

The two cathepsin B-like proteases of *Arabidopsis thaliana* are closely related enzymes with discrete endopeptidase and carboxydipeptidase activities

Andreas Porodko, Ana Cirnski, Drazen Petrov, Teresa Raab, Melanie Paireder, Bettina Mayer, Daniel Maresch, Lisa Nika, Martin L. Biniossek, Patrick Gallois, Oliver Schilling, Chris Oostenbrink, Marko Novinec and Lukas Mach

Supplementary material

Table S1 Polypeptide sequences encoded by AtCathB1 splice variants.

AtCathB1 (UniProt: F4HVZ1):

MADSCCIRLHLLASVFLLLFSSFNLQGIAAE[LSKQKLTSLILQNEIVKEVNENPNAGWKAAFNDRFA]
[ATVAEFKRLLGVIQTPKTAYLGVPIVRHDL]SLK↓LPKEFDARTAWSHCTSIRRIL[VG]YILNNVLLWSTITL
[WFW]LLGHCGS[C]WAFGAVESLSDRFCIKYNL[N]VSLSANDVIACCGLLCGFGCNGGFPMGAWLYFKYH
GVVTQECPYFDNTGCSHPGCEPTYTPKCERKCVSRNQLWGESKHYGVGAYRINPDPQDIMAEEVYKN
GPVEVAFTVYEDFAHYKSGVYKYITGTKIGGH[AV]KLIGWGTSDDGEDYWLLA[NQW][N]RSWGDDGYFK
IRRGTNECGIEQSVVAGLPSEKNVFKGITTSDDLLVSSV

AtCathB1 (open reading frame of transcript retaining introns 4 and 5):

MADSCCIRLHLLASVFLLLFSSFNLQGIAAE[LSKQKLTSLILQNEIVKEVNENPNAGWKAAFNDRFA]
[ATVAEFKRLLGVIQTPKTAYLGVPIVRHDL]SLK↓LPKEFDARTAWSHCTSIRRILGRF

AtCathB1 (open reading frame of transcript retaining intron 5):

MADSCCIRLHLLASVFLLLFSSFNLQGIAAE[LSKQKLTSLILQNEIVKEVNENPNAGWKAAFNDRFA]
[ATVAEFKRLLGVIQTPKTAYLGVPIVRHDL]SLK↓LPKEFDARTAWSHCTSIRRILDQVKHHFI

AtCathB1 (open reading frame of transcript with truncated exon 6):

MADSCCIRLHLLASVFLLLFSSFNLQGIAAE[LSKQKLTSLILQNEIVKEVNENPNAGWKAAFNDRFA]
[ATVAEFKRLLGVIQTPKTAYLGVPIVRHDL]SLK↓LPKEFDARTAWSHCTSIRRILDQVLHQI

Proregions are presented in magenta. The insertion present in UniProt entry F4HVZ1 is shown in cyan. Important active-site residues (yellow) and potential N-glycosylation sites (green) are highlighted by coloured backgrounds. The N-terminus of the catalytic domain is indicated by a red arrow.

Table S2 Sequences of AtCathB2 and AtCathB3.

AtCathB2:

MADNCIRLLHSASVFFCLGLLISSFNLLQGIAAE¹LSKQKLT²SWILQNEIVKEVNENPNAGWKASFNDRF³
A⁴ATVAEFKRL⁵LGVKPTPKTEFLGVP⁶IVSHDISL⁷K↓LPKEFDARTAWSQCTSIGRIL⁸DQGHCGS⁹CWAFGA¹⁰
VESLSDRFCIKYNM¹¹NVSLSVNDLLACCGFLCGQGCNGGYPIAAWRYFKHHGVVTEECDPYFDNTGCS¹²H¹³
PGCEPAYPTPKCARKCVSGNQLWRESKH¹⁴YGVSA¹⁵YKVRSH¹⁶PDDIMAEVYKNGPVEVAFTVYEDFAHYK¹⁷
SGVYKHITGTNIGG¹⁸HAVKLIGWGTSDDGEDYWLLAN¹⁹QW²⁰NR²¹SWGDDGYFKIRRG²²TNECGIEH²³GVVAGL²⁴
PSDRNVVKGITTSDDL²⁵VSSF

AtCathB3:

MAVYNTKLCLASVFLLLGLLLA¹FDLKGIEA²ESLTKQKLD³SKILQDEIVKKVNENPNAGWKA⁴AINDRFS⁵
ATVAEFKRL⁶LGVKPTPKKH⁷FLGVP⁸IVSHDPSL⁹K↓LPKAFDARTAWPQCTSIGNILD¹⁰QGHCGS¹¹CWAFGAV¹²
ESLSDRFCIQFGM¹³NISLSVNDLLACCGFRCGDGCDGGYPIAAWQYFSYSGVVTEECDPYFDNTGCSHPG¹⁴
CEPAYPTPKCSRKCVSDNKLWSESKHYSVSTYTVKSNPQDIMAEVYKNGPVEVSFTVYEDFAHYKSGV¹⁵
YKHITGSNIGG¹⁶HAVKLIGWGTSS¹⁷EGEDYWLMA¹⁸NQWNRGWGDDGYFMIRRG¹⁹TNECGIEDEPVAGLPSS²⁰
KNVFRVDTGSNDLPVASV

Proregions are presented in magenta. Important active-site residues (yellow), potential N-glycosylation sites (green) and amino acids subjected to site-directed mutagenesis (cyan) are highlighted by coloured backgrounds. The N-termini of the catalytic domains are indicated by red arrows.

Table S3 Hydrolysis of peptidyl-MCA substrates by wild-type AtCathB2 and AtCathB3.

Substrate	AtCathB2 [mU/mg]	AtCathB3 [mU/mg]
Z-Phe-Arg-MCA	78	69
Z-Leu-Arg-MCA	28	70
Boc-Val-Arg-MCA	10	69
Z-Arg-Arg-MCA	1.0	30
Bz-Phe-Val-Arg-MCA	77	155
Z-Val-Val-Arg-MCA	40	274
Z-Gly-Pro-Arg-MCA	0.6	10
Boc-Val-Leu-Lys-MCA	18	60
Z-His-Glu-Lys-MCA	0.4	12

All substrates were tested at a final concentration of 10 μ M. mU, nmol per min.

Table S4 Oligonucleotide primers used in this study.

AtCathB1_fw	5'-GCCTCTGTTTTCTTGCTC-3'
AtCathB1_rev	5'-CAGGATTGATTCTGTATGCG-3'
AtCathB2_fw	5'-GAATCTAGAGAAAATCTTTCCAAGCAG-3'
AtCathB2_rev	5'-AAAGGTACCTTAAAATGAGGAAACAAGAAGATC-3'
AtCathB3_fw	5'-GACTCTAGAGAAAGTCTTACCAAAC-3'
AtCathB3_rev	5'-ACAGGTACCTTAAACCGATGCAACCG-3'
D127A_fw	5'-ATTGGAAGGATCTTAGCTCAGGGTCACTGTGGT-3'
D127A_rev	5'-ACCACAGTGACCCTGAGCTAAGATCCTTCCAAT-3'
H207A_fw	5'-AATACTGGTTGCTCGGCCCCGGGATGTGAACCC-3'
H207A_rev	5'-GGGTTCACATCCCGGGGCCGAGCAACCAGTATT-3'
G336E_fw	5'-TGTGGCATTGAACATGAGGTTGTAGCTGGTTTA-3'
G336E_rev	5'-TAAACCAGCTACAACCTCATGTTCAATGCCACA-3'

Figure S1 Sequence alignment of the catalytic domains of AtCathB2, AtCathB3, rat and human cathepsin B.

AtCathB2	1	LPKEFDARTAWSQCTSIGRILDQGHCGSCWAFGAVESLSDRECIKYN--MNVSLSVNDLL
AtCathB3	1	LPKAFDARTAWPQCTSIGNILDQGHCGSCWAFGAVESLSDRECIQFG--MNLISLVNDLL
RnCathB	1	LPESFDAREQWSNCPTIAQIRDQGS CGSCWAFGAVEAMSDRICIHTNGRVNVEVSAEDLL
HsCathB	1	LPASFDAREQWPQCPTIKEIRDQGS CGSCWAFGAVEAISDRICIHTNAHVSVVEVSAEDLL

AtCathB2	59	ACCGFLCGQGCNGGYPIAAWRYFKHHGVVTE-----ECDPYFDNTGCSH-----PGC
AtCathB3	59	ACCGFRCGDGCDGGYPIAAWQYFSYSGVVTE-----ECDPYFDNTGCSH-----PGC
RnCathB	61	TCCGIQCGDGCNGGYPSGAWNFWTRKGLVSGGVYNISHIGCLPY-TIPPCEHHVNGSRPPC
HsCathB	61	TCCGSMCGDGCNGGYPAEAWNFWTRKGLVSGGLYESHVGC RPY-SIPPCEHHVNGSRPPC

