Structure, interdomain dynamics, and pH-dependent autoactivation of pro-rhodesain, the main lysosomal cysteine protease from African trypanosomes

Patrick Johé[1], Elmar Jaenicke[2], Hannes Neuweiler[3], Tanja Schirmeister[1], Christian Kersten[1,*], and Ute A. Hellmich[4,5,*]

From the [1]Institute of Pharmaceutical and Biomedical Sciences, [2]Institute for Molecular Physiology, Johannes Gutenberg-University, Mainz, Germany; [3]Department for Biotechnology and Biophysics, Julius-Maximilians-University, Würzburg, Germany; [4]Department of Chemistry, Biochemistry Division, Johannes Gutenberg-University, Mainz, Germany; [5]Centre for Biomolecular Magnetic Resonance (BMRZ), Goethe University, Frankfurt, Germany

Edited by George DeMartino

Rhodesain is the lysosomal cathepsin L-like cysteine protease of *Trypanosoma brucei rhodesiense*, the causative agent of Human African Trypanosomiasis. The enzyme is essential for the proliferation and pathogenicity of the parasite as well as its ability to overcome the blood–brain barrier of the host. Lysosomal cathepsins are expressed as zymogens with an inactivating prodomain that is cleaved under acidic conditions. A structure of the uncleaved maturation intermediate from a trypanosomal cathepsin L-like protease is currently not available. We thus established the heterologous expression of *T. brucei rhodesiense* pro-rhodesain in *Escherichia coli* and determined its crystal structure. The trypanosomal prodomain differs from nonparasitic pro-cathepsins by a unique, extended α-helix that blocks the active site and whose side-chain interactions resemble those of the antiprotozoal inhibitor K11777. Interdomain dynamics between pro- and core protease domain as observed by photoinduced electron transfer fluorescence correlation spectroscopy increase at low pH, where pro-rhodesain also undergoes autocleavage. Using the crystal structure, molecular dynamics simulations, and mutagenesis, we identify a conserved interdomain salt bridge that prevents premature intramolecular cleavage at higher pH values and may thus present a control switch for the observed pH sensitivity of proenzyme cleavage in (trypanosomal) CathL-like proteases.

Human African Trypanosomiasis (HAT) or African Sleeping Sickness is a so-called neglected tropical disease (NTD) as classified by the World Health Organization (www.who.int). HAT is caused by two subspecies of African trypanosomes, *Trypanosoma brucei gambiense* and *T. brucei rhodesiense* (1). These unicellular, protozoan parasites are transmitted to the host *via* the bite of a Tse-Tse fly. The disease progresses in two stages. In the early hemolymphatic stage, the parasites are present in the blood and in the lymphatic systems leading to unspecific symptoms such as fever and headache. In the later meningoencephalitic stage, the parasites cross the blood–brain barrier (BBB) and enter the central nervous system of the host. At this point, patients suffer from severe neurological symptoms, deregulation of sleep–wake cycles, coma, and ultimately death (2). The few available treatment options often have severe side effects (3).

T. brucei spp. parasites express two main lysosomal cysteine proteases that belong to the papain family, the cathepsin B-like *T. brucei* cathepsin B (TbCathB) and the cathepsin L-like rhodesain (also called TbCathL, brucipain, or trypanopain). Rhodesain plays an important role in the progression into the second disease stage as it is involved in the crossing of parasites into the central nervous system *via* the BBB of the host (4, 5). The expression of TbCathB mRNA varies strongly during the life cycle of the parasite, while rhodesain is constitutively expressed (6). In agreement with an important role in parasite survival, proliferation, and pathogenicity, RNAi-based rhodesain knockdown or inhibition of rhodesain but not TbCathB led to diminished parasitic growth and an increased sensibility to lytic factors in human serum (7, 8). Rhodesain is thus generally assumed to be the main lysosomal cysteine protease in trypanosomes and presents a promising antitrypanosomal drug target. In general, proteases are attractive drug targets because the affinities of inhibitors toward active site residues can be tuned and potential inhibitors can easily be identified in high-throughput *in vitro* assay screenings (9). Nevertheless, structural similarities to host proteins can lead to problems with off-target effects. Rhodesain shares 46% (60%) sequence identity (similarity) with human cathepsin L (hsCathL). The irreversible, vinylsulfone-based rhodesain inhibitor K11777 (*N*-methylpiperazine-urea-Phe-homophenylalanine-vinyl-sulfone-benzene), was shown to prevent *T. brucei* from overcoming a barrier formed by brain microvascular endothelial cells in an *in vitro* BBB model (4, 5). Encouragingly, while K11777 was indeed observed to not be biochemically selective, it was nontoxic to mammals, efficient against related

* For correspondence: Christian Kersten, kerstec@uni-mainz.de; Ute A. Hellmich, ute.hellmich@uni-jena.de.
Present address for Ute A. Hellmich: Faculty of Chemistry and Earth Sciences, Institute of Organic Chemistry and Macromolecular Chemistry, Friedrich-Schiller-University, Jena, Germany and Cluster of Excellence 'Balance of the Microverse', Friedrich-Schiller-University, Jena, Germany.

protozoans, and could cure mice from a *Trypanosoma cruzi* infection (10–12).

In the cell, cathepsins are expressed as inactive zymogens where the ~215 amino acid papain-fold core protease domain is preceded by a ~20 amino acid signal peptide required for translation of the protein into the endoplasmatic reticulum, and a ~100 amino acid prodomain required for correct folding and lysosomal targeting of the protease (13–17). The prodomain interacts with the active site of the protease and keeps it in an autoinhibited state until it has been successfully trafficked to the acidic lysosome lumen. Here, the prodomain is removed by proteolytic cleavage, which may occur intra- or intermolecularly (18–21). The cleavage site is located in the linker between the prodomain and the core protease domain (22, 23). While the rhodesain core protease domain has been crystallized in complex with different inhibitors including K11777 (PDB: 2P7U) (23–26), a structure of the zymogen as an important intermediate of the parasitic protease maturation process is currently not available and the molecular details of the pH-dependent cleavage process remain unclear.

Here, we present the 2.8 Å crystal structure of the active site C150A mutant of the uncleaved *T. brucei rhodesiense* cysteine protease pro-rhodesain under acidic conditions. The rhodesain prodomain shares many structural features with other CathL prodomains, but is unique in the presence of an extended α-helix in its C-terminal end. Using photoinduced electron transfer–fluorescence correlation spectroscopy (PET-FCS) (27), we investigated the interdomain dynamics of the proenzyme. At low pH values, where the protease undergoes autocleavage, the "blocking peptide" region of the prodomain covering the protease active site shows increased dynamics. We further found evidence that pro-rhodesain can be processed both intra- and intermolecularly in a pH-dependent manner and identified a highly conserved interdomain salt bridge that can prevent premature intramolecular cleavage at high pH and may thus act as a "delay switch" for pro-cathepsin autocleavage.

Results

Heterologous expression and purification of pro-rhodesain

Cathepsin-like cysteine proteases are expressed as an inactive proform. pH-dependent cleavage of the prodomain from the globular core protease domain is an important step in the protease maturation process contingent on the correct localization of the protein to the lysosome. For a better understanding of pro-rhodesain inhibition and activation, we determined the crystal structure of *T. brucei rhodesiense* pro-rhodesain. While rhodesain expression in *P. pastoris* relies on the secretion of the folded protein (22, 25), all currently available expression and purification protocols for rhodesain from *Escherichia coli* are based on its expression in inclusion bodies and subsequent protein refolding (28). To obtain the folded proenzyme for structural studies, we found both production routes to be ineffective. We thus optimized the purification of mature and pro-rhodesain from *E. coli* in a manner that circumvents secretion and refolding (see Figs. S1–S4 and Extended Materials and Methods for details).

Purified mature rhodesain proteins from *P. pastoris* and *E. coli* have similar structural and functional properties as elucidated by size-exclusion chromatography (SEC), circular dichroism (CD) spectroscopy (Fig. 1, *A* and *B*), and a fluorescent peptide-cleavage assay using the fluorescent peptide Z-Phe-Arg-AMC (Z-phenylalanine-arginine-7-amido-4-methylcoumarin) to probe rhodesain activity (see Extended Materials and Methods) (9). *E. coli* can thus be used as a reliable source for the production of active rhodesain for optimal flexibility in biophysical studies. To prevent autocleavage and to stabilize the proenzyme, we introduced a mutation into the active site of the protein, C150A. Pro-rhodesain C150A displays the expected increase in molecular weight during SEC compared with the mature protease as well as an additional small increase in α-helical content in CD spectroscopy measurements (Fig. 1, *B* and *C*).
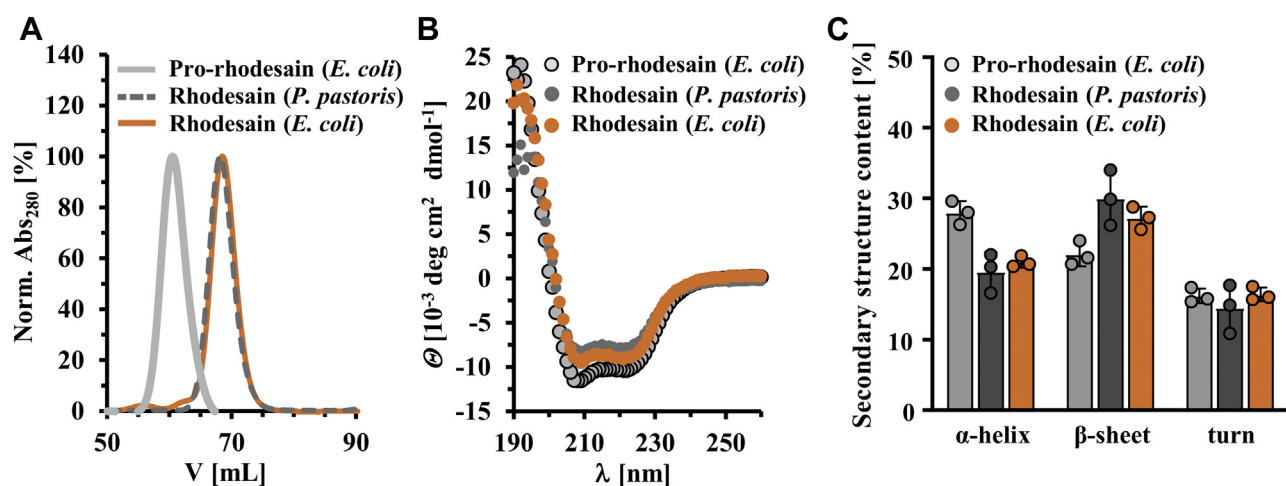

