THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284, NO. 6, pp. 3976–3984, February 6, 2009

Structure of Tripeptidyl-peptidase I Provides Insight into the Molecular Basis of Late Infantile Neuronal Ceroid Lipofuscinosis*[S]

Received for publication, September 8, 2008, and in revised form, November 10, 2008 Published, JBC Papers in Press, November 26, 2008, DOI 10.1074/jbc.M806947200

Aritra Pal[‡1], Ralph Kraetzner[§1], Tim Gruene[‡], Marcel Grapp[§], Kathrin Schreiber[§], Mads Grønborg[¶], Henning Urlaub[¶], Stefan Becker[‖], Abdul R. Asif[**], Jutta Gärtner[§], George M. Sheldrick[‡], and Robert Steinfeld[§2]

*From the [‡]Department of Structural Chemistry, University of Göttingen, Tammannstrasse 4, 34077 Göttingen, the [§]Department of Pediatrics and Pediatric Neurology, University of Göttingen, Robert-Koch-Strasse 40, 37075 Göttingen, the [¶]Bioanalytical Mass Spectrometry Group and the [‖]Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Fassberg 11, 37077 Göttingen, and the **Proteomics & Mass Spectrometry Facility, Department of Clinical Chemistry, University of Göttingen, Robert-Koch-Strasse 40, 37075 Göttingen, Germany*

Late infantile neuronal ceroid lipofuscinosis, a fatal neurodegenerative disease of childhood, is caused by mutations in the *TPP1* gene that encodes tripeptidyl-peptidase I. We show that purified TPP1 requires at least partial glycosylation for *in vitro* autoprocessing and proteolytic activity. We crystallized the fully glycosylated TPP1 precursor under conditions that implied partial autocatalytic cleavage between the prosegment and the catalytic domain. X-ray crystallographic analysis at 2.35 Å resolution reveals a globular structure with a subtilisin-like fold, a Ser^{475}–Glu^{272}–Asp^{360} catalytic triad, and an octahedrally coordinated Ca^{2+}-binding site that are characteristic features of the S53 sedolisin family of peptidases. In contrast to other S53 peptidases, the TPP1 structure revealed steric constraints on the P4 substrate pocket explaining its preferential cleavage of tripeptides from the unsubstituted N terminus of proteins. Two alternative conformations of the catalytic Asp^{276} are associated with the activation status of TPP1. 28 disease-causing missense mutations are analyzed in the light of the TPP1 structure providing insight into the molecular basis of late infantile neuronal ceroid lipofuscinosis.

Mutations in the *TPP1* gene (previously named *CLN2* gene) encoding tripeptidyl-peptidase I (TPP1,[3] EC 3.4.14.9) result in an autosomal recessive neurodegenerative disease of childhood called "classical late infantile neuronal ceroid lipofuscinosis" (late infantile neuronal ceroid lipofuscinosis, OMIM 204500). This enzyme was first purified from bovine pituitary glands and characterized as an acidic pepstatin-insensitive aminopeptidase (1). By exploiting its affinity to mannose 6-phosphate receptors, TPP1 was subsequently found to be missing in material from late infantile neuronal ceroid lipofuscinosis patients (2). Almost at the same time, high affinity substrates and inhibitors were characterized for TPP1 originating from rat spleen (3). Further studies demonstrated a lysosomal localization and an optimal pH of 4.0–5.0 for the tripeptidyl exopeptidase activity of TPP1. *In vitro*, TPP1 cleavage preferentially occurs between hydrophobic amino acids in the P1 and P1′ positions (according to the nomenclature of Schechter and Berger (4)) and is hampered by the presence of Pro or Lys in the P1 position or Pro in P1′ position (5). Further studies on the substrate specificity revealed that TPP1 prefers positively charged, small amino acids in positions P2 and P3 (6, 7). Additionally, TPP1 shows endopeptidase activity at a pH optimum of 3.0 (8) that mediates the autocatalytic processing of the 68-kDa TPP1 precursor to the mature 46-kDa active enzyme *in vitro* (9, 10). During processing, the primary eukaryotic 562-amino acid translation product is trimmed to the 368 C-terminal amino acids that contain the catalytic domain and all five potential *N*-glycosylation sites. The elimination of one particular glycosylation site at Asn^{286} results in complete loss of TPP1 activity, but the other glycosylation sites might also contribute to the stability of TPP1 (11, 12). At least 56 distinct disease-causing mutations in the *TPP1* gene have been reported up to now, many of which disrupt folding, processing, and trafficking of TPP1 (11, 13). The sequence homology of TPP1 to bacterial pepstatin-insensitive carboxyl peptidases suggests a similar catalytic mechanism involving a Glu, Asp, and Ser triad, with the conserved serine acting as catalytic nucleophile (14–19). We have crystallized the fully glycosylated TPP1 precursor expressed in mammalian cells at conditions permissive for autocatalytic processing, and we determined the structure by multiwavelength anomalous diffraction (MAD) analysis of the selenomethionine derivative and refined it against the 2.35 Å native data. We discuss here the structural features of TPP1 and the putative effect of mutations in the *TPP1* gene.

* This work was supported by the Fonds der Chemischen Industrie. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 3EE6) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[S] The on-line version of this article (available at http://www.jbc.org) contains Figs. S1–S3.

[1] Both authors contributed equally to this work.

[2] To whom correspondence should be addressed: Dept. of Pediatrics and Pediatric Neurology, University of Göttingen, Robert-Koch-Strasse 40, D-37075 Göttingen, Germany. Tel.: 49-551-39-8036; Fax: 49-551-39-6252; E-mail: rsteinfeld@med.uni-goettingen.de.

[3] The abbreviations used are: TPP1, tripeptidyl-peptidase I; Endo H, endoglycosidase H; LC-MS/MS, liquid chromatography-tandem mass spectrometry; MAD, multiwavelength anomalous diffraction; MALDI, matrix-assisted laser desorption ionization; TOF, time-of-flight; PNGase F, *N*-glycosidase F; PDB, Protein Data Bank.

EXPERIMENTAL PROCEDURES

Materials—Cell culture medium was purchased from Invitrogen. Oligonucleotides for the introduction of the C-terminal tag were synthesized by MWG-Biotech (Munich, Germany). Molecular weight protein markers were obtained from Amersham Biosciences. Endo H (cleaves within the core of high mannose oligosaccharides from *N*-linked glycoproteins leaving one *N*-acetyl-D-glucosamine residue bound after cleavage) and PNGase F (completely cleaves off all types of oligosaccharide chains from *N*-linked glycoproteins) were supplied by New England Biolabs (Frankfurt, Germany). All other reagents were purchased from Sigma.

Expression of TPP1—The amino acid sequence "RSHHHHHH" was introduced at the C terminus of the *TPP1* cDNA to facilitate the purification of TPP1. Transfection of the modified *TPP1* cDNA and selection of HEK 293 cells were performed as described previously (11). Both the native and selenomethionine derivative recombinant protein samples of TPP1 were purified from the cell culture supernatant. The medium was cleared by centrifugation at 3000 × *g* and 4 °C for 60 min and then filtered with a 0.2-μm pore membrane. After having added 20 mM K_2HPO_4, pH 7.5, 0.5 M NaCl, and 40 mM imidazole, the crude solution was loaded on a HisTrap HP column (GE Healthcare). Bound TPP1 was eluted with an imidazole gradient and was typically released at 70–90 mM imidazole. For crystallization the purified TPP1 fractions were pooled and concentrated using a centrifugal filter (Millipore, Schwalbach, Germany). The final solution contained 10 mg/ml TPP1 and was adapted to 20 mM KH_2PO_4, pH 4.6, and 50 mM NaCl.

Deglycosylation of TPP1—Endo H and PNGase F digestions were performed at nondenaturing conditions according to the supplier's protocol with one exception; the PNGase F treatment of autoprocessed TPP1 was done at pH 7.0 to minimize TPP1 degradation. TPP1 activity was measured according to Vines and Warburton (3). Total activity of autoprocessed TPP1 was defined as 100% and the activities of other TPP1 variants were related to it.

