Supplemental Materials

Molecular Biology of the Cell

Tort et al.

Supplemental Material

Supplemental figure legends

Figure S1. Structural alignment of CCPs and determination of amino acids of the active site. A) Structural alignment of the catalytic domain of hCCP3 with the CCP crystal structures of *Pseudomonas aeruginosa* (PDB 4a37) (Otero *et al.*, 2012), *Burkholderia mallei* (PDB 3k2k) and *Shewanella denitrificans* (PDB 3l2n) obtained with the RaptorX server. **B)** Comparison of the amino acids shaping the S1' binding pocket in different carboxypeptidases with different defined amino-acid specificities. Residues are indicated following the RasMol amino color scheme that colors amino acids accordingly to their properties. [a]Residue 207 is the major determinant of specificity for M14B carboxypeptidases. [b]Residue 255 is the major determinant of specificity for M14A carboxypeptidases. [c]Carboxypeptidase isolated from *Thermoactinomyces vulgaris* that is a model for broad substrate specificity. [d]M14D is the proposed classification for the M14 members that constitute the CCP subfamily. **C)** Amino acid sequence alignment of the region around position 255 of CCPs and canonical bovine CPA. The positions known to participate in the binding of the substrate C-terminal residue are indicated.

Figure S2. Sequence optimization of active mCCP2 and mCCP3. To confirm that our truncated 65-kD versions of mCCP2 and mCCP3 (Figure 2A, B) represent the shortest active version of these enzymes, we further truncated short sequences at their N- and the C-termini. **A)** Scheme of full-length of mCCP6, mCCP2, mCCP3 and several truncated forms of mCCP2 and mCCP3. The green boxes indicate in the conserved N-terminal domain (Nt) specific to CCPs, the blue box is the conserved carboxypeptidase domain (CP; compare with Figure 2A). Carboxypeptidase domain schemes were delimitated with Superfamily 1.73 database (Gough *et al.*, 2001). Gray lines are non-conserved sequences that were partially truncated in the optimization. **B)** Immunoblot analysis HEK293T cell extracts expressing different forms of YFP-mCCP2 and YFP-mCCP3 as shown in (**A**). The deglutamylase activity is monitored by the generation of Δ2-tubulin. 12G10 is used to control α-tubulin loading. **C)** Secondary structure prediction of an essential C-terminal fragment determined in

mCCP2 and mCCP3 performed with Coils server (Lupas *et al.*, 1991). The fragment present in mCCP2_Z1703 and mCCP3_Z1670, but deleted in mCCP2_1607 and mCCP3_Z1583 has a high probability for coiled-coil structure. Coiled-coil structures are known to participate in structural stabilization and oligomerization of proteins (Parry *et al.*, 2008). It is thus possible that the removal of these coiled-coil sequences from mCCP2 and mCCP3 resulted in misfolded or destabilized enzymes, and consequently in loss-of-activity. Alternatively, it could indicate that both enzymes need to oligomerize for activity.

Figure S3. **Specificity tests for truncated mCCP2 and mCCP3.** Immunoblot analysis of HEK293T protein extracts co-expressing truncated forms of YFP-mCCPs and YFP-telokin variants. **A)** Co-expression of active or dead truncated forms of YFP-mCCP2 and YFP-mCCP3 together with YFP-telokin variants ending in different acidic tails to test their ability to release single of multiple Asp and/or Glu residues. Generation of the Δ2-tubulin epitope on telokin is used to monitor C-terminal degradation. mCCP1 is used as positive control for deglutamylation. Note that only mCCP3-Z1670 is efficiently removing single and consecutive Asp residues. See also Figure 3B. **B)** Epitope mapping of polyG antibody on artificial C-terminal tails of YFP-telokin with different numbers of Gly residues. Note that only poly-Gly chains of four and longer are recognized by this antibody. (*non-specific band recognized by the polyG antibody). **C)** Schematic representation of the experimental setup used to identify deglutamylation and deglycylation activities. YFP-telokin with 3-Glu tails (detected by polyE) or 4-Gly tails (detected with polyG) are co-expressed with YFP-CCPs. The immunoblots show that truncated versions of YFP-mCCP2 and YFP-mCCP3 efficiently remove C-terminal Glu residues thus extinguishing the polyE signal. In contrast, no change in the polyG signal was detected in the deglycylation test.

Figure S4. qRT-PCR analyses expression levels of mCCP1, mCCP4, mCCP5 and mCCP6 in murine tissues and analysis of *Agbl2* and *Agbl3* KO mice. A) Relative expression levels of mCCP1, mCCP4, mCCP5 and mCCP6 in different organs of 4-month-old wild type mice as determined by

qRT-PCR. Average values relative to the *Tbp* gene expression are represented, and error bars represent standard deviation of three to five independent experiments. Experiments are complementary to Figure 4A. Values are used for Figure 4B. **B)** Comparative immunoblot analysis of protein extracts from different organs of 5-weeks old wild type (WT) and *Agbl2* or *Agbl3* KO mice with polyE. Total tubulin levels were detected with anti-α-tubulin antibody (12G10). Note that a 130-kDa substrate shows increased polyE signals stomach of *Agbl2* or *Agbl3* KO mice, and in oviduct of the *Agbl3* KO mouse.