Figure 1. Comparison of heterologously expressed rhodesain purified from *P. pastoris* or *E. coli*. *A*, size-exclusion chromatography of mature rhodesain from *P. pastoris* (*dashed-lines*, *dark gray*) or *E. coli* (*orange*) and pro-rhodesain C150A from *E. coli* (*light gray*). *B*, circular dichroism spectra of mature rhodesain (*P. pastoris*, *dark gray*; *E. coli*, *orange*) as well as pro-rhodesain C150A (*light gray*). *C*, secondary structure analysis of CD spectra shown in (*B*) using the BeStSel server (http://bestsel.elte.hu/index.php) (68).
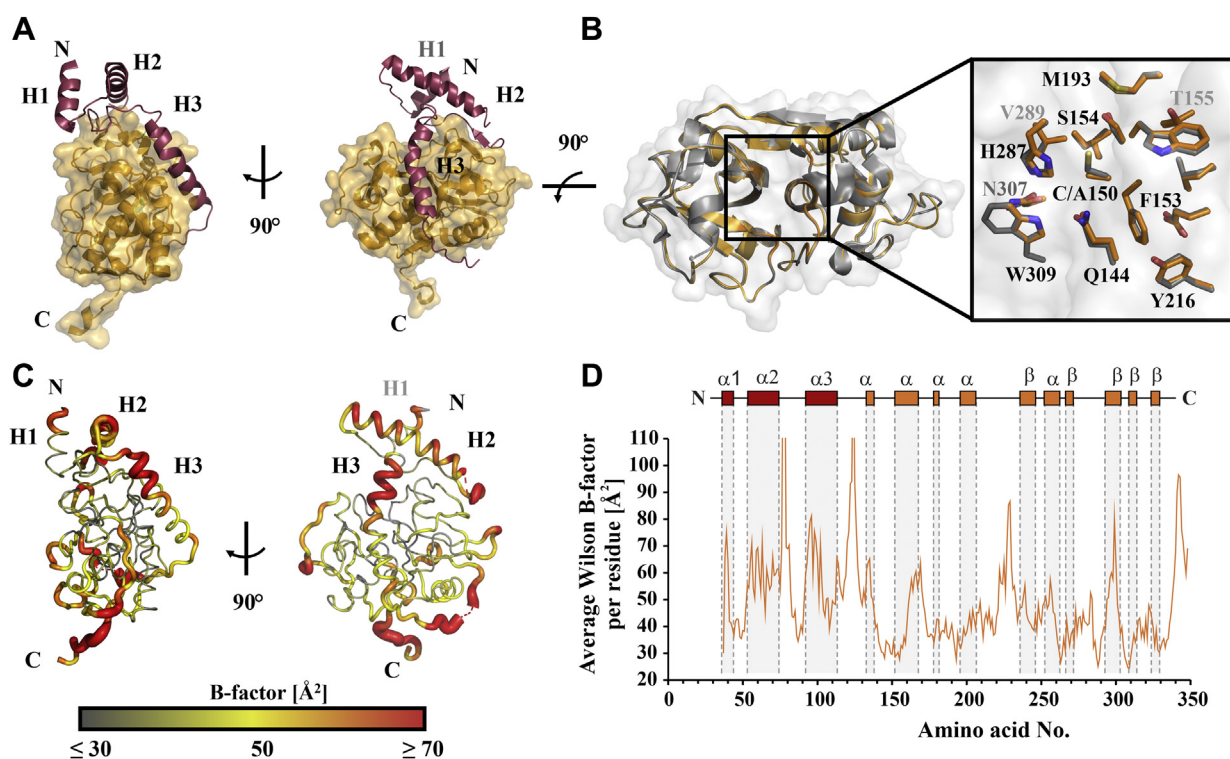
Figure 2. Crystal structure of *T. brucei rhodesiense* pro-rhodesain. *A*, 2.8 Å crystal structure of pro-rhodesain consisting of the prodomain (*red*) with three α-helices, H1, H2, and H3 and the core protease domain (*gold*). *B*, comparison of pro-rhodesain (*gold*) and mature rhodesain (*gray*, PDB: 2P7U) (25) including a close-up of the active site. Side chains within a distance of 8 Å around the active site cysteine or the C150A mutation are shown as *sticks*. *C* and *D*, average B factors per residue within the pro-rhodesain structure. Many low values (≤30 Å^2) in the catalytic domain indicate that it is mostly rigid with the exception of some flexible loop regions, thus resulting in an overall average B factor of 44.2 Å^2. In the prodomain, the average B factor is 55.6 Å indicating higher flexibility of the propeptide compared with the core protease domain. The residues with the highest B factors reside in the helix covering the active site cleft (H3) and the unstructured peptide chain connecting the pre- and the core protease domains.

T. b. rhodesiense *pro-rhodesain crystal structure*

We obtained a 2.8 Å crystal structure for *T. b. rhodesiense* pro-rhodesain at acidic pH (Fig. 2, Table 1, Fig. S5). The proprotease structure contains two parts, the N-terminal prodomain (residues E38 to T123 according to uniprot numbering, ID.: Q95 PM0) and the C-terminal core protease domain (residues A126 to P342). The first 18 amino acids belonging to the prodomain, a short loop between the first two helices (residues 77/78) and residues G124/R125 at the C-terminus of the prodomain, are not resolved. High disorder, in agreement with the crystallographic B factors in the C-terminal prodomain region (Fig. 2, *C* and *D*), coincides with the expected intermolecular cleavage site of pro-rhodesain at position T123 or R125 (22, 23). The entire C-terminus of the core protease up to residue P342 as well as the remaining residues of a TEV cleavage site (ENLYFQ), which was used to remove the C-terminal purification tag, could be reliably placed in the electron density. The remainder of the TEV cleavage site mediates crystal contacts to the neighboring protein's prodomain and may have thus aided crystallization of the construct but does not interfere with the interaction of the prodomain with its corresponding catalytic domain.

The catalytic core domain follows the typical two-lobed fold of the papain family of cysteine proteases with an α-helical L-domain and a β-sheet containing R-domain. The cleft between these domains harbors the active site triad residues C150 (mutated to alanine in our construct), H287 and N307 (Fig. 2B).

In the same position as in papain (PDBs: 9PAP, 3TNX) (29, 30), pro-rhodesain contains three intramolecular disulfide bonds between residues C147-C188 and C181-C226 in the L-domain as well as between C280–C328 in the R-domain (Fig. S5). As

Table 1

Crystallographic data collection and refinement statistics for *T. brucei rhodesiense* pro-rhodesain C150A

PDB entry	7AVM
Data collection	
Wavelength	1.54179
Space group	H3
Cell parameters	
a; b; c	123.19 Å; 123.19 Å; 53.91 Å
α; β; γ	90.00°; 90.00°; 120°
Resolution range [Å]	37.92–2.8 (2.9–2.8)
Unique reflections	7512 (729)
Multiplicity	20.6 (19.8)
Overall completeness [%]	97.23 (92.73)
$<I/\sigma(I)>$	3.63 (at 2.81 Å)
Refinement	
R_{work}, R_{free}	0.255, 0.264
R_{merge}	0.611
Average B-Factor [Å^2]	41.2
Protein residues	308
No. of solvent molecules	38
Ligand atoms	6
Total non-H atoms	2379
RMSD from ideal geometry	
Bond lengths [Å]	0.010
Bond angles [°]	1.60
Ramachandran	
Favored	91.7%
Allowed	8.3%
Outliers	0%

Numbers in parentheses characterize the highest-resolution shell.

apparent from the electron density, these residues all form disulfide bridges. The backbone RMSD between our structure of the pro-rhodesain core catalytic domain (PDB: 7AVM, residues 125–342) and the previously determined structure of mature rhodesain (PDB: 2P7U) (25) is 0.39 Å, thus the presence of the prodomain does not affect the structure of the catalytic domain, including the relative side-chain orientations of the residues of the active site catalytic triad (Fig. 2B).

The prodomain constitutes about one-third of the protein in immature cathepsin proteases and consists of three helices, H1–H3 (Fig. 2). In pro-rhodesain, the N-terminal helix 1 (H1, residues E38^{pro} -K47^{pro}) is shorter compared with H1 from other crystallized CathL zymogens. The lack of resolution for the first 18 amino acids indicates that this region may be highly flexible and potentially not structured in pro-rhodesain. H1 is connected to the orthogonal helix 2 (H2, residues A55^{pro}-A76^{pro}) *via* a long loop. H2 then leads into a long third helix (H3, residues E96^{pro}-K113^{pro}). Helix 3 covers the active site and constitutes the main difference between the prodomain of rhodesain and other CathL-like protease prodomains (see below).

CathL and CathB proteases both belong to the CA clan and the C1 family of proteases (31) and thus share a papain-like fold. Nevertheless, their prodomains differ. For instance, the conserved ER(I/V)FNIN motif is only found in CathL-like precursors within helix 2 of the prodomain and is slightly modified in the trypanosomal CathL proteases (32). Here, the ER(A/V)FNAA motif (consensus sequence $Ex_3Rx_2(A/V)(F/W)x_2Nx_3Ax_3A$, with x = any amino acid) (Fig. 3, Fig. S6) is involved in the correct folding of the prodomain and mediates intradomain interactions (33, 34). In most zymogens, H2 is longer than in our pro-rhodesain structure, and we do not observe density for the C-terminal end of this region (residues 77^{pro}–78^{pro}). Together with the elevated B-factor values for the C-terminal end of H2 indicate that this region may be flexible in the prodomain of pro-rhodesain as has been stated for human CathL (35).

H2 connects to Helix 3 (H3, residues 96^{pro}–113^{pro}) *via* a long linker. This linker partially forms a short antiparallel β-sheet (residues 82^{pro}–84^{pro}) with the propeptide binding loop (PBL, residues 271–275, Fig. 3C), a loop extending from the R subdomain of the catalytic core domain. Notably, both the short β-sheet in the H2/H3 linker of the prodomain and the interacting PBL in the protease domain display very low B-values indicating high rigidity (Fig. 2, C and D). In cathepsin prodomains, the β-strand in the H2/H3 linker is typically followed by a GNFD motif with the GxNxFxD (x = any amino acid) consensus sequence, which is required for proper protein folding and autoactivation (36). In pro-rhodesain, this sequence corresponds to the GNFD-like motif $G_{85}VT_{87}PF_{89}SD_{91}$. Residues in this motif form multiple contacts within both the prodomain and to the core protease domain (Fig. 3D). Our structure may thus provide some insights why the core protease domain's PBL β-strand harbors a di-glycine motif (G274/G275) that is highly conserved among CathL proteases (Fig. S6) and why this region could be crucial for protease folding (37) as larger amino acid side chains may sterically interfere with the PBL interactions (Fig. 3, C and D).

An interaction network involving mostly hydrophobic and aromatic residues positions the remainder of the prodomain H2/H3 linker whose P88^{pro} and F89^{pro} side chains dip into a hydrophobic basket formed by the side chains of F269, M270, W309, and W313 in the R-domain of the core protease (Fig. 3D). In addition, the backbone carbonyl of M270 forms a hydrogen bond with the side chain of T87^{pro} while its side chain forms a hydrophobic interaction with F97^{pro} in the first turn of H3, the long α-helix covering the entire length of the cleft between the L- and R-domains of the catalytic domain. In all available structures of other cathepsin proenzymes, this region is mainly unstructured (Fig. 4A). This α-helix therefore presents the most notable structural difference between pro-rhodesain and closely related nontrypanosomal proteases.

Helix H3 in pro-rhodesain differs from that of other nonparasitic pro-cathepsins

The residues in H3 facing the catalytic domain are involved in numerous, mostly hydrophobic interactions with the catalytic domain, thereby efficiently blocking the rhodesain active site (Fig. 4A).