Mass Spectrometry—Glycosylated and deglycosylated TPP1 were incubated with trypsin (Promega GmbH, Mannheim, Germany) overnight at 37 °C and analyzed by matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS) and electrospray ionization liquid chromatography tandem-mass spectrometry (LC-MS/MS). The MALDI-MS analysis was performed on a 4800 MALDI TOF/TOF instrument (Applied Biosystems, Frankfurt, Germany), and the LC-MS/MS analysis was done on an LTQ Orbitrap XL (Thermo Scientific, Hamburg, Germany).

Crystallization—All solutions used for crystallization purposes were sterile-filtered and contained 0.03% sodium azide. Crystals were grown by mixing 2 μl of 10 mg/ml TPP1 solution with 2 μl of reservoir solution containing 7% PEG4000, 0.02 M zinc sulfate, 0.1 M sodium acetate, pH 5.0, and 0.1 M ammonium sulfate. Diffraction quality crystals appeared after 7 days. Both the native and selenomethionine derivative crystals crystallized under the same conditions.

SDS-PAGE Analysis of Crystals—Selected crystals were washed in reservoir solution before dissolving them in 5 μl of water and loading them in SDS buffer. SDS-PAGE showed the coexistence of both the cleaved mature and uncleaved immature TPP1 molecules in the crystals.

Data Collection, Structure Solution, and Refinement—The crystals (both native and selenomethionine derivative) were

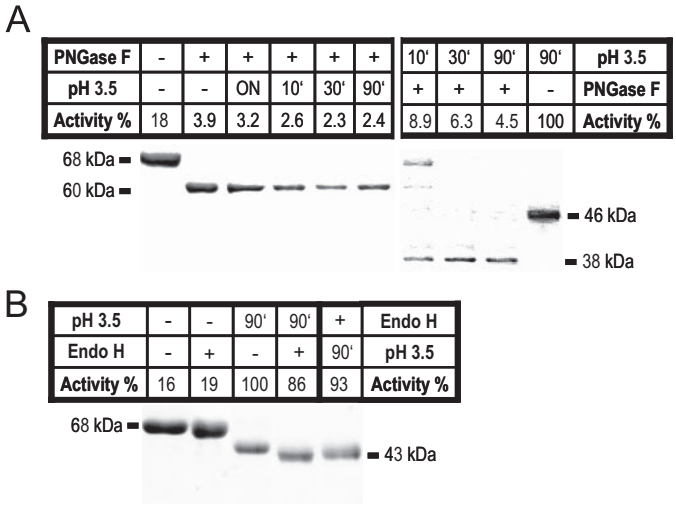

FIGURE 1. **Deglycosylation of TPP1.** *A,* purified TPP1 precursor (68 kDa) was completely deglycosylated with PNGase F for 3 h at 37 °C and then incubated at pH 3.5 overnight (*ON*) at 8 °C or for 10, 30, and 90 min at room temperature. After deglycosylation TPP1 precursor cannot autoprocess itself and shows no proteolytic activity (*left panel*). Alternatively, TPP1 precursor was first autoprocessed for 10, 30, and 90 min at room temperature before digestion with PNGase F for 3 h at 37 °C occurred. Activated TPP1 (46 kDa) becomes inactivate after deglycosylation with PNGase F (*right panel*). *B,* unprocessed or processed TPP1 was digested with Endo H for 3 h at 37 °C that had no significant effect on protease activity and autoprocessing. Shown are representative Coomassie-stained gels after SDS-PAGE. Proteolytic activity was measured with the fluorogenic substrate AAF-AMC according to Ref. 3. This experiment has been repeated at least eight times leading to similar results. Neither heat-inactivated PNGase F nor heat-inactivated Endo H affected TPP1 autoprocessing or protease activity (data not shown).

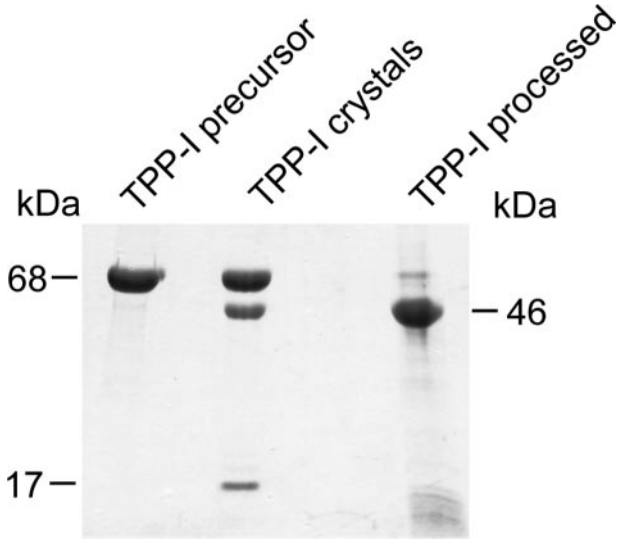

FIGURE 2. **Partial cleavage of TPP1 precursor during crystallization.** After 7 days of incubation at pH 4. 9 and room temperature, representative crystals were subjected to SDS-PAGE (*middle lane*) and compared with the unprocessed TPP1 precursor (68 kDa, *left lane*) and the autoprocessed TPP1 (46 kDa, *right lane*). About half of the crystallized TPP1 is cleaved into the catalytic domain and the prosegment (17 kDa). Hence all three proteins are present in the crystals, consistent with the crystallographic results. Autoprocessing of TPP1 was performed for 90 min at pH 3.5 and involves further degradation of the prosegment (*bottom of right lane*).