Figure S5. Analyses of the polyglutamylation levels in *Agbl2*/*Agbl3* double KO mice. A) Immuno blots of testes and sperm extracts from four of each, WT and *Agbl2*/*Agbl3* double KO mice (mice 1, 2, 3, 5, 6 and 7 were 4 months old; mice 4 and 8 were 5 weeks old). The polyE and 12G10 signals as represented here have been quantified using the software ImageJ, and polyE values have been adjusted to the total tubulin load detected with the antibody 12G10. Mean values have been plotted in Figure 4E. *Note that the 130-kDa polyE-positive protein band is only detected in the 5-week old mouse (which is shown in Figure 4D), but not in the older individuals.

Figure S6. Strategy for the generation of the knockout mice. Schematic representation in scale of the targeting vector used and all the possible alleles for *Agbl2* and *Agbl3* genes. Orange bar: genomic DNA. Black boxes: exons with their corresponding number. Green and purple arrowheads: LoxP and Flp sequences, respectively. White bar: *neo* cassette, with the neomycin resistance gene (white box). Blue lines: zone of sequence homology for homologous recombination, with the corresponding size in kbp. Black arrowheads: primers used for the PCR genotyping.

Supplemental Tables

Table S1. Primers used to clone or mutagenize CCP genes and their truncated forms.

Gene	Vector	Primers Fw/Rev
mCCP1	pcDNA3.1-EYFP	Fw: cgcggagtcgacACC ATGAGCAAGCTAAAAGTGGTGGGAGAG Rev: cgccgctgatca AATCAGGTGTGTTCTTGATACCTCAG
mCCP2	pEYFP	Fw: cgcggactcgagACC ATGAATGTCCTGCTTGAGATGGCTTTTC Rev: cgccgcagatct TGGGTATGTGTATATATGCAAGGATGGG
mCCP3	pEYFP	Fw: cgcggactcgagACC ATGTCAGAAGATTCAGAGGAGGAAGAC Rev: cgccgcggatcc CTGATGCTGTTGCAAGTTGGCTATC
mCCP3_opt	pEYFP	Synthetic gene optimized for bacterial codon usage (GeneCust).
hCCP3	pOPINFS	from N. Berrow (IRB, Barcelona, Spain)
mCCP4	pEGFP	Fw: cgaattctagccATGGCTGAACAAGAAGGCAGT Rev: ctggatccgggagacacagagatgtcac (non coding region)
mCCP5	pEYFPr	Fw: ccgcgactgcagACC ATGGAGCTGCGCTGTGGGGGATTGC Rev: ccgcgaggtacc TCCCTCTGCGAGTCGGCGGTGAGC
mCCP6	pEYFP	Fw: cgcggactcgagACC ATGGCGGAGCGGAGCCAGACAGCGCC Rev: ccgcagaTCT AAAGGGGGTTGATGGGTCTTTG
mCCP2_N2190	pEYFP	Fw: cgcggactcgagACC ATGAATGTCCTGCTTGAGATGGCTTTTC Rev: ccgcggatcc GCTCTTCTGGTACTGCTCATTCCG
mCCP2_N1992	pEYFP	Fw: cgcggactcgagACC ATGAATGTCCTGCTTGAGATGGCTTTTC Rev: ccgcggatcc TAAATCCATATCTTGTCCAAAGTTATTC
mCCP2_Z1703	pEYFP	Fw: cggactcgagACC atgGACTCACTTCTGCTGAGCTCGCC Rev: ccgcggatcc TAAATCCATATCTTGTCCAAAGTTATTC
mCCP2_Z1334	pEYFP	Fw: cggactcgagACC atgACACTGCAAGGGCCGGACGAC Rev: ccgcggatcc TAAATCCATATCTTGTCCAAAGTTATTC
mCCP2_Z1607	pEYFP	Fw: cggactcgagACC atgGACTCACTTCTGCTGAGCTCGCC Rev: ccgcggatcc GTCAGGATCACAGAAATCCAG
mCCP3_N1998	pEYFP	Fw: cggactcgagACC ATGAGCGAGGACTCTGAAGAAG Rev: ccgcggatcc GTAAACTTCCTGCATATTGTTCTGG
mCCP3_N1809	pEYFP	Fw: cggactcgagACC ATGAGCGAGGACTCTGAAGAAG Rev: ccgcggatcc GGTATCCGTGTTATCGGAGACGC
mCCP3_Z1670	pEYFP	Fw: cggactcgagACC atgGACCCGTTTTTCCCACGCACCAC Rev: ccgcggatcc GGTATCCGTGTTATCGGAGACGC
mCCP3_Z1325	pEYFP	Fw: cggactcgagACC atgGTGGATAACTGCGACAACACCC Rev: ccgcggatcc GGTATCCGTGTTATCGGAGACGC
mCCP3_Z1583	pEYFP	Fw: cggactcgagACC atgGACCCGTTTTTCCCACGCACCAC Rev: ccgcggatcc GTCCGGGTCGCAGTAGTCCAG
mCCP2_E593Q (dead version)	pEYFP	Fw: GCTACACCATGcAGTCTACCTTTGGC Rev: CAAAGGTAGACTgCATGGTGTAGCTG
mCCP3_E540Q (dead version)	pEYFP	Fw: CTTTCACCCTGcAAGCAACTTTCTGCG Rev: GAAAGTTGCTTgCAGGGTGAAAGAATTGC

