PHOSPHOPROTEOMIC ANALYSIS OF LEGUMAIN DEFICIENT MICE

Tilen Sever

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Assoc. Prof. Dr. Marko Fonović, Jožef Stefan Institute, Ljubljana, Slovenia

Evaluation Board:
Prof. DDr. Boris Turk, Chair, Jozef Stefan International Postgraduate School and Jožef Stefan Institute, Ljubljana, Slovenia
Prof. Dr. Maša Vodovnik, Member, Biotechnical Faculty, Univeristy of Ljubljana, Ljubljana, Slovenia
Prof. Dr. Janko Kos, Member, Jožef Stefan Institute and Faculty of Pharmacy, University of Ljubljana, Ljubljana, Slovenia

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL



Tilen Sever

PHOSPHOPROTEOMIC ANALISYS OF LEGUMAIN DEFICIENT MICE

Doctoral Dissertation

FOSFOPROTEOMSKA ANALIZA MIŠI Z IZBITIM GENOM ZA LEGUMAIN

Doktorska disertacija

Supervisor: Assoc. Prof. Dr. Marko Fonović

Ljubljana, Slovenia, July 2026

The cake is a lie.

Acknowledgments

I would like to thank Slovenian Research and Innovation agency for funding the presented work.

I would like to thank my supervisor Prof. Dr. Marko Fonović for guidance and advice through my work. I am also grateful for thorough review of this thesis.

I would like to Prof. DDr. Boris Turk for accepting me into his research group.

I would like to thank Dr. Robert Vidmar for help and advice for designing and preforming experimental work.

I would like to thank Tea Sinožić and Matej Kolarič for work with immunohistochemistry and Sara Fišer and Dr. Špela Baebler for mRNA sequencing.

I would like to thank our proteomics team for all the helpful and heated discussions.

I would like to thank all co-workers from our department B1 for help and support.

I express gratitude to all members of thesis evaluation board, for thoroughly reviewing this work.

Abstract

Protein phosphorylation is an important post-translational modification in cellular signal transmission. Perturbations in this mechanism can cause phenotypic abnormalities in organisms. Legumain, also known as asparagine-specific endopeptidase, is a cysteine proteinase found in plants and animals. Legumain regulates renal and haematopoietic homeostasis, as well as immunological responses. Its dysregulation is linked to a variety of malignancies and neurological disorders. Legumain knockout mice have a normal phenotype with the exception of impaired kidney function, hemophagocytic syndrome, and extramedullary haematopoiesis. In this work, we examined alterations in protein phosphorylation in legumain knock-out mice compared to wild-type controls to better understand how legumain deficiency impacts protein phosphorylation and associated cell signalling. Phosphopeptides from kidney and liver tissues were enriched and analysed using mass spectrometry, then verified using western blot and immunohistochemistry. We identified multiple differentially phosphorylated sites between wild type and legumain knock out mice. Among them, Y-box binding protein 1, an RNA and DNA binding protein was found to be differentially phosphorylated on three sites. In legumain-knockout mice kidneys as well as on HL-60 human cell line model of legumain knock out, a key location on the serine 100 residue of Y-box binding protein 1 activates the NF-κB pathway, leading to an increased inflammatory response. This was corroborated by increased expression of NF-κB genes and their downstream targets. Overall, this work sheds light on the role of legumain in mice and its effects on several cellular processes.

Povzetek

Fosforilacija proteinov je pomembna posttranslacijska modifikacija pri prenosu celičnih signalov. Motnje v tem mehanizmu lahko povzročijo fenotipske nepravilnosti pri zdravih organizmih. Legumain, znan tudi kot asparagin-specifična endopeptidaza, je cisteinska proteinaza, ki se nahaja v rastlinah in živalih. Legumain uravnava renalno in hematopoetsko homeostazo ter sodeluje pri imunskem odzivu organizma na patogene. Spremembe v njegovem delovanju so povezane z različnimi malignimi obolenji in nevrološkimi motnjami. Miši z izbitim genom za legumain imajo normalen fenotip, z izjemo oslabljene funkcije ledvic, hemofagocitnega sindroma in ekstramedularne hematopoeze. V tem delu smo preučevali spremembe v fosforilaciji proteinov pri miših z izbitim genom za legumain v primerjavi s kontrolami divjega tipa, da bi bolje razumeli, kako odsotnost legumaina vpliva na fosforilacijo proteinov in povezano celično signaliziranje. Fosfopeptidi iz ledvičnih in jetrnih tkiv so bili obogateni in analizirani z masno spektrometrijo, nato pa preverjeni s prenosom western in imunohistokemijo. Indentificirali smo več forsforiliranih mest. Med njimi so bila, Y-box vezavnem proteinu 1, ki veže RNA in DNA, odkrita tri fosforilirana mesta. Pri miših z izbitim genom za legumain in pri človeškem modelu, celični liniji HL-60, identificirano fosforilirano mesto na serinu 100 proteina Y-box aktivira NF-κB signalno pot, kar vodi v povečan vnetni odziv. To je bilo potrjeno s povečano ekspresijo NF-κB genov in njegovih tarč. To delo osvetljuje vlogo legumaina pri miših in njegovo vlogo pri celičnih in sistemskih procesih.

Contents

List of Figures **xvii**

Abbreviations **xxi**

1 Introduction **26**
1.1 Proteases 26
1.2 Cysteine proteases 26
1.3 Legumain 28
1.3.1 Classification and substrate specificity 28
1.3.2 Structure 29
1.3.3 Organism and tissue distribution 30
1.3.4 Sub-cellular location 32
1.3.5 Activity and forms 32
1.3.5.1 Asparagine endopeptidase (AEP) and asparaginyl-mono-carboxypeptidase (ACP) 32
1.3.5.2 Transcription factor, protein-protein interactions 35
1.3.5.3 Ligase activity 35
1.3.6 Regulation of activity 36
1.3.6.1 Inhibition 36
1.3.6.2 Activation 37
1.3.7 Physiological functions 37
1.3.7.1 Under normal conditions 37
1.3.7.2 Under pathological conditions 38
1.3.7.2.1 Cancer 38
1.3.7.2.2 Immune response 39
1.4 Protein phosphorylation 39
1.4.1 Function 39
1.4.2 Enzyme regulation 41
1.4.2.1 Protein kinases 41
1.4.2.2 Protein phosphatases 46
1.4.3 Methods for detection of protein phosphorylation sites 48
1.4.3.1 Mass spectrometry 48
1.4.4 Legumain and protein phosphorylation 50

2 Aims and Hypotheses **53**

3 Materials and Methods **55**
3.1 Materials 55
3.1.1 Chemicals 55
3.1.2 Commercial kits 55
3.1.3 Buffers, solutions and media 56

3.1.4 Plasmids 58
3.1.5 Antibodies 58
3.1.6 Primers 59
3.1.7 Mammalian cell lines 59
3.1.8 Equipment 59
3.2 Methods 60
3.2.1 Mice 60
3.2.1.1 Mice on the C57/BL6 background were maintained in accordance with national regulations authorized by the government's Ethical Committee and the Veterinary Administration of the Republic of Slovenia (VARS), with permit number U34401-40/2020/3. Legumain knockout mice were provided by the Thomas Reinheckl lab in Freiburg, Germany. Three mice each were kept in standard cages (Techniplast 1264C Eurostandard Type II) with bedding made of corncobs (Rehofix MK 2000). Access to food (Altromin 1324) and water was provided ad libitum. Kidneys and livers from both wild-type (WT) and AEP$^{-/-}$ (n= 3 for each genotype) mice were immediately frozen before further processing. Organ lysate preparation 60
3.2.2 Cell culture 61
3.2.2.1 Growing and passaging of cells 61
3.2.2.2 Legumain overexpression in HL-60 cells 61
3.2.2.3 Cell lysate preparation 61
3.2.3 Mass spectrometry 61
3.2.3.1 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) 61
3.2.3.1.1 In-gel digestion protocol 61
3.2.3.1.2 LC-MS/MS analysis 62
3.2.3.1.3 Data analysis 62
3.2.3.2 In-solution digestion protocol 63
3.2.3.3 Phosphoproteomic data analysis 64
3.2.4 Western transfer and immunoblotting 64
3.2.4.1 SDS-PAGE 64
3.2.4.2 Western blotting transfer and immunological detection 64
3.2.5 Real-Time quantitative polymerase chain reaction (RT-qPCR) 65
3.2.6 Immunohistology 65

4 Results 66
4.1 Phosphoproteomics analysis of legumain knock out mice kidney and liver 66
4.1.1 Phosphoproteomic analysis of mouse kidney 66
4.1.1.1 Prediction of subcellular location of identified phosphoproteins 67
4.1.1.2 Prediction of interacting kinases based on kinase recognition motifs 68
4.1.1.3 Gene Ontology (GO) term analysis 69
4.1.1.4 Protein interaction network analysis 70
4.1.2 Phosphoproteomic profile analysis of mouse liver 73
4.1.2.1 Prediction of subcellular location of identified phosphoproteins 76

4.1.2.2 Prediction of interacting kinases based on kinase recognition motifs....76
4.1.2.3 Gene Ontology (GO) term analysis....77
4.1.2.4 Protein interaction network analysis....79
4.2 Gene expression analysis....81
4.3 Immunohistochemistry....82

5 Discussion **86**

6 Conclusions **93**

Bibliography **114**

Biography **117**

List of Figures

Figure 1: Reaction mechanisms of peptide hydrolysis catalyzed by cysteine proteases 27
Figure 2: Schematic visualization of legumain structure 29
Figure 3: Legumain sequence from different organisms 31
Figure 4: Legumain activation 34
Figure 5: Legumain ligase activity 35
Figure 6: Catalytic mechanism of protein tyrosine phosphatases 48
Figure 7: Volcano plot of legumain kidney phosphoproteomic analysis. 67
Figure 8: Pie chart of subcellular locations of identified phosphoproteins in the kidneys.68
Figure 9: Motif analysis of identified phosphosites in kidneys 69
Figure 10: Gene onthology (GO) term analysis of phosphorylated proteins in kidneys....70
Figure 11: STRING network of phosphorylated proteins in kidneys 72
Figure 12: Volcano plot of legumain liver phosphoproteomic analysis 76
Figure 13: Pie chart of subcellular locations of identified phosphoproteins in the liver....76
Figure 14: Motif analysis of identified phosphosites in liver 77
Figure 15: Gene onthology (GO) term analysis of phosphorylated proteins in liver....78
Figure 16:STRING network of phosphorylated proteins in liver 81
Figure 17: RT-qPCR expression analysis of selected genes 81
Figure 18: mRNA seq expression analysis of selected genes 82
Figure 19: Western blot analysis of organ samples 83
Figure 20: Western blot analysis of lysates from HL-60 cells 84
Figure 21: Immunofluorescence staining of kidneys for macrophages 84
Figure 22: Immunofluorescence staining of kidneys for YBX1....85

Abbreviations

DNA	Deoxyribonucleic acid
ABC	Ammonium bicarbonate
ACN	Acetonitrile
ACP	Asparaginyl-mono-carboxypeptidase
AEP	Asparginyl endopeptidase
Akt	Protein kinase 4
AMPK	5' AMP-activated protein kinase
AOMK	Acyloxymethyl ketone
AP	Activation peptide
AP	Alkaline phosphatase
aPK	Atypical protein kinases
ATP	Adenosine triphosphate
BSA	Bovine serum albumine
C4,6S	Chondroitin 4,6-sulphate
C4S	Chondroitin 4-sulphate
C6S	Chondroitin 6-sulphate
CaMK	Ca2+/calmodulin-dependent protein kinase
CC	Cys-based class
CD	Cluster of differentiation
CDPK	Calcium-dependent protein kinases
CID	Collision-induced dissociation
CK	Casein kinase
CP	Cysteine protease
DSP	Dual specificity phosphatase
ECD	Electron-capture dissociation
ECM	Extracellular matrix
EGFR	Epidermal growth factor receptor
ePK	Eukaryotic protein kinases
ETD	Electron-transfer dissociation
FGF	Fibroblast growth factors
GAG	Glycosaminoglycans
GSK3	Glycogen synthase kinase-3 beta
GTP	Guanosine triphosphate
HAD	Haloacid dehalogenase
HCD	Higher-energy collisional dissociation

HEAT	Huntingtin-elongation-A subunit-TOR
HL-60	Human leukaemia 60
HP	Histidine phosphatase
HSP	Heat shock protein
JAK	Janus kinase
IR	Insulin receptor
IκB	I kappa B kinase
LEL	Legumain exosite loop
LSAM	Legumain stabilisation and activity modulation
MAPK	Mitogen-activated protein kinase
MAPKKK	Mitogen-activated protein kinase kinase kinase
MBP	Myelin basic protein
MHC	Major histocompatibility class
MLCK	Myosin light-chain kinase
mTOR	Mechanistic target of rapamycin
NFAT	Nuclear factor of activated T cells
$NF_{\kappa}B$	Nuclear factor kappa-light-chain-enhancer of activated B cells
PAPS	3'-phosphoadenosine- 5'-phosphosulphate
PCNA	Proliferating cell nuclear antigen
PDK	3-phosphoinositide-dependent kinase
pH	Power of hydrogen
PHP	Protein histidine phosphatase
PKA	Protein kinase A
PKC	Protein kinase C
PLK1	Polo-like kinase
PP2A	Protein phosphatase 2A
PPPLs	Phosphoprotein-like phosphatases
PTM	Post-translational
PTP	Protein tyrosine phosphatase
PVDF	Polyvinylidene difluoride
RAC	Ras-related C3 botulinum toxin substrate 1
RGD	Arginylglycylaspartic acid
RhK	Rho-associated protein kinase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH2	Src homology 2
STAT	Signal transducers and activators of transcription
STE	Homologs of yeast Sterile
TGFβ	Tumour growth factor beta
TK	Tyrosine kinase
TKL	Tyrosine kinase like
TLR	Toll-like receptors
TNF	Tumour necrosis factor
TRAF	TNF receptor-associated factors

VEP	Vacuolar processing enzyme
βARK	G-protein-coupled receptor kinase 2

Standard Amino Acid Abbreviations

Amino acid	Three-letter abbreviation	One-letter abbreviation`
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Phenylalanine	Phe	F
Glycine	Gly	G
Glutamine	Gln	Q
Glutamic acid	Glu	E
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Proline	Pro	P
Serine	Ser	S
Tyrosine	Tyr	Y
Threonine	Thr	T
Tryptophan	Trp	W
Valine	Val	V

Chapter 1

Introduction

1.1 Proteases

Proteases are enzymes that break down peptide bonds in proteins. They play a role in overall protein breakdown and also control many biological processes. These include managing protein location, activity, interactions with other proteins, and creating new bioactive molecules. They help in cellular information processing and assist in generating, transferring, and amplifying molecular signals. Proteases are classified into six catalytic classes; these are aspartic, cysteine, glutamic, metallo, asparagine, threonine, and serine proteases [1], [2].

Alternatively, proteases can be classified as endopeptidases if they recognize and cleave a specific amino acid in the middle of a peptide, or they can be categorized as exopeptidases if they cleave from either the N-terminus (aminopeptidases) or the C-terminus (carboxypeptidases) of the peptide [3].

Cell signalling involving specific and limited processing of certain protease substrates is most often an irreversible step in signal transduction, crucial in many biological processes such as DNA replication and transcription, cell proliferation and differentiation, tissue development and remodelling, angiogenesis, neurogenesis, ovulation, inflammation, immunity, and programmed cell death. Therefore, any alternations in protease activity or its regulation can lead to pathological conditions, including cardiovascular and inflammatory diseases, cancer, osteoporosis, and neurological disorders [4].

1.2 Cysteine proteases

Cysteine proteases, also known as thiol proteases, are a group of enzymes that are characterized by the presence of cysteine residues in their active domain. Papains and cathepsins are part of the largest cysteine protease family, and are described as an example of cysteine proteases here. These enzymes require both a nucleophile and a proton donor to function. For cysteine proteases, the proton donor is a histidine residue while cysteine residue acts as a nucleophile [5].

Cysteine proteases, especially cathepsins and papains, are highly conserved throughout evolution, with an optimal pH range on the acidic side (pH 4–6.5). They are classified into several families based on their primary amino acid sequences and secondary or tertiary structures. Clans consist of multiple families, which are believed to share a common evolutionary origin. There are ten clans of cysteine proteases: CA, CB, CC, CD, CE, CF, CL, CM, CN, and CO, collectively comprising 113 families (from C1 to C124). Each family is further divided into subfamilies based on similarities in the amino acid sequence at the enzyme's active site. Proteases within the same clan operate with comparable catalytic mechanisms, despite potentially lacking sequence

similarity and differing in their secondary or tertiary structure. The hydrolytic activity of cysteine proteases relies on the catalytic triad of Cys/His/Asn (or Asp). Often, glutamine (Gln) serves as the fourth catalytic residue, playing a crucial role in stabilizing the acyl intermediates formed during catalysis, taking precedence over the Cys active sites [6]. Essentially, the catalytic process occurs in two steps: first, acylation, where the thiolate anion performs a nucleophilic attack on the C1 carbonyl carbon (Figure 1) of the peptide bond, creating an intermediate. This intermediate is then deacylated through a general base-catalysed reaction, where the imidazole nitrogen from His (N2) activates a water molecule to attack the C1 atom of the substrate peptide, leading to the final products [7].

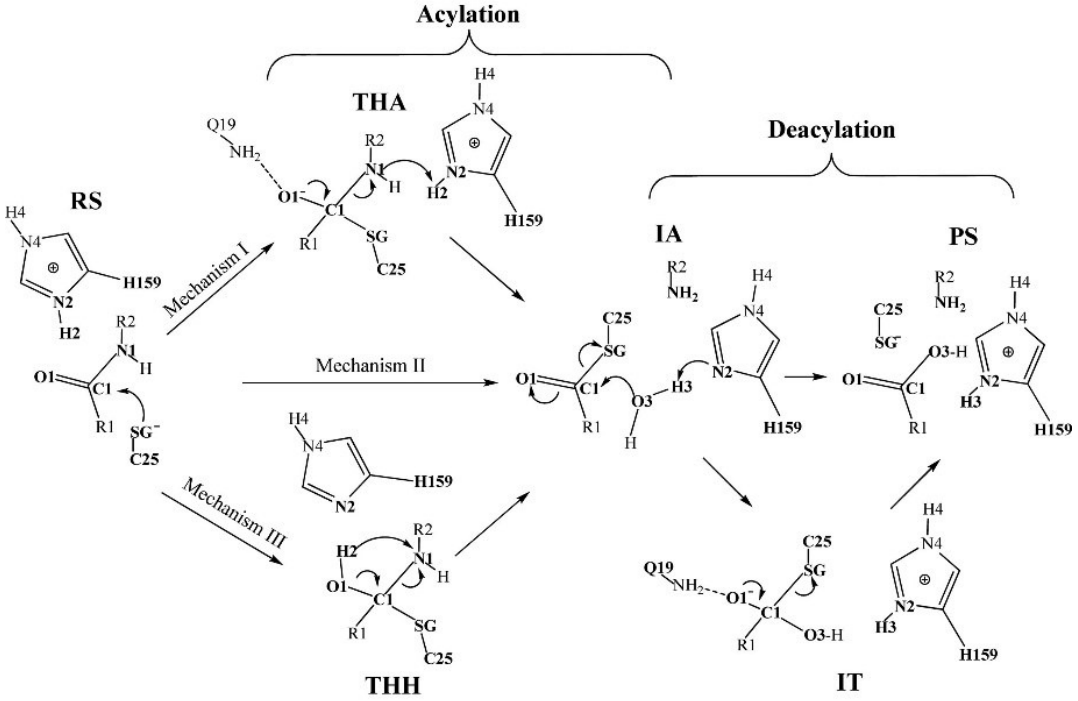


Figure 1: Reaction mechanisms of peptide hydrolysis catalysed by cysteine protease papain [7].

Cysteine proteases play many essential roles in the functioning of an organism. They are vital regulators of protein homeostasis, controlling autophagosome formation and degrading substrates within lysosomes. These enzymes are involved in both normal physiological and disease-related processes involving extracellular matrix (ECM) remodelling. Collagen, elastin, and other structural components are among the many ECM proteins that are non-specifically broken down by these enzymes. Under normal conditions, this activity is necessary for processes like bone resorption, where cathepsin K breaks down type I collagen, which constitutes the majority of the organic bone matrix. Similarly, the degradation of elastin in the vascular system depends on tightly regulated cysteine protease activity [8], [9]. This family of proteases is also closely involved in the immune response, playing a key role in Major Histocompatibility Complex (MHC) class II antigen presentation. The loading of antigenic peptides requires the degradation of the invariant chain (Ii) attached to MHC class II molecules; a process primarily carried out by cathepsin S in humans. Conversely, several cathepsins S and L may assist in breaking down the Ii chain in mice, depending on on organ [8], [10].

Cysteine proteases are not only part of a host's immune system but are also used by pathogens to weaken it. The complex processes by which infectious agents alter host immune responses are represented by pathogen-derived cysteine proteases. By targeting the same proteins as host

enzymes, many pathogens release cysteine proteases that interfere with normal immune function from the moment the pathogen is recognized through effector mechanisms. These pathogen proteases range from highly specific enzymes that target particular immune pathways to non-specific enzymes that broadly break down immune proteins. Understanding these interactions is essential for developing new treatment strategies for infectious diseases and for understanding how pathogens evade immune detection [11]. Cysteine proteases are also involved in tumour malignancy and cancer progression. They are upregulated and show altered trafficking patterns in various cancer types; higher expression levels typically correlate with increased malignancy and poorer patient prognosis. Both cancer cells and infiltrating immune cells release proteases that degrade intercellular adhesion molecules, promote angiogenesis, and facilitate metastasis of tumour cells [12].

Recent studies indicate that cysteine proteases have significant roles in various neurological diseases, including multiple sclerosis, amyotrophic lateral sclerosis, Parkinson's disease, Alzheimer's disease, and Huntington's disease. In these conditions, the exopeptidase-active cathepsins B, H, C, and X/Z are particularly vital. Cysteine cathepsins serve as important signalling molecules in neurodegeneration due to their involvement in multiple pathological processes and their presence outside lysosomes [13].

1.3 Legumain

1.3.1 Classification and substrate specificity

Legumain, also known as asparaginyl endopeptidase (AEP) or vacuolar processing enzyme (VPE) in plants, is a cysteine protease that belongs to the asparagine protease family (C 13, CD 3. 4. 22. 3.4.22.34 [14]). It is grouped together with caspases, metacaspases, clostripain, gingipain, and separin. They share a common six-stranded β-sheet domain at the centre of the molecule, made up of parallel strands. The central pair supports catalytic residues, His^{148} and Cys^{189}. Each catalytic site is preceded by a block of four hydrophobic residues. The active site histidine, if followed by a glycine and a cysteine, is preceded by an alanine residue [5]. This structure is surrounded by five α-helices [15]. Legumain exhibits strong P1 substrate specificity for Asn (and at lower pH, Asp is also acceptable), which is not located on structural elements [16]. This strict specificity is due to the zwitterionic nature of the S1 pocket. Arg^{44} and His^{45} create a positively charged environment that interacts with the side chain oxygen, while Asp^{231} and Glu^{187} provide a negative charge to the site, interacting with the amide side chain group. Normally, negatively charged aspartate would be rejected by the negative charge, but if protonated under low pH conditions (Asp side chain pKa is 3.90), its size and charge closely resemble asparagine, allowing it to bind to the S1 site of legumain. Residues Ser^{216} to Tyr^{220} on the βIV strand facilitate non-prime substrate binding along with Tyr^{228}, Trp^{232}, and Asn^{263} for P2 to P6. At this site, Tyr^{217} forms a somewhat hydrophobic S2 site, Tyr^{228} constitutes a polar S3 site, Tyr^{217} and Trp^{232} together form a hydrophobic S4 site, and Asn 236 determines interactions at S5 [15]. Experiments on human and Schistosoma legumain have shown a preference for Pro-Thr-Asn (in the former) and Thr-Ala-Asn (in the latter) at the P3-P1 positions. P1 specificity was studied with human, porcine, and *Schistosoma* legumain. At this position, Arg, Lys, and His are preferred, with small side chain residues also well tolerated; only Pro was found to be uncommon at this position [14]. S1' and S2' sites are formed by Gly^{149}, Ser^{150}, and Glu^{190} [15], [17] (Figure 2).

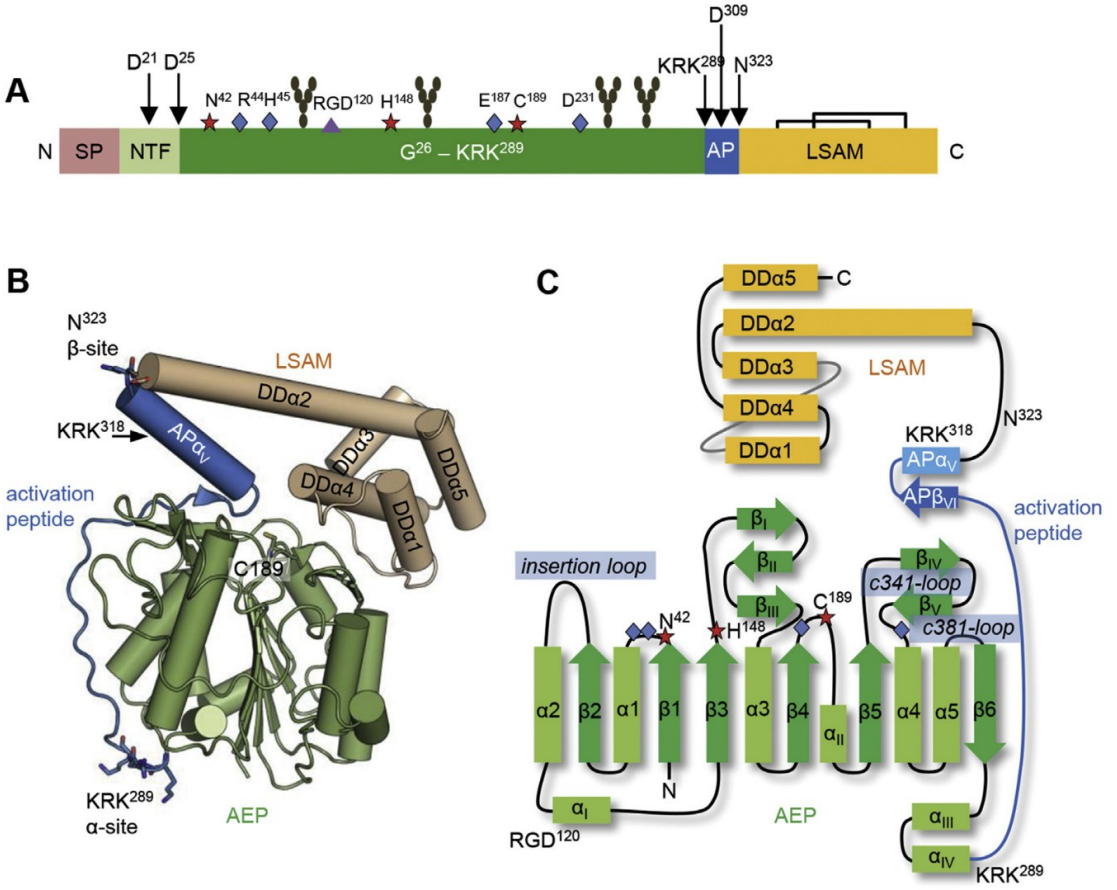


Figure 2: Schematic visualization of legumain structure: **(A)**: diagram of human prolegumain. Red: signal peptide (SP), light green: N-terminal fragment (NTF), dark green: catalytic domain, blue: activation peptide (AP; KRK^{289} - N^{323}), orange: Legumain Stabilisation and Activity Modulation (LSAM; D^{324}-Y^{433}) domain. Red stars mark catalytic residues, blue diamonds mark S1-specificity residues, and purple triangles mark the RGD^{120} motif. Arrows mark processing sites that are autocatalytic. There are disulphide bridges on the LSAM domain. **(B)**: The crystal structure of human prolegumain (PDB entry 4fgu). The caspase-like catalytic domain is in green, the AP is in blue, and the LSAM domain is in orange. The catalytic Cys^{189} marks the active site, while the C-terminal processing sites KRK^{289} and N^{323} are depicted as sticks. The KRK^{289}-KRK^{318} pattern on the AP might be a bipartite nuclear localization signal. Also, legumain may get ubiquitinated at Lys^{318}. **(C)**: A schematic of the topology of prolegumain. The AP is in blue, the catalytic domain is in green, and the LSAM domain is in orange. Red stars show catalytic residues, and blue diamonds show S1-specific residues [14].

1.3.2 Structure

Legumain is expressed as an inactive protein precursor. Starting from the N-terminal, there is a short signal peptide (Met^{1}-Ala^{17}) that is released during cellular transport [18]. Following that is a propeptide (Val^{18}-Asp^{25}), then the cysteine protease domain with the catalytic pocket (Gly^{26}-Asn^{323}), which includes the RGD motif and a C-terminal prodomain (Lys^{289}-Tyr^{433}). The prodomain is located above the protease domain, shielding access to the active site. The C-terminal domain can be divided into the activation peptide (AP) (Lys^{287}-Asn^{323}) and a death domain (DD)-like fold (Asp^{324}-Tyr^{433}), also known as the legumain stabilization and activity modulation (LSAM) domain. It consists of five alpha helices (DDα1-5) and is stabilized by two disulphide bridges. Enzyme activity is regulated by the AP and LSAM; the AP blocks substrate

access to the S'1 and non-primed sites, while LSAM blocks primed sites, keeping the zymogen inactive (Figure 2) [14], [15].

1.3.3 Organism and tissue distribution

Legumain or its ortologues have been confirmed to exist in plants, blastocysts, trematodes, mammals, and ticks [19]. In mammals, only one functional isoform is currently known, along with an additional pseudo-gene that serves no known purpose in humans. Other organisms, however, can produce a wide variety of isoforms. For example, *Arabidopsis thaliana* has four families of legumain isoforms (α , β, γ, and δ vacuolar processing enzyme (VPE)), sunflowers can produce at least five isoforms, and barley has eight. Depending on the species, ticks have one or two identified isoforms. Blastocysts have two isoforms [19], [20], and trematodes (including *Schistosoma mansoni*, *Fasciola gigantica*, and *Opisthorchis viverrini*) have one or two [21].

Plant legumains are distributed throughout the organism; αVPE and γVPE are more commonly found in vegetative parts of the plant, βVPE is localized to reproductive organs and particularly in the embryo, and δVPE is found in the seed coat. Since tick legumains play a vital role in digesting blood components, they are abundant in the midgut. Mammalian legumain is present in the spleen, liver, lungs, intestine, muscle, brain, testis, and most abundantly in the kidneys. It is also localized in some antigen-presenting cells, indicating its role in immune response and inflammation [14], [18], [22].

The primary sequence and structure of legumain are well conserved among species (Figure 3). Especially well conserved is the catalytic domain with active residues and disulphide bridges, the C-terminal less so. Although present in all, glycolisation sites, RGD-motif, TNF receptor associated factor 6 target sequence, signal peptide, sequence, and position of autocatalytic sites differ in different organisms [14].



Figure 3: Legumain sequence from different organisms. A sequence alignment of human (Q99538), mouse (O89017), Chinese hamster ovary (CHO, G3I1H5) and schistosome legumain (Q9NFY9), *Haemaphysalis longicornis* legumain isoform 1 (tick_1st; A4PF00) and isoform 2 (tick_2nd A9CQC1), *A. thaliana* αVPE (P49047), βVPE (Q39044), γVPE (Q39119), δVPE (Q9LJX8), and *Canavalia ensiformis* legumain (P49046), human GPI-transamidase (Q92643) and human caspase-1 (P29466) aligned with GPI-transamidase and caspase was prepared using ClustalW [23] and modified with Aline [24]. Autocatalytic processing sites are indicated with arrows, catalytic residues with red stars, glycosylation sites with green triangles, S1-specificity residues with blue diamonds, the double arginine motif (Arg^{342}, Arg^{403}) with light blue diamonds, the RGD^{120} motif is shaded in yellow, the TRAF6-binding site in purple ($A^{210}NPRESSY^{217}$), the ubiquitination site Lys^{318} is indicated by a purple circle, and the KRK^{289}-KRK^{318} motifs, constituting a potential nuclear localisation motif. The sequence numbers match to human legumain. Secondary structural elements are shown based on the crystal structure of human prolegumain (PDB entry 4fgu), with green parts representing the catalytic domain, blue parts representing the activation peptide, and orange parts representing the LSAM (legumain stabilisation and activity modulation) domain [14].

1.3.4 Sub-cellular location

Intracellularly, legumain is found in compartments with lower pH values and reducing conditions. In plants that include vacuoles, pericarp, and nucellar cell walls [25]. Ticks have legumain in the epithelium of midgut cells and extracellularly in the matrix surrounding ingested food. Mammalian legumain is directed to the endo-lysosomal system by the mannose-6-phosphate receptor pathway [14].

However, during pathophysiological conditions, legumain can also be detected extracellularly, in the cytoplasm, and in the nucleus. Another lysosomal protease, cathepsin L, can be expressed in the cytosol when the signal peptide is absent [26]. However, since glycosylation is crucial for legumain activity, it must first be processed in the Golgi apparatus [14].

In tumours, legumain is known to localize on the cell surface, most abundantly on the surface of tumour-associated macrophages. Excreted legumain is post-translationally modified, but not yet auto-activated [27]. The low extracellular pH values in the tumour microenvironment allow for activation and activity of legumain [28], [29], [30], [31]. Legumain features a double KRK motif (KRK^{289} and KRK^{318}), which mimics a nuclear localization sequence, important for facilitating nuclear localization of proteins [32]. Since the tumour microenvironment increases plasma membrane permeability by perforating it, this environment enables legumain to be translocated to the cytoplasm and then to the cell nucleus [14].

Since activated legumain is generally not stable at the neutral pH values typically found in these non-classical locations, it faces stability issues. The negatively charged surface of the AEP domain is stabilized by protonation at low pH, but becomes electrostatically repulsive at neutral pH, causing the protein to unfold [15]. To remain stable in neutral conditions, activated legumain can bind to other molecules, such as integrins. Integrin αVβ3 and β1 integrins are located on the cell surface and can stabilize active legumain through binding [15], [33]. Legumain inhibitors, such as cystatin F, C, and E/M, can stabilize it by binding in a canonical manner and may even act as an agonist [34]. As mentioned earlier, the cleaved LSAM domain remains electrostatically bound to the AEP domain, stabilizing it and defining ACP (Asparaginyl-mono-carboxypeptidase) activity [34]. Other complexes can also stabilize legumain; for example, HSP90α binds to ubiquitinated Lys^{63}, a modification mediated by E3 ligase TRAF6 [28].