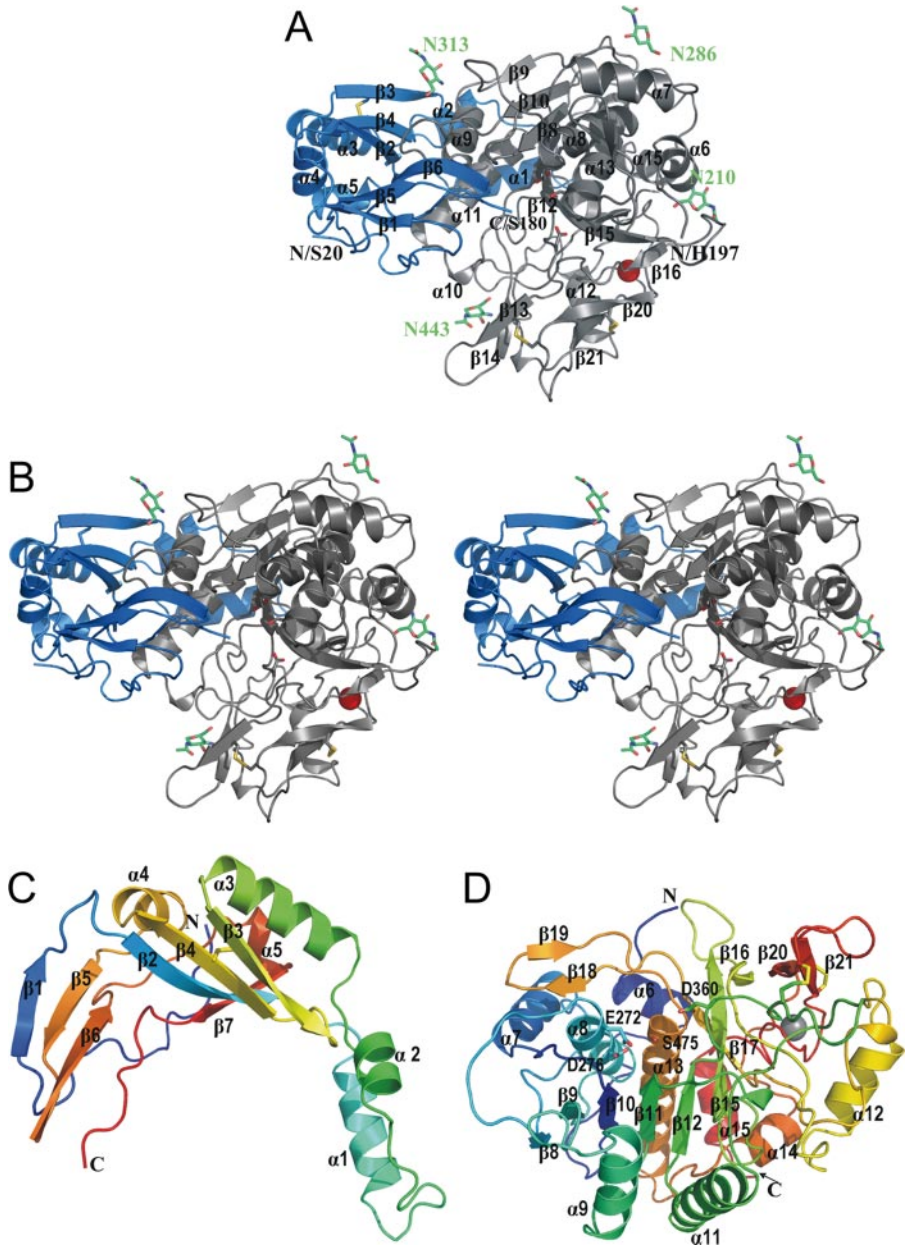

FIGURE 3. **Overall structure and fold of TPP1.** Cartoon structures of the C-α backbone of TPP1 are shown. *A,* TPP1 monomer is composed of a complex between prosegment (*blue*) and catalytic domain (*gray*). Indicated are the N terminus (Ser^{20}) as well as the ends of the prosegment (Ser^{180}) and the beginning of the catalytic domain (His^{197}). No density is observed for the linker region (Ser^{181}–Leu^{196}). The positions of the *N*-linked *N*-acetyl-D-glucosamine residues that could be seen in the density are labeled and highlighted (*green sticks*). Disulfide bonds are indicated by *yellow lines*. The Ca^{2+}-binding site is indicated by a *red sphere*. *B,* stereo view of *A*. *C,* prosegment (Ser^{20}–Ser^{180}) contains five α-helices (α1–5) and two sets of antiparallel β-strands (β1-β5-β6 and β7-β2-β4-β3). There is a disulfide bridge between Cys^{111} (located on β3) and Cys^{122} (located on β4). *D,* cartoon model of the catalytic domain rotated by 180° around the horizontal molecule axis in respect to *A*. The catalytic domain (His^{197}–Pro^{563}) contains 10 α-helices (α6–15) and 14 β-strands (β8–β21). A set of seven all-parallel β-strands (β8-β9-β10-β11-β12-β15-β17) are flanked by the helices α8, α9, α11, and α13 at four corners. There are three sets of antiparallel β-strands (β13-β14 β18-β19 and β20-β21); the β-strand β16 is parallel to β19. Two disulfide bridges are present, the first from Cys^{365} (located on β13) to Cys^{526} and the second from Cys^{522} to Cys^{537}. The Ca^{2+}-binding site is indicated by a *gray sphere*. The N terminus (His^{197}) is labeled by *N*, the C terminus (Pro^{563}) is indicated by *C*. *C* and *D* have been colored *blue-green-yellow-red* to facilitate following the protein backbone.

flash cryo-cooled using 25% glycerol and 5% MPD (in steps of 5% increasing gradient) using MiTeGen loops. A native dataset to 2.35 Å was recorded on the beamline PX2 at the Swiss Light Source on a MAR225 mosaic CCD detector. The selenomethionine derivative datasets (peak and inflection point at selenium edge) were collected with a MAR165 CCD detector on the beamline 14.2 at BESSY, Berlin, Germany. Data were processed, scaled, and analyzed using the programs HKL2000 (20), SADABS, and XPREP (Bruker AXS, Madison WI).

The peak and inflection point datasets collected at the absorption edge of selenomethionine were used for the initial phasing and model building of TPP1. The structure solution using SHELXD (21) identified eight selenium sites, consistent with the presence of a dimer in the asymmetric unit. With MAD phases calculated using these sites, the high resolution native dataset was used for phase extension with a new α test autotracing version of SHELXE (22) that was able to trace about half of the main chain. The crystal used for the MAD experiment diffracted poorly, accounting for the relatively high merging R values, but the $R_{\text{p.i.m.}}$ values of under 2% indicate that this was more than compensated for by the high redundancy of the data. This was confirmed by the very clear SHELXD solution for the selenium substructure with a correlation coefficient of 42.0% and a peak height ratio of 0.768:0.189 between the lowest correct site and the highest noise peak. Using the selenomethionine positions, bulky side chains, and biochemical properties as a guideline, the model was completed manually using COOT (23). Refmac_5.2.0019 (24) was used for structure refinement. The model was refined first isotropically to convergence and then treating each molecule in the asymmetric using as a translation, libration, and skew rigid group. The final model consists of 1050 amino acid residues, 8 *N*-acetyl-D-glucosamine substituents, 8 zinc ions, 2 calcium ions, and 2 sulfates. For the prosegment, amino acids Ser^{20} through Ser^{180} were observed, and for the catalytic domain residues, His^{197} through Pro^{563} and the first amino acid "Arg" of the purification tag was modeled.

Computer Graphics—The graphical representation of the TPP1 structure was performed using the computer program PyMOL 1_1.

TABLE 1
Data collection statistics
$R_{int} = 100\Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$ and $R_{p.i.m.} = 100\Sigma_{hkl}((1/(N-1))^{1/2}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|(/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where N is the number of equivalent reflections in each group of equivalents.

Data statistics	Native	Selenium peak	Selenium inflection
Wavelength	1.00000 Å	0.97971 Å	0.97987 Å
Beamline	SLS-PX2	BESSY-14.2	BESSY-14.2
Detector	MAR225 CCD	MAR165 CCD	MAR165 CCD
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
a	100.5 Å	100.4 Å	100.4 Å
b	113.5 Å	113.3 Å	113.1 Å
c	128.9 Å	128.4 Å	128.3 Å
Resolution	2.35 Å (2.45-2.35 Å)	2.70 Å (2.80-2.70 Å)	2.70 Å (2.80-2.70 Å)
Reflections measured	447,600	1,339,940	1,343,398
Unique reflections	60,714	40,695	40,750
Redundancy	7.26 (6.11)	32.70 (20.60)	32.75 (20.67)
Completeness (%)	98.4 (94.8)	99.3 (96.5)	99.3 (96.6)
Mean I/σ (I)	17.20 (3.15)	31.02 (4.62)	34.31 (4.44)
R_{int} (%)	5.55 (57.20)	11.30 (60.85)	11.02 (62.51)
$R_{p.i.m.}$ (%)	2.20 (22.78)	1.96 (13.35)	1.91 (13.70)

RESULTS

Glycosylation of TPP1 Is Essential for in Vitro Autoprocessing and Protease Activity—Previous studies have pointed to the importance of *N*-glycosylation for intracellular folding and trafficking of TPP1 (12, 25). In particular, replacement of Asn^{286} by Ser or Gln prevents intracellular processing of TPP1. To elucidate the importance of *N*-glycosylation for *in vitro* protease activity, we studied autoprocessing of purified TPP1 after complete deglycosylation and measured *in vitro* activity of autoprocessed TPP1 after deglycosylation. PNGase F-treated TPP1 precursor was unable to autoprocess at pH 3.5 and remained inactive, even after 24 h of incubation. Autoprocessed TPP1 almost completely lost its peptidase activity after deglycosylation with PNGase F (Fig. 1*A*). In contrast, partial deglycosylation by Endo H digestion did not prevent TPP1 autoprocessing nor did it significantly affect TPP1 protease activity (Fig. 1*B*). Control experiments with heat-inactivated PNGase F or Endo H ruled out that factors other than glycosidase activities were responsible for these effects (data not shown). These results indicate that TPP1 requires some *N*-linked glycosylation for cell-free peptidase activity. Tryptic digests of glycosylated and deglycosylated TPP1 were analyzed by MALDI-MS and electrospray ionization-LC-MS/MS for additional modifications that could explain the loss of TPP1 activity. No abnormalities were detected.