Table S2. Primary antibodies used in this study

Antibody name	Antigen	Type	Dilutions			Provider
			immuno blot		immuno cyto-chemistry	
			cell lines	tissue extracts		
12G10	α-tubulin	mouse monoclonal	1:1,000	1:400	-	from J. Frankel, E. M. Nelson, University of Iowa, USA
anti-β-tubulin	β-tubulin	mouse monoclonal	-	-	1:200	Sigma #T5201
anti-Δ2-tubulin	CEGEEEGE-COOH	rabbit polyclonal	1:5,000	1:5,000	1:1,000	our own production
anti-deTyr-tubulin	-CGEEEGEE-COOH	rabbit polyclonal	1:2,000	-	-	Millipore #AB3201
polyE	-CEEEEEEEEE-COOH	rabbit polyclonal	1:4,000	1:10,000	-	our own production
polyG	-CGGGGGGGGG-COOH	rabbit polyclonal	1:6,000	-	-	our own production
anti-GFP	GFP, YFP, CFP	rabbit polyclonal	1:5,000	-	-	Torrey Pines Biolabs, #TP401

Table S3. Search for C-terminal acidic amino acid stretches

	Number of proteins with uninterrupted E/D-stretches at the C-terminus				
Organism	**2**	**3**	**4**	**5**	**6+**
Homo sapiens	287	60	16	12	24
Mus musculus	218	55	9	4	23
Drosophila melanogaster	47	9	2	3	6
Caenorhabditis elegans	42	14	6	3	4

	Number of proteins uninterrupted E-stretches at the C-terminus				
Organism	**2**	**3**	**4**	**5**	**6+**
Homo sapiens	94	12	1	1	3
Mus musculus	76	7	1	1	4
Drosophila melanogaster	14	1	0	0	2
Caenorhabditis elegans	16	0	0	0	2

Table S4. Primers used for qRT-PCR

Specificity	Gene	Primer name	Sequence (5'-3')
mouse	CCP1	AGTPBP1-U1	TTCCACAGAGTCAGATACTGCCAGAT
		AGTPBP1-L1	CAGAACTTCCATGCCTGTAGAACCT
	CCP2	AGBL2-U2	AATCTGCAGAAAGCCGTCAGAGT
		AGBL2-L2	AGTGTGTTTGTCCGTGTAGAGGTCA
	CCP3	AGBL3-U1	CTGTTTACCCAAACTCCAAGGAAGAT
		AGBL3-L1	GGATGTTCGGTTACCCCCAACT
	CCP4	AGBL1-U2	GAGCTGTCCTGTAGCTTTGAGGAACT
		AGBL1-L2	AAGCAACACTTGGAATGTGGTGGT
	CCP5	AGBL5-U2	GCACCCAAAAGGTCAGCCAT
		AGBL5-L2	GCCGCCTTCTGTCTGAGCA
	CCP6	AGBL4-U2	CCAAGAGTCTTTACCGAGATGGGAT
		AGBL4-L2	CTGTGGTCTGGGCAGCGATAGT

Supplemental References

Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J. Mol. Biol. *313*, 903–919.

Lupas, A., Dyke, M. Van, and Stock, J. (1991). Predicting coiled coils from protein sequences. Science *252*, 1162–1164.

Otero, A., Rodríguez de la Vega, M., Tanco, S., Lorenzo, J., Avilés, F. X., and Reverter, D. (2012). The novel structure of a cytosolic M14 metallocarboxypeptidase (CCP) from Pseudomonas aeruginosa: a model for mammalian CCPs. FASEB J. *26*, 3754–3764.

Parry, D. a D., Fraser, R. D. B., and Squire, J. M. (2008). Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. J. Struct. Biol. *163*, 258–269.

