SUPPLEMENTARY MATERIAL

Figure S1. SDS-PAGE analysis of proTPP1 crystals. SDS-PAGE analysis of proTPP1 crystals was performed to determine the maturation state of the zymogen. Lane 1 – Endo H-deglycosylated proTPP1 preparation used for crystallization (5 μg); Lane 2 - one ~50 x 100 x 200 μm crystal; and Lane 3 – four ~50 x 100 x 200 μm crystals. Crystal samples for SDS-PAGE analysis were prepared by repeated washes of each crystal by passage through fresh 2 μl drops of reservoir solution containing 0.1 M citrate, 5% polyethylene glycol 6000 (pH 5.0), followed by dissolution in 2 μl water and addition of SDS loading buffer. Samples were subjected to electrophoresis in a Tris-HCl 7.5 % polyacrylamide gel for 45 min at 100 V at room temperature. Gels were stained with Coomassie blue for 20 min, destained overnight and photographed.

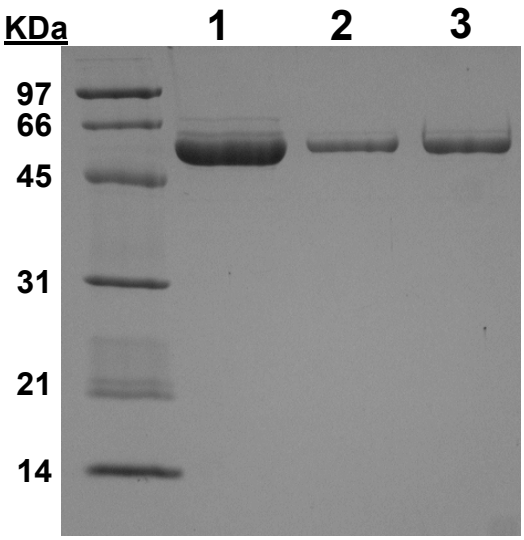


Figure S2. Observed and TLS calculated mean B factor per residue. TLSMD server output (1-3) with observed (black) and TLS calculated mean B factors plotted against residue number for three TLS groups (residues 20-183 in cyan, residues 184-197 in magenta and residues 198-563 in orange) are shown. The vibrational motion detected for the linker region is significantly higher than for the rest of the protein.

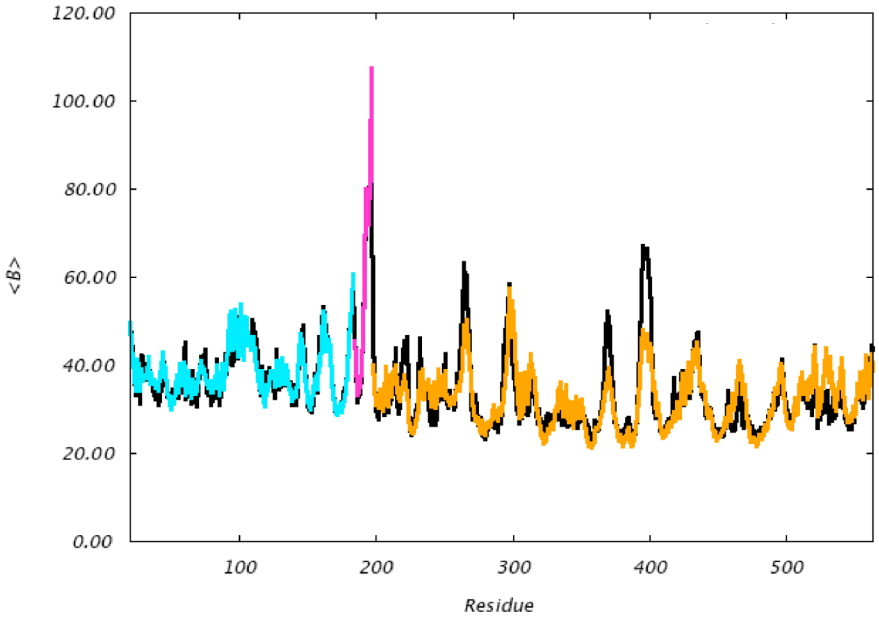
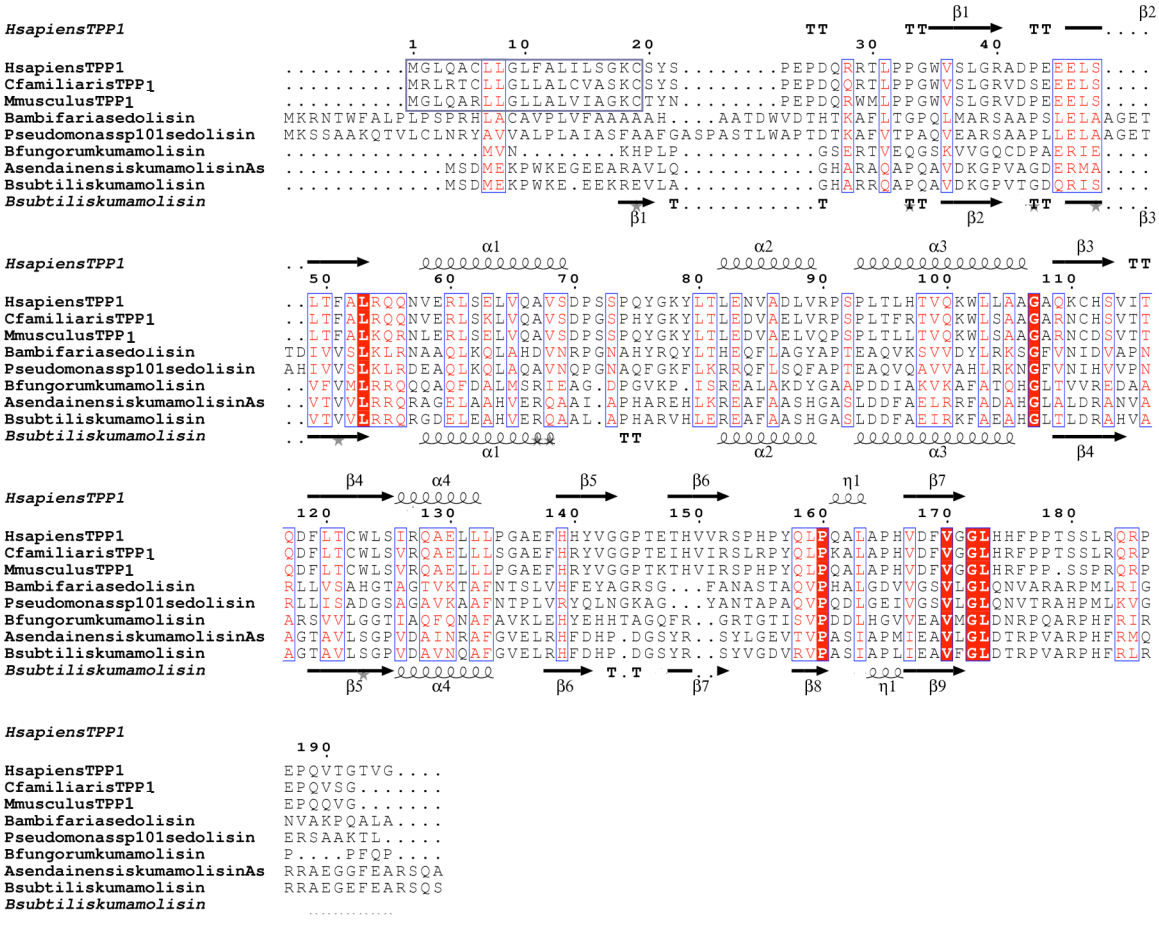