Partial Autoprocessing of TPP1 Precursor during Crystallization—Human TPP1 was engineered with a C-terminal His_6 tag, overexpressed in HEK 293 cells, and purified by metal ion affinity chromatography. Fully glycosylated TPP1 precursor was incubated at crystallization conditions that allowed slow autoprocessing. Crystals appeared after 7 days of incubation at pH 4.9 and room temperature. SDS-PAGE analysis of representative crystals showed the presence of TPP1 precursor as well as autoprocessed TPP1 along with the TPP1 prosegment (Fig. 2). This result points to the coexistence of unprocessed and processed TPP1 within the crystals and is consistent with experiments showing that TPP1 autoactivation is relatively slow at pH 4.5–5.5 and does not involve degradation of the TPP1 prosegment (10).

Overall Structure and Fold of TPP1—TPP1 is globular and has an overall fold similar to that of subtilisin and to members of the S53 family of sedolisin peptidases such as sedolisin and kumamolisin. The asymmetric unit contains two TPP1 monomers. Each monomer is a complex of two chains, the prosegment and the catalytic domain (Fig. 3*A*). The data collection and refinement statistics are displayed in Tables 1 and 2.

TABLE 2
Refinement statistics
The Ramachandran statistics were compiled using the program COOT (23).

Refinement statistics	TPP1
R-factor	22.1%
Free *R*-factor	26.2%
Root mean square deviations from ideal geometry	
Bond lengths	0.008 Å
Bond angles	1.250°
No. of protein residues/atoms	1050/8094
No. of sugar molecules/atoms	8/112
No. of calcium ions	2
No. of zinc ions	8
No. of chloride ions	2
No. of water molecules	71
Ramachandran plot (%)	
Residues in preferred regions	91.6
Residues in allowed regions	6.9
Residues in disallowed regions	1.5

The pro-segment spans residues Ser^{20} through Ser^{180} and has a calculated molecular mass of 17 kDa. It is composed of five α-helices formed by α1, α2, α3, α4, and α5 and two sets of antiparallel β-strands, formed by β1-β5-β6 and β7-β2-β4-β3, and stabilized by a disulfide bridge between Cys^{111} and Cys^{122} on β3 and β4, respectively (Fig. 3*B*).

The catalytic domain contains residues His^{197} through Pro^{563} and has a calculated molecular mass of 39.7 kDa. It contains 10 α-helices and 14 β-strands. The core of the TPP1 molecule includes seven all-parallel β-strands β8-β9-β10-β11-β12-β15-β17 and is sandwiched and prominently flanked by four helices α8, α9, α11, and α13 near the active site. There are two disulfide linkages, the first between Cys^{365} and Cys^{526} and the second between Cys^{522} and Cys^{537}, stabilizing the antiparallel α-strands β20 and β21 (Fig. 3, *C* and *D*). Two proline residues, Pro^{378} and Pro^{449}, are found in *cis* conformation. The schematic topology of the TPP1 structure is given in Fig. 4.

Each molecule has an octahedrally coordinated Ca^{2+}-binding site, coordinated by carboxylate oxygens of Asp^{517}

and Asp^{543}, the main chain carbonyls of Val^{518}, Gly^{539}, and Gly^{541}, and a water molecule (Fig. 5*A*). Also, there are four Zn^{2+}-binding sites in each of the TPP1 units that are quite likely to be nonphysiological and a result of the crystallization procedure (Fig. 5*B*). As expected, glycosylations were observed at residues Asn^{210}, Asn^{286}, Asn^{313}, and Asn^{443} and were modeled as *N*-acetyl-D-glucosamine. No electron density is observed for the glycosylation site at residue Asn^{222} in both monomers.

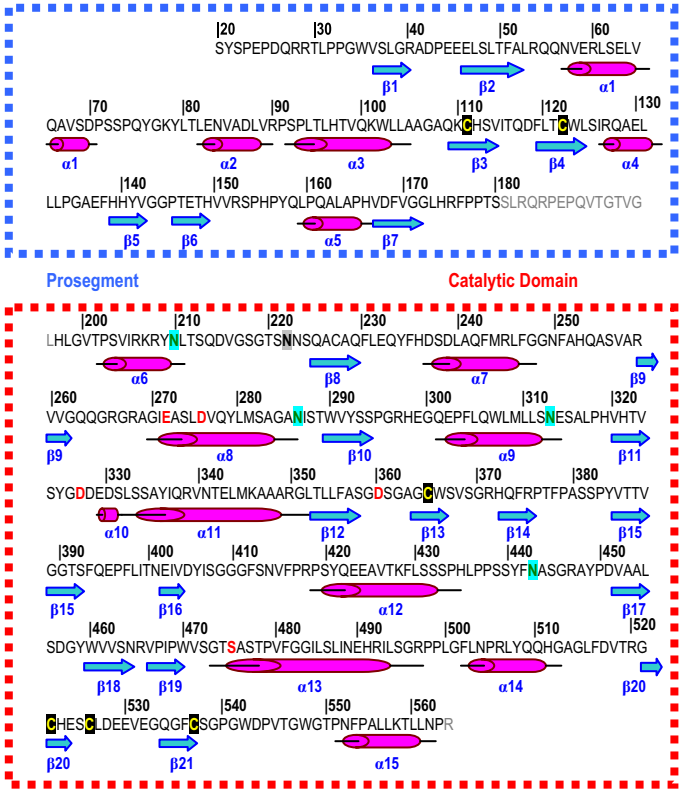

FIGURE 4. **Topology diagram of the TPP1 structure showing the prosegment (Ser^{20}–Ser^{180}) and the catalytic domain (His^{197}–Pro^{563}).** α-Helices are labeled α and are represented by *cylinders*, and β-strands are labeled β and represented by *arrows*. Residues that are involved in the catalytic mechanism are marked by *red letters*; the 6 cysteine residues that form disulfide bridges are colored *yellow* on *black background*, and the asparagine residues that carry *N*-linked oligosaccharides are highlighted by a *turquoise background*. No density could be observed for a possible glycosylation at Asn^{222} (highlighted with *gray background*). The linker region (Ser^{181}–Leu^{196}) was poorly defined (*gray letters*). The first amino acid *R* of the purification tag RSHHHHHH is displayed in *gray*.

The residues Ser^{181} to Leu^{196} could not be modeled as no density is observed for these residues (Fig. 3*A*). This might be explained by the variation of the cleavage sites at pH values 4.5–5.5 that result in a slow TPP1 processing and variable cleavage after residues Ser^{181}, Glu^{189}, and Gly^{195} (10). However, no other structural ambiguity was found in any of the regions near the active site, indicating that autoprocessing is not associated with any major change in the secondary structure of TPP1. Attempts to refine group occupancies of the prosegment and the catalytic domain indicate that both are probably fully occupied, suggesting that the prosegment is homogeneously located within the crystals no matter whether the TPP1 precursor is cleaved or not. These data are in agreement with the lack of differences in the CD spectra measured after TPP1 autoprocessing (10) and consistent with recent studies suggesting that the TPP1 prosegment remains tightly bound to the catalytic domain at pH 4.5 even after cleavage (26). Multiple molecular contacts between the prosegment and the catalytic domain explain this strong binding that covers ~21 and 15% of the total solvent accessible surface areas of the prosegment and the catalytic domain (if glycosylation is not taken into account), respectively (supplemental Fig. S1). The catalytic domain and the prosegment contribute 56 and 65 residues, respectively, to form the interface between the two domains. Salt bridges are formed predominantly between the guanidinium group of Arg^{339} and the carboxylates of Asp^{118}, the ϵ-amino group of Lys^{346} and carboxylates of Asp^{118}, carboxylates of Glu^{302}, and the guanidinium group of Arg^{175}. In addition, there are several hydrogen bonds stabilizing the complex between the prosegment and the catalytic domain (supplemental Table S1). The strongest interface interactions involve polar amino acid side chains and hence are sensitive to pH changes. Hence, lowering the pH will release the catalytic domain from the trap formed by the prosegment and will allow autoproteolytic cleavage of the linker.

