Postdoc Fellowships for non-EU researchers

Final Report

Name	Sebastian Tanco
Selection	2012
Host institution	Proteome Analysis and Bioinformatics Unit - VIB Department of Medical Protein Research, UGent
Supervisor	Prof. Kris Gevaert/Prof. Petra Van Damme
Period covered by this report	from 01/08/2013 to 31/01/2015
Title	Analysis of cytosolic metallo-carboxypeptidases (CCPs) mediated C-terminal processing by means of positional proteomics using *in vitro* and *in vivo* models

1. Objectives of the Fellowship (1/2 page)

The subfamily of cytosolic carboxypeptidases (CCPs), which is composed of six members, was reported to function in the C-terminal processing of tubulin. CCP1 is the best-characterized member of this subfamily, and gene disruption of mouse *Ccp1* leads to the ataxic Purkinje cell degeneration (*pcd*) mouse phenotype characterized by neuronal degeneration and infertility. Besides, CCP1 was found over-expressed during axonal regeneration in mice. The association of CCP1 with neuronal cell survival and regeneration makes the study of its mechanism of action of importance for a variety of neurodegenerative diseases.

In fact, the knowledge on CCP actions is very limited and apart from tubulin, their natural substrates are largely unknown. Recently, however, it was demonstrated that CCP1, CCP4, and CCP6 can shorten the gene encoded C-terminal glutamate stretch of two important regulators of myosin functions: myosin light chain and telokin. The possibility of finding new substrates for CCPs and concomitantly identify their substrate specificity profile *in vivo* and *in vitro* (*i.e.* in biological extracts), motivated our research proposal.

This Belspo postdoctoral fellowship project proposal comprised the following milestones:

1) Proteomics-based identification of novel CCP substrates using different *in vitro* and *in vivo* model systems.

More specifically, for the search of natural substrates of CCPs we proposed to apply C-terminal COFRADIC, a degradomics technology that enriches for C-termini from whole proteome digests, thereby allowing for the analysis of C-terminal protein modifications and C-terminal protein processing events.

Complementary to the degradomics studies, we proposed the identification of interacting partners of CCPs through affinity purification mass spectrometry (AP-MS) technologies, with the objective of identifying substrates and regulators of CCP function among their interactors. For this purpose, we applied Virotrap, a technology recently developed in our laboratory (Eyckerman *et al.*, submitted for publication).

2) Biochemical validation of selected CCP substrates and molecular characterization of the biological outcome of these post-translational modifications and their implication in cellular signaling.

2. Methodology in a nutshell (1/2/ page)

For the search of natural substrates of CCPs, we used C-terminal COFRADIC, a positional proteomics strategy that enables the identification of proteolytically generated C-termini (Van Damme *et al.*, PMID 20526345). C-terminal COFRADIC allows for the quantitative analysis of C-terminal peptides, by enriching these peptides from whole proteome digests thereby directly pointing to peptides that report protein processing of gene-encoded C-termini. The COFRADIC technology exploits (differential) peptide chromatographic behavior to specifically isolate peptides of interest from complex peptide mixtures. In C-terminal COFRADIC, not only C-terminal but also N-terminal peptides can be enriched for, and subsequently analyzed by mass spectrometry. Here, after blocking all primary amines by acylation, the proteome is digested with trypsin. By performing an SCX fractionation of the modified tryptic digests at low pH (pH 3.0), both blocked N-termini and α-carboxyl free C-termini carry a net charge of zero and are thus enriched in the SCX flow through fraction. In contrast, internal tryptic peptides are retained by the SCX resin. Subsequent chemical derivatization of primary α-amines (present at C-terminal peptides) using an N-hydroxysuccinimide (NHS) ester of butyrate, performed in between two consecutive RP-HPLC separations, allows for segregation of N- and C-terminal peptides prior to their analysis by LC-MS/MS.

In search of interaction partners of CCP1, we employed Virotrap, a new technology that exploits the spontaneous particle formation and multimerization by HIV-1 p55 GAG protein expression (Eyckerman *et al.*, manuscript in preparation). These virus-like particles are here used to purify bait proteins together with their interaction partners. This technology constitutes a novel concept in protein-protein interaction research, since it aims at extracting cellular protein complexes in their native context.

3. Results (6-8 pages)

Proteomics-based identification of novel CCP substrates using different *in vitro* and *in vivo* model systems.

We initially applied our degradomics technology to study CCP1 in a cellular system. Consequently, we used C-terminal COFRADIC to differentially analyze C-terminal peptides from HEK293T cells overexpressing CCP1 and using mock-transfected HEK293T control cells. We chose HEK293T cells since a previous study showed that this cell line constitute a simple and appropriate cellular system to study CCP1 function. Stable isotope-labeling using isotopic variants of N-hydroxysuccinimide (NHS)-butyrate (*i.e.* NHS ester of $^{12}C_4$-butyrate for the control sample and $^{13}C_4$-butyrate for the sample with CCP1 over-expression) enabled the relative quantification of isolated C-terminal peptides.

