Natively Inhibited *Trypanosoma brucei* Cathepsin B Structure Determined by Using an X-ray Laser

Lars Redecke,1,2* Karol Nass,3,4* Daniel P. DePonte,3 Thomas A. White,3 Dirk Rehders,1 Anton Barty,3 Francesco Stellato,2 Mengning Liang,3 Thomas R.M. Barends,5,6 Sébastien Boutet,7 Garth J. Williams,7 Marc Messerschmidt,7 M. Marvin Seibert,7 Andrew Aquila,3 David Arnlund,8 Sasa Bajt,5 Torsten Barth,10 Michael J. Bogan,11 Carl Caleman,3 Tzu-Chiao Chao,12 R. Bruce Doak,13 Holger Fleckenstein,5 Matthias Frank,14 Raimund Fromme,12 Lorenzo Galli,3 Ingo Grotjohann,12 Mark S. Hunter,12† Linda C. Johansson,8 Stephan Kassemeyer,5,6 Gergely Katona,9 Richard A. Kirian,3,15 Rudolf Koopmann,10 Chris Kupitz,12 Lukas Lomb,5,6 Andrew V. Martin,3 Stefan Mogk,10 Richard Neutze,8 Robert L. Shoeman,5,6 Jan Steinbrener,5,6 Nicusor Timneanu,15 Dingjie Wang,13 Uwe Weiерstall,13 Nadia A. Zatsepin,13 John C.H. Spence,13 Petra Fromme,12 Ilme Schlichting,5,6 Michael Duszenko,10 Christian Betzel,16‡ Henry N. Chapman3,4‡

1Joint Laboratory for Structural Biology of Infection and Inflammation, Institute of Biochemistry and Molecular Biology, University of Hamburg, and Institute of Biochemistry, University of Lübeck, at Deutsches Elektronen-Synchrotron (DESY), Notkestrasse 85, 22607 Hamburg, Germany. 2German Centre for Infection Research, University of Lübeck, 23538 Lübeck, Germany. 3Center for Free-Electron Laser Science (CFEL), DESY, Notkestrasse 85, 22607 Hamburg, Germany. 4Department of Physics, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. 5Max-Planck-Institut für medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany. 6Max Planck Advanced Study Group, Center for Free-Electron Laser Science (CFEL), DESY, Notkestrasse 85, 22607 Hamburg, Germany. 7Linac Coherent Light Source, Stanford Linear Accelerator Center (SLAC) National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. 8Department of Chemistry and Molecular Biology, University of Gothenburg, SE-405 30 Gothenburg, Sweden. 9Photon Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. 10Interfaculty Institute of Biochemistry, University of Tübingen, Hoppe-Seyler-Strasse4, 72076 Tübingen, Germany. 11PULSE Institute,SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. 12Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA. 13Department of Physics, Arizona State University, Tempe, AZ 85287, USA. 14Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA. 15Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, SE-75124 Uppsala, Sweden. 16Institute of Biochemistry and Molecular Biology, University of Hamburg, at DESY, Notkestrasse 85, 22607 Hamburg, Germany.

†Present address: Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA.
‡To whom correspondence should be addressed. E-mail: henry.chapman@desy.de (H.N.C.) or christian.betzel@uni-hamburg.de (C.B.)

*The Trypanosoma brucei* cysteine protease cathepsin B (TbCatB), which is involved in host protein degradation, is a promising target to develop new treatments against sleeping sickness, a fatal disease caused by this protozoan parasite. The structure of the mature, active form of TbCatB has so far not provided sufficient information for the design of a safe and specific drug against *T. brucei*. By combining two recent innovations, in vivo crystallization and serial femtosecond crystallography, we obtain the room-temperature 2.1 Å resolution structure of the fully glycosylated precursor complex of TbCatB. The structure reveals the mechanism of native TbCatB inhibition and demonstrates that new biomolecular information can be obtained by the “diffraction before destruction” approach of x-ray free-electron lasers from hundreds of thousands of individual microcrystals.