In agreement with structures from multiple CathL proenzymes, the pro-rhodesain H3 helix starts out as a regular α-helix. However, and so far unique to pro-rhodesain, a bulge forms around residues N103^{pro} and G104^{pro}, whose backbone amides are not involved in hydrogen bonds. The hydrogen bond acceptor group for the amide group of G104^{pro} in the context of a regular α-helix would be the oxygen of the backbone carbonyl of R100^{Pro}. This group, however, is 5.1 Å away from the G104^{pro} amide group, thus precluding formation of a hydrogen bond and instead locally distorting the α-helix in proximity to the active site of the core protease (Fig. 4B). Importantly, this bending of the α-helix enables the guanidino group of R100^{pro} to interact with the side chain of Q144 in the core protease domain. Notably, Q144 is conserved across all CathL proteases while R100^{pro} is conserved in the majority of homologous trypansomal CathL proproteases (Fig. S6). Below the bulge, H3 continues as a regular α-helix up to residue K113^{pro} from where the prodomain leads into the R-domain of the catalytic core domain *via* a C-terminal 15 amino acid unstructured linker (Figs. 2 and 4B). The electron density in our X-ray structure allows tracing the backbone of the entire H3 α-helix unambiguously for all residues in H3 (E95^{pro}–K113^{pro}) (Fig. S5B). The B-factor values for the two N-terminal helical turns of H3 are among the highest in the prodomain while in the entire remainder of this helix following the bulge, the B-factors are among the lowest (Fig. 2, C and D). The sequence of the N-terminal region of H3 preceding the bulge, which is α-helical across all available CathL prodomain structures, is generally well conserved across CathL enzymes (Fig. 4A, Fig. S6). In contrast to the continuous α-helix observed for H3 in pro-rhodesain, the structures of all other available, nontrypanosomal pro-CathL proproteases display a mostly unstructured peptide in the C-terminal end of H3 (Fig. 4A). Nonetheless, despite the differences in secondary structure within this blocking peptide, its overall position relative to the cleft of the core protease domain is highly
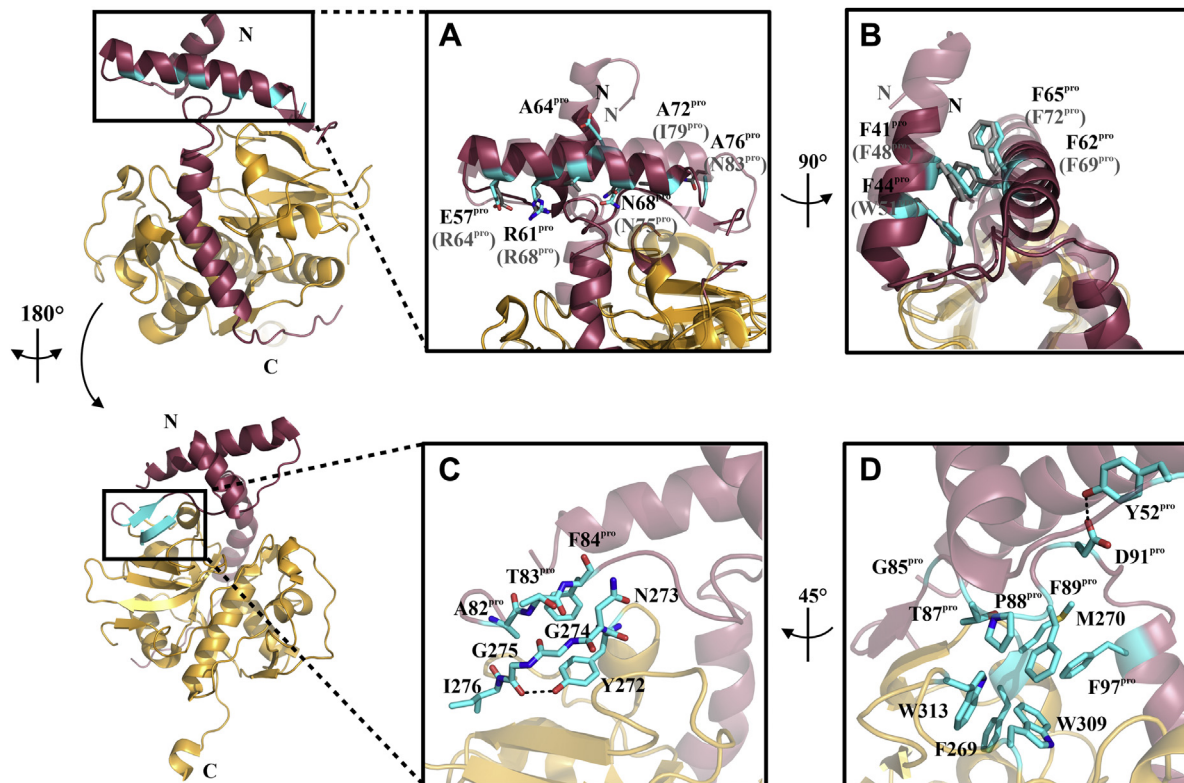
Figure 3. Conserved motifs within CathL-like proteases. Structure of pro-rhodesain is shown in two different orientations with the core domain in *gold*, the prodomain in *red*, and specific residues highlighted in *cyan*. *A*, zoom on the conserved ER(A/V)FNAA (*cyan*) motif. For comparison, the pro-papain structure (PDB: 3TNX) (30) is shown transparently. The amino acid numbers for both enzymes are based on their uniprot entries (ID.: Q95 PM0, pro-rhodesain; ID:P00784, pro-papain). Pro-papain residues are displayed in parentheses. *B*, conserved aromatic π-stacking between residues from H1 and H2 of the propeptide. Such interactions are also observed in *e.g.*, pro-papain (PDB: 3TNX) (30) and at least three of the four aromatic residues involved in this interaction are highly conserved across the prodomains of other cathepsin zymogens, including human pro-CathL (37) (Fig. S6). *C*, the propeptide binding loop (PBL) forms an antiparallel β-sheet with a loop in the R subdomain of the core protease. Additionally, the hydroxyl group of the Y272 sidechain forms an H-bond with the backbone carbonyl group of G275 while its aromatic ring stacks with the peptide bond between N273 and G274, thus providing further stabilizing interactions. *D*, interactions of the GNFD-like motif. D91^{pro} interacts *via* an H-bridge with Y52^{pro} of the first prodomain loop, while P88^{pro} and F89^{pro} are buried in a hydrophobic pocket formed by F269, W309, and W313 in the catalytic domain.

similar in all pro-CathL enzymes. However, this region shows no sequence similarity between *T. brucei* rhodesain and nonparasitic pro-CathL enzymes, but extremely high sequence identity across trypanosomal pro-cathepsins (Fig. S6).

The observation that the region of the prodomain blocking the protease active site is α-helical in pro-rhodesain, but not in any of the other pro-Cath(L) structures determined to date (Fig. 4*A*), prompted us to test whether the intrinsic propensity of this region to form an α-helix is higher in pro-rhodesain compared with related proteases. We used the isolated blocking peptides from *T. b. rhodesiense* rhodesain (residues R102^{pro}–K113^{pro}) and human CathL (residues N93^{pro}–R101^{pro}) to record CD spectra at pH 8 and pH 4 and with increasing amounts of trifluoroethanol (TFE) (Fig. 4, *C* and *D*, Fig. S7). TFE induces secondary structure formation by promoting hydrogen bond formation (38, 39). Both peptides are mainly unstructured in the absence of TFE. For the peptide from human CathL, even high TFE concentrations (up to 90% v/v) do not induce significant α-helical content (Fig. S7). This is in agreement with the highly disordered protein region observed in the human pro-CathL crystal structure (PDB: 1CJL) (37). In contrast, the isolated rhodesain prodomain peptide becomes α-helical already at low TFE concentrations

(Fig. 4, *C* and *D*). The variation in pH had only a small effect on secondary structure propensity of either blocking peptide. Likewise, the CD spectra of the purified pro-rhodesain C150A construct are also virtually identical at high and low pH (Fig. S8).

Propeptide binding site interactions resemble the binding mode of the K11777 inhibitor

Proteases recognize their substrates by engaging with the substrate amino acid side chains in specific pockets. These pockets are numbered outward from the cleavage site toward the substrate's N- (binding sites S1-Sn) and C-terminus (prime sites S1'-Sn') (40). In pro-cathepsins, the N- to C-terminal orientation of the prodomain blocking peptide is inverted with respect to the interaction of a putative peptide substrate. Additionally, the peptide backbone of the blocking peptide hovers "above" the active site cleft typically occupied by a substrate (compare Fig. 5, *A*, *C* and *E*). In the pro-rhodesain structure, the S2' and S1 sites are partially occupied by R100^{pro} and S1' by the side chain of Y101^{pro} while a glycerol molecule sits in the center of the substrate-binding site (Fig. 5, *A* and *B*). The side chains of F108^{pro} and Y107^{pro} occupy the S2 and S3 pockets, thus overlapping with
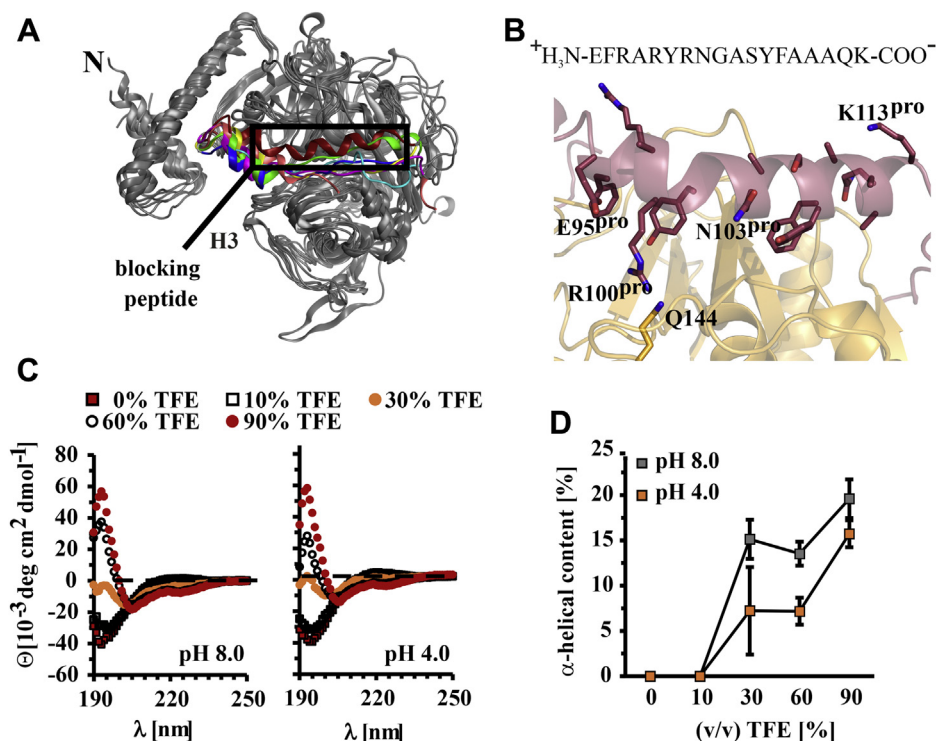
Figure 4. The trypanosomal pro-CathL blocking peptide adopts an α-helical structure not previously observed in nonparasitic CathL zymogens. *A*, comparison of the blocking peptide in pro-rhodesain and other CathL zymogens (*green = F. hepatica*, PDB: 2O6X; *blue = H.sapiens*, PDB: 1CJL; *pink = C. papaya*, PDB: 1PCI; *beige = T. molitor*, PDB: 3QT4; *turquoise = H. sapiens* (Pro-CathX), PDB: 1DEU; *yellow = H. sapiens* (Pro-CathK), PDB: 1BY8). *B*, zoom on helix 3 of the pro-rhodesain propeptide (*red*) covering the active site cleft of the core protease domain (*gold*). The amino acid sequence of H3 is shown on top and side chains are presented as sticks. H3 harbors a kink at residue N103[pro]. The side chains of R100[pro] and Q144 (within the catalytic domain, *gold*) interact. *C*, CD spectra of the trypanosomal blocking peptide at various TFE concentrations and different pH values (for CD spectra of the human CathL blocking peptide see Fig. S7). *D*, the α-helical content of the blocking peptides was calculated with the online tool BeStSel (http://bestsel.elte.hu/index.php) (68).

the binding sites of the phenyl- and *N*-methylpiperazine groups of the K11777 inhibitor observed in the crystal structure of rhodesain (PDB: 2P7U) (Fig. 5, *C* and *D*) (25). We also looked at the position of the peptidic substrate Z-Phe-Arg-AMC used for activity assays by molecular docking and observed a similar occupation of the rhodesain substrate-binding pockets (Fig. 5, *E* and *F*).

pH-dependent autocleavage of pro-rhodesain can occur intra- and intermolecularly

For the physiological activation of rhodesain and other CathL-like proteases, the prodomain has to undergo cleavage at low pH, and it has been suggested that this process can occur either inter- or intramolecularly (or both) (21, 41, 42). In our hands, folded, inactive pro-rhodesain C150A purified from *E. coli* is not cleaved, ruling out a role for other *E. coli* proteases in the cleavage process of heterologously expressed pro-rhodesain. Lowering the pH for purified wildtype pro-rhodesain leads to cleavage indicating that intramolecular autocleavage plays at least some role in pro-rhodesain processing (Fig. S4). However, this does not exclude the possibility that cleavage also occurs intermolecularly. Intermolecular protease cleavage can be easily probed by mixing catalytically inactivated enzymes with a wildtype protease (43). Incubating an excess of purified, catalytically inactive pro-rhodesain C150A with catalytically active, mature, wildtype rhodesain

led to a continuous decrease of the high molecular band corresponding to pro-rhodesain in a gel-shift assay (Fig. 6, *A* and *B*). Importantly, this process is pH-dependent and lower pH values result in more efficient autocleavage.

More efficient pro-rhodesain C150A cleavage at lower pH values could be due to a higher intrinsic activity of the mature wt rhodesain and/or changes in the structural dynamics of the proenzyme and consequently a higher accessibility of its cleavage site. To probe the first possibility, we tested the residual activity of mature rhodesain at different pH values by incubating it with the fluorogenic peptide-based substrate (Fig. 6C). While the activity of rhodesain slowly decreases at pH 8 over the course of multiple hours, this drop in activity is not sufficient to explain the near-absence of prodomain cleavage within the same timeframe and the differences in pro-rhodesain C150A cleavage at pH 6.5, pH 5.5, and pH 4 (Fig. 6). We therefore investigated pro-rhodesain dynamics at high and low pH values, in particular the relative motion of the prodomain to the core protease domain.