As a result of this analysis, 886 unique database-annotated protein C-termini and 75 processed protein C-termini (*i.e.* only considering peptides lacking up to 20 amino acids from database annotated C-terminus) were identified. Most of the latter were found at similar levels in both proteomes and thus point to proteins processed by proteases other than CCP1 or to artifacts (*i.e.* C-terminal ragging), generated during sample preparation. Interestingly, we identified 9 processed C-termini that were present at much higher levels (over 7-fold up, $p < 0.0001$) in the proteome of HEK293T cells over-expressing CCP1 (Table 1). We focused our analysis on these peptides, since they directly point to putative CCP1 substrates, as C-terminal peptides generated by proteolytic cleavage of CCP1 (*i.e.* neo-C-termini) are expected to be present at much higher levels in the proteome where CCP1 is over-expressed.

Table 1 – List of proteins processed by CCP1 at their gene-encoded C-terminus

Protein	Accession number	Assay	Identified neo-C-terminus	Cleaved amino acids	L/H ratio	Subcellular Localization	Function
α-tubulin 1A/1B	Q71U36/ P68363	C-terminal COFRADIC	EDMAALEKDYEEVGVDSVEGEGEEEG	EE(Y) [a]	N/A	Cytoplasm	Microtubules
α-tubulin 1C	Q9BQE3	C-terminal COFRADIC	EDMAALEKDYEEVGADSADGEDEG	EE(Y) [a]	N/A	Cytoplasm	Microtubules
Eukaryotic translation initiation factor 4H	Q15056	C-terminal COFRADIC	EEVVQKEQ	E	N/A	Cytoplasm	Translation initiation
Stathmin	P16949	C-terminal COFRADIC	KNKESKDPADETEA	D	0.029	Cytoplasm	Microtubules regulation
40S ribosomal protein S9	P46781	C-terminal COFRADIC	KNAKKGQGGAGAGDDE	EED	N/A	Cytoplasm	Translation initiation
		C-terminal COFRADIC	KNAKKGQGGAGAGDD	EEED	N/A		
TRAF-type zinc finger domain-containing protein	O14545	C-terminal COFRADIC	TAKAKPSKQQGAGDA	EEEEEE	0.14	Cytoplasm	Regulation TLR4 and RLH pathways
High mobility group protein B3	O15347	C-terminal COFRADIC	KKVEEEDE	EEEEEEEEEEEEDE	0.083	Nucleus/ Cytoplasm	Transcription regulation, chromatin
		C-terminal COFRADIC	KKVEEED	EEEEEEEEEEEEEDE	N/A		
High mobility group protein B2	P26583	Western blot		EDEEEEEEEEDEDEEEEDEDEE [b]		Nucleus/ Cytoplasm	Transcription regulation, chromatin
High mobility group protein B1	P09429	Western blot		EEEEDEEDEEDEEEEEDEEDEDEEEDDDDE [b]		Nucleus/ Cytoplasm	Transcription regulation, chromatin
Myosin light chain kinase 1/ Telokin [c]	Q6PDN3	Western blot		EEEEEE [d]		Cytoplasm	Regulator of myosin function

N/A, not applicable for the impossibility to calculate an *m/z* ratio due to the absence of light peptide ions (*i.e.* peptide not identified in the control sample).

[a] In the case of α-tubulin, we considered that CCP1 uses as substrate the pool of detyrosinated tubulin naturally present in the cell.

[b] The C-terminus of these proteins is displayed, although no CCP1 cleavage sites have been identified.

[c] This substrate was not identified in our screen, but the orthologous mouse substrate was previously reported by Rogowski *et al.* (PMID 21074048).

[d] Based on western blots showing an increase in Δ2-tubulin.

In particular, two of these neo-C-termini corresponded to processed C-termini of α-tubulin, a protein known to be processed by CCP1. C-terminal COFRADIC identified the processed C-terminus of α-tubulin 1A/1B (EDMAALEKDYEEVGVDSVEGEGEEEG; residues 423-448) as nearly exclusively present in the proteome of HEK293T cells over-expressing CCP1 (Fig. 1). Moreover, we also identified a neo-C-terminus (EDMAALEKDYEEVGADSADGEDEG; residues 423-446) of a different α-tubulin isotype (α-tubulin 1C) as being uniquely present in the CCP1 over-expression sample. Both peptides correspond to the C-terminus of Δ3-tubulin, a tubulin form that lacks the last three residues of α-tubulin. Coincident with previous reports, our data confirm the ability of CCP1 to generate Δ2-tubulin or Δ3-tubulin by removing 1 or 2 Glu residues from the C-terminus of detyrosinated α-tubulin and considering that another enzyme, different from CCP1, is responsible for releasing C-terminal Tyr from α-tubulin.