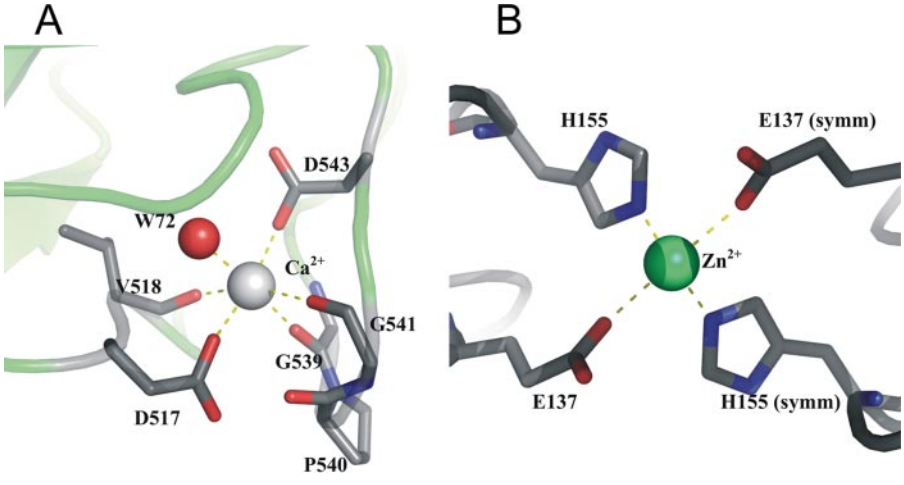

FIGURE 5. *A*, Ca^{2+}-binding site. One Ca^{2+} ion per TPP1 monomer is coordinated by six ligands, including the two carboxylate oxygen groups of Asp^{517} and Asp^{543}, the main chain carbonyl groups of Val^{518}, Gly^{539}, and Gly^{541}, and a water molecule (*W72*). The side chains of the binding residues are drawn on the cartoon model of the α-carbon backbone of TPP1 (*green*). *B*, Zn^{2+}-binding sites. The 8 Zn^{2+} ions, identified by their density in the anomalous Fourier synthesis, were almost certainly introduced by crystallization. Four of the Zn^{2+} sites, including the one shown in this figure, lie on 2-fold axes and introduce interactions between symmetry related molecules that may well be essential for crystallization.

Homology of TPP1 to Members of the Sedolisin Family of Serine Proteases—Although the amino acid sequence identity of TPP1 relative to other members of the S53 family of peptidases is rather low, there is a considerable structural homology. The TPP1 precursor shares only 18.7% of its amino acid sequence with pro-kumamolisin that is currently the only reported structure of a full-length S53 peptidase. However, superposition of the reported structure of pro-kuma-

molisin (17) with this full-length TPP1 structure reveals a 1.64 Å root mean square deviation of the 376 matching C-α atoms (supplemental Fig. S2). The TPP1 and kumamolisin prosegments (TPP1 residues 20–180) have a 1.67Å root mean square deviation of the 125 matching C-α atoms, and the two catalytic domains (TPP1 residues 197–563) have a 1.47 Å root mean square deviation of the 270 matching C-α atoms. This indicates that the structural homology between the TPP1 precursor and pro-kumamolisin is high and that the catalytic domains of the two proteases match even better than their prosegments. The latter finding is underlined by the fact that the catalytic domains are longer and yet show a lower deviation.

The crystal structure of TPP1 reveals a subtilisin like fold and contains a Ser^{475}–Glu^{272}–Asp^{360} catalytic triad similar to kumamolisin and sedolisin (15) (Fig. 6*A*). Ser^{475}, the nucleophile, resides on helix α13, and Glu^{272}, the general base catalyst along with Asp^{276}, lies on helix α8. Asp^{360} lies on a loop connecting the strands β12 and β13, whereas Asp^{327} is observed on the loop connecting strand β11 to helix α10 (Fig. 3*D*). Superposition of the catalytic center of TPP1 with that of kumamolisin, sedolisin, and subtilisin discloses the high degree of structural homology among them (Fig. 6, *A* and *B*). The overall structure of the catalytic domains of TPP1, kumamolisin, and sedolisin resemble with each other, and even the subtilisin fold contains similar regions (supplemental Fig. S3, *A*–*C*). Although the TPP1 prosegment bears good similarity to the kumamolisin prosegment, both differ significantly from the corresponding regions of subtilisin (supplemental Fig. 3*D*). The calcium ion is observed at a position homologous to that of other reported sedolisin structures (Fig. 3, *C* and *D*, and Fig. 5*A*).

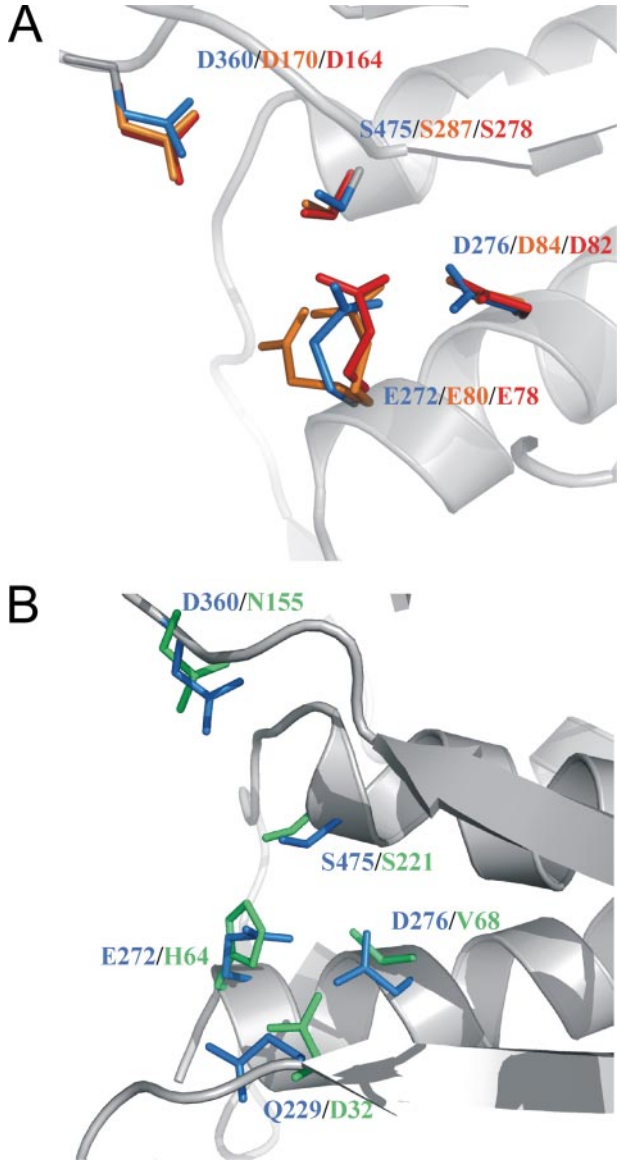

FIGURE 6. *A*, superposition of the catalytic residues of TPP1, kumamolisin (PDB ID 1KDV), and sedolisin (PDB ID 1SI0). The *stick diagram* displays the corresponding side chains of the catalytic residues of TPP1 (*blue*), kumamolisin (*red*), and sedolisin (*orange*) positioned on the cartoon model of the α-carbon backbone of TPP1 (*gray*). Note that Glu^{80} is displayed in two alternative conformations reported for sedolisin. The high structural similarity of the catalytic center of TPP1 with that of kumamolisin and sedolisin includes the type of amino acids as well as their relative positions to each other. *B*, superposition of the catalytic residues of TPP1 and subtilisin (PDB ID 1CSE). The *stick diagram* displays the corresponding side chains of the catalytic residues of TPP1 (*blue*) and subtilisin (*green*) positioned on the cartoon model of the α-carbon backbone of TPP1. The overall structure of TPP1 and subtilisin are similar. However, a number of structural details vary, and both proteases possess different catalytic triads. TPP1, Ser^{475}–Glu^{272}–Asp^{360} *versus* subtilisin Ser^{221}–His^{64}–Asp^{32}. Additionally, the positions of the catalytic residues to each other are not conserved.

