Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

Assessing proteolytic events in bioinformatic reanalysis of public secretome data from melanoma cell lines

Murilo Salardani, Uilla Barcick, André Zelanis*

Functional Proteomics Laboratory, Institute of Science and Technology, Federal University of São Paulo, UNIFESP, São José dos Campos, SP, 12231-280, Brazil

ABSTRACT

Autocrine and paracrine signals are of paramount importance in both normal and oncogenic events and the composition of such secreted molecular signals (i.e the secretome) designate the communication status of cells. In this context, the analysis of post-translational modifications in secreted proteins may unravel biological circuits regulated by irreversible modifications such as proteolytic processing. In the present study, we have performed a bioinformatic reanalysis of public proteomics data on melanoma cell line secretomes, changing database searching parameters to allow for the identification of proteolytic events generated by active proteases. Such approach enabled the identification of proteolytic signatures which suggested active proteases and whose expression profiles might be targeted in patient tissues or liquid biopsies, as well as their cleaved substrates. Although N-terminomics approaches continue to be the method of choice for the evaluation of proteolytic signaling events in complex samples, the simple approach performed in this work resulted in the gain of biological insights derived from shotgun proteomics data.

1. Introduction

Melanoma is an aggressive skin cancer, with increasing annual number of cases worldwide [1,2]. The prognosis of melanoma is worse when neoplastic cells spread deeper from the primary tumor and invade neighboring sites, such as lymph nodes and blood vessels. Such a process of invasion and metastasis strongly relies on the activity of proteases which are secreted from tumoral and stromal cells [3,4]. Autocrine and paracrine signals are of paramount importance in both normal and oncogenic events and the composition of such secreted molecular signals (i.e the secretome) designates the communication status of cells. In this context, the analysis of post-translational modifications in secreted proteins may unravel biological circuits regulated by irreversible modifications such as proteolytic processing [5–7]. Unlike protein degradation and turnover, where proteins are broken down into amino acids, proteolytic protein processing is an accurate signaling event, generating distinct protein species, which may modulate the activity of other proteins such as those involved in inflammation, immune response, cell cycle, cardiovascular diseases, blood clotting disorders and so on [5,6, 8]. Given the pervasiveness of proteolytic processes in signaling circuits related to cancer development and progression, the identification and annotation of protease cleavage sites as well as their respective substrates may reveal biological insights on the events responsible by melanoma progression. In this regard, publicly available proteomics data on cancer secretomes may encompass processed substrates whose identity have not been uncover until the proteomics searching parameters are not changed to allow for semi-specific cleavages. Such uninterrogated data source may reveal biological processes regulated by proteases as well as new substrates whose proteolytic status might be targeted for therapeutic interventions as well as a progression status.

2. Material and methods

2.1. Secretome data source

Data regarding secretomes obtained from different human melanoma metastatic cell lines (A375 and SH4) and a Cancer-associated Fibroblast (Cell line Hs895T), were downloaded from ProteomeXchange consortium website (http://www.proteomexchange.org/; for details on proteomeXchange files, see the Supplementary material section). Data from biochemically treated cells (i.e the treatment of tumoral cells with any chemical compound) were not considered.