1.3.5 Activity and forms

1.3.5.1 Asparagine endopeptidase (AEP) and asparaginyl-mono-carboxypeptidase (ACP)

In its full, zymogen form (in mammals that is 56 kDa), legumain doesn't show any activity. AP is positioned on the catalytic domain similarly to the substrate, covering S6-S2 and S1' sites. When the endosome matures to the lysosome, the pH value drops, first to 5.5 and then to 4.5. This triggers autocatalysis at both ends of the protein, at Asp^{25} and Asn^{323}, producing 47 and 46 kDa forms, respectively [35]. This cleavage at the C-termini is crucial for AEP peptidase activity, as it cleaves AP, partially exposing catalytic residues, allowing them to interact with substrates; however, although cleaved, LSAM remains electrostatically bound to the protein. Modification of the N-terminal is not significant for activity. Decreasing pH to 4.0 enables the release of the C-terminal at $Asp^{303\text{-}309}$, yielding maximum activity. While the release of this fragment doesn't increase legumain's catalytic activity, this is explained by the low pH itself, which allows for protonation of Glu and Asp substrate side chains, in addition to breaking up salt bridges [36]. Additional cleavage by an unknown protease creates the 36 kDa form, with a possible cleavage site at the KRK^{289} motif, permitting AP to dissociate from the enzyme completely. This last one

and the 46 kDa form show no difference in activity [21]. Instead, putative cleavage at KRK^{289} may influence legumains ubiquitination by TRAF6 on Lys^{318} [28].

The zymogen form is stable at physiological pH, but after activation, it favours lower pH conditions typical in lysosomes and becomes irreversibly inactivated at pH 6 or higher. Stabilization at neutral pH is achieved through electrostatic interactions with the LSAM domain. The AEP domain carries a strong negative charge, while LSAM is positively charged, balancing the entire complex. Removal of AP enables ACP activity. In this state, LSAM remains electrostatically attached to the AEP domain, covering the previously mentioned sites, which prevents endopeptidase activity and stabilizes it. Additionally, the LSAM domain contains a double Arg motif (Arg^{342} and Arg^{406} located on DDα2 and DDα4, respectively), which is typical for carboxypeptidases and flanks the S1' site [14], [15].

His^{148}, Cys^{189} and Gly^{147} form an oxyanion pocket for carbonyl oxygen of substrate P1. It was theorized by Brandstetter and Dall, that this triply polarized carbonyl oxygen can extract proton from His^{148} Nδ1, creating C-O-H bonding. This allows Cys^{189}-Sγ to deprotonate and perform nucleophilic attack. Peptide bond rupture is induced by keto-enol tautomerization (Figure 4) [14], [15].

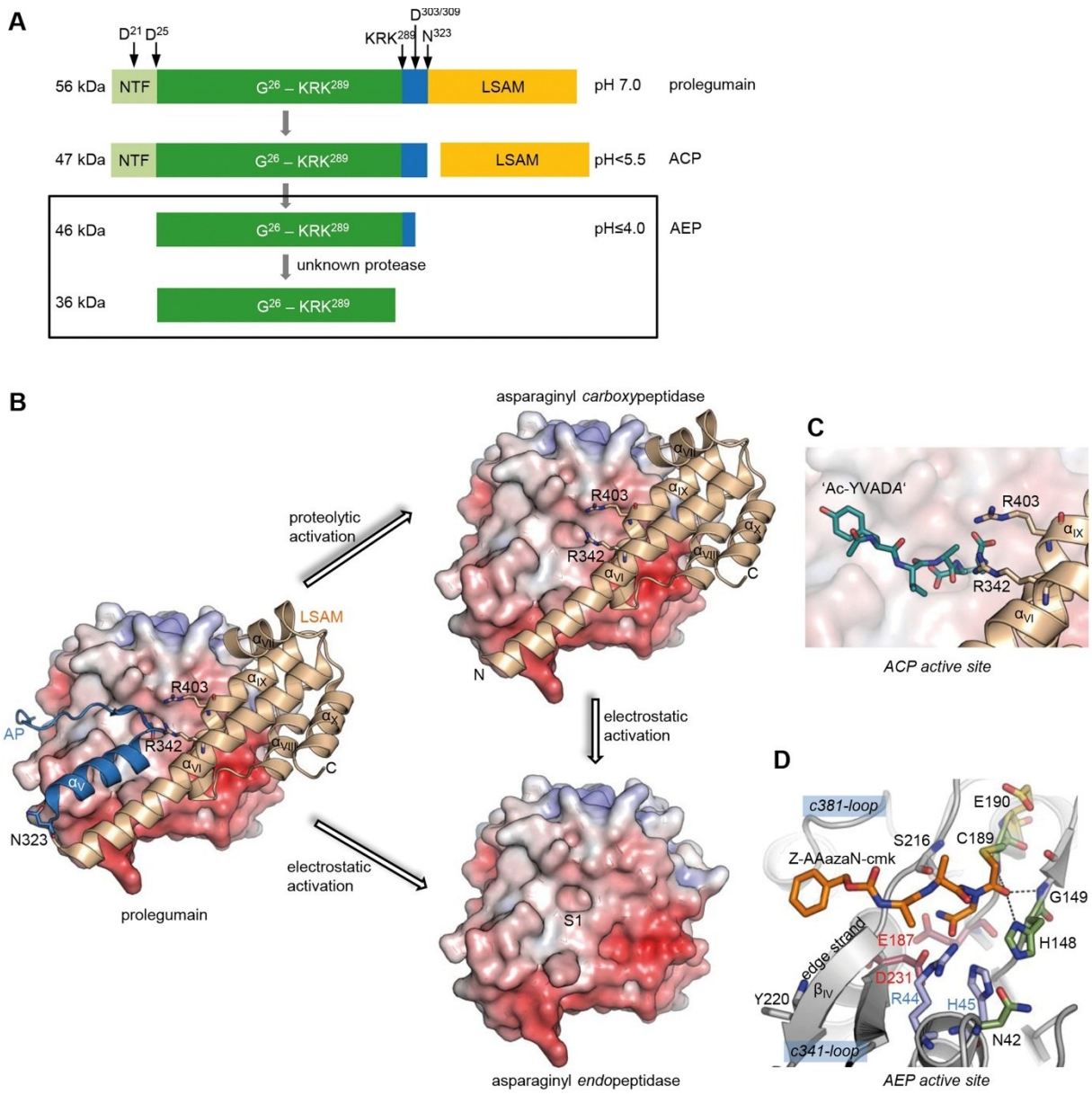


Figure 4: Legumain activation: **(A)**: Legumain is activated into asparaginyl-specific endopeptidase (AEP) or carboxypeptidase (ACP). Inactive prolegumain is synthesised at neutral pH and autocatalytically processed at the C-terminal Asn^{323} (pH < 5.5) and $D^{303/309}$ (pH ≤ 4.0) sites, as well as N-terminally after Asp^{21} and Asp^{25} residues. Another trans processing occurs following the KRK^{289} motif by an unknown protease. While N-terminal processing is not required for enzyme activation, C-terminal processing makes non-primed substrate binding sites accessible. The activation peptide (blue) creates ACP, but the Legumain Stabilisation and Activity Modulation (LSAM) domain is required for complete AEP activity. **(B)**: Legumain activation via proteolytic and electrostatic mechanisms. At pH > 6, legumain activates to an Asn-specific mono-carboxypeptidase (ACP) by proteolytic release of the activation peptide (AP). When pH drops below 5, the LSAM domain is electrostatically removed, releasing the active endopeptidase. **(C)**: Zoom-in picture of the ACP active site with a modelled substrate (Ac-YVAD|A; blue sticks). The substrate was simulated using the Ac-YVAD-chloromethylketone complex structure (pdb entry 4aw9). The double-arginine motif (Arg^{342}, Arg^{403}) is represented by orange sticks. **(D)**: Zoom-in image of the AEP active location. Catalytic residues are indicated in green sticks, S1-specificity is represented by blue (Arg^{44}, His^{45}) and red (Glu^{187}, Asp^{231}) sticks, and Glu^{190} by yellow sticks. Figure taken from [14].

1.3.5.2 Transcription factor, protein-protein interactions

It has been speculated that legumain may have functions other than its enzymatic functions. In tobacco and tomato plants, it functions as a transcription factor for 1-aminocyclopropane-1-carboxylic acid synthase after ubiquitination and transport to the nucleus [37]. The C-terminal domain acts as a paracrine factor inhibiting bone resorption [38]. The death domain-like architecture of the LSAM domain points to the possibility of other protein-protein interactions, as seen with proteins with death domains. For instance, caspases with their death domains direct signalisation in apoptosis, necrosis, inflammation, and immunity [14].

1.3.5.3 Ligase activity

Up till now, human, mouse, and plant legumains have been proven to have ligase activity. Among targets for ligation in mammals is also legumain, reverting Asn^{323} cleavage to yield inactive zymogen form. Apart from that, legumain can also ligate cystatins [34]. More substrates are known from plants. Conclavin A requires splicing by legumain for its maturation. The SFTI-1 precursor in sunflowers is cyclised before it can act as a trypsin inhibitor. The cyclotide protein family is involved in plant pathogenic response that needs processing by legumain [14].

Mammalian and plant ligation mechanisms are a bit different; for human legumain, it can be summed up as follows: there are two possible reaction mechanisms, transpeptidation or a 'one-step' reaction. First, let's take a look at the direct assault of N- on C-terminus. Like peptidase activities, ligase activity is pH-dependent, with ligation activity being possible at near-neutral conditions. In such conditions, catalytic Cys^{189} is protonated while His^{148} is not. For ligation to take place, the substrate requires a free C-terminus. Proton shift occurs between His^{148} and the C-terminus carboxylate of a substrate. Like that, carboxylate is bound to His^{148} with a hydrogen bond, and oxygen interacts with substrate 1 Ser. Proton moves to substrate 2 Asn to activate their N and C-terminus. Water is released from the latter, carbonyl is attacked by the N-terminus, resulting in the formation of a peptide bond between substrates (Figure 5) [39].

Transpeptidation occurs via thioester if there is no catalytic water present at the catalytic site, as there may be a competition for a proton from the substrate N-terminus. Water, being a stronger nucleophile, prevents the thioester from accepting it, favouring hydrolysis over ligation [39]. As for the peptidase mode of action, ligase substrates are also required to have Asn or Asp at the P1 position. At P1' position, all but Pro and acidic Asp or Glu are permitted. Lastly, P2' is more stringent, with only Cys, Glu, Ile, and Leu accepted [14].

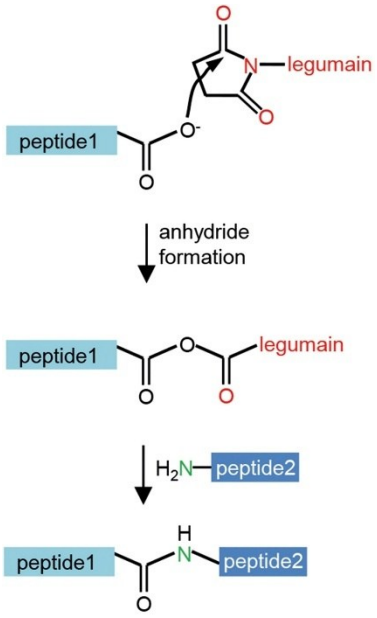


Figure 5: Legumain ligase activity. The substrate's C-terminal carboxylic acid is activated by the energy-rich Asu[147] located in the active site of legumain. Consequently, an anhydride intermediate is produced. In either situation, a peptide bond is formed, joining peptide 2 to peptide 1 via its free amino group at the C-terminus. The nucleophile in "peptide2" is a peptide's amino group or of a lysine side chain [14].

1.3.6 Regulation of activity

1.3.6.1 Inhibition

Numerous factors can obstruct legumains activity. Cystatins are a superfamily of cysteine protease inhibitors. Members of the cystatin family 2 can inhibit legumain, with their potency listed from highest to lowest: cystatin E/M [14], C [40] and F [41]. Cystatins interact with legumain in the same way as its substrates, with conserved Asn in cystatins acting as the P1 residue. Cystatin is then cleaved after Asn but remains bound to the protease, for the complex is stabilized through interaction with inhibitors legumain exosite loop (LEL) and mimics prolegumain structure. While cystatin C LEL is hydrophobic, cystatin E LEL has Arg residues, forming salt bridges with legumains Glu [34].

Clitocypin and macrocypins 1 and 3 are cysteine proteinase inhibitors found in fungi. They act as Kunitz-type inhibitors, acting on legumain in a manner similar to that of cystatins [42].

Unprocessed prolegumain is also inactive. C-terminal domain blocks access to the catalytic site, effectively inhibiting enzymatic activity. Autocatalytically activated legumain by cleavage at Asn^{323} can reverse the situation by ligation, again inhibiting itself [34].

Major factor in regulating legumain activity is pH. As described, lowering pH in the endo-lysosomal pathway is essential for initial activation, and when activated, the AEP domain by itself is inactive at $>$ pH 6. LSAM domain stabilises AEP at neutral values and dictates ACP and ligase activity. For the proposed hydrolysis mechanism to work, catalytic Cys189 must be in thiolate form to act as a nucleophile, which makes it available for oxidation. To prevent that, reducing conditions are preferred to sustain proteinase activity [14].

Synthetic low molecular weight inhibitors, manufactured as reversible aldehydes (Ac-YVAD-CHO, Ac-DEVD-CHO, Ac-WEHD-CHO), irreversible fluoro- or chloromethylketones (Z-VAD-fmk, Z-YVAD- fmk/cmk Z-DEVD-fmk/cmk), esomeprazole and omeprazole mimic legumain substrates successfully inhibiting it [43], [44]. Attaching Asp and Asn to acyloxymethyl ketone (AOMK) to synthesise activity-based probes specific for legumains P1 requirements, proves to be an efficient way of inhibition [45]. Bovine legumain is also inhibited by divalent metal ions like Hg^{2+} or Cu^{2+}, presumably because of their affinity to cysteine thiol groups [46].

1.3.6.2 Activation

As already described in previous paragraphs, LSAM electrostatically binds to and stabilises AEP domain at neutral pH values. Additionally, it may have a role in the proper folding of the whole protein after translation. Like the LSAM domain, cystatins bind to and shield the active site, stabilising it. Some proteinase activity remains detectable when inhibited by cystatins, due to the enzyme-inhibitor complex dissociating over time. Complex with cystatin C forms at pH 5.5 and remains stable at 6.5, but inhibition is reversed when pH drops back to 4.0. Therefore, cystatin C acts as a stabiliser extracellularly and readily leaves the complex upon internalisation into the cytosol and endo-lysosomal system, where proton concentrations are higher. Cystatin E differs here, as it remains bound at pH 4.0, also inhibiting legumain intracellularly [14].

Glycosaminoglycans (GAGs) are unbranched heteropolysaccharides covalently attached to proteins. They are assembled from repetitive disaccharide units with sulphate and carboxylic acid moieties, causing them to bear a negative charge. Chondroitin 4-sulphate (C4S), chondroitin 6-sulphate (C6S), chondroitin 4,6-sulphate (C4,6S), heparin, heparan sulphate, and chondroitin sulphate-derived disaccharides are able to accelerate legumain autoactivation at pH 4.0. C4S and C4,6S act as accelerators also at pH 5.0. Acceleration of auto-activation is caused by ionic interactions between the prolegumain and GAGs, which supposedly fit in the gap separating N and C-terminal propeptides, inducing conformational changes exposing the catalytic site, making auto-cleavage easier [47].

1.3.7 Physiological functions

1.3.7.1 Under normal conditions

Legumain is widely used in general protein turnover in lysosomes. In ticks, vesicles containing ingested haemoglobin from ingested blood are taken up by gut epithelial cells through receptor-mediated endocytosis. These vesicles are then fused with lysosomes, where legumain directly acts on haemoglobin in addition to mediating other connected processes [14]. Similar processes are set up in *Schistosomes* [48]. In plants, legumains play a crucial role in the activation of seed storage proteins during seed germination [49].

Mammalian legumain is vital for the processing of endogenous and invasive proteins for presentation to class II MHC antigen [50]. In that capacity, firstly, it generates peptides by processing foreign proteins for presentation on T lymphocytes, and secondly, it destroys these epitopes found on invading pathogens [51].

Cyclotides are plant proteins that protect plants from pathogens, especially insects. They are synthesized as linear molecules, but mature proteins have a cyclic backbone. Legumain is believed to be involved in this process by processing cyclotides during their maturation process, as indicated by the fact that the C-terminal domain of cyclotides ends with an Asn or Asp residue. Legumain catalyses the final head-to-tail backbone cyclization of cyclotides, a critical step that confers the cyclic peptide structure [52].

In plants, legumain is heavily involved in programmed cell death during the hypersensitive response upon infection with bacteria or viruses, as well as in developmental apoptosis during the early stages of embryogenesis. Different isoforms act similarly as caspases, for instance, promoting cell death in infected tobacco plants, chemically induced death in tomato, or nitric oxide-mediated death in *A. thaliana* [53].

As briefly mentioned, legumain acts as a transcription factor. In tomato, it induces transcription of the Acs2 gene, which encodes 1-aminocyclopropane-1-carboxylic acid synthase. The latter one causes production of ethylene, which in turn promotes plant growth and defence response [37]. Positively charged LSAM probably interacts with the negatively charged phosphate backbone of DNA [15].

LSAM domain is also linked to legumains' function in osteoclast remodelling, as a paracrine factor preventing bone resorption [14], [38].

Legumain is upregulated during ischemia in astroglia and microglial cells. It was speculated it is involved in tissue reorganisation and repair by degrading extracellular matrix, easing access to the site for immune response cells [54].

A few more functions could be deducted from the phenotype of legumain knockout mice. Otherwise, healthy, they have slower antigen processing (slower T lymphocyte response) and gain weight at a slower rate than wild-type animals. Knockouts accumulate procathepsins in the kidneys and have hemophagocytic syndrome symptoms [14]. Accumulation of macromolecules in the endo-lysosomal system has been observed, indicating legumain's fundamental role in their degradation [55].

1.3.7.2 Under pathological conditions

With legumain's role in immunity, it processes myelin basic protein (MBP), an HLA-DR2-restricted epitope. Otherwise, crucial for class II MHC presentation, overzealous legumain activity can destroy these peptides, preventing presentation, generating tolerance against them, leading to autoimmune diseases [56].

Hyperphosphorylation of the tau protein is typical for Alzheimer's disease. In the Alzheimer's brain, legumain is also located in the cytosol, where it interacts with the protein phosphatase 2 inhibitor protein phosphatase 2 inhibitor 2 [57]. It was shown that legumain activates this inhibitor, in turn increasing the degree of phosphorylation of tau [58]. Protein phosphatase 2 inhibitor 2 also functions as a DNase inhibitor; when legumain cleaves it, it causes DNA damage in the brain [59].

1.3.7.2.1 Cancer

Normally restricted to the endo-lysosomal system, legumain is overexpressed in cancerous cells and localised also extracellularly, in the cytosol and nucleus [14]. Increased quantities of legumain are not seen in tumorous cells alone, but also in tumour-associated macrophages [60]. Numerous factors contribute to this overexpression. Oncogenic transcription factor TBX2 promotes breast cancer proliferation by inhibiting expression of cystatin 6, an intracellular legumain inhibitor [61]. DJ-1 is another oncogene; it has many roles, including in anti-oxidative stress, chaperone functioning, protease functioning, mitochondrial regulation, and regulation of transcription, where it binds to numerous transcriptional factors as a coactivator, among others, to p53 [62]. It was shown that p53 binds to intron 1 of legumain (human), promoting its expression [63].

Legumain then influences remodelling of extracellular matrix by mediating the action of matrix-metalloproteinases 2 and 3 [64], fibronectin and annexin II, promoting metastasis and neoangiogenesis [65]. Legumain is key for processing of cathepsins B and L, both overexpressed in cancer, promoting metastasis [66].

1.3.7.2.2 Immune response

Toll-like receptors (TLRs) are key proteins for the recognition of invasive foreign species and initiation an immune response [67]. TLR3, 7, and 9 are found in endosomes of dendritic cells. TRL9 is a receptor for CpG-containing viral and bacterial DNA, and legumain plays a key role in its activation through proteolytic activation [68]. TLR7 single-stranded viral RNA triggers inflammation. For the receptor to be able to propagate a signal, processing at the C-terminal by legumain is essential [69]. TLR5 detects flagellin. Babier and co-workers found that upon infection with *Pseudomonas aeruginosa*, alveolar macrophages trigger a lowering of pH in the endosomal system through increased release of IL-1β following signalling by TLR5. Lower pH in endosomes enables activation of legumain and enables it to kill off invading bacteria through its proteolytic activity [69]. Legumain was found to be upregulated during autocrine stimulation of the complement system by CD46. Legumain interacts with α1-thymosin (further facilitates immune response) and cathepsin L. Following down the pathway, cathepsin L activates the C3 component of complement, producing C3a and inducing INF-γ production and Th1 stimulation of $CD4^+$ cells [70].

1.4 Protein phosphorylation

Post-translational modifications (PTMs) are covalent additions of different functional groups to amino acid residues, proteolysis of proteins, or ligation of proteins. These modifications change proteins' physical and chemical properties, modulate their stability, activity, localization, and conformation. This allows proteins to expand their array of functions and modes of action [71]. Disparities in the regulation of PTMs can lead to cell death or the development of diseases such as cancer, neurological conditions, and diabetes [72]. Hence, understanding of PTMs distribution and occurrence is crucial for the study of diseases. However, post-translational modifications can be difficult to detect and characterize [73].

Phosphorylation is one of the most common protein modifications and plays a role in many cellular processes such as metabolism, growth, division, differentiation, mobility, subcellular localization, organelle and membrane transport, and immune response [74]. Faults in the regulation of protein phosphorylation can lead to the development of various diseases, including cancer [75]. Phosphorylation is a reversible, covalent addition of a phosphate group (PO_4^{3-}) from a donor [76], which is usually ATP [77], but PAPS (3'-phosphoadenosine-5'-phosphosulphate), GTP (guanosine triphosphate), and phosphoenolpyruvate can also be used. The reaction is catalysed by kinases and reversed by phosphatases [78].

1.4.1 Function

Protein phosphorylation can be used by the cell to regulate several different processes:

Metabolism can be regulated through various mechanisms, including phosphorylation. This rapid process can serve as an on and off switch, as seen with glycogen phosphorylase B, an enzyme involved in glycogen metabolism. glycogen phosphorylase B is activated by phosphorylation at the Ser^{14} residue [79].

One of phosphorylation's main roles is in signal transduction, which can be initiated in different ways. Adding a phosphate group to kinase activation sites causes it to change conformation into an active or inactive form and subsequently triggers downstream signal transduction. Examples include phosphorylation of ERK on Thr^{202}/Tyr^{204}, which activates the enzyme, or inhibition of SRC by modification of Tyr^{527}. Phosphorylation on specific motifs can create binding sites for protein-protein interactions, which can trigger signalling, activation, localization, or cellular transport. For instance, these motifs include Src homology 2 (SH2) and phosphotyrosine-binding (PTB) domains for phosphorylated tyrosine residues. One example is the interaction between nephrin 1 and Grb2 after phosphorylation of nephrin by Src Fyn [80], [81].

The cell cycle is also partly regulated by protein phosphorylation. For example, phosphorylation of Akt on prevents cell death through following mechanism. Inhibiting glycogen synthase kinase 3 (GSK-3) through phosphorylation, which prevents GSK-3 from tagging important cell cycle proteins like cyclin D and cyclin E for degradation, thereby enhancing their stability. Cyclin D and cyclin E are crucial for the G1 to S phase transition of the cell cycle. activating mTORC1 via phosphorylation of TSC2 and PRAS40, which drives protein synthesis essential for cell growth and cell cycle progression. mTORC1 activation promotes translation of cyclin D1 and c-Myc, further supporting cell proliferation. Regulating cyclin-dependent kinase inhibitors p21^Cip1^ and p27^Kip1^ through phosphorylation, which modulates their localization and stability, thereby facilitating progression through the cell cycle [82]. Dephosphorylation of c-Src, a proto-oncogene, causes its inactivation and the loss of control over cell growth. c-Src is kept in an inactive conformation when phosphorylated at Tyr^{527}. Dephosphorylation at this site releases this inhibition, causing c-Src activation, which then leads to loss of normal growth regulation and can contribute to cancer development [76].

Signalization and cell survival can be directly affected by phosphorylation's impact on cellular transport. For example, death-associated protein, a Ser/Thr kinase, is transported from the cytoplasm to the nucleus when phosphorylated on Ser^{350}, which triggers apoptosis in T and B cells [76].

Phosphorylation can regulate other post translational modifications. For instance, when insulin receptor substrate-1 is dephosphorylated, it's prone to ubiquitination by subsequent degradation [83].

Since protein phosphorylation plays a crucial role in signalling cascades, many things can go wrong if its regulation is disrupted, commonly leading to the development of pathological states, including cancer. More than 1 000 kinases are known to be aberrant in cancerous conditions. One example is v-Src, a tyrosine kinase discovered as the main driver of sarcoma in chickens. Normally, v-Src is phosphorylated on Tyr^{527}, but a mutation on the C-terminal abolishes this modification, leading to conformational changes and auto-phosphorylation, which promote increased cell growth with downstream effects [76].

Protein phosphorylation is also important in regulating epithelial-mesenchymal transition (EMT), as a key signalling mechanism. For instance, in TGF-β/SMAD signalling cascade, receptor mediated phosphorylation activates SNAIL, ZEB and TWIST, which lead to stemness and therapy resistance of the tumour. Receptor tyrosine kinase (like EGFR or cMET) phosphorylation is needed for downstream PI3K/Akt and MAPK/ERK signalling pathways, activating EMT transcription factors that lead to EMT and further, cause extracellular matrix degradation through action of matrix metalloproteinases [84].

1.4.2 Enzyme regulation

Phosphorylation is carried out by protein kinases, while protein phosphatases remove the modification [85]. Because phosphorylation occurs so often, there are many enzymes involved. In eukaryotes, kinases make up 2–3% of genes, with 621 identified in the human genome [74], [86], [87]. Conversely, protein phosphatases, estimated at 264 in humans, account for just over 1% of the genome [88].

1.4.2.1 Protein kinases

Protein kinases consist of variable regulatory domains and a catalytic domain. About 250 amino acids long, the primary structures of the catalytic domain are well conserved across all kinase families, along with their tertiary structures. They can be further divided into 12 subunits based on typical amino acid sequences separated by larger residues. Within these domains, the key residues include two Gly in subdomain I, Lys in II, Glu in III, Asp and Asn in VIB, Asp and Gly in VII, Glu in VIII, Asp and Gly in IX, and Arg in XI. These subunits form two lobes: the smaller N-terminal lobe, made from subunits I through IV, consists of five antiparallel β sheets and is responsible for ATP binding [89]. The larger C-terminal lobe, also called the big lobe, covers subunits VIA through XI and is primarily composed of α helices. Subdomain V remains centrally located, forming a cleft between the two lobes, which serves as the catalytic and ATP-binding site. The nucleotide sits within this pocket, with the phosphate group facing outward toward the medium [90], [91]. A simple architecture with N- and C-terminal lobes is shown in Figure 2. All protein kinases contain the sequence Lys-Glu-Asp-Asp at their active site. When active, a salt bridge forms between Lys in the β4-strand and Glu in the αC-helix of the N-terminal lobe (the first two residues in the sequence). For activity, kinases require at least one divalent ion, such as Mg or Mn, which bridges the first Asp in the sequence to the β- and γ-phosphates of ATP. Another ion connects the α- and γ-phosphates [92].

Roughly, protein kinases can be divided into two main groups based on their target residues: Tyr-specific kinases, which make up about one-fifth of kinases, and Ser/Thr-specific kinases. This division is determined by the depth of the catalytic cleft, which is deeper in Tyr-targeting kinases to accommodate the larger tyrosine side chain compared to those of serine and threonine. However, some Ser/Thr kinases can attach a phosphate group to Tyr residues, and more rarely, some Tyr kinases can do the opposite [74].

Kinases are classified into numerous types based on their catalytic domain and function, as illustrated in Figure 6. At the highest level, they are divided into two superfamilies: eukaryotic protein kinases (ePK), which has eight groups, and atypical protein kinases (aPK), which contains thirteen groups. Each group is then divided into many families [86]. The first group, AGC, is made up of 14 families grouped into 21 subfamilies. These kinases typically prefer to target Ser/Thr residues near basic ones, such as Arg and Lys. Some examples are PKC, RAC, ribosomal S6 kinases, cyclic nucleotide-dependent kinases, βARK, and RhK. They regulate cell proliferation, division, metabolism, and survival. Similarly, CaMK kinases from the same family prefer to target substrates that have basic residues at specified positions. The family is slightly larger, with 17 families that include CaMK, MLCK, CDPK, and AMPK. The members of this group are implicated in calcium/calmodulin signalling, DNA repair, and cell cycle arrest [89], [93]. Two of the largest families in the following category, CMCG, are MAPK and CDK, along with 60 others. They appear to prefer substrates with a proline residue on the P1 position [94] and play an important role in cell proliferation, differentiation, and death [95]. Tyrosine kinase family (TK) is one of the most abundant, with around 100 recognized kinases in humans [86]. As the name suggests, they can only act on tyrosine residues [96]. In general, they are classified as receptor tyrosine kinases, which are transmembrane proteins situated on the plasma membrane,

and non-receptor tyrosine kinases, which are found in the cytoplasm. The extracellular domain of receptor tyrosine kinases is responsible for ligand binding and can fluctuate, whereas the transmembrane domain remains constant. Ligand binding causes oligomerization of the receptor and signal transduction inside the cell due to the activation of the receptor. Some of the families included are epidermal growth factor receptor (EGFR), insulin receptor (IR), and fibroblast growth factor (FGF). Modification of this receptors can lead to perturbations in phosphorylation signalling. When cathepsin L cleaves EGFR, this results in truncated form of the receptor with diminished capacity for downstream signalling via phosphorylation. When EGFR is truncated, downstream phosphorylation levels of ERK and Akt are reduced. This leads to activation of alternative signalling cascades, rendering EGFR inhibition ineffectual [97]. Non-receptor TKs are found in the cytoplasm, including members of the JAK family. Tyrosine kinase-like group (TKL) enzymes share a sequence with conventional tyrosine kinases but phosphorylate serine and threonine residues [98]. They, like TKs, can be receptor or non-receptor kinases [99]. RGC kinases are a smaller group with sequence resembling tyrosine kinases, but they lack the residues necessary for phosphate group transfer [86]. The STE group is critical for the activation of MAP kinases. Three recognized families activate each other before acting on MAPK. This comprises interleukin 1, rapid accelerated fibrosarcoma (RAF), and tumour growth factor beta (TGFβ) [98]. The casein kinase 1 group consists of serine/threonine kinases in sequences where the N terminal residue of the P0 site has already been phosphorylated. They have a key function in cytoskeleton repair and transcription [95]. The last category, dubbed 'others', is made up of families that cannot be classed together despite having conserved traits within their separate families. They include casein kinase 2 (CK2), IkappaB kinase (IκB), and nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) [86], [95]. Atypical protein kinases differ from eukaryotic protein kinases in their sequence, yet certain aPKs have similar three-dimensional structures. Examples include mTOR, an atypical protein kinase, which differs from other kinases in sequence, yet has a similar structure to them [86].

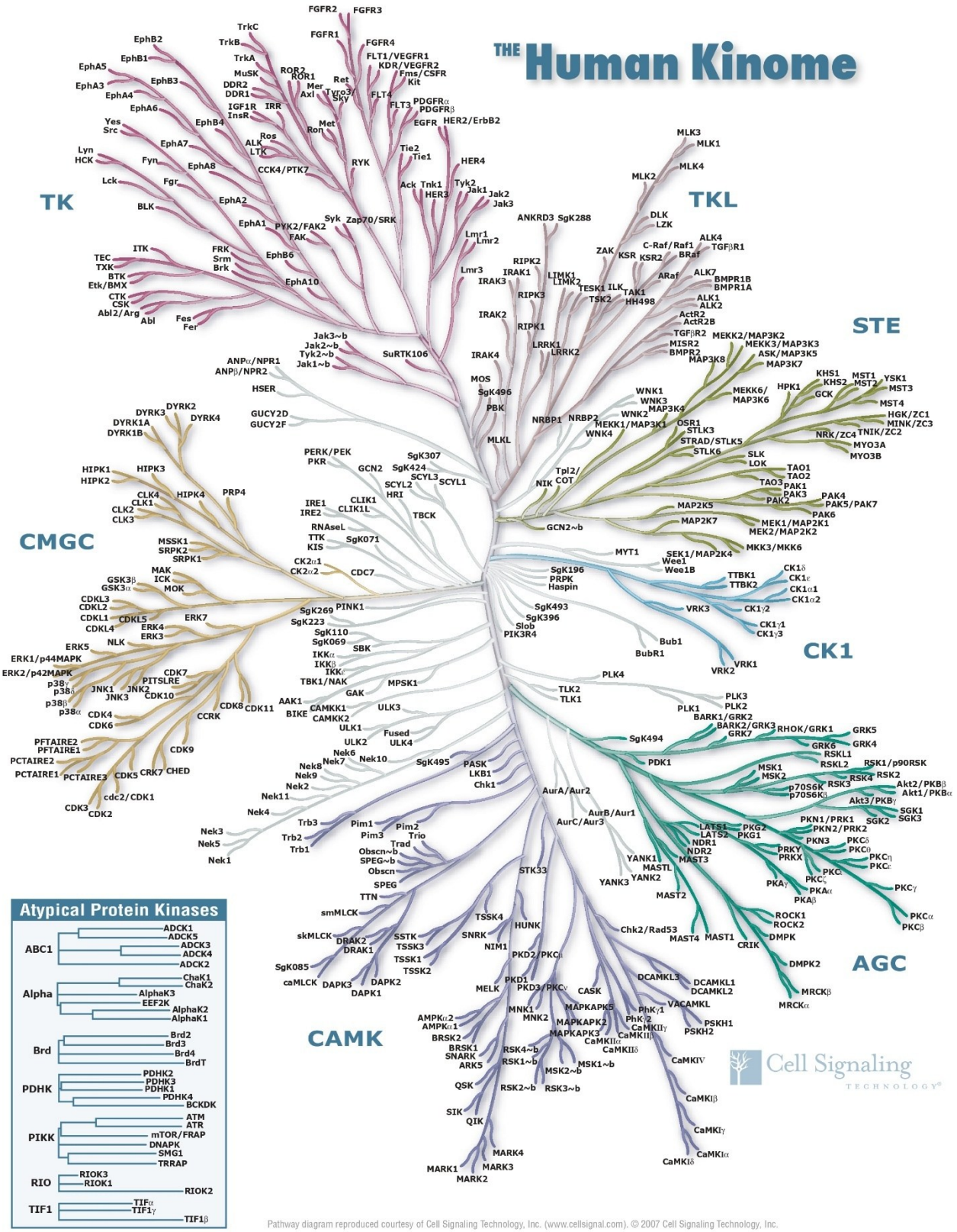


Figure 6: Human protein kinases overview with eukaryotic and atypical protein kinases. AGC (PKA, PKG, and PKC families), CAMK (calcium/calmodulin-dependent protein kinase), casein kinase 1 (CK1), CMGC (CDK, MAPK, GSK3, and CLK families), STE (homologs of yeast Sterile 7, 11, and 20), TK (tyrosine kinases), and TKL (tyrosine kinase-like). Image taken from Cell Signalling Technology [89].