Interdomain dynamics of pro-rhodesain probed by PET-FCS and MD simulations

To investigate the dynamics of the prodomain in relation to the core protease domain, we used PET-FCS (27). Here, the dynamics of a protein can be probed *via* contact-induced fluorescence quenching, *i.e.*, a PET reaction of an extrinsic
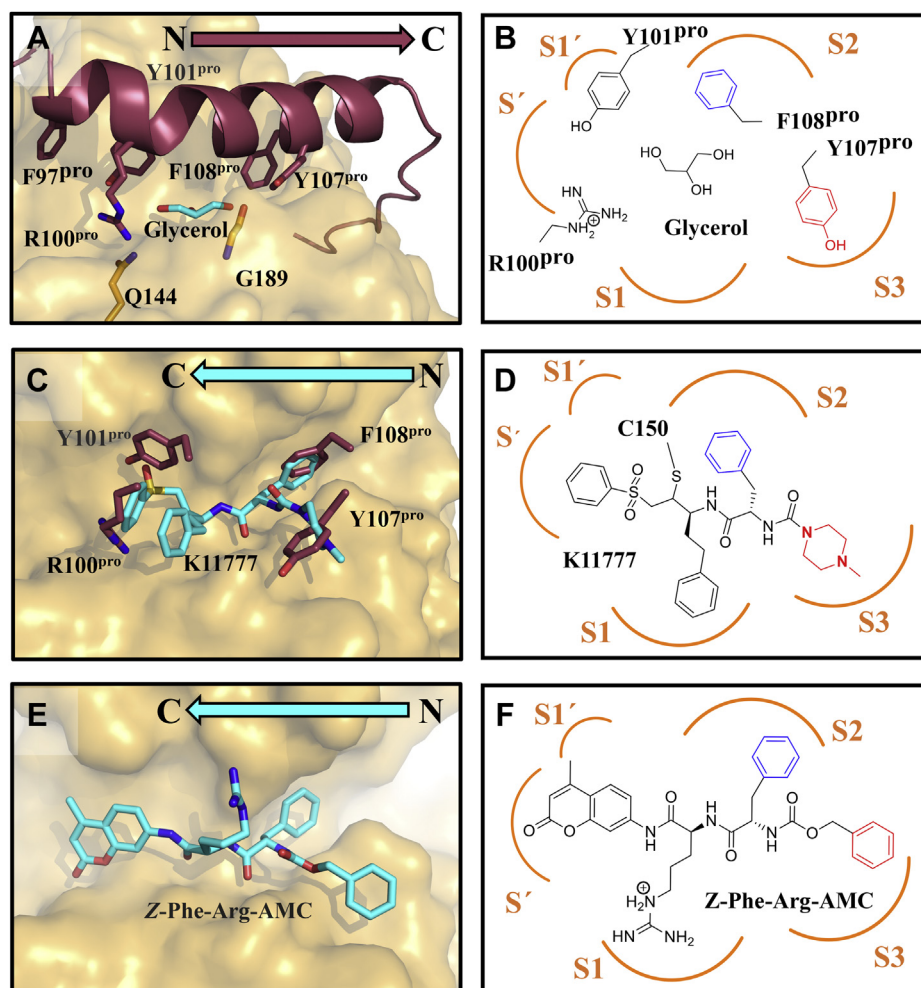
Figure 5. Side chain interactions of H3 with the protease active site cleft are similar to those of the peptide-based inhibitor K11777. A) Interactions between the active site cleft and the pro-peptide side chains. B) 2D-presentation of the H3 side chains within the active site cleft. C) Crystallographic binding mode of K11777 (cyan carbon atoms) in complex with rhodesain (*gold surface*) in superposition with pro-peptide residues occupying the S1', S2 and S3 pockets (*red sticks*). D) 2D-presentation of K11777 in the active site based on the crystal structure 2P7U. The orientation of the inhibitor within the active site resembles the binding mode of H3 in B) with regard to the placement of the sidechains. E, F) Binding mode and 2D-presentation of Z-Phe-Arg-AMC docked into the binding pocket of rhodesain (FlexX score -23.98 kJ/mol).

fluorophore attached to a cysteine side chain with the side chain of the natural amino acid tryptophan (44). Fluorophore and indole group form a van der Waals contact at distances below 1 nm, which can be disrupted or formed by conformational changes of a protein. Such distance changes are thus translated into fluorescence fluctuations, and the combination of the PET fluorescence reporters with fluorescence correlation spectroscopy facilitates the single-molecule detection of fast protein dynamics with nanosecond time resolution (27, 45, 46). Pro-rhodesain contains eight native cysteine residues: the three cysteine pairs forming disulfide bonds in the core protease domain (Fig. S5), the active site residue C150, and a cysteine in the prodomain (C22). To avoid unspecific labeling of the enzyme with a fluorophore, we mutated both the active site cysteine (C150A) to stabilize the proenzyme as well as C22^{pro} (C22S^{pro}).

For the pro-rhodesain C22S^{pro}, C150A double mutant, protein modified with an AttoOxa11-dye was hardly detected, showing that the remaining six cysteine residues were

protected against fluorophore labeling due to their involvement in disulfide bridges. Subsequently, two new cysteine residues were introduced in the C22S^{pro}, C150A background of pro-rhodesain, either at the C-terminal end of pro-domain helix 1 (V51C^{pro}) or right above the bulge in helix 3 in the vicinity of the active site (A99C^{pro}). For PET fluorescence measurements, a tryptophan mutation at position Q146 in the L domain of the core protease domain was introduced in the vicinity of V51C^{pro} and A99C^{pro} (Fig. 7). Neither the cysteine mutations nor the additional tryptophan mutation perturbed the secondary structure of pro-rhodesain and all constructs could be successfully labeled with an AttoOxa11-dye as assessed by steady-state fluorescence spectroscopy (Figs. 7 and 8, *A* and *B*). As expected, due to the large distance from the fluorophore as apparent from our structure, the eight native tryptophan residues in pro-rhodesain did not lead to fluorescence quenching of dyes attached to residues V51C^{pro} or A99C^{pro}, while introduction of Q146W did. The AttoOxa11-labeled V51C^{pro}/Q146W and A99C^{pro}/Q146W pairs can
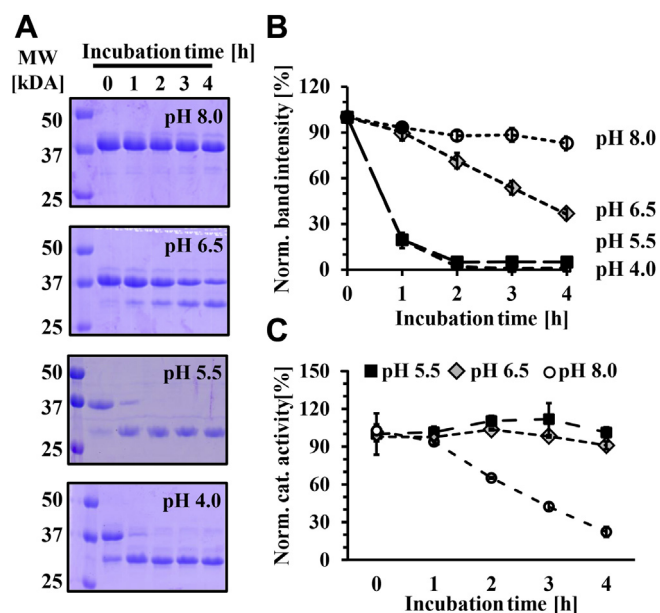
Figure 6. pH-dependent auto-cleavage of pro-rhodesain. *A,* coincubation of pro-rhodesain C150A with substoichiometric amounts of mature rhodesain (100:1 mol:mol) at different pH values leads to cleavage of the 37 kDa proenzyme yielding a 26 kDa fragment corresponding to the core protease as observed on SDS-PAGE. Lower pH values lead to more efficient cleavage. *B,* densitometric analysis of gels in (*A*). Shown is the decay in the bands at 37 kDa corresponding to pro-rhodesain C150A. Error bars depict SDE from three independent measurements. *C,* activity of wild-type rhodesain probed by a fluorescence peptide cleavage assays under the conditions used in (*A*). Shown are values normalized to rhodesain activity determined at pH 5.5 with a fluorogenic peptide.

thus be used as sensitive fluorescence reporters for interdomain dynamics between the prodomain and the protease domain of pro-rhodesain.

No changes in fluorescence intensity upon reducing the pH are observed in our fluorescence spectra of pro-rhodesain (Fig. 8, *A* and *B*). To probe whether there are nonetheless dynamic ramifications for pro-rhodesain in response to pH changes, we additionally carried out PET-FCS measurements with the dye/Trp pair containing proteins at pH 8 and pH 5.5 (Fig. 8, *C–F*). Here, the drop in pH led to a small increase of the diffusion time constant τ_D of all tested mutants indicative of a larger hydrodynamic radius under acidic conditions. Since this behavior is observed for all proteins, it might be interpreted as a global structural rearrangement of pro-rhodesain or "loosening" of the overall structure at low pH.

Interestingly, the pro-rhodesain construct with the A99C^{pro}/Q146W dye/quencher pair also showed a pH-dependent change in sub-ms fluorescence fluctuations (Fig. 8, *E* and *F*). Fluorophores attached to residue A99C^{pro}, within the "bulge" of H3 covering the protease active site, yield a monoexponential autocorrelation curve with a 1 µs time constant at pH 8. At pH 5.5, the same construct shows a biexponential trace with time constants of 0.21 and 2.1 µs (Fig. 8, *E* and *F*). The amplitudes are the same in both cases, showing that the total fluorescence quenching is similar at high and low pH, but that the local dynamics of the system are more heterogeneous under acidic conditions. In contrast, the pro-rhodesain construct harboring the V51C^{pro}/Q146W dye/quencher pair, with V51C^{pro} at the C-
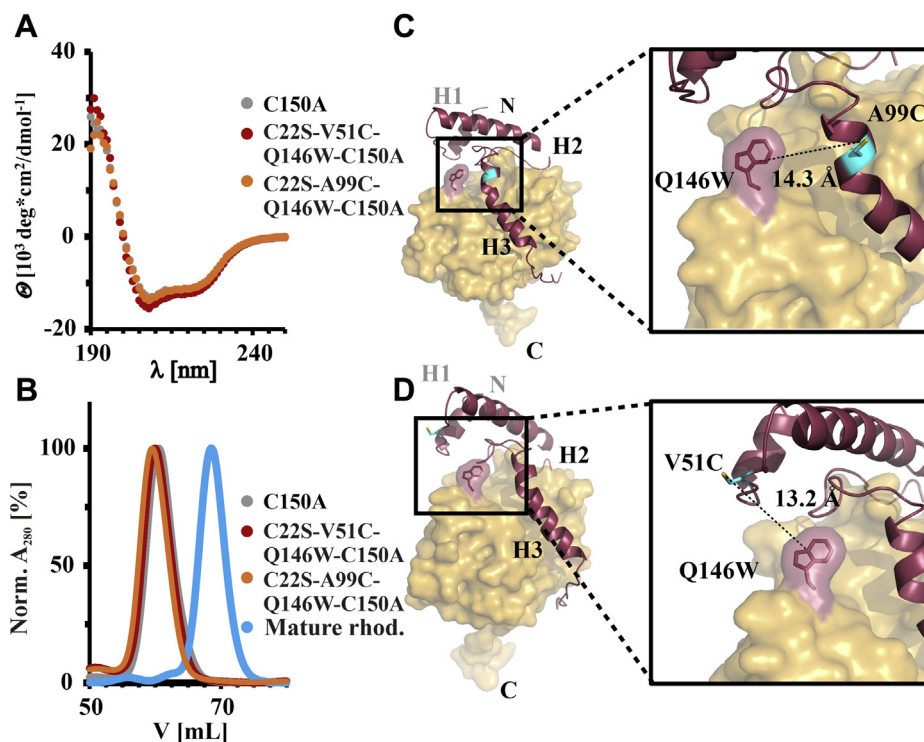

