SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PH.D.)

CONSTRUCTION OF ANTIMICROBIAL AND IMMUNOMODULATORY PEPTIDE DATABASE AND *IN SILICO* ANALYSIS OF THE ANTIMICROBIAL PEPTIDES IN DIFFERENT PATHOLOGICAL CONDITIONS

By Ajneesh Kumar

Supervisor: Dr. Éva Csősz

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLECULAR, CELLULAR AND IMMUNE BIOLOGY
DEBRECEN, 2023

Construction of antimicrobial and immunomodulatory peptide database and *in silico* analysis of the antimicrobial peptides in different pathological conditions

By Ajneesh Kumar

Supervisor: Éva Csősz, PhD, DSc

Doctoral School of Molecular Cell and Immune Biology, University of Debrecen

Reviewers:	Dr. László Márk, PhD Prof. Dr. Attila Bácsi, PhD, DSc
Head of the **Defense Committee**:	Prof. Dr. István Balogh, PhD, DSc
Members of the Defense Committee:	Dr. Lilla Turiák, PhD Prof. Dr. Ferenc Erdődi, PhD, DSc

The PhD Defense takes place at the Lecture Hall of Bldg, Department of Obstetrics and Gynecology, University of Debrecen at 1 p.m. on 11th of January 2023

1. INTRODUCTION

Multicellular organisms protect themselves from harmful microbes. The skin, eyes, respiratory tract and lungs, mouth, digestive tract, urinary tract, and reproductive tract are the best-known places for the first encounter with microbes. Antimicrobial and immunomodulatory peptides (AMPs) are the first line of defense in various organisms, protecting our bodies from potential pathogens. AMPs are present in all body fluids; they can be found in saliva, sweat, blood, tear, and urine. AMPs have a broad spectrum of action and can be anionic, cationic, and amphipathic. They can be divided into three categories based on their origin: (i) natural, ribosomally synthesized peptides, (ii) non-ribosomally produced natural peptides, or (iii) non-natural synthetic peptides. AMPs are the component of the first line of defense. Several AMPs are essential in the innate immune system in different organisms from lower eukaryotes to humans. They can generate various responses, i.e., cell differentiation, cell recruitments, antibacterial, antiviral and antifungal responses, etc.

Understanding the AMP's mechanism of action will drive their development as potential therapeutic agents against pathogens. AMPs have different modes of action; two major types of action are direct killing and immunomodulation. The cationic AMPs are direct killers. They interact with the negative charge of the bacterial cell wall, form a pore-like structure, and disrupt the cell wall killing the bacteria. On the other hand, AMPs can do immunomodulation by recruiting the immune cells to kill the pathogens and control the inflammation at the site of infection. There are many naturally occurring AMPs, and their classification can be challenging. AMPs have a broad range of lengths; usually, they can vary from 10 to 150 amino acids. The most common way to classify AMPs is structural classification. Based on the primary, secondary and tertiary structures, the AMPs are of four types: (i) linear alpha-helical peptides, e.g., cathelicidin LL-37, magainin-2, (ii) linear peptides rich in particular amino acids, e.g., indolicidin, antibacterial protein PR-39 (PR-39), (iii) beta-sheets containing peptides, e.g., protegrin, huma defensin-5 (HD-5) dimer, and (iv) peptides with alpha and beta structural elements, e.g., phorimicin, beta defensin-1 (HBD-1). AMPs typically have distinct structural domains, such as a defensin fold in defensins and a cathelicidin domain in cathelicidins providing extra stability even in a harsh environment, such as high salt concentration, low pH, etc., but some, such as cationic intrinsically disordered peptides lack a well-defined three-dimensional structure. Endogenous AMPs are present in human body fluids. Many AMPs have been identified in the blood, tear, saliva, urine, and cerebrospinal fluid.

The tear film that coats the cornea and conjunctiva serves vital functions. It lubricates

the ocular surface epithelia, helps to provide a smooth surface for refracting light, supplies oxygen, and is an important component of the eye's innate defense system, protecting against various potential pathogens. A variety of microorganisms can infect the cornea and conjunctiva. Because it is in direct contact with the external environment, the tear film serves as the first line of defense against pathogens attempting to breach the physical barrier created by the ocular surface epithelia. The tear film helps washing out invading pathogens in conjunction with blinking and reflex tearing. It contains a plethora of AMPs, such as lysozymes, lipocalins, lacritins, and lactoferrins produced by lacrimal gland, or dermcidin produced by epithelial cells to facilitate pathogen killing or to inhibit microorganism replication.

Microorganisms and pathogens easily access the oral cavity and the rest of the body through the gastrointestinal tract. The oral cavity has a high microbial load, which can lead to a variety of oral diseases. To combat the various types of oral diseases, saliva contains a number of AMPs having a role in the host-defense. Because of its antibacterial, antioxidant, and antifungal properties, saliva, and oral mucosa, both serve as a potential line of defense to encounter pathogens. As a response of host-defense mechanism, many AMPs are produced by salivary glands and found in the saliva. The most abundant ones are alpha-amylase, mucins, cystatins, proline-rich peptides, and serum albumin. Some other AMPs were identified in the saliva, such as histatins, defensins, LL-37 cathelicidins, and salvic.

Blood is an essential component of the human body. It is also a crucial component of the innate immune system. Blood-borne immune components circulate freely in the human body thanks to the help of blood cells. Some of these so-called blood-born immune components are the AMPs. From the approximately 12,000 identified serum proteins, more than 400 can act in the first line of host defense. Various types of AMPs are deployed by the blood cells at the site of infection. Antimicrobial peptides can be present in blood cells, the neutrophil granules for example, contain AMPs such as lactoferrin, cathelicidins and lysozymes. AMPs derived from bigger proteins can also be found in blood such as haemoglobin-derived peptides, the hemocidins.

Urine is a significant bodily fluid with an important excretory function. It can also be the ground for a variety of microbial infections present in urinary tract. These types of infections are known as urinary tract infection (UTIs). UTIs are the most common type of the human infections. To protect the human body from these UTIs, AMPs are produced in the urinary tract. Urine contains more than 7000 proteins and many of the urine proteins are part of the defense system. The AMPs present in urine are mostly produced by the epithelial cells

in the urinary tract. The most abundant AMPs in urine were found to be serum albumin, uromodulin, alpha-1-microglobulin, kininogen and various immunoglobulin chains. AMPs such as ribonuclease 6 and 7 were also found in the urine and aid in urinary tract sterility. Other AMPs found in urine are neuropeptide Y, protachykinin-1, and vasoactive intestinal peptides. Cathelicidin is an another AMP found in lower concentrations in adult human urine and higher concentrations in children with urinary tract infections.