Kinase target specificity is determined by the depth of the catalytic cleft, which influences Ser/Thr or Tyr specificity. Additionally, amino acids in the phosphorylation site itself play a crucial role in enzyme-substrate recognition. On average, the kinase's catalytic cleft interacts with four residues on each side of the phosphorylation site; residues farther away can interact with regions adjacent to the catalytic site. For example, cyclin-dependent kinase-2 (CDK2) recognizes phosphorylation sites within the consensus sequence S/T-P-X-K/R (where X is any residue). Proline is favoured at the P1 position because it cannot form a hydrogen bond, and positive residues at P3 can interact with the phosphorylated site on the kinase (Thr^{160}) [74]. Further kinase target specificity is governed by two mechanisms, each involving a part of the enzyme separated from the catalytic site, facilitating more flexibility during target recognition and modification. The first mechanism involves protein-protein interactions, where the kinase contains a separate interaction domain from the catalytic domain. Besides targeting proteins, this domain can also engage with other co-acting partners through short peptide motifs. Domains like SH2 and SH3, found in kinases such as Scr, Abl, Hck, and Csk, exemplify this mode of interaction, typically observed in tyrosine kinases. The second mechanism involves docking interactions, where a docking groove on the catalytic domain remains distinct from the active site. These grooves interact with short recognition peptides; the modification site itself is not part of this sequence. Ser/Thr kinases utilize docking interactions as well. Both strategies depend on recognizing short peptides, typically 3 to 10 amino acids long, enabling transient interactions suited to the rapid nature of signal transduction pathways. An example of such a recognition peptide is the D-motif (Ψ1-3x3-7-Φ-x-Φ, where Ψ is a positively charged amino acid, Φ is a hydrophobic residue, and x is any amino acid) in the MAP kinase signalling pathway, found on activators, inhibitors, and substrates. Positive residues bind to negatively charged ones, and hydrophobic residues interact with hydrophobic regions on the kinase's CD-site [100],[101].

Furthermore, other mechanisms also contribute to specificity. These interactions can be phosphorylation-dependent, as some kinases require a phosphorylated residue—either nearby or far from their target—before acting on the substrate. For example, PDK1 (3-phosphoinositide-dependent kinase 1) acts on several other AGC kinases that contain its recognition motif. However, PDK1 only binds to a phosphorylated recognition peptide: F-x-x-F-S/T-F/Y, where either Ser or Thr must be phosphorylated. Binding of the kinase to this motif increases its activity roughly fivefold. A similar situation occurs with GSK3 (glycogen synthase kinase-3), where the substrate must have a phosphate group on a residue following Ser or Thr (P4) [100]. Another example is PLK1 (polo-like kinase), which targets the sequence S-S/T-X, with the middle Ser or Thr being phosphorylated [74].

One peculiar case of interaction between kinases and other proteins is found with cyclin-dependent kinases. Cyclins act both as activators for dependent kinases and additionally carry docking groove for the substrate, acting as adaptors for their kinase [100],[74].

Where the kinase is located in the cell influences overall target specificity by restricting the substrates it can interact with. For example, if ERK2 (extracellular signal-regulated kinase-2) is in the nucleus, it affects substrates that control cell differentiation. However, if it's in the cytoplasm, it regulates mitogenesis [74].

Some kinases rely on forming complexes with partners to create a scaffold around them. These scaffolds determine the specific kinase substrate based on their composition. One kinase in the yeast MAPK signalling cascade, MAPKKK (Ste11), is involved in three different signalling pathways. The composition of the scaffold greatly influences which pathway the kinase will participate in [74].

Even if a kinase for a specific substrate is present in the environment, the substrate might not get phosphorylated if there's another substrate that the kinase prefers. If the preferred substrate's concentration is high enough, it can occupy all available enzymes, preventing other substrates from being modified [74].

There are two proposed mechanisms by which kinases mediate phosphate transfer to substrates. Protein kinase A (PKA) operates via a ternary complex, where ATP directly reacts with the protein substrate to produce ADP and phosphorylate the protein. The order of binding between ATP and the protein substrate varies among kinases. PKA binds ATP first, then the substrate, while Csk kinase shows no preference for the order. Research on p38 kinase varies widely, with possibilities including ATP first, substrate first, or no preferred sequence. The ternary complex mechanism seems more common than the next one, the ping-pong mechanism. In this mechanism, the enzyme forms an intermediate where the phosphate is covalently attached to the active site Asp before being transferred to the substrate. This mechanism has been observed in some atypical protein kinases. The transfer is believed to be dissociative, with the bond between the γ-phosphate and the rest of ATP breaking before the nucleophile bonds with it, creating a metaphosphate-like intermediate [102]. Using PKA as an example, the residues involved are part of the Lys-Glu-Asp-Asp sequence found in all protein kinases. When ATP and the substrate are bound to PKA, the phosphorylation site (serine residue) and the γ-phosphate of ATP are aligned. Asp^{166} removes a proton from serine, while Lys^{168} neutralizes the negative charge of the γ-phosphate. The free oxygen from serine then acts as a nucleophile, attacking the distal phosphate and forming metaphosphate. The complex dissociates, leaving ADP and the phosphorylated protein (Figure 7) [92].

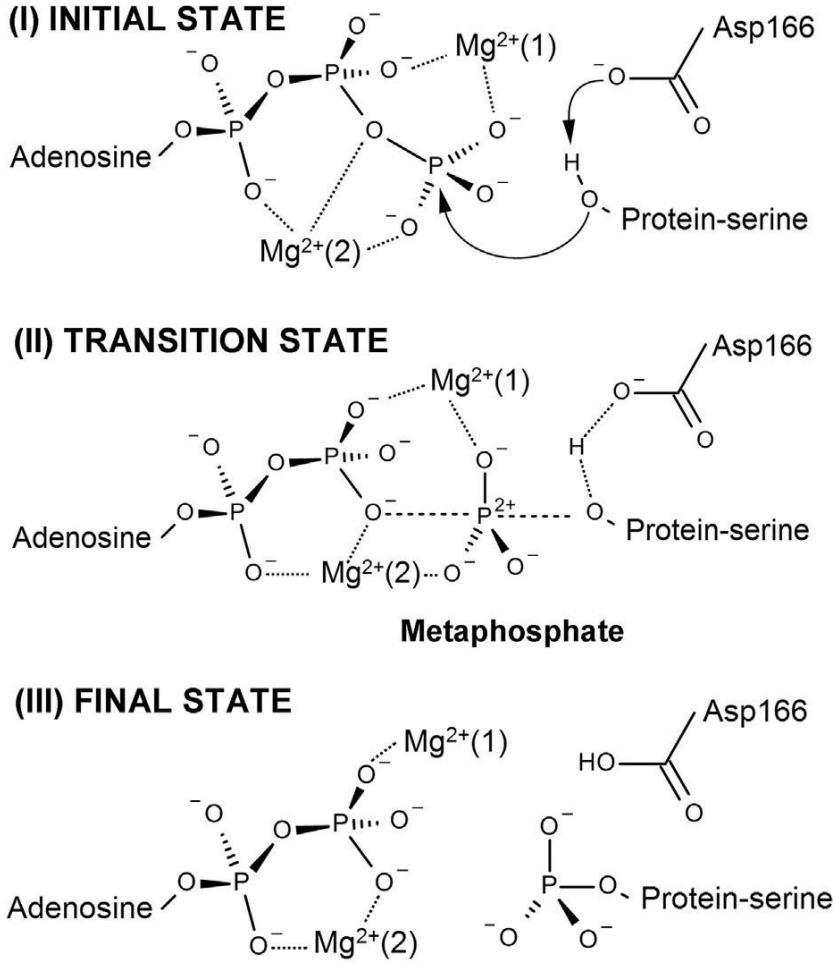


Figure 7: Mechanism of phosphate transfer, when ATP and protein substrate bind to the kinase, the γ-phosphate of ATP and the phosphorylated serine of the protein are in close proximity to each other **(I)**. Kinase's Asp^{166} attracts a proton from serine, enabling it to perform nucleophilic attack on phosphate group, creating a transition state with metaphosphate **(II)**. When transfer is over, ADP and phosphorylated protein are left **(III)** [92].

1.4.2.2 Protein phosphatases

Phosphatases oppose kinases by removing the phosphate group from the substrate through hydrolysis. Unlike kinases, protein phosphatases do not have such conserved features [88],[103]. Below is an example of a typical 3D structure of a phosphatase, like protein phosphatase 2A (PP2A). The centre of the molecule consists of a scaffolding subunit A and a catalytic subunit C. Subunit A is shaped like a half circle and is composed of two layers of α-helices with 15 HEAT (huntingtin-elongation-A subunit-TOR) repeats, each consisting of a pair of antiparallel helices, forming a ridge along the molecule. Subunit C, a globular structure made of α and β components, connects to A through this ridge on the concave side of the scaffolding subunit via its highly conserved C-terminal. Protein kinases greatly outnumber protein phosphatases, yet they can be highly specific. They achieve this through interactions with numerous regulatory subunits, which determine phosphatase localization and target specificity. In the case of PP2A, this is subunit B, which has at least 16 known different subunits across 4 subfamilies. Subunit B'-γ1 interacts with subunit A via the HEAT repeats ridge and with subunit C via its C-terminus, helix α5, and a larger surface area. This specific subunit B itself is composed of 18 α-helices (Figure 8) [104], [105].

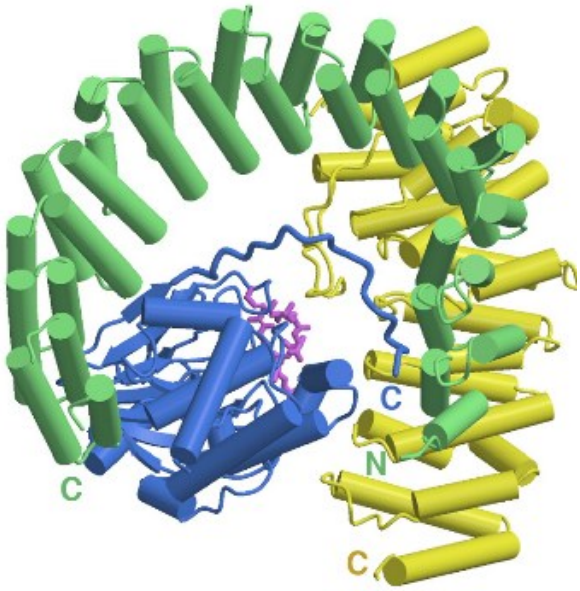


Figure 8: Crystal structure of protein phosphatase 2A (PP2A) holoenzyme. In green, forming half circle is scaffolding subunit A, yellow is catalytic subunit C, blue is regulatory subunit Bα, purple is microcystin-LR [104].

While protein kinases may make up to 3% of the human proteome, protein phosphatases account for about 0.84%, with 189 known enzymes categorized into 10 folds, which are further divided into 20 families. This classification is based on structural similarity, which is greater among phosphatases than kinases, since phosphatases evolved from several separate families. As a result, they are structurally and functionally diverse. The first three folds—Cys-based class I, II, and III (CC1–3)—all share a conserved catalytic motif His-Cys-X5-Arg. Most human protein phosphatases belong to CC1, including the protein tyrosine phosphatase (PTP) family and the dual specificity phosphatase (DSP) family [88]. PTPs can be divided into receptor-like PTPs (examples include CD45, LAR, and STEP) and non-transmembrane PTPs (such as PTP1B, SHP2, MEG2). They feature a conserved 280-residue catalytic domain [106] with 10 signature motifs and seven conserved residues. DSPs can dephosphorylate modified Ser, Tyr, and Thr residues, although their structures are less conserved aside from the Cys motif. Two DSPs are CDK and VHR [107]. The CC2 fold includes only two families in humans: the small CTD phosphatase family (Fcp1, Scp1) and the SSU72 family of RNA polymerase C-terminal phosphatases [108]. The last Cys-based fold contains three families, including CDC25 [88]. Metal-dependent protein phosphatases (PPM) are Ser and Thr-specific enzymes requiring metal ions (Mg^{2+} or Mn^{2+}). Phosphoprotein-like phosphatases (PPPLs) share substrate specificity with PPMs and also depend on metal ions, but they utilize regulatory subunits for substrate specificity, whereas PPMs have additional domains and motifs for this purpose [109]. Haloacid dehalogenase (HAD) fold phosphatases, characterized by little sequence similarity except for the conserved D-X-D-V/T catalytic site and Rossmann-like fold, are specific for Ser/Thr and Tyr phosphorylated substrates [110]. The alkaline phosphatase fold, as the name suggests, includes phosphatases with highest activity in basic conditions [111]. Histidine phosphatase fold encompasses enzymes with conserved His residues and the L-S/T-X-X-G motif [112]. The protein histidine phosphatase fold, represented by only one family (which includes LHPP), dephosphorylates histidine residues [113]. The final fold, RTR1, also has a single known member: Rtr1, a phosphatase targeting Ser and Tyr [114].

One way for protein phosphatases to locate their targets is through docking grooves on the enzyme's surface. Protein phosphatase 2B (PP2B) recognizes the sequence P-X-I-X-I-T on its target nuclear factor of activated T cells (NFAT) via a docking groove on the enzyme's surface. This feature is more common among Ser/Thr phosphatases, whereas phosphatases that prefer Tyr often use a different mechanism. As mentioned, phosphatases depend on regulatory subunits with interaction domains to determine their localization and specificity. For example, protein phosphatase 1 (PP1) has more than 50 known interaction partners, all sharing common motifs of conserved residues F-X-X-R-X-R or R-V-X-F, which serve as interaction motifs between the core enzyme and regulatory subunits [100].

Different families of phosphatases utilize different catalytic mechanisms. The following are briefly described principles of some groups. PPPs use various metal ions to activate water molecules, which in turn hydrolyse the phosphoester bond. PPM also uses metal ions, two Mn^{2+} ions, to activate water, which interacts with phosphate oxygens via hydrogen bonds and performs nucleophilic attack. Additional water molecule protonates now released phosphate. Conserved Cys residue in PTP family provides a nucleophile for hydrolysis while Arg stabilises the reaction (Figure 9). Members of HAD family use first Asp in conserved D-X-D-X-Z/V motif to act as nucleophile and then water to release phosphate from covalently bound phosphate – enzyme intermediate [115].

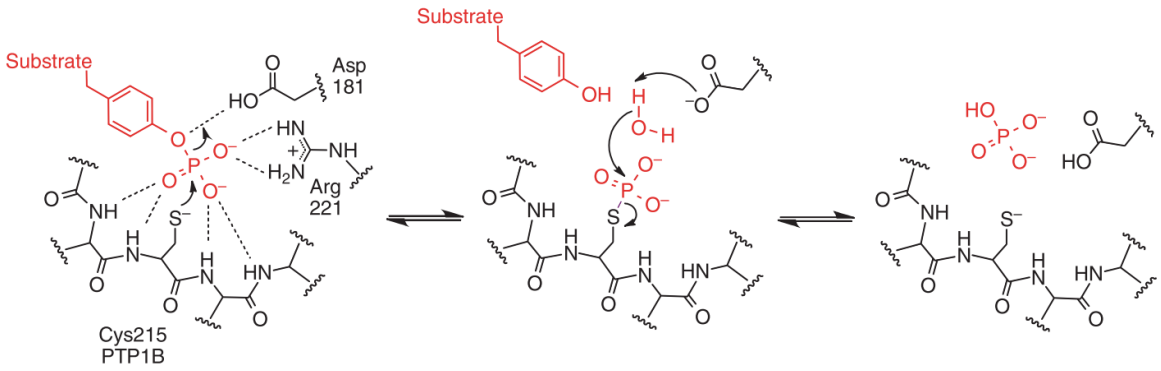


Figure 9: Catalytic mechanism of protein tyrosine phosphatases [115].

1.4.3 Methods for detection of protein phosphorylation sites

1.4.3.1 Mass spectrometry

When mass spectrometry methods first became available, the procedures were labour-intensive. With advances in both mass spectrometers and sample preparation methods, MS-based phosphoproteomics became a highly sensitive technique suitable for large-scale studies [116]. Due to technological limitations, early phosphoproteomics focused mainly on more abundant proteins and a few phosphorylation sites, using radioactive labelling and antibody-based methods in the 1960s and 1980s. Since sequencing techniques required large, highly purified samples, only the more abundant phosphoproteins could be studied. Although it had limitations for serine or threonine sites, the use of anti-phosphotyrosine (pY) antibodies in the early 1980s marked a significant advancement in detecting and analysing tyrosine-phosphorylated proteins [117]. In the late 1990s, mass spectrometry began to be used for phosphoprotein identification when tandem mass spectrometry (MS) was combined with gel-based methods for peptide sequencing and site mapping. Although peptide mass fingerprinting and MS/MS techniques allowed for more comprehensive protein mapping, research remained mostly qualitative and focused on individual proteins [117]. The adoption of enrichment techniques like immobilized metal affinity

chromatography (IMAC), titanium dioxide (TiO_2) chromatography, and stable isotope labelling for quantification (e.g., SILAC) by MS-equipped labs led to a major increase in the scale and accuracy of experiments in the 2000s. These techniques enabled global profiling of phosphorylation states in complex samples. Large-scale analyses, for example, profoundly impacted cell signalling research by quantifying thousands of phosphorylation sites in cell lines after stimulation with different molecules. The sheer size of these data sets made it more challenging to interpret biological significance, necessitating advances in bioinformatics [116]. Recent developments have aimed to miniaturize sample processing, improve MS sensitivity, and facilitate analysis of low-input samples. Phosphoproteomics is now applicable to rare cell populations and clinical samples thanks to new workflows like Phospho-SISPROT, tandem tip-based preparations, and data-independent acquisition (DIA) MS, which have significantly reduced sample requirements and preparation time. These technologies allow the regular identification and quantification of tens of thousands of phosphosites, enabling an unprecedented level of detail in cellular signalling and disease research [118]. Today, mass spectrometry-based phosphoproteomics is the premier tool for mapping protein phosphorylation globally, providing crucial information for understanding complex biological regulation and disease mechanisms. As a rapidly evolving field, phosphoproteomics continues to push the boundaries of sensitivity, throughput, and data analysis [119].

Phosphoproteomic analysis using a mass spectrometer can yield comprehensive results about the quantity and site of modification. However, there are several challenges one must confront before performing such analysis. Even though approximately 30% of proteins are phosphorylated at any given time, they might still get lost in the background of non-phosphorylated proteins. For that reason, phospho peptides should be enriched before further analysis [120], [121]. There are several platforms for enrichment available. During enrichment by chemical derivatization, phosphate is first removed by hydroxide-ion-mediated β-elimination and Michael addition of several possible affinity tags like biotin, which are then purified on an avidin column [122]. In phosphoprotein isotope-coded solid-phase tagging, a photosensitive linker is used in place of biotin, immobilized on aminopropyl glass beads. After enrichment, bound peptides are released from the beads with UV light [123]. Guanidinoethanethiol is an arginine-like tag; in the presence of barium phosphorylation, serine residues are converted into guanidinoethylcysteine. Trypsin digestion occurs at these converted sites as they mimic arginine, and peptides are recognized by increased intensities on the mass spectrometer caused by modification [122]. Rhodamine with cysteamine conjugate can be enriched on anti-rhodamine affinity columns. While the combination of β-elimination and Michael addition is sensitive and reproducible, a significant drawback of these methods is nonspecific labelling of some non-phosphorylated residues (cysteines and O-glycosylated residues) and the inability to enrich phosphotyrosine, where β-elimination is not possible [124].

However, chemical conversion, with the use of phosphoramidate phosphopeptides can be linked to iodoacetyl groups bound to solid base. However, relatively high number of chemical reactions needed leads to major loses during the whole procedure.

Converting phosphopeptides to their esters and then catching then on soluble dendrimer is very efficient, although not very practical due to lengthily protocol [125]. Next set of methods work on a principle of affinity to phosphorylated residues. Pulldown with antibodies is more feasible option for Tyr, since they are more specific than that for Ser and Thr. However, there is a limited number of such antibodies available. Some kits utilise phosphoprotein binding proteins for their purification, one must mind their high price and poor stability of proteins used [126].

Widely popular are immobilised metal affinity chromatography and metal oxide affinity chromatography. They take advantage of metal ions (Fe^{3+}, Al^{3+}, Ga^{3+}, Zr^{4+}) and oxides (Nb_2O_3, Fe_3O_4, TiO_2, $Al(OH)_3$, ZrO_2, SnO_2, HfO_2, Ta_2O_2, Ga_2O_3) respectively immobilised on solid surface with nitro acetic acid, imidodiacetic acid or tris-(carboxy-methyl)-ethylenediamine. Negatively charged phosphate readily interacts with metal ions, while ion exchange mechanism is speculated for metal oxide-phosphate interaction. There are also quite a few magnetic particles coated with metal oxides or other materials available. Under right conditions they prove to be more specific for phosphopeptides than above listed ions and oxides alone and can drastically speed up protocols, with digestion just one minute long [127].

Lately, phosphate affinity tags are gaining traction. Unlike metal-based methods, which rely on phosphates chemical properties alone, they recognize phosphate on peptides. These tags include two metal ions which interacts with phosphomonoester dianions on modified residues. The tag can be linked to biotin, acrylamide or agarose [127].

A combination of strong anion and cation exchange chromatography under right conditions can be used for crude preliminary purification of phosphoprotein from the whole sample. However, co-eluted non-phosphorylated proteins remain abundant [128], [129], [130], [131].

Once phosphopeptides are enriched, one faces another problem. Phosphorylated peptides' elution times are altered compared to their non-modified counterparts. Information gained from chromatogram can aid in peptide identification, so caution is required when analysing this data [132].

Negatively charged phosphate reduces ionization efficiency in positive ion mode, limiting their detection. Fragmentation also presents issues. Due to the labile nature of the modification, neutral loss of phosphoric or metaphosphoric acid is favoured over phosphate backbone fragmentation with CID as the fragmentation mode. The use of ETD, ECD, or HCD retains phosphate and promotes backbone fragmentation. Both of these problems have been addressed by treating the sample with phosphatases before MS analysis and comparing treated and untreated data [130], [131], [133].

1.4.4 Legumain and protein phosphorylation

Mice lacking legumain exhibit a range of phenotypic abnormalities affecting multiple organ systems. Compared to wild-type animals, their T-cell response is slower. Impaired lysosomal processing causes structural abnormalities in the kidneys' proximal tubules, leading to progressive kidney damage that includes proteinuria and reduced glomerular filtration rates. Additionally, these mice show an enlarged spleen and contain red blood cell-containing histiocytes, which are hallmarks of hemophagocytic syndrome [14].

Changes in protein phosphorylation are linked to some observed effects in legumain knockout mice [134], [135], [136], [137], [138], [139], [140], [141], including increased levels of EGFR in the kidneys [142] and activation of STAT3 signalling [143]. Additionally, phosphorylation of the PDH domain of podocin influences its dimerization, which affects podocyte filtration function in mouse kidneys [135]. In models of renal interstitial fibrosis, p38 MAPK phosphorylation was found to be increased by nuclear receptor Nr4a1, worsening renal fibrosis [144]. The JAK/STAT signalling pathway, which depends on phosphorylation for activation, has been identified as a key driver in myelofibrosis, which is directly connected to extramedullary haematopoiesis [145]. Increased phosphorylation of STAT1 has also been observed in patients with hemophagocytic

lymphohistiocytosis [146]. Hepatomegaly occurs after activation of NFR2, a transcription factor that regulates antioxidant expression and is activated by AKT kinase [147].

Since the exact mechanism by which legumain causes the described changes in the phenotype of knock-out mice is unknown, we investigated whether protein phosphorylation is the link connecting legumain to these changes.

Chapter 2

Aims and Hypotheses

Mice with knocked-out gene for legumain don't exhibit extreme phenotype, they have no behavioural aberrations and have normal viability and fertility. However, there are multiple differences when compared to their wild-type counterparts. Most notably, kidneys have altered morphology and function. One can observe an accumulation of endosomal substrates in proximal tubular cells, leading to endosome enlargement and hyperplasia of these cells. Morphological changes in the kidneys also include interstitial fibrosis and the formation of glomerular cysts. The function of the kidneys changes so that the glomerular filtration rate is lowered, which results in heightened levels of creatinine in plasma and proteinuria. Knockout mice develop some symptoms of hemophagocytic syndrome; enlarged histiocytes containing red blood cells, haematocrit values decline with aging, and the spleen is enlarged with active extramedullary haematopoiesis. Since there appears to be no direct connection between legumain and observed changes in legumain knock-out mice, we probed the possibility legumain influences cell signalling through protein phosphorylation. Some of the reported modifications of legumain knock-out mice have been linked to changes in protein phosphorylation, including higher levels of EGFR in the kidneys, and activation of STAT3 signalling. Per these previous observations, we tested the following hypothesis:

- Legumain deficiency alters phosphorylation-based signalling in murine organs
- Legumain deficiency affects function of proteins involved in regulation of DNA transcription and translation.
- Differences in phosphorylation in legumain knockout mice affect expression of proteins involved in pro-inflammatory response
- Activation of the NF-κB pathway is the main trigger of pro-inflammatory response in legumain deficient mice.

As legumain has a known role in some pathogenic states, especially in tumours environment and inflammation, further understanding into signalling mechanics connected to legumain's function will help develop treatment methods to ameliorate negative effects associated with legumain and possibly employ its positive effects to our good.

Chapter 2

Materials and Methods

3.1 Materials

3.1.1 Chemicals

- ABC (ammonium bicarbonate) (Sigma-Aldrich, 9830)
- ACN (acetonitrile) (Fisher chemicals, LS120-1)
- Bio-Rad Protein Assay Dye Reagent Concentrate (Bio-Rad, 5000006)
- BSA (Promega, R396D)
- DMSO (Dimethyl sulphoxyde) (Merck, 1.02931.1000)
- DPBS, (Sigma, D8662)
- DTT (Sigma-Aldrich, 43815-5G)
- EDTA (Fluka, 3677)
- FA (formic acid) (Fisher scientific, A117-50)
- FBS (foetal bovine serum) (Gibco, 10500-064)
- Glutamax (Gibco, 35050061)
- IAM (iodoacetamide) (Amersham Biosciences, RPN6302)
- Methanol (VWR, 414814)
- NaCl (Fisher Chemicals, S/3160/60)
- Igepal CA-630 (NP-40 substitute) (UBS, 19628)
- Penicillin/Streptomycin (Sigma-Adrich, P4333)
- Prestained Protein Ladder PageRuler (Thermo Scientific, 26620)
- Phosphatase arrest cocktail (G-Biosciences, 786-647)
- Protease inhibitor cocktail (Sigma, P8340-1ML)
- SDS (Thermo Fisher Scientific, BP 2436-1)
- Sequencing-grade porcine trypsin (Promega, PR_V5111)
- TEMED (Sigma, T9281)
- TRIS (Sigma, T6066)
- Triton X-100 (Sigma, 93424)
- Trypan Blue (Sigma-Aldrich, T8154-100ml)
- Tween-20 (Merck, P9416)
- Western Blotting Detection Reagents ECL™ (GE Healthcare, RPN2106)
- Urea (VWR, 97063-802)

3.1.2 Commercial kits

- Novex^{tm} 8-16 %, Tris-Glycine Plus WedgeWell^{tm} Gel (Invitrogen, XP08160BOX)
- Lenti-X lentiviral expression system (Invitrogen)

- QIAGEN RNeasy Mini Kit (Qiagen)

3.1.3 Buffers, solutions and media

Lysis buffer 1
20 mM TRIS/HCl pH 7.4
150 mM NaCl
1 mM EDTA
1% (v/v) NP-40
1x Protease inhibitor cocktail
1x Phosphatase inhibitor cocktail

Lysis buffer 2
50 mM TRIS pH 7.8
150 mM NaCl
1 mM EDTA
1% NP-40
0.1% SDS
0.5% Na-deoxycholate

Laemmli buffer (4x)
240 mM Tris-Cl
8% SDS
40% glycerol
20% β-mercaptoethanol
0.04% bromphenol blue

Solution for SDS-PAGE stacking gel
0.625 mL 40% acrylamide-bisacrylamide
1.25 mL 0.5 M Tris/HCl (pH 6.8)
3.075 mL H2O
0.1 mL 10% SDS
0.02 mL 10% APS
0.01 mL TEMED

Solution for SDS-PAGE (10%) resolving gel
2.5 mL 40% acrylamide-bisacrylamide
2.5 mL 1 M Tris-HCl (pH 8.8)
4.9 mL H2O
0.1 mL 10% SDS
0.04 mL 10% APS
0.02 mL TEMED

SDS-PAGE loading buffer (6x)
375 mM Tris-HCl
6% SDS
48% glycerol
100 mM DTT
0.03% bromophenol blue
pH 6.8

SDS-PAGE buffer
- 25 mM Tris
- 192 mM L-Gly
- 0.1% SDS

TBS
- 20 mM Tris
- 500 mM NaCl
- pH 7.5

TBST
- TBS
- 0.05% (v/v) Tween 20

Western transfer buffer
- 25 mM Tris
- 192 mM L-Gly
- 15 mM SDS
- 10% methanol

Staining solution
- 0.5% Coomassie Brilliant Blue R-250
- 50% methanol
- 10% acetic acid

Destaining solution
- 30% ethanol
- 10% acetic acid

Reduction buffer
- 10 mM DTT
- 25 mM ABC

Alkylation buffer
- 55 mM IAM
- 25 mM ABC

Trypsinisation solution
- 25 mM ABC
- 1 µg sequencing-grade porcine trypsin

Extraction solution
- 50% ACN
- 5% FA

Elution solution
- 60% ACN
- 0.1% FA

Phosphopeptide wash solution 1

50% ACN
1% TFA

Phosphopeptide wash solution 2
35% ACN
1% TFA

Phosphopeptide elution solution
150 mM NH4OH
25% ACN

Phosphopeptide reconstitution solution
1 M glycolic acid
80% ACN
5% TFA

Culture media
RPMI
10% FBS
1% penicillin/streptomycin
1% glutamine

PBS
6.6 mM Na2HPO4
13.4 mM NaH2PO4
150 mM NaCl
pH 7.4

Freezing media
complemented RPMI
40% FBS
10% DMSO

3.1.4 Plasmids

pLVX-IRES-Puro lentiviral expression vector (Clontech)

3.1.5 Antibodies

- GAPDH (Cell Signaling Technology, 97166S)
- IkBα (Cell Signaling Technology, 9241)
- IL-10 (Abcam, ab9722)
- YBX-1 (Abcam, ab12148)
- pSer 100 YBX-1 (Abcam, ab138654)
- F4/80 (Abcam, ab6640)
- HRP-conjugated goat anti-rabbit IgG (#111-035-045, Jackson Immunoreshearch)
- Alexa Fluor 555 anti-rat IgG (Invitrogen, A48270)

3.1.6 Primers

gene name *Homo sapiens*	forward	reverse
Nfkbia	5'-TCCTGAAGGCTACCAACTACA-3'	5'-CATTGACATCAGCACCCAAG-3'
Nfkb1	5'-CAGGAGACGTGAAGATGCTG-3'	5'-AGTTGAGAATGAAGGTGGATGA-3'
Nfkb2	5'-GAGAGAAGCCGCAACCAG-3'	5'-CGGCTCTGTCTAGTGGCT-3'
Il10ra	5'-TTCCGAGAGTATGAGATTGCC-3'	5-AGCGACAGATGGTTTCACC-3'
Il1b	5'-CAGCCAATCTTCATTGCTCAAG-3'	5'-GAACAAGTCATCCTCATTGCC-3'
Ccl5	5'-GCTGTCATCCTCATTGCTACT-3'	5'-TGCCACTGGTGTAGAAATACTC-3'
5LegEcoR1	5'-TAATTGAATTCATGGTTTGGAAAGTA-3'	
3LegXba1	5'-TTAATTTCTAGATCAGTAGTGACCAAG-3'	

gene name *Mus musculus*	forward	reverse
Il10ra	5'-TTTGCTCCCATTCCTCGTC-3'	5'-GGCTTCAAACCACACATAGGA-3'
Ccl2	5'-CATCCACGTGTTGGCTCA-3'	5'-AACTACAGCTTCTTTGGGACA-3'
Ccl3	5'-CCTTGCTGTTCTTCTCTGTACC-3'	5'-CGATGAATTGGCGTGGAATC-3'
Ccl5	5'-GCTCCAATCTTGCAGTCGT-3'	5'-CCTCTATCCTAGCTCATCTCCA-3'
Cxcl1	5-CCAAACCGAAGTCATAGCCA-3'	5'-GTGCCATCAGAGCAGTCT-3'
Ikbke	5'-AGTCTTTGCCTGATTCTGAGC-3	5'-CATCAGTATGCCACAGGTAGT-3'
Il1f6	5'-CGATGAGCTGCCTGTTCTG-3'	5'-GGTTGTACCACTCTTCTTGTGA-3'
Nfkb2	5'-CGGAAAGAAGTCGGAACCAG-3'	5'-TCGTAGCAATTGTCCATATCTCTG-3'
Serpina1d	5'-GAAAGATAGTTGAGGCGGTGA-3'	5'-CTCAGGATCGAATGGCTGTT-3'

3.1.7 Mammalian cell lines

- HL-60, human leukaemia cell line ATCC (98070106)

3.1.8 Equipment

- 1.5 mL and 2 mL micro tubes (Eppendorf)
- 15 mL and 50 mL centrifuge tubes (BD Biosciences)

- 96-well microtiter cell culture black plate with transparent bottom (Corning)
- Amicon 0.5 Ultra Centrifugal filters (Millipore)
- Automatic pipettes (Eppendorf)
- Cell line incubator (Binder)
- Centrifuges (Eppendorf)
- Concentrator 5301 (Eppendorf)
- Empore/C18 disks (Varian)
- Gel imaging system G:Box (Syngene)
- G: BOX Chemi XRQ gel dock system (Syngene)
- Laminar air flow cabinet for cell culture BSB 4A (Gelaire Flow Laboratories)
- Nitrocellulose membrane (GE healthcare)
- Orbitrap LTQ Velos mass spectrometer (Thermo Fisher Scientific) coupled to EASY-nanoLC II HPLC unit (Thermo Fisher Scientific)
- pH-meter MP 220 (Mettler Toledo)
- Power supply Western Lightning Volt™ Power Supply OSP-300 (Thermo Scientific)
- PHOS-Select iron-affinity gel (Merck)
- SigmaPrep spin columns (Sigma-Aldrich)
- Tecan Infinite M1000 plate reader (Tecan)
- TiO2 beads (Titansphere Phos-TiO 10 µm, GL Science)
- Vertical Electrophoresis Lab Apparatus MiniProtean Tetracell (Biorad)
- Western transfer apparatus TE22 Mighty Small Transphor Unit (GE Lifesciences)

3.2 Methods

3.2.1 Mice

3.2.1.1 Mice on the C57/BL6 background were maintained in accordance with national regulations authorized by the government's Ethical Committee and the Veterinary Administration of the Republic of Slovenia (VARS), with permit number U34401-40/2020/3. Legumain knockout mice were provided by the Thomas Reinheckl lab in Freiburg, Germany. Three mice each were kept in standard cages (Techniplast 1264C Eurostandard Type II) with bedding made of corncobs (Rehofix MK 2000). Access to food (Altromin 1324) and water was provided ad libitum. Kidneys and livers from both wild-type (WT) and AEP$^{-/-}$ (n= 3 for each genotype) mice were immediately frozen before further processing. Organ lysate preparation

Organs (kidneys and liver) were weighed and homogenized with 20 µL of lysis buffer containing a phosphatase inhibitor cocktail (G-Biosciences), using a Potter-Elvehjem PTFE homogenizer and drawn through 20G needles on ice. Lysates were shaken on ice for an additional 30 minutes. The samples were then centrifuged at 12,000 g for 30 minutes at 4 °C. Supernatants were collected, and total protein concentration was measured using the Bradford assay, which relies on the shift in absorption maximum of Coomassie Brilliant Blue G-250 from 465 nm to 595 nm after binding to aromatic amino acid residues. Bovine serum albumin (BSA) at concentrations of 1–16 µg/mL served as an external standard. A clear, flat-bottom 96-well plate was pipetted with the appropriate volumes of BSA stock solution and diluted samples, followed by the addition of dH2O to reach a final volume of 160 µL. Then, 40 µL of a 5× BioRad Reagent solution was added to each well. An Infinite

M1000 Tecan plate reader was used to measure absorbance at 595 nm (A595). Standard curve was generated from BSA measurements, and sample concentrations were calculated accordingly.

