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**[3], **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**[9], **Torsten Barth**[10], **Michael J. Bogan**[11], **Carl Caleman**[3], **Tzu-Chiao Chao**[12], **R. Bruce Doak**[13], **Holger Fleckenstein**[3], **Matthias Frank**[14], **Raimund Fromme**[12], **Lorenzo Galli**[3,4], **Ingo Grotjohann**[12], **Mark S. Hunter**[12,*], **Linda C. Johansson**[8], **Stephan Kassemeyer**[5,6], **Gergely Katona**[8], **Richard A. Kirian**[3,13], **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 Weierstall**[13], **Nadia A. Zatsepin**[13], **John C. H. Spence**[13], **Petra Fromme**[12], **Ilme Schlichting**[5,6], **Michael Duszenko**[10], **Christian Betzel**[16,‡], and **Henry N. Chapman**[3,4,‡]

[1]Joint 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.

[2]German Centre for Infection Research, University of Lübeck, 23538 Lübeck, Germany.

[3]Center for Free-Electron Laser Science (CFEL), DESY, Notkestrasse 85, 22607 Hamburg, Germany.

[4]Department of Physics, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany.

[5]Max-Planck-Institut für medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany.

[6]Max Planck Advanced Study Group, Center for Free-Electron Laser Science (CFEL), DESY, Notkestrasse 85, 22607 Hamburg, Germany.

[7]Linac Coherent Light Source, Stanford Linear Accelerator Center (SLAC) National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA.

‡To whom correspondence should be addressed. henry.chapman@desy.de or christian.betzel@unihamburg.de.
*Present address: Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA.

[8]Department of Chemistry and Molecular Biology, University of Gothenburg, SE-405 30 Gothenburg, Sweden.

[9]Photon Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany.

[10]Interfaculty Institute of Biochemistry, University of Tübingen, Hoppe-Seyler-Strasse 4, 72076 Tübingen, Germany.

[11]Photon Ultrafast Laser Science and Engineering (PULSE) Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA.

[12]Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA.

[13]Department of Physics, Arizona State University, Tempe, AZ 85287, USA.

[14]Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA.

[15]Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, SE-75124 Uppsala, Sweden.

[16]Institute of Biochemistry and Molecular Biology, University of Hamburg, at DESY, Notkestrasse 85, 22607 Hamburg, Germany.

[#] These authors contributed equally to this work.

Abstract

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 obtained the room-temperature 2.1 angstrom 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 ~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 antiparasitic 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 papainlike 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), which qualify 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 and forms 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) because of the lack of structural information on the mode of propeptide inhibition and the large extent of structural conservation at the active site between mammalian and trypanosome cathepsin B (10–12). Previously solved mature *T. brucei* and human CatB structures show 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 posttranslationally 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). Although 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, which limits 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 over-expression (11). Such crystals are extremely well suited for the new method of serial femtosecond crystallography (SFX) (17). X-ray free-electron laser (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 because of the "diffraction-before-destruction" principle (17, 18). Diffraction data are recorded for each pulse as crystals are continually replenished by a microcrystal suspension in aqueous buffer flowing across the FEL beam in a vacuum in a fine liquid jet.

The Coherent X-ray Imaging (CXI) beamline (19) at the Linac Coherent Light Source (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-infected *Spodoptera frugiperda* (baculovirus-Sf9) insect cell system (11) (Fig. 1, A and B). Crystals with average dimensions of about 0.9 by 0.9 by 11 μm^3 (fig. S1) were sent in a 4-μm-diameter column of buffer fluid at room temperature, at a flow rate of 10 μl/minute, by using a liquid microjet (21). X-ray pulses from the FEL were focused onto this column to a spot 4 μm in diameter, before 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 a 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 (4×10^{11} photons per pulse) with a duration of less than 40 fs gave an x-ray intensity above 10^{17} W/cm^2 and a maximum dose of about 31 MGy per crystal. This dose exceeds that tolerable at room temperature with conventional data collection approaches because of 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 snap-shots contained crystal diffraction (fig. S3), from which 178,875 (61%) diffraction patterns were indexed and combined into a three-dimensional data set 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 data set is indicated by an R_{split} of 10.2% (which is a quality measure for SFX instead of R_{merge}) (23). Data statistics are summarized in table S2, table S3, and fig. S4. The structure was solved by molecular replacement using the coordinates of the previously determined in vitro crystallized mature TbCatB structure (Protein Data Bank ID, 3MOR) (11) as a search model.