Over 60 million people are affected by human African trypanosomiasis (HAT), also known as sleeping sickness, which causes approximately 30,000 deaths per year (1). The protozoan parasite *Trypanosoma brucei*, transmitted by tsetse flies, infects the blood and the lymphatic system before invading the brain. Severe clinical manifestations occur within weeks or months. Current treatments of HAT rely on anti-parasitic drugs developed during the last century, without knowledge of the biochemical pathways. These treatments are limited in their efficacy and safety, and drug resistance is increasing (2–4). Thus new compounds that selectively inhibit vital pathways of the parasite without adverse affects to the host are urgently required. A promising strategy is to target lysosomal papain-like cysteine proteases that are involved in host-protein degradation, such as cathepsin B (5). The knockdown of this essential enzyme in *T. brucei* resulted in clearance of parasites from the blood of infected mice and cured the infection (6), qualifying cathepsin B as a suitable drug target. Cysteine proteases are synthesized as inactive precursors with N-terminal propeptides that act as potent and selective intrinsic inhibitors until the proteases enter the lysosome (7) where the propeptide is released, forming the mature active enzyme. Such native propeptide-inhibited structures have been used to develop species-specific protease inhibitors against proteases of other *Trypanosoma* species, e.g., cruzipain of *T. cruzi* (causing human Chagas disease in America), and congopain of *T. congolense* (causing nagana in cattle) (8, 9). This approach could not be explored for *T. brucei* cathepsin B (TbCatB) due to the lack of structural information on the mode of propeptide inhibition and the high extent of structural conservation at the active site between mammalian and trypanosomane cathepsin B (10–12). Previously-solved mature *T. brucei* and human CatB structures identified differences at the S2 and in part of the S1′ subsite of the substrate binding cleft (Fig. 1C), and have been suggested as possible targets for the development of species-specific CatB inhibitors (10). Together with the natively-inhibited human procathepsin B structure (13) our work fills the gap to understand the structural basis for species-specific inhibition.

The growth of large well-ordered protein crystals is one of the major bottlenecks in structure determination by x-ray crystallography with important biological targets such as integral membrane proteins and post-translationally modified proteins proving particularly challenging to crystallize (14). Sizable crystals are required to obtain measurable high-resolution diffraction data within an exposure that is limited by the accumulation of radiation damage (15). While microfocus beamlines enable the collection of diffraction data from micron-sized protein crystals...
(16), the tolerable dose limit of less than 30 MGy for cryogenically cooled protein crystals remains, limiting the achievable signal. The tolerable dose for room temperature measurements is about 1 MGy (15). We have previously shown that micron-sized crystals of glycosylated TbCatB spontaneously form in insect cells during protein overexpression (11). Such crystals are extremely well suited for the new method of serial femtosecond crystallography (SFX) (17). X-ray FEL pulses of less than 100 fs duration allow the dose to individual crystals to exceed the ~1 MGy limit by over a thousand times due to the "diffraction-before-destruction" principle (17, 18). Diffraction data are recorded for each pulse while crystals are continually replenished by flowing a microcrystal suspension in aqueous buffer across the FEL beam in vacuum in a fine liquid jet.