3.2.2 Cell culture

3.2.2.1 Growing and passaging of cells

HL-60 cells were purchased from ATCC (98070106) and were grown in RPMI 1460 medium with 20 % foetal bovine serum (FBS), 1 % penicillin/streptavidin (all Sigma-Aldrich), and 1 % GlutaMAXTM (Invitrogen). Cells were grown in a humidified incubator at 37 °C and 5 % CO2. When cells reached density of 10^6 / mL, they were centrifuged at 300 rcf for 5 min, washed twice with PBS and diluted to 0.2 * 106 in complemented RPMI and finally transferred to a new flask.

3.2.2.2 Legumain overexpression in HL-60 cells

Legumain cDNA was amplified using Q5 Hot Start DNA Polymerase (New England BioLabs) with 5' LegEcoRI and 3' LegXbaI primers, which also introduced EcoRI and XbaI restriction sites. The amplified DNA was then ligated into the pLVX-IRES-Puro lentiviral expression vector (Clontech). A control vector was created with a scrambled DNA sequence inserted.

Lentiviruses were produced using the Lenti-X lentiviral expression system (Invitrogen), following the manufacturer's instructions. The expression vector was transfected into Lenti-X 293T cells along with the Lenti HTX Packaging Mix to produce infectious virions. The virions containing pLVX-IRES-Puro with legumain and scrambled sequences were then used to infect HL-60 cells.

3.2.2.3 Cell lysate preparation

Lysates were prepared in lysis buffer containing phosphatase inhibitor cocktail (G-Biosciences) and protease inhibitor cocktail (Sigma). Cells were manually homogenized with a syringe and a 20 G needle on ice, then shaken on ice for an additional 30 minutes. Samples were then centrifuged at 12 000 rfc for 30 minutes at 4 °C. Supernatants were collected, and the total protein concentration was determined using the Bradford assay.

3.2.3 Mass spectrometry

3.2.3.1 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis was used for protein separation before preparing samples for mass spectrometry. Equal amounts of proteins were boiled at 95 °C for 5 minutes in the presence of 6× loading buffer. Proteins were separated on Novex WedgeWell 8-16% Tris-Glycine gel, 1.0 mm thick, with 10 wells (XP08160BOX, Invitrogen) in SDS-PAGE buffer. Pre-stained protein ladder, PAGE Ruler, was used as a standard (Invitrogen). Electrophoresis was run at a constant current of 35 mA per gel until the front reached the bottom. After separation, the gel was stained with Coomassie Brilliant Blue for 30 minutes and destained with 30% ethanol/10% acetic acid.

3.2.3.1.1 In-gel digestion protocol

For the control experiment without phosphopeptide enrichment, we used an in-gel procedure. The destained gel was cut into individual lanes, and each lane was divided into six pieces. Each piece was cut into approximately 1 mm^3 fragments and transferred into Eppendorf Lo-bind tubes. First, the samples were thoroughly destained with two incubations using 150 µL of 25 mM ammonium bicarbonate (ABC) in 50% acetonitrile (ACN), shaking at 1200 rpm for 30 minutes at room temperature. Then, the gel pieces were dehydrated in 100% ACN and vacuum dried using a SpeedVac. Cysteine residues were reduced by incubating the samples in 150 µL of 10 mM dithiothreitol (DTT) in 25 mM ABC at 56 °C for 30 minutes. Next, 150 µL of 55 mM iodoacetamide (IAM) in 25 mM ABC was added in the dark and incubated for 30 minutes at room temperature. To quench any remaining IAM, 150 µL of 20 mM DTT in 25 mM ABC was added for 30 minutes at 37 °C. Finally, the gel pieces were washed again with 150 µL of MS-grade water, followed by two additional washes with 150 µL of 100% ACN until they were completely dry.
The samples were digested with sequencing-grade modified porcine trypsin (Promega) at a final concentration of 1 µM in 0.1 M ABC buffer pH 7.8. Using 1 µg of trypsin per sample, the gel pieces were protected from drying out during digestion by adding 100 µL of 0.1 M ABC buffer pH 7.8 and incubating overnight at 37°C.

Peptides produced from trypsin digestion were extracted with 150 µL of 50% acetonitrile in 5% formic acid, agitating for 30 minutes at room temperature. The extracted peptide solution was transferred to a new tube, and the extraction process was repeated two more times. The pooled peptides were concentrated in a SpeedVac to approximately 100 µL to remove all ACN present in the samples. The peptides were desalted through a stack of three C18 discs (Empore) in a 200 µL pipette tip. The C18 columns were preconditioned with 100% methanol and 100% ACN/0.1% FA. The bound peptides were washed with 0.1% formic acid and eluted with 60% ACN/0.1% FA. ACN in the eluted samples was removed with vacuum concentration until the final volume was 20 µL.

3.2.3.1.2 LC-MS/MS analysis

The samples were analysed on an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) connected to an EASY-nanoLC II HPLC system. The peptide samples were diluted in solvent A (0.1% FA (v/v) in water) and loaded onto a C18 trapping column (Proxeon EASY-ColumnTM, Thermo Fisher Scientific), then separated on a C18 PicoFritTM AQUASIL analytical column. The peptides were eluted over a 90-minute linear gradient of 5 to 50% solvent B (100% ACN in 0.1% FA) at a flow rate of 300 nL/min. Full MS spectra were acquired using an Orbitrap mass analyser at a resolution of 30 000 and a mass range of 300-2 000 m/z. The nine most intense ions from each MS spectrum were selected for fragmentation with HCD. MS/MS spectra were collected at a resolution of 7 500. For MS/MS fragmentation, only precursor ions with a charge greater than one were selected. Dynamic exclusion was enabled with a repeat count of one, a 30-second repetition time, and a 30-second exclusion duration. Data was captured using Thermo Xcalibur version 2.2 SP1.48 (released on August 11, 2012).

3.2.3.1.3 Data analysis

Tandem mass spectra were analysed with MaxQuant (version 1.6.17.0, Max-Planck Institute for Biochemistry) using the integrated Andromeda search engine [148]. The searches were conducted against the mouse reference proteome available at UniProt (UniProtKB, *Mus musculus*, reviewed database with 16 996 entries, as of August 2019). Peptides generated by the protease were searched using trypsin-specific cleavage sites with

a maximum of two missed cleavages. Fixed modification was set to carbamidomethylation of cysteine residues (+57.0215 Da), while oxidation of methionine (+15.9949 Da) and acetylation of protein N-termini (+42.0106 Da) were set as variable modifications. The precursor ion and fragment ion mass tolerances were set to 4.5 ppm and 0.5 Da, respectively, and Label-free quantification was enabled. A reverse database search was performed with a false discovery rate (FDR) set to 1%. The resulting data were further processed in Perseus (v.1.6.14.). Reverse decoy matches, potential contaminants, and proteins detected with fewer than two peptides were filtered out before a $\log_2$ transformation was applied. Proteins identified in at least three biological replicates in at least one group (WT/AEP$^{-/-}$) were selected for further analysis, and missing values were imputed based on a downshifted normal distribution with default settings. For statistical analysis, Student's two-sample t-test was used for group comparisons, with default parameters.

3.2.3.2 In-solution digestion protocol

For phosphoprotein analysis, we used an in-solution sample preparation procedure. We used 2 mg of samples with equalized concentrations for the reaction. Urea was added to the samples at a final concentration of 6 M. DTT (10 mM final concentration) was used to reduce the samples, which were incubated at room temperature for 1 hour. Reduced samples were alkylated with 32 mM iodoacetamide in the dark at room temperature for one hour. The reaction was quenched by adding DTT to a final concentration of 38 mM and incubating for an additional hour at room temperature. Proteins were digested using a 1:100 (w/w) mixture of sequencing grade Lys-C/Trypsin (Promega) for 3 hours at 37 °C, then diluted to a urea concentration of 1 M with 50 mM Tris-HCl, 0.1% SDS, pH 7.4, and digested overnight at 37 °C. After digestion, the reaction was quenched with 0.5% trifluoroacetic acid (TFA), and the samples were centrifuged at 12 000 rcf for 15 minutes to remove debris. The samples were then concentrated using C18 reverse-phase columns (8B-S001-DAK, Phenomenex). Columns were first equilibrated with three volumes of methanol, 0.1% FA in ACN, and 0.1% FA in water. After loading, the samples were washed with three volumes of 0.1% FA, eluted with 400 µL of 60% ACN/0.1% FA, and concentrated using a SpeedVac. Concentrated samples were reconstituted in 0.1% TFA and 50% ACN to an approximate protein concentration of 4 µg/µL. Subsequently, phosphopeptide enrichment was performed using PHOS-Select Iron Affinity Gel (Sigma-Aldrich) (40 µL per sample). The PHOS-Select gel was placed in SigmaPrep spin columns (SC1000-1KT, Sigma-Aldrich), washed three times with 0.1% TFA and 50% ACN. The samples and gel were mixed in SigmaPrep columns and incubated for one hour at room temperature. After incubation, the flow-through was collected as the unbound fraction, and the gel was washed twice with 200 µL of 1% TFA and 35% ACN. The eluate from this process was combined with the unbound fraction and concentrated with a SpeedVac for further phosphopeptide enrichment using Titansphere Phos-TiO2 beads (5010-21315, GL Sciences). The iron affinity gel was washed with 250 µL of LC-MS grade water, and phosphopeptides were eluted with 150 mM NH_4OH and 25% acetonitrile, shaking for 5 minutes. The eluted phosphopeptides were concentrated using a SpeedVac and then transferred to C18 StageTips (20 discs, 3M Empore). The StageTips were equilibrated as described for the columns. Elution from StageTips was performed successively with 100 µL of elution solution containing 0%, 2%, 5%, 8%, 10%, and 40% ACN, with 0.1% FA at 1 400 rcf for 5 minutes. The concentrated unbound fraction from the iron affinity gel was reconstituted in 1 M glycolic acid, 80% ACN, and 5% TFA, reaching an approximate concentration of 4 µg/µL. TiO_2 beads were washed three times with 80% ACN and 0.4% TFA. The beads were added to the samples (1:8, w/w) and mixed for one hour at room

temperature. Following incubation, the samples were spun down and the beads transferred to C18 StageTips. The samples were washed first with 100 µL of 0.5% TFA, 30% ACN, and again with 0.4% TFA, 80% ACN. Sequential elution of phosphopeptides from the StageTips was carried out with 0%, 2%, 5%, 8%, 10%, and 40% acetonitrile with 15% NH_4OH (v/v). All eluted samples were concentrated in a SpeedVac and reconstituted with 0.1% FA in water before LC-MS analysis. MS/MS analysis was performed as described above.

3.2.3.3 Phosphoproteomic data analysis

Raw data files were processed using MaxQuant software (version 1.6.17.0, Max-Planck Institute for Biochemistry) with an integrated Andromeda search engine. The files were searched against the reviewed Swiss-Prot Mus musculus database containing 16 996 entries, as of August 2019. Aside from default parameters (described above), STY phosphorylation was added as a variable modification, and label-free quantification (LFQ) was enabled. Results from MaxQuant were further analysed using Perseus software (version 1.6.14). Potential contaminants, reverse hits, and phosphosites with a localization probability below 0.75 were removed prior to further analysis. Only phosphosites detected in at least two of three technical replicates and two of three biological replicates were included in the Student's t-test. Permutation-based FDR was used for truncation, with FDR set to 1%.

3.2.4 Western transfer and immunoblotting

3.2.4.1 SDS-PAGE

The setup for casting homemade gels involved 8 mL of the SDS-PAGE separating gel solution, which was covered with 0.5 mL of isopropanol for 30 minutes. After 30 minutes, the isopropanol was removed, and the gel was rinsed with distilled water. Then, 2 mL of the SDS-PAGE stacking gel solution was poured over it, and a comb was inserted. After another 30 minutes, the comb was removed, and the gel was transferred into the tank of the electrophoresis apparatus. Equal amounts of proteins were boiled at 95 °C for 6 minutes in the presence of 6x loading buffer, and 20 µg of each sample was loaded onto the gel. Electrophoresis was run in SDS-PAGE buffer at a constant current of 30 mA per gel until the standards separated. A pre-stained protein ladder, PAGE Ruler, was used as a standard.

3.2.4.2 Western blotting transfer and immunological detection

Western transfer was performed using the Bio-Rad Trans-Blot Turbo system according to the manufacturer's instructions for 10 minutes at 25 A onto nitrocellulose membranes. After transfer, the membranes were briefly washed in TBS buffer with 1% Tween-20 (TBS-T) and blocked with 5% skimmed milk in TBS-T for 1 hour at room temperature. Subsequently, the primary antibody solution in TBS-T BSA was applied to the membrane and incubated overnight at 4 °C. The membrane was then washed four times for 15 minutes in TBS-T before incubation with secondary antibodies for 1 hour at room temperature. Following this, the membrane was washed again four times for 15 minutes in TBS-T. Chemiluminescence detection was performed using the Amersham ECL Western Blotting Detection Reagents kit, following the manufacturer's instructions.

3.2.5 Real-Time quantitative polymerase chain reaction (RT-qPCR)

Total RNA from HL-60 cells with stable legumain expression and controls with a mock plasmid, as well as from mice kidneys and livers (wild-type and legumain knockout), was isolated using the QIAGEN RNeasy Mini Kit (Qiagen) following the manufacturer's instructions. The RNA concentration was measured with a NanoDrop One/One spectrophotometer (Thermo Scientific). One milligram of total RNA was used for the reverse transcription reaction with LunaScript RT Supermix (New England Biolabs). For the RT-qPCR reaction, 10 ng of cDNA per well was used with KAPA SYBR FAST (Merck) master mix. The primer concentration was 100 nM. RT-qPCR was performed on a Stratagen MX300P (Agilent) with the following settings: initial denaturation at 95 °C for 3 minutes; then 40 cycles of denaturation for 3 seconds at 95 °C and primer annealing and extension for 20 seconds at 60 °C. Data was collected at the end of each cycle. Melting curves were generated after the cycling. A set of housekeeping genes was tested for stability across all samples using the GNorm method [149], with GUSB identified as the most stable reference gene. Results were analyzed in R using the pcr package (version 1.2.2) [150] with the double delta Ct method [150]. Primers were purchased from Integrated DNA Technologies.

3.2.6 Immunohistology

After dissection, mouse kidneys were frozen in liquid nitrogen and kept at -80 °C till use. The organs were sliced into 5 µm slices with a cryostat (SLEE medical) and mounted onto glass slides. The samples were fixated with 4 % cold paraformaldehyde and 100 % cold acetone. Anti-F4/80 primary antibodies (1:100 dilution, Abcam, ab6640) for staining macrophages and Alexa Fluor 555 anti-rat IgG (1:200 dilution, Invitrogen, A48270) secondary antibodies were used for immunofluorescence labelling. Anti-fade mountant with DAPI (Thermo Fisher, P36931) was used for nuclear counterstaining, before coverslips were mounted onto glass slides. A fluorescence microscope (Olympus IX 81) was used to image the tissue samples and the Olympus cellSens Dimension (ver. 3.2) imaging software for image analysis. Five sections per sample were selected, and the proportion of the stained regions was measured after adjusting the colour threshold and normalizing the target protein signal to the DAPI signal using ImageJ software (ver. 1.49).

Chapter 4

Results

4.1 Phosphoproteomics analysis of legumain knock out mice kidney and liver

4.1.1 Phosphoproteomic analysis of mouse kidney

Altogether, we identified 3 183 phosphopeptides associated with 1 681 proteins. Most of the peptides, 3 827 (91.5%), were singly phosphorylated, 332 (7.9%) were doubly phosphorylated, and 245 were phosphorylated on at least three residues. Since one phosphosite can be identified multiple times, for example: once on a singly phosphorylated and one on a doubly phosphorylated peptides, number of identified phosphosites doesn't directly translate into number of identified phosphopeptides. The majority of the phosphorylation sites, 82.0%, were on serine residues, while 16.7% and 1.3% were on threonine and tyrosine residues, respectively. In total, we detected 5 082 phosphosites in the kidneys, with 3 089 having peptide identification confidence of at least 99% and site localization confidence greater than 75% (as recommended in [151]). Among these, 31 phosphosites were differentially regulated between knock-out and control groups. Of these, YBX1 and Srrm2 were the only ones to be multiphosphorylated (Figure 10).

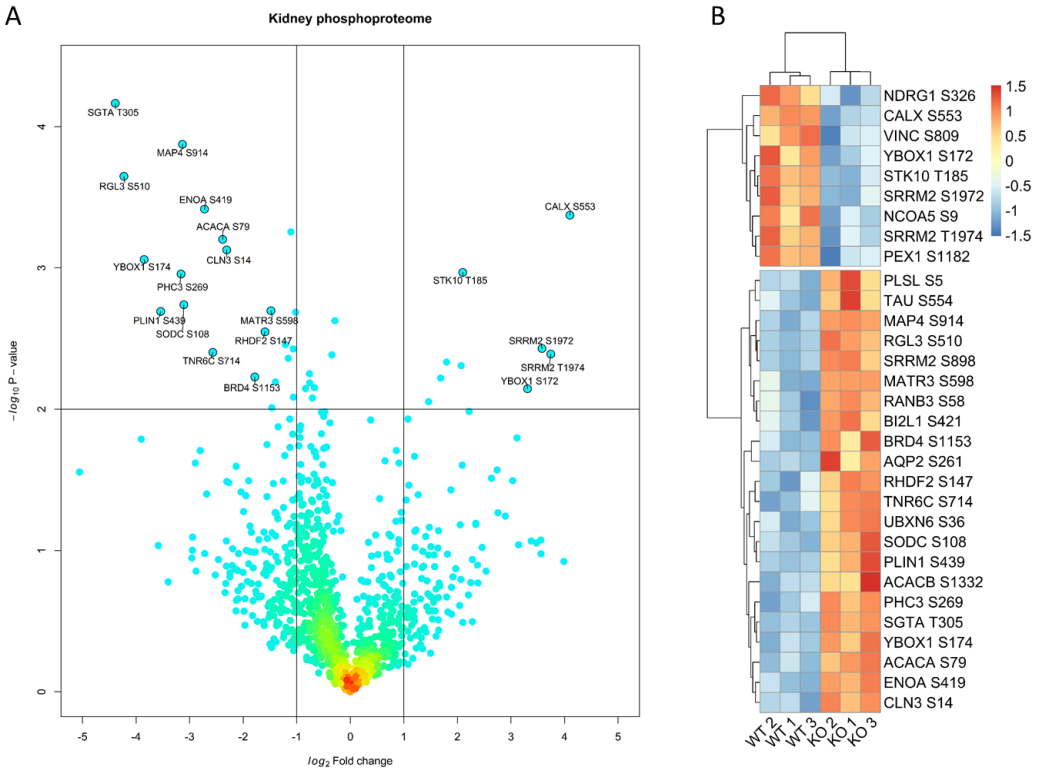


Figure 10: **(A)**: Volcano plot showing phosphosites over-represented (top left) and under-represented (top right) in legumain null C57/BL6 mice kidneys. Some significantly dysregulated phosphosites are labelled with gene names and modified amino acid residues. Horizontal lines denote a 2-fold change in abundance, horizontal line denotes a p-value of 0.01. **(B)**: Heat-map constructed with pheatmap package in R shows under-represented phosphorylation sites in AEP $^{-}/^{-}$ kidneys with blue and over-represented ones with orange. WT 1-3 are wild-type mice and KO 1-3 are knock-out mice.

To determine whether any proteins that are statistically significantly different from the whole proteome, which is not enriched for phosphopeptides, are also dysregulated in the phosphoproteome, we compared our phospho-enriched datasets with control experiments (non-enriched proteome). None of the proteins that were differentially regulated in the whole proteome matched those that were differentially phosphorylated. This confirms that the changes we observed are solely due to alterations in phosphorylation.

4.1.1.1 Prediction of subcellular location of identified phosphoproteins

The subcellular location of the identified dysregulated phosphoproteins was predicted using DeepLoc 2.0 [152]. This tool predicts protein subcellular localization based on homology with proteins in curated databases and sorting signal information. Half of the dysregulated proteins in the kidneys are located in the cytosol, and nearly a fifth are nuclear proteins. An additional 10% of kidney proteins are found either in the nucleus or cytoplasm. Other subcellular locations make up smaller fractions (Figure 11).

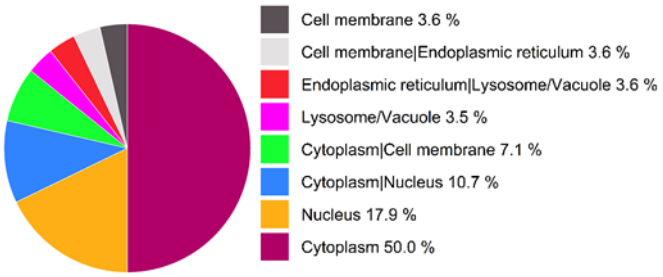


Figure 11: Pie chart showing predicted subcellular locations of identified phosphorylated proteins as identified by DeepLoc 2.0 algorithm in the kidneys.

4.1.1.2 Prediction of interacting kinases based on kinase recognition motifs

The Motif-X algorithm within the MEME suite was used to predict kinase-substrate relationships in identified phosphoproteins. We detected motifs phosphorylated by mitogen-activated protein kinases (MAPKs) (.pSP.) (686), AGC family kinases (RR.pS (66), and R..pS (117). Motifs pSD.E (72) and pS..D/E (79) are recognized by acidic kinases such as casein kinase I and II (CK1 and 2). In dysregulated phosphopeptides, the proline-rich .SP. motif was the most abundant (Figure 12).

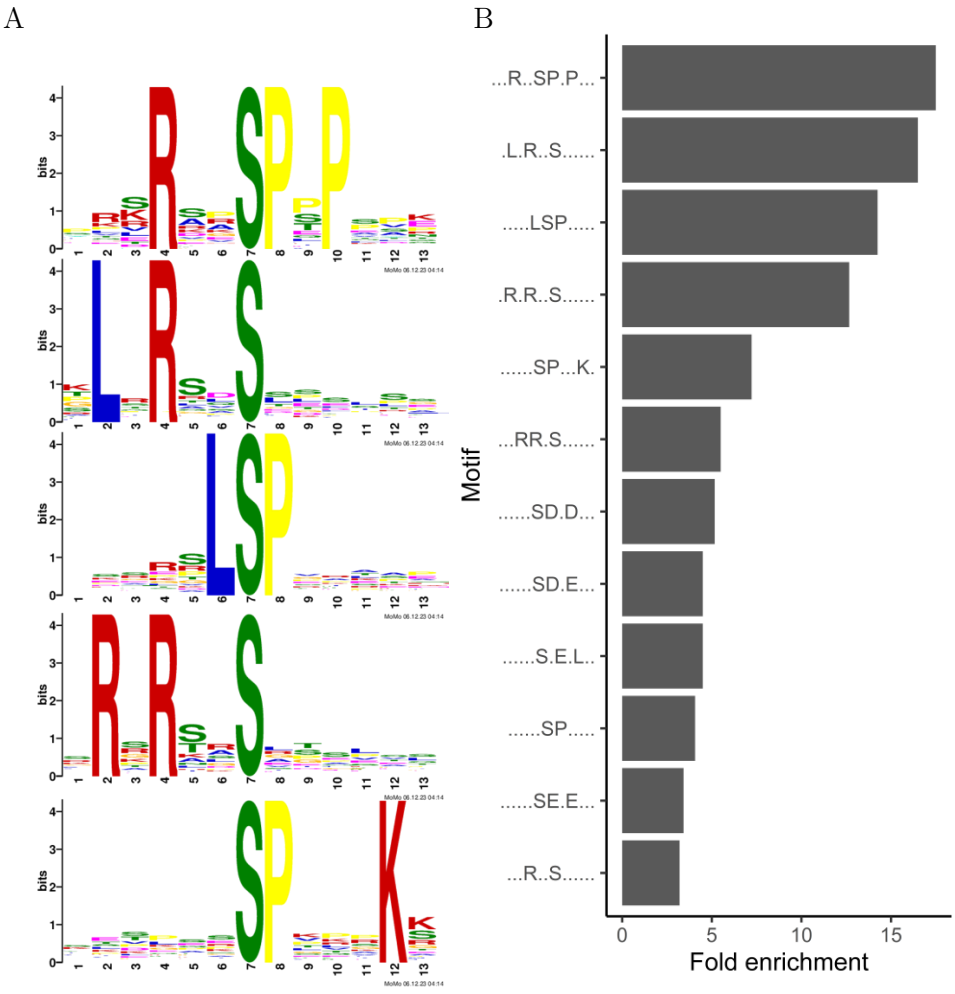


Figure 12: Motif analysis of identified phosphosites in C57/BL6 mouse kidneys with their fold enrichment **(B)**: total number of background peptides with central phosphorylation/number of peptides that match the motif. **(A)**: Ice logos of five most enriched motifs are shown.

4.1.1.3 Gene Ontology (GO) term analysis

G:profiler [153] was used to annotate differentially phosphorylated proteins in our dataset. The proteins were classified into biological process (BP), molecular function (MF), cellular component (CC), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The most common MF terms included GTPase binding and protein binding. For BP, cellular response to stress and malonyl-CoA biosynthetic process were the most enriched terms, while for CC, cytosol and cell junction were the most enriched localizations. Fatty acid biosynthesis was the most overrepresented KEGG pathway. Among upregulated phosphoproteins, the terms acetyl-CoA carboxylase activity, malonyl-CoA biosynthetic process, exocytic vesicle, and fatty acid biosynthesis were enriched in MF, BP, CC, and KEGG categories, respectively. For downregulated phosphoproteins, GTPase binding was enriched in MF, and dense plaque of desmosome was enriched in CC (Figure 13).

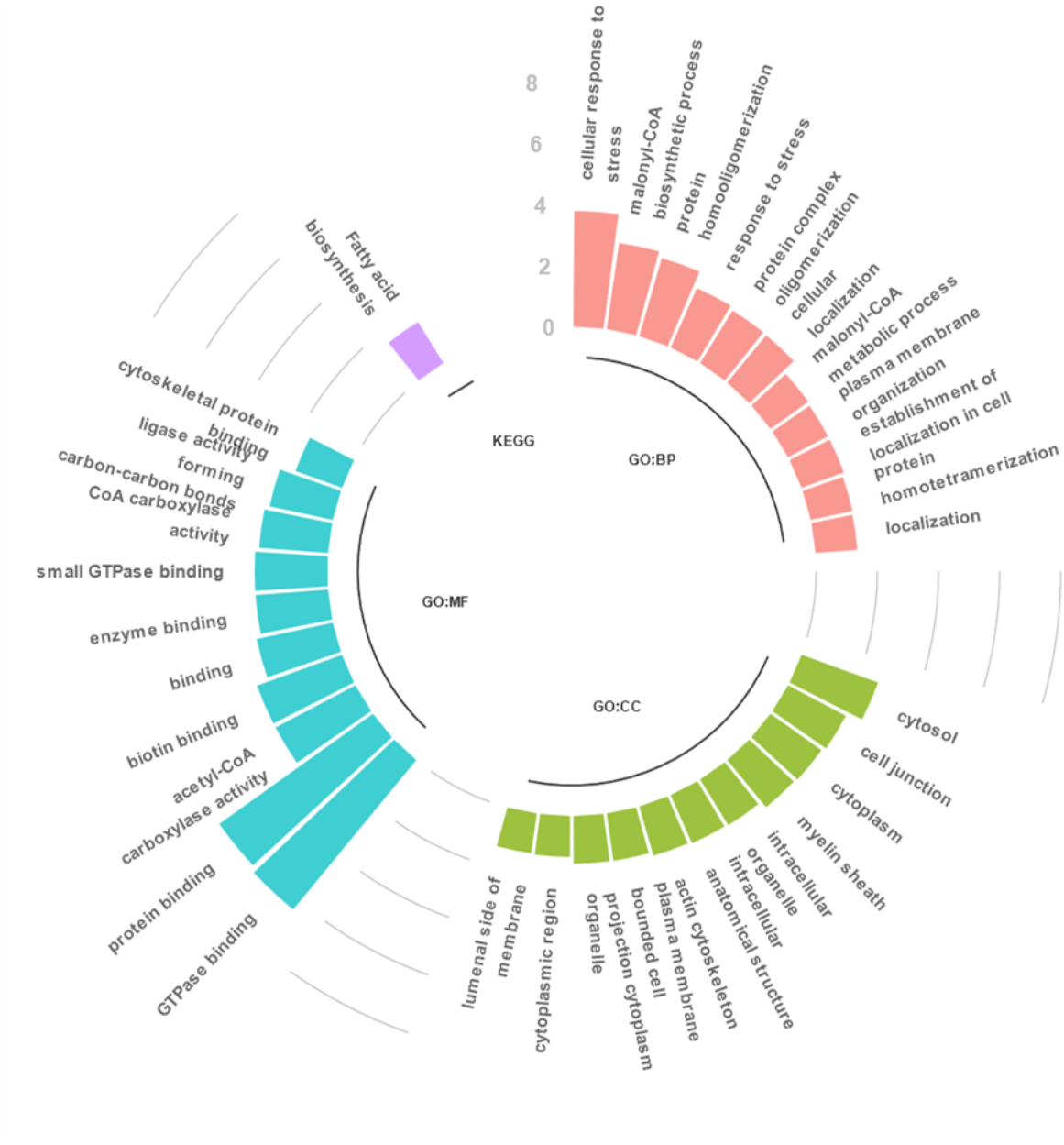


Figure 13: Gene ontology (GO) term analysis of differentially phosphorylated proteins from AEP$^{-/-}$ and wild-type C56/BL6 mice kidneys. Analysis was carried out with G:profiler. Biological process (BP), molecular function (MF), cellular component (CC) and Kyoto encyclopaedia of genes and genomes (KEGG).

4.1.1.4 Protein interaction network analysis

Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is a database of known and predicted protein-protein interactions. Interaction data are based on experimental results, *in-silico* prediction methods, and data mining [154]. We constructed a complete STRING network with dysregulated phosphoproteins, five kinases identified as acting enzymes on specific phosphosites from PhosphoSitePlus [155] (RSK, AMPKA2, ERK1, GSK3B, and CK2), and 50 nearest interacting proteins with a confidence level set at 0.7. Functional enrichment analysis was performed on the largest clusters of interacting

proteins. The highest-confidence GO term for the biggest cluster was ‘RNA binding’ (FDR 0.0011), while the neighbouring cluster was connected with the ‘Proteasomal protein catabolic process’ (FDR 2.97e-7). Next, we identified ‘Negative regulation of cellular metabolism’ (FDR 0.028), ‘Apoptotic process’ (FDR 0.031), ‘Fatty acid metabolism’ (FDR 2.87e-11), ‘Regulation of transcription elongation’ (FDR 1.5e-4), and ‘Lysosome organization’ (FDR 4.47e-6) in subsequent groups of proteins, respectively (Figure 14).