Figure S3. Primary and secondary structural alignments. ClustalW2 (v. 2.0.8) (4) multiple sequence alignments are shown for A. prodomains, and B. mature domains of TPP1 and selected bacterial homologs belonging to the S53 family using an ESPript output (5). Alignments of the secondary structural elements of proTPP1 (PDB ID 3EDY) and prokumamolisin (PDB ID 1T1E) with turns denoted as TT, β sheets as arrows, α helices and 3-10 helices (η helices) as squiggles are shown on top and bottom of the primary structural alignments, respectively. In panel B, the letter "*m*" before secondary structural elements symbolizes the mature domain. The accession numbers used for the alignments are as follows: *Homo sapiens* TPP1 (AAB80725), *Canis familiaris* TPP1 (AAD25043), *Mus musculus* TPP1 (AAD03083), *Burkholderia ambifaria* Sedolisin (EAO47687), *Psuedomonas sp. 101* Sedolisin (AAB36054), *Burkholderia fungorum* Kumamolisin (ZP_00030505), *Bacillus subtilis* Kumamolisin (BAB85637) and *Alicyclobacillus sendainensis* Kumamolisin-As (BAC41257). Conserved regions are boxed with strictly conserved residues shown in red background, well-conserved residues in red and the remainder in black. The *N*-terminal 19 residues of TPP1 encode the cleaved signal peptide. Residue numbers correspond to the *H. sapiens* proTPP1 sequence (AAB80725). Star symbols above the blocks indicate residues at the active site with blue stars used for the catalytic triad and black for the remaining residues, diamonds indicate the cis-Pro residues, and inverted triangles indicate the gylcosylation sites in TPP1.

A.



B.