Figure 7. Pro-rhodesain mutants used for PET-FCS measurements. Introduction of the required Trp and Cys mutations for PET-FCS does not change the protein secondary structure as assessed by (*A*) CD spectroscopy and (*B*) size-exclusion chromatography of rhodesain variants. *C* and *D,* position of the Trp/Cys pairs used for PET-FCS to probe interdomain dynamics between prodomain and core protease domain.
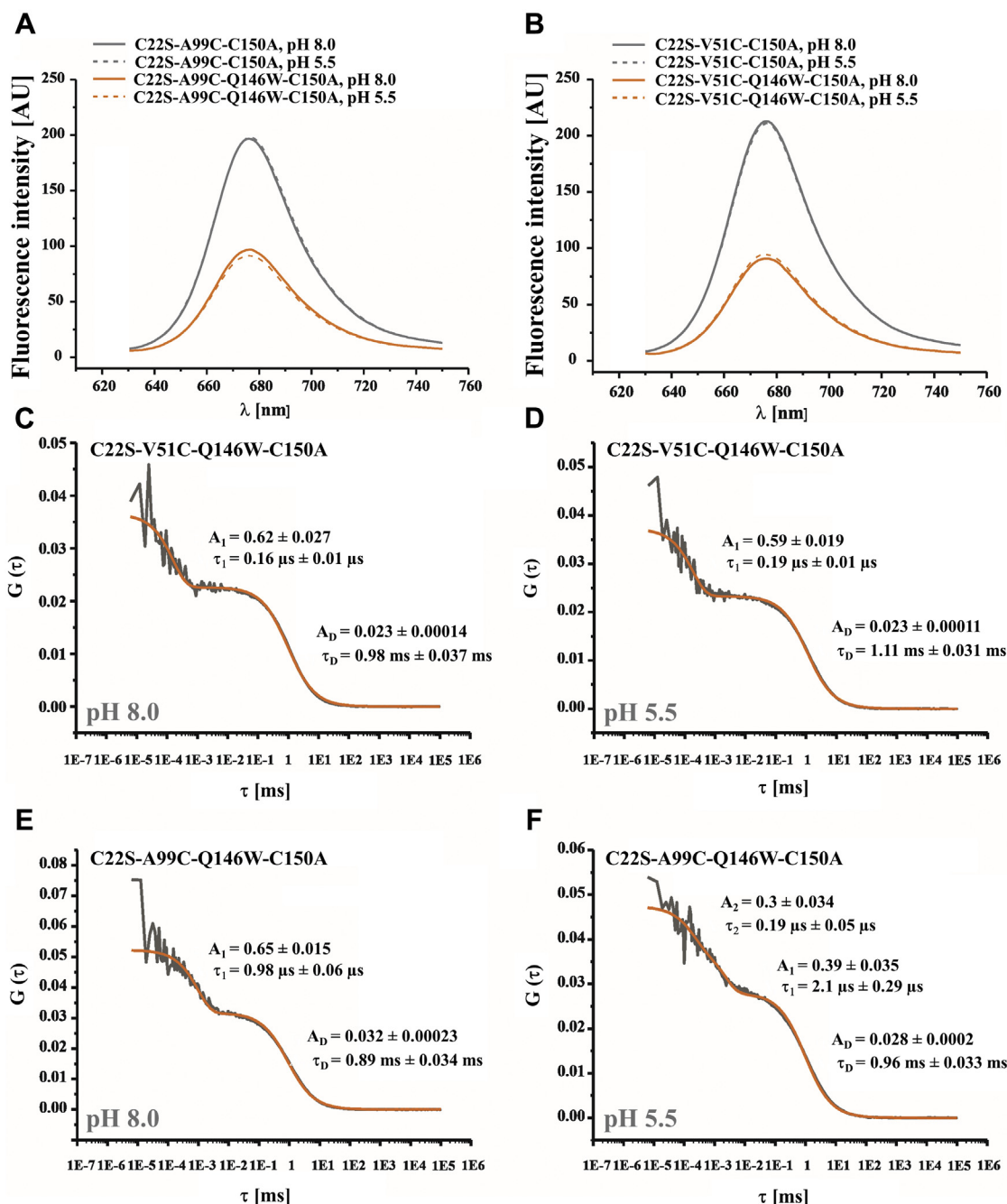
Figure 8. PET-FCS on pro-rhodesain. *A* and *B*, steady-state fluorescence spectra of pro-rhodesain constructs, labeled at position A99C[pro] or V51C[pro] using AttoOxa11, show efficient quenching of the label by the Trp residue Q146W in the core protease domain while changes in pH have a negligible effect on the spectra. *C–F*, FCS autocorrelation functions (G(τ)) measured from pro-rhodesain constructs at 37 °C under solution conditions specified in panels. Data shown in panels *C–E* were fit using a model for molecular diffusion containing a single exponential relaxation (*orange line*). Data shown in panel *F* were fit using a model for molecular diffusion containing two single-exponential relaxations (*orange line*). Amplitudes (A_n) and time constants (τ_n) extracted from fits are shown in panels.

terminal end of H1 in the prodomain shows only a modest response to the change in pH (Fig. 8, *C* and *D*).

To shed additional light on the pH-dependent autocleavage process, we carried out molecular dynamic (MD) simulations with the proenzyme at pH 4 and 8, respectively. While large conformational changes are beyond the timescale of the 20 ns MD simulations, "loosening" of the prodomain may be indicated by an increased flexibility and the reduction of interactions between residues sensitive to pH changes (35). All simulations showed a high stability within the given timeframe (Fig. S9*A*). The root mean square fluctuation (Fig. S9*B*) and deviation (RMSD, Fig. S9*C*) analysis indicated flexibility in the same protein regions as found by B-factor analysis (Fig. 2*D*). The evaluation of the interaction energies between the prodomain and main domain indicates significantly reduced interactions between the two domains at pH 4 compared with

pH 8 while higher interaction energies were found within the main domain at pH 4 (Table S1).

Since pro-rhodesain is cleaved efficiently at low pH and our PET-FCS data indicate that the dynamics of H3 increase under acidic conditions, we wondered whether protonatable groups mediating interactions between H3 and the core protease play a role in the cleavage process. We thus performed a detailed analysis of the H-bond interaction profile for the MD simulations with a focus on titratable groups (Tables S2–S7). This revealed some changes in the interaction patterns of specific amino acids for different protonation states. A salt bridge connecting the C-terminal end of H3 (R116^{Pro}) with the core protease (D194 and D242) was deemed especially interesting. For MD simulations at pH 4, D194 was protonated and therefore able to interact with D242 *via* a hydrogen bond (64–74% of the simulation time). When D194 was deprotonated during simulations at pH 8, this interaction was no longer possible and D194 and D242 were instead able to interact with the side chains of R114^{Pro} and R116^{Pro} *via* ionic interactions (Table 2). This provides a hint for putative stabilizing interactions between pro- and main domain at higher pH values that can be replaced by intradomain interactions at acidic pH.

Incidentally, the residues involved in this putative interdomain "lock" at pH 8 are highly conserved across trypanosomal pro-CathL proteins (Fig. 9A, Fig. S6). Despite the lack of overall sequence similarity, this interaction is even replicated in nontrypanosomal CathL proteins (Fig. 9A). In *C. papaya* pro-caricain, the positions of the corresponding basic and acidic residues are interchanged further underscoring the fact that these amino acids presumably play an important role and may coevolve.

To probe the effect of this putative interaction on pro-rhodesain autocleavage, we mutated the two aspartic acids into nonprotonatable asparagine residues in the C150A background yielding pro-rhodesain C150A, D194N, D242N (C150A DD/NN). Adding mature wt rhodesain to this mutant, we probed its digestion *via* the same gel-shift assay used previously to establish intermolecular cleavage (Fig. 9, *B* and *C*). As observed previously for the construct harboring only the C150A mutation, the inactive pro-rhodesain C150A DD/NN mutant was efficiently cleaved at low pH values. The interdomain interaction at this site therefore does not significantly affect the intermolecular cleavage process. To probe whether it may instead play a role in the intramolecular cleavage, both wt pro-rhodesain and pro-rhodesain DD/NN (with an intact active site cysteine) were expressed in *E. coli* and obtained in their proforms by purifying them at pH > 7 and in the presence of the protease inhibitor phenyl-methylsulfonyl fluoride (PMSF) to avoid autocleavage (47). After purification, the PMSF inhibition was reversed by the addition of DTT. Upon adjusting the pH to either pH 8 or pH 4, the enzymes' autoactivation was monitored using the fluorescence-based peptide cleavage assay. As expected, at pH 4 the overall activity is higher for both enzymes. While the DD/NN mutant exhibits a slightly reduced maximum activity compared with the wt protein, both enzymes show comparable activation kinetics and reach their activation plateau at the same time (Fig. 9D). At pH 8, the activity for both proteins is significantly lower. However, under these conditions the autoactivation of the DD/NN mutant is accelerated considerably compared with the wild-type enzyme (Fig. 9E).

Discussion

For survival, propagation, and overcoming the BBB in the host, African trypanosomes rely on proteases such as rhodesain. Similar to related cathepsin proteases, rhodesain is expressed as a zymogen with a prodomain that maintains its inactive state until the protein is correctly delivered into the acidic environment of the lysosome. Here, we present the first trypanosomal pro-CathL structure, *T. brucei rhodesiense* pro-rhodesain, as an important intermediate of the protease maturation process. While the overall architecture of pro-rhodesain agrees with related proenzymes including human cathepsin L or *C. papaya* papain (37, 48), pro-rhodesain displays a unique extended α-helix covering the entire cleft between the L- and R-lobes of the core protease domain (Fig. 2). A pronounced bulge in the region covering the protease active site disrupts the extended α-helix in H3 of the prodomain. In the structures of currently available homologous, nontrypanosomal proproteases, H3 becomes unstructured at this position (Fig. 4A). It thus seems possible that deviations from a regular α-helix are a unifying structural characteristic of pro-CathL proteins. As a word of caution, it should be noted that all available structures of pro-CathL proteases including pro-rhodesain were obtained by mutating the active site cysteine. This removes a thiolate anion from the active

Table 2
Relative occurrence of side-chain hydrogen bonds between selected pH-sensitive residues

Acceptor	Donor	% pH 4 (1)	% pH 4 (2)	% pH 4 (3)	% pH 4 (4)	% pH 8 (1)	% pH 8 (2)	% pH 8 (3)	% pH 8 (4)
Asp242	Asp194[a]	64.62	74.21	70.36	64.32	0	0	0	0
Asp242	Arg114^{Pro}	0	0	0	0.15	30.80	0	0	7.60
Asp242	Arg116^{Pro}	3.40	8.05	9.00	3.75	1.25	63.80	4.75	0
Asp242	Lys117^{Pro} [b]	21.15	15.85	0	0.05	7.25	34.70	24.69	19.94/17.04[c]
Asp242	Ser335	20.94	9.80	26.19	20.59	0.15	50.17	23.39	0
Asp242	His240	1.20	3.05	6.85	4.10	0	0	0.75	2.20
Asp194	Arg114^{Pro}	0	0	0	0	9.05	15.05	0.10	0.20
Asp194	Arg116^{Pro}	0.20	0.85	0.30	0.85	19.60	22.75	49.58	50.07
Asp194	Ser335	0.10	0.15	4.85	0.60	56.77	1.10	20.79	72.21