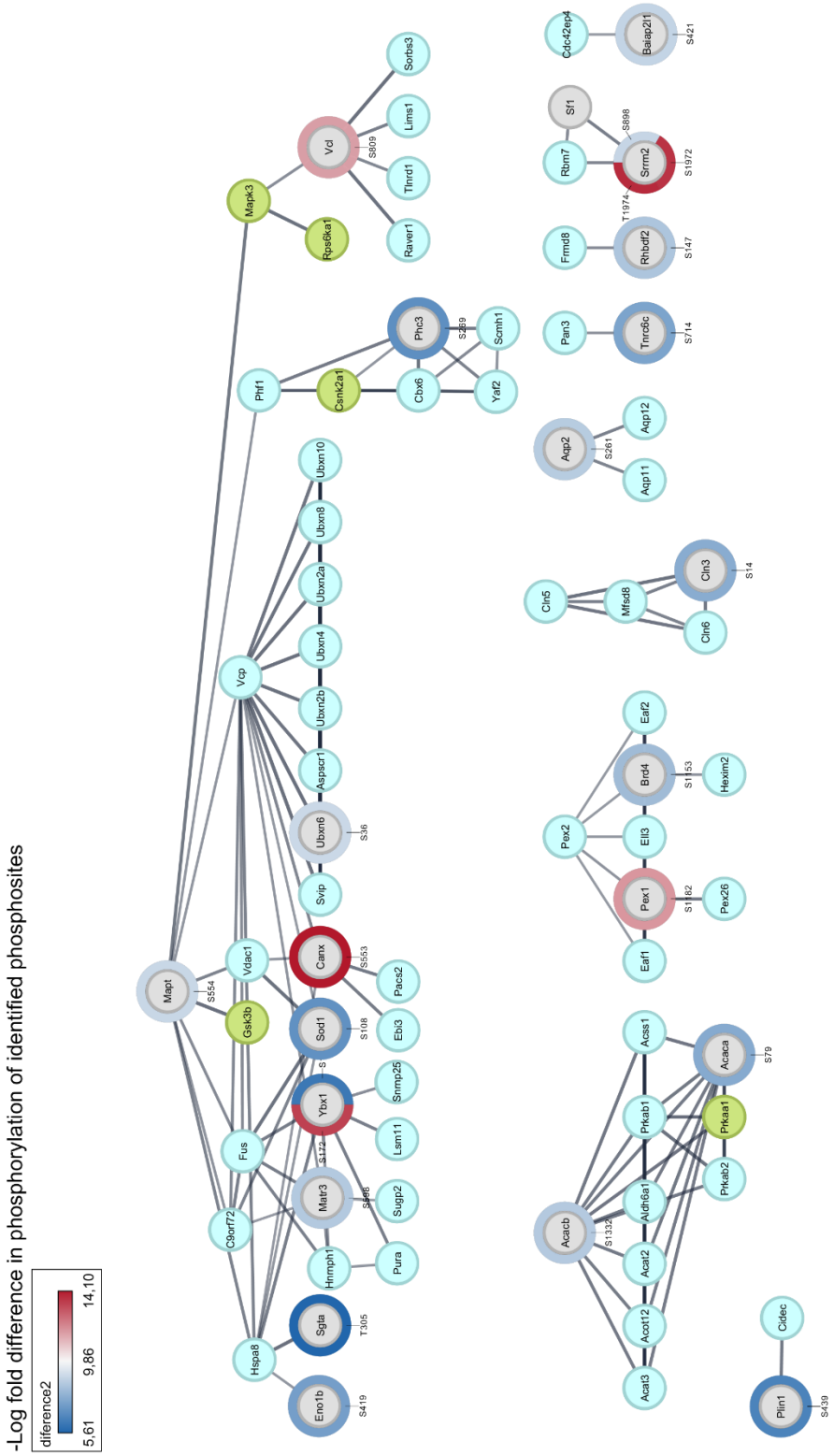


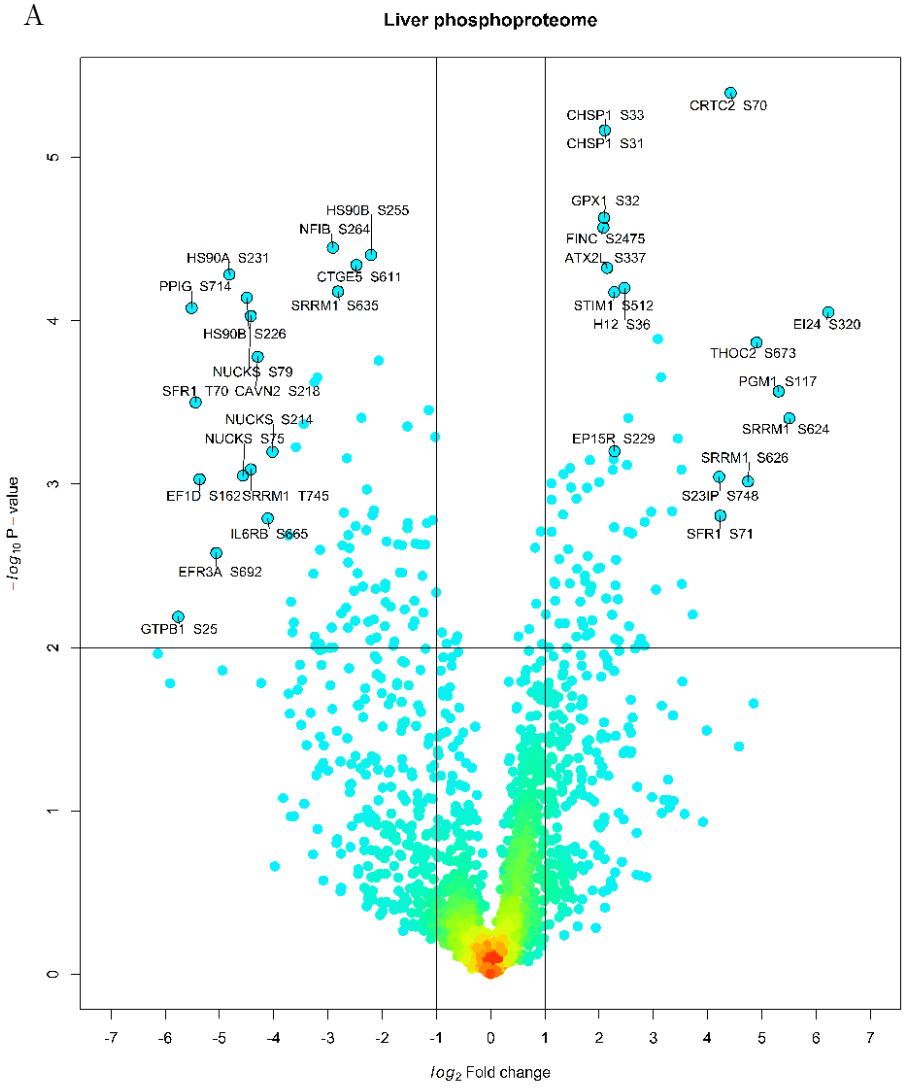
Figure 14: STRING network of significantly dysregulated phosphorylated proteins in kidneys of AEP^{-}/$^{-}$ and C57/BL6 wild-type mice (grey nodes). In kidney, identified phosphoprotein are

connected with their nearest 50 high confidence (confidence set at 0.7) interacting partner proteins (light blue nodes). Green nodes are known kinases that target identified phosphoproteins.

4.1.2 Phosphoproteomic profile analysis of mouse liver

Phosphosites in the liver were identified on 2 268 phosphopeptides, which were assigned to 2 789 proteins. Similar to the kidneys, most peptides were singly phosphorylated (1 954, 85.9 %), with 275 (12.0 %) doubly phosphorylated, and 47 (2.0 %) phosphorylated on at least three residues. Among the identified phosphorylation sites, 82.9 % (2 413) were phosphorylated on serine residues, while 16.1 % (469) and 1.0% (30) were on threonine and tyrosine residues, respectively. In the liver, a total of 2 912 phosphosites were identified, with 1 842 of them having over 75% localization confidence (class I sites). Additionally, 173 phosphosites in the liver were differentially regulated between wild-type and knock-out mice, 20 of which were phosphorylated on multiple sites (Figure 15).

A


Liver phosphoproteome
CRTC2 S70
CHSP1 S33
CHSP1 S31
GPX1 S32
FINC S2475
ATX2L S337
STIM1 S512
H12 S36
EI24 S320
THOC2 S673
PGM1 S117
SRRM1 S624
EP15R S229
SRRM1 S626
S23IP S748
SFR1 S71
HS90B S255
NFIB S264
HS90A S231
CTGE5 S611
PPIG S714
SRRM1 S635
HS90B S226
NUCKS S79
SFR1 T70 CAVN2 S218
NUCKS S214
NUCKS S75
EF1D S162 SRRM1 T745
IL6RB S665
EFR3A S692
GTPB1 S25
−log10 P−value
log2 Fold change

B

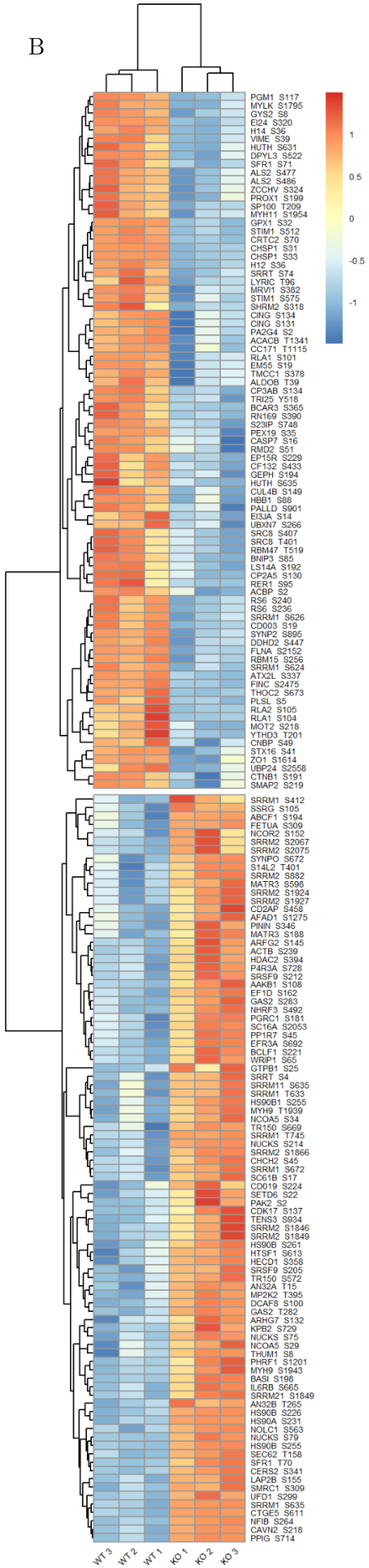

PGM1 S117
MYLK S1795
GYS2 S8
EI24 S320
H14 S36
VIME S39
HUTH S631
DPYL3 S522
SFR1 S71
ALS2 S477
ALS2 S486
ZCCHV S324
PROX1 S199
SP100 T209
MYH11 S1954
GPX1 S32
STIM1 S512
CRTC2 S70
CHSP1 S31
CHSP1 S33
H12 S36
SRRT S74
LYRIC T96
MRVI1 S382
STIM1 S575
SHRM2 S318
CING S134
CING S131
PA2G4 S2
ACACB T1341
CC171 T1115
RLA1 S101
EM55 S19
TMCC1 S378
ALDOB T39
CP3AB S134
TRI25 Y518
BCAR3 S365
RN169 S390
S23IP S748
PEX19 S35
CASP7 S16
RMD2 S51
EP15R S229
CF132 S433
GEPH S194
HUTH S635
CUL4B S149
HBB1 S88
PALLD S901
EI3JA S14
UBXN7 S266
SRC8 S407
SRC8 T401
RBM47 T519
BNIP3 S85
LS14A S192
CP2A5 S130
RER1 S95
ACBP S2
RS6 S240
RS6 S236
SRRM1 S626
CD003 S19
SYNP2 S895
DDHD2 S447
FLNA S2152
RBM15 S256
SRRM1 S624
ATX2L S337
FINC S2475
THOC2 S673
PLSL S5
RLA2 S105
RLA1 S104
MOT2 S218
YTHD3 T201
CNBP S49
STX16 S41
ZO1 S1614
UBP24 S2558
CTNB1 S191
SMAP2 S219
SRRM1 S412
SSRG S105
ABCF1 S194
FETUA S309
NCOR2 S152
SRRM2 S2067
SRRM2 S2075
SYNPO S672
S14L2 T401
SRRM2 S882
MATR3 S598
SRRM2 S1924
SRRM2 S1927
CD2AP S458
AFAD1 S1275
PININ S346
MATR3 S188
ARFG2 S145
ACTB S239
HDAC2 S394
P4R3A S728
SRSF9 S212
AAKB1 S108
EF1D S162
GAS2 S283
NHRF3 S492
PGRC1 S181
SC16A S2053
PP1R7 S45
EFR3A S692
BCLF1 S221
WRIP1 S65
GTPB1 S25
SRRT S4
SRRM11 S635
SRRM1 T633
HS90B1 S255
MYH9 T1939
NCOA5 S34
TR150 S669
SRRM1 T745
NUCKS S214
SRRM2 S1866
CHCH2 S45
SRRM1 S672
SC61B S17
CD019 S224
SETD6 S22
PAK2 S2
CDK17 S137
TENS3 S934
SRRM2 S1846
SRRM2 S1849
HS90B S261
HTSF1 S613
HECD1 S358
SRSF9 S205
TR150 S572
AN32A T15
MP2K2 T395
DCAF8 S100
GAS2 T282
ARHG7 S132
KPB2 S729
NUCKS S75
NCOA5 S29
THUM1 S8
PHRF1 S1201
MYH9 S1943
BASI S198
IL6RB S665
SRRM21 S1849
AN32B T265
HS90B S226
HS90A S231
NOLC1 S563
NUCKS S79
HS90B S255
SEC62 T158
SFR1 T70
CERS2 S341
LAP2B S155
SMRC1 S309
UFD1 S299
SRRM1 S635
CTGE5 S611
NFIB S264
CAVN2 S218
PPIG S714
1
0.5
0
-0.5
-1
WT 3
WT 2
WT 1
KO 1
KO 2
KO 3

Figure 15: **(A)**: Volcano plot showing phosphosites over-represented (top left) and under-represented (top right) in legumain null C57/BL6 mice liver. Some significantly dysregulated phosphosites are labelled with gene names and modified amino acid residues. Horizontal lines denote a 2-fold change in abundance, horizontal line denotes a p-value of 0.01. **(B)**: Heat-map constructed with pheatmap package in R shows under-represented phosphorylation sites in AEP $^{-}/^{-}$ liver with blue and over-represented ones with orange. WT 1-3 are wild-type mice and KO 1-3 are knock-out mice.

As in the kidneys, liver phosphoproteomic dataset was compared to control experiment, where phosphoproteins weren't enriched. here also we confirmed, the difference in phosphorylation is not due to the overall difference in whole protein abundance.

4.1.2.1 Prediction of subcellular location of identified phosphoproteins

According to analysis by DeepLoc 2.0, the most proteins in the liver were localized to the cytoplasm (33.1 %) and nucleus (21.3 %) with an additional 16.9 % assigned to the cytoplasm and nucleus. Roughly a tenth of proteins can be located in the endoplasmic reticulum with smaller fractions in other subcellular locations (Figure 16).

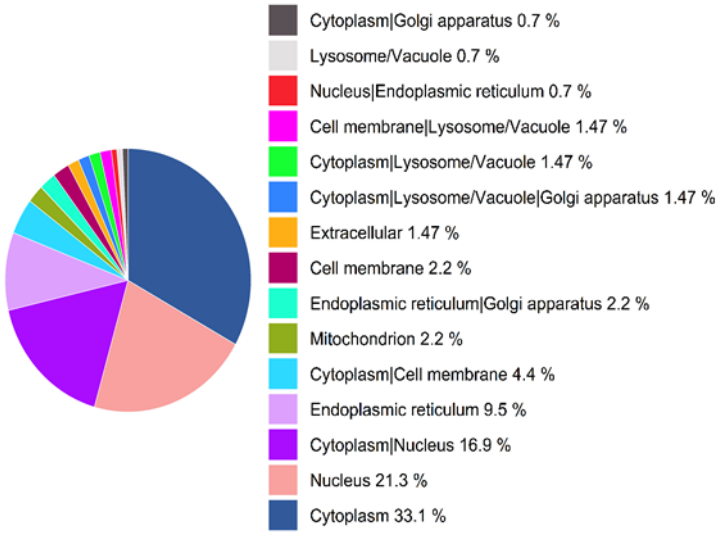


Figure 16: Pie chart showing predicted subcellular locations of identified phosphorylated proteins as identified by DeepLoc 2.0 algorithm in the liver.

4.1.2.2 Prediction of interacting kinases based on kinase recognition motifs

Dysregulated phosphopeptides identified in the liver most commonly included motifs .pSP., R..pS and pS..E. In downregulated peptides motif .pSP. was the most enriched while in upregulated peptides motifs .pSP. and pS..E. In the whole dataset, the most common motifs included .pSP. (425), R..S (361), S..E (181) and .S. (164) (Figure 17).

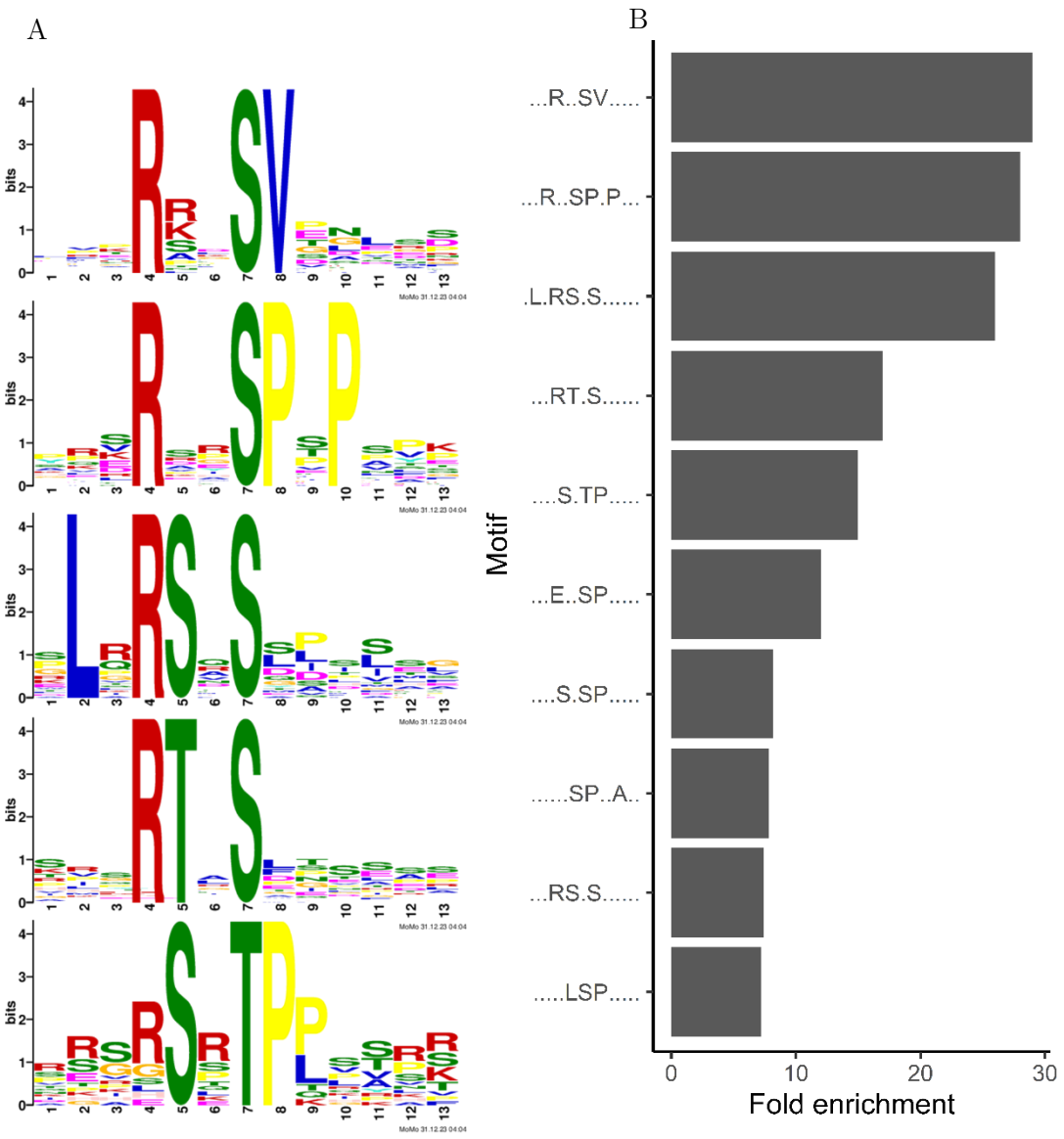


Figure 17: Motif analysis of identified phosphosites in C57/BL6 mouse liver with their fold enrichment **(B)**: total number of background peptides with central phosphorylation/number of peptides that match the motif. **(A)**: Ice logos for five most enriched motifs are shown.

4.1.2.3 Gene Ontology (GO) term analysis

Among all differentially phosphorylated proteins in the liver, the most enriched GO terms were protein binding (MF), positive regulation of biological process (BP), cytoplasm (CC), and Glucagon signalling pathway (KEGG). In upregulated proteins, these terms were protein binding, regulation of metabolic process, intracellular anatomical structure, and protein processing in the endoplasmic reticulum, corresponding to MF, BP, CC, and KEGG, respectively. For downregulated proteins, the most enriched GO terms in the same order were protein binding, catabolic process, and cytoplasm (Figure 18).



Figure 18: Gene onthology (GO) term analysis of differientaly phosphorylated proteins from AEP-/- and wild-type C56/BL6 mice liver. Analysis was carried out with G:profiler. Biological process (BP), molecular function (MF), cellular component (CC) and Kyoto encyclopaedia of genes and genomes (KEGG).

4.1.2.4 Protein interaction network analysis

A complete STRING network was built from dysregulated phosphorylated proteins identified in the liver, with a confidence level set at 0.7 and no additional interactors included. Functional enrichment analysis of the largest protein cluster revealed the GO:MF term 'Cytoskeletal protein binding' (FDR 2.65e-10) as the most significant. The next three clusters of interacting proteins showed GO:BP 'mRNA processing' (FDR 2.65e-7), GO:BF 'Cytoplasmic translation' (FDR 0.0024), and GO:BP 'Carbohydrate metabolic process' (FDR 0.0035) as the most enriched terms (Figure 19).

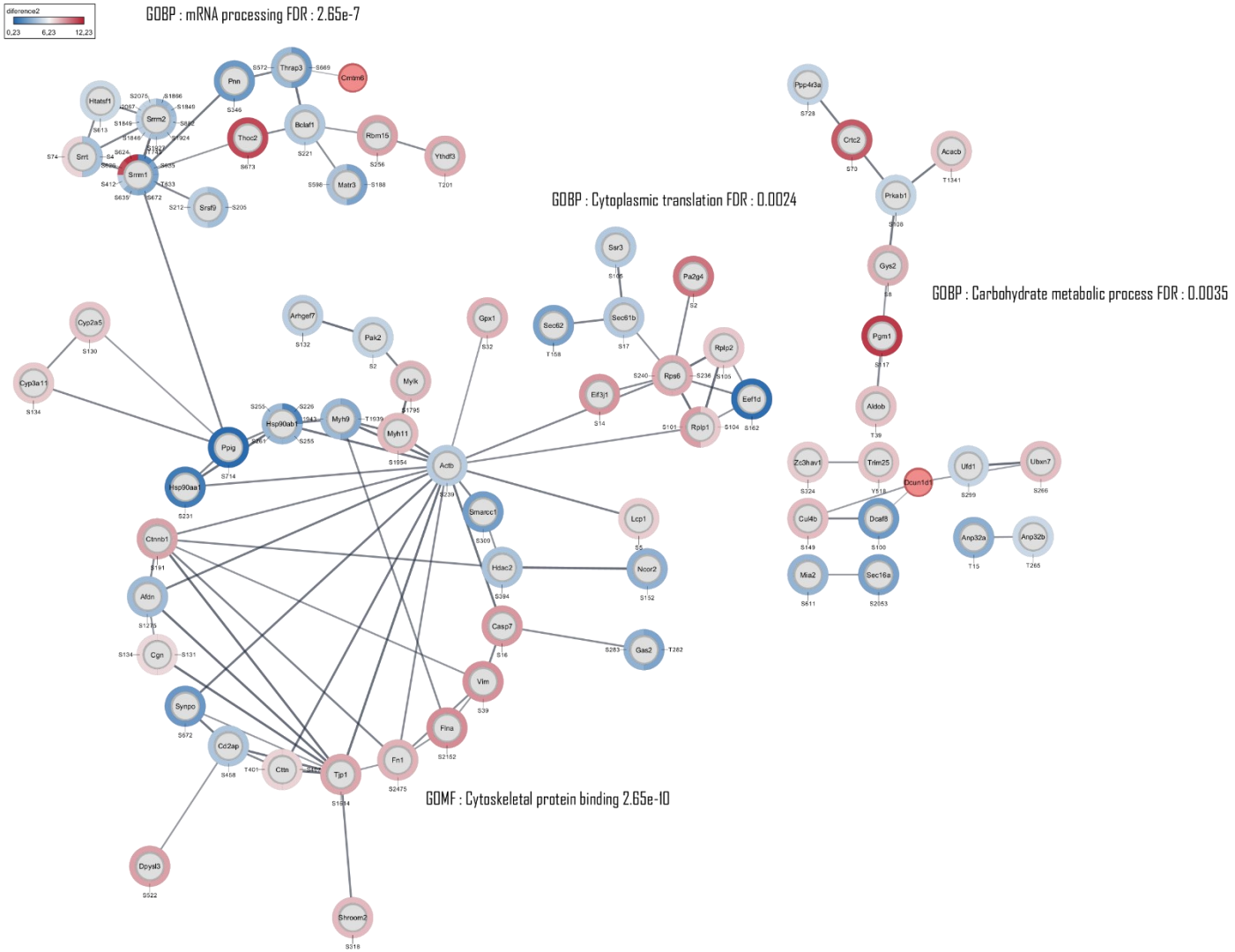

diference2
0.23
6.23
12.23
GOBP : mRNA processing FDR : 2.65e-7
GOBP : Cytoplasmic translation FDR : 0.0024
GOBP : Carbohydrate metabolic process FDR : 0.0035
GOMF : Cytoskeletal protein binding 2.65e-10
Phn
Thrap3
Cmtm6
Htatsf1
Srrm2
Bclaf1
Thoc2
Rbm15
Ythdf3
Srrt
Srrm1
Matr3
Srsf9
Ppp4r3a
Crtc2
Acacb
Prkab1
Gys2
Pgm1
Aldob
Sar3
Pa2g4
Sec61b
Sec62
Rplp2
Rps6
Eif3j1
Eef1d
Rplp1
Arhgef7
Pak2
Gpx1
Mylk
Cyp2a5
Cyp3a11
Hsp90ab1
Myh9
Myh11
Ppig
Actb
Hsp90aa1
Smarcc1
Lcp1
Ncor2
Hdac2
Ctnnb1
Afdn
Casp7
Gas2
Cgn
Vim
Synpo
Fina
Cd2ap
Cttn
Tjp1
Fn1
Dpysl3
Shroom2
Zc3hav1
Trim25
Dcun1d1
Ufd1
Ubxn7
Cul4b
Dcaf8
Anp32a
Anp32b
Mia2
Sec16a

Figure 19: STRING network of significantly dysregulated phosphorylated proteins in liver of AEP^{-}/$^{-}$ and C57/BL6 wild-type mice (grey nodes). Identified phosphoproteins are connected with their nearest 50 high confidence (confidence set at 0.7) interacting partner proteins (light blue nodes). Phosphosites are represented with donuts enveloping the nodes, marked with phosphosite amino acid and position. Donut colour presents if the phosphosite is upregulated (blue) or downregulated (red) in knock-outs.

4.2 Gene expression analysis

Based on connections from STRING analysis, we compared the expression of several gene targets using RT-qPCR. This helped us determine whether differences in abundance result from increased expression or decreased degradation of mature proteins. The results show that there is no expression of IL-10 in the kidney or liver. However, we observed a 1.7-fold increase in IL-10 receptor alpha expression in the kidneys of knockout samples compared to the control group. Most notably, Il36a was overexpressed in the kidneys of knockout mice. We also saw an increase in the expression of pro-inflammatory cytokines, especially CCL2 in the kidneys, with CXCL1, CCL3, and CCL5 showing smaller increases (about two times higher expression) (Figure 20).

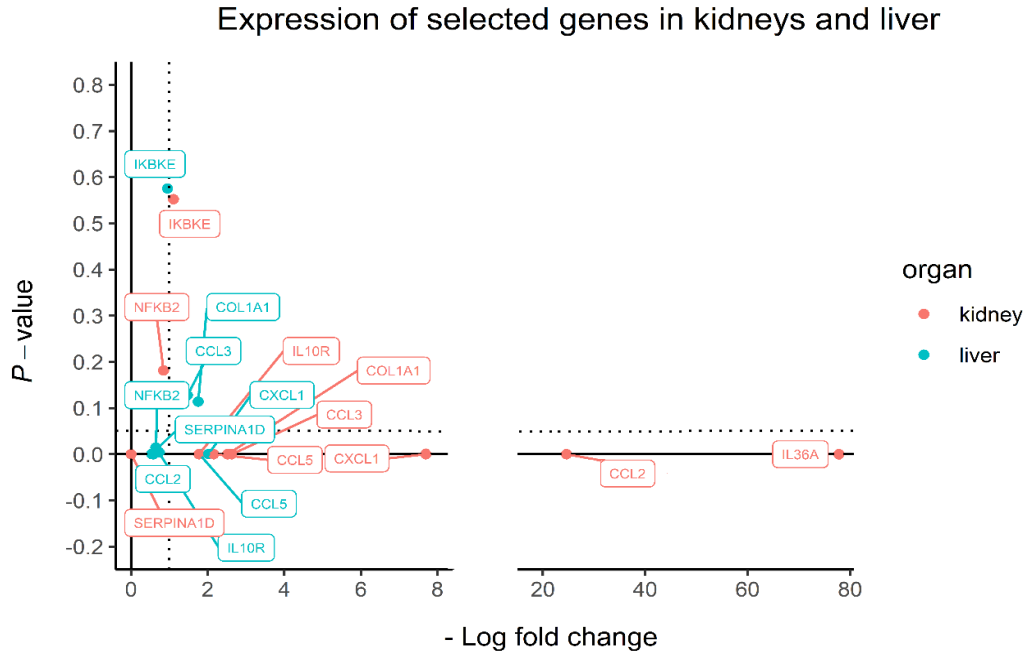


Figure 20: RT-qPCR expression analysis of selected genes in genes in kidneys and liver of AEP^{-}/$^{-}$ and wild-type C57/BL6 mice. Log fold change indicates an increase in expression in knock-out animals. A vertical dotted line indicates a 1-fold log change in expression, horizontal dotted line denotes a p-value of 0.05. Expression data from the kidney is marked with pink, and from the liver with blue.

Within the transcriptome data, we examined the expression of NF-κB genes and some of their downstream targets. Particularly in the kidneys, we observed increased expression of Nfkb1, Nfkb2, Nfkbia, Rel, Relb, Icam1, and Vcam1. Most notably, Il36a was overexpressed in these animals, which we also confirmed with RT-qPCR (Figure 21). Using both methods—transcriptomics and RT-qPCR—we noted an increase in the expression of pro-inflammatory cytokines, most prominently CCL2 in the kidneys.

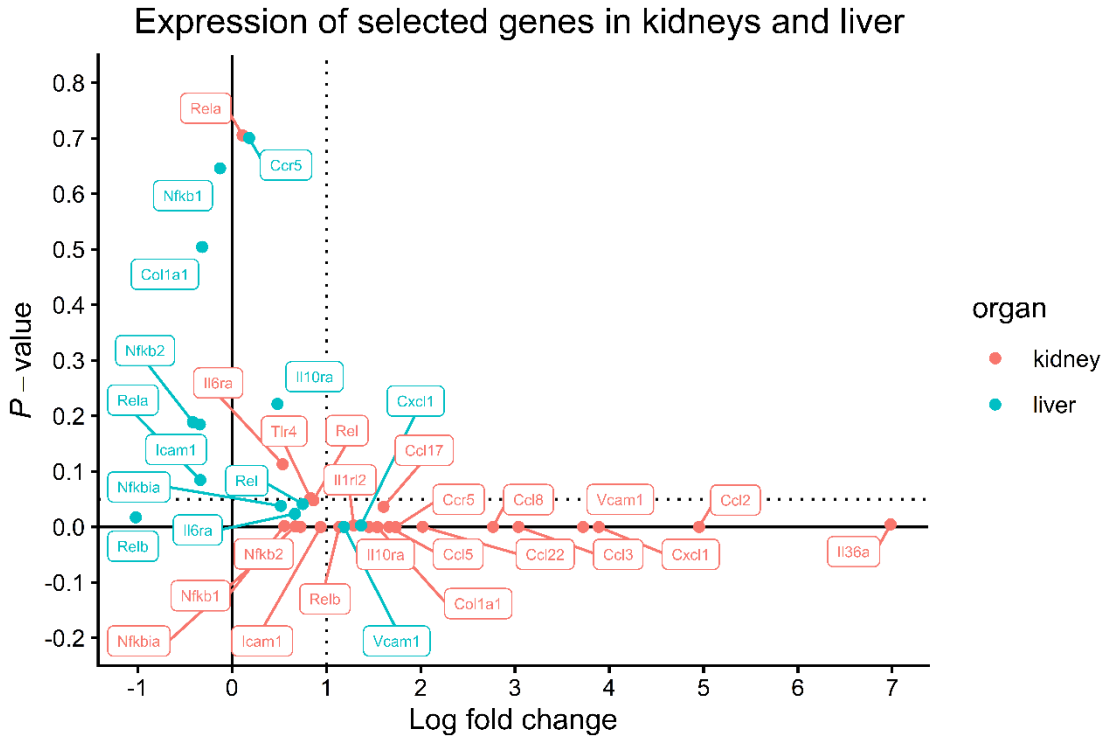


Figure 21: mRNA-Seq expression analysis of selected genes in genes in kidneys and liver of AEP$^{-/-}$ and wild-type C57/BL6 mice. Log fold change indicates an increase in expression in knock-out animals. A vertical dotted line indicates a 1-fold log change in expression, horizontal dotted line denotes a P-value of 0.05. Expression data from the kidney is marked with pink, and from the liver with blue.