A

```
3l2nA   1 P Y S Y E R H L D L I S A V Q - - - - L H P L V S T E H L G L T L D G R D M T L V K V G D D D P - - - 44
3k2kA   1 P Y S E E R H S E F L G A V Q - - - - Q M P Q A S V V E L G R T V E G R P M S L V V L G T P D - - - - 43
4a37A   1 P Y S R E R H A R L V E R A L - - - - G I E G V E R L A V G T S V Q G R D I E L L R V R R H P D - - - 44
hCCP3   1 P Y T Y T N L Q E Y L S G I N N D P V R S K F C K I R V L C H T L A R N M V Y I L T I T T P L K N S D 51

3l2nA  45 - - S K K S I W I T A R Q H P G E T M A E W L V E G L L N Q L L D - - - - N D C P T S K A L L D K A N 89
3k2kA  44 - - A K K K V W I I A R Q H P G E S M A E W F I E G L V K R L V G W G D W S G D P V A R K L Y D H A T 92
4a37A  45 - - S H L K L W V I A Q Q H P G E H M A E W F M E G L I E R L Q R - - - - P D D T E M Q R L L E K A D 89
hCCP3  52 S R K R K A V I L T A R V H P G E T N S S W I M K G F L D Y I L G - - - - - N S S D A Q L L R D T F V 97

3l2nA  90 F Y I V P N M N P D G S V R G H L R T N A V G A N L N R E W Q T P S L E R S P E V Y Y V V N K M H E T 140
3k2kA  93 F Y I V P N M N P D G S V H G N L R T N A A G A N L N R E W M E P D A E R S P E V L V V R D A I H A I 143
4a37A  90 L Y L V P N M N P D G A F H G N L R T N A A G Q D L N R A W L E P S A E R S P E V W F V Q Q E M K R H 140
hCCP3  98 F K V V P M L N P D G V I V G N Y R C S L A G R D L N R N Y T S L L K E S F P S V W Y T R N M V H R L 148

                                  K250                R255
3l2nA 141 - - - - G V D L F Y D V H G D E G L P Y V F L A G C E G I P N Y S D K L A S L Q Q D F V A A L S L A S 187
3k2kA 144 - - - - G C D L F F D I H G D E D L P Y V F A A G S E M L P G F T E Q Q R V E Q S A F I D S F K R A S 190
4a37A 141 - - - - G V D L F L D I H G D E E I P H V F A A G C E G N P G Y T P R L E R L E Q R F R E E L M A R G 187
hCCP3 149 M E K R E V I L Y C D L H G H S R K E N I F M Y G C D G S D R S - K T L Y L Q Q R I F P L M L S K N C 198

3l2nA 188 A D - F - - - Q T E F G Y D K D E P G K A N L T V A C N W V A N T F K C L S N T L E M P F K D N A N L 234
3k2kA 191 P D - F - - - Q D E H G Y P P G K Y R E D A F K L A S K Y I G H R F G C L S L T L E M P F K D N A N L 237
4a37A 188 - E - F - - - Q I R H G Y P R S A P G Q A N L A L A C N F V G Q T Y D C L A F T I E M P F K D H D D N 233
hCCP3 199 P D K F S F S A C K F N V Q K S K E G - - - - - T G R V V M W K M G I R N S F T M E A T F C G S T L G 244

3l2nA 235 A D P F Q G W S P E R S V Y F G E A S L I A M R A V I D K I G Q 266
3k2kA 238 P D E H I G W N G A R S A S L G A A M L G A I L E H V R A F - - 267
4a37A 234 P E P G T G W S G A R S K R L G Q D V L S T L A V L V D E L R - 264
hCCP3 245 N K R G T H F S T K D L E S M G Y H F C D S L L D Y C D P D R T 276
```