The cerebrospinal fluid (CSF) is found in ventricles, in the subarachnoid spaces of the cranium and spine. It helps in supplying nutrients, in waste elimination, and brain protection being a fluid coat and protecting the central nervous system. CSF is an essential fluid in the diagnosis of many neurodegenerative conditions, such as Alzheimer's disease and Parkinson's disease. The progression of certain neurodegenerative diseases can be monitored by examining CSF protein amounts. Until now, more than 4000 proteins have been identified in the human CSF. Approximately 8% of its proteins take part in host defense, for e.g., amyloid-beta precursor protein (APP), clusterin (CLU), and protein S100A7 (S100A7). It was also discovered that the level of LL-37 cathelicidin and defensins' in CSF is increased upon bacterial challenge.

Alzheimer's disease (AD) is a complex disease that, despite medical advances, is extremely difficult to diagnose. It is very critical to detect this disease in its early stages. The main pathological features of AD are amyloid-beta depositions in the brain and tau protein hyperphosphorylation, which results in neuropathological lesions. These neuropathological lesions cause/are the consequence of inflammation and can lead to synaptic loss, and neurodegeneration, which results in cognitive impairment, characteristic to AD. Several studies have identified some potential AMP biomarkers that might play a role in the pathological conditions related to AD, in addition to amyloid beta and tau protein. Presenillin (PSEN1, PSEN2), β-site APP -cleaving enzyme 1 (BACE1), alpha-1-antitrypsin (SERPINA1), apolipoprotein E (APOE), neurosecretory VGF (VGF), and complement components have been identified as potential biomarkers for AD. According to the pathogenic model of Alzheimer's disease, the presence of pathogens can cause the appearance of the disease. Aβ peptide has been found to activate microglia, resulting in increased release of pro-inflammatory cytokines such as interleukins, prostaglandins, and leukotrienes as part of an inflammatory response. Several scientific works have revealed that the APP plays a central role in AD. According to recent research, Aβ has a protective effect in the case of brain damage recovery and aids in the maintenance of the blood-brain barrier. Several studies have shown that amyloid-beta is active against at least eight different pathogens. However, in

the case of the AD brain, amyloid-beta also induces the action of LL-37 cathelicidin and other AMPs.

Oral cancer refers to a category of malignant tumors that affect the oral cavity, pharyngeal areas, and salivary glands. All of these malignancies are classified as oral squamous cell carcinoma (OSCC), accounting for up to 80-90 percent of all cancers of the head and neck region. Smoking, alcohol, UV-radiation, human papillomavirus (HPV) infection, *Candida albicans*, and dietary inadequacies are major risk factors of OSCC. OSCC is more common in elderly people and is diagnosed at an average age of 60 years. The OSCC's development takes several years. The buccal mucosa, the ventral and lateral sides of the tongue, the floor of the mouth, and the retromolar area are all common sites for OSCC in the oral cavity. OSCC can also be caused by precancerous abnormalities such as oral leukoplakia (OLK) or oral lichen planus (OLP). OLP is a disease characterized by inflammation affecting the mucous membranes of the mouth. OLP develops after viral infections such as herpes simplex, Epstein-Barr virus, human papillomavirus, and hepatitis C virus. OLK is described as firm, adhesive white patches on the oral mucosa that cannot be scratched. It can be of two kinds: white patches on the tongue or white patches on the tongue base. OLK and OLP are considered precancerous conditions, as in some circumstances, they can progress to OSCC. Regarding AMPs, many of them have been identified in the saliva, and some of the previously identified biomarkers for oral cancer and precancerous lesions belong to the AMP family. Such examples are S100A8, SLPI, LTF and CLU.

2. AIMS

Our main goal was to generate a unified human AMP database and to acquire information on the involvement of AMPs in different pathological conditions.

As far as AMPs have a role in innate immune response, which is involved in the pathophysiology of AD, and that some AMPs have been explicitly related to AD, our goal was to determine the extent to which AMPs may be involved in the pathogenesis of this neurodegenerative disease. We were excited to learn more about the AMPs discovered in the brain, cerebrospinal fluid, and blood in association with AD, and we wanted to investigate how AD affects the network of AMPs.

At the same time, we wanted to enhance data reutilization, as the reanalysis of genomics and proteomics datasets by bioinformatics approaches can be an important tool to examine large amounts of reliable data deposited in publicly accessible datasets. In our study,

we aimed to use high-quality proteomics data from publicly accessible sources to examine the presence and function of AMPs in AD.

In the case of the OSCC, OLP and OLK our main goal was to do an extensive network analysis of the salivary proteins and AMPs to obtain disease-related information. We also intended to do an extensive pathway analysis to find the biologically relevant information related to salivary proteins and AMPs.

3. METHODS

3.1 Generation of UDAMP antimicrobial and immunomodulatory peptide database

We have collected all human AMPs listed in Collection of Antimicrobial peptides (CAMP), Antimicrobial Peptide Database (APD), Database of Antimicrobial Peptides (dbAMP), Linking Antimicrobial Peptides (LAMP), and Database of Antimicrobial Activity and Structures of Peptides (DBAASP). We made a PubMed search between May 2020 and June 2020 using the keywords "human antimicrobial peptide" and listed all hits not present in online AMP databases. We have curated the retrieved data, removed the redundancies, and compiled a unified comprehensive human antimicrobial and immunomodulatory peptide database. We named our database as University of Debrecen Antimicrobial and Immunomodulatory Peptide (UDAMP) Database. UniProt IDs of the AMPs were used to generate the Gene IDs from the UniProt using its Applications Programming Interface (API). NCBI was used to find cross annotations and links between gene IDs and PubMed IDs. When we got our gene IDs, the "gene2pubmed" annotation file was used to retrieve the PubMed IDs of the genes through their gene IDs. E-utilities, a programming interface for the NCBI Entrez databases, was used to retrieve the article metadata based on PubMed IDs. We run a home-written script to retrieve an offline keyword search on the article metadata. As we got the high data redundancy in the AMP sequence, we crosschecked the individual protein/peptide sequences. They were matched against the UniProtKB/SwissProt proteome database using BLASTP with the following settings: the target organism was "Homo sapiens", the E threshold was set to 1e-10; scoring matrix: BLOSUM62 (default), filtering: none, gap cost: existence 11, extension: 1 (default), limit of hits: 250. For each AMP of UDAMP: protein name, gene name, UniProt ID, GI number, peptide sequence, length of peptide, antiviral, antibacterial or antifungal activity, Gram type in case of AMPs having antibacterial activity, method of validation, database ID from where the AMP was downloaded, and the Protein Data Bank (PDB) ID indicating the structure of the AMP was retrieved. The online source

databases of AMPs provided various identifiers for the listed AMPs. The UniProt and NCBI GenBank databases were queried to retrieve or verify the protein name, gene name, UniProt ID, and GI number. The Protein Databank ID was retrieved from the RCSB Protein Database; all other data were obtained from the online database of AMPs or a scientific article deposited in PubMed.

