

Scientific paper

Characterization of Legumain Degradome Confirms Narrow Cleavage Specificity

Robert Vidmar,¹ Matej Vizovišek,¹ Dušan Turk,^{1,2} Boris Turk,^{1,3}
and Marko Fonovič^{1,*}

¹ Jožef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology, Jamova 39,
SI-1000 Ljubljana, Slovenia

² Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Ljubljana, Slovenia, Jamova 39,
SI-1000 Ljubljana, Slovenia

³ Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia

* Corresponding author: E-mail: marko.fonovic@ijs.si
Phone: +386 1 477 3474

Received: 08-02-2018

Dedicated to the memory of Prof. Dr. Igor Kregar

Abstract

Legumain or asparagine endopeptidase is a unique cysteine endopeptidase with a distinctive specificity for the hydrolysis of peptide bonds after asparagine and to a lesser extent after aspartate. It is ubiquitously expressed in various tissues and besides its involvement in immune response and other physiological processes, it was also shown to play a role in pathological states such as inflammation and cancer. In order to improve our understanding of legumain substrate recognition we have performed proteomic profiling of legumain specificity on native proteins derived from MDA-MB-231 cells using two different N-terminal labelling methodologies (FPPS and ISIL). Our data revealed narrow cleavage specificity for P1 position combined with clear cleavage preference for unstructured secondary regions in the substrate proteins. No extended cleavage specificity on native proteins was observed. Moreover, a limited number of identified cleavages on individual substrates suggest its primary role in precision proteolysis and regulatory proteolytic events.

Keywords: Proteolysis; proteases; degradome; cleavage specificity; legumain

1. Introduction

Legumain or asparagine endopeptidase (AEP) is a lysosomal cysteine protease (C13 family, EC number 3.4.22.34) that catalyses the hydrolysis of peptide bonds after asparagines and to a lower extent also after aspartates.^{1–2} The catalytic site of legumain is composed of a catalytic dyad (His148 and Cys189). Consistent with its lysosomal localization, legumain has a pH optimum for substrate hydrolysis at acidic pH (pH = 5.8). In mammals, legumain is expressed in various organs and tissues, most abundantly in kidneys, testis and antigen-presenting cells. In contrast to other lysosomal proteases such as cathepsins with a high degree of functional redundancy, legumain exhibits unique substrate specificity and evolutionary highly conserved primary structure, which suggests that legu-

main may be involved in specific physiological processes. However, the exact physiological role of legumain is still not well understood. Although it was reported to regulate immune response through the activation of TLR receptors and antigen presentation it has not been shown to be crucial in these processes.^{3–5} In pathological conditions legumain is strongly associated with tumorigenesis, where its overexpression was shown in a number of human solid tumours such as carcinomas of the breast, colon and prostate.^{6–7} Recent reports have also shown its possible involvement in atherosclerosis⁸ and ischemic brain injury.⁹ Legumain has been also linked to the development of neuropathology in Alzheimer's disease, where it cleaves amyloid precursor protein and tau protein.^{10–11}

Nevertheless, the pathophysiological functions of legumain, its substrates and its association with disease de-

velopment and progression remain poorly understood. Conservation of legumain between evolutionary divergent plants and animals suggest that there is a biological need for legumain that has yet to be identified. A better understanding of features that govern the legumain substrate recognition is therefore expected to provide a novel insight in its biological role. Processing of legumain substrates *in vitro* was usually performed on peptides or denatured proteins and although these studies revealed its primary amino acid preference, it did not account for the possible influence of secondary and tertiary structure of the substrate.^{12–13} In this work, we studied legumain processing on a complex pool of structurally native proteins. We treated native human proteome with recombinant legumain and used in-gel stable isotope labelling (ISIL) and fast profiling of protease specificity (FPPS) for the identification of substrate cleavages. This enabled identification of primary cleavage specificity of legumain on native proteins as well as determination of secondary structure cleavage preference.

2. Experimental

2. 1. Materials

Antibodies against human prelamina-A/C (ab8984) were purchased from Abcam, United Kingdom. Recombinant human legumain was expressed in the baculovirus expression system as previously described.¹³ Ac(D3)-NHS was synthesized at the Faculty of Pharmacy (University of Ljubljana, Slovenia) according to the established protocol.¹⁴

2. 2. Cells and Cell Lysate Preparation

MDA-MB-231 breast cancer cells were routinely maintained in DMEM medium (Lonza) supplemented with 10% FBS (Sigma-Aldrich), 2 mM glutamine (Gibco) and 1% antibiotic stock solution (10.000 U/ml penicilin and 10 mg/ml streptomycin) (Sigma-Aldrich) at 37 °C and 5% CO₂. Before the harvesting, the cells were grown to confluency and washed with DPBS (Lonza). The cells were detached with enzyme-free dissociation solution (Millipore). After the centrifugation, the cell pellet was lysed on ice for 15 min in 20 mM sodium phosphate buffer pH 6.0, containing 150 mM NaCl, 0.05% NP-40, 1 mM EDTA, 25 μM E-64, 1 mM PMSF. The insoluble material was removed by centrifugation at 14 000 g for 10 min and the cleared cell lysate was used to determine protein concentration with the Bradford assay and portioned into aliquots containing 0.5 mg of total protein.