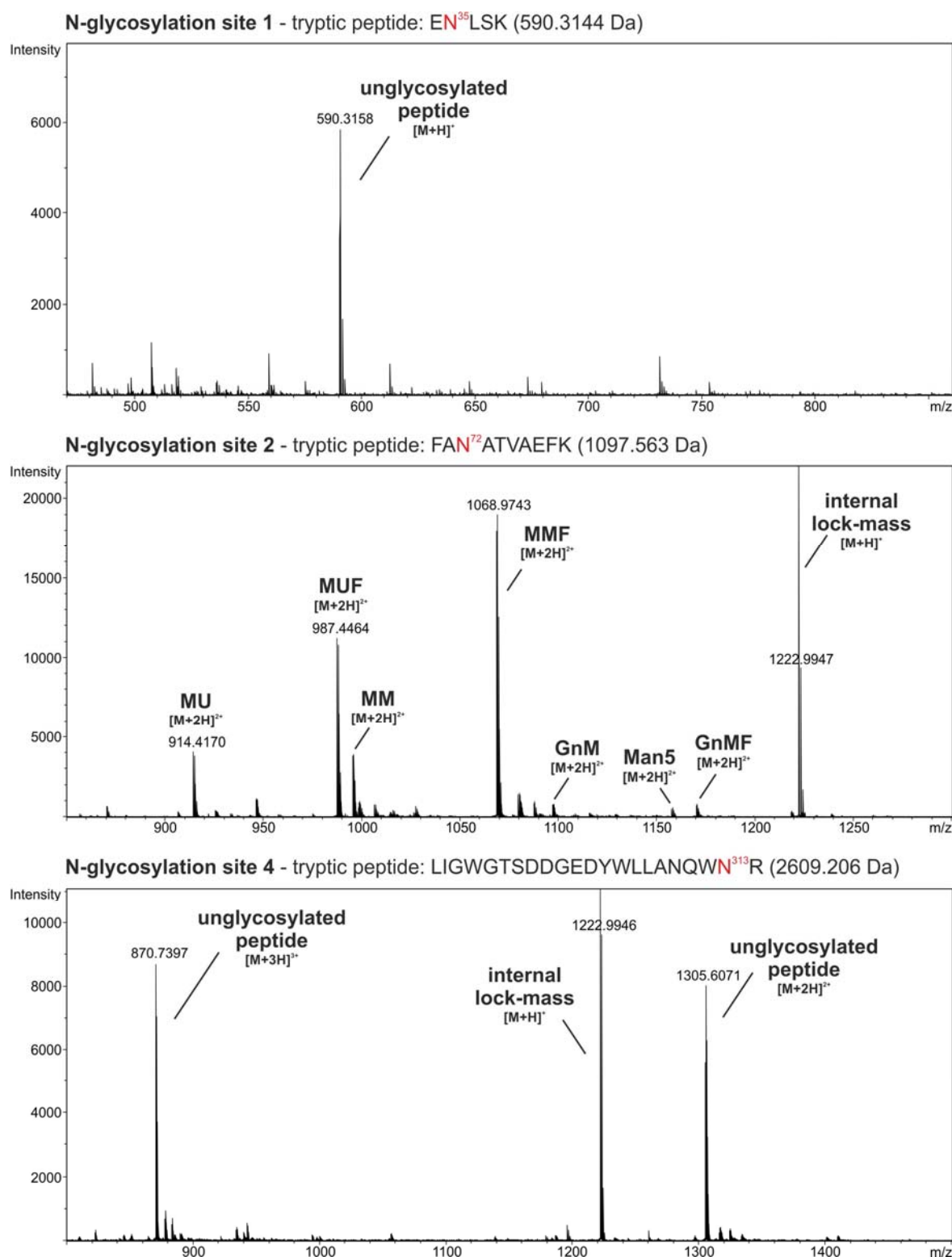
AtCathB2	106	EPAYPTPKCARKCVSGNQ-LWRESKHYGVSAYKVRSHPD DIMAEVYKNGPVEVAF TVYED
AtCathB3	106	EPAYPTPKCSRKCVSDNK-LWSESKHYSVSTYTVKSNPQD DIMAEVYKNGPVEVSE TVYED
RnCathB	120	TGEGDTPKCNKMCEAGYSTSYKEDKHYGYTSYSVSDSEKE IMAEIVYKNGPVEGAF TVESD
HsCathB	120	TGEGDTPKCSKICEPGYSPTYKQDKHYGYNSYSVSNSEK DIMAEIVYKNGPVEGAF SVYSD