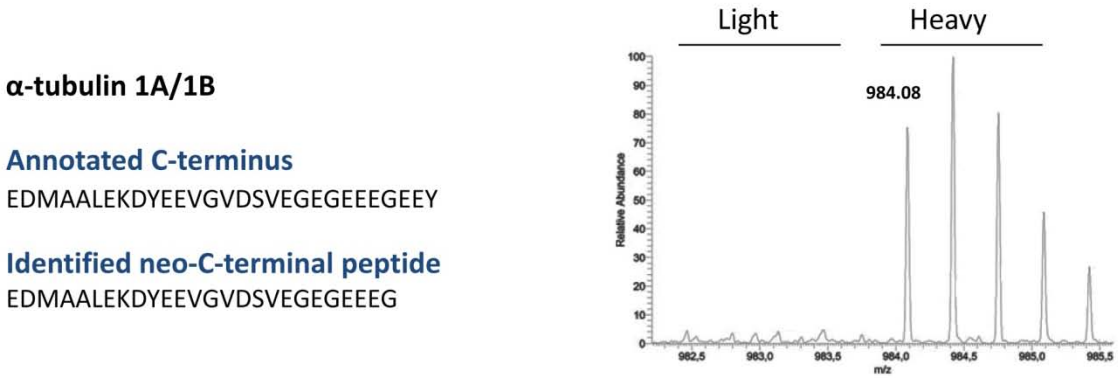

Fig. 1. Identification of C-terminal proteolytic processing events in cells overexpressing CCP1 by means of C-terminal COFRADIC. Identification of α-tubulin 1A/1B as CCP1 substrate. MS-spectrum of the triply charged neo-C-terminus of α-tubulin 1A/1B; But$^{13}C_4$-EDMAALEKDYEEVGVDSVEGEGEEEG (residues 423-448; But$^{13}C_4$, α-amino group modified by $^{13}C_4$-butyric acid).

Besides α-tubulin, we identified 7 neo-C-termini originating from 5 new putative CCP1 substrates (Table 1). Interestingly, in all cases the amino acids released by CCP1 are acidic amino acids, a substrate specificity in

line with previous reports. Two of these peptides indicated CCP1-mediated removal of a single amino acid from the protein's C-terminus: release of glutamate from the C-terminus in case of the eukaryotic translation initiation factor 4H (eIF4H), and of aspartate in the case of stathmin. Previous reports demonstrated the action of CCP1 over C-terminal glutamates, and it was suggested that CCP1 does not recognize aspartate. Our observed trimming of the C-terminus of stathmin indicates that CCP1 is indeed able to release C-terminal aspartates. Moreover, two peptides pointed to C-terminal processing of the 40S ribosomal protein S9 (RPS9). Here, CCP1 removed respectively 3 and 4 amino acids from the C-terminus of RSP9 (*i.e.* the C-terminal Asp and two or three subsequent Glu residues were removed, Table 1). Given that M14 MCPs release only one amino acid at a time, it is expected that CCP1 sequentially releases these subsequent amino acids. The TRAF-type zinc finger domain-containing protein (TRAD1) was also identified as a putative CCP1 substrate. TRAD1 contains a C-terminal stretch of six glutamates and CCP1 seems to be capable of releasing all of them. Remarkably, we also identified two neo-C-termini of the high mobility group protein B3 (HMGB3), which shows that up to 15 or 16 amino acids are removed by CCP1.

Interestingly, whenever identified, the levels of corresponding intact C-termini of these putative CCP1 substrates were not significantly altered. This might indicate that the extent of processing of these putative CCP1 substrates in cells is low.

We applied a similar strategy to compare the proteome of HEK293T cells where human CCP1 was knocked-down using shRNAs, with the proteome of scramble shRNA treated cells. However, with this approach no substrate of CCP1 could be univocally identified. It is likely that, as a result of the low stoichiometry of CCP1 processed C-terminal proteoforms, with respect to intact CCP1 substrates, neo-C-termini are scarce and thus fall below the detection limit. Similarly, no substrate(s) could be identified when comparing the C-terminomes of MEF cells derived from *pcd* mice with MEF cells derived from wt mice.

Biochemical validation of selected CCP substrates and molecular characterization of the biological outcome of these post-translational modifications and their implication in cellular signaling.