2. 3. *In vitro* Processing of the Cell Lysate

Recombinant prolegumain was first activated in 50 mM citric buffer pH 4.0, supplemented with 5 mM DTT

for 30 min at 37 °C. The active concentration of legumain was determined to be 16.7 μM.¹² In our degradomic workflow, recombinant human legumain was added to each aliquot of the cell lysate at 0.2 μM and 1.0 μM final concentration and the samples were incubated at 37 °C for 1h before further processing.

Immunological detection of prelamina-A/C was performed under identical conditions except that additional time-points were used for the *in vitro* processing (0, 10, 30 and 60 min). Western blot analysis was performed on nitrocellulose membrane using mouse monoclonal antibodies against prelamina-A/C according to manufacturer's recommendations.

2. 4. N-terminal Labelling with Trideutero-Acetylation

2. 4. 1. In-Gel Stable Isotope Labelling Protocol

After the *in vitro* processing, the samples were labelled with Ac(D3)-NHS according to the ISIL protocol described earlier with some modifications.¹⁵ Briefly, the samples were incubated in 6x SDS-PAGE loading buffer at 95 °C for 5 min and separated on a 12.5% SDS-PAGE gel (Lonza). The gel was stained with Coomassie brilliant blue and each of the protein lanes was cut into six bands and destained with 25 mM NH₄HCO₃ in 50% acetonitrile/dH₂O. The proteins were reduced with 10 mM DTT (56 °C, 45 min) and alkylated with 55 mM iodoacetamide in the dark at room temperature for 30 min. In-gel stable isotope labelling was performed with an addition of 1 mg Ac(D3)-NHS per sample (40 mM final concentration). Prior to the labelling step Ac(D3)-NHS was dissolved in 100 mM phosphate buffer, pH 8.5. The reaction was performed at 30 °C for 1h and the labelling step was repeated one more time. To reverse any potential threonine and serine O-acetylation 1 mg hydroxylamine was added to each sample and the samples were incubated at room temperature for 20 min. The excess Ac(D3)-NHS reagent was quenched by adding 1 mg of glycine per sample followed by an incubation at 30 °C for 1h. The gel pieces were washed with acetonitrile and vacuum dried before rehydrating in 80 μl 25 mM NH₄HCO₃ solution containing 1 μg of sequencing-grade modified porcine trypsin per sample, and the trypsinization was then performed over night at 37 °C. The next day the peptides were extracted from the gel using the extraction solution (50% acetonitrile, 5% formic acid). The samples were desalted before LC-MS/MS analysis on C18 tips as described elsewhere.¹²

2. 4. 2. Fast-Profiling of Protease Specificity Protocol

The samples after *in vitro* processing were prepared according to the FPPS protocol described elsewhere.¹⁶ Briefly, the samples were transferred to a 500 μl micro-fil-

ter device with a cut-off of 3000 Da (Millipore) and the buffer was exchanged with 100 mM phosphate buffer pH 8.5. Subsequently, 2 mg of Ac(D3)-NHS reagent was dissolved in the sample followed by 1 hour incubation at 30 °C. Afterwards, the labelling step was repeated. To reverse partial labelling of serines, threonines and tyrosines, hydroxylamine was added to the sample at 1 mM concentration and left at room temperature for 20 min. Afterwards, 8 M urea was added and proteins were reduced with 10 mM DTT for 1 hour at room temperature before addition of iodoacetamide at 50 mM final concentration for 1 hour in the dark at room temperature. After the free cysteines were alkylated, the unreacted iodoacetamide was quenched with 50 mM DTT for 30 min at room temperature before the buffer was exchanged with 25 mM ammonium bicarbonate pH 7.8. The sample volume was set to 250 μ l before the overnight trypsinization at a 1:100 (w/w, enzyme/substrate) ratio at 30 °C. The peptide-rich flow-through was collected the next day by spinning the micro-filters in the centrifuge and concentrated to 50 μ l. The peptide samples were fractionated using a SAX-C18 stage tip protocol as previously described.^{17–18} Accordingly, the samples were mixed with Britton & Robinson buffer (20 mM acetic acid, 20 mM phosphoric acid and 20 mM boric acid, pH 11) and the pH was set to 11 with 1 M NaOH. The SAX tips were prepared by stacking 6 discs of Empore/Disk Anion Exchange (Varian) in a 200 μ l pipet tip (Eppendorf) and the C18 tips were prepared by stacking 4 discs of Empore/C18 (Varian). The samples were applied to the SAX-C18 tip and the peptide fractions were eluted using buffers at pH 11, 10, 9, 8, 7, 6, 5, 4 and 3. The eluting peptides were captured on C18 tips and subjected to LC-MS/MS analysis.

2. 5. Mass Spectrometric Analysis

The LC-MS/MS analysis of the samples was performed using an Orbitrap LTQ Velos mass spectrometer (Thermo Fischer Scientific) coupled to an EASY-nanoLC II HPLC (Thermo Fischer Scientific) operated automatically via XCalibur software (Thermo Fischer Scientific). The samples containing 0.1% FA were loaded onto a C18 trapping column (Proxeon Easy-column, Thermo Fischer Scientific) and separated on a C18 PicoFrit Aquasil analytical column (New Objective). The peptides were eluted using a 5–40% (v/v) 50 min linear gradient of acetonitrile with 0.1% FA at a constant flow rate of 300 nl/min. The full MS mass spectra were acquired with the Orbitrap mass analyzer in the mass range of 300 to 2,000 m/z at resolution of 30,000 in the profile mode. The MS/MS spectra were obtained by HCD fragmentation of the nine most intense MS precursor ions and recorded at a resolution of 7,500 in the centroid mode. Only the precursor ions with assigned charge states (> 1) were chosen for MS/MS fragmentation. The dynamic exclusion was set to repeat count of 1, repeat duration of 30 s, and exclusion duration of 20 s.