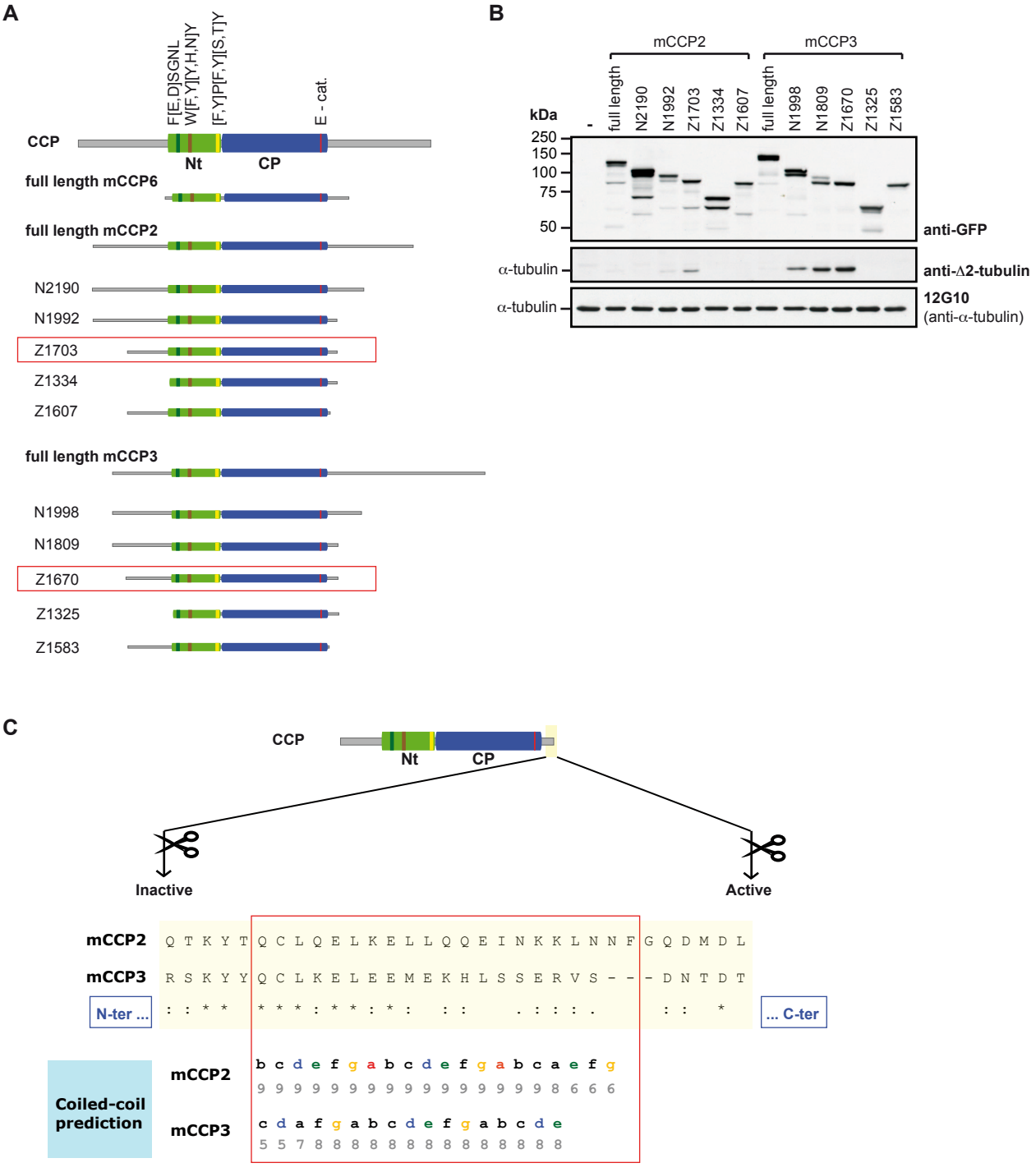
B

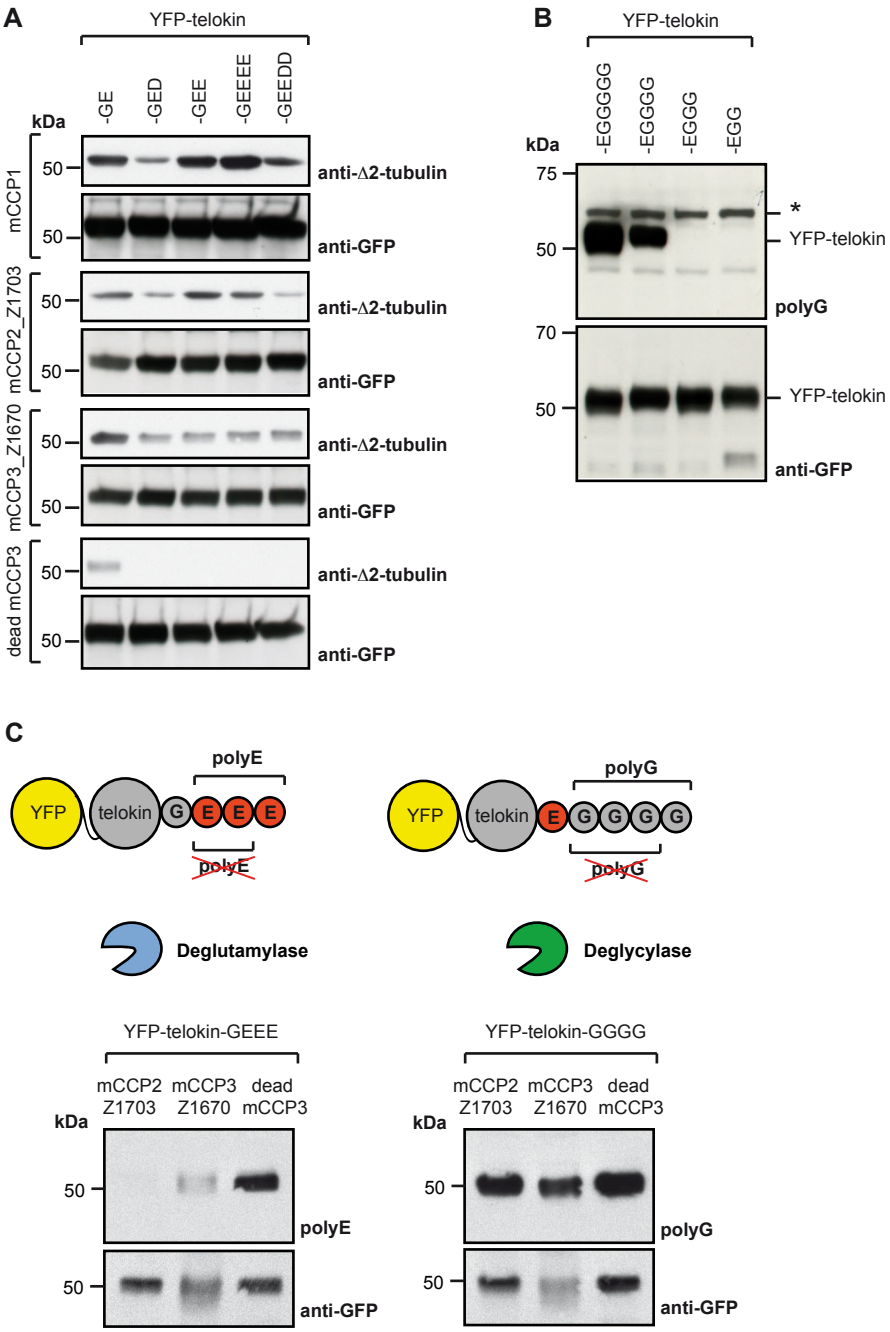
Position in bovine CPA (active form)	Homologous amino acid residues forming S1' subsite									Substrate specificity	MEROPS classification
	194	203	207[a]	243	247	250	253	255[b]	268		
Bovine CPA	Ser	Leu	Gly	Ile	Ile	Ala	Gly	Ile	Thr	Hydrophobic	M14A
Human CPA1	Ser	Met	Gly	Ile	Ile	Ala	Ser	Ile	Thr	Hydrophobic	
Human CPA2	Thr	Met	Gly	Ile	Ile	Ala	Gly	Ile	Ala	Hydrophobic	
Human CPA4	Asp	Met	Gly	Thr	Val	Ala	Ser	Ile	Thr	Hydrophobic	
CPT[c]	Thr	Leu	Gly	Gln	Leu	Thr	Asp	Thr	Thr	Broad	
Human CPB	Thr	Ile	Ser	Gly	Ile	Ala	Gly	Asp	Thr	Basic	
