. Notably, no disorder of these loops is seen in any of the three crystallographic copies of chagasin. Furthermore, loops L2 and L6 show invariant and stable conformation, unaffected by enzyme binding. Loop L4 does undergo a conformational change upon complex formation and, indeed, it is this region that shows the highest mobility in the NMR model. However, this highly dynamic region seen in the NMR structure extends from residue P60 to G69, whereas G69 (in strand β 6) in the crystallographic models forms antiparallel β -sheet hydrogen bonds with S28 of β 3 (also disordered in the NMR model). The highest structural variability in the crystallographic models is seen at loop L3. In the NMR model, this region of the non-inhibitory end of the molecule is also characterized by increased dynamic properties.

Comparison with a crystallographic model obtained using different crystallization conditions

The coordinates corresponding to the structure of chagasin reported by Figueiredo da Silva *et al.*¹⁵ were released by the PDB (PDB code 2H7W) during the final stage of preparation of this manuscript. The crystal used to determine that structure (referred to as the “PEG” structure) was grown from 15% PEG8000, 0.1 M ammonium sulfate, and 0.1 M cacodylate buffer at pH 6.5. Despite the different crystallization conditions, the structure is almost isomorphous with that determined by us and described in this paper. The space group is the same for crystals obtained under both sets of conditions. Whereas other cell dimensions are very similar, the c parameter is 4 Å shorter for the crystal grown from PEG. The resolution of both data sets is the same. The architecture of both molecules in the asymmetric unit is similar (C α r.m.s. deviation 0.38 Å and 0.46 Å for molecules A and B, respectively). The relative position of the monomers in the A/B unit is slightly different due to contraction of the unit cell. The present structure is more complete

and the model contains all residues, including all side chains. In the PEG structure, two residues at the N-terminus of chain A, one at the N-terminus, and one at the C-terminus of chain B are missing. Additionally, eight side chains have been truncated to alanine. The chagasin construct used for the PEG crystallization contained a His₆-tag with an additional linker region of 17 residues, not visible in the structure. In the PDB-deposited structure 2H7W, residue 104 is isoleucine instead of the threonine present in the native chagasin sequence at this position (GenBank/EMBL Acc. No. **AJ299433**).¹¹ The 2H7W structure was refined with anisotropic atomic displacement parameters, although the resolution of the experimental data (1.7 Å) is too low to justify utilization of such a model. It must be stressed that despite the unwarranted proliferation of refined parameters, the 2H7W model is characterized by significantly worse R (20.7% vs 18.6%) and R_{free} (23.8% vs 21.6%) factors. Since the structure of chagasin refined by us is complete and has superior statistical parameters, we have decided to include its discussion in this paper, and use it as a reference chagasin model.

Docking of chagasin in the active site of cruzipain

The complex described in the present study should be relevant to both physiology and pathology, as chagasin is present on the parasite surface or outside the cell,¹² where it can interact with human cysteine proteases, such as cathepsin L. Since cathepsin L is likely to be important in the host response to the invading parasite through its regulatory properties on antigen presentation,⁸ it is possible that chagasin-mediated inhibition of cathepsin L improves the chances of survival for the parasite or that the parasite needs chagasin to regulate the action of its own cathepsin L-like cysteine protease, cruzipain. Superposition of the catalytic part of cruzipain²⁶ (PDB code 1ME3) on the cathepsin L molecule in its chagasin complex, shows a possible interaction mode of chagasin with the parasite enzyme. The catalytic pockets of cathepsin L and cruzipain are superposed almost perfectly (Fig. 4a). This observation suggests that the interactions of loop L2 of chagasin in this area will be essentially identical for both enzymes.

The interaction of loop L4 of chagasin with cathepsin L is based on intramolecular β -sheet formation. The formation of the main-chain – main-chain hydrogen bonds is possible due to the presence of flexible glycine residues G67e and G68e in the cathepsin L molecule, which

correspond to G65e and G66e in cruzipain. The additional polar contact between the amino group of K63 and the hydroxyl group Y72e can be maintained because N70e is present in this position in cruzipain. In support of the importance of this interaction, small-molecule inhibitors developed for cruzipain²⁶ adopt a conformation that is almost identical to that of loop L4 and create the same pattern of hydrogen bonds mimicking the antiparallel β -sheet (Fig. 4b). The additional hydrophobic contact of L65 of chagasin with L69e of cathepsin L is preserved because cruzipain has L67e in this position. In the synthetic inhibitors, the position of L65 is occupied by a phenylalanine residue.

The interaction of loop L6 of chagasin with cruzipain should basically be the same as with cathepsin L. The interaction of W93 of chagasin with the aromatic cluster is maintained. The aromatic cluster of cathepsin L comprising W193e, W189e, and F143e is reproduced in cruzipain by a triad of tryptophan residues, W181e, W177e and W141e. The crucial polar interaction of R91 of chagasin with N18e of cathepsin L can be even stronger because cruzipain has D18e in this position.

Models of chagasin binding mode

Two recently published studies^{14,15} have attempted to model the interactions between chagasin and cruzipain to elucidate the binding utilized to achieve efficient enzyme inhibition. The studies used different techniques to identify inhibitor-enzyme interactions. In one of them, the residues involved in enzyme binding were pinpointed by recording differential NMR signals after addition of cruzipain to isotope-labeled chagasin in solution. Reportedly, some residues show a large chemical shift and change of signal intensity even when they are distant from the binding region. This approach properly identified the amino acids of loops L4, L2, and L6 as the binding region, but could not specify the exact nature of the interactions with cruzipain, leaving the detailed picture of the enzyme binding as an open question.

In the X-ray structural study of chagasin,¹⁵ the authors proposed a model of its mode of enzyme interaction based on docking of the inhibitor molecule to the high-resolution structure of the catalytic domain of cruzipain. With knowledge of the experimental results of the present study

we can now see that the overall architecture of the complex was predicted correctly by this modeling approach, but the details show significant deficiencies of the model. Most importantly, the modeling failed to predict the conformational adaptation of loop L4 on enzyme binding and the creation of the extended intermolecular β -sheet. The interaction of residue G66 of chagasin with D158e from cruzipain (corresponding to D162e in cathepsin L) was properly identified, but the polar interaction of K63 is not present in that model. For the central L2 loop the interactions of both threonine residues, T31 and T32, were not predicted correctly. The repulsive interaction with the catalytic H159e residue and the stabilizing interaction of R91 are also missing. For loop L6 the interaction of W93 with W181e is identified correctly in the model, but the more complex interaction with the cluster of aromatic residues (W181e, W177e, W141e) is not described. The crucial role of R91 in loop stabilization and enzyme interaction was not discussed in the model. Also, the role of P92 in the shaping of loop L6 and in the hydrophobic interactions with the enzyme has not been noticed.

The experimentally determined structure of the chagasin-cathepsin L complex described in the present study provides a more complete picture than what could be gleaned from the modeling studies. It explains the efficient inhibition of both cathepsin L and cruzipain through clear elucidation of the inhibitor-enzyme interactions at atomic detail and through precise description of the role of the individual residues. This complex structure should constitute a reliable target for the development of synthetic inhibitors of cruzipain, which could lead to the design of potential drugs for the treatment of Chagas disease.