2. 6. Database Search and Data Processing

For the identification of peptides we used the MaxQuant proteomic software¹⁹ and performed the database searches against the human proteome deposited in the UniProt/Swiss-Prot database (UniProtKB, *Homo sapiens*, canonical database, 20 336 entries). The settings for the database searches applied trideutero-acetylation of peptide N-termini (+45.029 Da) and methionine oxidation (+15.995 Da) as variable modifications, trideutero-acetylation of lysines (+45.029 Da) and carbamidomethylation of cysteines (+57.021 Da) as fixed modifications, Semi-ArgC/P as the enzyme specificity setting while allowing for one missed cleavage, precursor ion and fragment ion mass tolerances were set to 20 ppm and 0.5 Da, respectively. A reversed database search was performed and the false discovery rate (FDR) was set at 1% for peptide and protein identifications. Raw data and database search files are available via ProteomeXchange with identifier PXD010466.

After the database searches, the identified modification specific peptides were filtered to obtain the true positive hits for the legumain cleavage sites. Peptides with trideutero-acetylated N-termini identified only in the legumain-treated samples were considered to be the result of legumain cleavages. Reverse and contaminant peptides were removed from the peptide list. Additionally, the peptides were filtered for posterior error probability (PEP values above 0.05 were discarded) and score (values below 40 were discarded) as described previously.²⁰ The P1'–P4' positions were determined from the peptide N-terminus, whilst the P4–P1 positions were determined bioinformatically. The iceLogo representations were generated using the frequencies of positional amino acid occurrences normalised to the natural amino acid abundances in the human Swiss-Prot database.²¹

2. 7. Analysis of Structural Determinants of Legumain Substrates

To analyse the structural determinants of legumain substrate recognition we first prepared a list of the most reliable legumain substrates. Accordingly, we analysed the identified cleavages with ISIL and FPPS protocol and constructed Venn diagrams to identify the overlapping substrates and cleavages (identified in both experiments). These substrates were selected for further analysis to assign the identified cleavages to the secondary structure of each substrate using PSIPRED.^{22–24}

3. Results and Discussion

3. 1. Determination of Legumain Cleavage Specificity

For specificity profiling of the native proteome, we prepared protein lysates under mild lysis conditions to

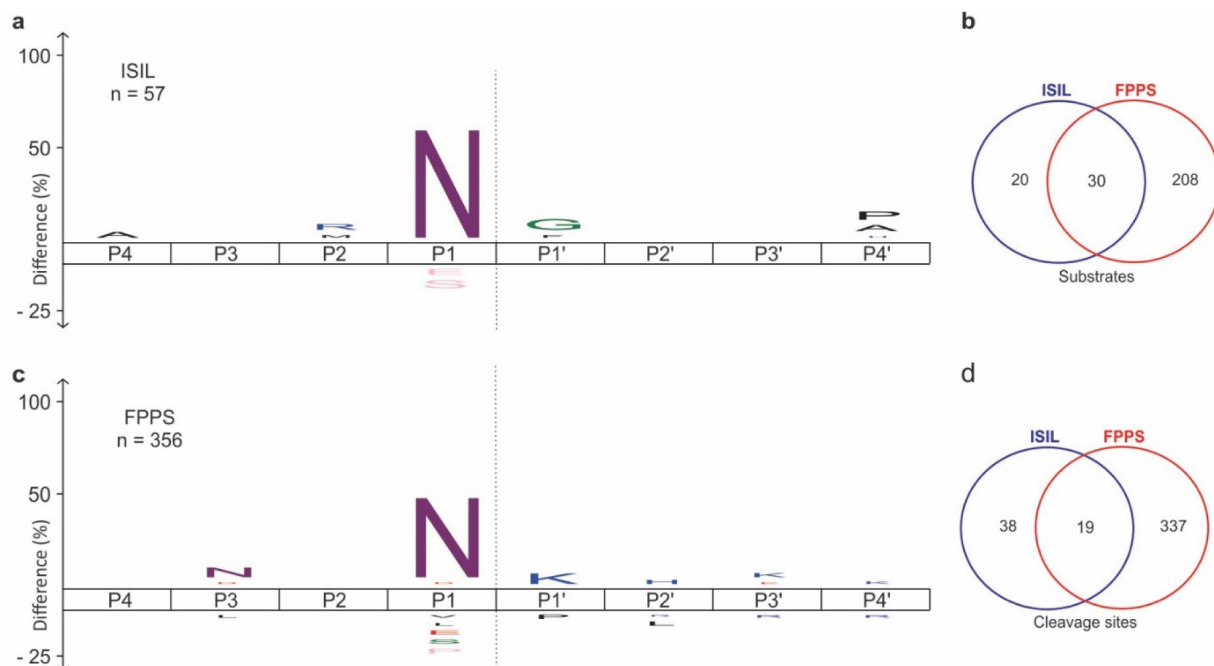


Figure 1. Legumain cleavage specificity profiling. (a) ISIL and (c) FPPS specificity profile of legumain for P4 to P4' positions presented as iceLogo²¹ with the representative number of determined cleavages used for the iceLogo construction (n-value). Venn diagrams of unique and shared legumain (b) substrates and (d) cleavage sites between ISIL and FPPS approaches.