3.2 Examination of AMPs in proteomic datasets related to Alzheimer's disease

We have retrieved the AD datasets including human donors from the ProteomeXchange repository (http://www.proteomexchange.org/). Datasets deposited before 2019 were selected, where the sample from AD and matched non-AD controls were investigated. We omitted the datasets involving only patients with AD and searched for comparisons between AD and healthy/control data. We have retrieved datasets originating from the analysis of brain, CSF, or blood samples. We checked which of the AMPs listed in the UDAMP database were present in the ProteomeXchange datasets and evaluated the changes in their levels. Our results were based on the statistical analysis performed by the authors. AMPs whose level showed a statistically significant change in the AD samples compared to control were defined with labels "increase" and "decrease", respectively. In contrast, AMPs without any statistically significant change were labelled as "no significant change". "0" was assigned to those AMPs absent from the ProteomeXchange datasets.

3.3 Network analysis of datasets related to Alzheimer’s disease

AMPs that were present in the UDAMP database and showed a statistically significant change in the downloaded datasets from the ProteomeXchange datasets were analyzed through network analysis. We have used Cytoscape with String-DB v11 and ClueGO v2.5.7, plus CluePedia v1.5.7 plugins. Query proteins were imported in String-DB, where networks were generated at a confidence score of 0.90. Networks containing AMPs with a statistically significant change in AD compared to controls and their 50 first shell interactors were retrieved. The networks were imported to the Cytoscape and analyzed using the "Analyze Network" tool available under the "Tools" menu. Pathway analysis was performed using the ClueGO tool using default settings, and the number of gene visualization thresholds was set to the highest number, i.e. 1000 in CluePedia. The betweenness centrality and the degree of distribution were obtained for the networks. The Cytoscape network was imported to ClueGO and CluePedia. The gene function analysis was performed using Cytoscape network in CluePedia by using default settings. The used interactions for the

network analysis were: activation, binding, co-expression, and inhibition selected from the Sting-DB. Pathway analysis was also performed using DAVID, and AMPs were further submitted for gene ontology analysis. The top 10 enriched pathways were selected by investigating the gene counts of the pathways.

3.4. Network analysis of data related to oral squamous cell carcinoma and precancerous conditions

Saliva sample were collected from OSCC, OLP, OLK, and control groups (3 donor/ group, mean age 62 years) and analysed by proximity extension assay (PEA) at the Olink proteomics facility (Uppsala, Sweden) by applying the Oncology II and Inflammation panels (www.olink.com). It was checked if the detected salivary proteins were AMPs, namely, if they were present in the UDAMP database. String-DB and CluePedia v1.5.7 were used to create the networks of differentially expressed proteins and AMPs. The network generation and examination of the generated networks was done as it was described in the case of AD, with one exception: in case of the String network generation the medium stringency (0.7) was applied.

4. RESULTS

4.1 Generation of the comprehensive human antimicrobial and immunomodulatory peptide database (UDAMP Database)

We created a large unified human AMP database and named it University of Debrecen Antimicrobial and Immunomodulatory Peptide database (UDAMP). UDAMP contains data from online databases as well as scientific literature.

In the year 2019, when the database was originally made, the UDAMP had 186 AMPs. Each AMP had a unique protein and gene identifier (gene and protein name, UniProt ID, GI number), as well as the database name from which it was obtained. Every AMP also had a peptide sequence, Protein Databank ID (PDB ID), where the 3D structure was available, and the type of antimicrobial activity. In the case of the antibacterial peptides, it was indicated if they are effective against Gram positive or Gram negative bacteria. The validation cell was used to categorize antimicrobial activity as experimentally validated or predicted. Meanwhile the database was updated, currently containing 670 entries. The database was uploaded to the Figshare and can be accessed freely through the weblink https://figshare.com/articles/dataset/UDAMP_Database/20926873.

4.2 Functional examination of the human AMPs

In order to get more information on the human AMPs, a functional analysis and pathway analysis, along with gene-gene interaction network analysis was carried out. As it was expected, the extensive GO analysis revealed pathways such as antimicrobial response, immune system-related functions, and signalling. To better understand the subtle relations among AMPs, a protein-protein interaction analysis using the Sting-DB and Cytoscape was carried out. A dense cluster of proteins was observed containing defensins linked by defensin alpha 4 (DEFA4) to the cluster of proteins containing AMPs like antileukoprotinase (SLPI), bacterial permeability increasing protein (BPI), lipocalin-2 (LCN2) and cathelicidin (CAMP). The other cluster contained mainly cytokines, cystatins, APP and other AMPs. Based on the betweenness centrality and degree of distribution, proteins such as amyloid beta precursor proteins (APP), DEFA4, growth regulated alpha protein (CXCL1), C-X-C motif chemokine 10 (CXCL10), kininogen 1 (KNG1), platelet precursor basic protein (PPBP), and platelet factor 4 (PF4) were playing a central role in the network. To have a functional overview, the gene-gene interaction analysis was performed using the Cytoscape plugin CluePedia. Gene-gene interaction analysis revealed many subnetworks and proteins which were playing significant role in the human AMPs network. Some regulatory subnetworks could be observed: VIP was activating the VGF, and neuropeptide Y (NPY), NPY was found to inhibit the pro-opiomelanocortin (POMC), and activating the resistin (RETN). RETN was found to activate the C-X-C motif chemokine 12 (CXCL12), and C-C motif chemokine 4 (CCL4), this later being the downstream target for both RETN and protachykinin 1 (TAC1), which in turn activated C-X-C motif chemokine 8 (CXCL8). CXCL8 was the target for several other upstream activators: TAC1, CAMP, neurotensin (NTS), and galectin-3 (LGALS3). An inhibitory loop showed that heparin cofactor 2 (SERPIND1) inhibited both prothrombin (F2) and fibrinogen alpha chain (FGA), and F2 inhibited FGA and SERPIND1. Thymic stromal lymphopoietin (TSLP) was found to activate the C-C motif chemokine 17 (CCL17) and SLPI, which inhibited the polymeric immunoglobulin receptor (PIGR). Interferon-induced with helicase C domain 1 (IFIH1) and interferon alpha-2 (IFNA2) were mutually activating each other, while IFNA2 was activating the CXCL10, and IFIH1 was inhibiting it. Apart from the main network, there were some independent pairs found such as alpha-antitrypsin (SERPINA1) was activating hepcidin (HAMP), DEFB4A the apolipoprotein B mRNA editing enzyme catalytic subunit 3G (APOBEC3G), and guanylate-binding protein 1 (GBP1) the interferon-induced GTP-binding protein Mx1 (MX1). APP and glyceraldehyde -3-phosphate dehydrogenase (GAPDH) were reciprocally inhibiting each other. The DAVID