Four respective MD simulations at pH 4 or pH 8 were carried out.
[a] Denotes donor functionality only in protonated state (pH 4).
[b] Interaction with NH backbone.
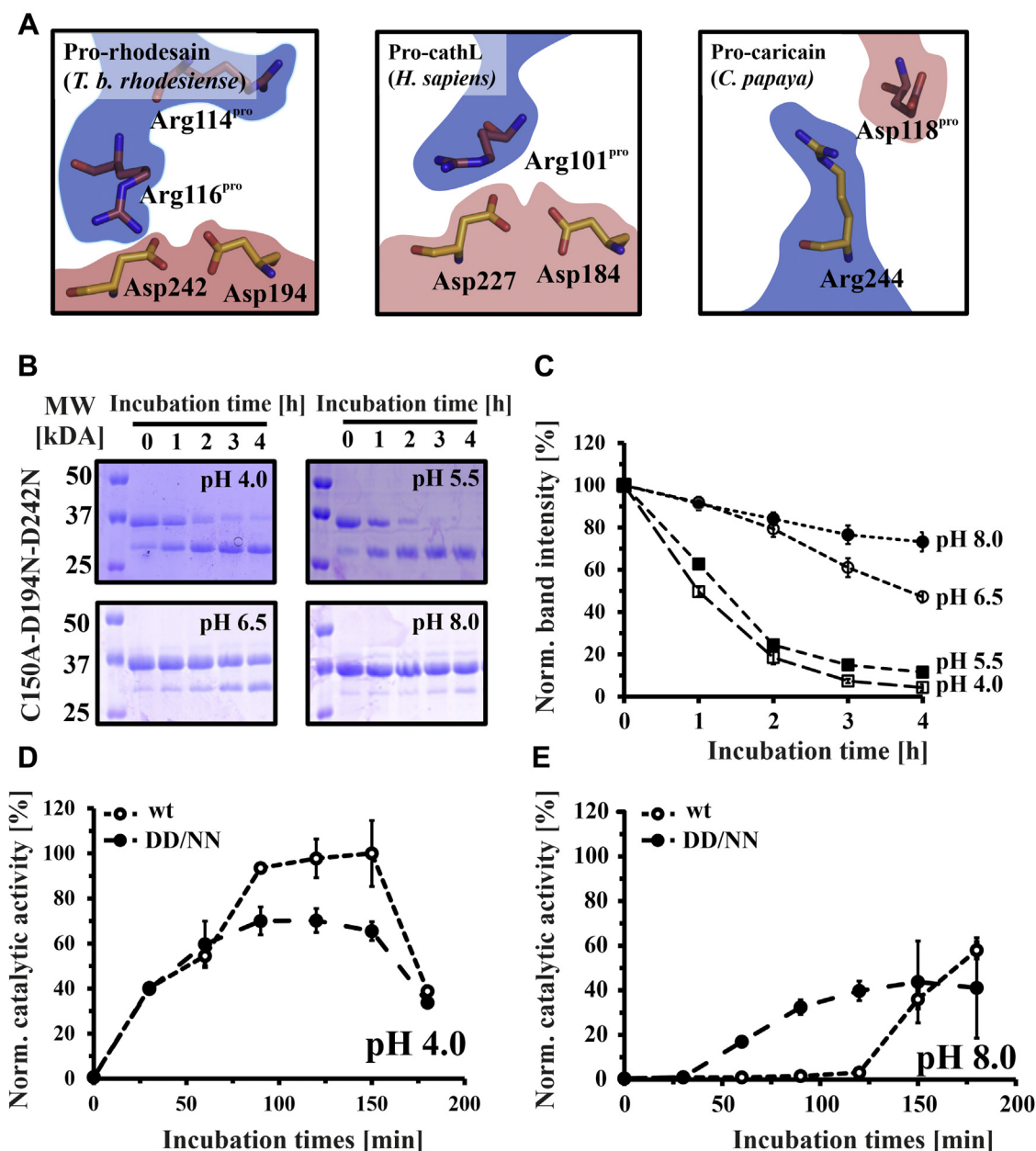[c] Interaction with side chain amine.

Figure 9. An interdomain salt bridge influences the pH sensitivity of pro-rhodesain autoactivation. *A*, conserved salt bridges between C-termini of blocking peptides and catalytic domains in crystal structures from pro-CathL proteases from different species (*T. b. rhodesiense*, PDB: 7AVM; *H. sapiens*, PDB: 1CJL; *C. papaya*, PDB: 1PCI). The amino acids of the prodomain (*red*) and the catalytic domain (*orange*) are shown as *sticks* and highlighted by their charge (*blue* = positive, *red* = negative). In the case of *C. papaya* pro-caricain, a polarity switch occurs. *B*, pH-dependent intermolecular (*trans*) cleavage of inactive pro-rhodesain C150A DD/NN (D194N, D242N) was probed by incubating the purified protein with a substoichiometric amount of mature rhodesain (100:1 mol:mol) at pH 4.0, 5.5, 6.5, or 8.0 at 37 °C. Samples were taken at indicated time points and run on 15% SDS-PAGE. Pro-rhodesain C150A DD/NN can be digested intermolecularly by mature rhodesain. *C*, intensities of the pro-rhodesain band at ~37 kDa were determined with ImageJ and compared with timepoint t_0. *D* and *E*, purified wt pro-rhodesain and the pro-rhodesain DD/NN variant, both harboring the active site cysteine to enable autoactivation, were incubated at pH 4.0 (*D*) and pH 8.0 (*E*) at 37 °C. Inhibition with PMSF was removed by adding DTT and waiting for the specified time before monitoring the protease catalytic activity *via* the cleavage of a fluorogenic peptide as a proxy for protease autoactivation. Data were normalized to the maximum activity of the wt enzyme at pH 4.

site and may thus locally distort the structure of the prodomain. In addition, we observe a glycerol molecule in the active site that could lead to deviations in the side chain orientations of residues within prodomain peptide interacting with the active site.

In the α-helical region following the bulge in H3, trypanosomal CathL proteases display very high sequence identity to each other but not to CathL enzymes from other organisms. The extended H3 helix may thus be a defining feature of

trypanosomal CathL proteases and as such present a structural template to further improve the affinity or selectivity of rhodesain inhibitors. We found that the side chains of the rhodesain prodomain blocking peptide partially resemble binding of K11777, a promising drug candidate against protozoan diseases (49), by occupying the S2 and S3 substrate-binding pockets with its aromatic and piperazine groups. In hindsight, this nicely rationalizes the original screening-based

design of the K11777 (50) and may provide an opportunity for further inhibitor optimizations.

In agreement with the literature (51), our data show that pro-rhodesain can be activated by pH-dependent intra- and intermolecular cleavage processes. Our PET-FCS measurements may provide a molecular explanation for the pH dependency of pro-rhodesain autoactivation due to changes in the interaction dynamics between the H3 helix in the prodomain and the core protease domain (Fig. 8). Low pH "loosens" the pro-rhodesain structure as indicated by the increase in the hydrodynamic radius and the observed reduced diffusion times compared with the proenzyme at pH 8. The acid-induced heterogeneity in conformational dynamics may be interpreted as local folding/unfolding events of H3 at low pH values. The high crystallographic B factors also indicate higher flexibility in this region compared with the remainder of the prodomain and the core protease domain.

Our MD simulations also suggest that the interactions between pro- and core protease domain are significantly reduced at lower pH values. Interestingly, a PET-FCS reporter pair informing on the relative motions between H1 from the propeptide and the core protease domain did not exhibit changes in dynamics under all conditions tested, thus suggesting that the degree of pH-dependent structural dynamics differs within the prodomain and that H3 may be particularly sensitive to pH changes.

Combining MD simulations and site-directed mutagenesis, we narrowed the pH-sensitive changes in H3 down to a set of protonatable groups, D194 and D242 in the protease domain. These residues can mediate ionic interactions with basic residues in the prodomain, R114^{pro} and R116^{pro}, and thus form a pH-dependent "interdomain lock." The participating residues of this salt bridge connecting the C-terminal end of H3 to the core protease domain are highly conserved across pro-cathepsins. In wild-type pro-rhodesain, where the lock is intact, autocleavage occurs efficiently at pH values of 5.5 or lower (Fig. 6B) when the aspartic acids become protonated. The typical pK$_a$ of free aspartic acid is ~4 although this value can be significantly lower in the context of a salt bridge in the interior of a protein (52). In pro-rhodesain, the salt bridge is solvent exposed, a feature that may be necessary to fine-tune pH-dependent autoactivation of pro-rhodesain in a physiological pH range. An elevated pK$_a$ of one of the two aspartic acid residues involved in the intradomain interaction is likely because they form a "handshake" interaction (Fig. 9) (53). Mutation of the two conserved aspartic residues breaks the lock and thus enables autoactivation of pro-rhodesain even at near neutral pH (Fig. 9E). This indicates that these residues act as a pH-dependent protective lock, presumably to prevent premature protease activation at higher pH values, *i.e.,* before the protein reaches the acidic lumen of the lysosome.

In agreement with models put forward for the related human CathL protease (21) as well as more recently for rhodesain itself (51), activiation of pro-rhodesain could follow a two-step mechanism. An initial, slow intramolecular cleavage step is delayed by the protective salt bridge between the prodomain and the core protease as well as strong interdomain interactions. This reduces interdomain dynamics and prevents premature autoactivation at higher pH values. As the proprotease progresses along the cellular trafficking routes to the lysosome and experiences continuously more acidic environments, the aspartic residues within the core protease domain are protonated, which leads to the opening of the "interdomain lock" by replacement of the interdomain ionic interaction with an intradomain D194-D242 hydrogen bond, followed by an increase in interdomain dynamics and resulting in autoactivation of the protease. Once liberated from the inhibitory prodomain, the mature protease is now available to also activate other pro-rhodesain protomers in an intermolecular fashion, thus speeding up the maturation process of the protease population significantly. The interdomain salt bridge and dynamics centered on the unique helix 3 of the trypanosomal prodomain are thus decisive elements to prevent premature activation and to correctly time pro-rhodesain cleavage under the right (*i.e.*, low pH) physiological conditions. Due to the high structural similarities among pro-CathL enzymes and the additional high sequence similarity of H3 among trypansomatid proteases, our structure in combination with the presented delayed activation model may potentially act as a blueprint for these important enzymes in trypanosomatids and beyond.

Experimental procedures

Rhodesain expression and purification from P. pastoris and E. coli

The expression of rhodesain from *P. pastoris* was performed as described previously (9). For purification from *E. coli* Rosetta2 DE3 pLysS (Novagen), the *T. b. rhodesiense* rhodesain sequence (uniprot-ID.: Q95 PM0) followed by a TEV cleavage site and a hexahistidine-tag was cloned between NdeI and BamHI restriction sites of a pET-11a vector by Genscript using codon optimization (Fig. S3). The plasmid was amplified in XL-10 cells. A GFP sequence was additionally introduced between the TEV site and the 6xHis-tag by Gibson assembly following extensive construct and purification optimization (see Extended Materials and Methods for details). The rhodesain plasmid and the GFP gene were amplified in a PCR using the primers 5′-caccaccat-caccaccactaagg-3′ and 5′-gctaccttgaaaatacagattctccggg-3′ for the vector and 5′-ctgtattttcaaggtagcgtgagcaagggcgaggagc-3′ as well as 5′-ggtggtgatggtggtgcttgtacagctcgtccatgccg-3′ for the insert.

Purification of pro-rhodesain C150A from E. coli

The active site C150A mutation was introduced by site direct mutagenesis using the primers 5′-GCGGTAGCGCGTGGGCGTTCAGCACC-3′ and 5′-CGCCCACGCGCTACCG-CATTGGCCCTG-3′. Transformation, protein expression, cell lysis, purification by Ni-NTA, and TEV-cleavage were performed as described in the Extended Materials and Methods. In short, after digestion of IMAC-purified pro-rhodesain with TEV protease, the cleavage products were loaded on 8 ml Ni-NTA gravity flow column preincubated with 50 mM Tris buffer (pH 8.5, 100 mM NaCl, 10 mM imidazole) to remove the TEV protease and GFP. Pro-rhodesain was eluted with 50 mM Tris buffer (pH 8.5, 100 mM NaCl, 10 mM imidazole) and combined with the flow through. After concentration by spin filtration, pro-rhodesain was loaded onto a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) and purified by SEC (0.5 ml/min, 20 mM Tris, pH 8.0,

200 mM NaCl). The protein in the peak fraction was dialyzed against H$_2$O at 4 °C, lyophilized, and stored at −80 °C until further use. To obtain wild-type pro-rhodesain, the purification protocol was adapted (see Supporting information for details).

Circular dichroism spectroscopy

CD spectra were measured using a Jasco-815 CD-spectrometer between 190 and 260 nm wavelength with 5 nm bandwidth, an interval of 1 nm, and a scanning speed of 50 nm/min. Measurements of mature rhodesain from *P. pastoris* and *E. coli* (5 μM in 3 mM sodium phosphate buffer, pH 6.8, 0.5 mM TCEP, 5 mM NaCl) and inactive pro-rhodesain C150A (5 μM in 10 mM Tris, pH 8.0, 3 mM NaCl) were performed in a 1 mm quartz curvet at 25 °C. Spectra were derived from the average of three measurements after subtraction of the baseline.