4.3 Immunohistochemistry

With mass spectrometry, we found two phosphorylation sites on Y-box binding protein 1 (YBX1). These are serines 172 and 174. The first was more abundant in knockout kidney samples while the latter was more abundant in wild-type kidney samples. This difference led us to investigate this protein further. However, there were no commercially available antibodies for any of these sited at the time of writing this thesis. The only phospho antibodies we were able to procure were for serine 100, therefore we continued our work with examination of this site.

To examine the potential differences in phosphorylation levels of YBX1 on serine 100, we conducted an immunological analysis of this site in mouse organ samples and HL-60 cells. Results show higher phosphorylation levels in legumain knock-out animals, or in the case of the human cell line, higher phosphorylation levels in the control group (cells without overexpression of legumain). Next, we examined differences in IL-10 levels. There was a significant increase in IL-10 in the kidneys of knock-out animals, while the difference in the liver was only slight, favouring the knock-out group (Figures 22, 21).

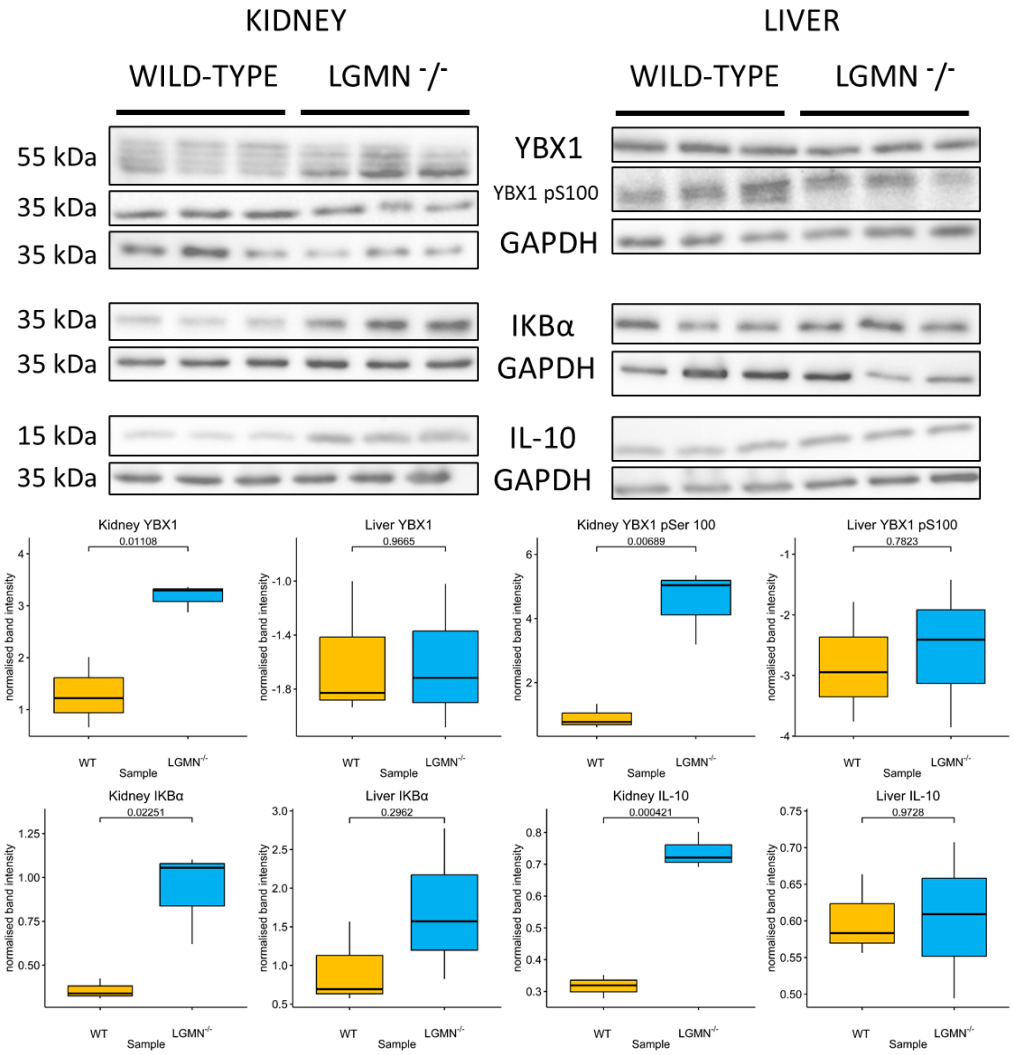


Figure 22: Western blot analysis of oof lysates from C57/BL6 wild-type (n=3) and LGMN $^{-/-}$ (n=3) kidneys (left) and liver (right) against YBX1, YBX1 pSer 100, IKBα, IL-10 and GAPDH as loading control. Band intensities were analysed with ImageJ software (ver. 1.49) and an unpaired two-sided t-test was performed in R. p-value is presented above horizontal line above box plots.

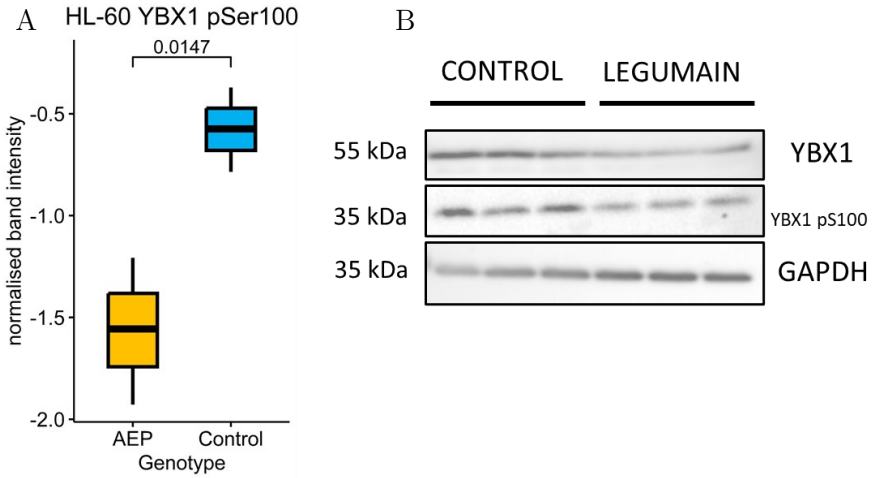


Figure 23: Western blot analysis of lysates from HL-60 cells with mock plasmid (control) (n=3) and legumain overexpression (n=3) against YBX1, YBX1 pSer 100, IKBα, IL-10 and GAPDH **(B)** as loading control. Band intensities were analysed with ImageJ software (ver. 1.49) and an unpaired two-sided t-test was performed in R; p-value is presented above horizontal line above box plots **(A)**.

We used immunofluorescent microscopy to label macrophages and determine if there were more of them in the kidneys of the knock-out animals compared to the wild-type mice. In the kidneys, we observed about twice the intensity of signal from labelled macrophages in knock-out organs compared to wild-type samples, suggesting an increased presence of macrophages in kidneys lacking legumain (Figure 24). We also checked whether YBX1 is concentrated in the nucleus of kidney cells in knock-out mice (Figure 25). There appear to be enucleations of YBX1 (red specks of YBX1 labelled with Alexa Flour 680).

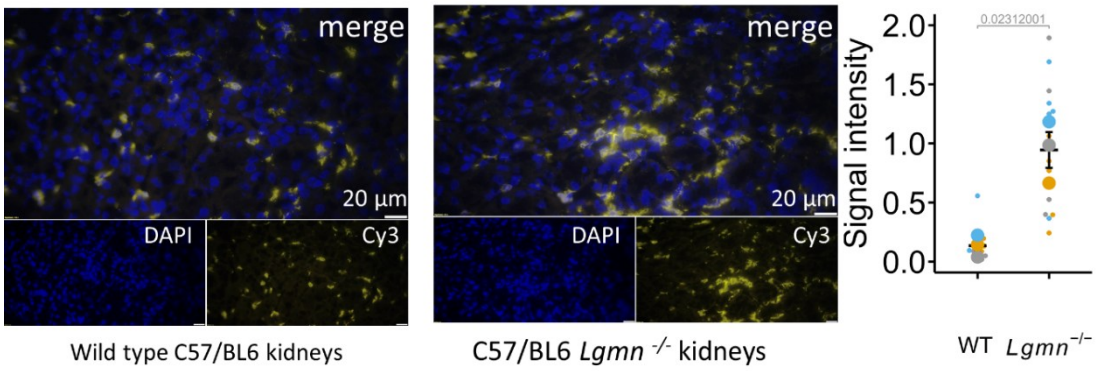


Figure 24: Immunofluorescence staining of macrophages in C57/BL6 wild-type (left) and Legumain$^{-/-}$(right) mouse kidneys. Cell nuclei are stained with DAPI (blue) and macrophage marker F4-80 with Cy3 (yellow). Quantification was performed with ImageJ software; five areas were chosen from each sample where the percentage of the stained area was quantified after adjustment for the colour threshold and normalization of the signal to the DAPI signal. Absence of legumain results in higher presence of macrophages in the kidneys of knock-out mice

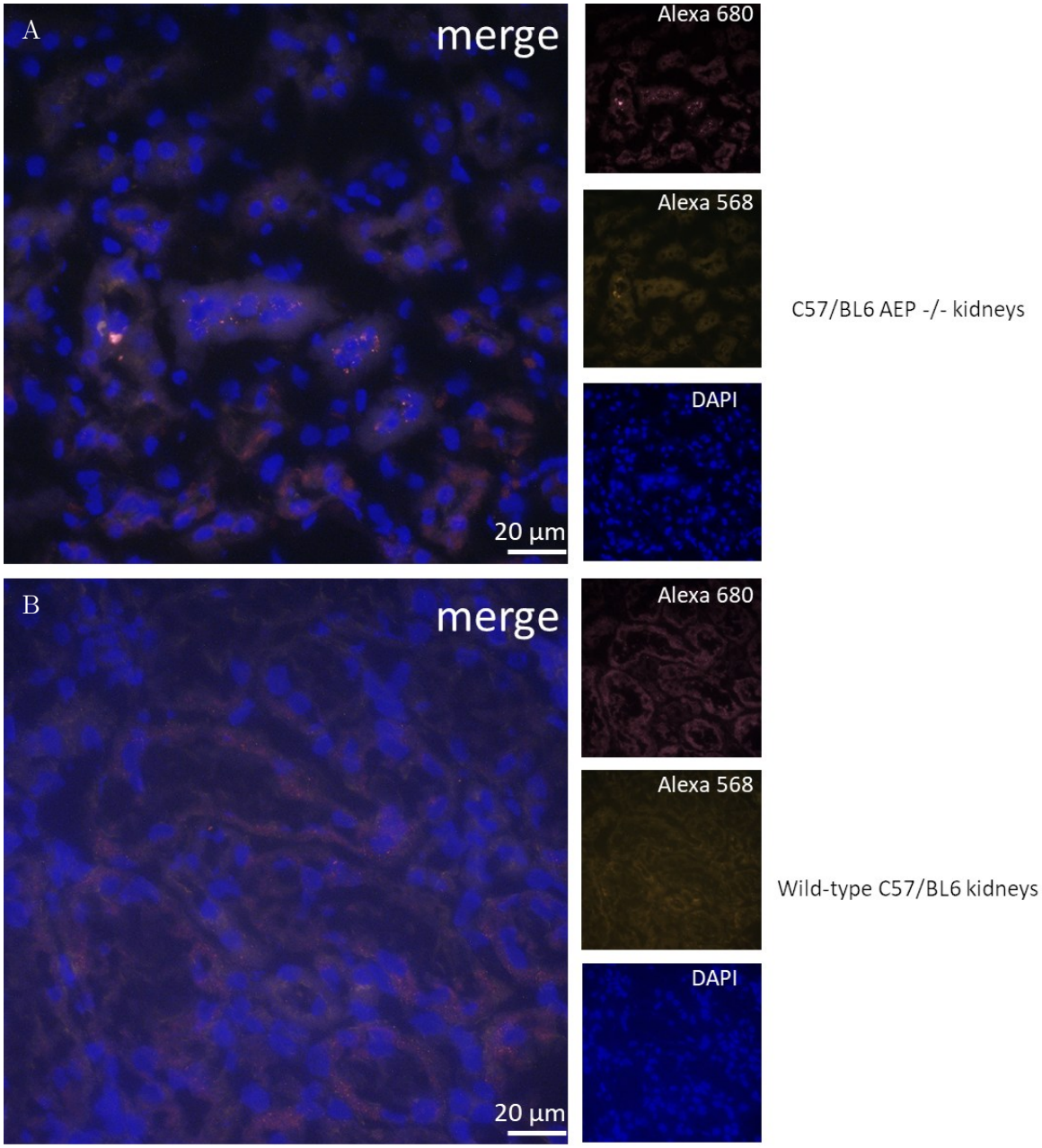


Figure 25: Immunofluorescence staining of YBX1 in C57/BL6 wild-type **(B)** and AEP^{-}/$^{-}$ **(A)** mouse kidneys. Cell nuclei are stained with DAPI (blue) and F-actin with Alexa Fluor™ 568 Phalloidin (yellow) and YBX1 (Alexa Fluor™ 680, red) marked with primary antibodies (Abcam ab12148) stained with secondary antibodies coupled with Alexa-PLUS-680 (Thermo Scientific, A32802). Aggregations of YBX1 in the nucleus can be observed in the knockout sample.

Chapter 5

Discussion

Legumain-deficient mice exhibit several phenotypic anomalies affecting multiple organ systems. Their T-cell response in the immune system is slower than that of wild-type specimens. Structural abnormalities in the proximal tubules of the kidneys result from impaired lysosomal processing, leading to progressive renal impairment characterized by decreased glomerular filtration rates and proteinuria. Additionally, the mice showed signs indicative of hemophagocytic syndrome, marked by enlarged histiocytes containing erythrocytes and splenomegaly [14].

Phosphorylation is one of the most important posttranslational modifications, serving as a key regulator of cellular signal transduction. Depending on the specific phosphorylation site, the target protein can be activated, inhibited, or enabled to interact with its binding partners [156]. Since EGFR, a receptor tyrosine kinase, was found to be upregulated in legumain-knockout mice, we decided to investigate how legumain deficiency affects protein phosphorylation in the kidneys and liver of these mice. Protein phosphorylation may be one of the pathways absence of legumain causes phenotype observed in legumain deficient mice.

In our analysis, we identified 31 differently regulated phosphosites in mouse kidneys and 173 in the liver (Figure 7). This seems to contradict the published pathological data, which showed many more abnormalities in the kidney than in the liver [142]. According to the gene ontology analysis, the most enriched molecular functional terms in the liver from phosphoproteomic data were "GTPase binding", "protein binding," and "binding." GTPases are a large superfamily of proteins that catalyse the hydrolysis of GTP to GDP. Their main role is to transmit signals from membrane receptors to effectors within cells [157], showing that legumain could influence cellular signalling networks through some possible interaction with GTPases. The prevalent mechanism that could cause several legumain knockout phenotypes includes the disruption of integrin-mediated signalling to Rho GTPases. Legumain serves as an intrinsic modulator of integrin $\alpha v\beta 3$ function by binding to and inhibiting this receptor in vascular smooth muscle cells . In wild-type mice, this legumain-integrin interaction diminishes downstream signalling to RhoA and possibly other small GTPases in vascular smooth muscle cells and alleviates extracellular matrix degradation. In legumain knockout mice, the lack of macrophage derived legumain results in increased or prolonged integrin-mediated activation of Rho GTPases [158].

Some of the enriched gene ontology terms in the kidneys were also related to fatty acid production (Figure 10). Fatty acids may act as intracellular signal transducers or hormones [159]. Knockout specimens had more proteins involved in fatty acid production, such as Acacb and Acaca. Gene ontology also indicates that the dysregulated proteins interact with the cytoskeleton, biotin, transcription factors, deoxyribonucleotides, and mRNA. Along with the enriched terms of specific clusters of interacting proteins identified by STRING network analysis—such as mRNA processing, cytoplasmic translation, and RNA binding—our findings could suggest that the phenotype observed in legumain-deficient

mice could be caused by changes in cellular signalling and the regulation of gene expression and/or translation. Ergo, while legumain may not affect protein phosphorylation directly through processing of kinases or phosphatases, it may cause changes in mRNA processing or translation that downstream lead to changes in phosphoproteome,

Identifying enriched phosphorylation motifs can help us discover upstream kinases and their pathways. Among dysregulated phosphorylation sites, a proline-rich motif .SP. was most common in the kidneys and the second most common in the liver. This consensus motif is shared by various kinases, including Cyclin-dependent kinases (CDKs), mitogen-activated protein kinases (MAPKs), and Glycogen Synthase Kinase 3 (GSK3). The R..S motif found in dysregulated phosphopeptides in the liver is most strongly associated with AGC kinases [160]. Of the 14 AGC kinase families, 8 are downstream effectors of growth factor signalling, while the others are involved in various other pathways [161]. The most common motif in the kidney and liver was proline-rich SP, suggesting that legumain may influence signalling via kinases targeting this pattern. According to the Phosphosite Plus [155] database, the most frequently predicted kinases based on enriched motifs in dysregulated phosphoproteins are Casein kinase 2, Protein kinase C, and Protein Kinase A. Casein kinase 2 is a Ser/Thr protein kinase involved in signalling pathways such as PI3K/Akt and NFκB. It is constitutively active. This kinase phosphorylates IκBα, an inhibitor of NFκB, thereby activating NFκB [162], as shown in our research. Mice with defective regulatory T-cells lacking legumain exhibit increased levels of TRAF6, which activates kinase TAK1, leading to phosphorylation and degradation of IκBα and activation of NFκB [163] Since we detected possible activation of NFκB in our samples based on gene expression analysis and casein kinase 2 is most confidently predicted kinase for our dataset, legumain may have some effect on the described pathway. Protein kinase C is a Ser/Thr protein kinase involved in antigen receptor signal transduction in T-cells, B-cells, and mast cells. Because legumain is involved in antigen processing, changes in the phosphorylation of PKC targets may occur if legumain is absent [163]. Protein kinase A is a family of cAMP-dependent kinases that target Ser and Thr residues [164]. It regulates kidney function by activating aquaporin 2 and water reabsorption [165], and it modulates immune system responses by affecting immune cell signalling and activity, possibly connecting it to the phenotype of legumain deficient mice, which have slightly slower immune response due to slower activation of T-cells [166], [167], [168].

Protein kinases are essential enzymes that regulate the function of proteins by adding phosphate groups, which in turn affects overall organism function. Some of the physiological changes we observed in legumain-deficient mice are linked to kinase activity. For example, Protein kinase A (PKA) plays a role in kidney function regulation [165]. In the kidney, PKA controls processes like water reabsorption by phosphorylating aquaporin-2 (AQP2) channels in the collecting ducts, which is vital for maintaining water balance and urine concentration. PKA also impacts cell proliferation and signalling pathways relevant to kidney development and disease. Notably, PKA signalling is involved in polycystic kidney disease (PKD). Increased PKA activity promotes cyst formation and disease progression by enhancing cell proliferation and fluid secretion in kidney tubules. Research indicates that specific PKA isoforms contribute to cyst development, and directly inhibiting PKA can decrease the severity and progression of PKD in experimental models, making PKA a promising therapeutic target for kidney diseases [169], [170]. Receptor tyrosine kinases participate in signalling during renal fibrosis [171]. Many kinases, such as CDK4/6, PCK, and AKT, regulate lysosome formation and function [172]. Some possible kinases predicted through enriched phosphorylation motifs in our dataset, including Protein

kinase A (PKA) [173], Protein Kinase C (PKC) [174], Calmodulin Kinase II [174], and MAP kinase [174], also play a role in antigen presentation. Information about legumain's direct interactions with kinases is limited. It is known that legumain cleaves Serine-Arginine Protein kinase 2 (SRPK2) in the Alzheimer disease brain and is also phosphorylated by SRPK2 [175] and is involved in initiating the AKT [175] pathway by promoting PI3K activation, which leads to enhanced Akt phosphorylation [176]. But there is little other solid evidence in existing literature and we have not detected any phosphorylation of legumain in the kidneys or liver of our samples.

Among the identified phosphorylated proteins, two were found to be differentially phosphorylated at two sites (SRRM and YBX1). Proteins SRRM1 and SRRM2 (serine and arginine repetitive matrix) are large (160 kDa and 300 kDa), extensively phosphorylated proteins located in nuclear speckles, where they participate in the regulation of mRNA splicing via the spliceosome [177]. SRRM1 is a splicing coactivator that enables RNA binding activity and mainly functions in regulating mRNA splicing through the spliceosome machinery. It is localized in nuclear speckles and operates as part of both pre- and post-splicing multiprotein mRNP complexes [178], [179]. SRRM2 is a large, intrinsically disordered nuclear protein that acts as the primary scaffold for nuclear speckle formation and plays essential roles in RNA splicing regulation. The protein contains multiple serine- and arginine-rich domains that facilitate liquid-liquid phase separation, enabling it to organize membraneless nuclear organelles where splicing factors congregate. SRRM2 functions as a splicing coactivator that favours the inclusion of cassette exons with weak splice sites, working alongside SRRM1 to regulate alternative splicing programs. Beyond its molecular functions, SRRM2 is vital for embryonic development and stem cell pluripotency, with its dysregulation linked to neurodevelopmental disorders, Parkinson's disease, and various cancers. The protein's dual role as both a structural organizer of nuclear architecture and a regulator of gene expression makes it a key hub for coordinating cellular responses to developmental and pathological stimuli [180], [181]. Phosphorylation of SRRM2 at sites that we identified, S1972, T1974, and S898 modulates the material properties of nuclear speckles by creating distinct charge blocks, which intensify intermolecular interactions but do not dramatically shift condensation thresholds; these modifications—primarily catalysed by casein kinase 2—facilitate nuclear speckle relaxation and enhance mRNA splicing efficiency, thereby supporting genome stability during DNA damage and regulating alternative splicing crucial for cell homeostasis, immune function, and disease prevention [180], [182] Since there is no evidence of direct interaction between legumain and SRRM2, some other, indirect pathway connects these two proteins. This could also indicate legumain influences mRNA splicing in an unknown way through changes in phosphorylation of SRRM2.

On the other hand, Y-box-binding protein (YBX1) is a 36 kDa protein that was found to be differentially phosphorylated only in murine kidneys. It is otherwise known to be phosphorylated on multiple sites, including its cold shock domain [155]. YBX1 (Y-box binding protein 1) is a highly versatile and evolutionarily conserved protein that acts as a master regulator of gene expression in cells. As a member of the cold shock domain (CSD) protein family, YBX1 plays crucial roles in nearly every aspect of cellular function, from basic RNA processing to complex disease development, making it one of the most extensively studied proteins in molecular biology [183]. YBX1 undergoes numerous post-translational modifications, with phosphorylation being the most common and functionally important. Phosphorylation occurs at multiple sites across the protein and serves as a key

regulatory mechanism that controls YBX1's subcellular localization, protein interactions, and various cellular functions [184], [185].

We identified two YBX1 phosphorylation sites using mass spectrometry (serines 172 and 174) and one additional site with immunoblotting (serine 100). YBX1 is a DNA and RNA binding protein with various functions, including regulation of translation and transcription, RNA stabilization, and mRNA splicing. It can act as a repressor or enhancer of the same gene depending on the cell environment [186]. YBX1 has 32 known phosphorylation sites, with five located within the Cold Shock Domain. Only 13 of these sites have been confirmed through low-throughput studies, while the rest have been identified via mass spectrometry [187], [188], [189], [190], [191], [192], [193], [194], [195], [196]. The phosphosite on serine 174, reportedly targeted by Casein Kinase 1 [187], is more abundant in knockouts, whereas phosphorylation on serine 172 is more prevalent in wild-type samples. Phosphorylation of YBX1 at serines 172 and 174 is also mediated by Polo-like kinase 1 and p90 Ribosomal-S6 Kinase (RSK) [197], which regulate YBX1 translocation into the nucleus in glioblastoma multiforme tumours, preventing DNA damage and apoptosis of cancer cells [198]. In glioblastoma, YBX1 promotes proliferation and survival by upregulating stemness genes such as CD133, Nestin, and SOX2 [198]. Similarly, apurinic/apyrimidinic endonuclease 1 phosphorylates YBX1 at these sites in ovarian cancer, leading to stress granule formation and increased cell survival [199]. Phosphorylation of YBX1 at serine 174 is essential for its translocation to the nucleus and subsequent NF-κB activation [199]. This finding aligns with our observations, where YBX1 is localized in the nuclei of kidneys from legumain knockout mice [142] (Figure 22). Previous reports also describe renal inflammation in legumain-deficient mice, showing slightly increased NF-κB gene expression in the kidneys, along with elevated levels of the NF-κB inhibitor IκBα. The latter may represent a compensatory response to NF-κB pathway overactivation [200]. To further support NF-κB activation, we observed increased IL-36α expression levels in the kidneys of legumain knockout mice through RT-qPCR, as its activation triggers NF-κB signalling [201], [202]. This activation subsequently induces the transcription of several downstream genes, including chemokines CCL2, CCL3, CCL5, and CXCL1 [203], [204]. These inflammatory chemokines may contribute to the heightened inflammatory state observed in legumain knockout mice, particularly in the kidneys [205]. They collectively orchestrate heightened inflammatory states through complementary mechanisms involving immune cell recruitment, activation, and self-amplifying inflammatory cascades. CCL2 is the primary monocyte recruiter, mobilizing inflammatory monocytes from bone marrow and guiding their infiltration into inflamed tissues via CCR2. Recruited monocytes differentiate into pro-inflammatory macrophages that produce additional cytokines (IL-1β, IL-6, TNF-α), creating a self-perpetuating cycle [206], [207]. CCL3 recruits diverse immune cells: monocytes, T cells, NK cells, and dendritic cells. Critically, it triggers production of additional pro-inflammatory cytokines from macrophages and fibroblasts, amplifying the inflammatory cascade [208], [209]. CCL5 mobilizes T cells, eosinophils, basophils, and NK cells, and sustains chronic inflammation through late expression (3-5 days post-activation). It inhibits macrophage apoptosis, allowing persistent inflammatory cell presence, and induces matrix metalloproteinases that cause tissue remodelling and damage [208], [210]. CXCL1 is the primary neutrophil chemoattractant, acting via its CXCR2 receptor. Recruited neutrophils produce reactive oxygen species, proteases, and matrix metalloproteinases that cause collateral tissue damage. CXCL1 also induces COX-2 expression, generating prostaglandins that amplify inflammation [211], [212], [213].

Previous studies indicated that YBX1 S100 phosphorylation regulates the release of many pro-inflammatory chemokines, including CCL2, CCL5, and CXCL1 that we identified as more expressed in legumain knockout kidneys. Evidence suggests that NF-κB activation triggers the production of these cytokines [214]. These cytokines are possibly the main drivers of renal inflammation in legumain-deficient animals. The chemokines listed above attract immune cells such as macrophages. Our findings reveal that macrophages are more abundant in the kidneys of legumain-deficient animals than in the wild-type, supporting this theory.

Since there are no commercially available antibodies for YBX1 phosphorylation sites that we identified with mass spectrometry, we used the only available phospho YBX1 on the market: phosphorylation at serine 100. The phosphorylation site at serine 100 is described by Mehta *et al.* as "difficult to detect" with mass spectrometry [197]. We agree with the authors of this paper [197], as we also failed to detect the site in our experiments. That's why we studied it only using immunological methods (Figure 19). Several studies have identified Akt1, p90RSK, p70S6K, ERK, and RSK2 as the kinases responsible for modifying this residue *in vivo* [215], [216], [217], [218]. Accordingly, we observed an increase in Serine 100 phosphorylation in the kidneys of knockout mice. However, there were no noticeable differences in S100 phosphorylation in the liver. This aligns with existing research, where YBX1 is involved in several liver conditions (hepatocellular carcinoma [219], metabolic dysfunction-associated steatotic liver disease, liver fibrosis , and hepatitis B infection) not seen in legumain knockout mice.

Wang and colleagues [222] found that S100 phosphorylated YBX1 increases IL-10 production in inflamed kidneys, thereby reducing and resolving chronic inflammation damage. We observed higher levels of IL-10 in the kidneys of legumain deficient mice (Figure 19). However, when we measured IL-10 expression levels using RT-qPCR, we detected no gene expression in the kidneys or liver. In legumain-deficient kidneys, IL-10 receptor expression was, however, upregulated. IL-10 expression was also elevated in the serum and in regulatory T-cell-specific legumain knockout C57BL/6 mice. In this context, IL-10 is involved in lowering hypertension [163]. Increased chemokine expression can attract immune cells, such as macrophages, to the kidneys. Therefore, increased IL-10 levels may reflect an effort to reduce inflammatory damage in the kidney of legumain knockout mice [223], [224], [225].

YBX1 Ser100 phosphorylation plays complex and context-dependent roles in kidney pathology, with significantly different effects depending on cellular localization and timing of injury [226]. During acute kidney injury, phosphorylated YBX1 at Ser1020localizes to the fourth intron of the IL-10 gene locus and enhances IL-10 mRNA expression. This represents a crucial renoprotective mechanism where YBX1 acts as a transcriptional activator of the anti-inflammatory cytokine IL-10. Mice with half-maximal YBX1 expression ($Yb1^{+/-}$) show reduced inflammatory responses and less kidney damage in the early stages (days 1 and 5) of ischemia-reperfusion injury, along with decreased IL-10 expression when exposed to LPS and increased immune cell infiltration and tubular cell CCL5 expression when YBX1 is deficient [222]. Paradoxically, the same mice with reduced YBX1 expression develop worse kidney injury and inflammation with less regeneration at later stages (day 21) of ischemia-reperfusion. This indicates that while Ser100 phosphorylation offers early protection, prolonged YBX1 activation may contribute to chronic injury [222]. Recent studies reveal that YBX1 Ser100 phosphorylation is essential for glomerular-tubular communication via a novel paracrine signalling mechanism [227].

Podocytes physiologically secrete YBX1, which requires post-translational acetylation rather than Ser100 phosphorylation specifically. However, Ser100 phosphorylation likely influences the nuclear localization of YBX1, affecting the pool available for secretion. Mice with deletion of YBX1 in their podocytes (Ybx1ΔPod mice) have increased inflammatory cell infiltration in their kidneys [227]. YBX1 Ser100 phosphorylation has conflicting effects on renal fibrosis based on its subcellular localization [228]. When phosphorylated at Ser100 and localized to the nucleus, YBX1 acts as a strong negative regulator of fibrotic genes. It suppresses actin expression in α-smooth muscle cells, represses Col1A1 gene transcription, and activates MMP-2 expression, thereby promoting collagen degradation [227]. YBX1 Ser100 phosphorylation significantly impacts kidney inflammatory responses through multiple pathways [229]. YBX1 modulates the inflammatory response by affecting TNFα binding to its receptors (TNFR) and altering TNFα-mediated signalling. Specifically, phosphorylated YBX1 can cooperate with progranulin, a growth factor that binds TNF receptors and inhibits TNFα signalling, to reduce TNFα binding to its receptors and suppress downstream NF-κB signalling. This pathway modulation leads to decreased pro-inflammatory signalling and macrophage infiltration in kidney inflammation, highlighting a regulatory role of YBX1 S100 phosphorylation in controlling inflammation via the TNFα-TNFR axis and NF-κB signalling pathway [229]. Specifically deleting YBX1 in monocytes/macrophages increases tissue damage and fibrosis following ureteral obstruction, supporting YBX1's role in guiding macrophages toward anti-inflammatory phenotypes. The protein promotes the resolution of inflammation over time by upregulating IL-10, which depends on its nuclear localization facilitated by Ser100 phosphorylation [230]. YBX1 also interferes with TNF-α binding to its receptors and modulates TLR4 signalling in tubular cells. This dual anti-inflammatory role occurs through YBX1 directly binding to TLR4 receptors, blocking pro-inflammatory signalling [227].

Collagen type 1A (Col1A1) is a component of the extracellular matrix, and its levels rise in the kidneys during renal fibrosis [231]. We found that the expression of Col1A1 is significantly higher in the kidneys of knockout animals, but not in the liver. Increased levels of YBX1 phosphorylated on tyrosine 97 have been shown to stabilize Col1A1 mRNA in the cytoplasm, facilitating renal fibrosis [228]. However, we did not identify this phosphosite in our study and unexpectedly, identified site on serine 100, when located in the nucleus, acts as suppressor of Col1A1 expression [227].

The absence of legumain in mice kidneys may be a cause for renal inflammation by affecting the activity of YBX1, which activates the NF-κB pathway and leads to the activation of certain downstream proteins. Chemokines guide immune cells to the kidneys, where they could attempt to reduce inflammation by producing anti-inflammatory cytokines, such as IL-10. Our findings provide a solid foundation for a better understanding of how legumain deficiency influences the phenotype of knockout mice.

Chronic liver diseases, such as cirrhosis and fibrosis, often involve the excessive build-up of extracellular matrix. Legumain is associated with fibrogenesis through the activation of transforming growth factor-β1 (TGF-β1), a key regulator of fibrosis. In chronic pancreatitis, legumain encourages the activation of pancreatic stellate cells by increasing the production of extracellular matrix proteins via the activation of matrix metalloproteinase-2 (MMP-2). MMP-2 cleaves the transforming growth factor-β1 (TGF-β1) precursor into its active form, which then promotes fibrogenesis. This process contributes to pancreatic fibrosis in chronic pancreatitis. Inhibiting or removing legumain

decreases pancreatic fibrosis by blocking the activation of the TGF-β1 precursor [232]. Accordingly, we don't observe liber fibrosis in our legumain knockout mice livers.

With this work, we confirmed our first hypothesis, that legumain deficiency does alter protein phosphorylation landscape in mice kidney and liver. However, we cannot claim as to the exact mechanism through which legumain is connected to it. It could be either through direct processing of kinases, phosphatases or transcription factors. Legumain could indirectly influence other signalling factors involved in protein phosphorylation or it could possibly act as a transcription factor itself. We did not confirm second hypothesis with complete certainty, as we did not find direct evidence showing legumain influences function of DNA or RNA regulating proteins. We can suggest changes in expression and protein abundance if inflammatory proteins is the result of changes in protein phosphorylation, specifically phosphorylation of YBX1. While our data strongly suggests activation of NF-κB pathway, it is not certain this pathway is the sole regulator of pro-inflammatory response observed in kidneys of legumain deficient mice.