MATERIALS AND METHODS

Expression, purification, and characterization of recombinant chagasin

Production of chagasin was accomplished in the glutathione S-transferase (GST) Gene Fusion Protein System with the vector pGEX-6P-1 (Amersham Biosciences, Uppsala, Sweden). A chagasin-encoding fragment for insertion into the plasmid was made by PCR using a full-length chagasin cDNA (Genbank/EMBL acc. no. **AJ299433**) as template with primers MA550 (5'-TAT

ATG GAT CCA TGT CCC ACA AGG TGA CGA A-3') and MA551 (5'-TAC ATG AAT TCT CAG TTT GCC TTG AGA TAT ACA-3') containing flanking restriction sites for *Bam*HI and *Eco*RI, respectively. The fragment was cloned between the *Bam*HI and *Eco*RI sites in the pGEX-6P-1 cloning region. Correct cloning and position of the insert was verified by sequencing (ABI Prism 310 Genetic Analyser) using vector primers MA866 (5'-GGG CTG GCA AGC CAC GTT TGG TG-3') and MA867 (5'-CCG GGA GCT GCA TGT GTC AGA GG-3'). The plasmid was transformed into competent *E. coli* BL21 cells (Novagen, VWR International AB, Stockholm, Sweden) for cytosolic overexpression of the protein. Single colonies were picked from agar plates and initially grown in YT medium at 37 °C. Expression was induced with 0.1 mM IPTG and the culture incubated for 3 h at 30 °C. After centrifugation the cells were lysed by sonication and centrifuged. The recombinant fusion protein was purified on a GST Trap FF column (Amersham Biosciences, Uppsala, Sweden) using 10 mM reduced glutathione in 50 mM Tris buffer at pH 8.0 for elution, and then cleaved with PreScission Protease (Amersham). Final purification of recombinant chagasin from GST and traces of uncleaved fusion protein was done by anion exchange chromatography on a Mono Q column (10x100 mm²; Amersham) in 20 mM ethanolamine, pH 9.0, using a gradient of the same buffer containing 1M NaCl for elution. Purity of the recombinant chagasin was verified by SDS-PAGE (4-12%; Novex; Invitrogen AB, Stockholm, Sweden) and 1% agarose gel electrophoresis.²⁷ The fractions of highest chagasin concentration were pooled and dialyzed against 100 mM Hepes buffer at pH 7.0. The protein solution was concentrated using a Vivaspin column with a cut-off limit of 5 kDa (Vivascience, Lincoln, UK) to a final concentration of 8 mg/ml.

Production of Se-Met chagasin

Production of Se-Met containing chagasin was accomplished in the system described above. The plasmid was transformed into competent *E. coli* B834 (DE3) cells (Novagen, VWR International AB, Stockholm, Sweden). Single colonies were picked and grown in minimal medium, as described earlier.²⁸ The cells were centrifuged and resuspended in minimal medium containing Se-Met for expression. The culture was grown in 30 °C, induced with 0.1 mM IPTG, and incubated for 4 h. The cells were harvested by centrifugation and lysed by sonication. The protein

was purified as described above and concentrated to 5 mg/ml. Successful Met→Se-Met substitution was confirmed with mass spectrometric analysis.

Production of recombinant human cathepsin L

By site-directed mutagenesis²⁹ the region coding for C25 in a wild-type human cathepsin L cDNA was mutated to an alanine codon using the oligonucleotide 5'-T CAG TGT GGT AGC GCT TGG GCT TTT AG-3' (mutated bases underlined). This mutation introduced an *Hae*II restriction site. The mutated cDNA was subcloned into the vector pPIC9 (Invitrogen; San Diego, CA, U.S.A) and introduced into the yeast *P. pastoris*, strain GS115 (Invitrogen), by electroporation. The mutant proenzyme was expressed by the yeast under similar conditions to those outlined by the manufacturer (Invitrogen) and to those previously described for the C25S/T110A mutant.³⁰ The culture medium was centrifuged and the resulting supernatant was concentrated using an Amicon YM10 membrane. The concentrated medium was dialyzed against 50 mM NaOAc, pH 4.7. After adjusting the pH to 5.0, the C25A procathepsin L mutant was processed to the mature form upon incubation with 100 nM of active recombinant wild-type cathepsin L³¹ in 50 mM NaOAc, 2 mM DTT at 37° C for 18 h. The addition of wild-type cathepsin L corresponded to approximately 0.6% $w_{\text{active}}/w_{\text{inactive}}$ protein. The activity of the wild-type enzyme was inhibited by addition of 24 μM E-64. The excess E-64 was removed when the processed C25A inactive enzyme was purified by weak cation exchange chromatography (CM Sepharose Fast Flow, GE Healthcare). Fractions free of propeptide, as analyzed by SDS-PAGE, were eluted in the wash. As estimated from the extinction coefficient, 4 L of buffered minimal glycerol-complex medium (BMGY) yielded approximately 9 mg of C25A cathepsin L.

Protein analyses

Protein concentration in preparations used for crystallization trials was estimated by Comma^sie Protein Assay (Pierce Biotechnology, Inc., Rockford, IL, USA). N-terminal sequencing was carried out after electrophoresis in agarose gels, blotting to a PVDF membrane (Millipore, Bedford, MA, USA) and staining with 0.05% Coomassie blue. Edman degradation was carried out by an Applied Biosystems 470A gas-liquid solid-phase sequenator at the Department of Clinical Chemistry, Malmö University Hospital, Sweden. MALDI-TOF mass spectrometry

(Reflex III, Bruker) analysis was used to verify the correct mass of recombinant chagasin and Se-Met chagasin as described earlier.³²

Enzyme inhibition assays

Active site titration of papain (with E-64, using Bz-DL-Arg-NHPhNO₂ as substrate; Bachem Feinchemikalien, Bubendorf, Switzerland) and titration of the molar papain-inhibitory concentration of chagasin was accomplished as earlier described for cystatin analysis.³³ Active inhibitor concentrations determined in this way (as this is the method normally used for other cysteine protease inhibitors) were used for calculation of K_i values. The fluorogenic substrate used for determination of equilibrium constants for dissociation (K_i) of complexes between chagasin and cathepsin L or other family C1 cysteine peptidases was Z-Phe-Arg-NHMec (10 mM; from Bachem) and the assay medium was 100 mM Na-phosphate buffer, pH 6.0, containing 1 mM dithiothreitol and 1 mM EDTA. Steady-state velocities were measured before and after addition of varying amounts of chagasin in assays at 37 °C, and K_i values were calculated according to Henderson.³⁴ Corrections for substrate competition were made using K_m values determined for the substrate batch used, under the assay conditions employed (60, 55, 3.2 and 1.9 μ M for papain, cathepsin B, cathepsin L and cruzipain, respectively). To determine the association rate constant for the chagasin-cathepsin L interaction, the observed pseudo-first-order rate constants (k_{obs}) in continuous-rate assays with different concentrations of chagasin were determined by non-linear regression. The association rate constant (k_{on}) was then calculated from the slope of a plot of the k_{obs} values vs. inhibitor concentrations.

Crystallization

All crystallization experiments were carried out at 18 °C using the hanging-drop vapor diffusion method. Initial screening for crystallization conditions of native chagasin was done using Crystal Screen 1 (Hampton Research). Drops were mixed from a 2 μ l protein solution at 8 mg/ml concentration and 2 μ l precipitant solution. Crystals of native chagasin were seen under different conditions, but those used for diffraction experiments were grown from 1.5 M Li₂SO₄ in 0.1 M HEPES buffer, pH 8.5. The crystals were triangular plates measuring 0.3x0.3x0.1 mm. A mixture

of the well solution with 10% (v/v) of glycerol was used as cryoprotectant. Crystals of Se-Met chagasin were grown after adjustment of the conditions used for native recombinant chagasin. The protein at a concentration of 3 mg/ml was mixed with the well solution in the proportion 2 μ l protein solution to 4 μ l of well solution (2.4 M Li_2SO_4 in 0.1 M HEPES buffer, pH 7.5). Crystals of recombinant chagasin and Se-Met chagasin appeared in three days. Due to a high concentration of salt, the crystal of Se-Met chagasin soaked in the well solution did not require additional cryoprotection.

Crystallization of chagasin in complex with cathepsin L was carried out using Crystal Screen 1, as described above. The complex was prepared by direct mixing in the drops using the following solutions: 1 μ l chagasin (8 mg/ml), 4 μ l cathepsin L (0.3 mg/ml), 2 μ l precipitant (20% v/v isopropanol, 20% w/v PEG4000, 0.1 M sodium citrate buffer, pH 5.6). A rod-shaped crystal of the chagasin-cathepsin L complex appeared after four weeks and grew to the dimensions of 0.20x0.05x0.05 mm in 15 weeks.