The obtained ellipticities θ were converted to the mean residue ellipticities $[\theta]_{mrw,\lambda}$ by the equation

$$[\theta]_{mrw,\lambda} = \frac{MRW \cdot [\theta]}{10 \cdot c \cdot d} \ ,$$

where c equals the concentration of the protein, d the pathlength through the sample, and MRW for the mean residue weight, which is calculated by

$$MRW = \frac{M}{N-1}$$

M corresponds to the molecular weight and N to the number of amino acids in the protein.

Circular dichroism spectroscopy of prodomain CathL peptides

Peptides were purchased from *peptides&elephants* (Henningsdorf, Germany). CD spectra were measured between 190 and 250 nm in sodium acetate (3 mM NaOAc, pH 4.0, 5 mM NaCl) or Tris buffer (3 mM Tris, pH 8.0, 5 mM NaCl) with TFE (0, 10, 30, 60 and 90% (v/v)) at a peptide concentration of 25 μM. Instrument settings, parameters, and data transformations were used as described above for the purified protein. Secondary structure content was calculated using the online server BeStSel (http://bestsel.elte.hu/index.php).

Size-exclusion chromatography

SEC runs were performed with a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) at a flow rate of 0.5 ml/min at 4 °C. In total, 4.5 ml samples were loaded using a 5 ml loop and 20 mM Tris (pH 8.0, 200 mM NaCl) as the running buffer. Absorption was measured at a wavelength of 280 nm and values were normalized to the absorption maximum.

Pro-rhodesain autocleavage (gel-shift)-assay

Inactive pro-rhodesain C150A was dissolved in the respective buffers: (a) 50 mM NaOAc pH 4.0, 235 mM NaCl, 5 mM EDTA, 5 mM DTT, 0.005% Brij35; (b) 50 mM NaOAc pH 5.5,

200 mM NaCl, 5 mM EDTA, 5 mM DTT, 0.005% Brij35; (c) 50 mM Bis-Tris pH 6.5, 229 mM NaCl, 5 mM EDTA, 5 mM DTT, 0.005% Brij35; or (d) 50 mM Tris pH 8.0, 221 mM NaCl, 5 mM EDTA, 5 mM DTT, 0.005% Brij35 to a concentration of 0.5 mg/ml. Mature, active rhodesain was preincubated in 50 mM sodium acetate buffer (pH 5.5, 200 mM NaCl, 5 mM EDTA, 5 mM DTT, 33 μg/ml) at room temperature for 10 min. In total, 11 μl of the active rhodesain solution was added to 100 μl of the different pro-rhodesain solutions each and the mixture was kept at 37 °C. Samples were taken at the given time points, denatured immediately, and analyzed by 15% SDS-PAGE. Band intensities were determined by integration with ImageJ (Version 1.52q) and normalized to the respective sample at time point zero (t$_0$).

X-ray crystallography

For crystallization, 1 μl of reservoir buffer (40 mM sodium citrate pH 3.5, 30% PEG-6000 (v/v), Pb(OAc)$_2$ (saturated)) was added to 1 μl of a pro-rhodesain C150A solution (0.7 mg/ml in 10 mM sodium citrate, pH 5). Crystals were grown at room temperature by the hanging drop method within 2 weeks. The crystals were preincubated in reservoir buffer with additional 10% (v/v) glycerol for 20 s before flash freezing in a 100 K N$_2$ cryo stream. Diffraction data were collected with a Bruker AXS Microstar-H generator and a MAR-scanner 345 mm image-plate detector in a distance of 350 mm and a X-ray wavelength of 1.5417 Å in 1°-steps after 10 min exposure. Data processing was performed by the program XDS in space group H1 (54). For structure determination, the previously determined structure of mature rhodesain in complex with a macrolactam inhibitor (PDB: 6EX8) was used as a model for molecular replacement with PhaserMR of the phenix work suites after removal of the ligand and all water molecules (24, 54–56). The prodomain was then built into the remaining electron density and final refinement was done with WinCoot (57).

MD simulations

MD simulations were performed with the crystal structure of pro-rhodesain described in this article but containing an active site cysteine residue instead of alanine. All amino acids and crystallographic water molecules were kept, while one glycerol molecule was removed from the structure. Missing residues within the structure were added and the structures were built using tleap of AmberTools17 (58, 59). For MD simulations mimicking the protein at pH 8, aspartate and glutamate residues were deprotonated and the catalytic diad was defined to form an ion pair of C150$^-$ (thiolate, CYM) and H287$^+$ (imidazolium, HIP). Additionally, residue H231 was protonated and H240 deprotonated (HIE). Protonation states of pH 4 were determined using predicted pK$_a$-values from MOE, resulting in protonation of all histidine residues, as well as protonated residues D143, 185, 194, 271, 289, 296 (ASH) and E38, 58, 66, 95, 144, 175, 232, 256, 283, 322, and 343 (GLH). Again, the catalytic C150 was deprotonated to form the active ion pair. The protein structures were energetically minimized with sander over 200 time steps. Counter ions (7 Cl$^-$ for pH 4 and 10 Na$^+$ for pH 8) to

neutralize the system and a TIP3P waterbox exceeding the protein by 10 Å in every dimension were added. System equilibration was performed over 500 ns with gradually releasing the constraints and 500 ns without constraint while heating from 100 to 300 K in an NVT ensemble (60). Subsequent production runs were performed in an NPT ensemble with periodic boundary conditions and a van der Waals cutoff of 14.0 Å over 20 ns. All simulations were performed with the AMBER force-field (ff14SB) on the high-performance computing (HPC) cluster MOGON of the University of Mainz with NAMD-2.12 (61, 62). Simulations were performed four times for each pH and results were analyzed using VMD-1.9.3 and related scripts (63). For the analysis of hydrogen bonds a cutoff distance of 3.0 Å and angle of 20° were selected.

Z-Phe-Arg-AMC docking

The predicted binding mode of the assay substrate in complex with rhodesain (Fig. 5) was generated by molecular docking. The receptor was prepared within LeadIT-2.3.2 (*BioSolveIT GmbH*, Sankt Augustin, Germany, 2017, www.biosolveit.de/LeadIT) using the crystal structure of the K11777-rhodesain complex, PDB-ID 2P7U (25, 64). The covalent ligand was untethered from the enzyme using MOE (Molecular Operating Environment, 2018.01; *Chemical Computing Group ULC*: Montreal, QC, Canada, 2018). For docking, the receptor was protonated using the Protoss module within LeadIT and the catalytic dyad was manually set to form the C25 thiolate and H162 imidazolium ion pair (corresponding to C150 and H287 in the proenzyme) (65). Water molecule 512 was kept as part of the binding site. Ligands were energetically minimized using omega2 (OMEGA 3.1.0.3, *OpenEye Scientific Software*, Santa Fe, NM. http://www.eyesopen.com) and the Merck molecular force field (MMFF94) (66, 67). The receptor was validated by redocking of K11777 (FlexX score—24.92 kJ/mol, RMSD of 1.7 Å). The substrate ligand Z-Phe-Arg-AMC was subsequently docked into the binding site.

PET fluorescence experiments

Steady-state fluorescence emission spectra were recorded using a Jasco FP-6500 spectrofluorometer. Sample temperature was set to 37 °C using a Peltier thermocouple. A sample concentration of 100 nM modified pro-rhodesain in a 0.5 ml quartz glass cuvette (Hellma) was used. Spectra were recorded in either 50 mM phosphate, pH 8.0, with the ionic strength adjusted to 200 mM using sodium chloride, or in 50 mM acetate, pH 5.5, with the ionic strength adjusted to 200 mM using sodium chloride. Both buffers contained 5 mM EDTA; and 0.3 mg/ml bovine serum albumine and 0.05% Tween-20 to prevent sticking of rhodesain to the glass surface.

PET-FCS was performed on a custom-built confocal fluorescence microscope setup that consists of an inverse microscope body (Zeiss Axiovert 100 TV) equipped with a diode laser emitting at 637 nm as the excitation source (Coherent Cube). The laser beam is coupled into an oil-immersion objective lens (Zeiss Plan Apochromat, 63x, NA 1.4) *via* a dichroic beam splitter (Omega Optics 645DLRP). The average

laser power was adjusted to 400 μW before entering the back aperture of the microscope using an optical density filter. The fluorescence signal was collected by the same objective, filtered by a band-pass filter (Omega Optics 675RDF50), and imaged onto the active area of two fibre-coupled avalanche photodiode detectors (APDs; PerkinElmer, SPCM-AQRH-15-FC) using a cubic nonpolarizing beam-splitter (Linos) and multimode optical fibres of 100 μm diameter. The signals of the APDs were recorded in the cross-correlation mode using a digital hardware correlator device (ALV 5000/60X0 multiple tau digital real correlator) to bypass detector dead time and afterpulsing effects. Fluorescently modified pro-rhodesain was diluted to 1 nM in the buffered solutions specified above for steady-state fluorescence spectroscopy. Samples were filtered through a 0.2 μm syringe filter before measurement, transferred onto a microscope slide, and covered by a cover slip. Sample temperature was set to 37 °C using a custom-built objective heater. For each sample the accumulated measurement time of the autocorrelation function, $G(\tau)$, was 15 min. $G(\tau)$ was fitted using an analytical model for translational diffusion of a molecule that exhibits chemical relaxations (27). ACFs fitted well to a model for diffusion of a globule with either one- or two single-exponential decays:

$$G(\tau) = \frac{1}{N\left(1 + \frac{\tau}{\tau_D}\right)}\left(1 + \sum_n a_n exp\left(-\frac{\tau}{\tau_n}\right)\right)$$

where τ is the lag time, N is the average number of molecules in the detection focus, τ_D is the diffusion time constant, a_n is the amplitude of the nth relaxation, and τ_n is the according time constants. The application of a model for diffusion in two dimensions was of sufficient accuracy because the two horizontal dimensions (x, y) of the detection focus were much smaller than the lateral dimension (z) in the applied setup (27). Errors are s.e. from regression analyses.

Data availability

Raw data for experiments including but not restricted to *e.g.*, CD spectroscopy, PET fluorescence experiments, X-Ray crystallography, or MD simulation trajectories are available upon request from the corresponding authors Ute A. Hellmich, Faculty of Chemistry and Earth Sciences, Institute of Organic Chemistry and Macromolecular Chemistry, Friedrich-Schiller-University, Jena, Germany: ute.hellmich@uni-jena.de or Christian Kersten, Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg-University, Mainz, Germany: kerstec@uni-mainz.de. The X-ray structure of *T. brucei rhodesiense* pro-rhodesain C150A has been deposited in the PDB under the accession number 7AVM.