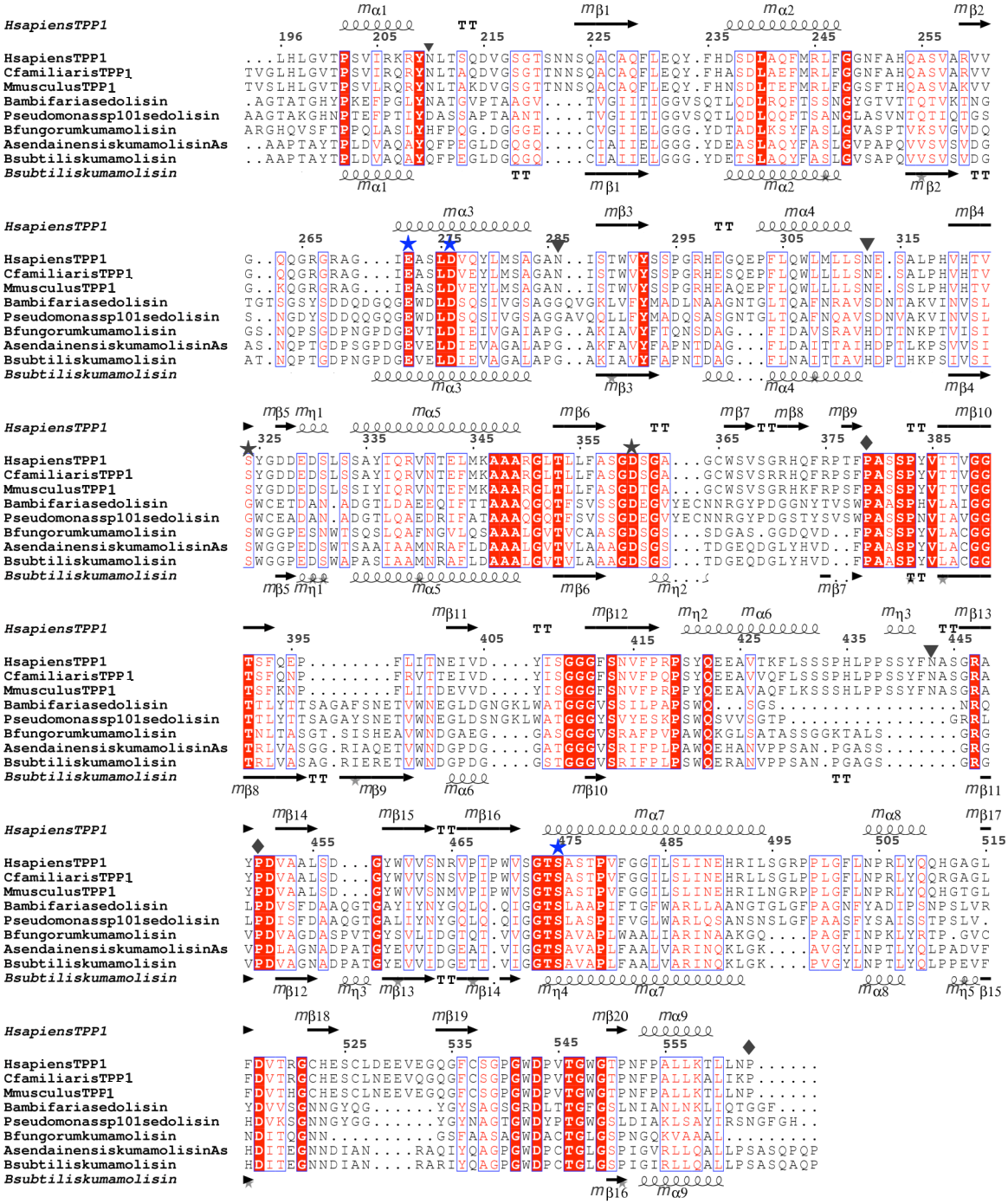
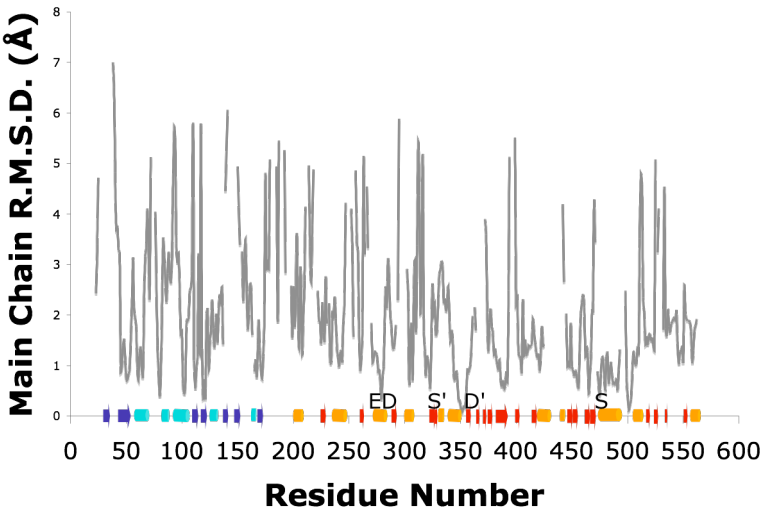

HsapiensTPP1
CfamiliarisTPP1
MmusculusTPP1
Bambifariasedolisin
Pseudomonassp101sedolisin
Bfungorumkumamolisin
AsendainensiskumamolisinAs
Bsubtiliskumamolisin
Bsubtiliskumamolisin

Figure S4. Comparison of proTPP1 and prokumamolisin. A. Main chain r.m.s. deviations of proTPP1 and prokumamolisin residues upon secondary structure matching. Secondary structure matching was performed with proTPP1 (PDB ID 3EDY) and prokumamolisin (PDB ID 1T1E) (6) using the protein structure comparison service SSM at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm) (7). The secondary structural elements of proTPP1 are shown on the x-axis with block arrows for β strands and cylinders for α helices. The propiece is shown in cyan and the mature domain in orange. The catalytic triad residues S475, E272 and D276, the oxyanion residue D360 and residue S324 at the active site are designated as S, E, D, D' and S', respectively. Note that in spite of high r.m.s.d. values throughout the entire length of the zymogen, the active center residues exhibit the lowest r.m.s.d. values, indicating the conservation of the C_α positions of these residues. B. Ribbon diagrams of proTPP1 and prokumamolisin upon superposition of C_α coordinates of mature domains. The C_α coordinates of the mature domain residues of proTPP1 and prokumamolisin matched using the SSM server were superposed using LSQKAB (8) from the CCP4 program suite (9). The respective propieces of proTPP1 and prokumamolisin are shown in cyan and deep blue, the linkers in magenta and green and the corresponding mature domains in orange and wheat, respectively. The endopeptidase cleavage sites are shown in ball-and-stick models in yellow for proTPP1 and in green for prokumamolisin. Residue labels are italicized for prokumamolisin. There are notable differences in the positions of the homologous cleavage sites. Whereas the carbonyl carbon of the scissile peptide bond of the endopeptidase cleavage site in prokumamolisin sits near the active site, the corresponding one in proTPP1 is 12.0 Å away from the active site S475 hydroxyl. Both differences in the conformation of the respective linkers as well as the distal β-sheets (boxed), creating a more open pocket in prokumamolisin as compared to proTPP1, as well as lattice contacts involving the linker region (Fig. S6) are likely to contribute to conformation of the linker and the position of the cleavage site #1 in the proTPP1 structure.