From the 5 new putative CCP1 substrates identified, two were selected for further validation; TRAD1 and HMGB3. TRAD1 contains a C-terminal stretch of 6 glutamates residues, and CCP1 is able to sequentially remove all of them (Fig. 2a). We, therefore, used the polyE antibody that was designed to recognize a polyglutamate side-chain of at least three consecutive glutamate residues in tubulin, but additionally recognizes gene-encoded C-termini of three or more glutamates. N-terminal Strep- and HA-tagged TRAD1 was over-expressed in HEK293F cells together with CCP1 or an inactive form of CCP1 (CCP1-E270Q). The latter contains a mutation of a key catalytic residue of M14 metallo-carboxypeptidases (E270Q using the numbering system of bovine carboxypeptidase A, E1102Q in human CCP1). This catalytic residue polarizes the metal-bound water molecule responsible for the attack of the scissile bond. Due to the low sensitivity of the polyE antibody, we enriched TRAD1 by means of its N-terminal Strep-tag. Subsequent Western blot analysis of TRAD1 with antibodies directed to its N-terminal HA-tag (Fig. 2b) showed that TRAD1 is enriched from both cell lines with similar efficiency. Note that CCP1 is co-purified because it also holds a Strep-tag at its C-terminus (Fig. 2b). Interestingly, the polyE antibody detected TRAD1's C-terminus when co-expressed with inactive CCP1, but not when co-expressed with active CCP1 (Fig. 2b). In addition, we directly incubated purified recombinant TRAD1 with purified recombinant CCP1 and found that TRAD1 was no longer recognized by the polyE antibody (Fig. 2c). However, in the presence of o-phenanthroline, a specific inhibitor of metallo-proteases, the polyE antibody was still capable of recognizing affinity purified TRAD1, overall indicating CCP1 dependent removal of the acidic tail of TRAD1. Hence, our data validate TRAD1 as a novel and direct CCP1 substrate.