X-ray data collection and analysis

Chagasin: The measurements were performed at 100 K using the EMBL beamlines of the DESY synchrotron in Hamburg. For native chagasin, a data set with 1.7 Å resolution and with good statistics was recorded (Table 2). Se-Met chagasin crystals showed inferior diffraction properties. Smearing of the reflections was the most characteristic defect of the diffraction pattern. However, a crystal that was accidentally dropped into the well solution, and as a consequence dehydrated because of the high salt concentration, showed significant improvement of diffraction quality and was, in the end, used to collect MAD data. The diffraction data were indexed, integrated and scaled using the HKL program package.³⁵

Chagasin-cathepsin L complex: A crystal of the complex was obtained from the screen solution without further optimization. The crystal was small, around 0.15 mm long, but allowed collection of a 1.75 Å resolution data set. The diffraction data were collected using the X13 EMBL beamline of the DESY synchrotron. The crystal was flash-frozen in nitrogen gas stream at 100 K without additional cryoprotection. The images were indexed, integrated and scaled using the HKL program package.³⁵ Table 4 shows the statistics of data collection and processing.

Structure solution and refinement

Chagasin: The structure of chagasin was solved by the Se-Met MAD method, with full experimental details to be published elsewhere. The model derived from the initial Se-Met-phased map was refined against a native data set extending to 1.70 Å resolution. The refinement was carried out with the Refmac5 program³⁶ from the CCP4 package³⁷ and included TLS parameters. Each chagasin molecule was divided into two rigid-body segments corresponding to the two β -sheets comprising the molecular core. Model rebuilding in the electron density maps was carried out using the program Coot.³⁸ Water molecules were either automatically identified by the ARP/wARP program³⁹ or manually introduced using the Coot software. The progress of the refinement was monitored and the model was validated using R_{free} ⁴⁰. The side chains of a number of residues were modeled in double conformations. In addition, the main chain of the E39-T42 fragment of loop L3 in molecule A also had dual conformation. The quality of the final structure was assessed with PROCHECK.⁴¹ The final statistics are shown in Table 4.

Chagasin-cathepsin L complex: The structure of the complex was solved by double molecular replacement. An initial molecular-replacement solution ($R=52.1\%$, correlation coefficient 55.9%) was obtained using the catalytic part of procathepsin L (PDB code 1CS8) as the search model in MolRep.⁴² The second run of molecular replacement used fixed contribution of the oriented and positioned cathepsin L molecule, whereas molecule A of the previously solved chagasin structure provided the search model ($R=45.8\%$, correlation coefficient 61.7%). Manual model rebuilding was done subsequently using Coot. For refinement, the Refmac5 program was used with TLS parameters refined for cathepsin L and chagasin (two rigid-body segments defined as above). Water molecules were either automatically identified by ARP/wARP or manually introduced using the Coot software. R_{free} was monitored using a randomly chosen subset of reflections comprising 5% of the unique data set. The side chains of a number of surface residues were modeled in two conformations. The quality of the final structure was assessed with PROCHECK. The final statistics are shown in Table 4.

All crystallographic calculations were performed using the CCP4 suite of programs. Molecular illustrations were prepared using PyMOL.⁴³

FIGURE LEGENDS

Fig. 1. Structural properties of chagasin

a, Amino acid sequence of chagasin with secondary structure elements as assigned by DSSP.⁴⁴ **b**, Superposition of the C α -traces of the two copies of free chagasin (molecules A and B in light and dark green, respectively) and of chagasin from the cathepsin L complex (orange), as determined in the present study. **c**, Main-chain hydrogen bonds in the core of the protein. Additionally, main-chain hydrogen-bond interactions with residues of cathepsin L (boxed) are shown.

Fig. 2. Structure of the chagasin-cathepsin L complex

a, A ribbon stereodiagram of the complex. The inhibitor molecule is colored gold and the enzyme blue. The surfaces of both proteins are marked in an analogous way. The view corresponds to a standard orientation used for cysteine proteases, along the interface between the left (L-) and right (R-) domains forming the sides of the active site cleft of cathepsin L. **b**, Interaction of chagasin in the catalytic cleft of cathepsin L, viewed perpendicular to the standard orientation, with the L-domain behind the inhibitor molecule and the R-domain in the front. The side chains of residues crucial for enzyme interactions are represented by sticks and balls (the disordered side chain of K63 is shown in both alternate conformations) **c, d, e**, Details of the enzyme interactions of loops L2, L4 and L6, respectively. In (**c**), side chains of the interacting residues are shown in 2F_o-F_c electron density contoured at the 1.2 σ level, to illustrate its quality.

Fig. 3. Chagasin mode of enzyme binding compared to those of other cysteine protease inhibitors (stereoviews)

a, Superimposed inhibitors bound to cathepsin L. The picture was prepared based on the superposition of the enzyme part: cathepsin L for chagasin and p41 fragment complexes and papain (not shown) for cystatin B complex, generated with ALIGN⁴⁵. The inhibitor molecules are presented in their original orientation relative to the enzyme. Color code: chagasin (gold), cystatin B (brown), p41 fragment (red). **b**, Shape of the binding loops of chagasin (gold) in the present complex with cathepsin L (blue) and of cystatin B (brown) in its complex with papain, after least-squares fit of the C α traces of the enzymes. **c**, Shape of the binding loops of chagasin

(gold) in the present complex with cathepsin L (blue) and of p41 fragment (red) from its complex with cathepsin L, after least-squares fit of the C α traces of the enzymes.

Fig. 4. Docking of chagasin into the active site of cruzipain

a, A ribbon stereodiagram of the superposition of the chagasin-cathepsin L complex (colored as in Fig. 2a) with cruzipain (burgundy), after least-squares fit of the C α traces of the enzymes. The view is in standard orientation along the active site cleft of the enzymes. **b**, A detailed view of the interactions showing the central part of the chagasin inhibitory wedge (gold) and the complementary fragment of the L-domain of cruzipain (burgundy). The side chains of residues crucial for the interactions are shown in ball-and-stick representation. The synthetic inhibitor P10 ([1-(3-hydroxy-2-oxo-1-phenethyl-propylcarbamoyl)-2-phenyl-ethyl]-carbamic acid pyridin-4-ylmethyl ester), from the complex with cruzipain (PDB code 1ME3) is shown in green.

Table 1. Inhibition of cysteine proteases by chagasin

Equilibrium constants for dissociation of chagasin complexes with relevant papain-like cysteine proteases were determined under steady-state conditions as described in Materials and Methods. The corresponding values for cystatins C, D, E/M and F, determined by similar methods,²³ are given for comparison. The presented K_i values were corrected for substrate competition in the assays, as described in the Materials and Methods section. N.D., not determined.

	K_i (nM)				
	<i>Chagasin</i>	<i>Cystatin C</i>	<i>Cystatin D</i>	<i>Cystatin E/M</i>	<i>Cystatin F</i>
<i>Papain</i>	0.023	0.000011	0.9	0.46	1.1
<i>Cruzipain</i>	0.0039	0.0048	N.D.	N.D.	N.D.
<i>Cathepsin B</i>	1.9 ^a	0.27	>1000	31	>1000
<i>Cathepsin L</i>	0.018 ^a	< 0.005	5.8	N.D.	0.31

^a Determined for a recombinant chagasin variant with a 16-residue N-terminal extension instead of the 5-residue extension in the variant used in the present study.

Table 2. Least-squares superpositions of various chagasin models

A, B – free chagasin (from the present study), Compl – chagasin from the cathepsin L complex (from the present study), NMR – the lowest-energy chagasin model derived by NMR spectroscopy (PDB code 2FO8). For each superposition, calculated in ALIGN,⁴⁵ the r.m.s. deviation (Å) between the N superposable C α atoms is given (N in parentheses). Also listed is the maximum deviation within the superposable set, and the largest discrepancy (Å) between the superposed molecules [residue numbers in brackets].