The refined SFX TbCatB structure (R factor = 18.1%, R_{free} = 21.4%) shares the papainlike fold that is characteristic of cathepsin B–like proteases (Fig. 1C and supplementary text S1) (24), with a root mean square deviation of 0.4 Å for equivalent C atoms of the mature TbCatB structure determined at 100 K and refined to 2.55 Å resolution (11). The molecular replacement solution reveals electron density that is not part of the search model, which we identified as the coordinated, cleaved main part of the propeptide (residues 26 to 72) (Fig. 2A), and as two-carbohydrate structures (Fig. 2, B and C). Proteolytic cleavage of the expressed precursor occurs within the propeptide between Ser[78] and Ile[79], as revealed by mass spectrometry, which leaves 15 propeptide residues bound to the N terminus of mature TbCatB (supplementary text S2). The preceding residues Lys[73] to Ser[78] are disordered in the crystal structure, owing to a rise in flexibility, which shows up as a gap in the electron density in this region of the propeptide. The cleavage may be part of the initial maturation step within the activation process of TbCatB. The final model of glycosylated TbCatB in complex with its processed, but still-bound, propeptide contains 62 propeptide residues and 247 mature enzyme residues, as well as 98 solvent and 5 carbohydrate molecules. No electron density is observed for 11 flexible amino acid side chains or the eight atoms of the carbohydrate structures.

The SFX TbCatB structure shows that the inhibitory mechanism observed for mammalian papainlike protease-precursors remains largely conserved in *T. brucei*, including the overall conformation of the propeptide (supplementary text S3 and fig. S5). The active site of TbCatB is blocked by the propeptide, which tightly binds in a direction the reverse of the substrate's (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 ~310 Å^2 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), which supplements the endopeptidase (nonterminal substrate cleavage) activity common to all papainlike proteases (26). In mature CatB, the occluding loop is in the "closed" conformation and buries an essential part of the prime subsite (S1′ and S2′ positions) at the substrate cleft (Fig. 3A) (27) and competes 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 confirmations observed in human CatB; however, the trypanosomal occluding loop is more rigid. The displaced loop segment comprises only 4 residues rather than the 10 observed in human CatB (13). This results in a narrower exposed S2′ subsite ~8.5 Å wide compared with ~11.9 Å for human CatB (supplementary text S5). In particular, the side chain of His[194] is only slightly shifted compared with the closed loop conformation and still extends into the open cleft. Thus, His[194] 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 His[189] 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 replaced 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), which supports 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 FELs 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 posttranslationally modified proteins, in their biologically functional form.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References and Notes