* Corresponding author. Department of Science and Technology, Federal University of São Paulo (ICT-UNIFESP), Rua Talim, 330, 12231-280, São José dos Campos, SP, Brazil.
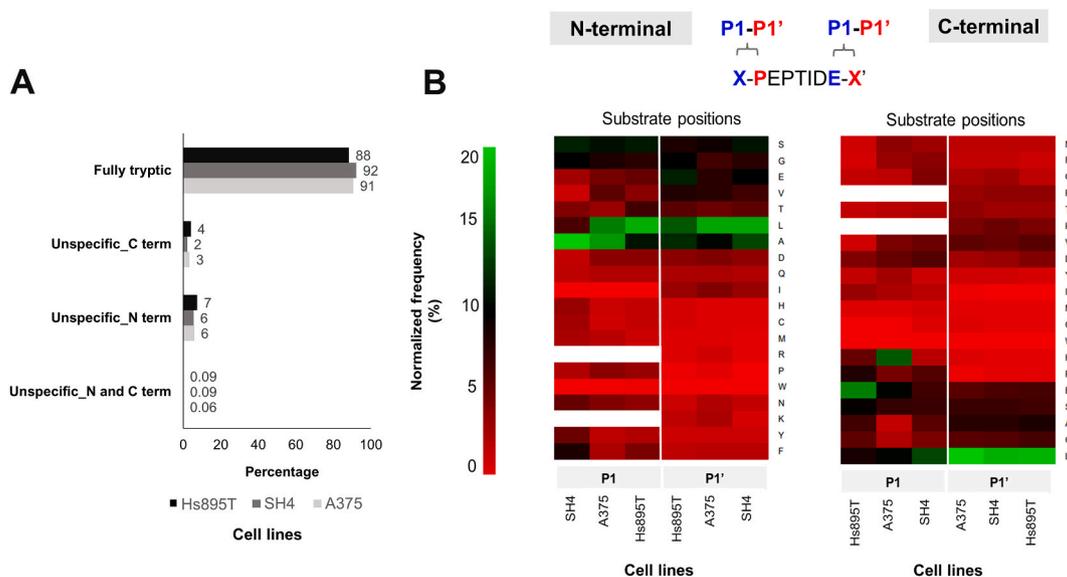*E-mail address:* andre.zelanis@unifesp.br (A. Zelanis).

Fig. 1. (A) Percentage of peptide identifications in all cell line secretomes. Fully tryptic peptides, peptides preceded by the primary specificity of trypsin (Unspecific C_term), peptides in which the trypsin specificity was restricted to C-terminus (Unspecific N_term), and peptides with no trypsin specificity (Unspecific_N and C term) are shown; **(B)** heat map showing the normalized amino acid frequency at the scissile bond (P1–P1′) for both N-terminal (left) and C-terminal (right) peptide positions in all cell line secretomes. Amino acid frequency was normalized over their natural abundance in *Homo sapiens*, according to the data available at the UniProt database (https://www.uniprot.org/).

2.2. Proteomics data analysis

Database searching of the LC-MS/MS spectra was carried out in Trans-Proteomic Pipeline (TPP) [9,10] using Comet search engine (version 2017.01 rev. 1) [11,12] restricted to *Homo sapiens* Uni-Prot/SwissProt database (UniProt/SwissProt release 2021_04; 42, 377entries; with 20,381 canonical entries and 21,996 isoforms). Considering trypsin as the searching enzyme, semi-specific searches were performed with no constraints at N- or C- terminus and at least two missed cleavages were allowed. Cysteine carbamidomethylation was selected as fixed modification whereas methionine oxidation and glutamine/asparagine deamidation were selected as variable modifications. In essence, we strictly followed the parameters specified in each original publication. Briefly, for spectra analyzed in the so-called 'high-low' mode, peptide identification was based on a search with an initial mass deviation of the precursor ion of 20 ppm and the fragment mass tolerance was set to 0.5 Da. For spectra acquired in 'high-high' mode, the fragment tolerance was set to 0.02 Da instead. Protein/Peptide identifications were accepted after estimating the False Discovery Rate calculated based on the score distributions in the output of the Comet search engine. Search results were filtered with PeptideProphet to a > 99% confidence interval, corresponding to a False Discovery Rate (FDR) of less than 1% at the peptide level (Supplementary Tables S1–S6). For each identified peptide, we mapped both the P1/P1′ positions at both N- and C- peptide terminus and the protease(s) potentially involved in the generation of such a cleavage. Cleavage site and mRNA expression analyses were represented as heat maps using the 'gplot' package in R [13]. For cleavage site analysis, the frequency of amino acids at the scissile bond (P1 and P1′ positions) were evaluated for both N- and C-termini cleavage sites. In addition, amino acid frequencies were normalized by their natural abundance in *Homo sapiens*. mRNA expression profile of selected proteases was carried out using whole exome sequencing data from pre-treatment tumors from 121 patients with metastatic melanoma, available at cBioPortal for cancer genomics (https://www.cbioportal.org; [14]). Selected proteases/substrates were analyzed in Reactome (version 75) (https://reactome.org; with the 'Analyse gene list' tool and Uniprot accession list) [15] to identify enriched biological pathways. All the settings were set as default.