	B	Compl	NMR
A	0.35 (104) 0.97 5.02 [1/1]	0.48 (100) 1.03 4.65 [63/63]	2.69 (99) 7.05 7.05 [30/31]
B		0.52 (103) 1.20 4.28 [63/63]	2.77 (101) 7.06 7.06 [65/62]
Compl			2.37 (95) 6.08 8.04 [30/31]

Table 3. Hydrogen bonds between chagasin and cathepsin L in the complex

chagasin - cathepsin L				Distance (Å)
Loop	(res)	atom	- atom (res)	
L2	(T31)	O	- Nε (Q19e)	3.34
L2	(T31)	O	- Nε1 (W189e)	2.98
β4	(Y37)	Oη	- Oε2 (E141e)	2.59
L3	(K43)	Nζ	- Oε1 (E141e)	2.82
L4	(K63)	Nζ	- Oη (Y72e)	2.74
L4	(L65)	N	- O (G68e)	2.90
L4	(L65)	O	- N (G68e)	3.06
L4	(G66)	N	- O (D162e)	2.85
L4	(A67)	N	- O (N66e)	3.69
L4	(G68)	N	- Oδ1 (N66e)	3.00
L4	(G68)	N	- Cl	2.84
		Cl ^b	- Nδ2 (N66e)	2.65
L6	(R91)	Nη1	- Oδ1 (N18e)	2.86
L6	(R91)	Nη2	- Oδ1 (N18e)	3.16
L6	(R91)	Nη2	- O (G20e)	3.39

^b Chloride anion

Table 4. Data collection and structure refinement statistics

	Chagasin	Chagasin-cathepsin L complex
<i>Data collection</i>		
Radiation source	X11, EMBL Hamburg	X13, EMBL Hamburg
Wavelength (Å)	0.8123	0.8078
Temperature of measurements (K)	100	100
Space group	P3 ₁ 12	P2 ₁
Cell dimensions (Å, °)	$a=49.55, b=49.55, c=187.58$ $\alpha=90, \beta=90, \gamma=120$	$a=59.27, b=50.06, c=65.01$ $\alpha=90, \beta=103.4, \gamma=90$
Resolution range (Å)	45.0-1.70 (1.73-1.70) ^c	40.0-1.75 (1.81-1.75)
Unique reflections	29434	37521
Completeness (%)	99.6 (100.0)	98.9 (92.5)
Redundancy	7.8 (7.2)	3.6 (3.0)
$\langle I \rangle / \langle \sigma I \rangle$	38.3 (4.0)	20.6 (3.3)
$R_{\text{int}}^{\text{d}}$ (%)	3.8 (44.4)	6.1 (39.9)
<i>Refinement</i>		
Independent molecules	2	1
No. of reflections in working set	28186	35622
No. of reflections in test set	1218	1891
$R^{\text{e}}/R_{\text{free}}$ (%)	18.6/21.6	14.8/18.8
Number of atoms, protein/solvent/Na ⁺ /Cl ⁻ /SO ₄ ²⁻ /other	1770/195/-/1/5/42	2665/564/1/1/-/28
r.m.s. deviations from ideal		
Bond lengths (Å)	0.014	0.015
Bond angles (°)	1.54	1.43
Average B factor (Å ²)	42.6	21.5
Ramachandran statistics (%)		
Most favored	92.9	87.6
Additional allowed	7.1	12.4

^c Values in parentheses correspond to the last resolution shell.

^d $R_{\text{int}} = \sum_h \sum_j |I_{hj} - \langle I_h \rangle| / \sum_h \sum_j I_{hj}$, where I_{hj} is the intensity of observation j of reflection h .

^e $R = \sum_h ||F_o| - |F_c|| / \sum_h ||F_o|$ for all reflections, where F_o and F_c are the observed and calculated structure factors, respectively. R_{free} is calculated analogously for the test reflections, randomly selected and excluded from the refinement.

REFERENCES

- ¹ McGrath, M.E., Eakin, A.E., Engel, J.C., McKerrow, J.H., Craik, C.S. & Fletterick, R.J. (2005). The crystal structure of cruzain: a therapeutic target for Chagas' disease. *J. Mol. Biol.* **24**, 251-259.
- ² Rawlings, N.D. & Barrett, A.J. (1990). Evolution of proteins of the cystatin superfamily. *J. Mol. Evol.* **30**, 60-71.
- ³ Lustigman, S., Brotman, B., Huima, T., Prince, A.M. & McKerrow J.H. (1992). Molecular cloning and characterization of onchocystatin, a cysteine proteinase inhibitor of *Onchocerca volvulus*. *J. Biol.* **267**, 17339-17346.
- ⁴ Hartmann, S., Kyewski, B., Sonnenburg, B., & Lucius, R. (1997). A filarial cysteine protease inhibitor down-regulates T cell proliferation and enhances interleukin-10 production. *Eur. J. Immunol.* **27**, 2253-2260.
- ⁵ Rawlings, N.D., Morton, F.R. & Barrett, A.J. (2006). *MEROPS*: the peptidase database. *Nucleic Acids Res.* **34**, D270-D272.
- ⁶ Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J. & Turk, V. (1988). The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *EMBO J.* **7**, 2593-2599.
- ⁷ Lindahl, P., Abrahamson, M. & Björk, I. (1992). Interaction of recombinant human cystatin C with the cysteine proteinases papain and actinidin. *Biochem. J.* **281**, 49-55.
- ⁸ Bevec, T., Stoka, V., Pungercic, G., Dolenc, I. & Turk, V. (1996). Major histocompatibility complex class II-associated p41 invariant chain fragment is a strong inhibitor of lysosomal cathepsin L. *J. Exp. Med.* **183**, 1331-1338.
- ⁹ Guncar, G., Pungercic, G., Klemencic, I., Turk, V. & Turk, D. (1999). Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO J.* **18**, 793-803.
- ¹⁰ Irvine, J. W., Coombs, G. H. & North, M. J. (1992). Cystatin-like cysteine proteinase inhibitors of parasitic protozoa. *FEMS Microbiol Lett.* **75**, 67-72.
- ¹¹ Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Wheeler, D.L. (2006). GenBank. *Nucleic Acids Res.* **34**(Database issue), D16-20.
- ¹² Monteiro, A. C. S., Abrahamson, M., Lima, A. P. C. A., Vannier-Santos, M. A. & Scharfstein, J. (2001). Identification, characterization and localization of chagasin, a tight-binding cysteine proteases inhibitor in *Trypanosoma cruzi*. *J. Cell Sci.* **114**, 3933-3942
- ¹³ Rigden, D.J., Monteiro, A.C. & Grossi de Sa, M.F. (2001). The protease inhibitor chagasin of

Trypanosoma cruzi adopts an immunoglobulin-type fold and may have arisen by horizontal gene transfer. *FEBS Lett.* **504**, 41-44.

¹⁴ Salmon, D., do Aido-Machado, R., Diehl, A., Leidert, M., Schmetzer, O., de Lima, A.A.P., Scharfstein, J., Oschkinat, H. & Pires, J.R. (2006). Solution structure and backbone dynamics of the *Trypanosoma cruzi* cysteine protease inhibitor chagasin. *J. Mol. Biol.* **357**, 1511-1521.

¹⁵ Figueiredo da Silva, A.A., Carvalho Vieira, L.D., Krieger, M.A., Goldenberg, S., Tonin Zanchin, N.I. & Guimaraes, B.G. (2007). Crystal structure of chagasin, the endogenous cysteine-protease inhibitor from *Trypanosoma cruzi*. *J. Struct. Biol.* [Epub ahead of print].

¹⁶ Smith, B.O., Picken, N.C., Westrop, G.D., Bromek, K., Mottram, J.C. & Coombs, G.H. (2006). The structure of *Leishmania mexicana* icp provides evidence for convergent evolution of cysteine peptidase inhibitors. *J. Biol. Chem.* **281**, 5821-5828.

¹⁷ Nicklin, M.J. & Barrett, A.J. (1984). Inhibition of cysteine proteinases and dipeptidyl peptidase I by egg-white cystatin. *Biochem J.* **223**, 245-253.

¹⁸ Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J.S. & Cygler, M. (1996). Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.* **15**, 5492-5503.

¹⁹ Groves, M.R., Coulombe, R., Jenkins, J. & Cygler, M. (1998). Structural basis for specificity of papain like cystein protease proregions toward their cognate enzymes *Proteins*, **32**, 504-514.

²⁰ UniProtKB/Swiss-Prot P07711 (CATL_HUMAN) Cathepsin L precursor (E.C. 3.4.22.15).

²¹ Stearns, N.A., Dong, J.M., Pan, J.X., Brenner, D.A. & Sahagian, G.G. (1990). Comparison of cathepsin L synthesized by normal and transformed cells at the gene, message, protein, and oligosaccharide levels. *Arch. Biochem Biophys.* **283**, 447-57.