A.


Main Chain R.M.S.D. (Å)
0
1
2
3
4
5
6
7
8
ED
S'
D'
S
0
50
100
150
200
250
300
350
400
450
500
550
600
Residue Number

B.

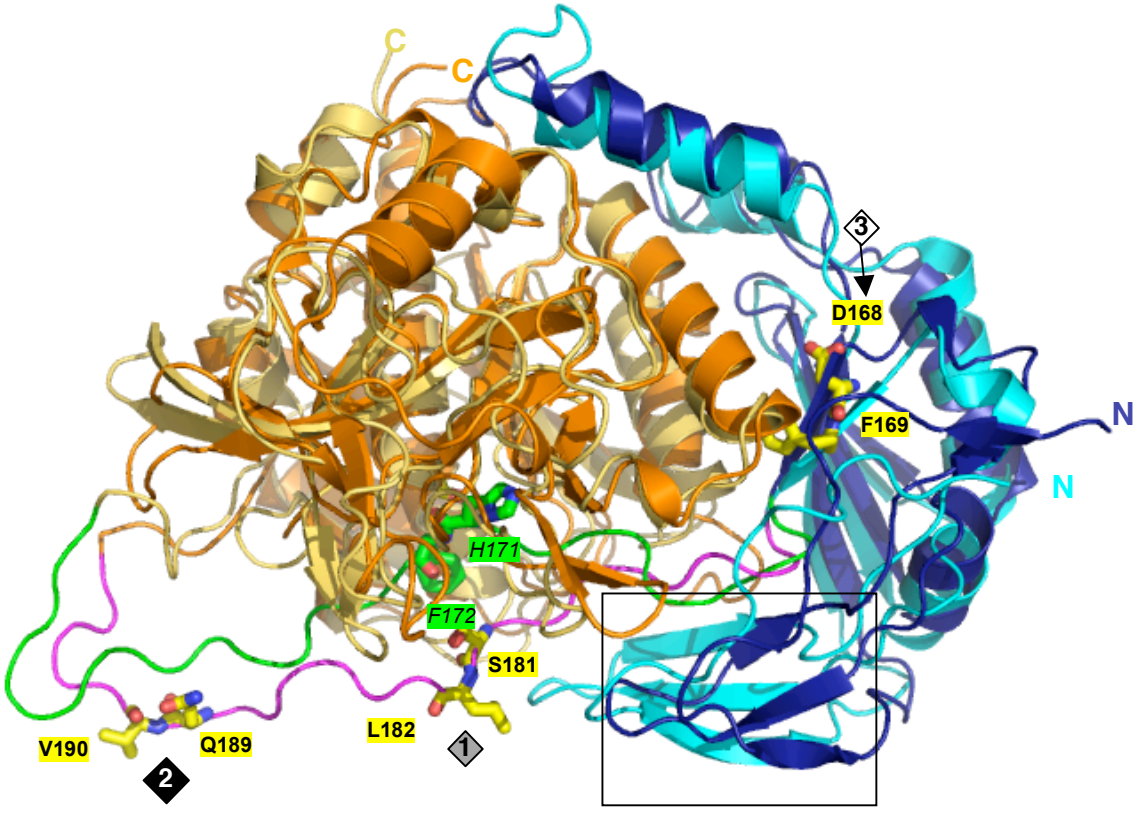

C
C
3
D168
F169
N
N
H171
F172
S181
L182
V190
Q189
2
1

Figure S5. Interdomain contacts in proTPP1. Ribbon representation of proTPP1 with specific contacts between the propiece and mature domain are shown. The propiece is shown in light cyan, the linker in magenta and the mature domain in wheat. The interface is divided into five regions I, II, III, IV and V, with corresponding insets showing specific interactions in each region. Residues involved in hydrophobic interactions are shown in stick models and dotted surfaces and those involved in hydrogen bond and salt bridge interactions as sticks. The ionizable interface residues with significant probabilities of protonation change over pH range from 2.5 to 8.0 (Table S4) are shown in surface filling models with carbon atoms colored as follows: E232 in light green, E299 in deep blue, E302 in green, E24 in yellow, D26 in cyan, D168 in orange, D118 in brown, E343 in pink and D70 in maroon. The electrostatic bonds are shown as dashes with corresponding distances given in Å.