and ClueGO was used to analyse GO functions. The top 10 GO functions were obtained in order to examine the regulated pathways in the network of AMPs characteristic to the brain, CSF and blood. Bacterial defense, inflammation, and innate immune response were the primarily enriched GO functions in the AMPs characteristic to the brain, CSF, and blood. Some of the GO functions were sample specific. The cellular metabolic process was enriched in all three sample types, but retinal homeostasis was only found in the brain. Platelet degranulation and negative regulation of endopeptidase activity were found in the brain and CSF, while the defense response to fungi was found in the brain and the blood. Gene-gene interaction analysis was performed by using the CluePedia plugin in the Cytoscape. This analysis revealed the presence of the different regulatory circuits among the AMPs present in all sample types. The network of AMPs observed in the brain, CSF and blood shared some similarities. For example the small subnetwork containing the reciprocal inhibition between APP and GAPDH was found in all sample types. The inhibitory loop involving F2, FGA, and SERPIND1 was present in CSF and blood and was only partially present in the brain, while the VIP—VGF—NPY regulatory subnetwork was present in brain and CSF and only partially in blood.

4.3 AMPs characteristic of Alzheimer's disease in the brain, CSF, and blood

To retrieve the AMPs in Alzheimer's disease, we examined AD datasets obtained from the ProteomXchange repository. For the data analysis, 11 datasets from brain tissue, 2 from CSF, and 2 from blood were used. Identified AMPs were used for the GO analysis and network analysis. We examined the AMPs that showed a statistically significant difference between the AD and control groups. There were nine AMPs increased in the brain characteristic to the AD. The increased AMPs were APP, SERPINA1, chromogranin-A (CHGA), CLU, FGA, protein S100A9 (S100A9), protein S100A12 (S100A12), haptoglobin (HP), and TAC1. The amount of three AMPs, FAU ubiquitin-like and ribosomal protein S30 (FAU), VGF, and beta-2-microglobin (B2M), decreased in AD. In the CSF, SLPI, B2M, DEFA3, hornerin (HRNR), TAC1, protein S100A7 (S100A7), ribonuclease 6 (RNASE6), and VIP were found in increased, whereas VGF in decreased amount in AD. The amount of APP and FGA in the blood samples was found to be increased in AD. On the other hand, the amount of SERPINA1, CLU, HP, PPBP, S100A9, and zinc-alpha-2-glycoprotein (AZGP1) was decreased.

ClueGo and DAVID were used to perform functional analysis of the AMPs characteristic of AD in different sample types. The GO terms that were enriched in different

sample types were identified using functional analysis. According to DAVID, several of the GO functions enriched in the brain for AMPs increased in AD were similar to those enriched in the overall AMP network: inflammatory response, innate immune response, and signal transduction. Blood coagulation (fibrinolysis, platelet degranulation, and blood coagulation), cellular metabolic processes, endocytosis, and negative regulation of endopeptidase activity, and apoptotic activity were newly enriched functions. Majority of the enriched GO terms in the CSF were related to antimicrobial defense, immune response, inflammation, and G-protein receptor signaling pathway, and also present in the main AMP network. Proteolysis and the response to yeast was the newly enriched functions. Blood coagulation (fibrinolysis, platelet activation, and degranulation), acute phase response, extracellular matrix organization, signal transduction, cellular protein metabolic process, and negative regulation of endopeptidase activity and of cell proliferation were the enriched functions in the blood. Inflammatory and innate immune response was shared by the brain and CSF. In contrast, blood clotting, negative regulation of endopeptidase activity, signal transduction, and cellular protein metabolism were shared by the brain and the blood. Regarding the AMPs found to be decreased in AD, the RNA metabolism, translation, viral transcription, cell adhesion were newly discovered enriched functions. The GO functions characteristic to CSF were related to lipid (lipoprotein and cholesterol) metabolism, cell adhesion, osteoblast differentiation, endopeptidase activity, platelet degranulation, inflammation, and cell adhesion. Blood coagulation, signal transduction, protein secretion, endopeptidase activity regulation, cell proliferation, transcription, and extracellular matrix organization were more prevalent in the blood. Cell proliferation regulation and endopeptidase activity, the extracellular matrix organization, and the platelet degranulation functions were general in both blood and CSF. The GO analysis with ClueGO revealed additional information to DAVID about functional enrichment. Aside from the enriched GO functions in DAVID in the case of AMPs increased in AD, the amyloid precursor protein metabolic process, neuropeptide and tachykinin signaling pathway, and antioxidant activity have been found only in brain samples. Regulation of lipid localization, protein-lipid complex remodelling, amyloid beta formation, and response to wounding were newly enriched GO functions in the blood in the case of AMPs increased in AD. Along with this, the cellular response to glucagon stimulus and regulation of osteoclast differentiation were considered new enriched GO functions in the CSF. When running ClueGO, the additional functions to those identified by DAVID in the case of AMPs decreased in AD, were antigen processing and DNA damage response in the brain, regulation of amyloid fibril formation in the blood, and response to axon injury,

regulation of insulin-like growth factor receptor signaling, cellular response to cAMP, post-translational protein modification, and membrane protein proteolysis in the CSF.

Given that there were no considerable differences between the gene interaction networks of AMPs in the UDAMP Database and those reported in the brain, CSF, and blood, we were keen to study the changes in gene-gene interaction networks in the case of AMPs related with AD in the next step. We wanted to find out which proteins could be the most important in the AMP network by identifying the key hub proteins. As a result, we investigated the protein - protein interaction networks and computed the betweenness centrality and degree of distribution. In line with earlier research, we discovered APP to be the primary hub protein in the AMP network observed in the brain.