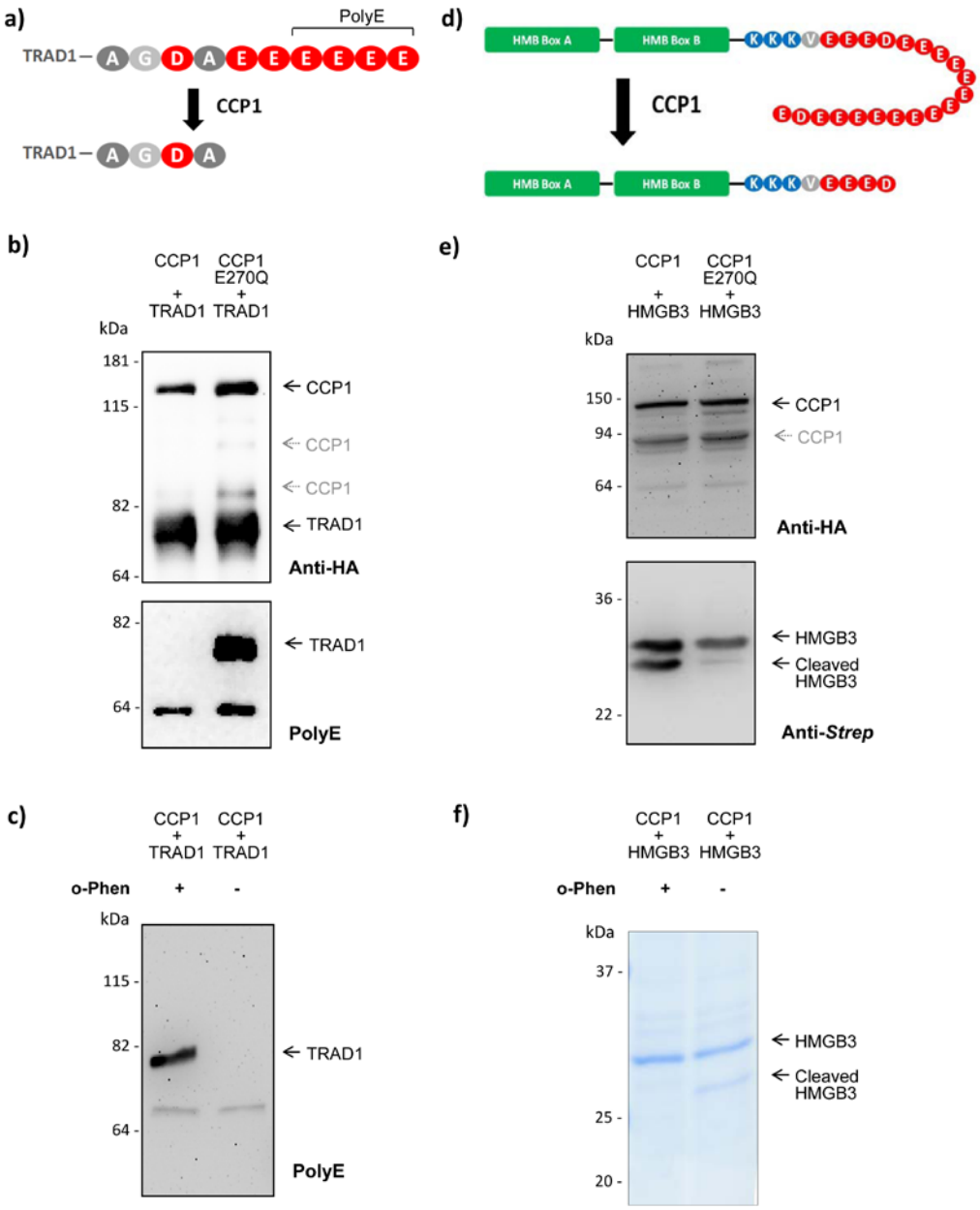

Fig. 2. Validation of TRAD1 and HMGB3 as direct CCP1 substrates. a) CCP1 is capable of removing all acidic C-terminal amino acids of TRAD1. The epitope recognized by the PolyE antibody is indicated. **b)** Active CCP1 processes the C-terminus of TRAD1. Strep/HA- tagged TRAD1 was cotransfected in HEK293F cells either with active CCP1 or inactive (E270Q) CCP1. TRAD1 was enriched from protein extracts using Strep-Tactin columns. Equal amounts of protein eluate were analyzed by Western blot using an anti-Strep antibody. CCP1 is also enriched using its C-terminal Strep-tag. The polyE antibody shows the integrity of TRAD1 C-terminus when co-expressed with inactive CCP1, but not when co-expressed with the active enzyme. **c)** *In vitro* validation of TRAD1 as a direct CCP1 substrate. Purified recombinant TRAD1 was incubated with CCP1 for 2 h at 37 °C, in the presence or absence of the MCP inhibitor o-phenanthroline (o-Phen), and analyzed by Western blot using the polyE antibody. **d)** HMGB3 is composed of two central DNA binding domains (HMG-boxes) and a long acidic C-terminal tail. CCP1 is able to sequentially release up to 16 acidic amino acids from its C-terminus. The size of the HMG boxes and the acidic tail are merely illustrative and does not represent their real sizes. **e)** HMGB3 is processed by CCP1. Strep-tagged HMGB3 was co-transfected with active or inactive (E270Q) CCP1. Equal amounts of each cell extract were analyzed by Western blot using an HA-tag antibody, to demonstrate equal expression levels of both CCP1 variants. Western blotting using an anti-Strep antibody shows the appearance of a degraded form of HMGB3 when co-expressed with active CCP1. **f)** *In vitro* validation of HMGB3 as a direct CCP1 substrate. Purified recombinant HMGB3 was incubated with CCP1 ON at 37 °C, in the presence or absence of 10 mM o-Phen, and analyzed by SDS-PAGE.

As for HMGB3, CCP1 seems to remove quite a long stretch of acidic amino acids from its C-terminus (Fig. 2d). Accordingly, we expected to observe a significant change in the MW of the HMGB3 precursor versus its processed counterpart(s) by means of SDS-PAGE. As such, N-terminally Strep-tagged HMGB3 was co-expressed in HEK293T cells with either active or inactive CCP1. Western blot analysis showed that in addition to the intact HMGB3 band, a proteolytic fragment of HMGB3 is found when HMGB3 is co-expressed with active CCP1 (Fig. 2e). The appearance of an HMGB3 fragment, with different electrophoretic mobility, is consistent with the CCP1-mediated processing identified by C-terminal COFRADIC. By assaying recombinantly produced and purified proteins, we could also show direct processing of HMGB3 by CCP1 only in the absence of o-phenanthroline (Fig. 2f). Finally, we used intact protein mass spectrometry to measure the masses of the purified HMGB3 protein(-fragments). Q-TOF analyses determined the molecular weight of intact HMGB3 purified from HEK293F cells to be 25,300 ± 12.65 Da, which indicates that this protein lacks its initiator methionine, is N-terminally acetylated and lacks a C-terminal glutamate (predicted MW 25303.3 Da). Peptide mass fingerprint analysis of digested HMGB3 confirmed the presence of these post-translational modifications. The mass of the CCP1 generated HMGB3 fragment is 23,377 ± 11.67 Da, which corresponds to the release of 15 additional C-terminal residues of HMGB3 by CCP1 (predicted MW 23380.6 Da). Hence, the *in vitro* cleavage product fully correlates with one of the HMGB3 neo-C-termini identified by C-terminal COFRADIC, and all of these results validate HMGB3 as a direct CCP1 substrate.