preserve the protein fold. During the lysis, we inhibited proteolytic background with general protease inhibitors in order to minimise potential false positive hits as a result of endogenous proteases. After the treatment of the native proteome with recombinant human legumain we applied two different approaches for chemical labelling of neo-N-terminal peptides in order to identify the corresponding cleavage events (Fig 1). Using ISIL we identified 57 cleavage sites in 50 proteins, FPPS enabled us to identify 356 cleavages in 238 legumain substrates (Fig. 1b, c). This difference can be attributed to a better performance of the in-solution approach and the extensive peptide fractionation using the anion exchange resin.

Our results confirmed that P1 asparagine is the main determinant of legumain substrate specificity and that even cleavages of native protein substrates showed no extended cleavage specificity. Both tested approaches resulted in similar cleavage specificity profiles with a strong preference for asparagine in the P1 position (ISIL 68.4% and FPPS 52.2% of cleavages, respectively) and less prominently for the P1 aspartate (ISIL 8.8% and FPPS 9.0% of cleavages, respectively). These results are consistent with the canonical specificity of legumain reported in the CutDB²⁵ and PMAP²⁶ databases, which is based on 15 proteolytic events deposited in both databases. The cleavage specificity determined is also in a good agreement with previously reported profiling experiments performed on the denatured protein samples, where primary structure was shown to be the leading factor of protease-substrate recognition.¹² In both experimental setups the specificity for P1 asparagine

was highly similar under native (52.2–68.4%) as well as under denaturing conditions (85%). This led us to the conclusion that regardless of the native/denatured state of the substrates, legumain has a relatively stringent requirement for an Asn in the P1 position. Such observations were reported also by studies using combinatorial peptide substrate libraries which concluded that the P1 position is crucial for recognition of legumain substrates, while other positions did not have a significant role.¹³

3. 2. Validation of Substrate Processing in the Case of Prelamin-A/C

Our specificity profiling of legumain showed a considerable overlap of identified cleavages between the ISIL and FPPS approach (Fig. 1d). Among the 19 overlapping cleavage events (33% of the total ISIL-determined cleavage sites) we selected the cleavage after Asn283 in human prelamina-A/C (Uniprot code P02545, gene LMNA) for further validation. Prelamin-A/C is an important constituent of the nuclear lamina that provides framework for nuclear envelope. Maturation of prelamina-A/C involves several steps, including farnesylation of C-terminal – CAAX motif and subsequent proteolytic maturation by zinc metalloprotease ZMPSTE24.²⁷ Failure of prelamina-A/C proteolytic maturation due to protease absence or mutations was shown to result in its accumulation and consequently to development of several diseases.²⁸

We identified prelamina-A/C cleavage after Asn283 using both labelling approaches. The incubation with in-

Table 1. The ion intensity of prelamina-A/C peptide AcD3-LVGAAHEELQQR determined by ISIL and FPPS approach.

Approach	PEP	Score	Ion intensity (*10 ⁶ a.u.)		
			Control	0.2 μM legumain	1.0 μM legumain
ISIL	1.28E-16	265.61	0.0	19.3	31.1
FPPS	6.21E-12	164.33	0.0	14.5	43.5

creased concentration of legumain led to an increased MS peak intensity of trideutero-acetylated peptide LVGAAHEELQQR, as shown in Table 1. This additionally confirmed that generation of the labeled peptide was directly related to the legumain activity.

To verify these results, we next performed immunological detection of prelamina-A/C processing (Figure 2). We confirmed that legumain processed prelamina-A/C at a single site, since the intensity of the single band corresponding to the cleaved prelamina-A/C fragment increased

with incubation time. The theoretical molecular weight of prelamina-A/C fragments is 32.5 kDa (N-terminal fragment, amino acid sequence 1–283) and 39.7 kDa (C-terminal fragment, amino acid sequence 284–664), respectively. Since the antibodies used in this assay recognized the epitope between residues 319 and 566, we concluded that the band observed at a molecular weight around 40 kDa corresponded to the C-terminal cleavage product (284–664) of prelamina A/C. Identification of prelamina-A/C as legumain substrate is based on *in vitro* experiments