1. Frearson JA, et al. Nature. 2010; 464:728. [PubMed: 20360736]

2. Barrett MP, Boykin DW, Brun R, Tidwell RR. Br. J. Pharmacol. 2007; 152:1155. [PubMed: 17618313]

3. Alsford S, et al. Nature. 2012; 482:232. [PubMed: 22278056]

4. Fairlamb AH. Trends Parasitol. 2003; 19:488. [PubMed: 14580959]

5. Bryant C, et al. Bioorg. Med. Chem. Lett. 2009; 19:6218. [PubMed: 19773167]

6. Abdulla MH, et al. PLoS Negl. Trop. Dis. 2008; 2:e298. [PubMed: 18820745]

7. Lecaille F, Kaleta J, Brömme D. Chem. Rev. 2002; 102:4459. [PubMed: 12475197]

8. Yamamoto Y, Kurata M, Watabe S, Murakami R, Takahashi SY. Curr. Protein Pept. Sci. 2002; 3:231. [PubMed: 12188906]

9. Dubin G. Cell. Mol. Life Sci. 2005; 62:653. [PubMed: 15770418]

10. Kerr ID, Wu P, Marion-Tsukamaki R, Mackey ZB, Brinen LS. PLoS Negl. Trop. Dis. 2010; 4:e701. [PubMed: 20544024]

11. Koopmann R, et al. Nat. Methods. 2012; 9:259. [PubMed: 22286384]

12. Tomoo K. Curr. Top. Med. Chem. 2010; 10:696. [PubMed: 20337581]

13. Podobnik M, Kuhelj R, Turk V, Turk D. J. Mol. Biol. 1997; 271:774. [PubMed: 9299326]

14. Bill RM, et al. Nat. Biotechnol. 2011; 29:335. [PubMed: 21478852]

15. Southworth-Davies RJ, Medina MA, Carmichael I, Garman EF. Structure. 2007; 15:1531. [PubMed: 18073104]

16. Riekel C. J. Synchrotron Radiat. 2004; 11:4. [PubMed: 14646120]

17. Chapman HN, et al. Nature. 2011; 470:73. [PubMed: 21293373]

18. Barty A, et al. Nat. Photonics. 2012; 6:35. [PubMed: 24078834]

19. Boutet S, Williams GJ. New J. Phys. 2010; 12:035024.

20. Boutet S, et al. Science. 2012; 337:362. [PubMed: 22653729]

21. Weierstall U, Spence JC, Doak RB. Rev. Sci. Instrum. 2012; 83:035108. [PubMed: 22462961]

22. Kirian RA, et al. Opt. Express. 2010; 18:5713. [PubMed: 20389587]

23. White TA, et al. J. Appl. Cryst. 2012; 45:335.

24. McGrath ME. Annu. Rev. Biophys. Biomol. Struct. 1999; 28:181. [PubMed: 10410800]

25. Chagas JR, Ferrer-Di Martino M, Gauthier F, Lalmanach G. FEBS Lett. 1996; 392:233. [PubMed: 8774851]

26. Illy C, et al. J. Biol. Chem. 1997; 272:1197. [PubMed: 8995421]

27. Schechter I, Berger A. Biochem. Biophys. Res. Commun. 1967; 27:157. [PubMed: 6035483]

28. Quraishi O, et al. Biochemistry. 1999; 38:5017. [PubMed: 10213604]

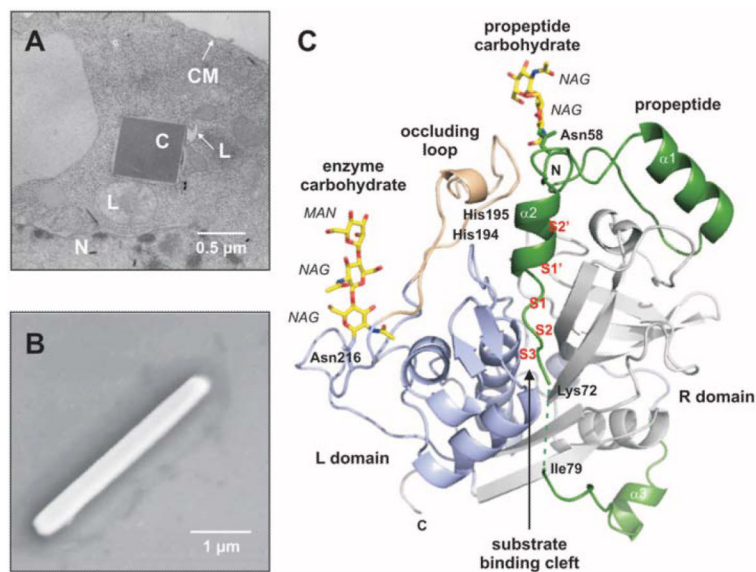29. Chen Y, Plouffe C, Ménard R, Storer AC. FEBS Lett. 1996; 393:24. [PubMed: 8804416]

Fig. 1.
In vivo grown crystals and three-dimensional structure of the TbCatB-propeptide complex.
(**A**) Transmission electron microscopy (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-like fold of cathepsin B–like proteases
(supplementary text S1). Gray, R domain; blue, L domain; beige, occluding loop. The native
propeptide (green) blocks the active site. The subsites of the substrate-binding cleft N-
terminal (nonprime: 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
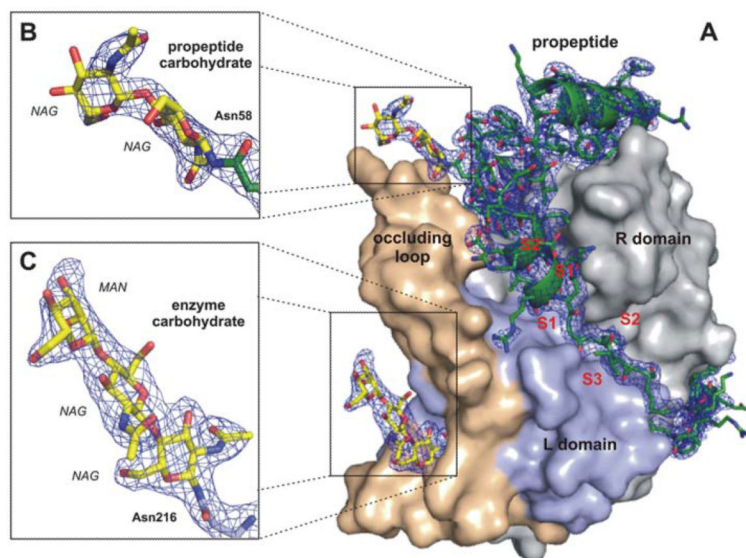atoms; blue, nitrogen atoms; red, oxygen atoms).