TAFI (CPU)	Ser	Val	Ser	Gly	Leu	Ala	Gly	Asp	Thr	Basic	
Human CPE	Asn	Asn	Asp	Gly	Trp	Val	Gly	Gln	Thr	Basic (Arg)	M14B
Human CPD	Asn	Asn	Asp	Asn	Phe	Val	Gly	Gln	Thr	Basic (Lys)	
Human CPO	Thr	Leu	Gly	Asn	Leu	Ser	Ser	Arg	Thr	Acidic	M14A
Human/ mouse CCP1	Asp	Val	Gly	Met	Ser	Lys	Thr	Arg	Thr	Acidic	M14D[d]
Human/ mouse CCP2	Asp	Val	Gly	Met	Ser	Arg	Thr	Arg	Thr	Acidic	
Human/ mouse CCP3	Asp	Ile	Gly	Phe	Asn	Lys	Thr	Arg	Thr	Acidic	
Human/ mouse CCP4	Asp	Ile	Gly	Phe	Lys	Lys	Thr	Arg	Thr	Acidic	
Human/ mouse CCP5	Asp	Cys	Gly	Phe	Asn	Lys	Ser	Arg	Thr	Acidic	
Human/ mouse CCP6	Asp	Gly	Gly	Tyr	Ser	Lys	Thr	Arg	Thr	Acidic	