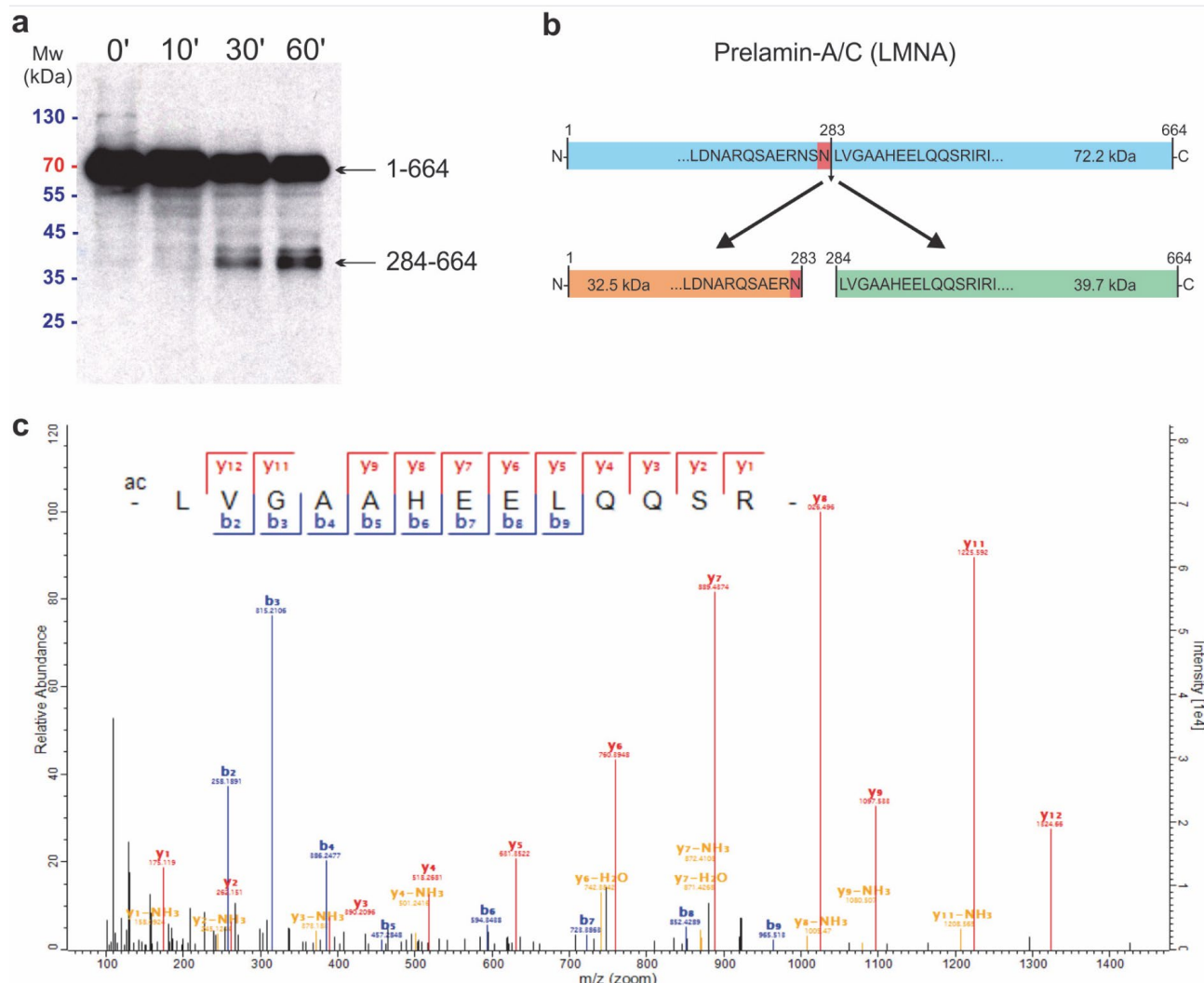


Figure 2. Legumain cleavage of prelamina-A/C. (a) Schematic representation of prelamina-A/C cleavage with legumain. (b) Immunological detection of prelamina-A/C processing with legumain. (c) MS/MS spectrum of prelamina-A/C peptide AcD3-LVGAAHEELQQR. y and b ions are shown in red and blue, respectively.

that investigate sequence and structure specific features of legumain substrates. Further *in vivo* studies are therefore required to determine possible physiological relevance of identified cleavages and since legumain has been shown to be localized in the nucleus²⁹ there is a potential spatial link present in living cells.

3. 3. Structural Preferences for Legumain Cleavages

In addition to the sequence-based substrate specificity interpretation of the legumain substrate repertoire we also investigated the structural features of the cleaved sub-

Table 2. Structural determinants of 19 identified legumain cleavage sites determined using disordered regions and secondary structure prediction algorithms.

Gene name	Uniprot code	neo-N-terminal peptide sequence	Cleavage site	2D structure at cleavage site	3D structure at cleavage site
ACTN4	O43707	GTLEDQLSHLKQYER	N680	unstructured	n.a.
FLNB	O75369	ETSSILVESVTR	N2483	unstructured	2EEC
CSDE1	O75534	IMLLKKKQAR	N175	α -helical	2YTX
PNP	P00491	STVPGHAGR	R58	unstructured	1RSZ
LMNA	P02545	LVGAAHEELQQR	N283	α -helical	n.a.
NPM1	P06748	DENEHQLSLR	N35	unstructured	n.a.
HSP90AB1	P08238	ASDALDKIR	N46	unstructured	3NMQ
HNRNPL	P14866	YDDPHKTPASPVVHIR	N91	unstructured	n.a.
EEF1D	P29692	EEEDKEAAQLR	N164	unstructured	2MVM
RPL12	P30050	EIKVVYLR	N8	unstructured	n.a.
RPL22	P35268	LGGGVVTIER	N55	unstructured	n.a.
RPS16	P62249	GRPLEMIEPR	N35	unstructured	n.a.
RPL7A	P62424	FGIGQDIQPKR	N38	unstructured	n.a.
EIF5A	P63241	GFVVLKGR	N28	unstructured	3CPF
ACTG1	P63261	SYVGDEAQSQR	D51	unstructured	n.a.
TUBA1B	P68363	AAIATIKTKR	N329	α -helical	n.a.
PCBP1	Q15365	STAASRPPVTLR	N89	unstructured	n.a.
CNN3	Q15417	KIASKYDHQAEDLR	N18	unstructured	n.a.
RPL36	Q9Y3U8	KGHKVTKNVSKPR	N12	unstructured	n.a.