The Coherent X-ray Imaging (CXI) beamline (19) at the LCLS enables high-resolution data collection using the SFX approach (20). We used this instrument to obtain diffraction data from in vivo grown crystals of TbCatB produced in the baculovirus-Sf9 insect cell system (11) (Fig. 1, A and B). Crystals with average dimensions of approx. 0.9 × 0.9 × 11 μm³ (fig. S1) were flowed in a 4 μm diameter column of buffer fluid at room temperature, at a flow rate of 10 μl/minute, using a liquid microjet (21). X-ray pulses from the FEL were focused onto this column to a spot of 4 μm diameter, prior to the breakup of the jet into drops (fig. S2). Single-pulse diffraction patterns of randomly oriented crystals, that by chance were present in the interaction region, were recorded at 120 Hz repetition rate by a Cornell-SLAC pixel array detector (CSPAD) (19, 20) at 9.4 keV photon energy (1.3 Å wavelength). An average pulse energy of 0.6 mJ at the sample (4 × 10¹¹ photons per pulse) with a duration of less than 40 fs gave an x-ray intensity above 10¹⁷ W/cm² and a maximum dose of about 31 MGy per crystal. This dose exceeds that tolerable at room temperature using conventional data collection approaches due to the radically different time scales and dose rates. The electron and photon beam parameters are summarized in table S1. Almost 4 million individual “snap-shot” diffraction patterns were collected. Of these, 293,195 snapshots contained crystal diffraction (fig. S3), from which 178,875 (61%) diffraction patterns were indexed and combined into a three-dimensional dataset of structure factors by “Monte Carlo” integration of partial reflections from each randomly oriented microcrystal (22, 23). The resulting complete set of structure factors contains 25,969 reflections in a resolution range from 20 to 2.1 Å. The high quality of the merged dataset is indicated by a 25,969 reflections in a resolution range from 20 to 2.1 Å. The high quality of the merged dataset is indicated by a

[(a) of the merged dataset is indicated by a](http://www.sciencemag.org/content/early/recent)

The SFX TbCatB structure shows that the inhibitory mechanism observed for mammalian papain-like protease-precursors remains largely conserved in *T. brucei*, including the overall conformation of the propeptide (supplementary text S3, fig. S5). The active site of TbCatB is blocked by the propeptide, which tightly binds in a reversed direction compared to the substrate (fig. S6) (25). A detailed comparison of the propeptide-enzyme contact area with that observed for human procathepsin B (Protein Data Bank ID, 3PBH) (13) indicates an interface enlarged by approximately 310 Å² within the TbCatB-propeptide complex (supplementary text S4). Tight binding of the *T. brucei* propeptide to the enzyme interface through three conserved epitopes is maintained by 21 intermolecular polar and ionic interactions (fig. S7). These are eight fewer interactions than for human procathepsin B.

The most significant difference between the structures of mature TbCatB and the natively inhibited propeptide complex occurs in the "occluding loop" region (residues 193 to 207) (fig. S8). This highly flexible loop is a structural element characteristic of cathepsin B-like enzymes that confers exopeptidase activity (removal of dipeptide units from the C terminus of the substrate), supplementing the endopeptidase (non-terminal substrate cleavage) activity common to all papain-like proteases (26). In mature CatB the occluding loop is in the “closed” conformation, burying an essential part of the prime substrate (S1’ and S2’ positions) at the substrate cleft (Fig. 3A) (27) and competing for binding with large substrates with an affinity that depends sensitively on pH (28). As a consequence of propeptide binding, the occluding loop is reoriented into an “open” conformation, exposing the entire S1’ and S2’ subsite of the substrate binding cleft in TbCatB (Fig. 3B). This mirrors the “open” and “closed” conformations observed in human CatB, however the trypanosomal occluding loop is more rigid, thus the displaced loop segment comprises only four residues rather than the ten observed in human CatB (13). This results in a narrower exposed S2’ subsite approximately 8.5 Å wide compared to approximately 11.9 Å for human CatB (supplementary text S5). In particular the side chain of His194 is only slightly shifted compared to the “closed” loop conformation and still extends into the open cleft. Thus, His194 not only establishes steric constraints for the substrates, but also provides a prominent polar anchor in the otherwise largely hydrophobic S2’ and S1’ subsites that are highly conserved between trypanosome and human CatB (fig. S9). In human CatB the larger exposed S2’ subsite in the "open" loop conformation is less restricted by the corresponding His189 residue. This suggests that smaller hydrophobic substituents could target the prime site (S1’ and S2’ positions) in TbCatB, which is also supported by the propeptide structures: the bulky Phe residue that sticks into the S2’ subsite of human CatB is substituted by the smaller Met of the *T. brucei* propeptide.