DISCUSSION

We have demonstrated the importance of *N*-glycosylation for *in vitro* autoprocessing and protease activity of TPP1. These results extend previous studies on the intracellular processing of TPP1 that showed the effect of mutagenesis of one or several *N*-glycosylation sites on intracellular folding and trafficking of TPP1 (12). Our experimental data reveal that *in vitro* unprocessed or processed TPP1 loses its protease activity after complete deglycosylation by PNGase F treatment. This loss of activity cannot be explained by instability, because deglycosylated TPP1 precursor showed no evidence for accelerated degradation (Fig. 1*A*). In addition, previously autoactivated and thus presumably properly folded TPP1 protease becomes inactive after complete deglycosylation with PNGase F. Interestingly, partial deglycosylation by Endo H treatment is not associated with a disturbed autoprocessing and loss of protease activity. This suggests that either the complex *N*-linked oligosaccharides that are not cleaved by Endo H or the terminal *N*-acetyl-D-glucosamine residues that remain after Endo H cleavage are critical for autoprocessing and protease activity of TPP1. This finding is supported by the fact that Endo H-deglycosylated TPP1 precursor is indistinguishable from fully glycosylated TPP1 in terms of autocatalytic processing of the precursor and enzymatic properties of the mature protease (accompanying paper by Guhaniyogi *et al.* (28)). It can be speculated that complete deglycosylation might lead to a significant conformational alteration or that *N*-linked oligosaccharides are involved in the catalytic process that is mediated by TPP1. Further studies are required to determine the precise oligosaccharide components and molecular interactions that are essential for TPP1 activity.

We have crystallized fully glycosylated TPP1 at conditions that allow slow autoprocessing and obtained an ~50:50% mixture of unprocessed and processed TPP1 within the crys-

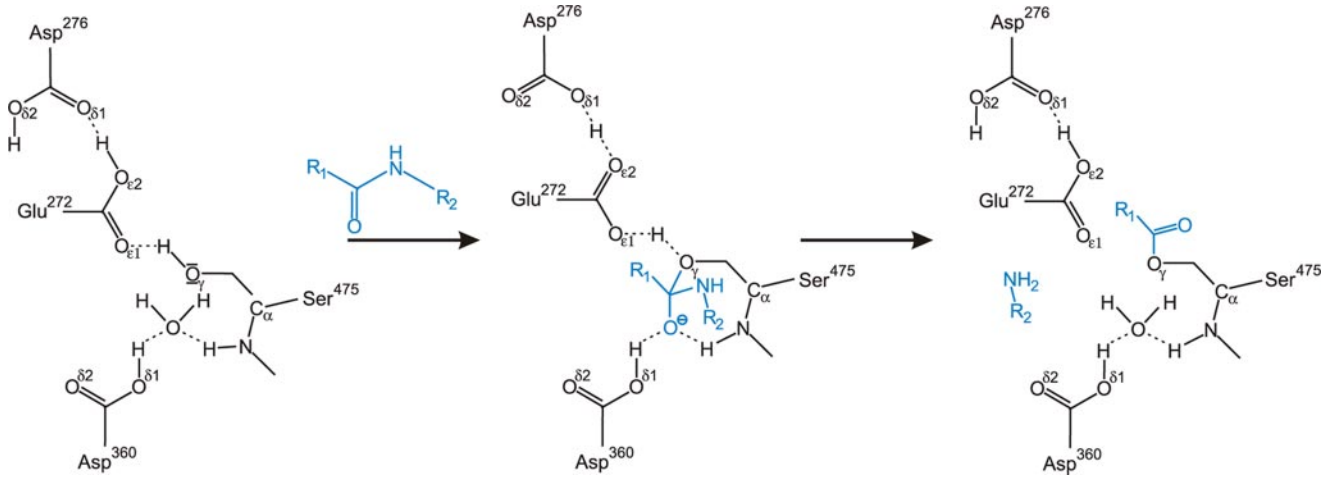

FIGURE 7. **Suggested proton transfer during the acylation step mediated by the catalytic residues of TPP1.** The peptide substrate (colored *blue*) is inserted into the active site of TPP1 such that the carbonyl carbon of the scissile bond is positioned near the nucleophilic Ser^{475}. The Ser^{475} hydroxyl group transfers the O-γ proton to the carboxylic oxygens of Glu^{272}. This enables the nucleophilic Ser^{475} O-γ to attack the polarized carbonyl group of the scissile peptide bond to form a tetrahedral intermediate and to pass through the transition state, with simultaneous proton transfer from the acidic Ser^{475} oxonium cation via the Glu^{272} carboxylic group to Asp^{276}. The electron pair from the double bond of the carbonyl oxygen moves to the oxygen to form an oxyanion and the Asp^{360} side chain helps to create the oxyanion hole, stabilizing the tetrahedral intermediate of the reaction. With the cleavage of the peptide bond the electron pair moves back from the oxyanion to recreate the carbonyl oxygen double bond, generating the acyl-TPP1 intermediate. Not shown is the hydrolysis reaction that ultimately leads to regeneration of the Ser^{475} and formation of the peptide C terminus. Note that the three letter code for amino acids has been chosen to facilitate distinction from the single letter symbols for chemical elements.

tals. Nevertheless, the resulting crystal structure did not suffer from any major structural ambiguities, and only the linker region carrying the potential cleavage sites for autoprocessing was poorly defined. As pointed out by Guhaniyogi *et al.* (28), there is a good agreement between the structures of the glycosylated TPP1 precursor and the Endo H-deglycosylated one (PDB ID 3EDY). Furthermore, the crystal structure of pro-kumamolisin, the most homologous bacterial member of the sedolisin family, showed that the connecting linker runs through the entire active site cleft (17). Superposition of pro-kumamolisin with the active kumamolisin revealed that the catalytic domain of the proenzyme exhibits an almost identical structure and is already properly prefolded within the proenzyme (18). Considering the high degree of structural homology between kumamolisin and TPP1, it can be assumed that also the TPP1 linker runs through the active site cleft of TPP1 and hence is partially cleaved and therefore disordered within our crystals. However, the cleaved off prosegment remains bound to the catalytic domain at pH 4.5 (or higher pH), as has been demonstrated recently (26). In conclusion, our data suggest that uncleaved and cleaved off TPP1 prosegments show virtually identical structures and positioning and that the catalytic domains of unprocessed and processed TPP1 possess very similar structures as well. The lack of significant structural differences between unprocessed and processed TPP1 is supported by the absence of alterations in CD spectroscopy observed after TPP1 processing (10). However, we find densities for two conformations of Asp^{276}, one position superposing with the equivalent residues in kumamolisin (Asp^{82}) and sedolisin (Asp^{84}) and the other superposing with the Endo H-deglycosylated TPP1 precursor (PDB 3EDY) that is not autoactivated (Fig. 8). Although we did not model alternative conformations in view of the modest resolution, it is likely that these two conformations reflect the active and inactive state of the catalytic center. As described below, Asp^{276} is involved in the catalytic process that requires the proper positioning in the catalytic center.

The high degree of structural homology between TPP1 and other members of the sedolisin family also implies that the catalytic mechanism that mediates the proteolytic cleavage is very similar. The active site of TPP1 includes the three amino acids Ser^{475}, Glu^{272}, and Asp^{360} from which the name of the SEDolisin family of peptidases is derived. During the catalytic process, the Ser^{475} hydroxyl group loses the O-γ proton and so becomes a stronger nucleophile. The only plausible proton acceptors are the carboxylic oxygens of Glu^{272}. In the approaching polypeptide substrate, the scissile peptide bond is oriented toward Ser^{475} O-γ with its carbonyl group inserting into the oxyanion hole, fixed by subsite interactions of both flanking peptide moieties (Fig. 7). This would allow the nucleophilic Ser^{475} O-γ to attack the polarized carbonyl group of the scissile peptide bond to form a tetrahedral intermediate and to pass through the transition state with simultaneous proton transfer from the acidic Ser^{475} oxonium cation via the Glu^{272} carboxylic group to Asp^{276}. The Asp^{360} side chain helps to create the oxyanion hole, stabilizing the tetrahedral intermediate of the reaction. Apart from the nucleophilic Ser^{475}, all other residues participating in the catalytic mechanism are acidic amino acids. To mediate the catalytic electron transfer in both directions, these residues have to exist at equilibrium between the protonated and deprotonated states. Thus the surrounding pH should be close to their pK_a values, which explains the requirement for an acidic environment for catalysis.