3. Results and discussion

The semi-specific searches allowed for the identification (at 1% FDR) of 3,840, 1,547 and 842 proteins in A375, SH4 and Hs895T cell line secretomes, respectively. The number of identified proteins/peptides was higher in A375 cell line secretome, as this cell line had the highest number of RAW files available at ProteomeXchange repository. Unique peptide identifications also varied among the samples analyzed; A375 cell line secretome displayed the highest number (43,354), followed by SH4 cell line secretome (6,225 unique peptides) and Hs895T (2,925 unique peptides). These values are higher than reported in the original publications, a likely (and expected) consequence of the semi-specific search, in which proteins are identified by peptides that are not fully tryptic in addition to the tryptic ones, therefore increasing the overall number of protein/peptide identifications. Such peptides derived from semi-specific searches may have been generated by activated proteases and otherwise would not be identified using full enzyme specificity. Regardless of cell line secretome, as expected, most identified peptides (~90%) were derived from trypsin activity (Fig. 1A; Supplementary Tables S1–S3). This is an expected feature since secretome samples were subjected to trypsin activity. A similar percentage of semi-specific peptides was found in all cell lines, including peptides in which the trypsin specificity was restricted to C-terminus (Unspecific N-terminus), or peptides preceded by the primary specificity of trypsin (Unspecific C-terminus). A small number (< 1%) of peptides with no trypsin specificity (N/C terminus) was also observed (unspecific N/C-terminus). This latter set of peptides likely corresponds to earlier proteolytic events (i.e occurred before or just after secretome harvesting, and prior to trypsin addition). As observed in our