The occluding loop conformation is further stabilized by two carbohydrate structures identified in the TbCatB complex, as shown in Fig. 4. The enzyme carbohydrate chain interacts with both strands of the occluding loop at the loop termini (Fig. 4A), supporting the increased loop rigidity in TbCatB mentioned above. The propeptide carbohydrate connects the tip of the "open" occluding loop and stabilizes the "open" conformation (Fig. 4B). Although N-linked oligosaccharide substitution has been detected in human procathepsin B, the predicted glycosylation sites differ from our observations in TbCatB (28, 29). Therefore, it is unlikely that the occluding loop is stabilized in a similar way in the human case (supplementary text S6). Differential glycosylation between the human and *T. brucei* precursors along with the differences in the occluding loop conformation could be exploited for synthetic parasite-specific inhibition.

As illustrated by the room-temperature glycosylated TbCatB-propeptide structure determined here, the combination of in vivo grown
microcrystals with the “diffraction before destruction” technique of x-ray free-electron lasers provides a compelling path to obtain macromolecular structures from challenging samples. This methodology could vastly speed up structure determination by removing the need for large well-diffracting crystals and providing a suitable amount of crystals of post-translationally modified proteins, in their biologically functional form.

References and Notes
Acknowledgments: Experiments were carried out in February 2011 at the Linac Coherent Light Source (LCLS), a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. This work was supported by the following agencies: the German Federal Ministry for Education and Research (grants 01KX0806 and 01KX0807); the Hamburg Ministry of Science and Research and Joachim Herz Stiftung as part of the Hamburg Initiative for Excellence in Research and the Hamburg School for Structure and Dynamics in Infection (SDI); the DFG Cluster of Excellence “Inflammation at Interfaces” (EXC 306); the Deutsche Forschungsgemeinschaft (DFG); the Landesgraduiertenförderung Baden-Württemberg; the Max Planck Society, the Swedish Research Council; the Swedish Strategic Research Foundation; the Swedish Foundation for International Cooperation in Research and Higher Education; the U.S. Department of Energy Office of Basic Energy Sciences through the Photon Ultrafast Laser Science and Engineering (PULSE) Institute at the Stanford Linear Accelerator Center (SLAC); the U.S. Department of Energy through Lawrence Livermore National Laboratory under the contract DE-AC52-07NA27344 and supported by the UCOP Lab Fee Program (award no. 118036); the U.S. National Science Foundation (award MCB-1021557 and MCB-1120997) and the National Institutes of Health (award 1R01GM095583). H.N.C., J.C.H.S., S.B., P.F., I.S., M.D., L.R., and C.B. conceived the experiment, which was designed with A.B., R.A.K., J.C.H.S., D.P.D., U.W., R.B.D., M.J.B., R.L.S. and H.F.; R.K., S.M., and T.B. performed the in vivo crystallization experiments under supervision of M.D.; FEL samples were prepared and characterized by F.S., K.N., L.R, and D.R. under supervision of C.B. and H.N.C.; SFX experiments were carried out by K.N., L.R., H.N.C., D.P.D., F.S., M.L., T.A.W., A.A., M.J.B., U.W., A.B., L.G., S. Bait, R.A.K., R.B.D., R.L.S., L.L., D.A., L.C.J., C.C., R.N., G.K., C.K., P.F., D.W., I.G., R.F., T.C., N.A.Z., N.T., M.S.H., M.F., J.S., S.B., M.M., M.M.S. and G.J.W. Beamline setup was done by S.B., G.J.W., and M.M. The development and operation of the sample delivery system was performed by R.B.D., D.P.D., U.W., R.L.S., L.L., J.S., and J.C.H.S.; K.N., T.A.W., R.A.K., A.A., A.V.M., L.L., S.K., T.R.M.B., I.S., and H.N.C. analyzed the data. K.N., L.R., and C.B performed molecular replacement, refined the structure and calculated the electron density maps. The manuscript was prepared by L.R., C.B., K.N., M.D, I.S., A.B., and H.N.C. with discussions and improvements from all authors. The structure factors and coordinates have been deposited with the Protein Data Bank (accession code 4HWY). The Arizona Board of Regents, acting for and on behalf of Arizona State University and in conjunction with R.B.D., U.W., D.P.D., and J.C.H.S., has filed U.S. and international patent applications on the nozzle technology applied herein. One of the patents was granted on 25 September 2012, U.S. 8,272,576 “Gas dynamic virtual nozzle for generation of microscopic droplet streams.”