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 ($2F_{\text{obs}} - F_{\text{calc}}$, 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), which confirms that the phases are not biased by the search model. Color codes correspond to Fig. 1C.
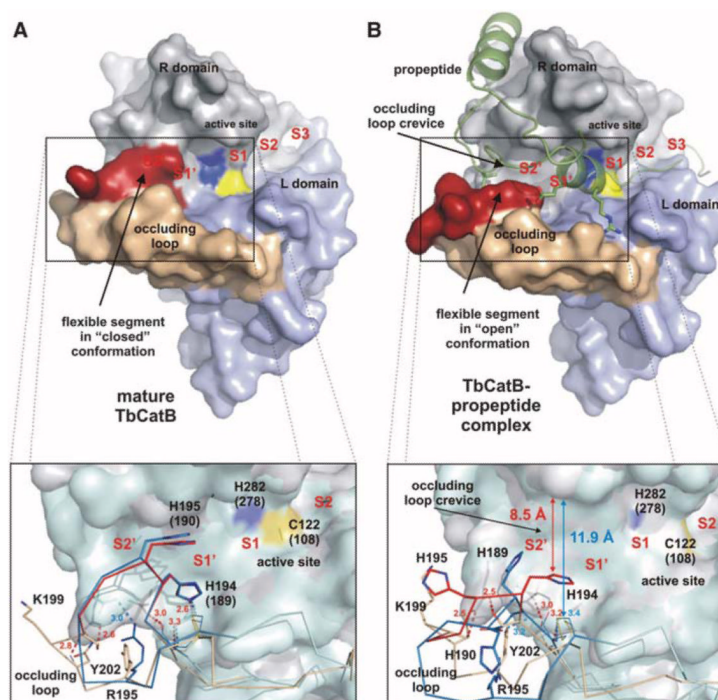
Fig. 3.
Occluding loop conformations of mature and propeptide-inhibited TbCatB. (**A**) Surface representation of mature TbCatB (11) showing the occluding loop (rigid part, beige; flexible part, red) in the closed conformation. A loop segment blocks the S2 site and part of the S1 site of the substrate-binding cleft. (**B**) Propeptide binding (green) shifts the flexible occluding loop segment (red) into an open conformation, which exposes the entire prime subsite. The insets compare the Tb (gray) and human (cyan) surface representations of mature and propeptide bound CatB. Almost superimposable C chains of the human (blue) and Tb (beige/red) occluding loops indicate a conservation of the closed loop conformation in the mature enzymes (A), whereas 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, which opens a crevice of ~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 10 residues and exposes an enlarged occluding loop crevice (supplementary text S5). Residues are labeled according to TbCatB, referring to the corresponding human CatB numbering in parentheses, if applicable.
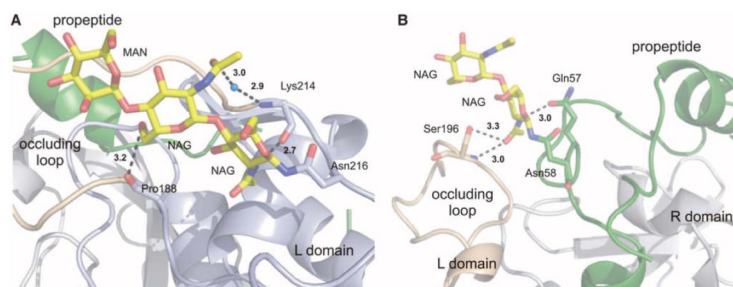
Fig. 4.
Glycosylation of the TbCatB-propeptide complex. (**A**) Enzyme carbohydrate structure comprising two NAG and one MAN residue (yellow) N-linked to Asn[216] C-terminal of the occluding loop (beige). The carbohydrate structure connects both occluding loop strands by two direct and one water-bridged H bond (black dashed lines). (**B**) Propeptide glycosylation site comprising two NAG units (yellow) at Asn[58] within the kinked region of the propeptide (green). The propeptide carbohydrate structure forms an H bond to Gln[57] of the propeptide and two H bonds to Ser[196] at the tip of the occluding loop (beige), which stabilize its open conformation. Color codes correspond to Fig. 1C.