n.a. – not available

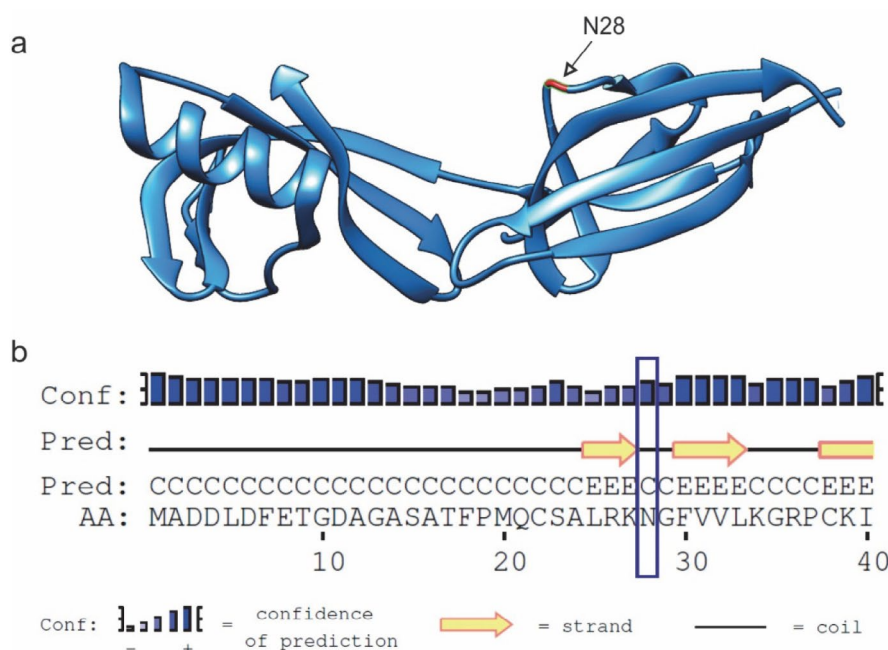


Figure 3. Example of a legumain cleavage site in a 3D protein structure. (a) The structure of eukaryotic translation initiation factor 5A-1 with annotated scissile site at asparagine 28 (For preparing this image the structure with PDB 3CPF was used). (b) An example of secondary structure prediction algorithm output for the region around Asn28 in eukaryotic translation initiation factor 5A-1 (marked in blue).

strates. We performed a detailed structural analysis on 19 substrates identified by both ISIL and FPPS approaches (Table 2, Appendix 1) and combined the exact cleavage site information to the secondary structural characteristics of each substrate using a 2D structure and disordered region prediction tool (i.e. PSIPRED protein sequence analysis workbench).²² The close inspection of individual secondary structure features has shown that the large majority of legumain cleavage sites are not located in the structured regions like alpha-helices or beta-sheets (84%). Since the unstructured elements are devoid of regular secondary structures they often serve as flexible linkers or loops connecting the organized secondary structure elements. Our observation that unstructured elements are more easily accessible to legumain proteolysis is in good agreement with the central proteolytic paradigm stating that most cleavages occur in easily accessible protein regions. Moreover, similar observations were also reported by other studies showing that largest number of proteolytic cleavages occurs in the loops, followed by alpha-helices and beta-sheets.^{30–33}

Among the identified substrates, 6 have known 3D structures of the cleavage site regions and we were able to confirm our observation, that cleavage sites are mainly located on unstructured regions characteristic for loops between protein domains or exposed tails near protein termini (Supplementary Figures 1–5). For example, in the crystal structure of eukaryotic translation initiation factor 5A-1 legumain cleaved the protein after an asparagine located in the unstructured region connecting two neighbouring beta-sheets (Fig 3), consistent with our predictions. For a more precise analysis a larger pool of cleavage sites and 3D structures would be required.

4. Conclusions

Determination of protease specificity provides the most basic information about protease-substrate interactions and helps us to understand why certain proteins are cleaved or processed by proteases. Whether a substrate cleavage will occur depends on several factors: (i) subsite cleavage specificity of a protease that recognizes an amino acid motif in a protein substrate, (ii) structural arrangement of a potential cleavage site(s) in a substrate (iii), presence of potential protease's exosites located outside the active site cleft and (iv), spatiotemporal co-localization of a protease and a corresponding substrate. Our study was focused on the first two factors and it showed that P1 asparagine (or less preferably aspartate) is the most important determinant of legumain specificity even in the processing of native proteins. The identified protein cleavages also revealed that legumain has a high preference for cleavages outside secondary structure elements, since only a minor portion of cleavages was found in the alpha-helical regions and none in the beta-sheets. The low number of

identified cleavage events in individual substrates showed that legumain is a highly selective protease, which promotes limited proteolysis rather than general protein degradation. Within the lysosomes, such limited proteolytic processing could serve as the first step in protein degradation, where legumain cleavage would increase exposure of proteins to other proteases. However, since legumain is also known to be secreted to extracellular space and translocated to the cell nucleus, its limited proteolysis could play an important role in cell signalling and other highly specific cellular processes.