The fundamental difference between TPP1 and other members of the sedolisin family of peptidases is the orientation of cleavage. Whereas all known prokaryotic members of the sedolisin family are putative carboxyl peptidases, TPP1 preferentially cleaves tripeptides from the N-terminal end of proteins.

This preference can be explained by a unique hydrogen bond between Tyr^{325} and Asp^{327} (distance 3.24 Å) that can sterically constrain the P4 residue of the substrate (Fig. 8). Tyr^{325} interacts with Phe^{304} by π-stacking and is localized within a hydrophobic pocket with several stacked aromatic systems. All these interactions occlude the putative S4 subsite and thus explain the preference for substrates with P1-P3 residues and a free unsubstituted N terminus at P3. Crystallization of complexes between TPP1 and substrates/inhibitors might prove this hypothesis.

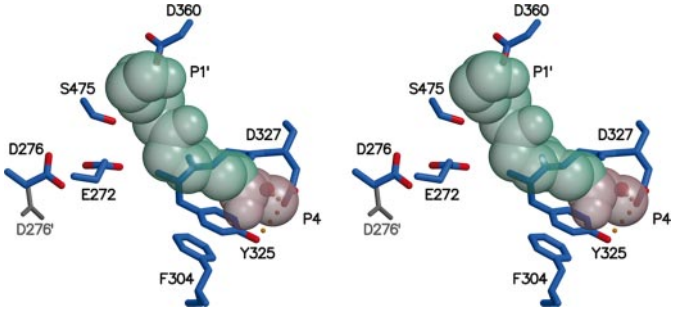

FIGURE 8. **Stereo plot of the catalytic residues with a modeled pentapeptide substrate.** The side chains (*stick model*) of the residues involved in the catalytic process are represented together with a pentapeptide substrate (cpk model in *green* containing P1′-P1-P2-P3-P4, P4 is highlighted in *pink*). The position of the pentapeptide is derived from the eglin c backbone in complex with subtilisin Carlsberg (27). The hydrogen bonding between Asp^{327} and Tyr^{325} (distance 3.24 Å) constrains the space occupied by the P4 substrate residue resulting in a clash. For the Asp^{276} residue an additional secondary conformation is indicated by $Asp^{276\prime}$. Interestingly, this conformation is the only one present in the structure of Endo H-deglycosylated TPP1 (PDB ID 3EDY), which does not autoprocess itself. In the $Asp^{276\prime}$ position, this residue cannot participate in the proton transfer of the catalytic mechanism and hence most likely reflects the inactive conformation. Ser^{475} = S475, Glu^{272} = *E*272, Asp^{276} = D276, Asp^{360} = D360.

More than 50% of the known *TPP1* mutations represent single amino acid substitutions that result in the biosynthesis of TPP1 mutants. None of these missense mutants shows any measurable residual TPP1 activity, and all are associated with severe symptoms such as developmental regression, epileptic seizures, ataxia, and progressive loss of motor abilities, speech and vision. On the basis of the resolved TPP1 structure, we have described the molecular effects of all currently known missense mutations (Table 3 and Fig. 9). The majority of these mutations would disrupt folding leading to instability and rapid degradation. The amino acid substitutions V277M, Q248P, G284V, G473R, and S475L probably compromise the active center and result in loss of proteolytic activity. Only two mutations, G77R and S153P, are located within the prosegment and are likely to disturb processing of TPP1. The N286S substitution results in loss of one glycosylation site, which leads to almost complete loss of protease activity without obvious alterations. Also, no apparent conformational destabilization is observed for the missense mutations V216M, R266Q, and K428N. Further biochemical and structural analyses are underway to elucidate the mech-

TABLE 3

Molecular effect of single amino acid substitutions

Reported amino acid substitutions are listed from the N terminus to the C terminus of full-length TPP1. The positions of the affected residues are described in respect to the TPP1 structure; α corresponds to α-helix and β to β-strand. Most of the amino acid substitutions lead to destabilization of folding and compromise the proteolytic activity. Two alleles, R127Q and H175R, which are associated with an almost normal TPP1 activity, have not been included in the table.

Mutation	Structural position	Putative functional consequence
G77R	Position in the last turn of α1, side chain is oriented outwards	Disturbed binding/interaction between mutant prosegment and catalytic domain (26)
S153P	Position at the end of β6, hydrogen bond to Glu^{173}	Disruption of β-strand conformation by proline, loss of hydrogen bond to Glu^{173} might dislocate β6 and carbon backbone
P202L	Position at the start of α5	Unclear
R206C	Position within α6, interacts with Thr^{212}, Ser^{282}, and Asp^{215}	Destabilization of α6 might cause folding defect
V216M	Position at the protein surface	Unclear
R266Q	Position at the protein surface	Unclear
V277M	Position within α8, that also harbors Glu^{272} and Asp^{276}	Clash with Val^{291} and/or Cys^{227}, destabilization of α8 might affect active site
Q278P	Position within α8, which also harbors Glu^{272} and Asp^{276}	Kink within α8 might affect active site
G284V	Position at the end of α8, which also harbors Glu^{272} and Asp^{276}	Clash with Gly^{483} of adjacent α13, disruption of fold
N286S	Position at the protein surface between α8 and β10	Loss of *N*-linked glycosylation
I287N	Inward oriented position between α8 and β10	Hydrogen-bonding of Asn^{287} reduces flexibility of pocket
R339Q	Position within α11, hydrogen bond to Asp^{118}, Leu^{53}, Asp^{168}, and Glu^{343}	Disruption of important interaction with E343, Asp^{168}, Leu^{53}, Asp^{118}, disturbed interaction with prosegment might affect folding
E343K	Position within α11, hydrogen bond to Gln^{55} and Arg^{339}	Clash with Leu^{305} and Phe^{119}, disturbed folding and interaction with prosegment
T353P	Position within β12	Disruption of β12, folding defect
L355P	Position within β12, part of hydrophobic pocket	Disruption of β12, folding defect
C365R/C365Y	Disulfide-bond with Cys^{526}	Loss of disulfide-bond, folding defect
V385D	Position within β15, part of hydrophobic pocket	Disruption of hydrophobic pocket and β15
G389E	Position within β15	Clash with Pro^{450}, Val^{545}, Ser^{585}, folding defect
K392N	Position at the protein surface	Possible interaction with autoinhibitory loop
Q422H	Position within α12	Disturbed interaction with prosegment
K428N	Position at the protein surface within α12	Unclear
R447H	Hydrogen bonds to Val^{385}, Asp^{451}, Ser^{382}, Ser^{381}	Loss of hydrogen bonds destabilizes fold
A454E	Position at the end of β17 in hydrophobic pocket	Disruption of hydrophobic pocket destabilizes fold, clash with adjacent residues Tyr^{209} and Phe^{481}
G473R	Position at the beginning of α13	Possible hydrogen bonding with Asp^{360} of the active center, disturbed catalytic activity
S475L	Position within α13, part of the active center	Loss of catalytic activity
V480G	Position within α13 pointing into hydrophobic pocket	Destabilization of α13 and fold
F481C	Position within α13 pointing into hydrophobic pocket (as Val^{452} and Pro^{551})	Destabilization of α13 and fold
P544S	Part of a tight β-turn	Destabilization of β-turn, folding defect