analysis, semi-specific/unspecific peptides comprised a small amount of the total set of identified peptides (~10%); therefore, reinforcing the need for N-termini enrichment using more robust analytical approaches such as Terminal Amine Isotopic Labeling of Substrates (TAILS; [16]) or COmbined FRActional DIagonal Chromatography (COFRADIC; [17]), for example. Such enrichment methods should be the preferred choice for the investigation of N-terminomes in complex samples. However, reanalyzing shotgun proteomics data with slight alterations in searching parameters provided information on proteolytic processing events in samples that were not processed for such a purpose. Proteolytic processing events are indeed present in such
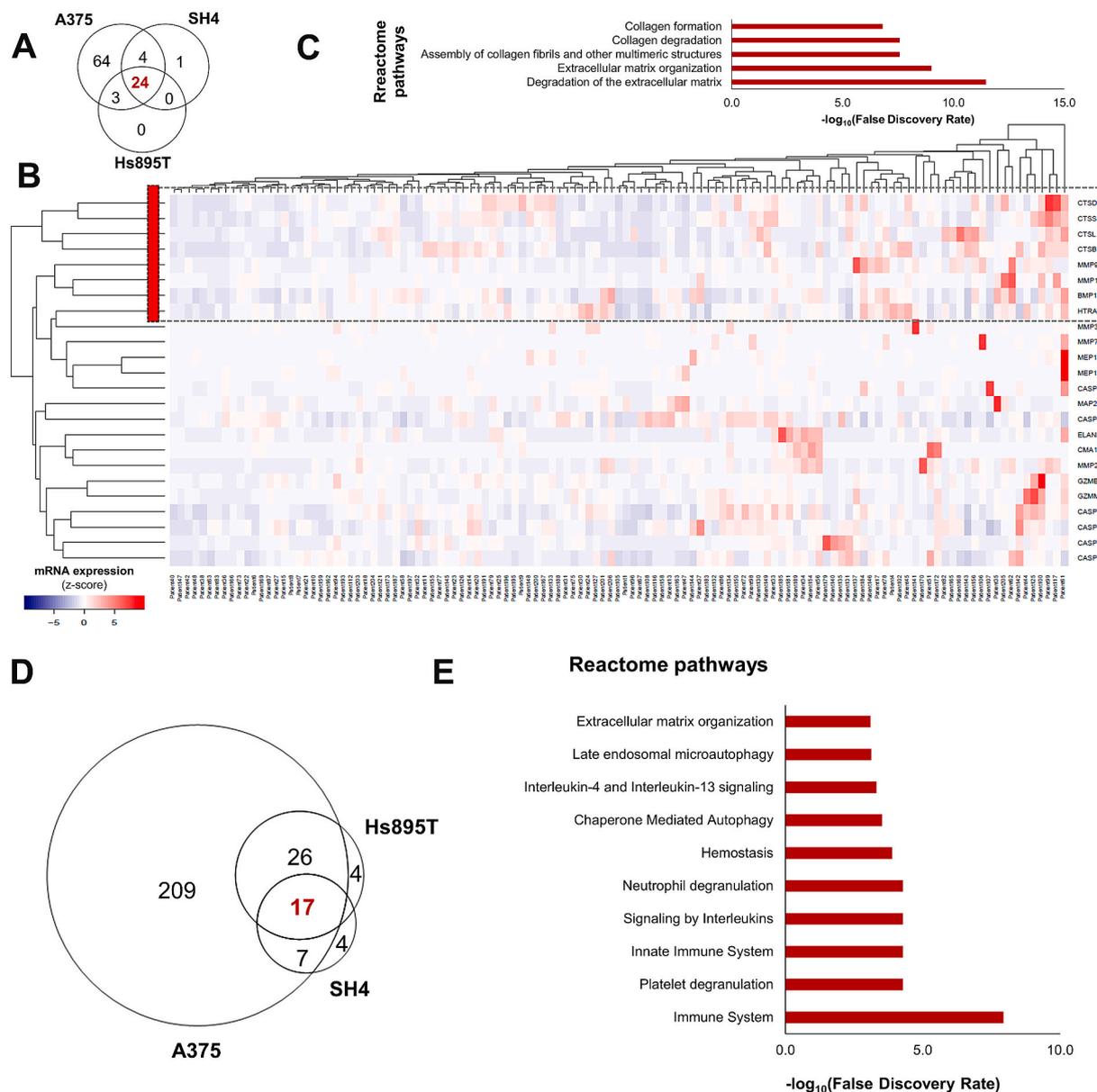
Fig. 2. Semi specific cleavage sites suggested active proteases in all cell line secretomes. **(A)** Venn diagram showing potentially active proteases in each cell line secretome according to data obtained in topFIND database (https://topfind.clip.msl.ubc.ca/topfinder) (B) heat map showing the normalized mRNA expression values (z-scores) of the 24 proteases (inferred from topFIND database analysis) derived from whole exome sequencing from pre-treatment tumors from 121 patients with metastatic melanoma (data retrieved from cBioPortal for cancer genomics https://www.cbioportal.org; [14]); the cluster of similar expression values for the proteases CTSD, CTSS, CTSL, CTSB, MMP9, MMP11, BMP1 and HTRA2 is highlighted (in red) and its enrichment (Reactome) pathways are presented in **(C)**. **(D)** Venn diagram showing, proteins (potential substrates; Supplementary Tables S4–S6) that were identified by the semi-specific/unspecific peptide set. **(E)** Functional enrichment analysis (reactome pathways analysis; https://reactome.org/PathwayBrowser/#TOOL=AT) of the 17 shared substrates. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