Next we generated the gene-gene interaction network of all proteins altered in AD, independent of the direction of change (increase or decrease) or the location of identification (brain, CSF, or blood). We could observe that the main sites of activations were vascular endothelial growth factor A (VEGFA), and plasminogen activator inhibitor 1 (SERPINE1). A cluster with multiple interactions among its members containing the apolipoproteins APOA1, APOA2, and APOE, as well as CLU, HP, mitogen-activated protein kinase 1 (MAPK1), and APP was present, whereas proteins implicated in blood clotting such as FGA, fibrinogen chain beta and gamma (FGB, FGC), fibronectin (FN1), alpha-2-antiplasmin (SERPINF2), and thrombospondin-1 (THBS1) were found in another cluster. VEGFA was the major hub protein, followed by epidermal growth factor (EGF), insulin-like growth factor 1 (IGF1), SERPINE1, and FN1, as proteins activated or inhibited by many other proteins.

4.4 Examination of antimicrobial peptides in oral squamous cell carcinoma and precancerous conditions

Saliva is a bodily fluid with great diagnostic potential. Besides being used for diagnosis, it can also be used for screening, further increasing the diagnostic value of this body fluid. Considering the presence of AMPs in saliva, our aim was to evaluate their role in OSCC and precancerous conditions such as OLP and OLK. A study was carried out by our group to observe the altered proteins characteristic of the OSCC and the other precancerous conditions: OLK and OLP. Saliva sample taken from 4 groups of donors (OSCC, OLP, OLK, control) were analysed by proximity extension assay. Three donors were recruited to each group and the donors were age matched. 176 out of examined 184 proteins were detected in salvia and it was checked if they were present in the UDAMP database. 15 proteins were present in the UDAMP database being AMPs. The AMPs and their 50 first shell of

interactors were used to generate gene-gene interaction networks. In order to better examine the activation and the binding events among the proteins in the network, from the composite network containing all interactions only the activation and the binding events, respectively, were extracted. The network analysis indicated activation hubs, showing proteins which were activated by many other proteins in the network. A highly interconnected cluster with multiple activations containing chemokines and chemokine receptors could be observed. Interleukin (IL) 6 has functioned as the activation hub for numerous proteins, as evidenced by the fact that it was activated by several proteins and vice versa. IL6 activated IL4, IL1beta, and IL17A, while tumor necrosis factor (TNF) activated IL6. The activations and interactions among the FAS-associated death domain protein (FADD) and members of the TNF receptor family (TNFRSF10A, TNFRSF10C, TNFRSF10D, and TNFRSF10B) showed another highly interconnected cluster. The degree of distribution analysis regarding the protein-protein interactions revealed CXCR1, CXCR2, CCR1, CCR5 and CCR7 as hub proteins in the binding network while taking into account the activation, the main hub proteins were CCL2, IL6, CXCL8, CCR5 and CCR7.

The amount of proteins that were present in the more than 80% of the samples were also examined. 37 proteins showed a statistically significant change between at least two groups. Most of the differentially expressed proteins between OLP and control group were chemokines, such as CCL28, CCL11, CXCL5, CXCL11, CX3CL1, CXCL10, MCP1/CCL2, and TNF ligand superfamily member 12 (TWEAK), melanoma- derived growth regulatory protein (MIA), midkine (MK/MDK), r-spondin-3 (RSPO3), tissue factor pathway inhibitor 2 (TFPI2), and TNF ligand superfamily member 10 (TNFSF10). Adenosine deaminase (ADA), amphiregulin (AREG), eotaxin (CCL11), CXCL11, transmembrane glycoprotein NMB (GPNMB), IL1, IL6, matrix metalloproteases MMP1 and MMP10, and toll like receptor 3 (TLR3) were found to be differentially expressed between the OSCC and control groups. Only two proteins were discovered to be changed in a statistically significant manner between the OLK and control groups, the chemokines CXCL11 and CCL11. Many chemokines such as CCL20, CX3CL1, CXCL1, CXCL9, CXCL10, CXCL11, and MCP1/CCL2, mucin-16 (CA125/MUC16), t-cell surface glycoprotein-5 (CD5), CCN family member 1 (CYR61), epidermal growth factor (EGF), endothelial cell-specific molecule 1 (ESM1), leukemia inhibitory factor receptor (LIFR), podocalxyin (PODXL), RSPO3, band 3 anion transport protein 1 (SLC4A1), TLR3, TWEAK, and WAP four-disulfide core domain protein 2 (WFDC2) were the proteins that differed between OSCC and OLP. Annexin-1 (ANXA1), cystatin-5 (CST5), eukaryotic translation initiation factor 4E-binding protein 1

(EIF4EBP1), GPNMB, IL6, MMP1, MMP10, and WFDC2 were the proteins that were altered differently between OLK and OSCC. On the other hand, only four proteins were found to be differently expressed between OLK and OLP: CYR61, MCP1/CCL2, MIA, and MK/MDK.

Proteins and AMPs with statistically significant differences between two groups were examined for protein-protein interactions in order to gain a better understanding of the OSCC and precancerous conditions. The protein-protein interaction and GO analysis of proteins showing statistically significant changes between two groups was performed using string-DB. The proteins and AMPs that were differentially expressed between the OSCC and controls were associated with functions such as inflammatory response, regulation of growth factor production, collagen catabolism, positive regulation of intracellular signaling, and cellular proliferation. In the case of OLK vs control, the CXCL11 and CCL11 were differentially expressed. These proteins were involved in biological processes like chemokine-signaling, antimicrobial humoral immune response, and lymphocyte chemotaxis. Regarding the OLP vs control groups, the enriched GO terms were related to the signalling pathways and immune response. Ten of the 14 proteins were shown to be involved in cell migration, which corresponds to histological observations of T-cell and lymphocyte infiltration in the epithelium and lamina propria. Proteins involved in positive stimulus response, stress response, and signalling may indicate the link between viral infection and OLP induction, as well as the link between autoimmune disorders and OLP development. The comparison of OSCC and OLP groups revealed, that the bulk of the proteins in the network are members of the cytokine-cytokine receptor interaction KEGG pathway. The biological processes of the leukocyte chemotaxis and inflammatory response were the enriched functions between the OSCC and OLP. The enriched functions between OSCC and OLK were immune response, cell migration, negative regulation of cell death, collagen catabolism, and the regulation of endopeptidase activity. There were no interactions found in the proteins differentially expressed between OLP and OLK. A CluePedia search was undertaken to acquire further information on the proteins that showed statistically significant changes between the groups, and the types of interactions were investigated. The relationships found with String-DB analyses were validated in the gene-gene interaction networks, which added more information to the data collected by the examination of the network using only String-DB. In the OSCC vs control, it was observed that AMP IL6 was activated by TLR3 and IL1A, while IL6 further activated the amphiregulin (AREG) and MMP1 and inhibited MMP10. In addition to the previously known information, activation of CCL20 by TNFSF12 and of

CCL2 by CYR61/CCN1 was identified in the networks of proteins with statistically significant differences between OSCC and OLP. The network between OLK and OSCC, on the other hand, revealed that ANXA1 inhibits IL6, which then activates MMP1 and inhibits MMP10.