Table 2 – List of human proteins containing a stretch of at least 5 consecutive acidic residues in their gene-encoded C-terminus

Number	Name	UniProt	C-terminus
1	High mobility group protein B1	P09429	SKKKK**EEEEDEEDEEDEEEEEDEEDEDEEEDDDDE**
2	Putative high mobility group protein B1-like 1	B2RPK0	SKKKK**EEEEDEEDEEDEEEEDEEDEEDDDDE**
3	High mobility group protein B2	P26583	KKNEP**EDEEEEEEEEDEDEEEEDEDEE**
4	High mobility group protein B3[a]	O15347	ARKKV**EEEDEEEEEEEEEEEEEEDE**
5	Homeobox protein Hox-A7	P31268	AADKA**DEEDDDEEEEDEEE**
6	Cytosolic purine 5′-nucleotidase	P49902	ITHCH**DEDDDEEEEEEEE**
7	Translocation protein SEC63 homolog	Q9UGP8	FEDSF**EEEEEEEEDDD**
8	mTERF domain-containing protein 2	Q7Z6M4	EAEDN**DEDEDDDEEE**
9	60S ribosomal protein L22	P35268	FQINQ**DEEEEEDED**
10	Calsequestrin-1	P31415	GEINT**EDDDDDDDD**
11	Ubiquitin thioesterase ZRANB1	Q9UGI0	SLSDG**EEDEDDEDE**
12	Calcium-dependent secretion activator 1	Q9ULU8	SMKDS**DEEDEEDD**
13	Putative high mobility group protein B3-like	P0C6E5	AQKKV**EEEDEDEE**
14	RNA polymerase-associated protein LEO1	Q8WVC0	KYVIS**DEEEEDDD**
15	LIM domain and actin-binding protein 1	Q9UHB6	RNRYY**DEDEDEE**
16	Single-pass membrane protein with aspartate-rich tail 1, mitochondrial	Q9H4I9	DIFVP**EDDDDDD**
17	Calsequestrin-2	O14958	DNDDS**DDDDDE**
18	Cell cycle control protein 50B	Q3MIR4	RYQDQ**DDDDEE**
19	Dipeptidyl aminopeptidase-like protein 6	P42658	TVTAK**EDEEED**
20	Homeobox protein Hox-B7	P09629	DRAEA**EEEEEE**
21	Metaxin-1	Q13505	LGMA**EEDEEE**
22	40S ribosomal protein S9[a]	P46781	GAGAG**DDEEED**
23	Protein timeless homolog	Q9UNS1	KRYQI**EDDEDD**
24	TRAF-type zinc finger domain-containing protein 1[a]	O14545	GAGDA**EEEEEE**
25	Acidic leucine-rich nuclear phosphoprotein 32 family member E	Q9BTT0	AEDDG**EEEDD**
26	Interferon alpha-inducible protein 6	P09912	KYLDS**EEDEE**
27	Myosin light chain kinase, smooth muscle[a]	Q15746	GEGEG**EEEEE**
28	Choline-phosphate cytidylyltransferase A	P49585	AYDIS**EDEED**
39	Protein SSX1	Q16384	EISDP**EEDDE**
30	Protein SSX2	Q16385	EISDP**EEDDE**
31	Protein SSX3	Q99909	EISDP**EEDDE**
32	Protein SSX4	O60224	EISDP**EEDDE**
33	Protein SSX7	Q7RTT5	EISDP**EEDDE**
34	Protein SSX8	Q7RTT4	EIRDP**EEDDE**
35	Protein SSX9	Q7RTT3	EISDP**EEDDE**
36	Transmembrane protein 190	Q8WZ59	EETEG**EEEED**

[a] Proteins identified as CCP1 substrates.

Our findings indicate that CCP1's substrate specificity is restricted to Glu and Asp residues. We exploited this characteristic to perform an *in silico* scan for other potential CCP1 targets using the ScanProsite tool, and using the pattern [ED](n)> where n represents the number of consecutive acidic residues at a protein's C-terminus, which was allowed to vary from 1 to 32. The pattern search was performed in the Homo sapiens taxon of the 2014_03 UniProtKB/Swiss-Prot database, using the default settings (except that no splice variants were allowed). Our search returned a list of 16, 36 and 111 human proteins containing respectively 7, 5 and 3 or more consecutive acidic amino acids at their C-terminus. Table 2 shows the protein hits with 5 or more consecutive acidic C-terminal residues. The number of protein hits listed in Table 2 is 76 times higher than expected by chance alone; among the 20,257 human proteins in the 2014_03 UniProtKB/Swiss-Prot database less than one protein (or 1 out of 43159 proteins) is expected to have 5 or more consecutive acidic C-terminal residues (considering that the natural frequencies of glutamate and aspartate for human proteins in Swiss-Prot are 7.1% and 4.7%, respectively). More so, finding proteins with 10 or more consecutive C-terminal acidic residues is highly unlikely (1 in $1.86x10^9$ proteins); however, 9 human proteins have acidic stretches of at least 10 residues.