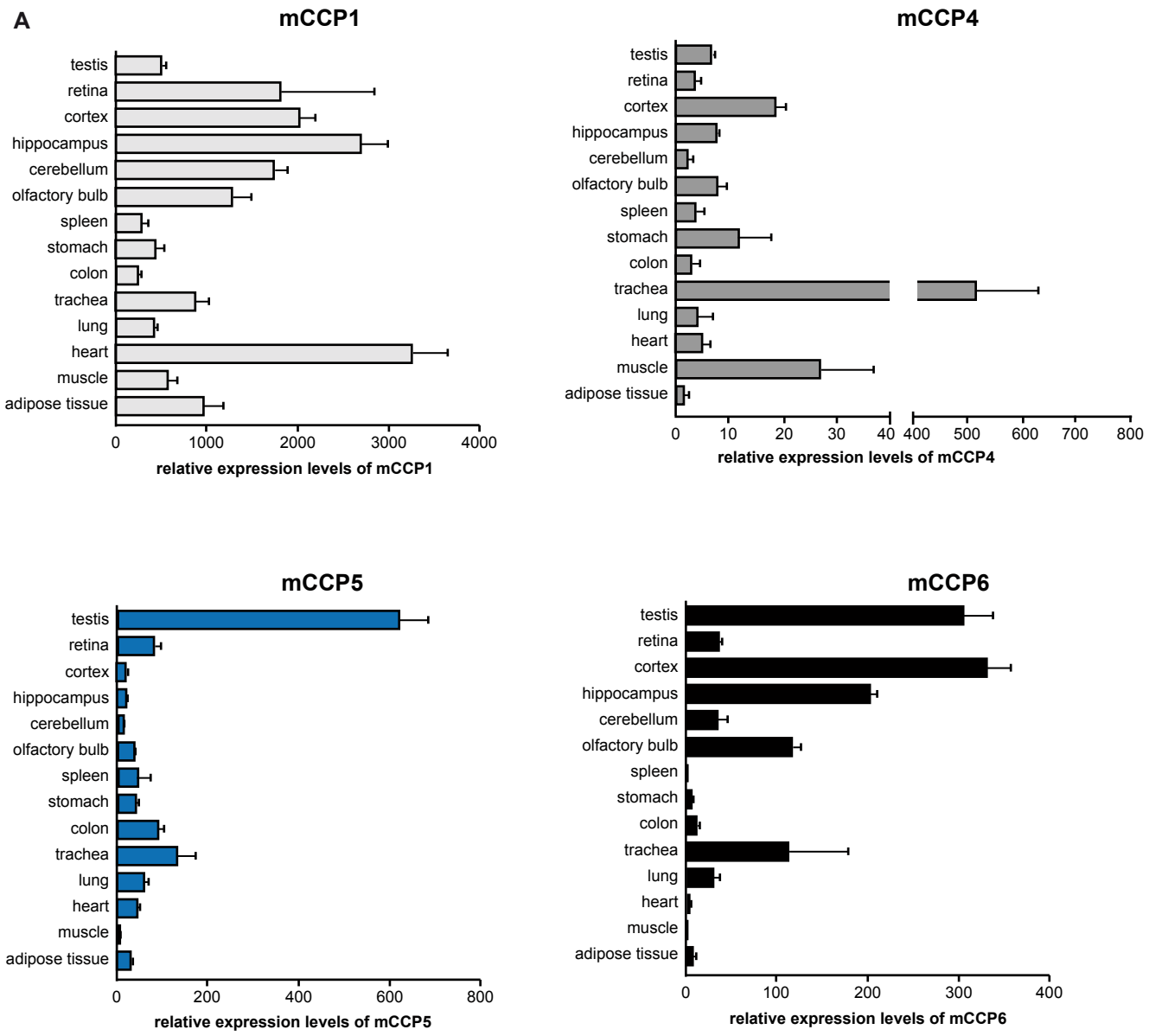
C

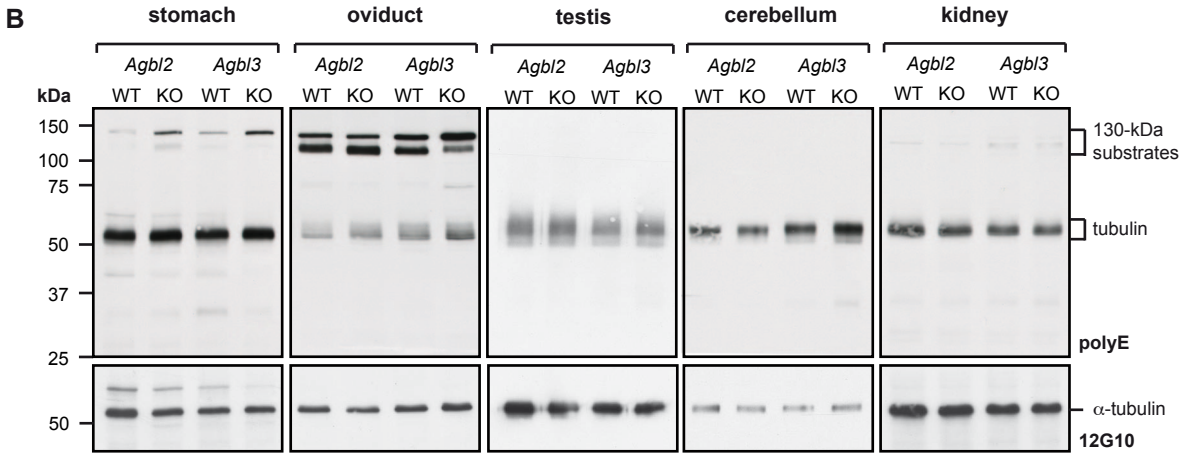
```
                  243 247                 250 253 255             268 270
CPA1_BOVIN  T-SYKYG-SIITTIYQ--------------ASGGSIDWSYNQ--GIKYSFTFELRDTGR
CCP1        I-APA-F-CMSSCSFVVEK----------SKESTAR VVVWREIGVQRSYTMESTLCGC
CCP4        L-APA-F-TMSSCSFLVEK----------SRASTAR-VVVWREMGVSRSYTMESSYCGC
CCP2        N-APDKF-SFHSCNFKVQK----------CKEGTGR-VVMWR-MGILNSYTMESTFGGS
CCP3        N-CPDKF-SFSACKFNVQK----------SKEGTGR-VVMWK-MGIRNSFTMEATFCGS
CCP5        N-SAH-F-DFQGCNFSEKNMYARDRRDGQSKEGSGR-VAIYKASGIIHSYTLECNYNTG
CCP6        N-AED-F-SYSSTSFNRDA----------VKAGTGRRFLGGLLDHTSYCYTLEVSFYSY
```