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1229663/DC1
Materials and Methods
Supplementary Text S1 to S6
Figures S1 to S10
Tables S1 to S5
References (30–42)

03 September 2012; accepted 14 November 2012
Published online 29 November 2012
10.1126/science.1229663
Fig. 1. In vivo grown crystals and 3D structure of the TbCatB-propeptide complex. (A) Transmission EM of an infected SF9 insect cell showing a crystal of overexpressed TbCatB inside the rough endoplasmic reticulum that is cut perpendicular to its long axis. N, nucleus; L, lysosome; C, crystal; CM, cell membrane. (B) Scanning EM of a single TbCatB crystal after isolation. (C) Cartoon plot of the TbCatB-propeptide complex exhibiting the typical papain-fold of cathepsin B-like proteases (supplementary text S1). Grey, R-domain; blue, L-domain; beige, occluding loop. The native propeptide (green) blocks the active site. The subsites of the substrate binding cleft N- (non-prime: S2, S3) and C-terminal (prime: S1′, S2′) to the active site (S1) have been identified by comparison with the human CatB structure (13) and labeled (red) according to Schechter and Berger (27). Two N-linked carbohydrate structures (yellow) consist of N-acetylglucosamine (NAG) and mannose (MAN) residues (yellow, carbon atoms; blue, nitrogen atoms; red, oxygen atoms).

Fig. 2. Quality of the calculated electron density. (A) Surface representation of the TbCatB-propeptide complex solved by molecular replacement using the mature TbCatB structure (11) as a search model. The solution revealed additional electron density ($2F_o - F_c$, 1σ, blue) of the propeptide (green) that is bound to the V-shaped substrate binding cleft and of two carbohydrate structures (yellow) N-linked to the propeptide (B) and to the mature enzyme (C). The propeptide as well as both carbohydrates are well-defined within the electron density map (blue), confirming that the phases are not biased by the search model. Color codes correspond to Fig. 1C.
Fig. 3. Occluding loop conformations of mature and propeptide inhibited TbCatB. (A) Surface representation of mature TbCatB (11) showing the occluding loop (rigid part, beige; flexible part, red) in the "closed" conformation. A loop segment blocks the S2' site and part of the S1' site of the substrate binding cleft. (B) Propeptide binding (green) shifts the flexible occluding loop segment (red) into an "open" conformation, exposing the entire prime subsite. The insets compare the Tb (grey) and human (cyan) surface representations of mature and propeptide bound CatB. Almost superimposable Cα chains of the human (blue) and Tb (beige/red) occluding loops indicate a conservation of the "closed" loop conformation in the mature enzymes (A), while the "open" conformations show significant differences (B). Four H-bonds maintained during conformational transition restricted the flexible loop segment to four residues (red) in TbCatB, opening a crevice of approximately 8.5 Å. In contrast, only one H-bond is maintained in the "open" loop conformation of human procathepsin B (13). Thus, the mobile segment of the human occluding loop (blue) comprises ten residues, exposing an enlarged occluding loop crevice (supplementary text S5). Residues are labeled according to TbCatB, referring to the corresponding human CatB numbering in parenthesis, if applicable.
Fig. 4. Glycosylation of the TbCatB-propeptide complex. (A) Enzyme carbohydrate structure comprising two N-acetylglucosamine (NAG) and one mannose (MAN) residue (yellow) N-linked to Asn216 C-terminal of the occluding loop (beige). The carbohydrate structure connects both occluding loop strands by two direct and one water-bridged H-bond (black dashed lines). (B) Propeptide glycosylation site comprising two NAG units (yellow) at Asn58 within the kinked region of the propeptide (green). The propeptide carbohydrate structure forms an H-bond to Gln57 of the propeptide and two H-bonds to Ser196 at the tip of the occluding loop (beige), stabilizing its “open” conformation. Color codes correspond to Fig. 1C.