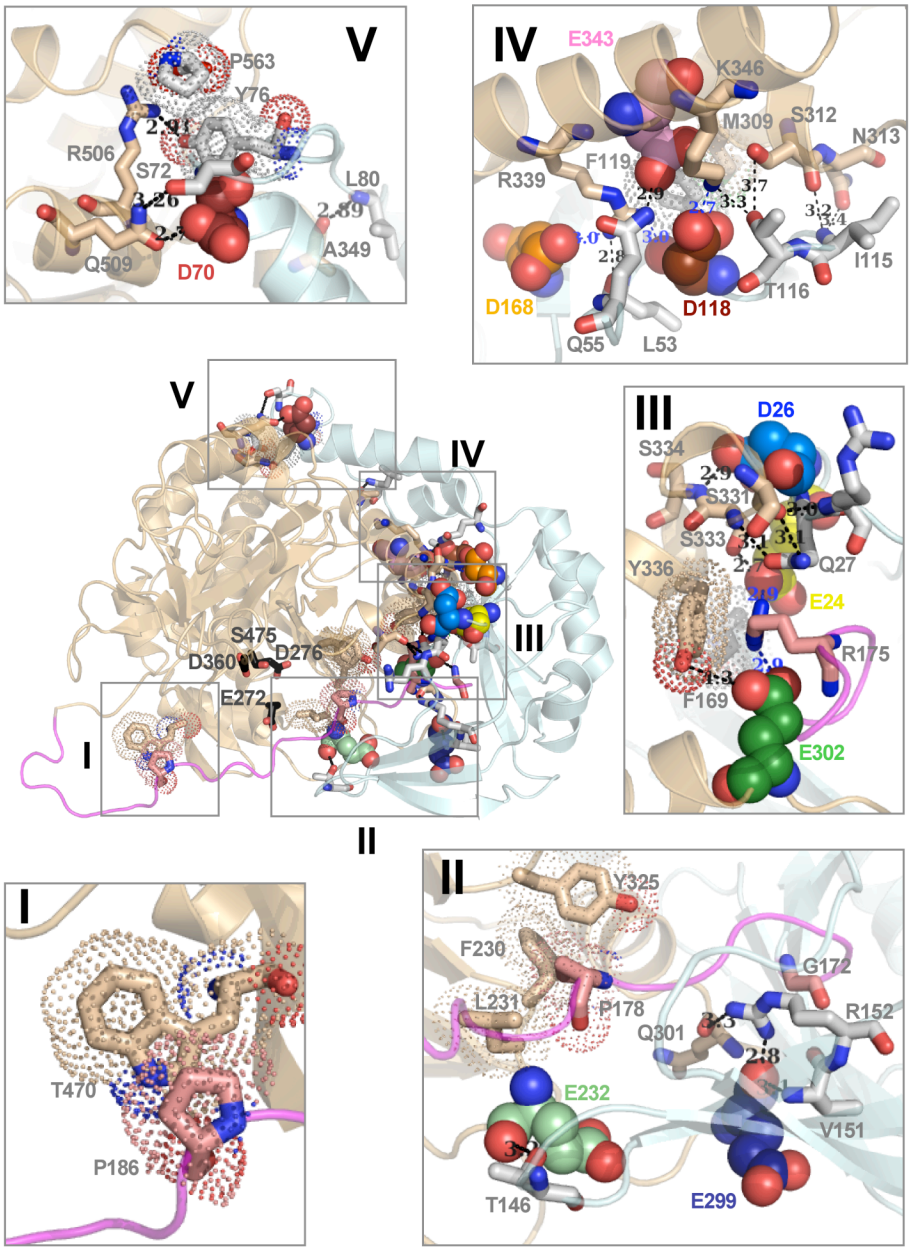


Figure S6. Lattice contacts involving the linker region. Ribbon representations of three symmetry related proTPP1 molecules (#1, #2 and #3) involved in specific contacts (inset) with the linker in molecule #1 are shown with the linker regions and the propieces colored magenta and cyan, respectively. The mature domains of molecules #1, #2 and #3 are shown in orange, yellow and wheat, respectively. In molecule #1, the active site S475 and the three endopeptidase cleavage sites, 1 (S181/L182), 2 (Q189/V190) and 3 (D168/F169), are shown with backbone atoms shown in CPK mode and labels as in Fig. 6. Hydrogen bonds are shown as dashed lines and the interacting residues in ball-and–stick mode, with linker residue labels in yellow background, labels of interacting residues from molecule #2 in green background and one from molecule #3 in blue background. The linker of molecule #1 is involved in seven hydrogen bond interactions with molecule #2 and two with molecule #3. These interactions, which bury cleavage site 2 at the interface with a symmetry related molecule, contribute to the positioning of cleavage site 2 and also, to a lesser extent, cleavage site 1, distant to the active site S475.