5. DISCUSSION

The purpose of this study was to look into the potential involvement of AMPs in different pathological conditions. AMPs, as part of the innate immune system, control immune responses, while also directly reducing infections, either by killing pathogens or inhibiting their proliferation.

Our aim was to discover which previously identified AMPs by proteomics methods in the brain, blood, and CSF, could play a role in the pathogenesis of AD. To examine this subject, we first gathered all accessible human AMPs by creating the UDAMP Database. The database presently has 670 AMPs and is accessible at Figshare https://figshare.com/articles/dataset/UDAMP_Database/20926873, but at the time of the analysis, it contained 186 AMPs.

The network analysis of all human AMPs identified three clusters of interacting proteins. The most important biological roles were related to defense, innate immune system, and signal transduction. As expected, the examination of the network of human AMPs recapitulated knowledge from the scientific literature regarding their activities and roles in human beings.

The UDAMP Database listed 46 human AMPs discovered in the brain, 46 in the CSF, and 75 in the blood. The collection of AMPs found in each sample type overlapped but was not identical. In terms of network analysis, each sample type had certain distinct characteristics; nevertheless, in terms of enriched GO functions, AMPs discovered in the brain, CSF, or blood exhibited general AMP functions linked to antimicrobial defense, immune system activation, and blood coagulation.

The situation was different when it came to the proteins that showed a statistically significant alteration in AD. The levels of nine AMPs were discovered to be increased in a statistically significant manner in the brain, and two of the AMPs were also found in the blood. These two proteins (FGA and APP) may have diagnostic potential in AD from blood, and there are already several diagnostic techniques for using FGA and APP as biomarkers. According to our findings and those presented in the scientific literature, APP is the key hub protein with the greatest number of interactions in AD. In terms of CSF, only TAC1

overlapped with AMPs elevated in the brain and no overlap with AMPs increased in the blood was observed. When the AMPs that showed a statistically significant decrease in AD were analysed, four AMPs that were increased in the brain were found to be decreased in the blood (CLU, HP, S100A9, SERPINA1). A high level of CLU (or ApoJ) and HP was already linked to AD, but their higher levels were found not only in the brain but also in the blood. Given that the blood collection period in the different tests varied in relation to the beginning of AD, additional study with consistent sample collection times is likely needed to clarify the situation regarding the dynamics of the change in the amount of these proteins in the blood. B2M levels rose in the CSF while decreased in the brain. FAU levels in the brain decreased, as the level of VGF in both the brain and the CSF. Data published in peer-reviewed journals show higher B2M and FAU levels in the brain, suggesting them as viable targets for AD treatment. FAU, or 40S ribosomal protein S30, plays a role in DNA repair and function, and its levels in the AD brain appeared to be decreased according to another study. Our results on TAC1 and VGF are consistent with previously reported data: TAC1-derived neuropeptides have a neuroprotective role in AD, and VGF was found as a major driving gene/protein of AD in a mouse model. It was discovered that increasing VGF levels helped by reducing memory impairment symptoms in mice. A network study revealed that a gradual decrease in the amount of VGF, along with other proteins, was diagnostic of asymptomatic AD and AD. Aside from the AMPs stated above, there are other AMPs that either considerably increased or decreased in AD patients. CHGA has been demonstrated to play a role in activating microglia, which can emit neurotoxins and cause neuronal death. Some of the other AMPs have already been related to AD. S100A7 is a recently identified possible biomarker for AD and PPBP and AZGP1 were discovered to be downregulated in the serum of patients with AD. Our findings support their downregulation in CSF. The rise in VIP can be viewed as an attempt to slow the progression of neurodegeneration since VIP inhibits the activation of microglial cells, preventing the production of additional inflammatory cytokines. When the activities of the differentially expressed AMPs that are not connected to the innate immune system were investigated, it was revealed that the majority of the proteins had protective effects, such as the antioxidant HP, which is responsible for heme binding, and CLU, which inhibits the development of amyloid fibrils and reduces protein aggregation caused by stress. FGA, SERPINA1, and PPBP all play a role in hemostasis preventing blood loss. In terms of the network analysis of AMPs changed in AD, our findings support the role of inflammation and innate defense systems in the illness's progression. Aside from these broad roles, changes influencing protein post-translational modification may be associated to tau

hyperphosphorylation, ubiquitination, glycosylation, or citrullination of brain proteins, all characteristic to AD. Dietary excess of lipids and lipid metabolism impairment, as well as insulin intolerance along with a lack of neurotrophic factors, can all contribute to the cognitive deterioration seen in AD. Our findings add to the evidence for changes in lipoprotein metabolism and cholesterol metabolism seen in CSF samples obtained from patients with AD. We could also highlight the importance of hemostasis via network analysis. The function of blood clotting in the development of AD was investigated, and it was discovered to be damaged. Some clotting factors may damage synapses, providing new possible candidates for diagnostic and therapeutic intervention. Changes in GO functions such as APP metabolism, amyloid fibril formation, inflammation, post-translational protein modification, translation, antimicrobial defense, hemostasis, response to axonal injury, regulation of IGFR signaling, cellular response to cAMP, and others suggest that AMPs play a central role in the disease's development. Our findings regarding changed activities linked to extracellular matrix architecture observed in CSF and blood, as well as altered RNA metabolism observed in the brain and blood is in agreement with previously published scientific data. RNA-binding proteins, alterations linked to RNA metabolism have been observed in AD, and RNA splicing has been shown to be significant in the exacerbation of AD symptoms. Other studies have identified proteostasis, RNA homeostasis, immune response, neuroinflammation, synaptic transmission, vesicular transport, cell signaling, cellular metabolism, lipid homeostasis, mitochondrial function, cytoskeleton organization, and myelin-axon interactions as key players in the pathology of AD, and the majority of these functions have been demonstrated in our study as well. On the basis of composite netwrok analysis, VEGF was identified as a functional hub protein modulated by the greatest number of proteins, and it has been shown to have a role in AD by other groups. Similarly, the other hub proteins identified: IGF-1, EGF, SERPINE1, fibronectin were linked earlier to aging, neurodegeneration, and AD.

Despite their small quantity (0.5-1% of all discovered proteins), the network analysis of AMPs might indicate the crucial activities mentioned previously in AD. Our findings highlight the necessity of AMP testing and the function of the innate immune system and blood clotting in AD. At the same time, we could find further evidence indicating that AMPs play a significant role in AD. As a result, AMPs may be suitable candidates for additional mechanistic research aimed at understanding their precise function in the complicated pathophysiology of AD and these may be used as therapeutic targets in the future.