Chapter 6

Conclusions

Our work highlights how legumain influences protein phosphorylation landscape in organism, influencing signalling pathways, expressing in changes of phenotype in legumain knock-out mice. Key findings include:

- We identified numerus differentially phosphorylation sites when comparing wild type C57/BL6 and legumain knock out mice
- Y box binding protein 1 (YBX1), an RNA binding protein, is phosphorylated on serines 100, 172 and 174 in the kidneys of legumain knock out mice
- This phosphorylation event may lead to activation of NF-κB pathway in the kidneys, as shown by higher levels of IKBα in knock out mice
- There is higher expression of several macrophage attracting chemokines in the kidneys of legumain knock out mice
- There are more macrophages present in the kidneys of legumain knock out mice when compared to wild type
- There are higher levels of IL-10 and its receptor in the kidneys of legumain knock out mice

Phosphoproteomic profiling of legumain-deficient mice revealed that loss of legumain profoundly reshapes the phosphorylation landscape in kidney and liver, uncovering multiple signalling perturbations beyond its canonical protease function. In the kidney, altered phosphorylation of YBX1 at Ser100 drives NF-κB activation and a heightened inflammatory response directed through upregulated expression of chemokines CCL2, CCL3, CCL5 and CXCL1. Kinase-substrate motif analysis implicates MAPKs, AGC kinases, and casein kinases in mediating these changes. In the liver, differential phosphorylation of metabolic regulators points to compensatory shifts in fatty acid biosynthesis and catabolism. Network analyses connect dysregulated phosphoproteins to processes including RNA binding, proteasomal turnover, apoptosis, and lysosomal organization. Together, these findings establish legumain as a critical modulator of cellular signalling, linking its absence to inflammation, altered RNA metabolism, and metabolic remodelling. This work lays the foundation for targeted mechanistic studies of legumain's non-proteolytic roles and highlights potential kinase pathways for therapeutic intervention in legumain-related pathologies.

References

[1] D. V. Jeyaraju, A. Sood, A. Laforce-Lavoie, and L. Pellegrini, 'Rhomboid proteases in mitochondria and plastids: Keeping organelles in shape', 2013, *Elsevier B.V.* doi: 10.1016/j.bbamcr.2012.05.019.

[2] C. López-Otín and J. S. Bond, 'Proteases: Multifunctional enzymes in life and disease', Nov. 07, 2008, *American Society for Biochemistry and Molecular Biology Inc.* doi: 10.1074/jbc.R800035200.

[3] V. H. J. Van Der Velden and A. R. Hulsmann, 'Peptidases: structure, function and modulation of peptide-mediated effects in the human lung'.

[4] B. Turk, 'Targeting proteases: Successes, failures and future prospects', Sep. 2006. doi: 10.1038/nrd2092.

[5] Barrett A J and R. N. D, 'Evolutionary Lines of Cysteine Peptidases', *Biol. Chem.*, vol. 382, no. May, p. 727, 2001, doi: 10.1515/BC.2001.088.

[6] N. Yang, M. A. Matthew, and C. Yao, 'Roles of Cysteine Proteases in Biology and Pathogenesis of Parasites', *Microorganisms*, vol. 11, no. 6, Jun. 2023, doi: 10.3390/microorganisms11061397.

[7] K. Arafet, S. Ferrer, and V. Moliner, 'Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease', *ACS Catal.*, vol. 7, no. 2, pp. 1207–1215, Feb. 2017, doi: 10.1021/acscatal.6b03096.

[8] H. A. Chapman, R. J. Riese, and G.-P. Shi, 'EMERGING ROLES FOR CYSTEINE PROTEASES IN HUMAN BIOLOGY', *Annu. Rev. Physiol.*, vol. 59, no. 1, pp. 63–88, Oct. 1997, doi: 10.1146/annurev.physiol.59.1.63.

[9] E. Vidak, U. Javoršek, M. Vizovišek, and B. Turk, 'Cysteine cathepsins and their extracellular roles: Shaping the microenvironment', Mar. 01, 2019, *MDPI.* doi: 10.3390/cells8030264.

[10] J. Bania *et al.*, 'Human cathepsin S, but not cathepsin L, degrades efficiently MHC class II-associated invariant chain in nonprofessional APCs'. [Online]. Available: www.pnas.orgcgidoi10.1073pnas.1131604100

[11] S. Donnelly, J. P. Dalton, and M. W. Robinson, 'HOW PATHOGEN-DERIVED CYSTEINE PROTEASES MODULATE HOST IMMUNE RESPONSES'.

[12] O. C. Olson and J. A. Joyce, 'Cysteine cathepsin proteases: Regulators of cancer progression and therapeutic response', Nov. 24, 2015, *Nature Publishing Group.* doi: 10.1038/nrc4027.

[13] V. Stoka, O. Vasiljeva, H. Nakanishi, and V. Turk, 'The Role of Cysteine Protease Cathepsins B, H, C, and X/Z in Neurodegenerative Diseases and Cancer †', Nov. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI).* doi: 10.3390/ijms242115613.

[14] E. Dall and H. Brandstetter, 'Structure and function of legumain in health and disease', *Biochimie*, vol. 122, pp. 126–150, 2016, doi: 10.1016/j.biochi.2015.09.022.

[15] E. Dall and H. Brandstetter, 'Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation', *Proceedings of the National Academy of Sciences*, vol. 110, no. 27, pp. 10940–10945, 2013, doi: 10.1073/pnas.1300686110.

[16] R. Vidmar, M. Vizovišek, D. Turk, B. Turk, and M. Fonović, 'Characterization of Legumain Degradome Confirms Narrow Cleavage Specificity', pp. 50–57, 2019, doi: 10.17344/acsi.2018.4632.

[17] R. Vidmar, M. Vizovišek, D. Turk, B. Turk, and M. Fonović, ' Protease cleavage site fingerprinting by label-free in-gel degradomics reveals pH -dependent specificity switch of legumain ', *EMBO J.*, vol. 36, no. 16, pp. 2455–2465, Aug. 2017, doi: 10.15252/embj.201796750.

[18] J.-M. Chen *et al.*, 'Cloning, Isolation, and Characterization of Mammalian Legumain, an Asparaginyl Endopeptidase', *Journal of Biological Chemistry*, vol. 272, no. 12, pp. 8090–8098, 2002, doi: 10.1074/jbc.272.12.8090.

[19] O. Hajdušek *et al.*, 'Profiling of proteolytic enzymes in the gut of the tick Ixodes ricinus reveals an evolutionarily conserved network of aspartic and cysteine peptidases', *Parasit. Vectors*, vol. 1, no. 1, p. 7, 2012, doi: 10.1186/1756-3305-1-7.

[20] T. Miyoshi *et al.*, 'Developmental Stage- and Organ-Specific Expression Profiles of Asparaginyl Endopeptidases/Legumains in the Ixodid Tick Haemaphysalis longicornis', *Journal of Veterinary Medical Science*, vol. 70, no. 12, pp. 1363–1366, 2009, doi: 10.1292/jvms.70.1363.

[21] N. D. Rawlings, A. J. Barrett, P. D. Thomas, X. Huang, A. Bateman, and R. D. Finn, 'The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database', *Nucleic Acids Res.*, vol. 46, no. D1, pp. D624–D632, 2018, doi: 10.1093/nar/gkx1134.

[22] A. J. Barrett, P. M. Dando, M. Fortunato, J.-M. Chen, and R. A. E. Stevens, 'Cloning and expression of mouse legumain, a lysosomal endopeptidase', *Biochemical Journal*, vol. 335, no. 1, pp. 111–117, 2015, doi: 10.1042/bj3350111.

[23] F. Sievers *et al.*, 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Mol. Syst. Biol.*, vol. 7, 2011, doi: 10.1038/msb.2011.75.

[24] C. S. Bond and A. W. Schüttelkopf, 'ALINE: A WYSIWYG protein-sequence alignment editor for publication-quality alignments', *Acta Crystallogr. D Biol. Crystallogr.*, vol. 65, no. 5, pp. 510–512, Apr. 2009, doi: 10.1107/S0907444909007835.

[25] C. Linnestad *et al.*, 'Nucellain, a Barley Homolog of the Dicot Vacuolar-Processing Protease, Is Localized in Nucellar Cell Walls', *Plant Physiol.*, vol. 118, no. 4, pp. 1169–1180, 2002, doi: 10.1104/pp.118.4.1169.

[26] B. Goulet *et al.*, 'A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor', *Mol. Cell*, vol. 14, no. 2, pp. 207–219, 2004, doi: 10.1016/S1097-2765(04)00209-6.

[27] R. Solberg, R. Smith, M. Almlöf, E. Tewolde, H. Nilsen, and H. T. Johansen, 'Legumain expression, activity and secretion are increased during monocyte-to-macrophage differentiation and inhibited by atorvastatin', *Biol. Chem.*, vol. 396, no. 1, pp. 71–80, Jan. 2015, doi: 10.1515/hsz-2014-0172.

[28] Y. Lin *et al.*, 'Functional role of asparaginyl endopeptidase ubiquitination by TRAF6 in tumor invasion and metastasis', *J. Natl. Cancer Inst.*, vol. 106, no. 4, 2014, doi: 10.1093/jnci/dju012.

[29] M. Izumi, M. Tachibana, T. Hashimoto, Y. Ohno, J. Nakashima, and M. Ohori, 'Association of legumain expression pattern with prostate cancer invasiveness and aggressiveness', *World J. Urol.*, vol. 31, no. 2, pp. 359–364, 2012, doi: 10.1007/s00345-012-0977-z.

[30] X. Liu *et al.*, 'Overexpression of asparaginyl endopeptidase is significant for esophageal carcinoma metastasis and predicts poor patient prognosis', *Oncol. Lett.*, vol. 15, no. 1, pp. 1229–1235, 2018, doi: 10.3892/ol.2017.7433.

[31] Q. Zhu, M. Tang, and X. Wang, 'The expression of asparaginyl endopeptidase promotes growth potential in epithelial ovarian cancer', *Cancer Biol. Ther.*, vol. 18, no. 4, pp. 222–228, 2017, doi: 10.1080/15384047.2017.1294290.

[32] S. Kosugi *et al.*, 'Six classes of nuclear localization signals specific to different binding grooves of importinα', *Journal of Biological Chemistry*, vol. 284, no. 1, pp. 478–485, 2009, doi: 10.1074/jbc.M807017200.

[33] H. Sierra, M. Cordova, C.-S. J. Chen, and M. Rajadhyaksha, 'Confocal Imaging–Guided Laser Ablation of Basal Cell Carcinomas: An Ex Vivo Study', *Journal of Investigative Dermatology*, vol. 135, no. 2, pp. 612–615, Feb. 2015, doi: 10.1038/jid.2014.371.

[34] E. Dall, J. C. Fegg, P. Briza, and H. Brandstetter, 'Structure and mechanism of an aspartimide-dependent peptide ligase in human legumain', *Angew. Chem. Int. Ed.*, vol. 54, no. 10, pp. 2917–2921, 2015, doi: 10.1002/anie.201409135.

[35] D. Mazzeo, A. N. Antoniou, S. P. Matthews, C. Watts, and D. N. Li, ' Multistep Autoactivation of Asparaginyl Endopeptidase in Vitro and in Vivo ', *Journal of Biological Chemistry*, vol. 278, no. 40, pp. 38980–38990, 2003, doi: 10.1074/jbc.m305930200.

[36] E. Dall and H. Brandstetter, 'Activation of legumain involves proteolytic and conformational events, resulting in a context-and substrate-dependent activity profile', *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, vol. 68, no. 1, pp. 24–31, 2012, doi: 10.1107/S1744309111048020.

[37] N. Matarasso, 'A Novel Plant Cysteine Protease Has a Dual Function as a Regulator of 1-Aminocyclopropane-1-Carboxylic Acid Synthase Gene Expression', *the Plant Cell Online*, vol. 17, no. 4, pp. 1205–1216, 2005, doi: 10.1105/tpc.105.030775.

[38] S. J. Choi, N. Kurihara, Y. Oba, and G. D. Roodman, 'Osteoclast inhibitory peptide 2 inhibits osteoclast formation via its c-terminal fragment', *Journal of Bone and Mineral Research*, vol. 16, no. 10, pp. 1804–1811, 2001, doi: 10.1359/jbmr.2001.16.10.1804.

[39] B. Elsässer, F. B. Zauner, J. Messner, W. T. Soh, E. Dall, and H. Brandstetter, 'Distinct Roles of Catalytic Cysteine and Histidine in the Protease and Ligase Mechanisms of Human Legumain As Revealed by DFT-Based QM/MM Simulations', *ACS Catal.*, vol. 7, no. 9, pp. 5585–5593, 2017, doi: 10.1021/acscatal.7b01505.

[40] P. M. Dando, M. Alvarez-Fernandez, J. Ni, A. J. Barrett, B. Gerhartz, and M. Abrahamson, 'Inhibition of Mammalian Legumain by Some Cystatins Is Due to a Novel Second Reactive Site', *Journal of Biological Chemistry*, vol. 274, no. 27, pp. 19195–19203, 2002, doi: 10.1074/jbc.274.27.19195.

[41] M. P. Nanut, J. Sabotič, U. Švajger, A. Jewett, and J. Kos, 'Cystatin F affects natural killer cell cytotoxicity', *Front. Immunol.*, vol. 8, no. NOV, 2017, doi: 10.3389/fimmu.2017.01459.

[42] J. Sabotič, T. Popovič, V. Puizdar, and J. Brzin, 'Macrocypins, a family of cysteine protease inhibitors from the basidiomycete Macrolepiota procera', *FEBS Journal*, vol. 276, no. 16, pp. 4334–4345, 2009, doi: 10.1111/j.1742-4658.2009.07138.x.

[43] J. Rozman-Pungerčar *et al.*, 'Inhibition of papain-like cysteine proteases and legumain by caspase-specific inhibitors: When reaction mechanism is more important than specificity', *Cell Death Differ.*, vol. 10, no. 8, pp. 881–888, 2003, doi: 10.1038/sj.cdd.4401247.

[44] T. Zhao *et al.*, 'Esomeprazole inhibits the lysosomal cysteine protease legumain to prevent cancer metastasis', *Invest. New Drugs*, vol. 39, no. 2, pp. 337–347, Apr. 2021, doi: 10.1007/s10637-020-01011-3.

[45] K. B. Sexton, M. D. Witte, G. Blum, and M. Bogyo, 'Design of cell-permeable, fluorescent activity-based probes for the lysosomal cysteine protease asparaginyl endopeptidase (AEP)/legumain', *Bioorg. Med. Chem. Lett.*, vol. 17, no. 3, pp. 649–653, 2007, doi: 10.1016/j.bmcl.2006.10.100.

[46] T. Yamane *et al.*, 'Legumain from bovine kidney: Its purification, molecular cloning, immunohistochemical localization and degradation

of annexin II and vitamin D-binding protein', *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, vol. 1596, no. 1, pp. 108–120, 2002, doi: 10.1016/S0167-4838(02)00209-1.

[47] S. O. Kolset, L. Berven, R. Solberg, A. B. C. Samuelsen, and H. T. Johansen, 'Autoactivation of prolegumain is accelerated by glycosaminoglycans', *Biochimie*, vol. 95, no. 4, pp. 772–781, 2012, doi: 10.1016/j.biochi.2012.11.002.

[48] P. J. Brindley *et al.*, 'Proteolytic degradation of host hemoglobin by schistosomes', *Mol. Biochem. Parasitol.*, vol. 89, no. 1, pp. 1–9, 1997, doi: 10.1016/S0166-6851(97)00098-4.

[49] K. Müntz and A. D. Shutov, 'Legumains and their functions in plants', *Trends Plant Sci.*, vol. 7, no. 8, pp. 340–344, 2002, doi: 10.1016/S1360-1385(02)02298-7.

[50] S. P. Matthews, I. Werber, J. Deussing, C. Peters, T. Reinheckel, and C. Watts, 'Distinct Protease Requirements for Antigen Presentation In Vitro and In Vivo', *The Journal of Immunology*, vol. 184, no. 5, pp. 2423–2431, Mar. 2010, doi: 10.4049/jimmunol.0901486.

[51] M. A. West *et al.*, 'Creation versus Destruction of T Cell Epitopes in the Class II MHC Pathway', *Ann. N. Y. Acad. Sci.*, vol. 987, no. 1, pp. 9–14, 2009, doi: 10.1111/j.1749-6632.2003.tb06028.x.

[52] D. J. Craik, 'Host-defense activities of cyclotides', *Toxins (Basel).*, vol. 4, no. 2, pp. 139–156, 2012, doi: 10.3390/toxins4020139.

[53] I. Hara-Nishimura, N. Hatsugai, S. Nakaune, M. Kuroyanagi, and M. Nishimura, 'Vacuolar processing enzyme: An executor of plant cell death', *Curr. Opin. Plant Biol.*, vol. 8, no. 4, pp. 404–408, 2005, doi: 10.1016/j.pbi.2005.05.016.

[54] T. Ishizaki *et al.*, 'The asparaginyl endopeptidase legumain after experimental stroke', *Journal of Cerebral Blood Flow and Metabolism*, vol. 30, no. 10, pp. 1756–1766, 2010, doi: 10.1038/jcbfm.2010.39.

[55] M. Asano *et al.*, 'Biosynthetic Processing of Cathepsins and Lysosomal Degradation Are Abolished in Asparaginyl Endopeptidase-deficient Mice', *Journal of Biological Chemistry*, vol. 278, no. 35, pp. 33194–33199, 2003, doi: 10.1074/jbc.m302742200.

[56] G. Mazza *et al.*, 'Destructive processing by asparagine endopeptidase limits presentation of a dominant T cell epitope in MBP', *Nat. Immunol.*, vol. 3, no. 2, pp. 169–174, 2002, doi: 10.1038/ni754.

[57] W. S. Ho *et al.*, 'Pharmacologic inhibition of protein phosphatase-2A achieves durable immune-mediated antitumor activity when combined with PD-1 blockade', doi: 10.1038/s41467-018-04425-z.

[58] I. Grundke-Iqbal, K. Iqbal, F. Liu, Y. C. Tung, and G. Basurto-Islas, 'Activation of Asparaginyl Endopeptidase Leads to Tau Hyperphosphorylation in Alzheimer Disease', *Journal of Biological Chemistry*, vol. 288, no. 24, pp. 17495–17507, 2013, doi: 10.1074/jbc.m112.446070.

[59] J. Peng *et al.*, 'Neuroprotective Actions of PIKE-L by Inhibition of SET Proteolytic Degradation by Asparagine Endopeptidase', *Mol. Cell*, vol. 29, no. 6, pp. 665–678, 2008, doi: 10.1016/j.molcel.2008.02.017.

[60] W. Su *et al.*, 'Legumain: A biomarker for diagnosis and prognosis of human ovarian cancer', *J. Cell. Biochem.*, vol. 113, no. 8, pp. 2679–2686, 2012, doi: 10.1002/jcb.24143.

[61] Z. C. D'Costa *et al.*, 'TBX2 represses CST6 resulting in uncontrolled legumain activity to sustain breast cancer proliferation: a novel cancer-selective target pathway with therapeutic opportunities', *Oncotarget*, vol. 5, no. 6, pp. 1609–1620, 2014, doi: 10.18632/oncotarget.1707.

[62] T. Yamane, Y. Yamamoto, Y. Nakano, T. Nakagaki, I. Ohkubo, and H. Ariga, 'Expression and protease activity of mouse legumain are regulated by the oncogene/transcription co-activator, DJ-1 through p53 and cleavage of annexin A2 is increased in DJ-1-knockout cells', *Biochem. Biophys. Res. Commun.*, vol. 467, no. 3, pp. 472–477, 2015, doi: 10.1016/j.bbrc.2015.10.032.

[63] T. Yamane *et al.*, 'Transcriptional regulation of the legumain gene by p53 in HCT116 cells', *Biochem. Biophys. Res. Commun.*, vol. 438, no. 4, pp. 613–618, 2013, doi: 10.1016/j.bbrc.2013.08.007.

[64] X. Liu *et al.*, 'Overexpression of asparaginyl endopeptidase is significant for esophageal carcinoma metastasis and predicts poor patient prognosis', *Oncol. Lett.*, vol. 15, no. 1, pp. 1229–1235, 2018, doi: 10.3892/ol.2017.7433.

[65] T. Yamane *et al.*, 'Transcriptional regulation of the legumain gene by p53 in HCT116 cells', *Biochem. Biophys. Res. Commun.*, vol. 438, no. 4, pp. 613–618, 2013, doi: 10.1016/j.bbrc.2013.08.007.

[66] R. Solberg *et al.*, 'Intra- and extracellular regulation of activity and processing of legumain by cystatin E/M', *Biochimie*, vol. 94, no. 12, pp. 2590–2599, 2012, doi: 10.1016/j.biochi.2012.07.026.

[67] S. Akira and K. Takeda, 'Toll-like receptor signalling', *Nat. Rev. Immunol.*, vol. 4, no. 7, pp. 499–511, 2004, doi: 10.1038/nri1391.

[68] F. E. Sepulveda *et al.*, 'Critical Role for Asparagine Endopeptidase in Endocytic Toll-like Receptor Signaling in Dendritic Cells', *Immunity*, vol. 31, no. 5, pp. 737–748, Nov. 2009, doi: 10.1016/j.immuni.2009.09.013.

[69] D. Barbier *et al.*, 'Toll-like receptor 5 (TLR5), IL-1 secretion, and asparagine endopeptidase are critical factors for alveolar macrophage phagocytosis and bacterial killing', *Proceedings of the National Academy of Sciences*, vol. 109, no. 5, pp. 1619–1624, 2012, doi: 10.1073/pnas.1108464109.

[70] S. Freeley *et al.*, 'Asparaginyl endopeptidase (Legumain) supports human Th1 induction via cathepsin L-mediated intracellular C3

activation', *Front. Immunol.*, vol. 9, no. OCT, pp. 1–7, 2018, doi: 10.3389/fimmu.2018.02449.

[71] T. Lee, H. Huang, J. Hung, and H. Huang, 'dbPTM : an information repository of protein post-translational modification', vol. 34, pp. 622–627, 2006, doi: 10.1093/nar/gkj083.

[72] O. N. Jensen, 'Modification-specific proteomics : characterization of post-translational modifications by mass spectrometry', doi: 10.1016/j.cbpa.2003.12.009.

[73] O. Nørregaard Jensen, 'Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry', *Curr. Opin. Chem. Biol.*, vol. 8, no. 1, pp. 33–41, Feb. 2004, doi: 10.1016/j.cbpa.2003.12.009.

[74] J. A. Ubersax and J. E. Ferrell, 'Mechanisms of specificity in protein phosphorylation', *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 7, pp. 530–541, 2007, doi: 10.1038/nrm2203.

[75] J. V. Olsen *et al.*, 'Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks', *Cell*, vol. 127, no. 3, pp. 635–648, 2006, doi: 10.1016/j.cell.2006.09.026.

[76] F. Ardito, M. Giuliani, D. Perrone, G. Troiano, and L. Lo Muzio, 'The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review)', *Int. J. Mol. Med.*, vol. 40, no. 2, pp. 271–280, 2017, doi: 10.3892/ijmm.2017.3036.

[77] K. Miki, R. Nagata, M. Fujihashi, T. Sato, and H. Atomi, 'Identification of a pyrophosphate-dependent kinase and its donor selectivity determinants', *Nat. Commun.*, vol. 9, no. 1, pp. 1–8, 2018, doi: 10.1038/s41467-018-04201-z.

[78] P. Cohen, 'The origins of protein phosphorylation', *Nat. Cell Biol.*, vol. 4, no. 5, 2002, doi: 10.1038/ncb0502-e127.

[79] L. N. Johnson, 'The regulation of protein phosphorylation', *Biochem. Soc. Trans.*, vol. 37, pp. 627–641, 2009, doi: 10.1042/BST0370627.

[80] F. Ardito, M. Giuliani, D. Perrone, G. Troiano, and L. Lo Muzio, 'The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review)', *Int. J. Mol. Med.*, vol. 40, no. 2, pp. 271–280, 2017, doi: 10.3892/ijmm.2017.3036.

[81] E. K. Day, N. G. Sosale, and M. J. Larazzara, 'Cell Signaling Regulation by Protein Phosphorylation: A Multivariate, Heterogeneous, and Context-dependent Process', vol. 40, pp. 185–192, 2016, doi: 10.1016/j.copbio.2016.06.005.

[82] B. M. Manning and L. C. Cantley, 'AKT/PKB Signaling: Navigating Downstream Brendan', *Cell*, vol. 129, no. 7, pp. 1261–1274, 2007, doi: 10.1016/j.cell.2007.06.009.AKT/PKB.

[83] X. Xu *et al.*, 'The CUL7 E3 Ubiquitin Ligase Targets Insulin Receptor Substrate 1 for Ubiquitin-Dependent Degradation', *Mol. Cell*, vol. 30, no. 4, pp. 403–414, May 2008, doi: 10.1016/j.molcel.2008.03.009.

[84] B. Majc, T. Sever, M. Zarić, B. Breznik, B. Turk, and T. T. Lah, 'Epithelial-to-mesenchymal transition as the driver of changing carcinoma and glioblastoma microenvironment', *Biochim. Biophys. Acta Mol. Cell Res.*, vol. 1867, no. 10, p. 118782, 2020, doi: 10.1016/j.bbamcr.2020.118782.

[85] T. Hunter, 'Protein Kinases and Phosphatases : The Yin and Yang of Protein Phosphorylation and Signaling', vol. 80, pp. 225–236, 1995.

[86] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam, 'Materials and Methods for "The protein kinase complement of the human genome"', *Science*, vol. 298, no. December, 2002.

[87] Kinase.com, 'Kinase.com'. Accessed: May 28, 2019. [Online]. Available: http://kinase.com/web/current/

[88] M. J. Chen, J. E. Dixon, and G. Manning, 'Genomics and evolution of protein phosphatases', *Sci. Signal.*, vol. 10, no. 474, pp. 1–18, 2017, doi: 10.1126/scisignal.aag1796.

[89] S. K. Hanks and T. Hunter, 'The eukaryotic protein kinase superfamily : (catalytic) domam structure and classification of the', *FASEB Journal*, 1988.

[90] S. K. Hanks and T. Hunter, 'The eukaryotic protein kinase superfamily : (catalytic) domam structure and classification of the', *FASEB Journal*, 1988.

[91] P. Sérgio *et al.*, 'Revisiting protein kinase-substrate interactions: Toward therapeutic development The Catalytic Domain of Eukaryotic Protein Kinases', *Biochemistry*, vol. 9, no. 420, p. DOI: 10.1126/scisignal.aad4016, 2016, [Online]. Available: www.SCIENCESIGNALING.org

[92] R. Roskoski, *Enzyme Structure and Function*☆, no. August. Elsevier Inc., 2014. doi: 10.1016/B978-0-12-801238-3.05007-8.

[93] U. Knippschild, A. Gocht, S. Wolff, N. Huber, J. Löhler, and M. Stöter, 'The casein kinase 1 family: Participation in multiple cellular processes in eukaryotes', *Cell. Signal.*, vol. 17, no. 6, pp. 675–689, 2005, doi: 10.1016/j.cellsig.2004.12.011.

[94] M. Varjosalo *et al.*, 'The Protein Interaction Landscape of the Human CMGC Kinase Group', *Cell Rep.*, 2013, doi: 10.1016/j.celrep.2013.03.027.

[95] U. Knippschild, A. Gocht, S. Wolff, N. Huber, J. Löhler, and M. Stöter, 'The casein kinase 1 family: Participation in multiple cellular processes in eukaryotes', *Cell. Signal.*, vol. 17, no. 6, pp. 675–689, 2005, doi: 10.1016/j.cellsig.2004.12.011.

[96] T. Hunter and J. A. Cooper, 'Protein-tyrosine kinases 1', *Annual Reviews Biochemistry*, vol. 54, pp. 897–930, 1985.

[97] M. Grozdanić *et al.*, 'Cathepsin L-mediated EGFR cleavage affects intracellular signalling pathways in cancer', *Biol. Chem.*, vol. 405, no. 4, pp. 283–296, Apr. 2024, doi: 10.1515/hsz-2023-0213.

[98] Y. Zhou, W. Li, Y. Tseng, J. Zhang, and J. Liu, 'Developing slow-off dickkopf-1 aptamers for early-diagnosis of hepatocellular carcinoma', *Talanta*, vol. 194, no. October 2018, pp. 422–429, 2019, doi: 10.1016/j.talanta.2018.10.014.

[99] K. C. Duong-Ly and J. R. Peterson, 'The Human Kinome and Kinase Inhibition as a therapeutic strategy', *Curr Protoc Pharmacolo.*, vol. 60, no. 1, pp. 2.9.1-2.9.14, 2013, doi: 10.1002/0471141755.ph0209s60.The.

[100] A. Reményi, M. C. Good, and W. A. Lim, 'Docking interactions in protein kinase and phosphatase networks', *Curr. Opin. Struct. Biol.*, vol. 16, no. 6, pp. 676–685, 2006, doi: 10.1016/j.sbi.2006.10.008.

[101] Á. Garai *et al.*, 'Europe PMC Funders Group Specificity of Linear Motifs That Bind to a Common Mitogen- Activated Protein Kinase Docking Groove *', *Sci. Signal.*, vol. 5, no. 245, pp. 1–28, 2013, doi: 10.1126/scisignal.2003004.Specificity.

[102] Z. Wang and P. A. Cole, 'Catalytic mechanisms and regulation of protein kinases', *Methods Enzymol.*, vol. 548, no. C, pp. 1–21, 2014, doi: 10.1016/B978-0-12-397918-6.00001-X.

[103] W. Peti, A. C. Nairn, and R. Page, 'Structural basis for protein phosphatase 1 regulation and specificity', *FEBS Journal*, vol. 280, no. 2, pp. 596–611, 2013, doi: 10.1111/j.1742-4658.2012.08509.x.

[104] Y. Xu *et al.*, 'Structure of the Protein Phosphatase 2A Holoenzyme', *Cell*, vol. 127, no. 6, pp. 1239–1251, 2006, doi: 10.1016/j.cell.2006.11.033.

[105] P. Seshacharyulu, P. Pandey, K. Datta, and S. K. Batra, 'Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer', *Cancer Lett.*, vol. 335, no. 1, pp. 9–18, 2013, doi: 10.1016/j.canlet.2013.02.036.

[106] N. K. Tonks, 'Protein tyrosine phosphatases: From genes, to function, to disease', *Nat. Rev. Mol. Cell Biol.*, vol. 7, no. 11, pp. 833–846, 2006, doi: 10.1038/nrm2039.

[107] A. Jannik *et al.*, 'Structural and Evolutionary Relationships among Protein Tyrosine Phosphatase Domains', *Mol. Cell. Biol.*, vol. 21, no. 21, pp. 7117–7136, 2001, doi: 10.1128/MCB.21.21.7117.

[108] Y. Zhang, M. Zhang, and Y. Zhang, 'Crystal structure of Ssu72, an essential eukaryotic phosphatase specific for the C-terminal domain of RNA polymerase II, in complex with a transition state analogue', *Biochemical Journal*, vol. 434, no. 3, pp. 435–444, 2011, doi: 10.1042/bj20101471.

[109] Y. Shi, 'Serine/Threonine Phosphatases: Mechanism through Structure', *Cell*, vol. 139, no. 3, pp. 468–484, 2009, doi: 10.1016/j.cell.2009.10.006.

[110] A. Seifried, J. Schultz, and A. Gohla, 'Human HAD phosphatases: Structure, mechanism, and roles in health and disease', *FEBS Journal*, vol. 280, no. 2, pp. 549–571, 2013, doi: 10.1111/j.1742-4658.2012.08633.x.

[111] G. Swarup, S. Cohen, and D. L. Garbers, 'Selective dephosphorylation of proteins containing phosphotyrosine by alkaline phosphatases.', *Journal of Biological Chemistry*, vol. 256, no. 15, pp. 8197–8201, 1981.

[112] D. J. Rigden, 'The histidine phosphatase superfamily: structure and function', *Biochemical Journal*, vol. 409, no. 2, pp. 333–348, 2008, doi: 10.1042/bj20071097.

[113] S. K. Hindupur *et al.*, 'The protein histidine phosphatase LHPP is a tumour suppressor', *Nature*, vol. 555, no. 7698, pp. 678–682, 2018, doi: 10.1038/nature26140.

[114] P. L. Hsu *et al.*, 'Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr1 and Ser5 on the RNA Polymerase II CTD', *J. Mol. Biol.*, vol. 16, no. 426, pp. 2970–2981, 2014, doi: doi:10.1016/j.jmb.2014.06.010.

[115] S. R. Offley and M. C. Schmidt, 'Protein phosphatases of Saccharomyces cerevisiae', *Curr. Genet.*, vol. 65, no. 1, pp. 41–55, 2019, doi: 10.1007/s00294-018-0884-y.

[116] A. Nita-Lazar, H. Saito-Benz, and F. M. White, 'Quantitative phosphoproteomics by mass spectrometry: Past, present, and future', Nov. 2008. doi: 10.1002/pmic.200800231.

[117] J. R. Porambo, A. M. Salicioni, P. E. Visconti, and M. D. Platt, 'Sperm phosphoproteomics: Historical perspectives and current methodologies', Oct. 2012. doi: 10.1586/epr.12.41.

[118] G. Muneer, C. S. Chen, and Y. J. Chen, 'Advancements in Global Phosphoproteomics Profiling: Overcoming Challenges in Sensitivity and Quantification', Jan. 01, 2025, *John Wiley and Sons Inc.* doi: 10.1002/pmic.202400087.

[119] M. A. Jünger and R. Aebersold, 'Mass spectrometry-driven phosphoproteomics: Patterning the systems biology mosaic', Jan. 2014. doi: 10.1002/wdev.121.

[120] A. B. Iliuk, J. V. Arrington, and W. A. Tao, 'Analytical challenges translating mass spectrometry-based phosphoproteomics from discovery to clinical applications', Dec. 01, 2014, *Wiley-VCH Verlag.* doi: 10.1002/elps.201400153.

[121] N. M. Lancaster *et al.*, 'Fast and deep phosphoproteome analysis with the Orbitrap Astral mass spectrometer', *Nat. Commun.*, vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-51274-0.

[122] J. Fílla and D. Honys, 'Enrichment techniques employed in phosphoproteomics', Jan. 01, 2012, *Springer-Verlag Wien*. doi: 10.1007/s00726-011-1111-z.

[123] W. J. Qian, M. B. Goshe, D. G. Camp, L. R. Yu, K. Tang, and R. D. Smith, 'Phosphoprotein Isotope-Coded Solid-Phase Tag Approach for Enrichment and Quantitative Analysis of Phosphopeptides from Complex Mixtures', *Anal. Chem.*, vol. 75, no. 20, pp. 5441–5450, Oct. 2003, doi: 10.1021/ac0342774.