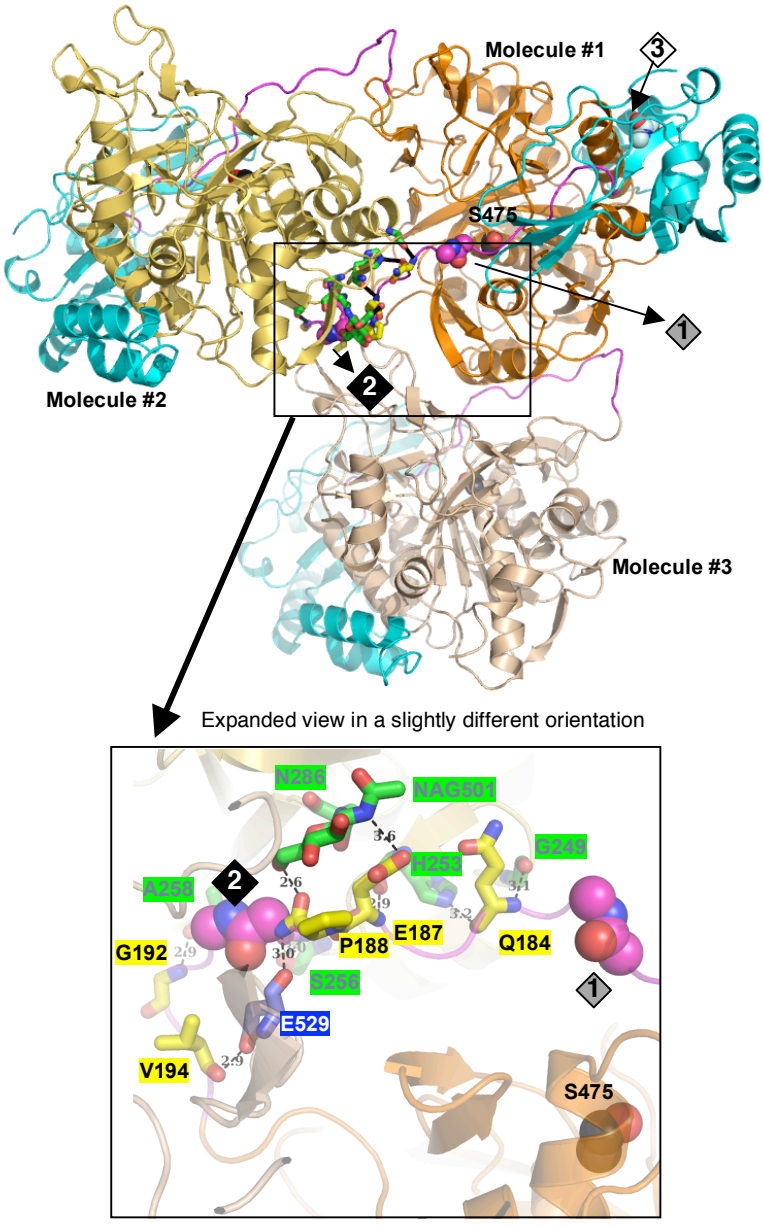


Figure S7. Urea enhanced autoactivation of TPP1 zymogen. Activation reactions were conducted at pH 4.0 in the presence of different concentration of urea using 30 nM proTPP1. **A**. and **B**. Separate short (**A**) and long (**B**) time courses were conducted with the following urea concentrations: 5 M, filled circles; 3 M, open circles; 2 M, filled squares; 1 M, open squares; 0.5 M, filled triangles, 0 M, open triangles. Appearance of TPP1 activity was fit with a single phase exponential association function using Prism v5.01 except for the long time course for the 5 M urea reaction, which was fit to an exponential decay function to model urea-induced denaturation of activated TPP1. **C**. Association rate constants determined using the short (filled diamonds) and long (open diamonds) reaction time courses are shown.

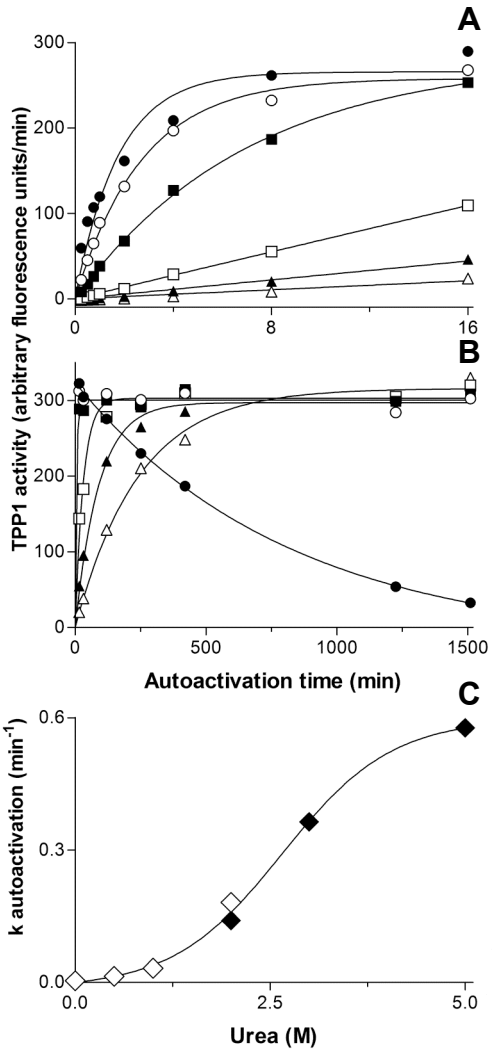


Figure S8. **Mapping of *cln2* missense mutations on the structure of proTPP1.** C_α positions of missense mutation sites using compilation and nomenclature from the neuronal ceroid lipofuscinosis (NCL) mutation database website (http://www.ucl.ac.uk/ncl/cln2.shtml) are shown as spheres on a ribbon representation of proTPP1 crystal structure. Secondary structural elements are colored as follows: Propiece – β strands (blue) and α helices (red); Linker loop (magenta); Mature domain – β strands (green) and α helices (orange). Mutation sites with identifiers indicated within parentheses are: 1) G77R (cln2.012; 2) R127Q (cln2.052); 3) S153P (cln2.043); 4) H175R (cln2.005); 5) P202L (cln2.049); 6) R206C (cln2.041); 7) V216M (cln2.075); 8) V277M (cln2.067); 9) Q278P (cln2.068); 10) I287N (cln2.068); 11) N286S (cln2.063); 12) G284V (cln2.054); 13) R266Q (cln2.074); 14) R339Q (cln2.074); 15) E343K (cln2.018); 16) T353P (cln2.064); 17) L355P (cln2.060); 18) C365R (cln2.002), C365Y (cln2.003); 19) V385 (cln2.020); 20) G389E (cln2.021); 21) Q422H (cln2.022); 22) V426V (cln2.076); 23) K428N (cln2.055); 24) R447H (cln2.006); 25) A454E (cln2.023); 26) W460X (cln2.056); 27) G473R (cln2.057); 28) S475L (cln2.024); 29) V480G (cln2.078); 30) F481C (cln2.066); 31) P544S (cln2.058).