²² Maes, D., Bouckaert, J., Poortmans, F., Wyns, L. & Looze, Y. (1996). Structure of chymopapain at 1.7 Å resolution. *Biochemistry*, **35**, 16292-16298.

²³ Abrahamson, M., Alvarez-Fernandez, M. & Nathanson, C-M. (2003). Cystatins. *Biochem. Soc. Symp.* **70**, 179-199.

²⁴ Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. & Turk, V. (1990). The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO J.* **9**, 1939-1947.

²⁵ Jenko, S., Dolenc, I., Guncar, G., Dobersek, A., Podobnik, M. & Turk, D. (2003). Crystal structure of stefin A in complex with cathepsin H: N-terminal residues of inhibitors can adapt to the active sites of endo- and exopeptidases. *J. Mol. Biol.* **326**, 875-885.

- ²⁶ Huang, L., Brinen, L.S. & Ellman, J.A. (2003). Crystal structures of reversible ketone-based inhibitors of the cysteine protease cruzain. *Bioorg. Med. Chem.* **11**, 21-29.
- ²⁷ Jeppsson, J. O., Laurell, C. B. & Franzén, B. (1979). Agarose gel electrophoresis. *Clin. Chem.* **25**, 629-638.
- ²⁸ Kozak, M., Jankowska, E., Janowski, R., Grzonka, Z., Grubb, A., Alvarez Fernandez, M., Abrahamson, M. & Jaskolski, M. (1999). Expression of a selenomethionine derivative and preliminary crystallographic studies of human cystatin C. *Acta Cryst. D*, **55**, 1939-1942.
- ²⁹ Kunkel, T.A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA*, **82**, 488-492.
- ³⁰ Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J.S. & Menard, R. (1996). Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry*, **35**, 8149-8157.
- ³¹ Menard, R., Carmona, E., Takebe, S., Dufour, E., Plouffe, C., Mason, P. & Mort, J.S. (1998). Autocatalytic processing of recombinant human procathepsin L. Contribution of both intermolecular and unimolecular events in the processing of procathepsin L *in vitro*. *J. Biol. Chem.* **273**, 4478-4484.
- ³² Vincents, B., Önnarfjord, P., Potempa, J. & Abrahamson, M. (2007). Down-regulation of human extracellular cysteine protease inhibitors by the secreted staphylococcal cysteine proteases, staphopain A and B. *Biol Chem.* (in press).
- ³³ Abrahamson, M. (1994). Cystatins. *Methods Enzymol.* **244**, 685-700.
- ³⁴ Henderson, P. J. F. (1972). A linear equation that describes the steady-state kinetics of enzymes and subcellular particles interacting with tightly bound inhibitors. *Biochem. J.* **127**, 321-333
- ³⁵ Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
- ³⁶ Murshudov, G.N., Vagin, A. & Dodson, E. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D*, **53**, 240-255.
- ³⁷ Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta. Cryst. D*, **59**, 1131-1137.
- ³⁸ Emsley, P. & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics *Acta Cryst. D*, **60**, 2126-2132.

- ³⁹ Perrakis, A., Harkiolaki, M., Wilson, K.S. & Lamzin, V.S. (2001). ARP/wARP and molecular replacement. *Acta Crystallogr. D. Biol Crystallogr.* **57**, 1445-1450.
- ⁴⁰ Brünger, A. T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature.* **355**, 472-474.
- ⁴¹ Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
- ⁴² Vagin, A.A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallog.* **30**, 1022-1025.
- ⁴³ DeLano, W.L. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. <http://www.pymol.org>
- ⁴⁴ Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.
- ⁴⁵ Cohen, G.H. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallogr.* **30**, 1160-1161.