trypsin-digested proteomes, however, unless searching parameters are changed, these features will remain unexplored.

We used all semi-specific/unspecific set of peptides to map amino acid frequencies at the N- and C-terminal peptide scissile bonds and such an analysis revealed a striking recurrence of Leucine (Leu) and Alanine (Ala) at the N-terminal P1 and P1′ positions, respectively, (Fig. 1B) and almost exclusively Leu at P1′ C-terminal position, regardless the cell line secretome. The preference of Leu at P1′ is a 'signature' of matrix metalloproteases specificity, including MMP2 and MMP9 [18]. Indeed, such metalloproteases are highly expressed in melanoma during radial and vertical growth phases [4]. We also subjected the set of semi-specific peptides to a cleavage site database analysis (topFIND database [19]) (Supplementary Tables S4–S6), which allowed their mapping to

previously reported cleavage sites, therefore, suggesting active proteases likely responsible for such cleavage events. Accordingly, the mapping of cleavage sites to protease activity suggested 24 active proteases, in all cell line secretomes, as observed in Fig. 2A. We next examined the expression mRNA profile of such proteases using whole exome sequencing data from pre-treatment tumors from 121 patients with metastatic melanoma, available at cBioPortal for cancer genomics (https://www.cbioportal.org; [14]). Interestingly, among the 24 proteases, 8 of them (cathepsins D, S, L and B, MMP9, MMP11, BMP1 and HTRA2) clustered in a group of similar expression values among metastatic melanoma patients (Fig. 2B). More importantly, this set of proteases actively participate in biological processes related to cell migration and invasion, such as extracellular matrix degradation, as

revealed by the pathway enrichment analysis (Fig. 2C). In addition, not all the 24 proteases were identified in cell line secretomes (Supplementary Tables S1–S3) indicating that, although their expression levels might be low, the identification of their cleavage products may be regarded as 'footprints' of their activities. On the other hand, proteins that were identified by the semi-specific/unspecific peptide set could be regarded as potential substrates of active proteases in the secretome samples (Supplementary Tables S4–S6). In this context, 17 proteins were identified in the secretome all samples (Fig. 2D). Basement membrane-specific heparan sulfate proteoglycan core protein, perlecan (UniProt ac# P98160), proteases such as procathepsin L (P07711) and cathepsin B (P07858) and Vascular Endothelium Growth Factor A (P15692) were identified among the shared substrates. Functional analysis revealed an enrichment in key processes that are known to be regulated by protease activity, such as extracellular matrix organization, signaling by interleukins, autophagy, among others (Fig. 2E).

In conclusion, by reanalyzing public proteomics data with slight adjustments in searching parameters (allowing semi-specific cleavages), we were able to identify proteolytic signatures which suggested active proteases and whose expression profiles might be targeted in patient tissues or liquid biopsies as well as their cleaved substrates. Such features might be of prognostic value for melanoma patients, mainly concerning the metastatic potential of this cancer. Finally, although N-terminomics approaches continue to be the method of choice for the evaluation of proteolytic signaling events in complex samples, the simple approach used in this work resulted in the gain of biological insights derived from shotgun proteomics data.

Author contributions

M.S, and U.B analyzed the data, and all authors wrote and reviewed the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2022.101259.

References

[1] Global Cancer Observatory, (n.d.).

[2] D. Holmes, The cancer that rises with the sun, Nature 515 (2014) S110–S111, https://doi.org/10.1038/515S110a.

[3] K. Kessenbrock, V. Plaks, Z. Werb, Matrix metalloproteinases: regulators of the tumor microenvironment, Cell 141 (2010) 52–67, https://doi.org/10.1016/j.cell.2010.03.015.

[4] N. Moro, C. Mauch, P. Zigrino, Metalloproteinases in melanoma, Eur. J. Cell Biol. 93 (2014) 23–29, https://doi.org/10.1016/j.ejcb.2014.01.002.

[5] C.M. Overall, C.P. Blobel, In search of partners: linking extracellular proteases to substrates, Nat. Rev. Mol. Cell Biol. 8 (2007) 245–257, https://doi.org/10.1038/nrm2120.

[6] B. Turk, D. Turk, V. Turk, Protease signalling: the cutting edge, EMBO J. 31 (2012) 1630–1643, https://doi.org/10.1038/emboj.2012.42.

[7] N. Fortelny, J.H. Cox, R. Kappelhoff, A.E. Starr, P.F. Lange, P. Pavlidis, C.M. Overall, Network analyses reveal pervasive functional regulation between proteases in the human protease web, PLoS Biol. 12 (2014), https://doi.org/10.1371/journal.pbio.1001869.

[8] P.F. Lange, C.M. Overall, Protein TAILS: when termini tell tales of proteolysis and function, Curr. Opin. Chem. Biol. 17 (2013) 73–82, https://doi.org/10.1016/j.cbpa.2012.11.025.