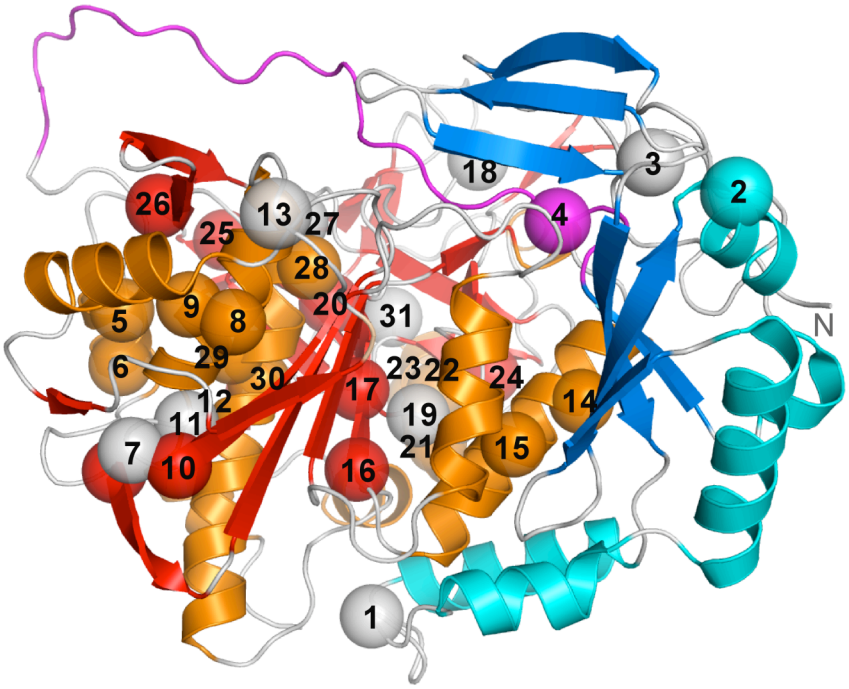


Table S1. Data collection statistics for data sets used for cross-crystal averaging

Values corresponding to highest resolution shells are shown in parentheses.

Crystal	**Native-1**	**Native-2**	**Hg**	**Pb**
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Wavelength (Å)	0.97908	0.97908	0.97908	0.97908
Cell a (Å)	59.8	58.8	59.7	59.6
b (Å)	93.0	88.5	96.7	92.6
c (Å)	102.4	99.0	106.8	101.2
$\alpha=\beta=\gamma$ (°)	90	90	90	90
Resolution (Å)	50.00–1.83 (1.90–1.83)	50.00–1.95 (2.02–1.95)	50.00–2.10 (2.18–2.10)	50.00–1.99 (2,06–1.99)
[a]R_{sym}	0.066 (0.334)	0.095 (0.351)	0.081 (0.218)	0.059 (0.321)
Completeness (%)	99.7 (98.7)	99.7 (98.6)	98.3 (99.9)	99.1 (98.5)
I/σI	34.0 (3.6)	22.4 (5.5)	32.6 (9.2)	37.0 (5.1)

[a] $R_{sym} = \frac{\sum | I_{obs} - I_{avg} |}{\sum I_{avg}}$, where I_{obs} = observed integrated intensity and I_{avg} = average integrated intensity from multiple measurements.

Table S2. List of specific domain contacts in proTPP1.
Contacts between the prodomain and mature domain were calculated using the WHATIF web interface (http://swift.cmbi.ru.nl/servers/html/index.html). A contact is defined as two atoms for which the distance between the van der Waals surfaces is less than 1.0 Å. WHATIF uses the following van der Waals radii: C:1.8Å; O:1.4 Å; N:1.7 Å; S:2.0 Å. Ionizable residues with pKa's within physiological pH range for proTPP1 are indicated in bold.

Hydrophobic interactions				
Pro	**Mature**	**Distance Range (Å)[a]**		
F119	L305	4.05 - 4.38		
F119	M309	3.80 - 4.68		
F169	Y336	3.64 - 4.58		
P178	F230	3.54 - 4.46		
P178	L231	3.38 - 4.60		
P178	Y325	3.89 - 4.59		
Hydrogen bonds				
Pro	**Mature**	**Distance (Å)**	**Energy[b]**	**Type[c]**
E24 OE1	S333 OG	2.65	0.81	S-S
D26 OD2	S334 N	2.85	0.67	S-B
Q27 NE2	S331 O	3.09	0.13	S-B
Q27 OE1	S333 N	3.14	0.49	S-B
R28 N	S331 O	2.97	0.77	B-B
L53 O	R338 NH1	2.82	0.46	B-S
Q55 NE2	**E343** OE2	2.94	0.30	S-S
D70 OD1	Q509 NE2	3.57	0.14	S-S
S72 OG	Q509 NE2	3.26	0.57	S-S
Y76 OH	R506 NH1	2.91	0.74	S-S
L80 N	A349 O	2.89	0.52	B-B
I115 N	S312 O	3.21	0.36	B-B
I115 N	N313 OD1	3.36	0.41	B-S
T116 OG1	S312 OG	3.71	0.63	S-S
T116 OG1	K346 NZ	3.26	0.91	S-S
D118 O	R339 NH2	3.00	0.70	B-S
T146 OG1	**E232** O	3.22	0.69	S-B
V151 N	**E299** O	3.15	0.70	B-B
R152 NH1	**E299** O	2.82	0.49	S-B
R152 NH2	Q301 O	3.29	0.55	S-B
G172 N	**E302** OE2	2.82	0.84	B-S
R175 NE	**E302** OE1	2.90	0.72	S-S
Salt bridges				
Pro	**Mature**	**Distance (Å)**	**Energy[b]**	**Type[c]**
D118 OD2	K346 NZ	2.73	0.86	S-S
D168 OD2	R339 NH1	2.97	0.63	S-S
R175 NH2	**E302** OE2	2.94	0.26	S-S