The other pathological condition where the importance of AMPs was examined during our study was related to oral cancer and precancerous lesions. OSCC, along with the precancerous conditions OLK and OLP, are responsible for a high number of pathological conditions appearing in patients. Saliva can be a sample of choice when we would like to examine patients efficiently and possibly to screen the patients with high risk for OSCC or precancerous conditions. In order to be able to perform screening, first proper biomarkers are needed. Our group applied PEA for the thorough examination of saliva and could demonstrate that the PEA technique is suitable for saliva analysis as well and it can be the technique of choice when extensive saliva analysis is required. Our aim during this study was to examine the AMPs showing differential expression between the OSCC, OLK, OLP and control groups, respectively.

The differentially expressed proteins between the OSCC and the control groups were mostly involved in biological functions having a role in carcinogenesis. The inflammatory response, regulation of growth factor production, collagen catabolism, positive regulation of intracellular signalling, and regulation of cellular proliferation are all well-known to be associated with tumor formation and progression. Between the OLK and control groups, only two proteins were seen to be differentially expressed. Both of them were chemokines, involved in many general processes like chemokine-signaling, antimicrobial humoral immune response, and lymphocyte chemotaxis. The 14 proteins differentially expressed between in OLP compared to the control group were involved in signaling pathway and immune response, being possibly related to the chronic inflammatory background observed in the case of OLP. Ten of the proteins were responsible for cell migration, being in line with the histopathological findings of T-cell and lymphocyte infiltration in the epithelium and lamina propria observed in OLP. Proteins responsible for positive response to stimulus, stress response, and signaling may reflect the association between viral infection and OLP induction, and OLP and autoimmune conditions. Further evidence highlights the inflammatory background seen in OLP. Among proteins differentially expressed between OSCC and OLP, many were involved in leukocyte chemotaxis and inflammatory response, correlating with the data observed by histopathology, namely, the infiltration of various leukocyte types in the epithelium and lamina propria. Considering the difference between OSCC and OLK the differentially expressed proteins indicate the presence of the malignant transformation. The enriched functions were related to immune response, cell migration, negative regulation of cell death, collagen catabolism, and the regulation of endopeptidase activity. The increase in MMPs expression could also be observed. The MMPs are

responsible for the catabolism of collagen in the basement membrane and the disassembly of extracellular matrix. Another important differentially expressed protein was ANXA1, which is involved in a large number of biological functions. ANXA1 and GPNMB are important in malignant transformation, being involved in procarcinogenic biological events, such as negative regulation of cell death and regulation of G1/S transition of mitotic cell cycle. It was also shown, that GPNMB is overexpressed in many cancer types, such as brain and breast cancers. The IL6, was shown earlier by our and other groups as characteristic to OSCC, being involved in inflammatory- and carcinogenesis-related biological functions. The proteins differentially expressed between the OLP and OLK groups had a role in signalling, but did not show interactions among each other. No specific common biological functions could be drawn from these unrelated proteins. This further confirmed the understanding that OLK and OLP are two unrelated conditions that separately progress to OSCC.

As far as most of these proteins belong to the AMP family, it is worth examining the AMPs in relation to oral cancers and premalignant conditions as well.

Our study indicates that AMPs constitute a functionally important class of proteins and their analysis can give valuable insights on the pathogenesis of Alzheimer's disease and oral cancer and precancerous conditions. Their utilization potential should be tested on other diseases as well, but according to our data, their analysis can provide with useful information about the disease and potential biomarkers able to help the diagnosis or the progression of the diseases.

7. NOVEL FINDINGS

- A comprehensive database containing the human AMPs was generated. The UDAMP database currently has 670 entries.
- We have successfully reutilized proteomics data deposited in publicly available databases such as ProteomeXchange (www.proeomexchange.org).
- We have examined the AMPs characteristic to AD in brain, CSF and blood. According to our data the AMPs make up 0.5–1% of all identified proteins, but their network analysis could give information on the functions related to AD, previously reported in the scientific literature.
- Our data emphasize the key role of the innate immune system and blood clotting in the development of AD and suggests the deep involvement of AMPs in AD. They can be

candidates for further mechanistic studies aiming to understand their exact role in the complex role in the pathophysiology of AD and may serve as targets for future therapies.

- We thoroughly examined the proteomics data arising from the analysis of saliva samples originating from patients with OSCC, OLK or OLP and discovered that many of the expressed proteins, such as CCL20, IL6, and CCL2, were AMPs.
- We discovered that inflammation and malignant transformation are two major pathways expressed by AMPs. Our findings shed light on the central role of IL6 in the OSCC, OLK, and OLP.
- We found no interaction between the proteins characteristic to OLP or OLK groups, indicating that these two conditions are independent of one another and progress to OSCC.

8. SUMMARY

This thesis focuses on the examination of the role of AMPs in two different pathological conditions: namely Alzheimer's disease (AD) and oral cancer along with two precancerous conditions.

Given the inflammatory nature of AD, we aimed the examination of antimicrobial and immunomodulatory peptide (AMP) family in human proteome datasets deposited in the publicly available proteomics repositories. First, a unified, complete human antimicrobial and immunomodulatory peptide database (UDAMP) was created, coupled with the examination of datasets downloaded from publicly available sources containing high-quality proteomics data derived from AD and control groups. Extensive network analysis was used to study the possible implications of AMPs in the pathogenesis of AD. AMPs made up less than 1% of all detected proteins in the brain, but their study could replicate the abnormalities previously identified in AD. Our findings emphasize the significance of the innate immune system and blood clotting in the progression of AD and can provide with new potential therapeutic targets.

Saliva is a bodily fluid with great diagnostic potential aiming to help diagnose oral malignancies. Apart from oral squamous cell carcinoma (OSCC), it is critical to examine precancerous conditions such as oral lichen planus (OLP) and leukoplakia (OLK) as well. Our aim was to assess the changes in salivary AMPs in case of patients with OSCC, OLK or OLP. Our research revealed pathways and proteins characteristic to each condition and

highlighted the distinct natures of the OLK and OLP.

Our study indicates that AMPs constitute a functionally important class of proteins and their analysis can give valuable insights on the pathogenesis of AD, OSCC, OLK and OLP. The examination of AMPs can provide with useful information about the disease and potential biomarkers able to help the diagnosis or the progression of the diseases.