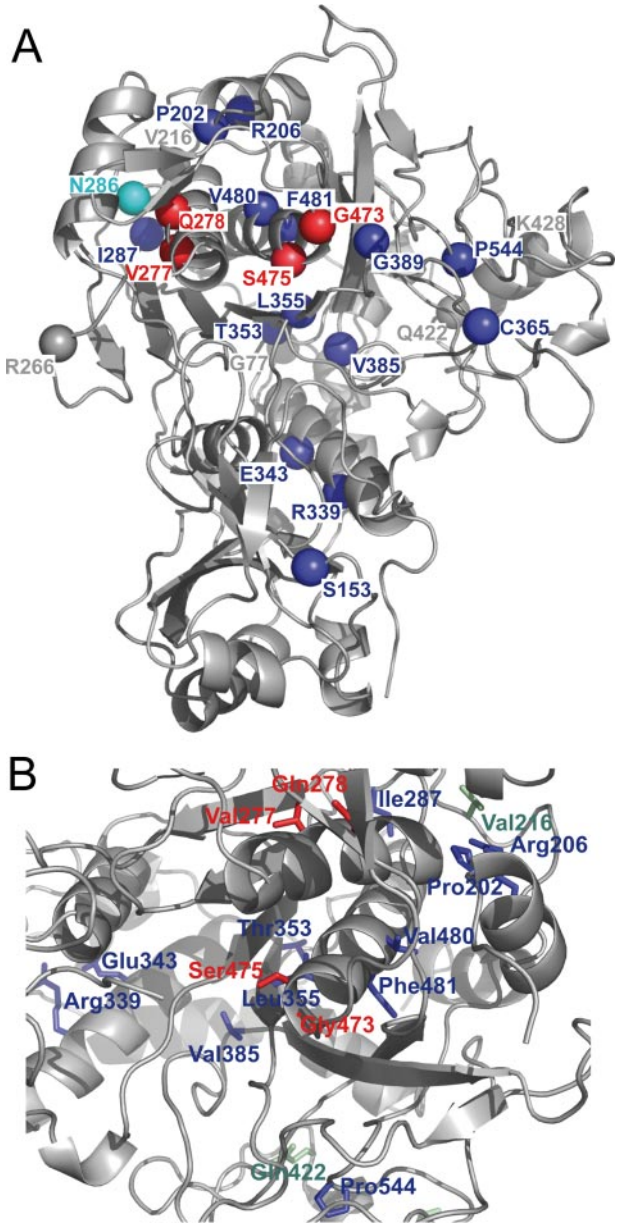

FIGURE 9. **Molecular basis of late infantile neuronal ceroid lipofuscinosis.** *A,* view from the top of the cartoon model in Fig. 3*D* and rotated. The currently known pathogenic *TPP1* missense mutations are mapped onto the TPP1 structure as labeled *spheres* and are colored according to their impact on the TPP1 structure. *Red colored spheres* point to mutations compromising the catalytic activity; *blue colored spheres* indicate conformational destabilization, and *gray colored spheres* designate unclear structural consequences. The Asn^{286} residue (*turquoise*) corresponds to one of the five *N*-glycosylation sites. *B,* view of the catalytic cleft of the cartoon model in Fig. 3*D*. *TPP1* missense mutations are shown as *stick models*, coloring scheme corresponds to *A,* except for mutations with unclear effect that are colored in *green*.

anisms that explain the functional loss of TPP1. The thorough biochemical and structural analyses of these mechanisms are fundamental for the design of small molecule compounds that are meant to rescue TPP1 activity and hence to prevent or protract disease progression. The resolution of the TPP1 structure has not only extended our understanding of the molecular mechanisms underlying late infantile neuronal ceroid lipofuscinosis but could also lead to novel therapeutic options for this devastating disorder.

Acknowledgments—We thank the Lobel and Stock laboratories for sharing their manuscript and coordinates of the Endo H-deglycosylated TPP1 structure prior to publication. We are grateful to Dr. Ehmke Pohl (SLS), Roland Pfoh (Göttingen), and Dr. Uwe Müller (Bessy) for help with the synchrotron data collection. We acknowledge the technical assistance of Tanja Wilke, Nicole Holstein, and Sven Hagen.

REFERENCES

1. Doebber, T. W., Divor, A. R., and Ellis, S. (1978) *Endocrinology* **103,** 1794–1804
2. Sleat, D. E., Donnelly, R. J., Lackland, H., Liu, C. G., Sohar, I., Pullarkat, R. K., and Lobel, P. (1997) *Science* **277,** 1802–1805
3. Vines, D., and Warburton, M. J. (1998) *Biochim. Biophys. Acta* **1384,** 233–242
4. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27,** 157–162
5. Warburton, M. J., and Bernardini, F. (2001) *FEBS Lett.* **500,** 145–148
6. Oyama, H., Fujisawa, T., Suzuki, T., Dunn, B. M., Wlodawer, A., and Oda, K. (2005) *J. Biochem.* (*Tokyo*) **138,** 127–134
7. Tian, Y., Sohar, I., Taylor, J. W., and Lobel, P. (2006) *J. Biol. Chem.* **281,** 6559–6572
8. Ezaki, J., Takeda-Ezaki, M., Oda, K., and Kominami, E. (2000) *Biochem. Biophys. Res. Commun.* **268,** 904–908
9. Lin, L., Sohar, I., Lackland, H., and Lobel, P. (2001) *J. Biol. Chem.* **276,** 2249–2255
10. Golabek, A. A., Wujek, P., Walus, M., Bieler, S., Soto, C., Wisniewski, K. E., and Kida, E. (2004) *J. Biol. Chem.* **279,** 31058–31067
11. Steinfeld, R., Steinke, H. B., Isbrandt, D., Kohlschutter, A., and Gartner, J. (2004) *Hum. Mol. Genet.* **13,** 2483–2491
12. Wujek, P., Kida, E., Walus, M., Wisniewski, K. E., and Golabek, A. A. (2004) *J. Biol. Chem.* **279,** 12827–12839
13. Sleat, D. E., Gin, R. M., Sohar, I., Wisniewski, K., Sklower-Brooks, S., Pullarkat, R. K., Palmer, D. N., Lerner, T. J., Boustany, R. M., Uldall, P., Siakotos, A. N., Donnelly, R. J., and Lobel, P. (1999) *Am. J. Hum. Genet.* **64,** 1511–1523
14. Rawlings, N. D., and Barrett, A. J. (1999) *Biochim. Biophys. Acta* **1429,** 496–500
15. Wlodawer, A., Durell, S. R., Li, M., Oyama, H., Oda, K., and Dunn, B. M. (2003) *BMC Struct. Biol.* **3,** 8
16. Wlodawer, A., Li, M., Gustchina, A., Tsuruoka, N., Ashida, M., Minakata, H., Oyama, H., Oda, K., Nishino, T., and Nakayama, T. (2004) *J. Biol. Chem.* **279,** 21500–21510
17. Comellas-Bigler, M., Maskos, K., Huber, R., Oyama, H., Oda, K., and Bode, W. (2004) *Structure* (*Lond.*) **12,** 1313–1323
18. Comellas-Bigler, M., Fuentes-Prior, P., Maskos, K., Huber, R., Oyama, H., Uchida, K., Dunn, B. M., Oda, K., and Bode, W. (2002) *Structure* (*Lond.*) **10,** 865–876
19. Wlodawer, A., Li, M., Dauter, Z., Gustchina, A., Uchida, K., Oyama, H., Dunn, B. M., and Oda, K. (2001) *Nat. Struct. Biol.* **8,** 442–446
20. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
21. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58,** 1772–1779
22. Sheldrick, G. M. (2008) *Acta Crystallogr. Sect. A* **64,** 112–122
23. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132
24. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–255
25. Tsiakas, K., Steinfeld, R., Storch, S., Ezaki, J., Lukacs, Z., Kominami, E., Kohlschutter, A., Ullrich, K., and Braulke, T. (2004) *Glycobiology* **14,** C1–C5
26. Golabek, A. A., Dolzhanskaya, N., Walus, M., Wisniewski, K. E., and Kida, E. (2008) *J. Biol. Chem.* **283,** 16497–16504
27. Bode, W., Papamokos, E., and Musil, D. (1987) *Eur. J. Biochem.* **166,** 673–692
28. Guhaniyogi, J., Sohar, I., Das, K., Stock, A. M., and Lobel, P. (2009) *J. Biol. Chem.* **284,** in press

Protein Structure and Folding:

Structure of Tripeptidyl-peptidase I Provides Insight into the Molecular Basis of Late Infantile Neuronal Ceroid Lipofuscinosis

Aritra Pal, Ralph Kraetzner, Tim Gruene, Marcel Grapp, Kathrin Schreiber, Mads Grønborg, Henning Urlaub, Stefan Becker, Abdul R. Asif, Jutta Gärtner, George M. Sheldrick and Robert Steinfeld

J. Biol. Chem. 2009, 284:3976-3984.
doi: 10.1074/jbc.M806947200 originally published online November 26, 2008