Fig. 1a

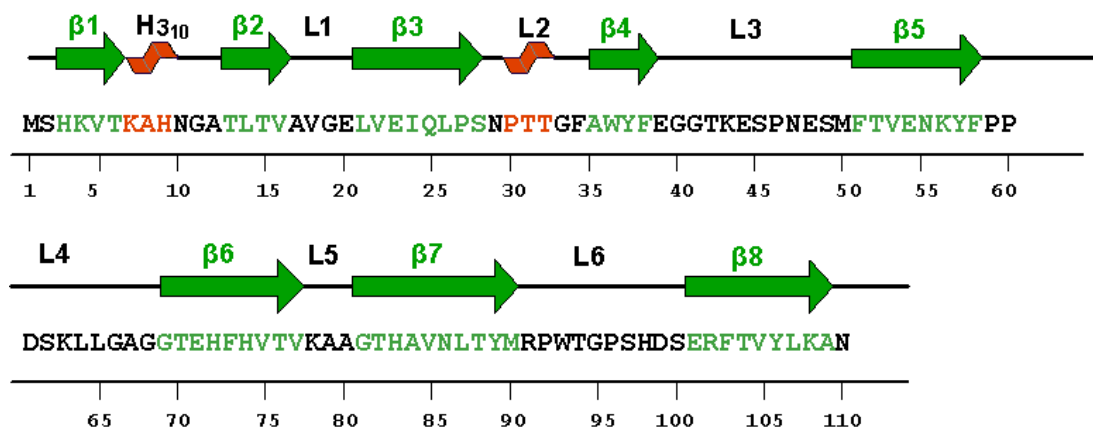


Fig. 1b



Fig. 1c

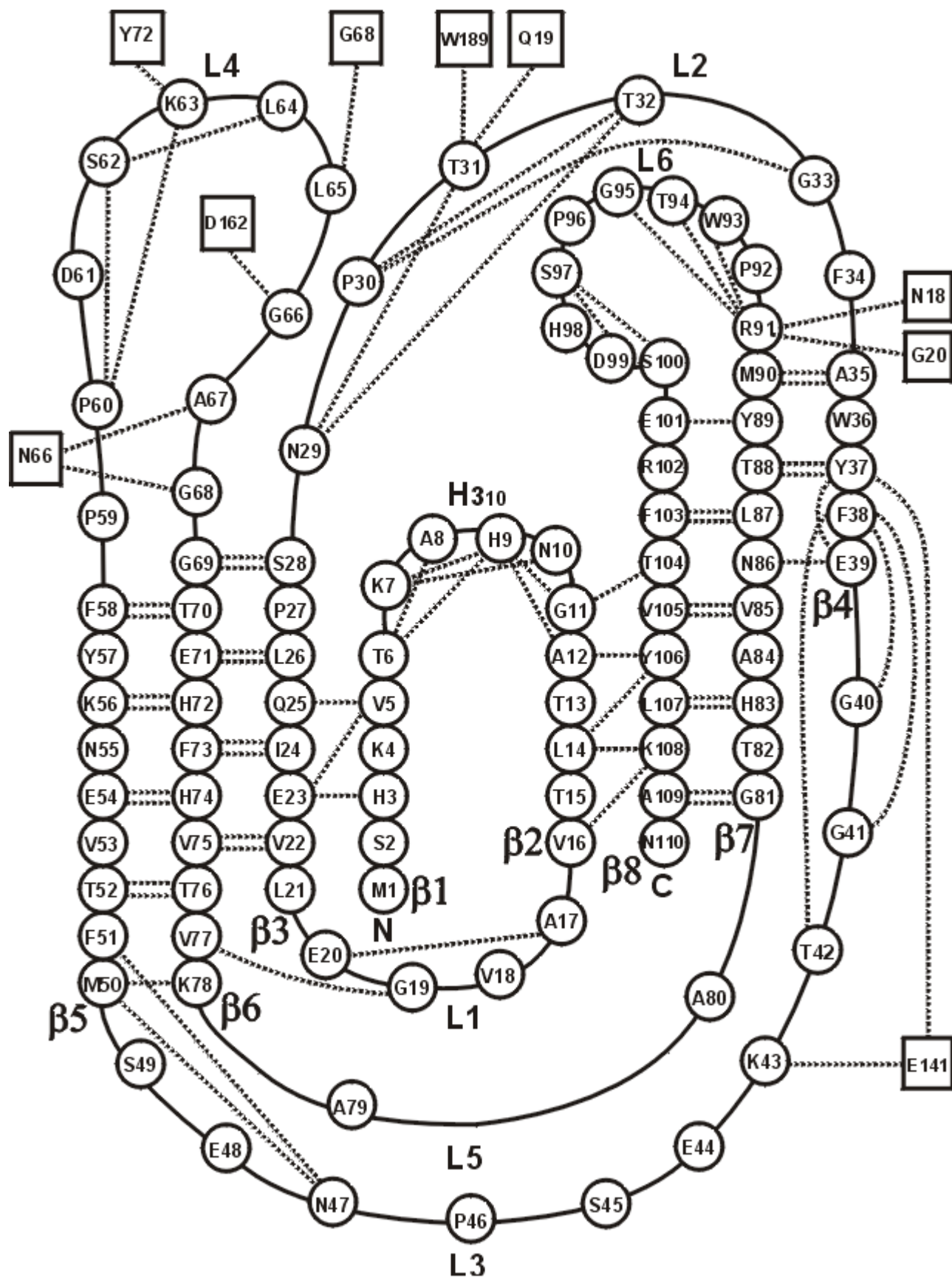


Fig. 2a

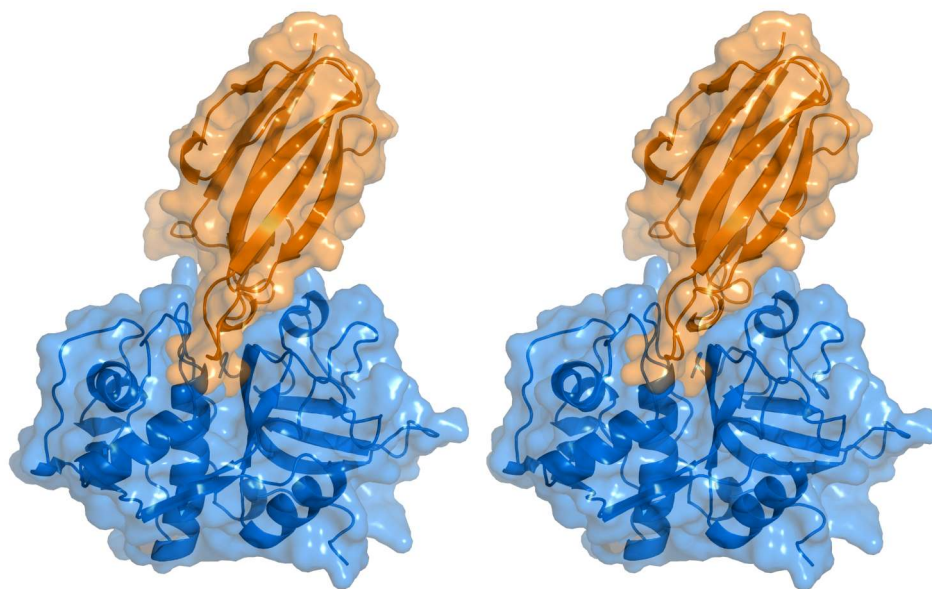


Fig. 2b