[9] E.W. Deutsch, L. Mendoza, D. Shteynberg, T. Farrah, H. Lam, Z. Sun, E. Nilsson, B. Pratt, B. Prazen, J.K. Eng, B. Daniel, A. Nesvizhskii, R. Aebersold, A Guided Tour of the TPP 10 (2011) 1150–1159, https://doi.org/10.1002/pmic.200900375.A.

[10] E.W. Deutsch, L. Mendoza, D. Shteynberg, J. Slagel, Z. Sun, R.L. Moritz, Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics, Proteomics -, Clin. Appl. 9 (2015) 745–754, https://doi.org/10.1002/prca.201400164.

[11] J.K. Eng, T.A. Jahan, M.R. Hoopmann, Comet: an open-source MS/MS sequence database search tool, Proteomics 13 (2013) 22–24, https://doi.org/10.1002/pmic.201200439.

[12] J.K. Eng, M.R. Hoopmann, T.A. Jahan, J.D. Egertson, W.S. Noble, M.J. MacCoss, A deeper look into Comet - implementation and features, J. Am. Soc. Mass Spectrom. 26 (2015) 1865–1874, https://doi.org/10.1007/s13361-015-1179-x.

[13] R. Ihaka, R. Gentleman, R: a language for data analysis and graphics, J. Comput. Graph Stat. 5 (1996) 299–314, https://doi.org/10.1080/10618600.1996.10474713.

[14] D. Liu, B. Schilling, D. Liu, A. Sucker, E. Livingstone, L. Jerby-Amon, L. Zimmer, R. Gutzmer, I. Satzger, C. Loquai, S. Grabbe, N. Vokes, C.A. Margolis, J. Conway, M.X. He, H. Elmarakeby, F. Dietlein, D. Miao, A. Tracy, H. Gogas, S.M. Goldinger, J. Utikal, C. Blank, R. Rauschenberg, D. von Bubnoff, A. Krackhardt, B. Weide, S. Haferkamp, F. Kiecker, B. Izar, L. Garraway, A. Regev, K. Flaherty, A. Paschen, E.M. Van Allen, D. Schadendorf, Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma, Nat. Med. 2512 (25) (2019) 1916–1927, https://doi.org/10.1038/s41591-019-0654-5, 2019.

[15] B. Jassal, L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, R. Haw, F. Loney, B. May, M. Milacic, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The reactome pathway knowledgebase, Nucleic Acids Res. 48 (2020) D498–D503, https://doi.org/10.1093/nar/gkz1031.

[16] O. Kleifeld, A. Doucet, A. Prudova, U. Auf Dem Keller, M. Gioia, J.N. Kizhakkedathu, C.M. Overall, Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates, Nat. Protoc. 6 (2011) 1578–1611, https://doi.org/10.1038/nprot.2011.382.

[17] K. Gevaert, M. Goethals, L. Martens, J. Van Damme, A. Staes, G.R. Thomas, J. Vandekerckhove, Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides, Nat. Biotechnol. 21 (2003) 566–569, https://doi.org/10.1038/NBT810.

[18] U. Eckhard, P.F. Huesgen, O. Schilling, C.L. Bellac, G.S. Butler, J.H. Cox, A. Dufour, V. Goebeler, R. Kappelhoff, U. auf dem Keller, T. Klein, P.F. Lange, G. Marino, C.J. Morrison, A. Prudova, D. Rodriguez, A.E. Starr, Y. Wang, C.M. Overall, Active site specificity profiling of the matrix metalloproteinase family: proteomic identification of 4300 cleavage sites by nine MMPs explored with structural and synthetic peptide cleavage analyses, Matrix Biol. 49 (2016) 37–60, https://doi.org/10.1016/j.matbio.2015.09.003.

[19] N. Fortelny, S. Yang, P. Pavlidis, P.F. Lange, C.M. Overall, Proteome TopFIND 3.0 with TopFINDer and PathFINDer: database and analysis tools for the association of protein termini to pre- and post-translational events, Nucleic Acids Res. 43 (2015) D290–D297, https://doi.org/10.1093/nar/gku1012.