9. LIST OF PUBLICATIONS PREPARED BY THE KENÉZY LIFE SCIENCE LIBRARY

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/422/2022.PL
Subject: PhD Publication List

Candidate: Ajneesh Kumar
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology

List of publications related to the dissertation

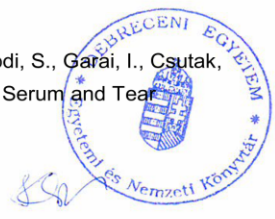
1. **Kumar, A.**, Doan-Xuan, Q. M., Kunkli, B., Csősz, É.: Construction of unified human antimicrobial and immunomodulatory peptide database and examination of antimicrobial and immunomodulatory peptides in Alzheimer's disease using network analysis of proteomics datasets.
   Front. Genet. 12, 1-15, 2021.
   DOI: http://dx.doi.org/10.3389/fgene.2021.633050
   IF: 4.772

2. Scholtz, B., Vo Minh, D., Kiss, C., Tar, I., **Kumar, A.**, Tőzsér, J., Csősz, É., Márton, I.: Examination of Oral Squamous Cell Carcinoma and Precancerous Lesions Using Proximity Extension Assay and Salivary RNA Quantification.
   Biomedicines. 8 (12), 1-16, 2020.
   DOI: http://dx.doi.org/10.3390/biomedicines8120610
   IF: 6.081

List of other publications

3. Kalló, G., **Kumar, A.**, Tőzsér, J., Csősz, É.: Chemical Barrier Proteins in Human Body Fluids.
   Biomedicines. 10 (7), 1-26, 2022.
   DOI: http://dx.doi.org/10.3390/biomedicines10071472
   IF: 4.757 (2021)

4. Nokhoijav, E., Guba, A., **Kumar, A.**, Kunkli, B., Kalló, G., Káplár, M., Somodi, S., Garai, I., Csutak, A., Tóth, N., Emri, M., Tőzsér, J., Csősz, É.: Metabolomic Analysis of Serum and Tear Samples from Patients with Obesity and Type 2 Diabetes Mellitus.
   Int. J. Mol. Sci. 23, 1-19, 2022.
   DOI: http://dx.doi.org/10.3390/ijms23094534
   IF: 6.208 (2021)



UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

5. Linkner, T. R., Ambrus, V. A., Kunkli, B., Szojka, Z., Kalló, G., Csősz, É., **Kumar, A.**, Emri, M., Tőzsér, J., Mahdi, M.: Cellular Proteo-Transcriptomic Changes in the Immediate Early-Phase of Lentiviral Transduction.
Microorganisms. 9 (11), 1-21, 2021.
DOI: http://dx.doi.org/10.3390/microorganisms9112207
IF: 4.926

Total IF of journals (all publications): 26,744
Total IF of journals (publications related to the dissertation): 10,853

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

08 September, 2022

DEBRECENI EGYETEM * Egyetemi és Nemzeti Könyvtár *

Oral presentations:

Ajneesh Kumar: *Examination of antimicrobial and immunomodulatory peptides in Alzheimer's disease using network analysis of proteomics datasets and the DEAMP Database,* Winter Symposium of the Molecular and Cellular Immune Biology Doctoral School, Debrecen, Hungary, Jan 9 – 10, 2020.

Ajneesh Kumar: *Examination of antimicrobial and immunomodulatory peptides in Alzheimer's disease using network analysis of proteomics datasets and the DEAMP Database,* Winter Symposium of the Molecular and Cellular Immune Biology Doctoral School, Debrecen, Hungary, Jan 7 – 8, 2021.

Ajneesh Kumar: *Construction of Unified Human Antimicrobial and Immunomodulatory Peptide Database and Examination of Antimicrobial and Immunomodulatory Peptides in Alzheimer's Disease Using Network Analysis of Proteomics Datasets, "*Learn from each other", meeting of the Proteomics Division of the Hungarian Biochemical Society, online, May 19, 2021

Ajneesh Kumar: *Network analysis of atheroma and complicated lesions in human atherosclerosis,* Winter Symposium of the Molecular and Cellular Immune Biology Doctoral School, Debrecen, Hungary, Jan 6 – 7, 2022.

Poster Presentations:

Ajneesh Kumar, Eva Csosz: *In-Silico Construction of Homo Sapiens Antimicrobial and Immunomodulatory Peptide (AMP) Database,* Network Tools and Applications in Biology NETTAB/BCC2019, Salerno, Italy, Nov 11 – 14, 2019.

Ajneesh Kumar, Eva Csosz: *In-Silico Construction of Homo Sapiens Antimicrobial and Immunomodulatory Peptide (AMP) Database DEAMP,* Human Proteome Organization HUPO Connect 2020 conference, online, Oct 19 – 22, 2020.

Ajneesh Kumar, Gergő Kalló, László Potor, Zoltán Hendrik, Gábor Méhes, Csaba Tóth, Péter Gergely, József Tőzsér, József Balla, Éva Csősz: *Proteomics analysis followed by network analysis reveals proteins and functions characteristic to atheroma and complicated*

lesions in human atherosclerosis, PROTEOMIC FORUM 2022 XIV. Annual Congress of the European Proteomics, Leipzig, Germany, Apr 3 – 7, 2022.

Ajneesh Kumar, Gergő Kalló, László Potor, Zoltán Hendrik, Gábor Méhes, Csaba Tóth, Péter Gergely, László Prókai, József Tőzsér, József Balla, Éva Csősz: *Network analysis of proteomics data in human atherosclerosis,* Annual Meeting of the Hungarian Biochemical Society, Pecs, Hungary, Aug 25 – 27, 2022.

10. KEYWORDS

APP, Alzheimer's disease, UDAMP database, network analysis, oral cancer, saliva, proximity extension assay.

11. ACKNOWLEDGMENT

Among all those who contributed to my Ph.D. thesis, I would like to thank my supervisor, Dr. Éva Csősz, for her continual support, scholarly conversations, and the chances she provided me during my studies.

I am thankful to Prof. József Tőzsér, head of the Department of Biochemistry and Molecular Biology, for the opportunity to conduct research in a well-equipped environment.

I am grateful to my previous and recent colleagues at the Proteomics Core Facility, notably Dr. Gergő Kalló, Erdenetsetseg Nokhoijav, Andrea Guba, Azhar Orynbek, Renáta Kovács, Julianna Kökényesiné Csáki, and VoMinh Doan.

I also want to thank the Tempus Public Foundation for giving me the Stipendium Hungaricum scholarship, which allowed me to study in Hungary.

Last but not least, I want to thank my wife, parents, and grandmother for all their aid, love, and encouragement throughout my life.

This work was supported by the Stipendium Hungaricum Scholarship Program, by the GINOP-2.3.3-15-2016-00020, and the Hungarian Scientific Research Fund (OTKA) K105034.