[a]Distance range for hydrophobic atomic contacts are given.
[b]For hydrogen bonds and salt bridges, the hydrogen bond "energy" is listed as calculated with WHATIF. This "energy" can range from 0.01 for a very poor hydrogen bond to 1.0 for a perfect hydrogen bond.
[c]"Type" indicates whether the contact involves atoms from the backbone (B) or sidechain (S).

Table S3. List of hydrogen bonds between the proTPP1 linker region and symmetry-related proTPP1 molecules.

Lattice contacts were calculated using the WHATIF web interface (http://swift.cmbi.ru.nl/servers/html/index.html). A contact is defined as two atoms for which the distance between the van der Waals surfaces is less than 0.5 Å. WHATIF uses the following van der Waals radii: C-1.8 Å; O-1.4 Å; N-1.7 Å; S-2.0 Å.

Hydrogen bonds				
ProTPP1	**Symmetry Molecule (#1 or 2)[a]**	**Distance (Å)**	**Energy[b]**	**Type[c]**
Q184 N	G249 O (#1)	3.14	0.59	B-B
Q184 O	H253 NE2 (#1)	3.22	0.44	B-S
Q187 N	H253 O (#1)	2.89	0.76	B-B
E187 OE2	NAG501 N2 (#1)	3.41	0.32	S-Su
P188 O	NAG501 06 (#1)	2.63	0.56	B-Su
Q189 O	E529 0 (#2)	2.97	0.72	B-B
Q189 NE2	S256 O (#1)	3.27	0.43	S-B
G192 N	A258 O (#1)	3.04	0.22	B-B
V194 O	E529 OE1 (#2)	2.75	0.58	B-S

[a]The proTPP1 linker region makes hydrogen bonds and van der Waals contacts with two symmetry-related molecules, indicated as #1 or #2.

[b]The hydrogen bond "energy" is listed as calculated with WHATIF. This "energy" can range from 0.01 for a very poor hydrogen bond to 1.0 for a perfect hydrogen bond.

[c]"Type" indicates whether the contact involves atoms from the backbone (B), sidechain (S) or N-acetyl glucosamine moiety (Su).

Table S4. Comparison of predicted pKa values of selected ionizable residues.
PROPKA v 1.0 (10) was used to predict pKa values of individual ionizable residues both in the zymogen (pKa^Z) and free domains (pKa^P for propiece residues and pKa^M for mature enzyme residues). Ionizable residues at the interdomain interface (see Fig. S5 for nomenclature), the active site and the substrate binding pocket have been included in the analysis.

Residue	Domain[a]	pKa^Z	pKa^P	pKa^M
Interdomain interface				
Region II				
E299	M	1.35		4.50
E232	M	4.95		4.57
Region III				
D26	P	2.39	3.06	
E24	P	3.18	3.82	
E302[b]	M	3.05		4.57
Region IV				
E343	M	3.53		3.14
D118[b]	P	-0.65	2.24	
D168[b]	P	1.78	3.94	
Region V				
D70	P	2.94	2.81	
Active site				
E272	M	4.92		3.59
D276	M	6.6		6.37
D360	M	4.69		4.03
Substrate bindng pocket				
D327	M	6.17		3.9

[a]Domain indicates whether residue is part of the propiece (P) or mature domain (M).
[b]Residues involved in salt bridge interactions. The remaining residues are involved in hydrogen bonds.

References

1. Painter, J., and Merritt, E. A. (2005) *Acta Cryst. Sect. D* **61**, 465-471
2. Painter, J., and Merritt, E. A. (2006) *Acta Cryst. Sect. D* **62**, 439-450
3. Painter, J., and Merritt, E. A. (2006) *J. Appl. Cryst.* **39**, 109-111
4. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) *Bioinformatics* **23**, 2947-2948
5. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics* **15**, 305-308
6. Comellas-Bigler, M., Maskos, K., Huber, R., Oyama, H., Oda, K., and Bode, W. (2004) *Structure* **12**, 1313-1323
7. Krissinel, E., and Henrick, K. (2004) *Acta Cryst. Sect. D* **60**, 2256-2268
8. Kabsch, W. (1976) *Acta Cryst. Sect. A* **32**, 922-923
9. Collaborative Computational Project, N. (1994) *Acta Cryst. Sect. D* **50**, 760-763
10. Li, H., Robertson, A. D., and Jensen, J. H. (2005) *Proteins* **61**, 704-721