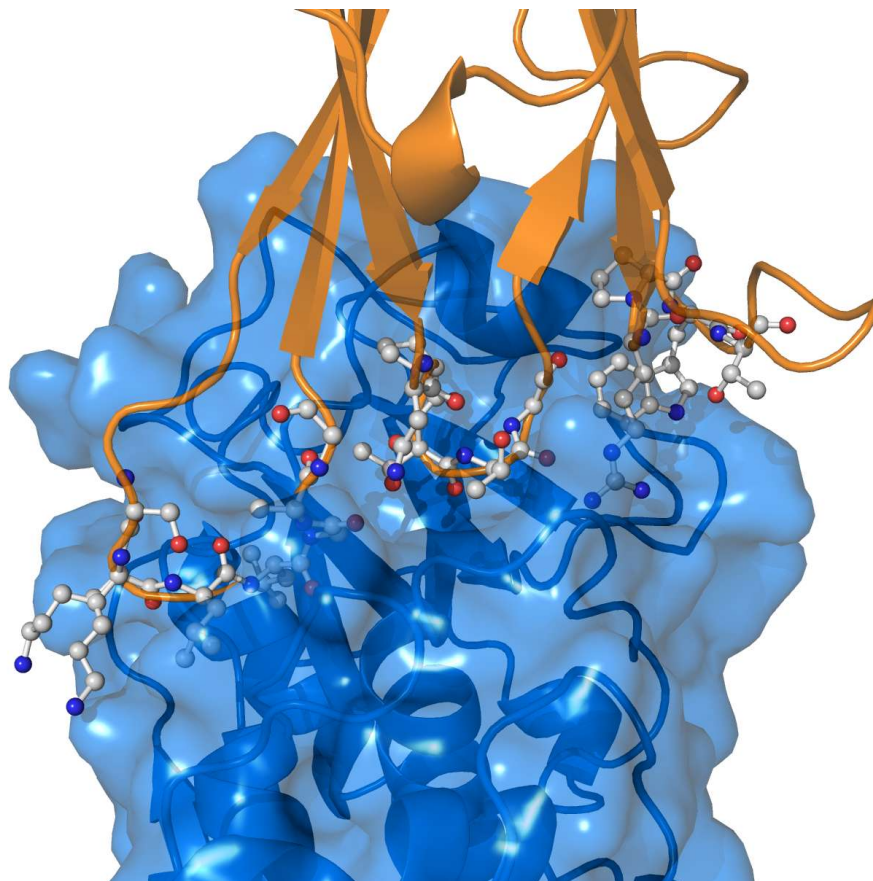


Fig. 2c

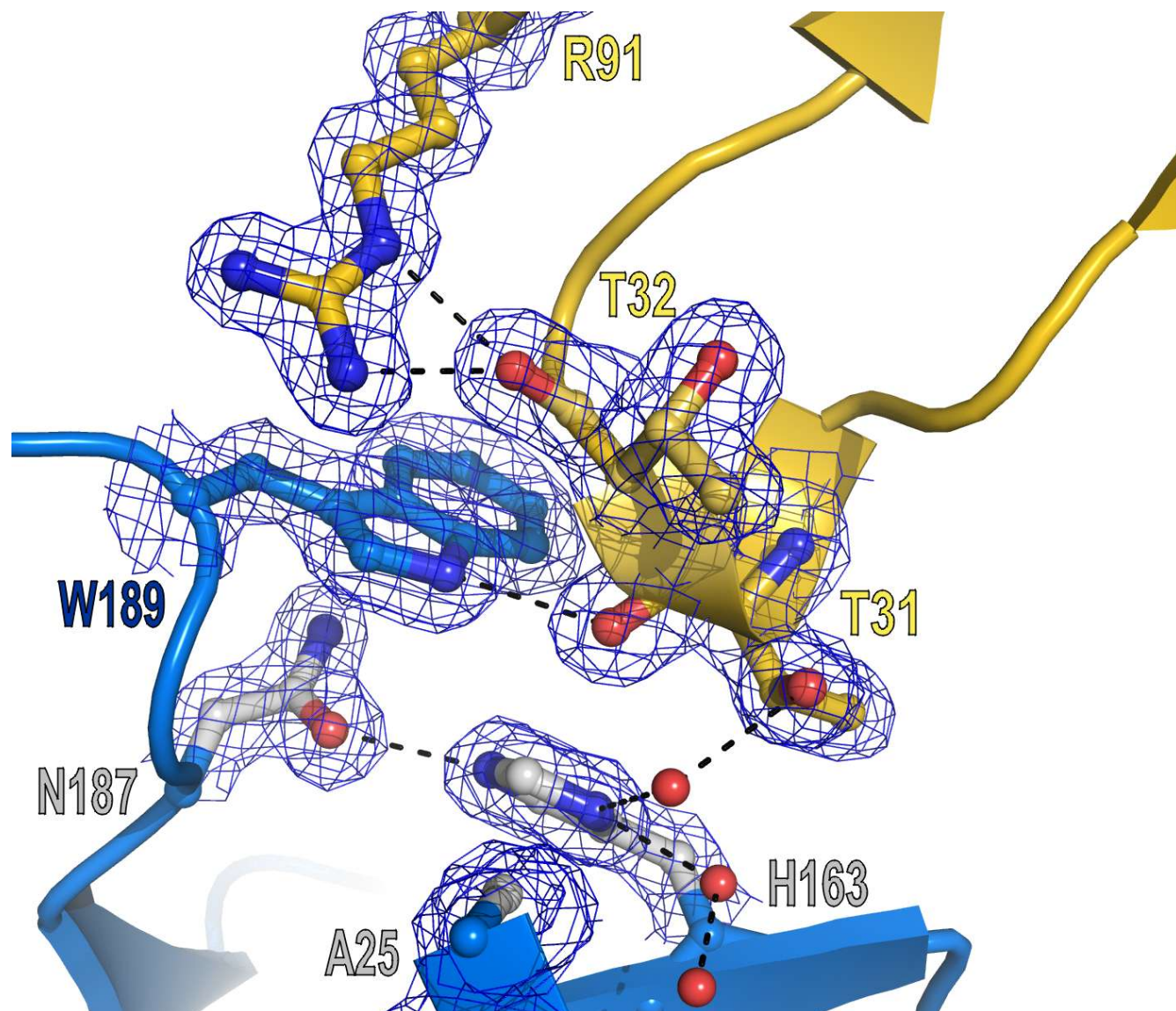


Fig. 2d

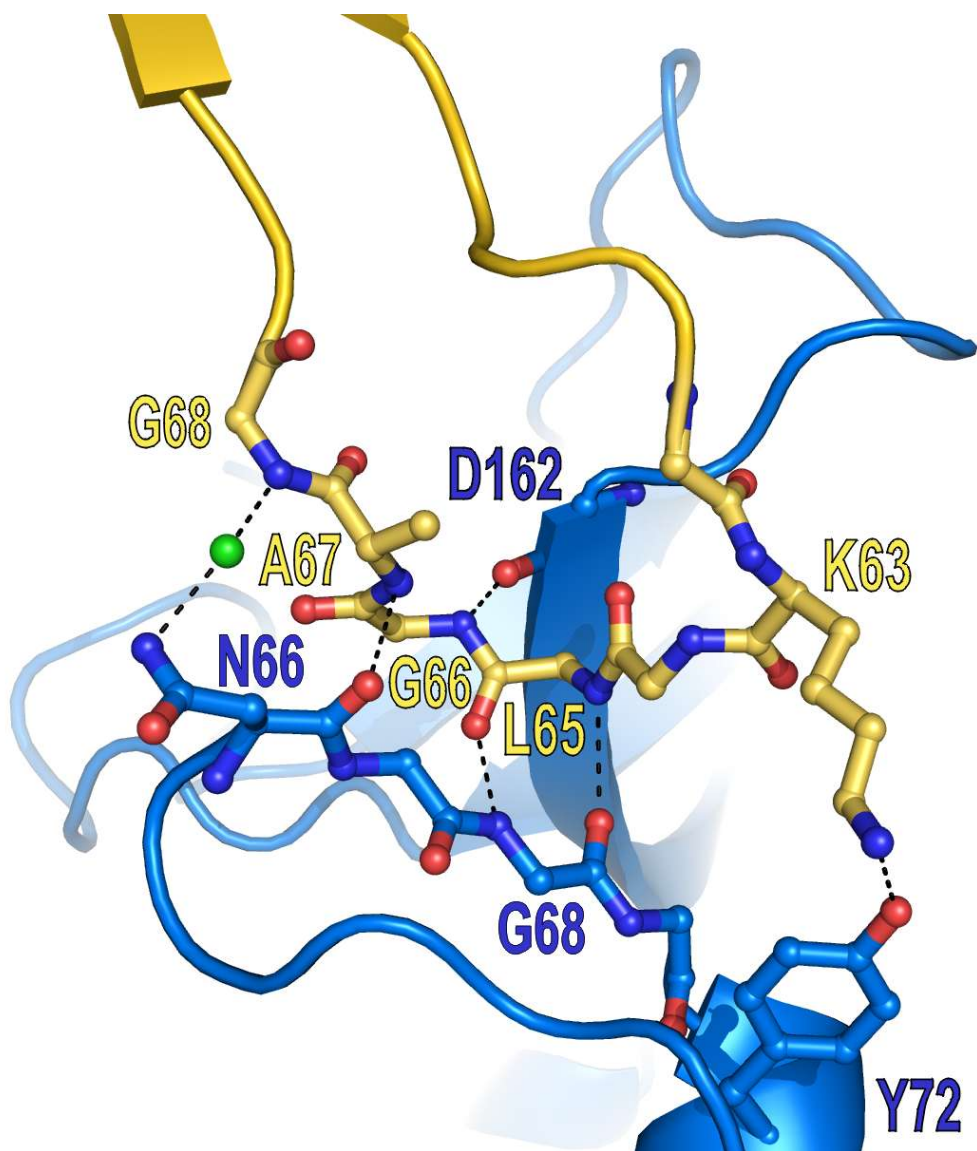


Fig. 2e

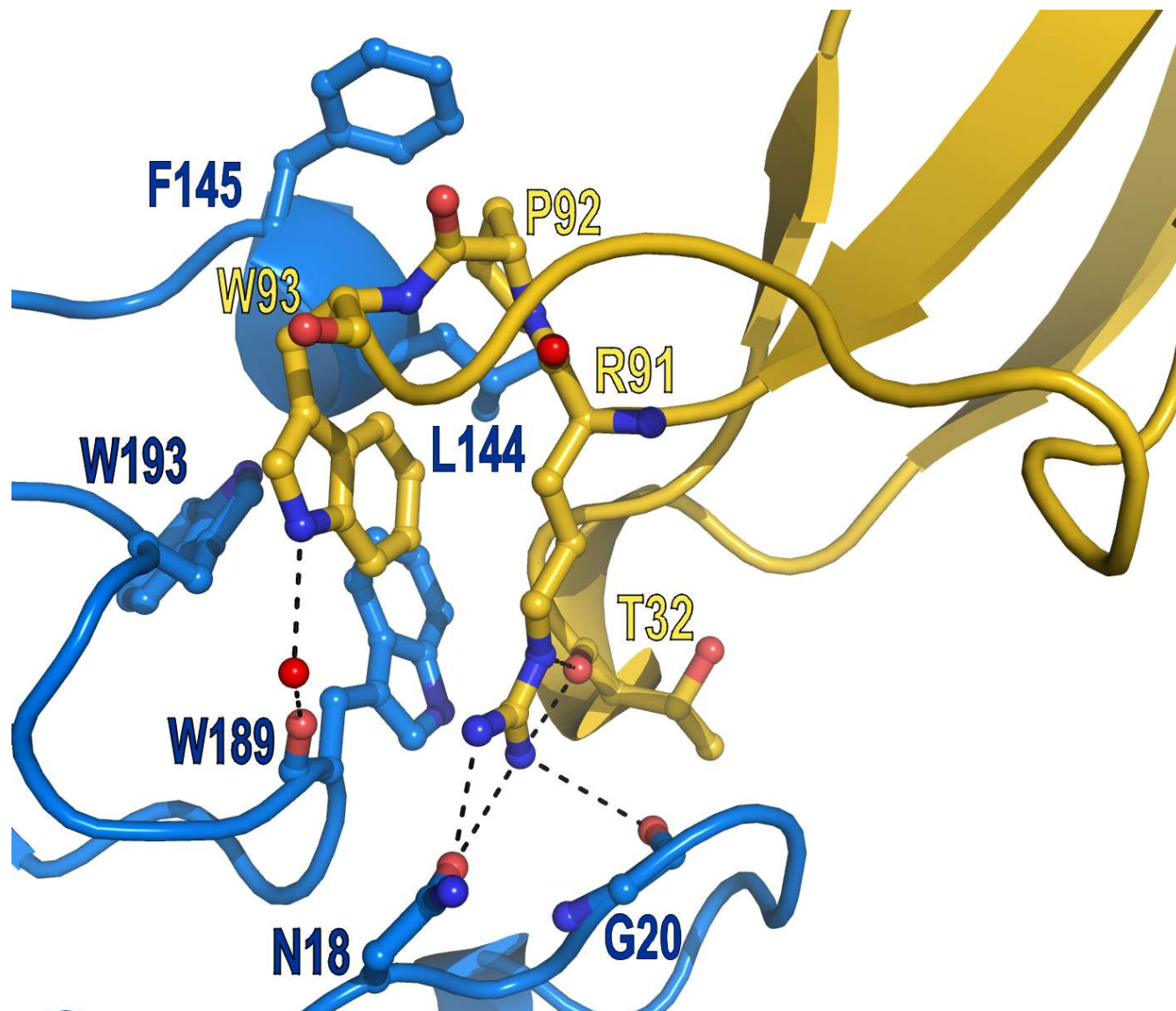


Fig. 3a