References

1. Simarro, P. P., Cecchi, G., Franco, J. R., Paone, M., Diarra, A., Ruiz-Postigo, J. A., Fèvre, E. M., Mattioli, R. C., and Jannin, J. G. (2012) Estimating and mapping the population at risk of sleeping sickness. *PLoS Negl. Trop. Dis.* **6**, 1–12

2. Kennedy, P. G. E. (2004) Human African trypanosomiasis of the CNS: Current issues and challenges. *J. Clin. Invest.* **113**, 496–504

3. Pépin, J., and Milord, F. (1994) The treatment of human African trypanosomiasis. *Adv. Parasitol.* **33**, 1–47

4. Grab, D. J., Garcia-Garcia, J. C., Nikolskaia, O. V., Kim, Y. V., Brown, A., Pardo, C. A., Zhang, Y., Becker, K. G., Wilson, B. A., De A Lima, A. P. C., Scharfstein, J., and Dumler, J. S. (2009) Protease activated receptor signaling is required for African trypanosome traversal of human brain microvascular endothelial cells. *PLoS Negl. Trop. Dis.* **3**, 1–13

5. Nikolskaia, O. V., De A Lima, A. P. C., Kim, Y. V., Lonsdale-Eccles, J. D., Fukuma, T., Scharfstein, J., and Grab, D. J. (2008) Blood-brain barrier traversal by African trypanosomes requires calcium signaling induced by parasite cysteine protease. *J. Clin. Invest.* **116**, 2739–2747

6. Mackey, Z. B., O'Brien, T. C., Greenbaum, D. C., Blank, R. B., and McKerrow, J. H. (2004) A cathepsin B-like protease is required for host protein degradation in Trypanosoma brucei. *J. Biol. Chem.* **279**, 48426–48433

7. Alsford, S., Currier, R. B., Guerra-Assunção, J. A., Clark, T. G., and Horn, D. (2014) Cathepsin-L can resist lysis by human serum in trypanosoma brucei brucei. *PLoS Pathog.* **10**, 1–12

8. Steverding, D., Sexton, D. W., Wang, X., Gehrke, S. S., Wagner, G. K., and Caffrey, C. R. (2012) Trypanosoma brucei: Chemical evidence that cathepsin L is essential for survival and a relevant drug target. *Int. J. Parasitol.* **42**, 481–488

9. Schirmeister, T., Kesselring, J., Jung, S., Schneider, T. H., Weickert, A., Becker, J., Lee, W., Bamberger, D., Wich, P. R., Distler, U., Tenzer, S., Johé, P., Hellmich, U. A., and Engels, B. (2016) Quantum chemical-based protocol for the rational design of covalent inhibitors. *J. Am. Chem. Soc.* **138**, 8332–8335

10. McKerrow, J. H. (2018) Update on drug development targeting parasite cysteine proteases. *PLoS Negl. Trop. Dis.* **12**, 4–7

11. Doyle, P. S., Zhou, Y. M., Engel, J. C., and McKerrow, J. H. (2007) A cysteine protease inhibitor cures Chagas' disease in an immunodeficient-mouse model of infection. *Antimicrob. Agents Chemother.* **51**, 3932–3939

12. Clayton, J. (2010) Chagas disease: Pushing through the pipeline. *Nature* **465**, 12–15

13. Huete-Pérez, J. A., Engel, J. C., Brinen, L. S., Mottram, J. C., and McKerrow, J. H. (1999) Protease trafficking in two primitive eukaryotes is mediated by a prodomain protein motif. *J. Biol. Chem.* **274**, 16249–16256

14. Tao, K., Stearns, A., Dong, J., Wu, Q., and Sahagian, G. G. (1994) The proregion of cathepsin L is required for proper folding, stability and ER exit. *Arch. Biochem. Biophys.* **311**, 19–27

15. Eakin, A. E., Mills, A. A., Harth, G., McKerrow, J. H., and Craik, C. S. (1992) The sequence, organization, and expression of the major cysteine protease (cruzain) from Trypanosoma cruzi. *J. Biol. Chem.* **267**, 7411–7420

16. Smith, S. M., and Gottesman, M. M. (1989) Activity and deletion analysis of recombinant human cathepsin L expressed in Escherichia coli. *J. Biol. Chem.* **264**, 20487–20495

17. Lecaille, F., Kaleta, J., and Brömme, D. (2002) Human and parasitic papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design. *Chem. Rev.* **102**, 4459–4488

18. Nishimura, Y., Kawabata, T., Furuno, K., and Kato, K. (1989) Evidence that aspartic proteinase is involved in the proteolytic processing event of procathepsin L in lysosomes. *Arch. Biochem. Biophys.* **271**, 400–406

19. Nishimura, Y., Kawabata, T., and Kato, K. (1988) Identification of latent procathepsins B and L in microsomal lumen: Characterization of enzymatic activation and proteolytic processing *in vitro*. *Arch. Biochem. Biophys.* **261**, 64–71

20. Wiederanders, B., and Kirschke, H. (1989) The processing of a cathepsin L precursor *in vitro*. *Arch. Biochem. Biophys.* **272**, 516–521

21. Ménard, R., Carmona, E., Takebe, S., Dufour, É., Plouffe, C., Mason, P., and 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

22. Caffrey, C. R., Hansell, E., Lucas, K. D., Brinen, L. S., Alvarez Hernandez, A., Cheng, J., Gwaltney, S. L., Roush, W. R., Stierhof, Y. D., Bogyo, M., Steverding, D., and McKerrow, J. H. (2001) Active site mapping, biochemical properties and subcellular localization of rhodesain, the major cysteine protease of Trypanosoma brucei rhodesiense. *Mol. Biochem. Parasitol.* **118**, 61–73

23. Kerr, I. D., Wu, P., Marion-Tsukamaki, R., Mackey, Z. B., and Brinen, L. S. (2010) Crystal structures of TbCatB and rhodesain, potential chemotherapeutic targets and major cysteine proteases of Trypanosoma brucei. *PLoS Negl. Trop. Dis.* **4**, 1–9

24. Giroud, M., Dietzel, U., Anselm, L., Banner, D., Kuglstatter, A., Benz, J., Blanc, J. B., Gaufreteau, D., Liu, H., Lin, X., Stich, A., Kuhn, B., Schuler, F., Kaiser, M., Brun, R., *et al.* (2018) Repurposing a library of human cathepsin L ligands: Identification of macrocyclic lactams as potent rhodesain and trypanosoma brucei inhibitors. *J. Med. Chem.* **61**, 3350–3369

25. Kerr, I. D., Lee, J. H., Farady, C. J., Marion, R., Rickert, M., Sajid, M., Pandey, K. C., Caffrey, C. R., Legac, J., Hansell, E., Mckerrow, J. H., Craik, C. S., Rosenthal, P. J., and Brinen, L. S. (2009) Vinyl sulfones as antiparasitic agents and a structural basis for drug design. *J. Biol. Chem.* **284**, 25697–25703

26. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2002) The protein data bank. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 899–907

27. Sauer, M., and Neuweiler, H. (2014) Probing rapid structural fluctuations of proteins and nucleic acids by single-molecule fluorescence quenching. *Fluoresc. Spectrosc. Microsc.* **1076**, 371–417

28. Pamer, E. G., Davis, C. E., and So, M. (1991) Expression and deletion analysis of the Trypanosoma brucei rhodesiense cysteine protease in Escherichia coli. *Infect. Immun.* **59**, 1074–1078

29. Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A., and Drenth, J. (1984) Structure of papain refined at 1.65 Å resolution. *J. Mol. Biol.* **179**, 233–256

30. Roy, S., Choudhury, D., Aich, P., Dattagupta, J. K., and Biswas, S. (2012) The structure of a thermostable mutant of pro-papain reveals its activation mechanism. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1591–1603

31. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2012) MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **40**, 343–350

32. Karrer, K. M., Peiffer, S. L., and Ditomas, M. E. (1993) Two distinct gene subfamilies within the family of cysteine protease genes. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3063–3067

33. Martínez, J., Henriksson, J., Rydåker, M., JoséCazzulo, J., and Pettersson, U. (1995) Genes for cysteine proteinases from Trypanosoma rangeli. *FEMS Microbiol. Lett.* **129**, 135–141

34. Mottram, J. C., North, M. J., Barry, J. D., and Coombs, G. H. (1989) A cysteine proteinase cDNA from Trypanosoma brucei predicts an enzyme with an unusual C-terminal extension. *FEBS Lett.* **258**, 211–215

35. Reif, M. M., MacH, L., and Oostenbrink, C. (2012) Molecular insight into propeptide-protein interactions in cathepsins L and O. *Biochemistry* **51**, 8636–8653

36. Vernet, T., Berti, P. J., De Montigny, C., Musil, R., Tessier, D. C., Menard, R., Magny, M. C., Storer, A. C., and Thomas, D. Y. (1995) Processing of the papain precursor. The ionization state of a conserved amino acid motif within the pro region participates in the regulation of intramolecular processing. *J. Biol. Chem.* **270**, 10838–10846

37. Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S., Cygler, M., *et al.* (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.* **15**, 5492–5503

38. Roccatano, D., Colombo, G., Fioroni, M., and Mark, A. E. (2002) Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: A molecular dynamics study. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12179–12184

39. Goretzki, B., Glogowski, N. A., Diehl, E., Duchardt-Ferner, E., Hacker, C., Gaudet, R., and Hellmich, U. A. (2018) Structural basis of TRPV4 N terminus interaction with syndapin/PACSIN1-3 and PIP 2. *Structure* **26**, 1583–1593.e5

40. Harper, E., and Berger, A. (1972) On the size of the active site in proteases: Pronase. *Biochem. Biophys. Res. Commun.* **46**, 1956–1960

41. Nomura, T., and Fujisawa, Y. (1997) Processing properties of recombinant human procathepsin L. *Biochem. Biophys. Res. Commun.* **146**, 143–146

42. Mcdonald, J. K., and Emerick, J. M. C. (1995) Purification and characterization of procathepsin L, a self-processing zymogen of Guinea pig spermatozoa that acts on a cathepsin D assay substrate. *Arch. Biochem. Biophys.* **323**, 409–422

43. von Hammerstein, F., Lauth, L. M., Hammerschmidt, S., and Wagner, A. (2019) Cis autocatalytic cleavage of glycine-linked Zika virus NS2B-NS3 protease constructs. *FEBS Lett.* **593**, 2204–2213

44. Doose, S., Neuweiler, H., and Sauer, M. (2009) Fluorescence quenching by photoinduced electron transfer: A reporter for conformational dynamics of macromolecules. *ChemPhysChem* **10**, 1389–1398

45. Heiby, J. C., Goretzki, B., Johnson, C. M., Hellmich, U. A., and Neuweiler, H. (2019) Methionine in a protein hydrophobic core drives tight interactions required for assembly of spider silk. *Nat. Commun.* **10**, 4378

46. Johe, P., Jung, S., Endres, E., Kersten, C., Zimmer, C., Ye, W., Sönnichsen, C., Hellmich, U. A., Sotriffer, C., Schirmeister, T., and Neuweiler, H. (2021) Warhead reactivity limits the speed of inhibition of the cysteine protease Rhodesain. *ACS Chem. Biol.* (accepted) https://doi.org/10.1021/acschembio.0c00911

47. Powers, J. C., Asgian, J. L., and James, K. E. (2002) Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem. Rev.* **102**, 4639–4750

48. Groves, M. R., Taylor, M. A. J., Scott, M., Cummings, N. J., Pickersgill, R. W., and Jenkins, J. A. (1996) The prosequence of procaricain forms an α-helical domain that prevents access to the substrate-binding cleft. *Structure* **4**, 1193–1203

49. Ettari, R., Tamborini, L., Angelo, I. C., Micale, N., Pinto, A., De Micheli, C., and Conti, P. (2013) Inhibition of rhodesain as a novel therapeutic modality for human African trypanosomiasis. *J. Med. Chem.* **56**, 5637–5658

50. Palmer, J. T., Rasnick, D., Klaus, J. L., and Brömme, D. (1995) Vinyl sulfones as mechanism-based cysteine protease inhibitors. *J. Med. Chem.* **38**, 3193–3196

51. Koeller, C. M., and Bangs, J. D. (2019) Processing and targeting of cathepsin L (TbCatL) to the lysosome in Trypanosoma brucei. *Cell. Microbiol.* **21**, 1–16

52. Gunner, M. R., Mao, J., Song, Y., and Kim, J. (2006) Factors influencing the energetics of electron and proton transfers in proteins. What can be learned from calculations? *Biochim. Biophys. Acta Bioenerg.* **1757**, 942–968

53. Flocco, M. M., and Mowbray, S. L. (1995) Strange bedfellows: Interactions between acidic side-chains in proteins. *J. Mol. Biol.* **254**, 96–105

54. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132

55. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877

56. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., *et al.* (2010) Phenix: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221

57. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132

58. Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260

59. Case, D. A., Cerutti, D. S., Cheatham, III, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Greene, D., Homeyer, N., Izadi, S., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Lin, C., *et al.* (2017) *AMBER*, University of California, San Francisco

60. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935

61. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theor. Comput* **11**, 3696–3713

62. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38

63. Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., Buch, R., Fiorin, G., Hénin, J., Jiang, W., McGreevy, R., Melo, M. C. R., Radak, B. K., Skeel, R. D., Singharoy, A., *et al.* (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**

64. Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261**, 470–489

65. Bietz, S., Urbaczek, S., Schulz, B., and Rarey, M. (2014) Protoss: A holistic approach to predict tautomers and protonation states in protein-ligand complexes. *J. Cheminform.* **6**, 1–12

66. Halgren, T. A. (1996) Merck molecular force field. *J. Comput. Chem.* **17**, 490–519

67. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., and Stahl, M. T. (2010) Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *J. Chem. Inf. Model.* **50**, 572–584

68. Micsonai, A., Wien, F., Kernya, L., Lee, Y. H., Goto, Y., Réfrégiers, M., and Kardos, J. (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3095–E3103