AtCathB2	165	FAHYKSGVYKHITGTNIGGHAVKLI GWGTSDDGEDYWLLANQOWNRSWGDDGYFKIRRGTN
AtCathB3	165	FAHYKSGVYKHITGSNIGGHAVKLI GWGTSSEGEDYWLMANQOWNRGWGDDGYFMIRRGTN
RnCathB	180	FLTYKSGVYKHEAGDVMGGHAIRILGWGI-ENGVPYWLVANSWNVWDGDNCEFFKILRGEN
HsCathB	180	FLLYKSGVYQHVTGEMMGGHAIRILGWGV-ENGTPYWLVANSWNTDWGDNCEFFKILRGQD

AtCathB2	225	ECGIEHGVVAGLPSDRNVVKGITTSDDLVSSEF
AtCathB3	225	ECGIEDEPVAGLPSSKNVERVDTGSDNLPVASV
RnCathB	239	HCGIESEIVAGIPRTQQYWGRF-----
HsCathB	239	HCGIESEVVAGIPRTDQYWEKI-----

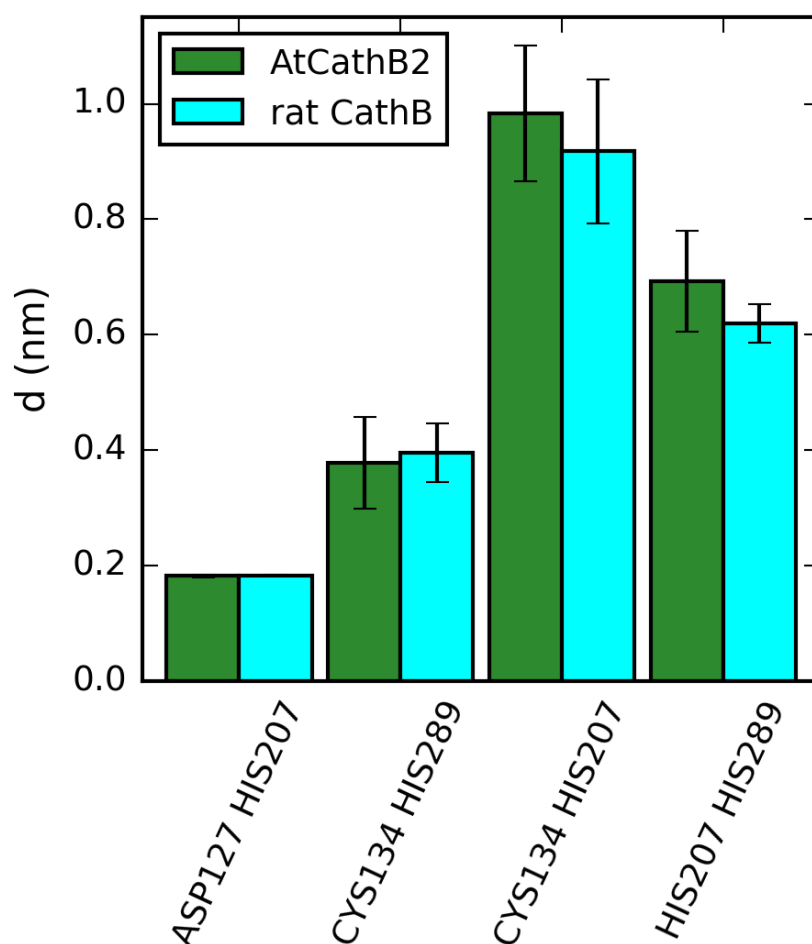
Important active-site residues are highlighted by black triangles. The occluding loop of human cathepsin B (HsCathB) is underlined, and the two histidines contributing to its exopeptidase activity are marked with asterisks. Amino acids contributing to the S2 and S1' subsites are indicated with black circles and white triangles, respectively. RnCathB, rat cathepsin B.

Figure S2 Analysis of AtCathB2 N-glycosylation site occupancy.



The tryptic peptide containing Asn¹⁵⁴ (N-glycosylation site 3) could not be identified. MU, Man₂GlcNAc₂; MUF, Man₂GlcNAc₂Fuc; MM, Man₃GlcNAc₂; MMF, Man₃GlcNAc₂Fuc; GnM, Man₃GlcNAc₃; Man5, Man₅GlcNAc₂; GnMF, Man₃GlcNAc₃Fuc. See www.proglycan.com for linkage details of the detected N-glycan forms.

Figure S3 Distances between key residues of AtCathB2.



Residues are numbered according to the AtCathB2 preproprotein sequence. Note that only the carboxyl group of Asp¹²⁷, the sulphur atom of Cys¹³⁴ and the imidazole rings of His²⁰⁷ and His²⁸⁹ were used to define the minimal distances (d) between the respective amino acids. The equivalent residues in mature rat cathepsin B (CathB) are Asp²², Cys²⁹, His¹¹⁰ and His¹⁹⁹.