[124] A. Silva, R. E. F. Boto, C. T. Tomaz, J. A. Queiroz, and P. Almeida, 'Rhodamine B as Ligand for Affinity Chromatography: Chromatographic Studies on Derivatized Beaded Cellulose', *J. Chromatogr. Sci.*, vol. 48, no. 3, pp. 240–244, Mar. 2010, doi: 10.1093/chromsci/48.3.240.

[125] S. Alamos-Musre *et al.*, 'From Structure to Function: The Promise of PAMAM Dendrimers in Biomedical Applications', Jul. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/pharmaceutics17070927.

[126] K. Engholm-Keller and M. R. Larsen, 'Technologies and challenges in large-scale phosphoproteomics', *Proteomics*, vol. 13, no. 6, pp. 910–931, 2013, doi: 10.1002/pmic.201200484.

[127] C. Yang, X. Zhong, and L. Li, 'Recent advances in enrichment and separation strategies for mass spectrometry-based phosphoproteomics', Dec. 01, 2014, *Wiley-VCH Verlag*. doi: 10.1002/elps.201400017.

[128] I. L. Batalha, C. R. Lowe, and A. C. A. Roque, 'Platforms for enrichment of phosphorylated proteins and peptides in proteomics', *Trends Biotechnol.*, vol. 30, no. 2, pp. 100–110, 2012, doi: 10.1016/j.tibtech.2011.07.004.

[129] A. Leitner, 'Phosphopeptide enrichment using metal oxide affinity chromatography', *TrAC - Trends in Analytical Chemistry*, vol. 29, no. 2, pp. 177–185, 2010, doi: 10.1016/j.trac.2009.08.007.

[130] Y. Zhang, B. R. Fonslow, B. Shan, M. C. Baek, and J. R. Yates, 'Protein analysis by shotgun/bottom-up proteomics', *Chem. Rev.*, vol. 113, no. 4, pp. 2343–2394, 2013, doi: 10.1021/cr3003533.

[131] K. Engholm-Keller and M. R. Larsen, 'Technologies and challenges in large-scale phosphoproteomics', *Proteomics*, vol. 13, no. 6, pp. 910–931, 2013, doi: 10.1002/pmic.201200484.

[132] J. Kim, K. Petritis, Y. Shen, G. D. Camp II, R. J. Moore, and R. D. Smith, 'Phosphopeptide elution times in reversed-phase liquid chromatography', *J. Chromatogr. A*, vol. 1172, no. 1, pp. 9–18, 2007.

[133] R. F. Dreier, E. Ahrné, P. Broz, and A. Schmidt, 'Global ion suppression limits the potential of mass spectrometry based phosphoproteomics.', *J. Proteome Res.*, 2018, doi: 10.1021/acs.jproteome.8b00812.

[134] H. Ding, L. X. Li, P. C. Harris, J. Yang, and X. Li, 'Extracellular vesicles and exosomes generated from cystic renal epithelial cells promote cyst growth in autosomal dominant polycystic kidney disease', *Nat. Commun.*, vol. 12, no. 1, Dec. 2021, doi: 10.1038/s41467-021-24799-x.

[135] M. M. Rinschen *et al.*, 'Phosphoproteomic analysis reveals regulatory mechanisms at the kidney filtration barrier', *Journal of the American Society of Nephrology*, vol. 25, no. 7, pp. 1509–1522, Jul. 2014, doi: 10.1681/ASN.2013070760.

[136] E. Drummond *et al.*, 'Phosphorylated tau interactome in the human Alzheimer's disease brain', *Brain*, vol. 143, no. 9, pp. 2803–2817, Sep. 2020, doi: 10.1093/brain/awaa223.

[137] M. Adli, E. Merkhofer, P. Cogswell, and A. S. Baldwin, 'IKKα and IKKβ each function to regulate NF-κB activation in the TNF-induced/canonical pathway', *PLoS One*, vol. 5, no. 2, Feb. 2010, doi: 10.1371/journal.pone.0009428.

[138] C. Brandts *et al.*, 'Activation of AKT1 in Hematopoietic Stem Cells Causes a Myeloproliferative Disease in a Tet-Inducible Mouse Model.', *Blood*, vol. 112, no. 11, pp. 1356–1356, Nov. 2008, doi: 10.1182/blood.v112.11.1356.1356.

[139] S. M. Siddique *et al.*, 'PKN1 kinase-negative knock-in mice develop splenomegaly and leukopenia at advanced age without obvious autoimmune-like phenotypes', *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-50419-2.

[140] A. Aguilar-Valles *et al.*, 'Translational control of depression-like behavior via phosphorylation of eukaryotic translation initiation factor 4E', *Nat. Commun.*, vol. 9, no. 1, Dec. 2018, doi: 10.1038/s41467-018-04883-5.

[141] Y. Yan *et al.*, 'Src inhibition blocks renal interstitial fibroblast activation and ameliorates renal fibrosis', *Kidney Int.*, vol. 89, no. 1, pp. 68–81, Jan. 2016, doi: 10.1038/ki.2015.293.

[142] G. Miller, S. P. Matthews, T. Reinheckel, S. Fleming, and C. Watts, 'Asparagine endopeptidase is required for normal kidney physiology and homeostasis', *The FASEB Journal*, vol. 25, no. 5, pp. 1606–1617, May 2011, doi: 10.1096/fj.10-172312.

[143] J. Martínez-Fábregas *et al.*, 'Lysosomal protease deficiency or substrate overload induces an oxidative-stress mediated STAT3-dependent pathway of lysosomal homeostasis', *Nat. Commun.*, vol. 9, no. 1, p. 5343, Dec. 2018, doi: 10.1038/s41467-018-07741-6.

[144] Y. Tao, C. Tang, J. Wei, Y. Shan, X. Fang, and Y. Li, 'Nr4a1 promotes renal interstitial fibrosis by regulating the p38 MAPK phosphorylation', *Molecular Medicine*, vol. 29, no. 1, Dec. 2023, doi: 10.1186/s10020-023-00657-y.

[145] M. K. Song, B. B. Park, and J. E. Uhm, 'Understanding splenomegaly in myelofibrosis: Association with molecular pathogenesis', Mar. 18, 2018, *MDPI AG*. doi: 10.3390/ijms19030898.

[146] A. Pascarella *et al.*, 'Monocytes From Patients With Macrophage Activation Syndrome and Secondary Hemophagocytic Lymphohistiocytosis Are Hyperresponsive to Interferon Gamma', *Front. Immunol.*, vol. 12, Mar. 2021, doi: 10.3389/fimmu.2021.663329.

[147] F. He *et al.*, 'NRF2 activates growth factor genes and downstream AKT signaling to induce mouse and human hepatomegaly', *J. Hepatol.*, vol. 72, no. 6, pp. 1182–1195, Jun. 2020, doi: 10.1016/j.jhep.2020.01.023.

[148] J. Cox and M. Mann, 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nat. Biotechnol.*, vol. 26, no. 12, pp. 1367–1372, 2008, doi: 10.1038/nbt.1511.

[149] J. Vandesompele *et al.*, 'Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes', 2002. [Online]. Available: http://genomebiology.com/2002/3/7/research/0034.1Correspondence:.rankSpeleman.

[150] M. Ahmed and D. R. Kim, 'pcr: An R package for quality assessment, analysis and testing of qPCR data', *PeerJ*, vol. 2018, no. 3, 2018, doi: 10.7717/peerj.4473.

[151] J. V. Olsen *et al.*, 'Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks', *Cell*, vol. 127, no. 3, pp. 635–648, Nov. 2006, doi: 10.1016/j.cell.2006.09.026.

[152] V. Thumuluri, J. J. Almagro Armenteros, A. R. Johansen, H. Nielsen, and O. Winther, 'DeepLoc 2.0: multi-label subcellular localization prediction using protein language models', *Nucleic Acids Res.*, vol. 50, no. W1, pp. W228–W234, Jul. 2022, doi: 10.1093/nar/gkac278.

[153] L. Kolberg, U. Raudvere, I. Kuzmin, P. Adler, J. Vilo, and H. Peterson, 'G:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update)', *Nucleic Acids Res.*, vol. 51, no. W1, pp. W207–W212, Jul. 2023, doi: 10.1093/nar/gkad347.

[154] D. Szklarczyk *et al.*, 'STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets', *Nucleic Acids Res.*, vol. 47, no. D1, pp. D607–D613, Jan. 2019, doi: 10.1093/nar/gky1131.

[155] P. V. Hornbeck, B. Zhang, B. Murray, J. M. Kornhauser, V. Latham, and E. Skrzypek, 'PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations', *Nucleic Acids Res.*, vol. 43, no. D1, pp. D512–D520, Jan. 2015, doi: 10.1093/nar/gku1267.

[156] K. Pang *et al.*, 'Role of protein phosphorylation in cell signaling, disease, and the intervention therapy', Dec. 01, 2022, *John Wiley and Sons Inc.* doi: 10.1002/mco2.175.

[157] H. R. Bourne, D. A. Sanders, and F. McCormick, 'The GTPase superfamily: conserved structure and molecular mechanism', *Nature*, vol. 349, no. 6305, pp. 117–127, Jan. 1991, doi: 10.1038/349117a0.

[158] A. Sun *et al.*, 'Legumain Is an Endogenous Modulator of Integrin αvβ3 Triggering Vascular Degeneration, Dissection, and Rupture', *Circulation*, vol. 145, pp. 659–674, 2022, doi: 10.1161/CIRCULATIONAHA.121.056640.

[159] Z. Papackova and M. Cahova, 'Fatty acid signaling: The new function of intracellular lipases', Feb. 10, 2015, *MDPI AG.* doi: 10.3390/ijms16023831.

[160] H. Yoshizaki and S. Okuda, 'Large-scale analysis of the evolutionary histories of phosphorylation motifs in the human genome', *Gigascience*, vol. 4, no. 1, 2015, doi: 10.1186/s13742-015-0057-6.

[161] A. E. Leroux, J. O. Schulze, and R. M. Biondi, 'AGC kinases, mechanisms of regulation and innovative drug development', Feb. 01, 2018, *Academic Press.* doi: 10.1016/j.semcancer.2017.05.011.

[162] C. Borgo, C. D'Amore, S. Sarno, M. Salvi, and M. Ruzzene, 'Protein kinase CK2: a potential therapeutic target for diverse human diseases', Dec. 01, 2021, *Springer Nature.* doi: 10.1038/s41392-021-00567-7.

[163] Y. He *et al.*, 'CD4+ T-Cell Legumain Deficiency Attenuates Hypertensive Damage via Preservation of TRAF6', *Circ. Res.*, vol. 134, no. 1, pp. 9–29, Jan. 2024, doi: 10.1161/CIRCRESAHA.123.322835.

[164] Y. Liu, J. Chen, S. K. Fontes, E. N. Bautista, and Z. Cheng, 'Physiological and pathological roles of protein kinase A in the heart', Feb. 01, 2022, *Oxford University Press.* doi: 10.1093/cvr/cvab008.

[165] F. Ando, Y. Hara, and S. Uchida, 'The Journal of Physiology Identification of protein kinase A signalling molecules in renal collecting ducts', *J Physiol*, vol. 602, pp. 3057–3067, 2024, doi: 10.1113/JP284178#support-information-section.

[166] A. Filippone *et al.*, 'Protein Kinase Inhibitors as a New Target for Immune System Modulation and Brain Cancer Management', Dec. 01, 2022, *MDPI.* doi: 10.3390/ijms232415693.

[167] V. L. Wehbi and K. Taskén, 'Molecular mechanisms for cAMP-mediated immunoregulation in T cells - role of anchored protein kinase a signaling units', Jun. 08, 2016, *Frontiers Research Foundation.* doi: 10.3389/fimmu.2016.00222.

[168] D. Pérez-Pérez, L. Santos-Argumedo, J. C. Rodríguez-Alba, and G. López-Herrera, 'Role of Protein Kinase A Activation in the Immune System with an Emphasis on Lipopolysaccharide-Responsive and

Beige-like Anchor Protein in B Cells', Feb. 01, 2023, *MDPI*. doi: 10.3390/ijms24043098.

[169] I. R. Gupta *et al.*, 'Protein kinase A is a negative regulator of renal branching morphogenesis and modulates inhibitory and stimulatory bone morphogenetic proteins', *Journal of Biological Chemistry*, vol. 274, no. 37, pp. 26305–26314, Sep. 1999, doi: 10.1074/jbc.274.37.26305.

[170] X. Wang *et al.*, 'Protein Kinase A Downregulation Delays the Development and Progression of Polycystic Kidney Disease', *Journal of the American Society of Nephrology*, vol. 33, no. 6, pp. 1087–1104, Jun. 2022, doi: 10.1681/ASN.2021081125.

[171] F. Liu and S. Zhuang, 'Role of receptor tyrosine kinase signaling in renal fibrosis', Jun. 20, 2016, *MDPI AG*. doi: 10.3390/ijms17060972.

[172] C. Yang and X. Wang, 'Lysosome biogenesis: Regulation and functions', 2021, *Rockefeller University Press*. doi: 10.1083/jcb.202102001.

[173] R. V. Schillace *et al.*, 'A-kinase anchoring in Dendritic cells is required for antigen presentation', *PLoS One*, vol. 4, no. 3, Mar. 2009, doi: 10.1371/journal.pone.0004807.

[174] M. Majewski, T. O. Bose, F. C. M. Sillé, A. M. Pollington, E. Fiebiger, and M. Boes, 'Protein kinase C delta stimulates antigen presentation by Class II MHC in murine dendritic cells', *Int. Immunol.*, vol. 19, no. 6, pp. 719–732, Jun. 2007, doi: 10.1093/intimm/dxm034.

[175] Z. H. Wang, P. Liu, X. Liu, S. P. Yu, J. Z. Wang, and K. Ye, 'Delta-secretase (AEP) mediates tau-splicing imbalance and accelerates cognitive decline in tauopathies', *Journal of Experimental Medicine*, vol. 215, no. 12, pp. 3038–3056, Dec. 2018, doi: 10.1084/jem.20180539.

[176] S. U. Khan *et al.*, 'Role of LGMN in tumor development and its progression and connection with the tumor microenvironment', Feb. 07, 2023, *Frontiers Media S.A.* doi: 10.3389/fmolb.2023.1121964.

[177] İ. A. Ilık, M. Malszycki, A. K. Lübke, C. Schade, D. Meierhofer, and T. Aktaş, 'Son and srrm2 are essential for nuclear speckle formation', *Elife*, vol. 9, pp. 1–48, Oct. 2020, doi: 10.7554/eLife.60579.

[178] B. E. Aubol, P. Serrano, L. Fattet, K. Wüthrich, and J. A. Adams, 'Molecular interactions connecting the function of the serine-arginine-rich protein SRSF1 to protein phosphatase 1', *Journal of Biological Chemistry*, vol. 293, no. 43, pp. 16751–16760, Oct. 2019, doi: 10.1074/jbc.RA118.004587.

[179] S. McCracken, M. Lambermon, and B. J. Blencowe, 'SRm160 Splicing Coactivator Promotes Transcript 3′-End Cleavage', *Mol. Cell. Biol.*, vol. 22, no. 1, pp. 148–160, Jan. 2002, doi: 10.1128/mcb.22.1.148-160.2002.

[180] S. Xu, S. K. Lai, D. Y. Sim, W. S. L. Ang, H. Y. Li, and X. Roca, 'SRRM2 organizes splicing condensates to regulate alternative

splicing', *Nucleic Acids Res.*, vol. 50, no. 15, pp. 8599–8614, 2022, doi: 10.1093/nar/gkac669.

[181] S. Carvalho *et al.*, 'SRRM2 splicing factor modulates cell fate in early development', *Biol. Open*, vol. 13, no. 4, Apr. 2024, doi: 10.1242/bio.060415.

[182] M. Zhang *et al.*, 'Phosphorylation-dependent charge blocks regulate the relaxation of nuclear speckle networks', *Mol. Cell*, vol. 85, no. 9, pp. 1760-1774.e7, May 2025, doi: 10.1016/j.molcel.2025.03.016.

[183] V. Evdokimova, 'Y-box Binding Protein 1: Looking Back to the Future', Jan. 01, 2022, *Pleiades journals.* doi: 10.1134/S0006297922140024.

[184] L. Prabhu, A. V. Hartley, M. Martin, F. Warsame, E. Sun, and T. Lu, 'Role of post-translational modification of the Y box binding protein 1 in human cancers', Sep. 01, 2015, *Chongqing yi ke da xue, di 2 lin chuang xue yuan Bing du xing gan yan yan jiu suo.* doi: 10.1016/j.gendis.2015.05.001.

[185] Q. Yin, M. Zheng, Q. Luo, D. Jiang, H. Zhang, and C. Chen, 'YB-1 as an Oncoprotein: Functions, Regulation, Post-Translational Modifications, and Targeted Therapy', Apr. 01, 2022, *MDPI.* doi: 10.3390/cells11071217.

[186] B. W. Sutherland *et al.*, 'Akt phosphorylates the Y-box binding protein 1 at Ser102 located in the cold shock domain and affects the anchorage-independent growth of breast cancer cells', *Oncogene*, vol. 24, no. 26, pp. 4281–4292, Jun. 2005, doi: 10.1038/sj.onc.1208590.

[187] E. M. Sogorina *et al.*, 'Yb-1 phosphorylation at serine 209 inhibits its nuclear translocation', *Int. J. Mol. Sci.*, vol. 23, no. 1, Jan. 2022, doi: 10.3390/ijms23010428.

[188] A. Kariminia *et al.*, 'Y-box-binding protein 1 contributes to IL-7-Mediated survival signaling in B-cell precursor acute lymphoblastic leukemia', *Oncol. Lett.*, vol. 13, no. 1, pp. 497–505, Jan. 2017, doi: 10.3892/ol.2016.5437.

[189] J. Zhang *et al.*, 'Structural basis of DNA binding to human YB-1 cold shock domain regulated by phosphorylation', *Nucleic Acids Res.*, vol. 48, no. 16, pp. 9361–9371, Sep. 2020, doi: 10.1093/nar/gkaa619.

[190] E. Kwon *et al.*, 'The RNA-binding protein YBX1 regulates epidermal progenitors at a posttranscriptional level', *Nat. Commun.*, vol. 9, no. 1, Dec. 2018, doi: 10.1038/s41467-018-04092-0.

[191] X. Zhong *et al.*, 'ERK/RSK-mediated phosphorylation of Y-box binding protein-1 aggravates diabetic cardiomyopathy by suppressing its interaction with deubiquitinase OTUB1', *Journal of Biological Chemistry*, vol. 298, no. 6, Jun. 2022, doi: 10.1016/j.jbc.2022.101989.

[192] L. Prabhu *et al.*, 'Critical role of phosphorylation of serine 165 of YBX1 on the activation of NF-κB in colon cancer', 2015. [Online]. Available: www.impactjournals.com/oncotarget/

[193] N. T. M. Dinh, T. M. Nguyen, M. K. Park, and C. H. Lee, 'Y-Box Binding Protein 1: Unraveling the Multifaceted Role in Cancer Development and Therapeutic Potential', Jan. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/ijms25020717.

[194] T. Nöthen *et al.*, 'DNA-Dependent Protein Kinase Mediates YB-1 (Y-Box Binding Protein)-Induced Double Strand Break Repair', *Arterioscler. Thromb. Vasc. Biol.*, vol. 43, no. 2, pp. 300–311, Feb. 2023, doi: 10.1161/ATVBAHA.122.317922.

[195] W. Zhang, Y. Liu, Z. Zhao, Y. Zhang, Y. Liang, and W. Wang, 'YBX1: A Multifunctional Protein in Senescence and Immune Regulation', Dec. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/cimb46120841.

[196] L. Prabhu, A. V. Hartley, M. Martin, F. Warsame, E. Sun, and T. Lu, 'Role of post-translational modification of the Y box binding protein 1 in human cancers', Sep. 01, 2015, *Chongqing yi ke da xue, di 2 lin chuang xue yuan Bing du xing gan yan yan jiu suo*. doi: 10.1016/j.gendis.2015.05.001.

[197] A. S. Mehta *et al.*, 'Critical role for cold shock protein YB-1 in cytokinesis Running title : Cold shock protein YB-1 controls cytokinesis', 2020, doi: 10.1101/2020.03.18.997817.

[198] X. Li *et al.*, 'PLK1 inhibition promotes apoptosis and DNA damage in glioma stem cells by regulating the nuclear translocation of YBX1', *Cell Death Discov.*, vol. 9, no. 1, Dec. 2023, doi: 10.1038/s41420-023-01302-7.

[199] S. Mao *et al.*, 'Apurinic/apyrimidinic endodeoxyribonuclease 1 (APE1) promotes stress granule formation via YBX1 phosphorylation in ovarian cancer', *Cellular and Molecular Life Sciences*, vol. 81, no. 1, Dec. 2024, doi: 10.1007/s00018-023-05086-y.

[200] T. Liu, L. Zhang, D. Joo, and S. C. Sun, 'NF-κB signaling in inflammation', 2017, *Springer Nature*. doi: 10.1038/sigtrans.2017.23.

[201] Y. Carrier *et al.*, 'Inter-regulation of Th17 cytokines and the IL-36 cytokines in vitro and in vivo: Implications in psoriasis pathogenesis', *Journal of Investigative Dermatology*, vol. 131, no. 12, pp. 2428–2437, 2011, doi: 10.1038/jid.2011.234.

[202] J. E. Towne *et al.*, 'Interleukin-36 (IL-36) ligands require processing for full agonist (IL-36α, IL-36β, and IL-36γ) or antagonist (IL-36Ra) activity', *Journal of Biological Chemistry*, vol. 286, no. 49, pp. 42594–42602, Dec. 2011, doi: 10.1074/jbc.M111.267922.

[203] A. M. Foster *et al.*, 'IL-36 Promotes Myeloid Cell Infiltration, Activation, and Inflammatory Activity in Skin', *The Journal of Immunology*, vol. 192, no. 12, pp. 6053–6061, Jun. 2014, doi: 10.4049/jimmunol.1301481.

[204] P. P. Tak and G. S. Firestein, 'NF-κB: A key role in inflammatory diseases', 2001, *The American Society for Clinical Investigation*. doi: 10.1172/JCI11830.

[205] D. C. arolina T. Palomino and L. C. avalheiro Marti, 'Chemokines and immunity', Jul. 01, 2015. doi: 10.1590/S1679-45082015RB3438.

[206] B. Z. Qian *et al.*, 'CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis', *Nature*, vol. 475, no. 7355, pp. 222–225, Jul. 2011, doi: 10.1038/nature10138.

[207] T. Yoshimura, T. Baba, K. S. Midwood, M. Gschwandtner, and R. Derler, 'More Than Just Attractive: How CCL2 Influences Myeloid Cell Behavior Beyond Chemotaxis', *Frontiers in Immunology | www.frontiersin.org*, vol. 10, p. 2759, 2019, doi: 10.3389/fimmu.2019.02759.

[208] B. Wolszczak-Biedrzycka *et al.*, 'Chemokines as potential biomarkers for predicting the course of COVID-19 – a review of the literature', *Front. Immunol.*, vol. 16, Sep. 2025, doi: 10.3389/fimmu.2025.1662643.

[209] C. Wang, J. Wang, Z. Zhu, J. Hu, and Y. Lin, 'Spotlight on pro-inflammatory chemokines: regulators of cellular communication in cognitive impairment', 2024, doi: 10.3389/fimmu.2024.1421076.

[210] Y. K. Takada, M. Fujita, and Y. Takada, 'Pro-Inflammatory Chemokines CCL5, CXCL12, and CX3CL1 Bind to and Activate Platelet Integrin αIIbβ3 in an Allosteric Manner', *Cells*, vol. 11, no. 19, Oct. 2022, doi: 10.3390/cells11193059.

[211] C. H. Hou, P. C. Chen, and J. F. Liu, 'CXCL1 enhances COX-II expression in rheumatoid arthritis synovial fibroblasts by CXCR2, PLC, PKC, and NF-κB signal pathway', *Int. Immunopharmacol.*, vol. 124, Nov. 2023, doi: 10.1016/j.intimp.2023.110909.

[212] K. V Sawant *et al.*, 'Chemokine CXCL1 mediated neutrophil recruitment: Role of glycosaminoglycan interactions', 2016, doi: 10.1038/srep33123.

[213] K. Thomas *et al.*, 'Alveolar epithelial and vascular CXCR2 mediates transcytosis of CXCL1 in inflamed lungs', doi: 10.1038/s41467-025-60174-w.

[214] L. Hanssen *et al.*, 'YB-1 Is an Early and Central Mediator of Bacterial and Sterile Inflammation In Vivo', *The Journal of Immunology*, vol. 191, no. 5, pp. 2604–2613, Sep. 2013, doi: 10.4049/jimmunol.1300416.

[215] A. L. Stratford *et al.*, 'Y-box binding protein-1 serine 102 is a downstream target of p90 ribosomal S6 kinase in basal-like breast cancer cells', *Breast Cancer Research*, vol. 10, no. 6, Nov. 2008, doi: 10.1186/bcr2202.

[216] A. G. Bader and P. K. Vogt, 'Phosphorylation by Akt disables the anti-oncogenic activity of YB-1', *Oncogene*, vol. 27, no. 8, pp. 1179–1182, Feb. 2008, doi: 10.1038/sj.onc.1210719.

[217] J. Lopez *et al.*, 'The ribosomal S6 kinase 2 (RSK2)–SPRED2 complex regulates the phosphorylation of RSK substrates and MAPK signaling', *Journal of Biological Chemistry*, vol. 299, no. 6, Jun. 2023, doi: 10.1016/j.jbc.2023.104789.

[218] K. Jürchott *et al.*, 'Identification of Y-box binding protein 1 as a core regulator of MEK/ERK pathway-dependent gene signatures in colorectal cancer cells', *PLoS Genet.*, vol. 6, no. 12, pp. 1–19, Dec. 2010, doi: 10.1371/journal.pgen.1001231.

[219] L. Z. Liao *et al.*, 'Y-box binding protein-1 promotes epithelial-mesenchymal transition in sorafenib-resistant hepatocellular carcinoma cells', *Int. J. Mol. Sci.*, vol. 22, no. 1, pp. 1–20, Jan. 2021, doi: 10.3390/ijms22010224.

[220] Z. Tang, B. Lin, W. Li, X. Li, F. Liu, and X. Zhu, 'Y-box binding protein 1 promotes chromatin accessibility to aggravate liver fibrosis', *Cell. Signal.*, vol. 109, Sep. 2023, doi: 10.1016/j.cellsig.2023.110750.

[221] E. R. Verrier *et al.*, 'Cell-based cccDNA reporter assay combined with functional genomics identifies YBX1 as HBV cccDNA host factor and antiviral candidate target', *Gut*, vol. 72, no. 9, pp. 1745–1757, Sep. 2023, doi: 10.1136/gutjnl-2020-323665.

[222] J. Wang *et al.*, 'YB-1 orchestrates onset and resolution of renal inflammation via IL10 gene regulation', *J. Cell. Mol. Med.*, vol. 21, no. 12, pp. 3494–3505, 2017, doi: 10.1111/jcmm.13260.

[223] E. Cantero-Navarro *et al.*, 'Role of Macrophages and Related Cytokines in Kidney Disease', Jul. 08, 2021, *Frontiers Media S.A.* doi: 10.3389/fmed.2021.688060.

[224] W. Wei, Y. Zhao, Y. Zhang, H. Jin, and S. Shou, 'The role of IL-10 in kidney disease', Jul. 01, 2022, *Elsevier B.V.* doi: 10.1016/j.intimp.2022.108917.

[225] I. Sinuani, I. Beberashvili, Z. Averbukh, and J. Sandbank, 'Role of IL-10 in the progression of kidney disease', *World J. Transplant.*, vol. 3, no. 4, p. 91, 2013, doi: 10.5500/wjt.v3.i4.91.

[226] B. Ke, C. Fan, W. Tu, and X. Fang, 'The Role of Y-Box Binding Protein 1 in Kidney Injury: Friend or Foe?', Apr. 01, 2018, *S. Karger AG.* doi: 10.1159/000488432.

[227] R. Rana *et al.*, 'Glomerular-tubular crosstalk via cold shock Y-box binding protein-1 in the kidney', *Kidney Int.*, vol. 105, no. 1, pp. 65–83, Jan. 2024, doi: 10.1016/j.kint.2023.09.014.

[228] J. Wang *et al.*, 'Therapeutic nuclear shuttling of YB-1 reduces renal damage and fibrosis', *Kidney Int.*, vol. 90, no. 6, pp. 1226–1237, Dec. 2016, doi: 10.1016/j.kint.2016.07.008.

[229] C. L. Hessman *et al.*, 'YB-1 interferes with tnfα–tnfr binding and modulates progranulin-mediated inhibition of TNFα signaling', *Int. J. Mol. Sci.*, vol. 21, no. 19, pp. 1–17, Oct. 2020, doi: 10.3390/ijms21197076.

[230] W. Lv, G. W. Booz, Y. Wang, F. Fan, and R. J. Roman, 'Inflammation and renal fibrosis: Recent developments on key signaling molecules as potential therapeutic targets', Feb. 05, 2018, *Elsevier B.V.* doi: 10.1016/j.ejphar.2017.12.016.

[231] M. Baues *et al.*, 'A collagen-binding protein enables molecular imaging of kidney fibrosis in vivo', *Kidney Int.*, vol. 97, no. 3, pp. 609–614, Mar. 2020, doi: 10.1016/j.kint.2019.08.029.

[232] Y. C. Ren *et al.*, 'Legumain promotes fibrogenesis in chronic pancreatitis via activation of transforming growth factor β1', *J. Mol. Med.*, vol. 98, no. 6, pp. 863–874, Jun. 2020, doi: 10.1007/s00109-020-01911-0.

Bibliography

Publications Related to the Thesis

Journal Articles

Sever, T., Sinožić, T., Kolarič, M., Turk, B., & Fonović, M. (2026). Phosphorylation of YBX1 in the Kidneys is Altered in Legumain Knockout-Mice. *Journal of Proteome Research*.

MAJC, Bernarda, SEVER, Tilen, ZARIĆ, Miki, BREZNIK, Barbara, TURK, Boris, LAH TURNŠEK, Tamara.Epithelial-to-mesenchymal transition as the driver of changing carcinoma and glioblastoma microenvironment.Biochimica et biophysica acta. BBA, Molecular cell research. [Print ed.]. Oct. 2020, vol. 1867, iss. 10, str. 118782,1-14, ilustr. ISSN 0167-4889. DiRROS - Digitalni repozitorij raziskovalnih organizacij Slovenije, DOI: 10.1016/j.bbamcr.2020.118782

GROZDANIĆ, Marija, SOBOTIČ, Barbara, BIASIZZO, Monika, SEVER, Tilen, VIDMAR, Robert, VIZOVIŠEK, Matej, TURK, Boris, FONOVIĆ, Marko. Cathepsin L-mediated EGFR cleavage affects intracellular signalling pathways in cancer. Biological chemistry. Apr. 2024, vol. 405, iss. 4, str. 283-296, ilustr. ISSN 1431-6730. Repozitorij Univerze v Ljubljani – RUL, DOI: 10.1515/hsz-2023-0213

Conference Paper

SEVER, Tilen, VIDMAR, Robert, TURK, Boris, FONOVIĆ, Marko. Phosphoproteomic analysis of legumain deficientmice. In: KISOVEC, Matic (ed.). Book of abstracts = Knjiga povzetkov. 13th Meeting of the Slovenian BiochemicalSociety with International Participation = 13. srečanje Slovenskega biokemijskega društva z mednarodno udeležbo,Dobrna, 24 - 27 September 2019. Ljubljana: Slovenian Biochemical Society, 2019. Str. 67. ISBN 978-961-93879-8-6.
SEVER, Tilen, VIDMAR, Robert, REINHECKEL, Thomas, TURK, Boris, FONOVIĆ, Marko. Phosphoproteomicanalysis of legumain deficient mice. FEBS open bio. 2021, vol. 11, iss. s1, str. 163. ISSN 2211-5463. DOI:10.1002/2211-5463.13205

SEVER, Tilen, VIDMAR, Robert, REINHECKEL, Thomas, TURK, Boris, FONOVIĆ, Marko. Phosphoproteomicanalysis of legumain deficient mice. In: KOZAK, Andreja (ed.), et al. Proteolysis at the interface between health and disease : FEBS ICGEB 2022 Advanced Course : 18-21 September, Bled, Slovenia : book of abstracts. Ljubljana: Jožef Stefan Institute, 2022. Str. 60. ISBN 978-961-264-229-7.

SEVER, Tilen, VIDMAR, Robert, REINHECKEL, Thomas, TURK, Boris, FONOVIĆ, Marko. Phosphoproteomic analysis of legumain deficient mice. In: SEPČIĆ, Kristina (ed.), PAVŠIČ, Miha (ed.). 15th Meeting of the Slovenian Biochemical Society with International Participation = 15. srečanje Slovenskega biokemijskega društva z mednarodno udeležbo :

book of abstracts = zbornik povzetkov : Portorož, 20-23 September 2023 = [20.–23. september 2023, Portorož, Slovenija]. Ljubljana: Slovenian Biochemical Society: = Slovensko biokemijsko društvo, 2023. Str. 141. ISBN 978-961-95941-1-7

SEVER, Tilen, SINOŽIĆ, Tea, KOLARIČ, Matej, TURK, Boris, FONOVIĆ, Marko. Phosphoproteomic analysis of legumain deficient mice. In: STOKA, Veronika (ed.), et al. Proteolysis : at the interface between health and disease :FEBS 2024 advanced course : Bled, Slovenia 14 - 18 September, 2024 : programme and book of abstracts. FEBS 2024 advanced course, Bled, Slovenia 14 - 18 September, 2024. Ljubljana: Jožef Stefan Institute, 2024. Str. 54. ISBN 978-961-264-295-2

Other Publications (optional)

Kolarič, M., Ivanovski, S., Sever, T., Turk, B., & Fonović, M. (2026). PeptideVisualizer: A Novel Software Solution for PROTOMAP Analysis. Journal of Proteome Research, 25(5), 2591-2597.

Biography

The author of this thesis obtained his Master's Degree in Molecular and Structural Biology in September 2018 at the Biotechnical Faculty, University of Ljubljana. In October of the same year, he enrolled in the Nanosciences and nanotechnologies programme at the Jožef StefanInternational Postgraduate School under the supervision of Prof. Dr. Marko Fonović. At the same time, he started working as a young researcher at the Jožef Stefan Institute at the Department of Biochemistry, Molecular and Structural Biology.