Our ScanProsite analysis pointed to different high mobility group proteins as potential CCP1 targets (Table 2). In particular, the high mobility group protein B1 (HMGB1) and B2 (HMGB2) contain an unusually high number of acidic residues at their C-termini (30 and 22 residues respectively). To confirm that in line with HMGB3, these HMGB proteins could represent CCP1 substrates, N-terminally Myc-tagged versions of both proteins were co-expressed in HEK293T cells with either active or inactive CCP1. Western blot analyses showed that CCP1 indeed processes the C-terminus of HMGB2. CCP1 was also capable of processing the C-terminus of HMGB1, but to a lower extent when compared to HMGB3 or HMGB2. In the latter case, processing leads to several protein fragments.

The putative natural substrates here identified together with the previously reported substrate MLCK1/telokin (Table 1), support a role for CCP1 in the proteolytic processing of gene-encoded acidic C-termini. To further support a role of CCP1 in gene-encoded C-terminal processing, we searched for reports describing the natural occurrence of the C-terminal modifications shown in Table 1. In the case of α-tubulin, a pool of Δ2-tubulin is known to be naturally present in the cell. Richter-Cook *et al.* described the purification of rabbit eIF4H from rabbit reticulocyte lysates and identified a form of the protein with a more basic pI and postulated that this protein form might result from proteolysis of its C-terminal glutamate (Richter-Cook *et al.*, PMID 9516461). Rusconi *et al.* analyzed by MALDI-TOF-MS the C-terminus of chicken telokin and found forms of the protein from which one to six C-terminal glutamate residues had been removed (Rusconi *et al.*, PMID 9283094). Furthermore, we searched for evidence of CCP1 being the responsible carboxypeptidase *in vivo* for some of these C-terminal processing events. Berezniuk *et al.* performed CCP1 knock-down experiments in HEK293T cells and observed that Δ2-tubulin levels were significantly decreased (Berezniuk *et al.*, PMID 22170066). Comparative Western blot analysis of wild type and *pcd* mouse show that MLCK1 and telokin have intact C-termini in the absence of CCP1 (*i.e.* in *pcd* mouse), but processed C-termini in wild type mice (Rogowski *et al.*, PMID 21074048).

HMGB proteins can provide clues to the functional implications of CCP1-mediated C-terminal processing, given that comprehensive research is done on how the lack of their acidic C-terminal tails affects their function. HMGBs are non-histone proteins that are involved in various nuclear chromatin-associated processes such as transcription, replication, recombination, DNA repair and genomic stability. Mammalian HMGBs are composed of two central DNA binding domains (HMG-boxes A and B) and a negatively charged C-terminal region, which is highly conserved among different species (Fig. 2d). The acidic tail of HMGB1 and HMGB2 negatively affects binding to both linear and supercoiled DNA mediated by intra-molecular interactions of the acidic tail with the HMGB boxes and thereby shielding these domains from other interactions. However, these acidic tails seem to modulate the biological functions of HMGBs in many other ways. For instance, they are involved in the nucleocytoplasmic transport of HMGB1 and HMGB2 and affect their localization. In addition, the C-terminal tail of HMGB1 is crucial for transcription stimulation, essentially by its interaction with histones H1 and H3. Furthermore, the HMGB1 C-terminal tail was reported to play a major role in DNA repair, chromatin remodeling, and DNA replication. Marintcheva *et al.*

built a model to explain the role of the acidic tail of the ssDNA-binding protein of bacteriophage T7, which can be extended to other proteins such as ribosomal proteins, translation factors and HMG proteins (Marintcheva *et al.*, PMID 18238893). According to this model, the acidic tail of these proteins would regulate the affinity to their binding partners (DNA or other proteins) by binding and thereby shedding their basic binding clefts. In this context, CCP1, by shortening the length of these acidic tails would regulate (the avidity of) protein-protein and DNA-protein interactions. In fact, this is in line with the proposed role for PTMs affecting the C-terminus of tubulin; modulation of motor proteins and MAPs binding to the C-terminal region of tubulin. Similarly, Rusconi *et al.* proposed that the observed heterogeneity of telokin's C-terminus might regulate its interaction with other proteins (Rusconi *et al.*, PMID 9283094).