5. Acknowledgments

The work was funded by the Slovenian Research Agency (P1-0140, J1-3602, and N1-0022 to B.T.; J1-0185 and J1-5449 to M.F.; and I0-0048 to D.T.) and the COST Action PROTEOSTASIS BM1307. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE³⁴ partner repository with the dataset identifier PXD010466.

6. References

- Dall, E.; Brandstetter, H., *Proc Natl Acad Sci* **2013**, *110* (27), 10940–5. DOI:10.1073/pnas.1300686110
- Dall, E.; Brandstetter, H., *Biochimie* **2016**, *122*, 126–50. DOI:10.1016/j.biochi.2015.09.022
- Sepulveda, F. E.; Maschalidi, S.; Colisson, R.; Heslop, L.; Ghirelli, C.; Sakka, E.; Lennon-Dumenil, A. M.; Amigorena, S.; Cabanie, L.; Manoury, B., *Immunity* **2009**, *31* (5), 737–48. DOI:10.1016/j.immuni.2009.09.013
- Maschalidi, S.; Hassler, S.; Blanc, F.; Sepulveda, F. E.; Tohme, M.; Chignard, M.; van Endert, P.; Si-Tahar, M.; Descamps, D.; Manoury, B., *PLoS Pathog* **2012**, *8* (8), e1002841. DOI:10.1371/journal.ppat.1002841
- Descamps, D.; Le Gars, M.; Balloy, V.; Barbier, D.; Maschalidi, S.; Tohme, M.; Chignard, M.; Ramphal, R.; Manoury, B.; Salenave, J. M., *Proc Natl Acad Sci U S A* **2012**, *109* (5), 1619–24. DOI:10.1073/pnas.1108464109
- Liu, C.; Sun, C.; Huang, H.; Janda, K.; Edgington, T., *Cancer Res* **2003**, *63* (11), 2957–64.
- Luo, Y.; Zhou, H.; Krueger, J.; Kaplan, C.; Lee, S. H.; Dolman, C.; Markowitz, D.; Wu, W.; Liu, C.; Reisfeld, R. A.; Xiang, R., *J Clin Invest* **2006**, *116* (8), 2132–2141. DOI:10.1172/JCI27648
- Lunde, N. N.; Holm, S.; Dahl, T. B.; Elyouncha, I.; Sporsheim, B.; Gregersen, I.; Abbas, A.; Skjelland, M.; Espevik, T.; Solberg, R.; Johansen, H. T.; Halvorsen, B., *Atherosclerosis* **2017**, *257*, 216–223. DOI:10.1016/j.atherosclerosis.2016.11.026
- Ishizaki, T.; Erickson, A.; Kuric, E.; Shamloo, M.; Hara-Nishimura, I.; Inacio, A. R.; Wieloch, T.; Ruscher, K., *J Cereb Blood Flow Metab* **2010**, *30* (10), 1756–66. DOI:10.1038/jcbfm.2010.39