Tort et al., Supplementary Figure S2

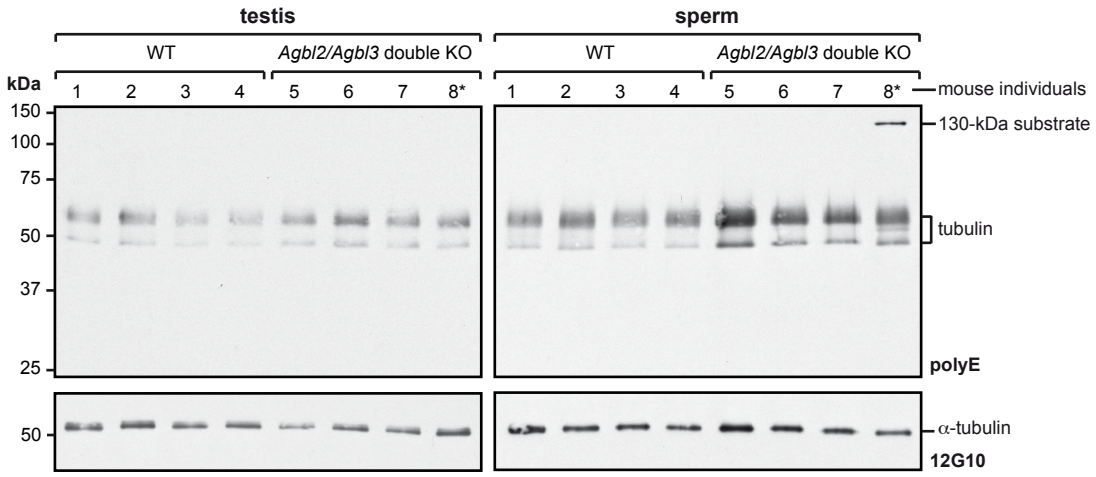



A
YFP-telokin
-GE
-GED
-GEE
-GEEEE
-GEEDD
kDa
mCCP1
50
anti-Δ2-tubulin
50
anti-GFP
mCCP2_Z1703
50
anti-Δ2-tubulin
50
anti-GFP
mCCP3_Z1670
50
anti-Δ2-tubulin
50
anti-GFP
dead mCCP3
50
anti-Δ2-tubulin
50
anti-GFP
B
YFP-telokin
-EGGGGG
-EGGGG
-EGGG
-EGG
kDa
75
*
YFP-telokin
50
polyG
70
YFP-telokin
50
anti-GFP
C
YFP
telokin
G
E
E
E
polyE
polyE
Deglutamylase
YFP
telokin
E
G
G
G
G
polyG
polyG
Deglycylase
YFP-telokin-GEEE
mCCP2 Z1703
mCCP3 Z1670
dead mCCP3
kDa
50
polyE
50
anti-GFP
YFP-telokin-GGGG
mCCP2 Z1703
mCCP3 Z1670
dead mCCP3
kDa
50
polyG
50
anti-GFP


A
mCCP1
testis
retina
cortex
hippocampus
cerebellum
olfactory bulb
spleen
stomach
colon
trachea
lung
heart
muscle
adipose tissue
0
1000
2000
3000
4000
relative expression levels of mCCP1
mCCP4
0
10
20
30
40
400
500
600
700
800
relative expression levels of mCCP4
mCCP5
0
200
400
600
800
relative expression levels of mCCP5
mCCP6
0
100
200
300
400
relative expression levels of mCCP6


B
stomach
oviduct
testis
cerebellum
kidney
Agbl2
Agbl3
WT
KO
kDa
150
100
75
50
37
25
50
130-kDa substrates
tubulin
polyE
α-tubulin
12G10

Tort et al., Supplementary Figure S5



Tort et al., Supplementary Figure S6