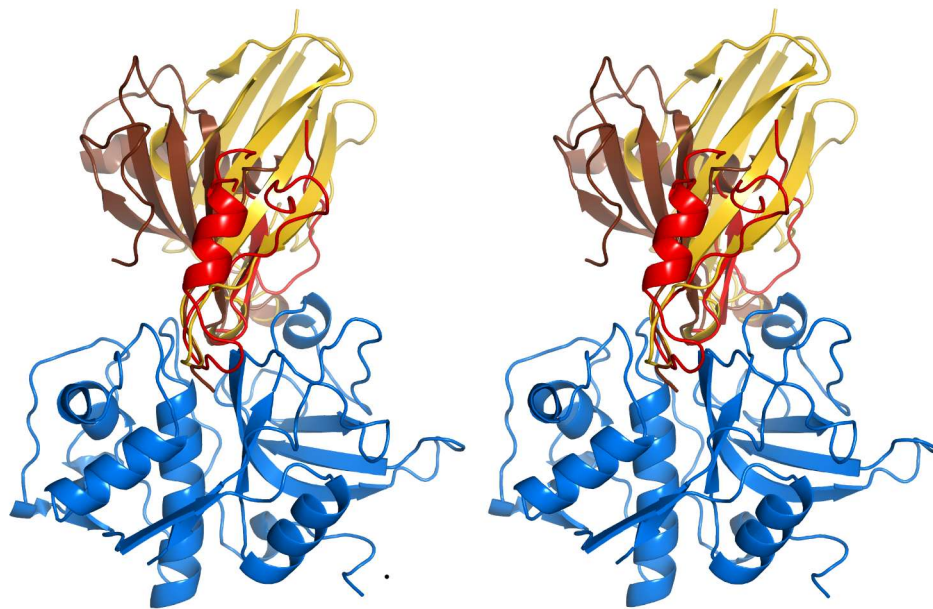


Fig. 3b

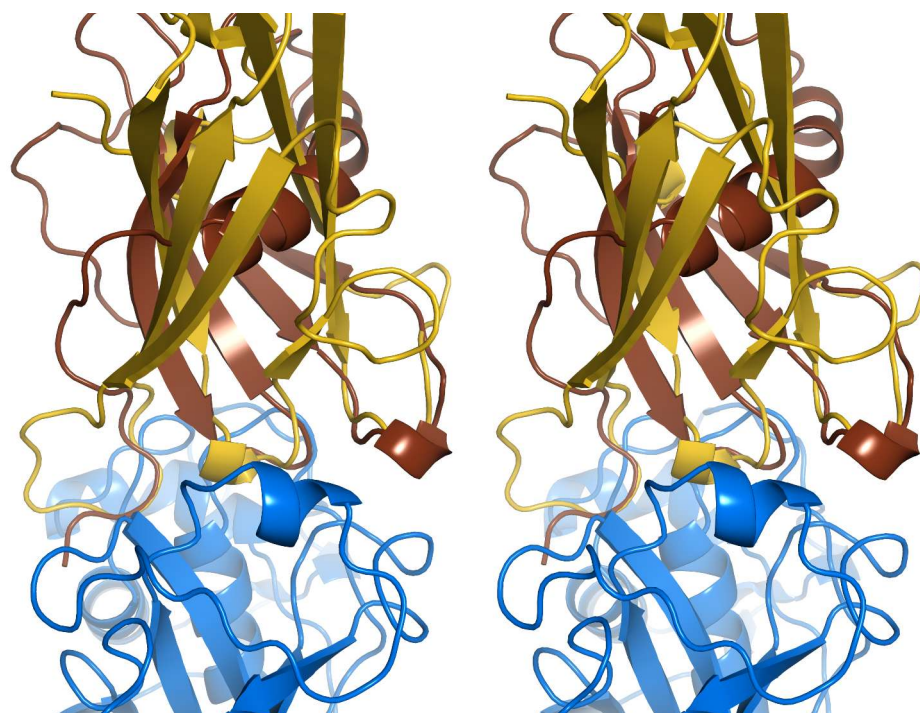


Fig. 3c

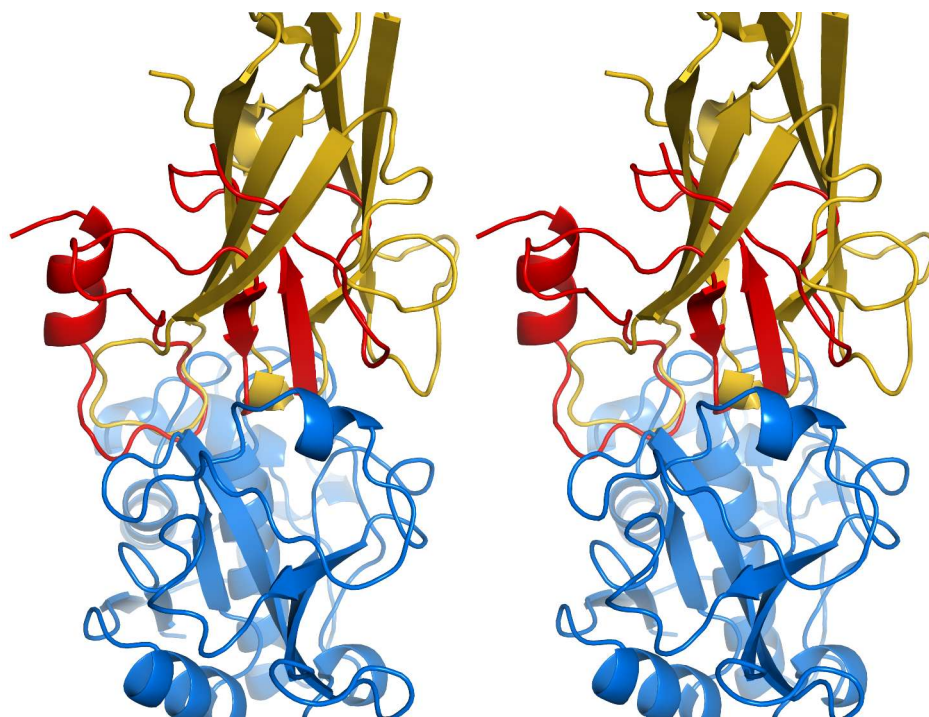


Fig. 4a

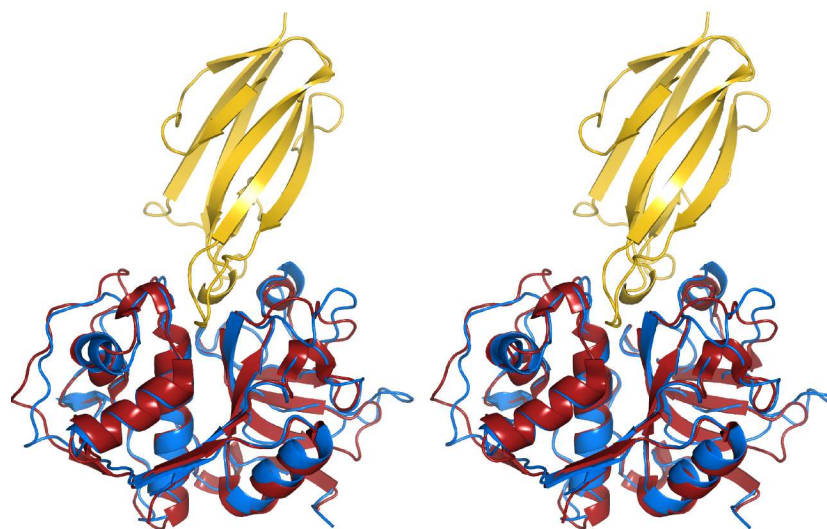


Fig. 4b