10. Zhang, Z.; Song, M.; Liu, X.; Kang, S. S.; Kwon, I. S.; Duong, D. M.; Seyfried, N. T.; Hu, W. T.; Liu, Z.; Wang, J. Z.; Cheng, L.; Sun, Y. E.; Yu, S. P.; Levey, A. L.; Ye, K., *Nat Med* **2014**, *20* (11), 1254–62. DOI:10.1038/nm.3700
11. Zhang, Z.; Song, M.; Liu, X.; Su Kang, S.; Duong, D. M.; Seyfried, N. T.; Cao, X.; Cheng, L.; Sun, Y. E.; Ping Yu, S.; Jia, J.; Levey, A. L.; Ye, K., *Nat Commun* **2015**, *6*, 8762. DOI:10.1038/ncomms9762
12. Vidmar, R.; Vizovisek, M.; Turk, D.; Turk, B.; Fonovic, M., *EMBO J* **2017**, *36* (16), 2455–2465. DOI:10.15252/embj.201796750
13. Poreba, M.; Solberg, R.; Rut, W.; Lunde, N. N.; Kasperkiewicz, P.; Snipas, S. J.; Mihelic, M.; Turk, D.; Turk, B.; Salvesen, G. S.; Drag, M., *Cell Chem Biol* **2016**, *23* (8), 1023–35.
14. Staes, A.; Impens, F.; Van Damme, P.; Ruttens, B.; Goethals, M.; Demol, H.; Timmerman, E.; Vandekerckhove, J.; Gevaert, K., *Nat Protoc* **2011**, *6* (8), 1130–41. DOI:10.1038/nprot.2011.355
15. Asara, J. M.; Zhang, X.; Zheng, B.; Christofk, H. H.; Wu, N.; Cantley, L. C., *J Proteome Res* **2006**, *5* (1), 155–63. DOI:10.1021/pr050334t
16. Vizovisek, M.; Vidmar, R.; Van Quickenberghe, E.; Impens, F.; Andjelkovic, U.; Sobotic, B.; Stoka, V.; Gevaert, K.; Turk, B.; Fonovic, M., *Proteomics* **2015**, *15* (14), 2479–90. DOI:10.1002/pmic.201400460
17. Wisniewski, J. R.; Zougman, A.; Mann, M., *J Proteome Res* **2009**, *8* (12), 5674–8. DOI:10.1021/pr900748n
18. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., *Nat Methods* **2009**, *6* (5), 359–62. DOI:10.1038/nmeth.1322
19. Cox, J.; Mann, M., *Nat Biotechnol* **2008**, *26* (12), 1367–72. DOI:10.1038/nbt.1511
20. Vizovisek, M.; Vidmar, R.; Fonovic, M., *Methods Mol Biol* **2017**, *1574*, 183–195. DOI:10.1007/978-1-4939-6850-3_13
21. Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K., *Nat Methods* **2009**, *6* (11), 786–7. DOI:10.1038/nmeth1109-786
22. Buchan, D. W.; Minneci, F.; Nugent, T. C.; Bryson, K.; Jones, D. T., *Nucleic Acids Res* **2013**, *41*, W349–57. DOI:10.1093/nar/gkt381
23. McGuffin, L. J.; Bryson, K.; Jones, D. T., *Bioinformatics* **2000**, *16* (4), 404–5. DOI:10.1093/bioinformatics/16.4.404
24. Jones, D. T.; Cozzetto, D., *Bioinformatics* **2015**, *31* (6), 857–63. DOI:10.1093/bioinformatics/btu744
25. Igarashi, Y.; Eroshkin, A.; Gramatikova, S.; Gramatikoff, K.; Zhang, Y.; Smith, J. W.; Osterman, A. L.; Godzik, A., *Nucleic Acids Res* **2007**, *35* (Database issue), D546–9. DOI:10.1093/nar/gkl813
26. Igarashi, Y.; Heureux, E.; Doctor, K. S.; Talwar, P.; Gramatikova, S.; Gramatikoff, K.; Zhang, Y.; Blinov, M.; Ibragimova, S. S.; Boyd, S.; Ratnikov, B.; Cieplak, P.; Godzik, A.; Smith, J. W.; Osterman, A. L.; Eroshkin, A. M., *Nucleic Acids Res* **2009**, *37* (Database issue), D611–8. DOI:10.1093/nar/gkn683
27. Casasola, A.; Scalzo, D.; Nandakumar, V.; Halow, J.; Recillas-Targa, F.; Groudine, M.; Rincon-Arango, H., *Nucleus* **2016**, *7* (1), 84–102. DOI:10.1080/19491034.2016.1150397
28. Rankin, J.; Ellard, S., *Clin Genet* **2006**, *70* (4), 261–74. DOI:10.1111/j.1399-0004.2006.00677.x
29. Haugen, M. H.; Johansen, H. T.; Pettersen, S. J.; Solberg, R.; Brix, K.; Flatmark, K.; Maelandsmo, G. M., *PLoS One* **2013**, *8* (1), e52980. DOI:10.1371/journal.pone.0052980
30. Mahrus, S.; Trinidad, J. C.; Barkan, D. T.; Sali, A.; Burlingame, A. L.; Wells, J. A., *Cell* **2008**, *134* (5), 866–76. DOI:10.1016/j.cell.2008.08.012
31. Timmer, J. C.; Zhu, W.; Pop, C.; Regan, T.; Snipas, S. J.; Eroshkin, A. M.; Riedl, S. J.; Salvesen, G. S., *Nat Struct Mol Biol* **2009**, *16* (10), 1101–8. DOI:10.1038/nsmb.1668
32. Kazanov, M. D.; Igarashi, Y.; Eroshkin, A. M.; Cieplak, P.; Ratnikov, B.; Zhang, Y.; Li, Z.; Godzik, A.; Osterman, A. L.; Smith, J. W., *J Proteome Res* **2011**, *10* (8), 3642–51. DOI:10.1021/pr200271w
33. Ratnikov, B. I.; Cieplak, P.; Gramatikoff, K.; Pierce, J.; Eroshkin, A.; Igarashi, Y.; Kazanov, M.; Sun, Q.; Godzik, A.; Osterman, A.; Stec, B.; Strongin, A.; Smith, J. W., *Proc Natl Acad Sci* **2014**, *111* (40), E4148–55. DOI:10.1073/pnas.1406134111
34. Vizcaino, J. A.; Csordas, A.; Del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; Xu, Q. W.; Wang, R.; Hermjakob, H., *Nucleic Acids Res* **2016**, *44* (22), 11033. DOI:10.1093/nar/gkw880

Povzetek

Legumain je po svoji specifičnosti edinstvena lizosomalna proteaza, ki cepi C-terminalno od asparagina in v redkih primerih tudi za aspartatom. V eksperimentalnem delu smo izvedli profiliranje specifičnosti legumaina s proteomskim pristopom na osnovi kemijskega označevanja N-koncev z uporabo trideuteroacetilacije. Raziskava predstavlja prvo globalno analizo nativnih legumainskih substratov pri čemer smo potrdili visoko specifičnost legumaina za cepitev peptidne vezi za asparaginom. Na osnovi proteomskih rezultatov smo z bioinformatično analizo podatkov raziskovali tudi povezavo med identificiranimi proteolitičnimi cepitvami ter sekundarno in tridimenzionalno strukturo substratov legumaina.