The functions of the here identified putative CCP1 substrates are quite diverse (Table 1), including microtubule-related proteins and proteins involved in transcription and chromatin remodeling. The latter could help to understand, for instance, the progressive transcriptional silencing or the large-scale reorganization of chromatin in *pcd* mice. However, further work is needed to fully validate these proteins as *in vivo* CCP1 substrates. Studies in the *pcd* knockout mouse might provide the ultimate validation of these substrates. Future research should additionally assess the possible implications of C-terminal processing of these proteins on their function.

Search of interacting partners of CCPs using Virotrap

Virotrap relies on the HIV-1 p55 GAG protein to trap a bait protein together with its physiological binding partner(s) into virus-like particles (VLPs). These VLPs are budded from human cells, as such enabling the native isolation of protein complexes followed by their subsequent analysis by mass spectrometry. When considering protease/substrate interactions they may be of transient nature, and thus the Virotrap technology seems well suited to search for CCP substrates, given its ability to detect weaker interactors. We tested in Virotrap two different CCPs (*i.e.* CCP2 and CCP3) that were chosen because they were already cloned in a compatible vector, ensuring straightforward subcloning into the Virotrap vector. Two independent tests showed that this technology is not suitable for the study of the interaction partners of CCPs, due to the very low expression levels observed for the GAG-CCP2/3 fusion protein. This low expression levels might be explained by the high molecular weights of these proteins (72 and 101 kDa, for CCP3 and CCP2 respectively), making them difficult to express recombinantly and probably interfering with the correct formation of the VLPs.

To potentially overcome this difficulty we cloned the armadillo domain present in the N-terminus of CCP1, expecting a better performance in Virotrap as a result of their lower molecular weight and increased solubility. However, in this case, again the GAG fusion protein showed low expression levels.

4. Perspectives for future collaboration between units (1 page)

The study of the proteolytic enzymes from *Trypanosoma cruzi* has been for many years one of the main research topics in the Laboratory of Professor Juan José Cazzulo, at the Instituto de Investigaciones Biotecnologicas in Argentina (home institution). Future collaboration between the units will focus in the determination of natural substrate-repertoire enzymes in the parasite as well as the *in vitro* substrate-specificity profile of some of these proteases (some of them carboxypeptidases) using proteomic tools like the ones described above and other methods available in the host institution. While the former should provide insights into the function of these enzymes in the parasite, the latter will be useful for the design of new inhibitors which may turn into leading compounds for the development of new drugs against the Chagas disease.

In particular, three members of the subfamily of cytosolic carboxypeptidases are present in *Trypanosoma cruzi*. This opens the possibility to extend some of the studies here performed for human CCPs to the enzymes of the parasite, and determine if these enzymes participate in similar protein post-translational modifications (so far not described in *Trypanosoma cruzi*).

The degradomics tools available in the host laboratory will also be used to search for natural substrates of an M32 carboxypeptidase from *Trypanosoma brucei*, a protozoan parasite that is the etiological agent of African sleeping sickness. A KO parasite in this carboxypeptidase shows reduced growth kinetics when compared to the wt parasite. Further, the observed parasitemia in infected mice is also reduced for the KO parasite.

5. Valorisation/Diffusion (including Publications, Conferences, Seminars, Missions abroad...

PUBLICATIONS

S. Tanco, O. Tort, H. Demol, FX Aviles, K. Gevaert, P. Van Damme, Julia Lorenzo. C-terminomics screen for natural substrates of cytosolic carboxypeptidase 1 reveals processing of acidic protein C-termini. Mol Cell Proteomics. (2015) 14, 177-190

S. Tanco, K. Gevaert, P. Van Damme. C-terminomics: targeted analysis of natural and post-translationally modified protein and peptide C-termini. Proteomics. (2014) 15, 903-914

CONFERENCES

VIB seminar 2014. 28-29 April 2014, Blankenberge, Belgium. Positional proteomics identification of natural substrates of cytosolic carboxypeptidases (CCPs). S. Tanco, O. Tort, FX Aviles, J Lorenzo, K Gevaert, P Van Damme.

6. Skills/Added value transferred to home institution abroad (1/2 page)

One of the main added values transferred to the home institute is the knowhow in degradomics techniques and validation skills acquired by the researcher. Nowadays, one of the critical tasks in the field of protease research is the elucidation of protease function, which is intrinsically linked to the determination of the degradome of the protease. This knowledge in degradomics includes C-terminal COFRADIC, a positional proteomics approach developed in the host laboratory that enables the study of processing at protein C-terminus. Additionally, this knowhow together with the experience gathered from applying other degradomics technologies can be applied to study the natural substrate repertoires of endoproteases. Finally, the analysis of proteomics data implicates handling large and complex data sets. The researcher brings to the home institution a thorough knowledge of the strategies and bioinformatics tools that allow interpreting these large data sets.